UNIVERSITÉ DE TOURS

École Doctorale MIPTIS

Laboratoire d'Informatique Fondamentale et Appliquée de Tours LIFAT (EA 6300)

THÈSE présentée par :

Frédérick BISONE

soutenue le : 31 mai 2021

pour obtenir le grade de : Docteur de l'Université de Tours

Discipline/ Spécialité : Informatique

Extraction de trajectoires sémantiques à partir de données multi-capteurs : application à des véhicules de secours

THÈSE DIRIGÉE PAR :

DEVOGELE Thomas	Professeur, Université de Tours

RAPPORTEURS :

PERRET Julien	Directeur de Recherche, IGN Saint Mandé
JOSSELIN Didier	Directeur de Recherche, UMR ESPACE, Avignon

JURY :

PERRET Julien	Directeur de Recherche, IGN Saint Mandé	Rapporteur
JOSSELIN Didier	Directeur de Recherche, UMR ESPACE, Avignon	Rapporteur
RENAUDIN Valérie	Directeur de Recherche, IFSTTAR Nantes	Examinatrice
ZEITOUNI Karine	Professeur, Université de Versailles Saint-Quentin	Examinatrice
DE RUNZ Cyril	Maître de conférences HDR, Université de Tours	Examinateur
PERALTA Veronika	Maîtresse de conférences, Université de Tours	Co-encadrante
ETIENNE Laurent	Enseignant-Chercheur, ISEN, Brest	Co-encadrant
DEVOGELE Thomas	Professeur, Université de Tours	Directeur

Remerciements

A la mémoire de mon frère,

Dans ces quelques lignes, je vais tenter de remercier les gens, qui, de prés ou de loin, ont contribué à me faire avancer, de différentes manières, chacun a sa façon. Pardonnez moi par avance pour ceux que j'aurais oublié, ou pas assez remercié dans ces quelques lignes. Merci à mes encadrants pour m'avoir fait confiance depuis le début de cette aventure, de la découverte et la recherche du sujet, en particulier le travail de Laurent, jusqu'à la finalisation du manuscrit. Un grand merci spécialement à Cyril, pour son soutien et son aide extrêmement précieuse pendant toute la période de rédaction. Un grand merci à mon directeur de thèse, Thomas, qui a su tout au long de ma thèse trouver les mots pour toujours me tirer vers le haut, ainsi qu'à Verónika pour son implication tant au niveau scientifique que moral. Votre implication et votre disponibilité m'a été très précieuse.

Je tiens également à remercier les membres du jury pour avoir accepté la relecture de mon manuscrit. Pour leurs remarques et commentaires pertinents, ainsi que leur présence à ma soutenance, suivis de questions réponses extrêmement enrichissantes. Un grand merci au groupe Gruau pour m'avoir donné l'opportunité de travailler sur ce sujet et particulièrement à Zouheir et Basher pour votre implication dans ma thèse jusqu'à la soutenance de celle-ci. Je tiens à remercier chaleureusement les pompiers, et plus particulièrement les pompiers du CSP Sud-Agglo pour leur accueil chaleureux et bienveillant pendant presque un mois. En particulier, merci au commandant Philippe Breton, pour son implication personnelle dans le projet des ambulances connectées. Ton implication dans ce projet m'a permis d'acquérir à la fois beaucoup d'idées et facilité grandement ces années de recherche. Un grand merci aux membres de l'équipe BDTLN, particulièrement à Arnaud Giacometti, qui m'a fait confiance lors de mon arrivée à l'université en me donnant d'abord un stage puis finalement un travail pour me permettre de continuer mes études, ainsi que toutes les personnes travaillant au département informatique de Blois. Merci pour ces moments passés pendant la pause café ou au détour d'un bureau ainsi que ces échanges enrichissants.

Merci à tous mes amis, qui, de prés ou de loin ont été la pour moi, me soutenir et m'encourager. Plus particulièrement, merci à Mahfoud et Adnan, je rejoins enfin notre groupe fermé de "docteurs", et c'est un honneur et un soulagement immense.

Merci à Aneeksha, Christine et Catherine, pour ces moments, ces cafés, ces encouragements durant ces si longues années. Et je n'oublie pas Clément, merci à toi pour tout, ces discutions où l'on avait tendance à refaire le monde, ces petits moments qui reboostaient le moral. Un énorme merci à mes proches, particulièrement mon père, ma soeur Virginie et ma chérie Alexandra, votre fierté à mon égard ainsi que votre soutien sont un moteur perpétuel qui me permet d'avancer.

Merci à toi, Jacko, pour tout ce que tu étais, et finalement, tout ce qui m'a fait moi. Dieu sait la fierté immense que tu aurais ressenti de m'entendre durant 45 minutes présenter mes travaux de recherche, répondre aux questions du jury puis enfin d'entendre la présidente du jury m'accorder le titre de docteur. Je sais cette fierté, je l'ai vue également à travers les yeux de notre soeur. Du plus profond de mon coeur, j'espère que tu es mieux là haut. Tu resteras à jamais dans nos coeurs et nos pensées.

Résumé

Les positions issues de capteurs GPS sont régulièrement employées pour analyser des déplacements d'objets mobiles. L'emploi exclusif des positions GPS s'avère insuffisant afin de détecter finement les comportements d'objets mobiles pour des raisons d'imprécisions et de perte d'informations mais aussi du fait de la richesse sémantique de ces activités. Les objets mobiles sont maintenant équipés de plus en plus de capteurs complémentaires (génériques ou spécifiques à des types de véhicules). Cette thèse introduit un processus généralisable de montée en sémantique forte de trajectoires obtenues à partir de données brutes issues de différents capteurs. L'instanciation de ce processus fournit un outil de fouille de trajectoires, s'appuyant sur l'analyse de qualité des données acquises, et permettant d'identifier les activités des véhicules lors de déplacements ou d'arrêts. Ce processus a été mis en oeuvre pour l'analyse des déplacements d'ambulances connectées des sapeurs-pompiers d'Indre et Loire. Dans ce contexte, nous avons exploité des méthodes à base de règles expert définies à partir de l'analyse métier.

Mots clés : objet mobile, fouille de trajectoires, trajectoires sémantiques, capteurs GPS, analyse de mobilité, véhicule de secours

Abstract

Positions from GPS sensors are regularly used to analyze the movements of mobile objects. The exclusive use of GPS positions is currently insufficient to detect the behavior of mobile objects because of inaccuracies and loss of information but also because of the semantic richness of these activities. Mobile objects are now equipped with more and more sensors (generic or specific to vehicle types). This thesis introduces a generalizable process of strong semantic enhancement of trajectories obtained from raw data from different sensors. The instantiation of this process provides a tool for semantic trajectory mining, based on the quality analysis of the acquired data, and allowing to identify the activities of vehicles during travel or stops. This process has been implemented for the analysis of the movements of connected ambulances for the Indre et Loire fire department in France. In this context, we used methods based on expert rules defined from the analysis of an ambulance's working cycle.

Keywords : moving object, trajectory data mining, semantic trajectory, GPS sensor, analysis, connected ambulance

Table des matières

Introduction **23**

1.1 Contexte de la thèse 23
1.1.1 Contexte scientifique 24
1.1.2 Contexte industriel 25
1.2 Motivations 27
1.3 Verrous technologiques et scientifiques 28
1.4 Démarche scientifique 31
1.5 Plan 32

2 État de l'art **33**

2.1 Modélisation de trajectoires 33
2.1.1 Modèle issue de la Time geography 34
2.1.2 Trajectoires brutes 36
2.1.3 Modèles de trajectoires enrichies 38
2.2 Processus de segmentation 42
2.2.1 Méthode utilisant des informations géographiques externes . . 43
2.2.2 Méthode utilisant l'extraction de caractéristiques issues des trajectoires 44
2.2.3 Méthode utilisant le clustering basé sur la densité spatiale et temporelle 45
2.3 Montée en sémantique 47
2.3.1 Enrichissement via propriétés 48
2.3.2 Enrichissement contextuel 49
2.3.3 Enrichissement d'activités 50
2.4 Analyse statistique et fouille de trajectoires 52
2.4.1 Analyse statistique 52
2.4.1.1 Analyse statistique classique 52

2.4.1.2 Analyse statistique orientée graphe 53
2.4.2 Fouille de données 54
2.4.2.1 Classification 55
2.4.2.2 Clustering 56
2.4.2.3 Extraction de motifs 57
2.4.3 Discussion 58
2.5 Conclusion 59

3 Principes et modèle générique de l'extraction de trajectoires sémantiques à partir de données multi-capteurs 61
3.1 Introduction 61
3.2 Trajectoire brute et trajectoire brute étendue 63
3.2.1 Trajectoire brute 63
3.2.2 Trajectoire brute étendue 65
3.3 Trajectoire sémantique 66
3.3.1 Sous-trajectoire 67
3.3.2 Trajectoire sémantique faible 69
3.3.3 Trajectoire sémantique forte 73
3.4 Processus opérationnel 75
3.4.1 Pré-traitement : extraction des trajectoires brutes étendues et de la liste des POI 75
3.4.1.1 Extraction des POI 75
3.4.1.2 Extraction des trajectoires brutes étendues 76
3.4.1.3 Identification et étiquetage des données aberrantes . 77
3.4.2 Montée en sémantique faible 77
3.4.3 Montée en sémantique forte 78
3.5 Discussion 80
3.6 Conclusion 81

4 Cas d'étude et phase de pré-traitement 83
4.1 Introduction 83
4.2 Déplacements des ambulances 84
4.3 Récolte des données d'étude 88
4.3.1 Système d'acquisition 88
4.3.1.1 Présentation 89
4.3.1.2 Limites 89

4.3.2 Capteurs disponibles 92
4.3.3 Acquisition des données 94
4.4 Pré-traitement des données d'étude 95
4.4.1 Correction des étiquettes temporelles 96
4.4.2 Identification des positions GPS problématiques 97
4.4.2.1 Détection des données GPS aberrantes 97
4.4.2.2 Résultats 99
4.4.3 Extraction des lieux d'intérêt 104
4.5 Conclusion 106

5 Extraction des trajectoires sémantiques fortes à l'aide de données GPS et frein à main 109
5.1 Introduction 109
5.2 Construction des trajectoires brutes étendues 110
5.2.1 Choix des capteurs 110
5.2.2 Trajectoires brutes étendues 111
5.2.3 Analyse des éléments de trajectoires acquis 111
5.3 Montée en sémantique faible des trajectoires brutes étendues 111
5.3.1 Segmentation en sous-trajectoires 112
5.3.1.1 Problèmes rencontrés 112
5.3.1.2 Méthode de détection basée sur le frein à main . . . 112
5.3.1.3 Discussion par rapport à l'état de l'art 113
5.3.1.4 Analyse des sous-trajectoires identifiées 115
5.3.2 Extraction des STOP et des MOVE 117
5.3.2.1 Méthode proposée pour la fonction arret 117
5.3.2.2 Choix des seuils pour le cas d'étude 121
5.3.2.3 STOP et MOVE détectés 121
5.3.3 Appariement STOP - POI 123
5.3.3.1 Méthode 123
5.3.3.2 Appariements obtenus 126
5.4 Montée en sémantique forte 128
5.4.1 Construction de la liste des activités 128
5.4.1.1 Liste des STOPActivityType 130
5.4.1.2 Liste des MOVEActivityType 131
5.4.2 Association des sous-trajectoires sémantiques faibles et activités 131
5.4.2.1 Association des STOP 132

5.4.2.2 Association des MOVE . . . 132
5.4.3 Affectation des activités . . . 135
5.4.3.1 Affectation des activités aux STOP . . . 135
5.4.3.2 Affectation des activités aux MOVE . . . 138
5.5 Analyse des trajectoires sémantiques obtenues . . . 141
5.5.1 Identification des arrêts fréquents . . . 141
5.5.2 Analyse des tournées . . . 143
5.6 Discussion . . . 146
5.6.1 Qualité des trajectoires obtenues . . . 146
5.6.2 Limites du processus . . . 147
5.6.2.1 Détection des STOP . . . 147
5.6.2.2 Labellisation des activités . . . 148
5.6.2.3 Segmentation . . . 148
5.7 Conclusion . . . 149

6 Extraction des trajectoires sémantiques fortes à l'aide de données issues des IOT 151
6.1 Introduction . . . 151
6.2 Construction des trajectoires brutes étendues . . . 153
6.2.1 Capteurs candidats . . . 153
6.2.2 Choix des capteurs . . . 153
6.2.3 Trajectoires brutes étendues . . . 155
6.2.4 Analyse des éléments de trajectoires obtenus . . . 155
6.3 Montée en sémantique faible des trajectoires des trajectoires brutes étendues . . . 156
6.3.1 Segmentation en sous-trajectoires : choix de la méthode . . . 157
6.3.1.1 Méthode de segmentation . . . 157
6.3.1.2 Analyse des sous-trajectoires identifiées . . . 159
6.3.2 Extraction des STOP et MOVE . . . 160
6.3.2.1 Choix de la fonction arrêt . . . 160
6.3.2.2 STOP et MOVE détectés . . . 161
6.3.3 Appariement des STOP et des lieux d'intérêts . . . 163
6.4 Montée en sémantique forte . . . 164
6.4.1 Construction de la liste d'activités . . . 164
6.4.1.1 Liste des STOPActivityType . . . 164
6.4.1.2 Liste des MOVEActivityType . . . 165

6.4.2 Association des sous-trajectoires sémantiques faibles et activités 167
6.4.2.1 Association des STOP . . . 168
6.4.2.2 Identification et traitement des reroutages . . . 168
6.4.2.3 Association des MOVE . . . 170
6.4.3 Affectation des activités . . . 173
6.4.3.1 Affectation des activités aux STOP . . . 173
6.4.3.2 Traitement des reroutages durant un MOVE . . . 176
6.4.3.3 Affectation des activités aux MOVE . . . 177
6.5 Analyse des trajectoires sémantiques obtenues . . . 180
6.5.1 Identification des arrêts fréquents . . . 180
6.5.2 Analyse des tournées . . . 182
6.5.3 Vers un entrepôt de données multi-dimensionnel . . . 183
6.6 Discussion . . . 184
6.6.1 Segmentation . . . 185
6.6.2 Appariement . . . 185
6.6.3 Utilisation de règles métiers . . . 185
6.7 Conclusion . . . 186

7 Conclusion générale **189**

Conclusion **189**
7.1 Rappel des contributions . . . 189
7.2 Travaux futurs . . . 190

Annexes **205**

A Capteurs **205**
A.1 capteurs . . . 205

Liste des tableaux

1.1 Liste des différents verrous levés dans la thèse 30

4.1 Symbologie des STOP et MOVE métiers 84
4.2 Volume des éléments étiquetés comme aberrant selon leurs types et véhicule de provenance 103
4.3 Volume des éléments normaux selon leur possible répétition 104

5.1 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires extraites 116
5.2 Répartition des données aberrantes selon les types de sous-trajectoires 118
5.3 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires STOP ou MOVE extraites . . . 123
5.4 Répartition des données outliers selon la sémantique faible des sous-trajectoires 124
5.5 Répartition des STOP selon la qualité de leur positionnement et des types de POI associés 130
5.6 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des STOP selon l'activité 136
5.7 Répartition des données aberrantes selon les types d'activités des STOP137
5.8 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des MOVE selon l'activité 138
5.9 Répartition des données aberrantes selon les types d'activités des MOVE140
5.10 Répartition des STOP en fonction de leur cluster et de l'activité associée144
5.11 Nombre de tournées contenant au moins une fois les différents types de STOP 144
5.12 Nombre de tournées contenant au moins une fois les différents types de MOVE 144

6.1 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires extraites 159

6.2 Nombre de données où les gyrophares sont activés suivant les arrêts ou déplacements potentiels 160
6.3 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires STOP ou MOVE extraites . . . 162
6.4 Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des STOP selon l'activité 176
6.5 Nombre et durées minimum, moyenne, médiane et maximum, en (minutes :secondes), des STOP de type reroutage selon l'activité 178
6.6 Nombre et durées minimum, moyenne, médiane et maximum, en (minutes :secondes), des MOVE (Hors reroutage) selon l'activité 178
6.7 Répartition des STOP en fonction de leur cluster et de l'activité associée 181
6.8 Motifs des tournées les plus fréquents et durées calculées, en (minutes :secondes) 182

Table des figures

1.1 Aperçu du processus général du traitement des données 27

2.1 Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813 34
2.2 Exemple de représentation de trajectoires à l'aide d'un space-time cube. 35
2.3 Impact de la position des satellites sur la qualité de la mesure de positionnement 37
2.4 Trajectoire sémantique d'un touriste issue de [Parent et al., 2013]. . . 40
2.5 Le modèle conceptuel CONSTAnT issu de [Bogorny et al., 2014] . . . 40
2.6 Exemple d'instance d'un touriste visitant Rome à l'aide de CONSTAnT. 42
2.7 Exemple d'application de SMoT sur une trajectoire issue de [Alvares et al., 2007] 44
2.8 Représentation du fonctionnement typique de DBSCAN. 46
2.9 Différents types de montée en sémantique en fonction des données externes utilisées. 48
2.10 Exemple des motifs principaux (daily mobility pattern) et de leur fréquence d'observation en fonction du jeu de données étudié, présenté dans [Schneider et al., 2013] 54
2.11 Processus générique de fouille de trajectoires 55
2.12 Schéma du processus de traitement des données brutes jusqu'à l'analyse de trajectoires sémantiques 59

3.1 Schéma de la montée en sémantique forte 62
3.2 Illustration du principe de montée en sémantique forte : Trajectoire brute étendue (a), Extraction des lieux d'arrêts(b), Étiquetage des STOP et des MOVE (c) 62
3.3 Représentation de la trajectoire brute de l'exemple 1 64
3.4 Modélisation d'une trajectoire brute 64
3.5 Représentation de la trajectoire brute étendue de l'exemple 2 66
3.6 Modélisation d'une trajectoire brute étendue 67

3.7 Modélisation d'une trajectoire brute étendue avec des sous-trajectoires 68
3.8 Représentation de la trajectoire sémantique faible de l'exemple 3 . . 70
3.9 Exemple de représentation d'un STOP 71
3.10 Extrait du diagramme de classe d'une trajectoire sémantique faible . 72
3.11 Représentation de la trajectoire sémantique forte de l'exemple 4 . . . 73
3.12 Diagramme de classes UML des trajectoires sémantiques fortes des véhicules 74
3.13 Schéma étendu de la montée en sémantique forte 75
3.14 Processus de pré-traitement 76
3.15 Processus de montée en sémantique faible 77
3.16 Diagramme d'activités de la montée en sémantique forte à partir d'une trajectoire sémantique faible. 78
3.17 Diagramme d'activités du processus global de la montée en sémantique forte à partir des données d'objets mobiles connectés 79

4.1 Tournée typique 1 d'une ambulance pompier 85
4.2 Tournée typique 2 d'une ambulance pompier avec prise en charge à l'hôpital 86
4.3 Tournées typiques 3 et 4 d'une ambulance pompier avec un arrêt logistique avant le retour à la caserne 87
4.4 Tournée fictive avec reroutage et arrêt logistique 87
4.5 Représentation cartographique d'une tournée fictive avec reroutage et arrêt logistique 87
4.6 Exemple de l'effet pelote résultant de données acquises par un capteur GPS à l'intérieur d'un bâtiment. 90
4.7 Illustration de l'effet de rebond entre des bâtiments 90
4.8 Illustration de l'effet de crénelage sur les données émises par le système 91
4.9 Représentation géographiques des données collectées 95
4.10 Extrait du format des données collectées 95
4.11 Extrait d'un fichier issu d'une ambulance montrant un problème sur l'étiquetage temporel 96
4.12 Répartition géographique des données répétées pour v1 100
4.13 Répartition des longueurs des périodes de répétitions 101
4.14 Répartition des longueurs des périodes de répétitions inférieures à 10 101
4.15 Répartition géographique des données caduques pour V1 102
4.16 Répartition des vitesses calculées inter-éléments de trajectoire supérieures à 5km/h 102

4.17 Répartition des données aberrantes au regard des données normales pour les deux ambulances 103
4.18 Répartition des données normales au regard de leurs possibles répétitions pour les deux ambulances 104
4.19 Positions des POI issus de l'extraction à partir d'OpenStreetMap . . 106

5.1 Segments détectés comme entrées / sorties de la caserne des pompiers 114
5.2 Représentation de l'ensemble des données collectées dans la zone . . . 115
5.3 Représentation des potentiels arrêts sur une carte OSM 116
5.4 Représentation de la durée en secondes des déplacements des véhicules d'une durée inférieure à 30s. 122
5.5 Appariement des STOP avec des POI 125
5.6 Représentation des STOP appareillés avec la caserne des sapeurs-pompiers de Tours centre 127
5.7 Représentation des STOP appareillés avec l'hôpital Trousseau 128
5.8 Représentation des STOP appareillés avec une station-service 129
5.9 Représentation des MOVE liés à une activité "Retour à la caserne" partant d'un hôpital 139
5.10 Cluster regroupant des STOP près de la caserne de pompiers 142
5.11 Cluster regroupant des STOP près d'une station-service 143
5.12 Clusters regroupant des STOP près de l'hôpital Trousseau 145

6.1 Répartition du pourcentage de la durée d'allumages des gyrophares pendant des STOP 173
6.2 Répartition du pourcentage de la durée d'allumages des gyrophares pendant des MOVE 174
6.3 Répartition de la durée avant le premier allumage des gyrophares lors d'un MOVE 177
6.4 Schéma en étoile de la création de l'entrepôt de données 184

7.1 Aperçu des contributions et des futurs travaux 190

Introduction générale

1.1 Contexte de la thèse

Aujourd'hui, d'énormes volumes de données sur notre vie quotidienne sont recueillis provenant de multiples sources et sous de nombreux formats. En 2018, on estime à environ 2,5 quintillions d'octets la production de données quotidienne dans le monde [1].

Parmi les différentes sources de données, on distingue une catégorie de données générée "directement" par les utilisateurs de différents systèmes d'informations.

D'autre part, les données générées automatiquement par différents objets. Ces objets sont inter-connectés via différents types de réseaux (internet, intranet, ...). L'interconnexion entre ces réseaux et ces objets, lieux ou environnements physiques est appelé IoT, pour Internet of Thing. De la brosse à dents connectée, au véhicule autonome, en passant par les compteurs (gaz, électricité, eau courante...) ou même les villes connectées (feux tricolores, conteneur de recyclages, ...), ces différents équipements produisent une quantité de données qui ne cesse de croître. Selon une étude de l'agence GSMAIntelligence [2], il y avait 8 milliards d'objets connectés en 2018. Les spécialistes de cette agence en prévoient près de 31 milliards en 2025.

En 2001, un rapport de recherche [Laney, 2001] du META Group (devenu Gartner) définit les principaux enjeux inhérents à la croissance des données. Ces enjeux, aussi appelés dimensions, sont le **Volume**, la **Vélocité** et la **Variété**.

Le Volume est une dimension relative du Big Data, en effet, les données produites hier peuvent être exploitées sur les ordinateurs personnels de demain. Cependant, la croissance de la production de données reste en expansion permanente et constitue un défi majeur dans le traitement de celles-ci.

La Vélocité représente la fréquence à laquelle les données sont à la fois générées, capturées, partagées et mises à jour mais également la nécessité de traiter et analyser ces flux d'informations avant la production d'un nouveau cycle de données.

1. https://www.domo.com/learn/data-never-sleeps-5?aid=ogsm072517_1&sf100871281=1

2. https://data.gsmaintelligence.com/research/research/research-2019/iot-connections-forecast-the-rise-of-enterprise

La dernière dimension, la Variété, représente l'hétérogénéité de ces données. Les sources de données provenant du monde entier et sous des formes variées. Ainsi, par exemple, une brosse à dents d'un modèle A envoie une trame de données sous un certain format et à une certaine fréquence qui ne seront pas forcément identiques avec ceux d'une autre brosse à dents d'un modèle B. Également, lors de l'utilisation de données externes (réseaux sociaux, données publiques, ...) le format et la disponibilité de celles-ci peuvent évoluer au fil du temps. Par exemple, un tweet peut être supprimé, des données publiques peuvent être enrichies, changées de format ou simplement être renommées. Le challenge est donc de traiter et de structurer ces données afin de permettre leur exploitation.

Au fil du temps, le modèle des "3V" à ensuite été étendu pour y introduire les dimensions de Véracité ainsi que de Valeur.

La Véracité fait référence à la qualité des données collectées, telles que les données aberrantes ou manquantes (bien que ce problème puisse être amoindri par la masse de données collectées), mais aussi la confiance qui est accordée à ces données. Cette dernière dépend d'une part du mode de recueil des données. D'autre part, les données peuvent aussi être volontairement faussées. Par exemple, lors de questionnaires obligatoires en ligne, un utilisateur peut rentrer des informations erronées. La Véracité fait aussi référence aux problèmes éthiques liés à l'utilisation de celles-ci [Mittelstadt and Floridi, 2016].

La valeur fait référence à la valeur ajoutée de ces données une fois analysées et synthétisées. En effet ces analyses n'ont d'intérêt que si elles apportent une réelle valeur en terme scientifique, commercial ou marketing.

1.1.1 Contexte scientifique

Classiquement, l'utilisation d'un seul capteur s'avère suffisante pour mesurer l'état d'une propriété d'un objet réel (allumé ou éteint, température, humidité...). Cependant, la compréhension fine d'un environnement complexe nécessite l'utilisation de plusieurs capteurs. Par exemple, pour savoir si un lieu d'habitation est agréable, il est nécessaire de combiner les valeurs provenant des capteurs d'humidité, de température et de luminosité.

Un des défis majeurs actuels est de gérer les grandes quantités de données issues de capteurs variés. En effet, ces volumes empêchent les outils utilisés couramment pour capturer, gérer et traiter les données, d'apporter des réponses en un temps "acceptable" et nécessitent de nouvelles méthodes et de nouveaux algorithmes ([Rajaraman and Ullman, 2011]) afin de passer à l'échelle. Ces nouveaux algorithmes doivent prendre en compte des données générées par plusieurs capteurs retraçant les activités d'un même objet. Ces données sont complémentaires mais ne sont pas facilement intégrables du fait que les capteurs n'ont généralement ni les mêmes précisions ni les mêmes fréquences.

Le deuxième défi est l'exploitation et l'analyse de ces données. En science des données, de nombreuses méthodes automatiques définies dans le cadre de la fouille de données permettent d'extraire les connaissances à partir des données collectées [Chen et al., 1996]. Elles reposent sur l'utilisation des statistiques et de l'intelligence artificielle. Leurs finalités sont multiples, par exemple :

— identifier des similitudes ou des comportements analogues entre des individus

— regrouper les objets similaires entre eux

— définir des comportements prototypiques synthétisant un ensemble d'objets similaires,

— détecter des objets aberrants,

— prédire des états futurs.

Dans le domaine des objets mobiles, l'analyse de ces volumes de données permet de mieux comprendre le comportement des objets étudiés [Renso et al., 2013]. Un biologiste analyse les flux de données captés pour mieux comprendre la migration des animaux. Un logisticien étudie les trajectoires des véhicules de sa flotte pour optimiser leur utilisation. Dans ce domaine, les méthodes de fouille de données peuvent être utilisées mais doivent être fortement adaptées. En effet, il est nécessaire de prendre en compte l'évolution spatio-temporelle des objets mobiles étudiés.

1.1.2 Contexte industriel

Dans le domaine de l'industrie automobile, les véhicules sont maintenant équipés de systèmes intelligents s'appuyant sur de nombreux capteurs. Ces capteurs dialoguent à l'aide d'un bus de communication : CAN (Controller Area Network). Les données issues de ces capteurs sont de plus en plus précises et transmises à des fréquences toujours plus élevées. L'analyse en temps réel de ces données à l'aide de calculateurs permet d'optimiser le système et de détecter certains dysfonctionnements. La sauvegarde des flux de données issus de ces capteurs produit des masses de données encore non exploitées. En ce qui concerne les véhicules professionnels (secours, sécurité, collecte/livraison, travaux,...), le nombre de capteurs est encore plus important pour recueillir les informations spécifiques à ces véhicules (gyrophares, chargement, bras motorisés, réfrigération...).

L'entreprise Gruau est un acteur incontournable de l'aménagement et de la transformation de véhicules utilitaires sur mesure. Petit-Picot est une des filiales du groupe Gruau basée à Joué-les Tours dans le département de l'Indre et Loire. Cette filiale est spécialisée dans l'aménagement de véhicules de secours et travaille en étroite collaboration avec le Service Départemental d'Incendie et de Secours d'Indre et Loire (SDIS37). Le SDIS37 s'est récemment équipé d'une flotte de 5 véhicules de secours et d'assistance aux victimes (VSAV) connectés qui, à terme, devront transmettre en temps réel différents flux de données.

Ces flux sont actuellement stockés dans des bases de données de trajectoires. Outre les données produites par le châssis du véhicule, Petit-Picot équipe ses ambulances de nombreux capteurs et systèmes auxiliaires spécifiques aux besoins des services de secours (système de communication, GPS, centrale inertielle, avertisseurs sonores et lumineux, matériel médical, brancard, régulation de température, éclairage...). La collecte et la sauvegarde conjointe des données provenant des véhicules et des systèmes auxiliaires ouvrent des perspectives de recherche en termes de fouille de trajectoires.

La fouille de données classique a pour objectif l'extraction d'un savoir issu de masses de données. Dans notre cas d'étude, le résultat recherché est un modèle spatio-temporel décrivant l'usage spécifique des véhicules de secours ainsi que celui des infrastructures routières empruntées. La création d'un tel modèle permettra l'extraction de connaissances issues de ces véhicules de secours. Ces connaissances essentielles pour le groupe Gruau sont indispensables à la définition du concept d'ambulance intelligente.

Ce concept d'ambulance intelligente permet de proposer de nouveaux services aux clients du groupe.

— Un outil d'analyse et de gestion de flotte spécifique aux véhicules de secours. Cet outil permettra aux gestionnaires de mieux comprendre et appréhender l'usage professionnel de leurs véhicules ainsi que d'optimiser leur utilisation.

— Un outil permettant la déclaration automatique des actions effectuées par les pompiers secouristes (départ, arrivée sur les lieux, arrivée à l'hôpital, ...). Ces déclarations, actuellement manuelles, sont obligatoires. L'automatisation de ce processus permettra aux équipages de se focaliser pleinement sur leurs interventions.

— Les systèmes de navigation GPS classiques ne sont pas adaptés aux véhicules d'urgence (non prise en compte des vitesses, rue étroite ou une ambulance ne peut pas dépasser les autres véhicules, . . .). L'utilisation d'ambulances connectées permettra la mise en place de nouveaux systèmes de navigation adaptés aux métiers de l'urgence, et des conditions de transport des patients.

— Les véhicules transformés subissent des contraintes beaucoup plus fortes que lors d'une utilisation classique. C'est pourquoi les recommandations constructeurs pour les entretiens périodiques ne sont pas adaptées. L'exploitation des données d'utilisation des véhicules connectés permettra de définir de nouvelles recommandations d'entretiens suivant l'utilisation réelle du véhicule.

Les services d'urbanisme des villes où évoluent ces ambulances connectées pourront également tirer profit de l'exploitation de ces connaissances (carrefours dangereux, circulation difficiles,...) lors de ré-aménagements du réseau routier.

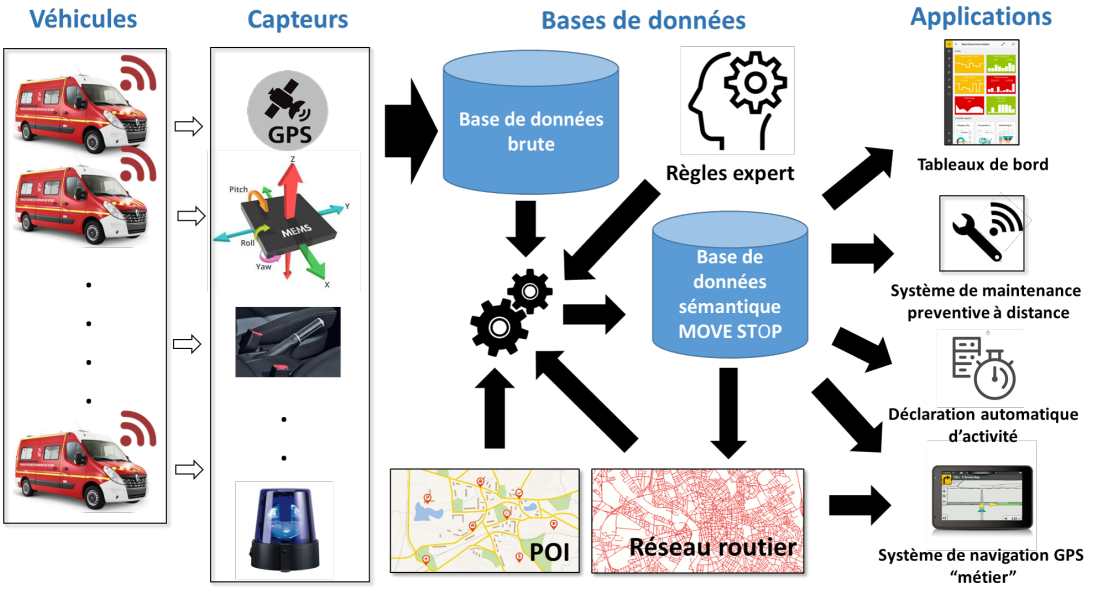


FIGURE 1.1 – Aperçu du processus général du traitement des données

Des recherches sur les déplacements de véhicules d'urgence ont déjà étés menés dans le cadre de l'ANR POSAMU (Performances et Optimisation Systémique de l'Aide Médicale Urgente) [Perret et al., 2013]. Dans celles-ci, les auteurs se focalisent sur l'estimation du temps de parcours d'une ambulance (déplacement routier et à pied) afin d'optimiser l'organisation du SAMU.

1.2 Motivations

Le projet, dans lequel s'inscrit cette thèse, a pour objet d'offrir des possibilités analytiques fortes dans le cadre des ambulances intelligentes. Ces possibilités analytiques permettrons, à plus long terme, d'avoir un processus automatisé de déclaration d'activité des ambulances, de mettre en place des tableaux de bord pour les gestionnaires de flottes, de proposer un système de maintenance préventive, ainsi qu'un système de navigation GPS intelligent enrichit des connaissances métiers extraites.

La figure 1.1 représente le processus général envisagé du traitement des données brutes jusqu'à l'exploitation de ces connaissances pour les différentes applications suscitées. La première phase de ce processus est la collecte des données issues des véhicules connectés. Ces données brutes sont alors stockées dans une base de données. À l'aide d'informations externes (centres d'intérêt liés au métier – *Point Of Interest (POI)* –, réseau routier) et de règles expert, des données, ou trajectoires,

sémantiques (STOP, MOVE) décrivant les activités spatio-temporelles des véhicules sont extraites. Ces connaissances, portant essentiellement sur les activités des objets mobiles, sont enfin restituées à travers différentes applications.

La motivation de cette thèse est l'extraction automatique des comportements métiers à partir des traces des ambulances. Cette extraction devra s'appuyer sur les données issues de différents capteurs, car nous faisons l'hypothèse que les informations des différents capteurs disponibles peuvent faciliter et améliorer la détection des STOP et des MOVE ainsi que des activités qui leurs sont associées. Elle devra également être suffisamment générique pour être adaptée à d'autres types de véhicules spécifiques et d'autres agglomérations.

Cette thèse et le projet de manière plus générale abordent des verrous allant de l'acquisition des données à leur restitution en passant par l'extraction des trajectoires sémantiques.

1.3 Verrous technologiques et scientifiques

Pour mettre en place ces différents processus, nous avons besoin de lever différents verrous technologiques et scientifiques.

Le tableau 1.1 présente les différents verrous et indique ceux levés dans cette thèse et ceux à résoudre lors de futurs travaux. Le premier verrou technologique porte sur l'acquisition et le traitement des données. Ce verrou a été traité partiellement avant le début de cette thèse du fait de la mise en place d'un système d'acquisition global dans chaque véhicule qui synchronise la collecte des données émise par le capteur (cf. section 4.3). Les différents capteurs des véhicules vont fournir des données fortement hétérogènes. Les véhicules sont équipés de capteurs différents. Chaque capteur fournit des informations avec des fréquences différentes mais le processus de transmission des ambulances "normalise" les données renvoyées ainsi que les fréquences de transmission. De plus, la qualité de ces données doit être vérifiée avant le processus de montée en sémantique (**VT1**).

Le premier verrou scientifique est la modélisation globale du processus. Celui-ci doit permettre à notre processus de garder une certaine généricité. Le spectre de cette modélisation doit aller des capteurs disponibles dans les véhicules jusqu'aux trajectoires montées en sémantique (**VS1**). Une trajectoire montée en sémantique, aussi appelée trajectoire sémantique, est une séquence d'activités de type STOP (arrêt) et MOVE (déplacement).

L'extraction de trajectoires sémantiques à partir des données brutes lève deux verrous particulièrement complexes. Dans un premier temps, les trajectoires doivent être segmentées afin de définir lorsque le véhicule est considéré en mouvement ou non. En effet, il faut être capable de distinguer les arrêts courts (attente à un feu,...) liés à la mobilité d'une part, et d'autre part les arrêts liés à l'activité métier de l'objet

mobile (intervention, plein de carburant,...). Les endroits où le véhicule est considéré arrêté doivent ensuite être reliés à des points d'intérêt (POI) différents (casernes, hôpitaux, cliniques, lieux d'intervention ...) qui doivent être identifiés (**VS2**). Nous appellerons dans cette thèse la résolution de ce verrou **montée en sémantique faible**. Dans un deuxième temps, l'enrichissement sémantique doit être complété. Celui-ci va concerner essentiellement la détection du type de mission (départ vers une intervention, transport de patient, retour à la caserne...). Une méthode pour assigner une activité en fonction des règles métiers et des données acquises doit également être définie (**VS3**). **Cette tâche est appelée montée en sémantique forte.**

La transmission et le stockage de données ayant un coût non négligeable dans le domaine de l'Internet des Objets (aussi appelé *Internet of Things* (IoT)), il faut également être capable de déterminer quels capteurs sont nécessaires et indispensables à une montée en sémantique fiable (**VS4**).

Une étude poussée de l'évolution des données capteurs, couplée à la mise en place de nouveaux capteurs sur le véhicule permettra, à terme, de pouvoir déterminer l'usure de certains composants du véhicule en fonction de son utilisation réelle et ainsi recommander l'entretien du véhicule (**VT2**). Une fois le processus de montée en sémantique validé, l'application de celui-ci en temps réel permettra la définition d'une déclaration automatique d'activité des équipes d'intervention (**VT3**). La mise en place d'un entrepôt de données doit offrir aux experts (gestionnaires, décideurs,...) des tableaux de bords résumant les indicateurs liés aux métiers (**VT4**).

Finalement, un algorithme d'optimisation multicritères d'itinéraire doit être adapté au problème des véhicules de secours, potentiellement en s'inspirant des approches faites pour le transport à la demande [Genre-Grandpierre and Josselin, 2005]. Pour chacun des 3 types de missions (intervention, transport, retour), les modes de conduite reposent sur des choix stratégiques de déplacements complètement différents qui cherchent à optimiser certains facteurs (temps de trajet le plus court lors du départ en intervention, mouvements souples évitant les cahots lors du transport sanitaire). En fonction de ces préférences, un itinéraire optimal est calculé. Il se base sur les connaissances terrain qui ont été extraites. Ainsi, un chemin optimal spécifique au type de mission tenant compte des horaires devra être proposé. L'itinéraire proposé sera donc différent de ceux indiqués par les systèmes de navigation GPS classiques (**VS5**).

Le dernier verrou levé dans cette thèse est un verrou méthodologique. En effet, une des principales difficultés de notre méthodologie est qu'il n'existe pas de vérité terrain évidente. Les déclarations existantes souffrent de problèmes de qualité, elles sont remplies manuellement par des professionnels focalisés sur des interventions urgentes. Il est donc difficile de pouvoir valider des résultats (taux de précision, rappel,...) avec certitude. Pour valider les résultats, il est nécessaire de réaliser une campagne de relevés précise et rigoureuse directement sur le terrain (**VM1**).

Verrous technologiques (T) - scientifiques (S) - méthodologiques (M)	Section	Travaux futurs
VT1 - Captation, stockage et pré-traitement des données	4.4	-
VT2 - Déterminer l'usure des véhicules pour préconiser des entretiens	-	X
VT3 - Déclarer automatiquement et en temps réel les activités effectuées par l'équipe d'intervention	5 et 6	Temps réel
VT4 - Permettre la création aisée de tableau de bords métiers	6.5.3 (POC)	X
VS1 - Modélisation globale des données brutes jusqu'aux trajectoires sémantiques	3.1	-
VS2 - Montée en sémantique faible	5.3 et 6.3	-
VS3 - Montée en sémantique forte	5.4 et 6.4	-
VS4 - Quels capteurs pour détecter quelles activités ?	5 et 6	-
VS5 - Comment intégrer les données extraites pour obtenir un système de navigation GPS métier ?	-	X
VM1 - Validation terrain	5.6.1 (pré-étude)	X

TABLE 1.1 – Liste des différents verrous levés dans la thèse

1.4 Démarche scientifique

Conformément aux motivations présentées précédemment, nous nous intéressons dans cette thèse à l'extraction des trajectoires sémantiques riches. La philosophie globale étant de déterminer comment les capteurs peuvent aider et améliorer les processus de montée en sémantique.

Dans cette idée, nous cherchons premièrement à définir et modéliser, dans un contexte multi-capteurs, les trajectoires à différents niveaux (de brutes à sémantiques riches). Nous proposons également un processus générique permettant leurs extractions. Le processus pour l'extraction de trajectoires sémantiques les plus évoluées s'appuie sur des règles métiers.

Nous cherchons ensuite à déployer ce processus, en accord avec les modèles et définitions précédentes, dans notre contexte industriel. Pour cela, nous proposons, dans un premier temps, une analyse du cas d'étude et des pré-traitements à effectuer. Dans ce cadre, l'étude des processus métiers liés aux déplacements des ambulances nous permet d'identifier des règles métier sur les modalités des déplacements des ambulances ainsi que des motifs de tournée suite à des départs en intervention typiques du métier. Par ailleurs, l'étude de la qualité et de la pertinence des données récoltées par le système, qui a été déployé sur les ambulances dans un but différent de celui de ce projet, montre qu'il y a des traitements nécessaires à effectuer sur les données afin de rendre les données cohérentes.

De cette analyse, nous proposons deux approches différentes pour la montée en sémantique à partir des données récoltées.

La première se base sur l'idée d'utiliser uniquement le frein à main comme capteur supplémentaire au GPS. Ce capteur a été choisi car, d'une part, il est déjà présent dans le système déployé sur les véhicules, et, d'autre part, il devrait être, sauf dans certains cas particuliers, relâché lors des déplacements et serré lors des arrêts du véhicule. Il devrait donc permettre de détecter des STOP et les MOVE à un coût computationnel réduit. Pour cela, nous définissons de nouvelles méthodes de segmentation des trajectoires. Pour ce qui est de la montée en sémantique de plus haut niveau, nous nous appuyons sur un ensemble de règles métiers extraites de l'analyse de notre cas d'étude. Cependant, notre expérimentation montre des limites de cette première approche. L'usage des seuls capteurs GPS et frein à main ne sont pas suffisant pour segmenter et enrichir les STOP et les MOVE.

Aussi, notre seconde approche vise à pallier ces limites par l'usage de capteurs complémentaires informant :

— de la vitesse au compteur,

— du contact ou non du moteur,

— de l'état des gyrophares.

Les deux premiers sont utilisés en complément du frein à main pour la détection des STOP et MOVE. Le capteur sur l'état des gyrophares est quant à lui utilisé d'une part pour enrichir la partie sémantique en complément du GPS. D'autre part, il est aussi employé pour identifier de possibles sous-déplacements successifs répondant à des activités différentes dans un déplacement initial (ex. reroutage d'une ambulance vers une intervention). Les résultats obtenus sont prometteurs pour lever, à terme, les verrous non traités dans cette thèse.

L'organisation de ce manuscrit présentée dans la section suivante, reflète cette démarche scientifique.

1.5 Plan

Ce mémoire est organisé en 7 chapitres. Après cette introduction, le deuxième chapitre est consacré aux travaux de la littérature sur les trajectoires. Ce chapitre d'état de l'art se décompose en quatre parties (la modélisation, la segmentation, la montée en sémantique et enfin l'analyse et la fouille de trajectoires). Le troisième chapitre présente notre modèle générique. Il englobe les données multi-capteurs ainsi que les trajectoires sémantiques fortes. Le quatrième chapitre décrit le cas d'étude dans lequel s'inscrit ce mémoire. Le domaine d'application ainsi que les pré-traitements effectués sur les données y sont présentés. Le cinquième chapitre développe le processus complet de montée en sémantique forte avec l'utilisation de données classique GPS et le capteur de frein à main. Des analyses des résultats obtenus sont également discutés. Le sixième chapitre étend les résultats du précédent en intégrant dans la montée en sémantique des capteurs issus de l'IoT. Le dernier chapitre résume et conclut les travaux de cette thèse. Il ouvre également la discussion sur les verrous qu'il reste à traiter et de nouvelles perspectives.

Chapitre 2

État de l'art

2.1 Modélisation de trajectoires

Les trajectoires sont étudiées scientifiquement depuis 1869. La première étude reconnue concerne les déplacements de l'armée Napoléonienne pendant la campagne de Russie. La carte de la figure 2.1 a été créée par Charles Joseph Minard en 1869. Cette carte représente le déplacement des troupes françaises lors de l'intervention de Napoléon 1er en Russie. La trajectoire en beige correspond au déplacement des troupes à l'aller, tandis que la noire représente le trajet retour. Cependant, cette représentation à la particularité de présenter plusieurs dimensions de cette trajectoire. En effet, en plus de représenter le chemin emprunté par les troupes de l'armée française, l'épaisseur du trait représente le nombre de soldats de l'armée. Tandis qu'en bas de la carte, se trouvent les données de température de l'air au cours de la retraite (correspondant donc à la trajectoire en noire, mais qui donne une idée relative de la variation de température suivant les zones traversées). En résumé, c'est la première carte publiée qui représente, en plus d'une simple trajectoire (dimension géographique), une dimension quantitative liée au temps (évolution du nombre de soldats) ainsi que des données externes liées à la trajectoire (température de l'air lors du chemin retour).

Depuis, d'autres approches macroscopiques représentant des phénomènes comme des congestions urbaines [Al-Mallah, 2018] ou les migrations [Ramon-Gonen and Gelbard, 2017] sont réalisées fréquemment. Aujourd'hui, nombre de ces études sont réalisées en temps réel à l'aide d'instruments de comptage (boucle de comptage sur les routes, tourniquets d'accès à des sites, etc.) connectés via des réseaux.

Les études de déplacements individuels (microscopiques) sont plus récentes. La section suivante présente un des premiers modèles s'intéressant aux déplacements individuels sous forme de trajectoires brutes et enrichies.

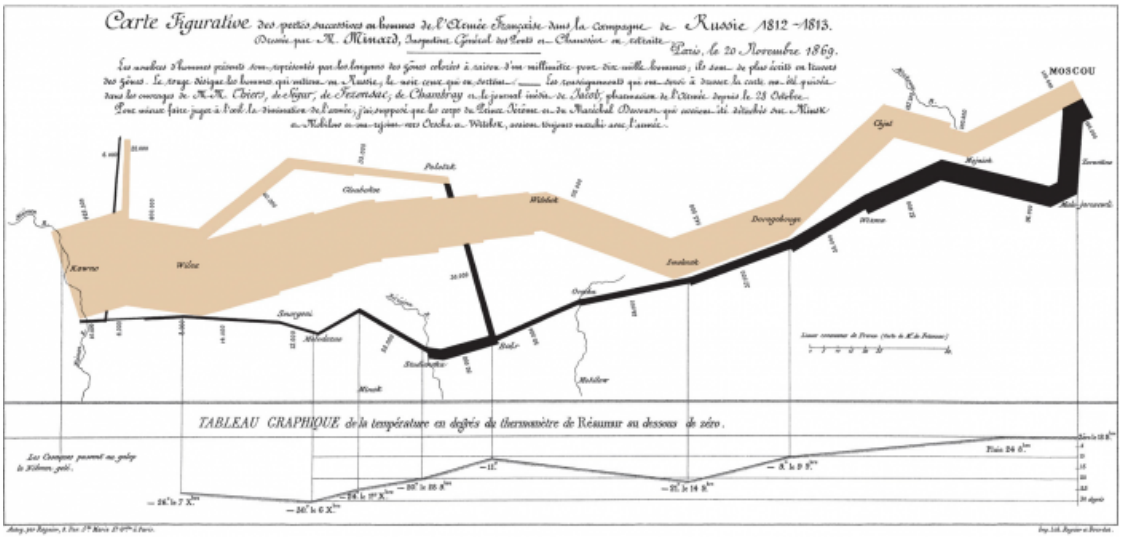


FIGURE 2.1 – Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813

2.1.1 Modèle issue de la Time geography

Les premières études scientifiques modernes qui ont considéré les déplacements individuels ont été réalisées dans le cadre de l'école de la Time geography encore appellée time-space geography [Hägerstrand, 1970]. Cette école avait pour objectif de représenter les déplacements afin d'identifier les interactions sociales. Pour chaque individu, un chemin est représenté selon une échelle spatio-temporelle. Pour la dimension temporelle cette échelle va classiquement de la journée à la vie d'individus. Pour la dimension spatiale, cette échelle concerne l'espace géographique dans lequel les individus se déplacent.

Un langage visuel est associé : le Space Time cube qui est en 3D. Il inclut la dimension spatiale en 2D et la dimension temporelle sur la troisième dimension. Chaque chemin est une suite de bundles. Un *bundle* est une partie du chemin individuel, il peut être statique si l'individu ne se déplace pas ou dynamique sinon. Classiquement les bundles statiques sont nommés STOP et les bundles dymamiques MOVE. Ce type de représentation permet l'étude et la comparaison visuelle de déplacements individuels (mode de déplacement, durée, chemin emprunté, etc.).

La figure 2.2, issue de [Parkes and Thrift, 1980] présente un exemple de représentation de trajectoires à l'aide d'un Space Time cube. Sur celui-ci, la hauteur du schéma représente l'évolution dans le temps, tandis que la surface plane représente une carte en deux dimensions. On peut y distinguer deux trajectoires différentes, qui évoluent au long d'une journée. La première est celle d'une femme, qui part de son domicile tôt le matin, à vélo, accompagnée de son enfant qu'elle dépose à l'école. Ensuite, elle effectue à nouveau un court trajet à vélo. Elle effectue également un

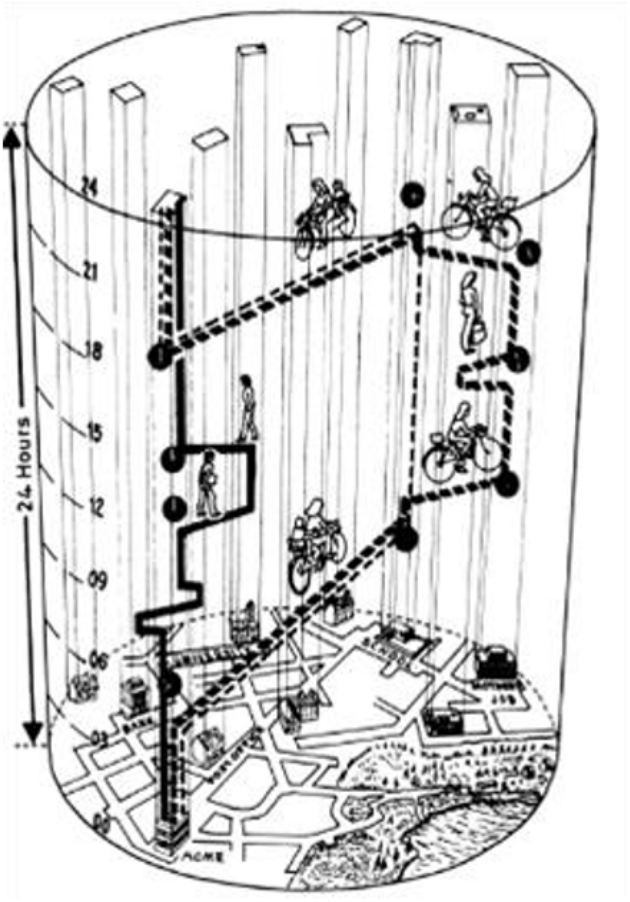


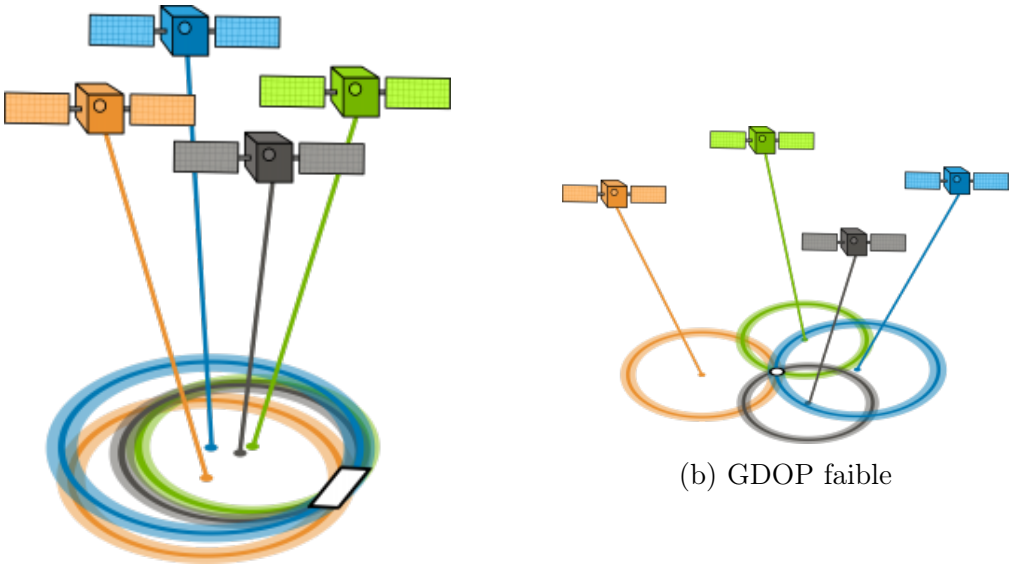
FIGURE 2.2 – Exemple de représentation de trajectoires à l'aide d'un space-time cube.

trajet à pied, puis retourne chercher son enfant à l'école et retourne à son domicile à vélo. La deuxième trajectoire est celle d'un homme, qui quitte son domicile plus tard. Il effectue ensuite quelques déplacements à pieds puis retourne à son domicile jusqu'au soir. Bien que précis et détaillé autant sur la durée des déplacements que les modes de transport, ce type de représentation devient rapidement complexe lorsque la zone d'étude s'agrandit ou que les trajectoires sont nombreuses.

Cependant, ces modèles géographiques et sociologiques étaient peuplés de données acquises manuellement. L'arrivée du système GPS et des autres GNSS (Géolocalisation et Navigation par un Système de Satellites), a permis de capter automatiquement la dimension spatiale des déplacements d'un grand nombre d'individus ou d'objets mobiles.

2.1.2 Trajectoires brutes

Une trajectoire brute est un ensemble de points de l'espace ordonnés dans le temps. Ces points représentent les coordonnées d'un objet dans l'espace. La position de ces objets est acquise à l'aide d'un système de positionnement par satellites également appelée GNSS. Le premier système de positionnement par satellites accessibles au grand public, et le plus connu, est le Global Positioning System (GPS), devenu opérationnel en 1995. Son mode de fonctionnement est aujourd'hui encore utilisé par d'autres systèmes de positionnements développés par d'autres pays (GLONASS par la Russie, Galileo par l'Europe, COMPASS par la Chine, etc.). L'utilisation des systèmes GNSS, autrefois employés à des fins purement militaires est aujourd'hui utilisé dans de nombreux domaines (agriculture, transport maritime, aérien et routier, etc.). Le mode de fonctionnement d'un GNSS est le suivant : plusieurs satellites sont en orbite autour de la Terre à chaque instant. Ces satellites diffusent, entre autres, leur position précise dans le ciel, ainsi que l'heure. A l'aide des messages transmis par ces satellites, un système de réception GNSS est capable de déterminer avec précision sa position (longitude, latitude et altitude), l'heure exacte et également sa vitesse de déplacement. En effet, connaissant la position précise des satellites (le récepteur doit "capter" le signal d'au moins 4 satellites pour déterminer précisément sa position), l'heure de transmission des messages et la vitesse de diffusion des ondes émises, le récepteur est capable par multilatération de calculer sa position sur le globe. Cependant, plusieurs facteurs peuvent perturber ce compte, et donc, le résultat de la position finale. Tout d'abord, la qualité du récepteur peu impacter la qualité de la mesure (performance de l'antenne, fréquence d'acquisition des données, réactivité du système effectuant les calculs, etc.). D'autres phénomènes viennent également perturber la mesure. Tout d'abord, les conditions météorologiques ont un impact certain. En effet, les récepteurs utilisent une variable fixe comme vitesse de propagation des ondes dans l'atmosphère. Or, les ondes ne se propagent pas à la même vitesse lorsque qu'elles traversent des nuages par exemple, rendant faux le calcul de distances avec les satellites. La nature des ondes émises par les satellites rendent très difficile, voire impossible leur transmission à travers des surfaces épaisses (bâtiments couverts par exemple), allant jusqu'à rendre le système GNSS inopérant à l'intérieur de certains bâtiments. Les ondes peuvent aussi rebondir sur de grandes surfaces, c'est notamment le cas lorsque le récepteur se trouve dans un couloir entre de hauts bâtiments (canyon urbain), et ainsi fausser le calcul de positionnement du récepteur (effet rebond). Du fait de ces sources d'imprécisions, d'autres données sont généralement stockées afin de déterminer la qualité de la donnée acquise. Tout d'abord, le nombre de satellites utilisées pour le calcul de la position est stocké. Plus ce nombre est important, plus la mesure peut être considérée comme fiable. Ensuite, il existe un facteur, calculé par le récepteur pour mesurer la qualité du signal reçu. Il s'agit de la dilution de précision (DOP en anglais). Cette mesure permet de cal-



(a) GDOP élevé

(b) GDOP faible

FIGURE 2.3 – Impact de la position des satellites sur la qualité de la mesure de positionnement

culer l'erreur potentielle de la mesure en fonction de la position des satellites dans l'espace. Plusieurs types de DOP [US Office of the Department of Defense, 2020] peuvent être calculés :

— HDOP - dilution de précision horizontale

— VDOP - dilution verticale de précision

— PDOP - dilution de précision la position (3D)

— TDOP - dilution de temps de précision

— GDOP - dilution géométrique de précision

La figure 2.3, issue de gisgeography.com [1] présente comment la position des satellites peut impacter le calcul de la position. Dans la figure 2.3a, les satellites sont très proches, le rectangle blanc, visible sur la figure, est la position potentielle du récepteur en fonction des données reçues. Ici, le DOP géométrique (GDOP) est élevé, ce qui signifie une mauvaise qualité de positionnement. A l'inverse, dans la figure 2.3b, les satellites sont positionnés de sorte que la position potentielle du récepteur (cercle blanc sur la figure) soit très petit, et de fait, la qualité de la mesure considérée comme fiable (GDOP faible).

1. `https://gisgeography.com/gps-accuracy-hdop-pdop-gdop-multipath/`

Afin d'améliorer la précision des mesures, une amélioration du GPS a été inventée, le DGPS (pour Differential Global Positioning System). Le principe du DGPS est simple, il s'agit de considérer qu'en des points voisins, les erreurs de mesures issues soit du dérèglement de l'horloge des satellites, soit des conditions de propagation des ondes (nuages, etc.) sont identiques.

De ce fait, un réseau de stations fixes (ayant une position connue précise) à été créé. Ces stations mesurent la variation entre leurs positions exacte et leurs positions mesurées puis les transmet aux GPS environnants (jusqu'à plusieurs centaines de kilomètres à la ronde). Ceci permet aux GPS recevant ces informations de corriger leur propre position et ainsi améliorer leur précision.

Finalement, les données brutes sont des tuples de données (au minima longitude, latitude et temps) potentiellement enrichis de différents indices permettant d'apprécier la qualité de la mesure. Une trajectoire brute est donc une séquence de points (x,y) ou de positions (x,y,t) brutes ordonnées temporellement.

Aujourd'hui, nous disposons d'un très grand nombre de trajectoires brutes employées dans des domaines très variés. Afin de pouvoir les étudier finement, en fonction de leur contexte applicatif, des modèles informatiques ont été développés. Ils reprennent les concepts définis dans la Time Geography. Ces modèles sont présentés dans la section suivante.

2.1.3 Modèles de trajectoires enrichies

La représentation des trajectoires brutes ne s'appuie que sur des dimensions spatiales et temporelles. Les modèles de trajectoires enrichis ont proposé d'intégrer de nouvelles dimensions. Par exemple, nous pouvons citer l'ajout des dimensions suivantes :

— Activités
— Mode de déplacement
— Données météorologiques
— Points d'intérêts (POI)

Les modèles les plus significatifs de la littérature vont être décrits par la suite.

Dans une optique d'analyse fine des données GPS collectées, [Alvares et al., 2007] proposent un des premiers modèles permettant la détection et l'extraction d'épisodes dans une trajectoire. Ces épisodes sont de type MOVE lorsque l'objet mobile est considéré en déplacement, et de type STOP lorsqu'il ne l'est pas. Dans ce modèle, une liste de géométries classées par catégories (hôtels, lieux touristiques, etc.) est donnée en entrée, et leur algorithme propose de définir si pour chaque point donné de la trajectoire, ceux-ci constituent un STOP ou un MOVE en fonction du temps passé dans une zone de STOP potentiel. Les trajectoires sémantiques finalement obtenues (séquences de STOP annotés sémantiquement et de MOVE) permettent

une analyse de celles-ci et l'utilisation des méthodes de data mining classique (règles d'associations, découverte de motifs fréquents) mais également de pouvoir interroger directement au travers de requêtes les données collectées (combien de fois cet hôtel a-t-il été visité ? pendant combien de temps en moyenne ?). Bien qu'ajoutant un nom à ces zones de STOP potentiels, cette modélisation de trajectoires reste cependant très orientée géométrie.

C'est de ce constat que [Spaccapietra et al., 2008] proposent un modèle pour les trajectoires sémantiques. Le modèle proposé est conceptuel. Les auteurs proposent ici de non plus se concentrer uniquement sur l'aspect géométrique d'une trajectoire, mais plutôt sur le côté sémantique de celles-ci. L'idée émergente du modèle proposé est de montrer qu'une trajectoire sémantique n'est pas une suite de données GPS brutes collectées, mais plutôt une séquence de STOP (où l'on effectue une activité pendant un certain temps, dans un lieu définit) connectés entre eux par des MOVE. Finalement, les MOVE n'ont ici pour fonction que de modéliser le déplacement entre deux STOP disjoints géographiquement. Cependant, ce modèle conceptuel n'est pas associé à un modèle logique. La même équipe a donc par la suite développé un modèle logique : SeMiTri [Yan et al., 2011]. Il est associé à un framework. Ce dernier permet l'ajout d'annotations sémantiques à chaque position de la trajectoire.

Par la suite, [Parent et al., 2013] ont formalisé informatiquement les concepts d'une trajectoire sémantique. Les STOP sont alors décrit par une activité, et les MOVE par un mode de déplacement. Une fois ces trajectoires sémantiques extraites, il est ainsi possible d'extraire le comportement des objets mobiles étudiés. La figure 2.4 représente une trajectoire sémantique de touriste. Dans cette figure, on peut voir le trajet effectué par un touriste dans la ville de Paris. Il part de l'hôtel Zola. Il va en bus à la Tour Eiffel. Il se déplace ensuite en bateau pour aller visiter le Palais Bourbon. Il alterne ensuite 2 activités séparées par des déplacements, deux visites et un repas au restaurant avant de retourner à son hôtel. Les activités (visites, repas, etc.) sont modélisées par des STOP en fonction du POI lié (Si le STOP est effectué dans un musée, alors l'activité sera une visite.) Les déplacements par des modes de transport.

Le modèle CONSTAnT (pour CONceptual model of Semantic TrAjectories) proposé par [Bogorny et al., 2014] intègre les concepts définis précédemment. Il propose en effet un modèle complet pour les trajectoires sémantiques, prenant en considération de multiples dimensions. La figure 2.5 présente le modèle conceptuel de trajectoires sémantique présenté. Dans ce modèle, une trajectoire sémantique est composée de sous trajectoires sémantiques, elles mêmes constituées de points sémantiques. Tout d'abord, l'objet mobile étudié ainsi que l'appareil ayant servit à récolter les données GPS sont associés à la trajectoire sémantique. Cela permet notamment de conserver des mesures de qualité et de précision des traces GPS récoltées ainsi que l'objet mobile de l'étude. Ensuite, une autre dimension du modèle porte sur le comportement lié à l'objet mobile étudié (objectif du déplacement, déplacement seul ou en groupe, motifs de déplacements, mode de déplacement, etc.). L'objectif

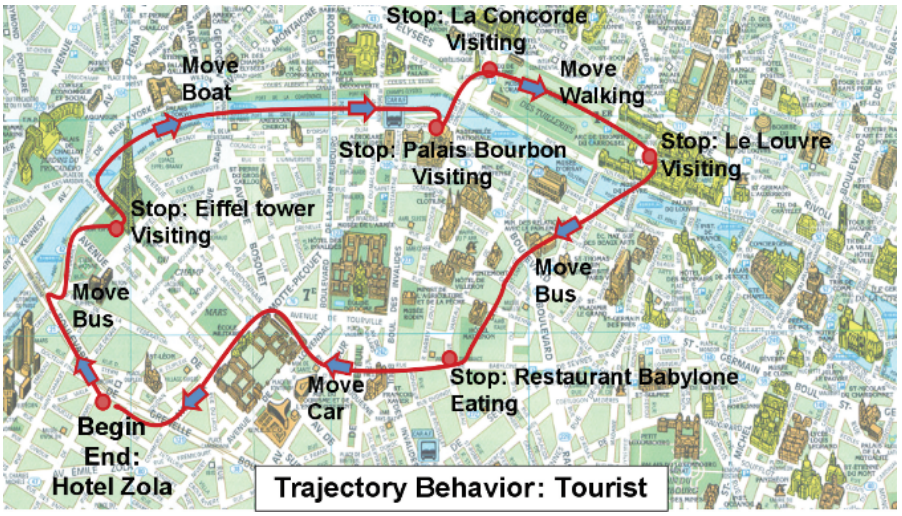


FIGURE 2.4 – Trajectoire sémantique d'un touriste issue de [Parent et al., 2013].

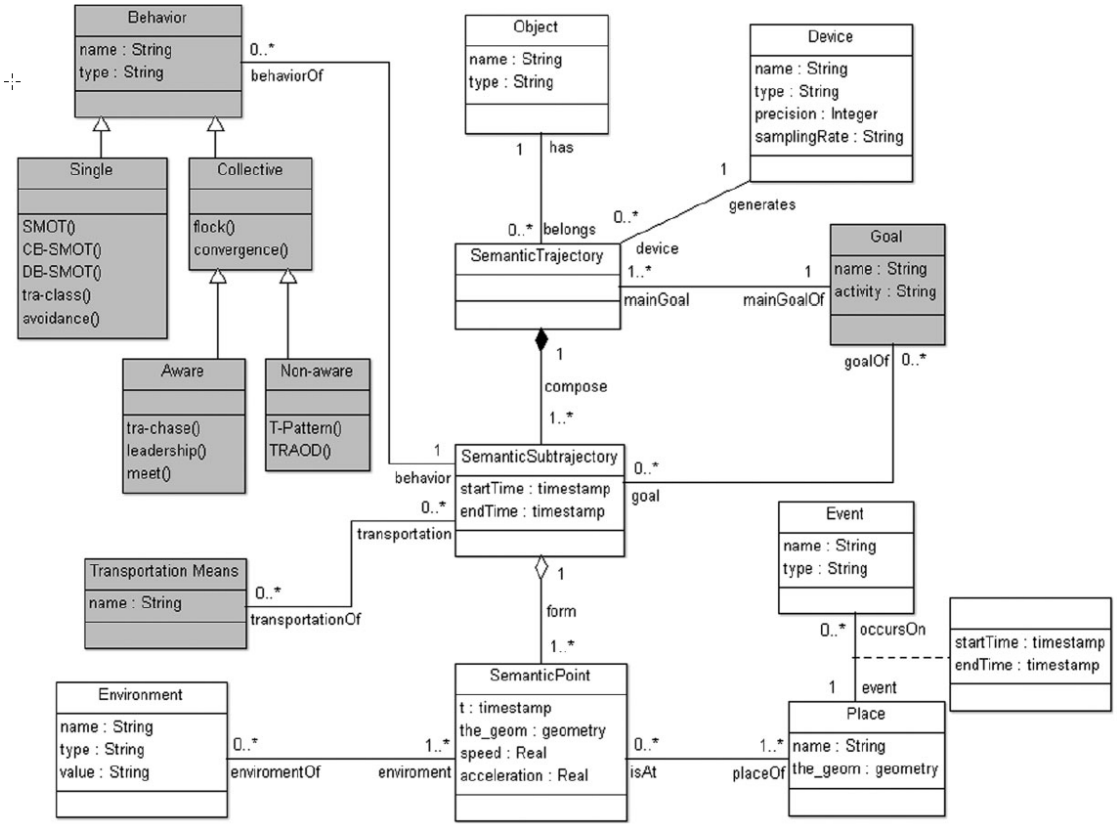


FIGURE 2.5 – Le modèle conceptuel CONSTAnT issu de [Bogorny et al., 2014]

du déplacement peut porter à la fois sur la trajectoire sémantique globale, ou bien sur une sous trajectoire sémantique. La dernière partie du modèle est l'annotation sémantique des points qui constituent une trajectoire. D'une part l'environnement physique est décrit par des valeurs captées. D'autre part, un POI (place) peut être associé à ce point de trajectoire sémantique. A travers différents exemples, les auteurs illustrent l'emploi du concept d'environnement d'un point de trajectoire sémantique. Par exemple, il est possible de définir un environnement de type température, et ainsi modéliser pour chaque position la température relevée. Les POI permettent de représenter le lieu d'intérêt où se trouve le point, mais également de les associer à des évènements qui s'y déroulent. Cette association est temporelle.

Un exemple concret de représentation sémantique à l'aide de CONSTAnT est présenté dans la figure 2.6. L'objet étudié est ici une personne qui se nomme John Smith. Les données récoltées l'ont été à des intervalles de 10 secondes avec un GPS ayant une précision de 5 mètres. L'objectif de la trajectoire sémantique étant la visite de Rome. Cette trajectoire est constituée de plusieurs sous trajectoires sémantiques. Dans cette figure, seule la première instance de chaque objet est indiquée. Cette sous trajectoire commence le 2 avril 2012 à 13h50 et se termine à 15h10 le même jour. L'objectif (goal) de cette instance est "participation à un événement". Le comportement de déplacement a pour valeur "non-aware joint movement" que nous pouvons traduire par déplacement groupé non intentionnel. En d'autre termes, l'individu s'est déplacé avec d'autres personnes. Ils ne se connaissaient pas forcement, mais ils parcourent ensemble le même chemin. Le déplacement des autres individus n'impacte donc pas le déplacement de l'individu étudié. Ce déplacement à été effectué à pied. Enfin, au niveau des points sémantiques (la encore, par soucis de lecture, un seul point est représenté). Un point relevé le 2 avril 2012 à 14h00 est modélisé. Lors de ce relevé, l'individu se déplaçait à une vitesse de 0.4km/h. La météo extérieure était ensoleillée. La position géométrique de ce point est liée au Colisée de Rome. Pendant cette période, une exposition photo durant toute la journée avait lieu au Colisée.

Finalement d'autres modèles spécifiques à un type de trajectoires ont aussi été proposés. Par exemple, pour les trajectoires de véhicules sur un réseau, un modèle s'appuyant sur le graphe modélisant le réseau est défini par [Popa et al., 2015]. Ce type de modèle est très efficace (requêtes rapides, compression des trajectoires) mais ne permet pas de gérer des déplacements dans un espace ouvert.

Pour plus de détails sur les concepts définis à travers ces modèles de trajectoires sémantiques, vous pouvez vous référer à l'état de l'art très détaillé de la thèse de [Paiva Nogueira, 2017].

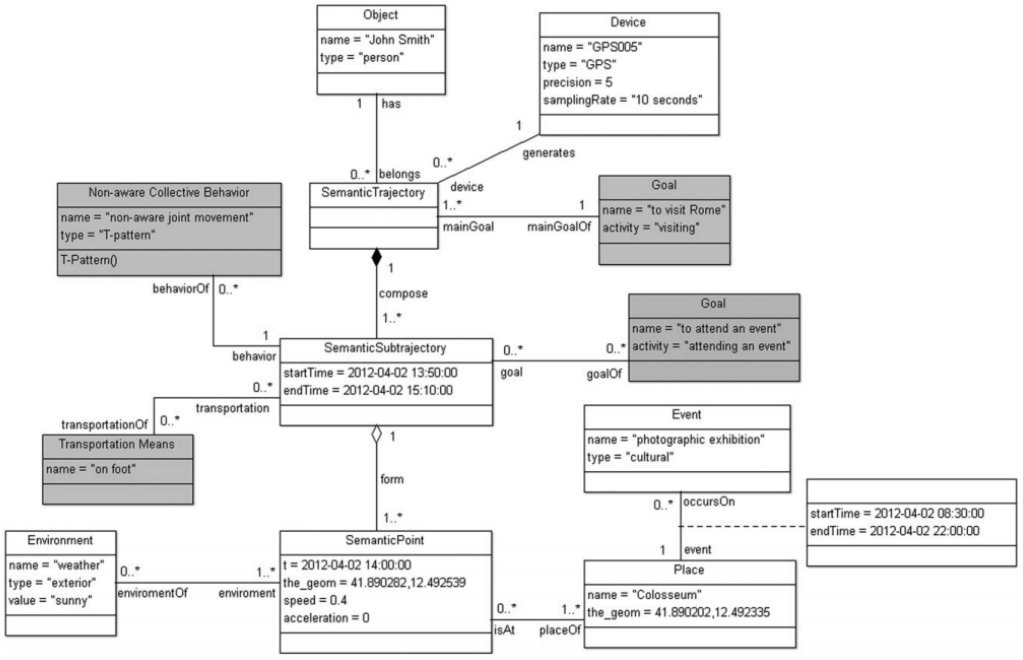


FIGURE 2.6 – Exemple d'instance d'un touriste visitant Rome à l'aide de CONSTAnT.

2.2 Processus de segmentation

Au cours du temps, le comportement d'un objet mobile change. Afin de décrire ce dernier, il est nécessaire de procéder à la segmentation de la trajectoire. Il s'agit de pouvoir décomposer la trajectoire en différentes sous trajectoires homogènes. Il existe plusieurs méthodes de segmentation de trajectoires, la plus utilisée est la segmentation en épisodes de STOP et de MOVE. Cependant, ces méthodes peuvent varier en fonction du type de segments recherchés. Ces méthodes sont régulièrement utilisées dans un objectif de compression de trajectoires. L'idée principale de ces méthodes étant de représenter une trajectoire par des segments représentatifs de celle-ci. Dans [Douglas and Peucker, 1973], les auteurs proposent un algorithme (noté DP) afin de compresser une trajectoire. Au départ, un premier segment est formé à l'aide du point de départ et d'arrivée de la trajectoire à simplifier. Ensuite, le point intermédiaire le plus éloigné de ce segment est sélectionné (en utilisant la distance euclidienne perpendiculaire DEP), si cette distance est inférieure à un certain seuil, alors l'algorithme se termine. Sinon, ce point est sélectionné pour former deux nouveaux segments. Cette procédure se répète récursivement sur chacun des segments jusqu'à qu'il n'y ai plus de points ayant une distance DEP supérieure au seuil. La complexité de cet algorithme est $O(n^2)$, avec n le nombre de points de la trajectoire.

Plusieurs extensions de DP ont été proposées afin d'améliorer l'algorithme et réduire sa complexité. [Hershberger and Snoeyink, 1992] a utilisé les propriétés de l'enveloppe convexe pour accélérer la détection du point le plus éloigné, avec une complexité de $O(n \log n)$. L'inconvénient majeur des méthodes précédentes est qu'elles ignorent les données temporelles en raison de l'utilisation d'erreur spatiale. [Meratnia and Rolf, 2004] étend DP pour considérer la dimension temporelle en utilisant la distance euclidienne synchrone (DES). [Gudmundsson et al., 2009] a développé une version de DP qui permet de prendre en compte la présence de boucles dans les trajectoires.

Également, [Popa et al., 2015] proposent une méthode de compression de trajectoires en utilisant des données externes. En s'appuyant sur le réseau routier, les trajectoires sont d'abord appareillées au réseau. Puis la trajectoire est segmentée de telle sorte que chaque point puisse être interpolé (si la vitesse de l'objet mobile est constante sur le segment, alors sa position à un instant t peut facilement être calculée). La méthode proposée permet ainsi de représenter une trajectoire par des segments au lieu de nombreux points tout en conservant un taux d'erreur lors de la représentation globale très faible.

Plus de détails sur la compression de trajectoires peuvent être trouvés dans [Meratnia and Rolf, 2004, Popa et al., 2015, Zhang et al., 2018].

Dans le cadre de cette thèse, et dans un objectif de montée en sémantique, l'état de l'art de la segmentation de trajectoire portera sur la détection d'épisodes de STOP et MOVE comme définit par [Spaccapietra et al., 2008] et [Alvares et al., 2007]. Plusieurs approches sont utilisées afin de déterminer quand un objet mobile peut être considéré en mouvement ou à l'arrêt. Ces approches peuvent être divisées en plusieurs catégories [Zhang and Lin, 2019, Bermingham and Lee, 2018, Longgang and Xiaotian, 2016]. Nous présenterons une description de ces catégories ainsi que leur mode de fonctionnement. Pour chacune de ces catégories, nous présenterons une ou plusieurs méthodes représentatives de la littérature.

2.2.1 Méthode utilisant des informations géographiques externes

Cette méthode s'appuie sur l'utilisation de données géographiques externes afin de détecter des STOP dans les trajectoires. Les données externes ont une représentation spatiale, et, en fonction du temps passé par l'objet mobile dans la zone, il est possible de déterminer que celui-ci effectuait ou non un STOP.

Parmi les travaux les plus significatifs, [Alvares et al., 2007] proposent un modèle pour l'enrichissement de trajectoires sémantique à l'aide de données géographiques. Dans cet article, ils proposent également un algorithme nommé SMoT (pour Stops and Moves of Trajectories) qui, permet la détection de STOP dans une trajectoire à l'aide de zones d'intérêts. Dans celui-ci, chaque point de la trajectoire est étudié un

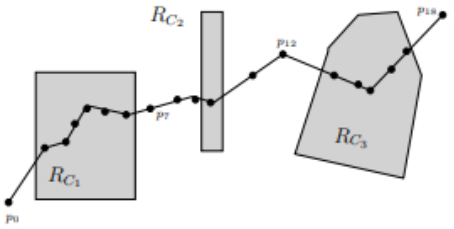


FIGURE 2.7 – Exemple d'application de SMoT sur une trajectoire issue de [Alvares et al., 2007]

à un. Si un sous ensemble de points consécutifs est à l'intérieur de l'une des zones d'intérêts et que la cardinalité de cet ensemble est supérieure à un seuil δ défini, alors ce sous ensemble est identifié comme un STOP. Un concept similaire est utilisé par [Zheng et al., 2009] afin de détecter les STOP. Ici, les auteurs utilisent également un seuil de distance pour déterminer si les points de la trajectoire peuvent être considérés comme des STOP ou non. La figure 2.7 illustre un exemple d'utilisation de SMoT sur une trajectoire. Sachant trois zones géographiques d'intérêt déterminées (RC_1, RC_2 et RC_3), un seuil minimum de taille $\delta = 2$ et une trajectoire T constituée de points $p_0, p_1, p_2, ..., p_{18}$, alors, les points inclus dans RC_1 et RC_3 sont considérés comme des STOP car au moins δ points consécutifs sont inclus dans une zone d'intérêt. Ce qui n'est pas le cas des points inclus dans RC_2 : il n'y a qu'un seul point inclus dans cet élément, les points ne sont donc pas considérés comme des STOP par l'algorithme. Ce dernier nécessite un échantillonnage régulier.

L'avantage de ces méthodes est qu'il est possible d'attribuer une notion sémantique au STOP, en fonction du type de zone dans lequel il s'effectue. Cependant, bien qu'il soit possible d'extraire des zones d'intérêts automatiquement (OSM, Google Maps, etc.), des données sont incomplètes, voire manquantes, suivant le domaine de l'objet mobile étudié, rendant cette méthode partiellement inopérante.

2.2.2 Méthode utilisant l'extraction de caractéristiques issues des trajectoires

Pour détecter les épisodes de STOP et de MOVE, il est possible d'utiliser des caractéristiques issues des données brutes afin de déterminer si les points font partie d'un STOP ou d'un MOVE. Il est par exemple possible de calculer la vitesse entre deux points et la direction de l'objet mobile afin de déterminer si l'objet mobile était en mouvement ou non. Dans [Stopher et al., 2005], les auteurs utilisent en plus d'autres contraintes, comme une durée minimum et une variation maximale de la valeur des coordonnées.

L'extraction de ces caractéristiques permet ensuite, via l'utilisation de règles liées à l'objet mobile étudié, une extraction des épisodes de STOP et de MOVE des trajectoires.

Par exemple, dans [Rocha et al., 2010], les auteurs calculent la variation de l'angle de la direction d'un bateau. Un clustering est ensuite appliqué sur les données constituant la trajectoire. Les points ayant une variation faible sont considérés comme des MOVE, alors que les points ayant une forte variation pendant une durée minimum définie sont considérés comme des STOP.

Ce type de segmentation ne nécessite que peu d'informations liées aux trajectoires. Les résultats obtenus permettent d'identifier des séquences de STOP et de MOVE cohérentes avec l'objet mobile étudié (bateaux, véhicules d'urgences, etc.). Cependant, ces approches sont très dépendantes du type d'objet mobile, du capteur GPS et du système de récoltes mis en œuvre. Ces méthodes nécessitent une caractérisation ad hoc et montrent donc des limites en termes de généricité pour son adaptation à d'autres types d'objets mobiles. De plus, l'étape de clustering est très coûteuse.

D'un autre côté, il est de plus en plus courant que les récepteurs GPS soient intégrés à des solutions plus globales (smartphone, véhicule connecté, etc.) étant capables de fournir de nombreuses informations en plus de la position (accélération, vitesse, caméra, etc.). Il serait ainsi possible d'utiliser ces données comme des caractéristiques propres à la trajectoire afin d'extraire des épisodes de STOP et de MOVE. Par exemple, les données issues du réseau CAN de véhicules connectés [iso, 2016], à l'instar de celles sur l'état du moteur (allumé/éteint) ou de celui du frein à main (relâché ou non), pourraient être très pertinente afin de déterminer quand celui-ci est à l'arrêt ou en déplacement. C'est l'approche privilégiée dans cette thèse.

2.2.3 Méthode utilisant le clustering basé sur la densité spatiale et temporelle

Cette méthode utilise les principes de base d'un clustering spatial (uniquement avec des coordonnées X et Y), mais en y ajoutant une dimension temporelle. De ce fait, les points se trouvant géographiquement et temporellement proches sont regroupés dans un même cluster et forment un STOP. Les points ne faisant pas partie d'un cluster sont considérés comme des MOVE.

Généralement, les travaux de la littérature utilisent un dérivé de la méthode de clustering DBSCAN [Ester et al., 1996]. C'est un algorithme de clustering qui ne nécessite pas de connaître le nombre attendu de clusters et n'oblige pas à associer à un cluster les valeurs isolées. Il utilise 2 paramètres en entrée : une distance ϵ, ainsi qu'un nombre minimum de points $MinPts$. Cependant, elle nécessite de définir manuellement ces deux valeurs.

Pour qu'un point soit considéré comme un centre de cluster (ou regroupement), appelé point centre, il faut que le point ait dans son voisinage de rayon ϵ un nombre minimum de points $MinPts$ (par exemple le point A de la figure 2.8).

Pour qu'un point soit inclus dans un cluster, il doit être soit un centre de cluster, soit être couvert par un centre de cluster de rayon ϵ (points B de la figure 2.8). Les centres de clusters positionnés à une distance inférieure à ϵ sont regroupés pour ne former qu'un seul cluster (les points rouges de la figure 2.8 sont tous des centres de cluster qui forment un ensemble plus grand).

Les points ne remplissant aucune des deux conditions sont considérés comme des points isolés (points M et N de la figure 2.8).

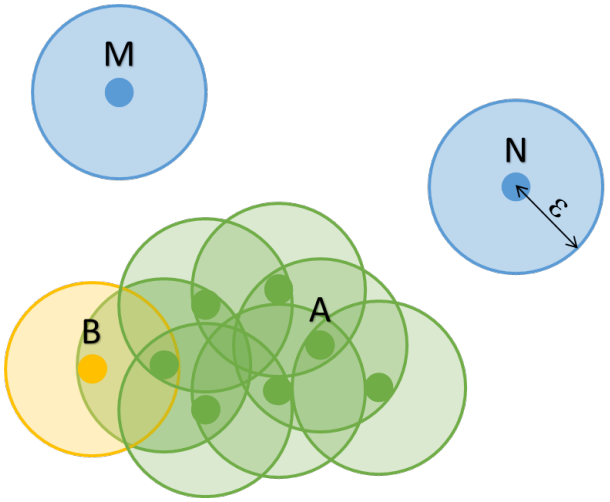


FIGURE 2.8 – Représentation du fonctionnement typique de DBSCAN.

Cette méthode semble ainsi idéale pour la découverte d'épisodes de STOP une fois la prise en compte de la dimension temporelle effectuée. C'est dans l'optique de prendre en considération cette dimension temporelle que plusieurs améliorations de l'algorithme initial ont été produites, chacun ayant des paramètres d'entrée ou des complexités différents. Nous pouvons par exemple citer ST-DBSCAN [Birant and Kut, 2007] qui prend en données d'entrée 4 paramètres.

— Une distance maximum pour les attributs spatiaux

— Une distance maximum pour les attributs non spatiaux

— Un nombre minimum de points pour former un cluster

— Un seuil afin d'éviter la fusion des clusters (cf. [Birant and Kut, 2007])

Dans cet algorithme, la dimension temporelle est prise en compte par le fait que le voisinage des points est étudié un à un, ordonnés temporellement. T-DBSCAN [Chen et al., 2014] et TrajDBSCAN [Tran et al., 2011] sont également des extensions de la méthode initiale. T-DBSCAN propose une optimisation de la méthode de

clustering initiale en restreignant la recherche de points voisins par deux paramètres. Le premier étant un seuil de distance maximum et le second un seuil de temps maximum (le processus de recherche de voisins ne sera pas appliqué au delà de ces seuils). TrajDBSCAN modifie l'algorithme initial avec les concepts suivants :

— Le voisinage doit être obligatoirement un point suivant temporellement

— Les centres de cluster sont définis non pas par un nombre minimum de points, mais par une durée minimum (deux points peuvent former un cluster s'ils se suivent temporellement pendant cette durée)

L'avantage de cette méthode de segmentation est que de nombreux algorithmes de segmentation ont été développés à partir de l'algorithme de clustering initial. De ce fait, il est possible de trouver de nombreuses méthodes suivant les contraintes du jeu de données utilisé. Cependant, malgré des optimisations de l'algorithme, celui-ci à une complexité rendant difficile l'exécution sur de très gros volumes de données. Également, la perte de signal ou la mauvaise réception entraîne des erreurs de segmentation dans la plupart de ces algorithmes.

2.3 Montée en sémantique

La montée en sémantique de trajectoires est le processus consistant, à partir de données brutes, à obtenir une trajectoire enrichie. Cet enrichissement peut porter sur la trajectoire complète ou bien sur des sous-segments de celle-ci. Plusieurs types d'enrichissements peuvent être considérés. La figure 2.9 présente un schéma montrant les différents types d'enrichissements, ainsi que les informations nécessaires pour obtenir une trajectoire sémantique. Sont visibles sur ce schéma en bleu, les données acquises via le GPS ou des capteurs placés sur l'objet mobile (accéléromètre, température, etc...), en orange les différents types d'enrichissement des trajectoires sémantiques et en jaune les informations externes utilisées. Les flèches noires symbolisent l'utilisation d'une ressource. À partir des données brutes, il est possible d'effectuer un premier type d'enrichissement que nous appelons dans le schéma enrichissement via propriétés. Un exemple d'enrichissement via propriétés est, par exemple, l'ajout de la vitesse moyenne calculée à l'aide des points GPS, une accélération, un angle de direction... Ce type d'enrichissement ne nécessite pas de données externes, car toutes les valeurs sont calculées à partir des données récoltées. Ensuite, l'enrichissement contextuel utilise des données externes afin d'ajouter des étiquettes sémantiques aux trajectoires. La liste des données externes utilisables n'est pas exhaustive sur notre schéma, mais permet de comprendre le type d'informations nécessaires à cet enrichissement. Il est ainsi possible d'étiqueter des trajectoires de type "STOP dans une station-service", "déplacement sur la RN10", "MOVE pendant un épisode orageux" etc... Enfin, le dernier type d'enrichissement est l'enrichissement d'activité. Ce type d'enrichissement nécessite également des trajectoires potentiellement déjà

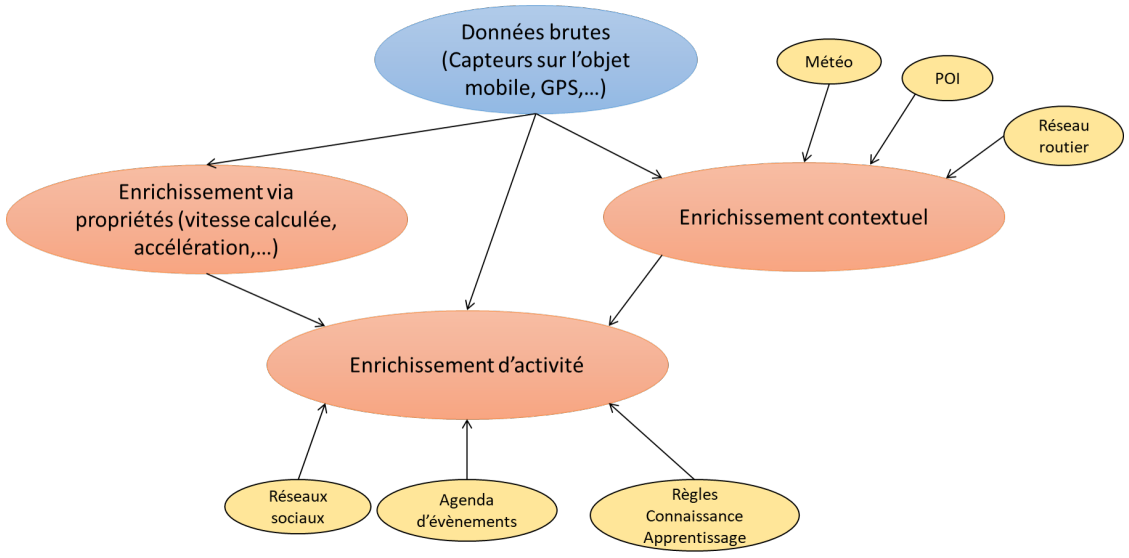


FIGURE 2.9 – Différents types de montée en sémantique en fonction des données externes utilisées.

enrichies via des propriétés ou via le contexte. L'objectif de cet enrichissement est d'obtenir des étiquettes sémantiques du type "Conduire rapidement une voiture sur la RN10", "Assister au match France - Brésil au Stade de France", etc.

La section suivante présente les trois types d'enrichissements possibles pour une trajectoire ainsi que l'usage de ces enrichissements dans la littérature.

2.3.1 Enrichissement via propriétés

L'enrichissement de trajectoires par des propriétés est relativement simple. Une fois la trajectoire segmentée, on ajoute des étiquettes sémantiques issues de données calculées. Ces données peuvent provenir du GPS lui-même, ou bien de capteurs placés sur l'objet mobile. Ainsi, une trajectoire segmentée en épisodes de STOP et de MOVE auxquels on ajoute la vitesse moyenne de l'objet mobile sur chaque MOVE est déjà une trajectoire sémantique. Comme nous l'avons vu précédemment, [Rocha et al., 2010] calculent par exemple l'angle de direction d'un bateau par rapport au nord. Ces points, enrichis de l'angle de direction forment ainsi une trajectoire sémantique (même si dans cet article, les positions sont ensuite utilisées afin de faire du clustering). Bien que la liste ne soit pas exhaustive, on peut citer comme type de propriétés la vitesse, la direction ou bien l'altitude.

Nous pouvons également à partir de ces informations extraire des variations, moyennes, maximums... Cependant, ce type d'enrichissement est rarement utilisé tel quel dans la littérature. En effet, le calcul de propriétés est en général une étape ayant pour but la segmentation, l'enrichissement d'activités ou même l'analyse statistique de trajectoires.

2.3.2 Enrichissement contextuel

L'enrichissement contextuel consiste à utiliser des données externes afin de qualifier des segments de trajectoires [Parent et al., 2013, Yan et al., 2011]. Ces données peuvent être de plusieurs types. Cependant, pour être associées aux différents segments de la trajectoire, celles-ci doivent avoir une représentation géographique liée. Nous nous attardons dans cet état de l'art à trois types de données contextuelles pouvant être utilisées dans la montée en sémantique de trajectoires.

— POI / ROI (Points ou régions d'intérêts). L'enrichissement de trajectoires à l'aide de POI est la méthode la plus répandue dans la littérature. Le principe est de relier géographiquement un épisode de trajectoire (STOP, MOVE, ou segment) à un lieu identifié. Les difficultés pour associer ces épisodes à des POI sont nombreuses. Tout d'abord, il s'agit de trouver une liste de POI / ROI corrélée avec l'activité de l'objet mobile (une liste de lieux touristiques pour une étude de trajectoires de touristes, par exemple). Afin d'obtenir de bonnes associations, cette liste doit être le plus exhaustive possible. Ensuite, il s'agit de trouver une méthode d'association géographique entre un point ou un segment et le POI ou le ROI. Si les POI / ROI sont très proches spatialement, des risques de confusion lors de l'association sont possibles. Un exemple de trajectoire obtenue est le suivant :
  — Départ de l'hôtel Villecroix
  — Déplacement
  — Arrêt à la station de bus Leclerc
  — Déplacement
  — Arrêt à la station de bus Parc des Mées
  — Déplacement dans le Parc des Mées
  — Fin
— Réseau routier : De même, il est possible d'annoter chaque segment de trajectoires en fonction du chemin emprunté par l'objet mobile. Ce processus d'association entre un réseau et une trajectoire brute est nommé map-matching. Un très bon état de l'art sur les processus de map-matching est présenté dans [Quddus et al., 2007]. Un autre, plus récent, est réalisé dans [Huang et al., 2020]. Ce processus s'appuie le plus souvent sur des chaînes de Markov. Un exemple de trajectoire obtenue est le suivant :

  - Début
  - Déplacement sur la RN10
  - Déplacement sur la RD895
  - Déplacement sur la RD662
  - Déplacement sur la A10
  - Fin
- Météo Enfin, à l'aide de données météorologiques, il est possible d'associer pour chaque point d'une trajectoire (ou segment) la météo associée (température, soleil, pression atmosphérique ...). De nombreuses sources de données existant sur Internet, il est facile d'associer cette information. Ainsi, à l'aide de la position du point à un instant donné, il est relativement aisé d'intégrer une étiquette sémantique météo à celle-ci. Un exemple de trajectoire obtenue est le suivant :
  - Début
  - Arrêt pendant un épisode nuageux à une température ambiante de 18 degrés
  - Déplacement pendant un épisode ensoleillé à une température ambiante de 25 degrés
  - Déplacement pendant un épisode pluvieux à une température ambiante de 22 degrés
  - Arrêt pendant un épisode pluvieux à une température ambiante de 19 degrés
  - Fin

Dans [Amici et al., 2014], dans le cadre d'une analyse épidémiologique, les auteurs enrichissent leurs données par des informations météorologiques.

2.3.3 Enrichissement d'activités

L'enrichissement d'activités est le type de montée en sémantique le plus avancé. Il s'agit d'associer, pour chaque segment de trajectoire, une activité. Dans le cas d'un déplacement, cette étiquette peut être l'objectif du déplacement, le mode de transport ou même une qualification (rapide, lent, dangereux, descendre des escaliers, courir, ...). Dans le cas d'un arrêt, cette étiquette est l'activité effectuée pendant celui-ci. Pour enrichir les segments de trajectoires avec des activités, les données brutes ainsi que les différents enrichissements présentés précédemment peuvent être utilisés (cf. figure 2.9). Concernant les données externes pouvant être utilisées, il existe les trois principaux types suivants dont nous présenterons les deux premiers :

- Les réseaux sociaux [Wu et al., 2015, Beber et al., 2016, Beber et al., 2017]

— Les règles / connaissances / l'apprentissage [Renaudin et al., 2007, Klotz et al., 2018]

— Les agendas d'évènements [Bogorny et al., 2014]

Dans [Beber et al., 2016, Beber et al., 2017], les auteurs proposent une méthode qui associe des activités à des trajectoires en utilisant des données issues de réseaux sociaux. Ici, les données provenant de tweets [2] géolocalisés constituent une trajectoire. Les auteurs construisent ensuite un classifieur qui permet de déterminer l'activité associée au tweet. Cette méthode utilise également des données issues de Foursquare [3]. Ce média social permet à l'utilisateur d'indiquer où il se trouve et de commenter l'activité qu'il effectue (restaurant, shopping, etc...). L'étude des mots utilisés dans les tweets, couplées aux données issues de Foursquare permet ensuite de créer le classifieur qui détermine l'activité effectuée en fonction du lieu, du temps passé et du contenu du tweet.

Dans [Renaudin et al., 2007], les auteurs équipent des sapeurs-pompiers de différents types de capteurs (gyroscopes, accéléromètres, magnétomètres, système de positionnement indoor). Ensuite, à l'aide des données récoltées et de règles de logique floue, il est possible de qualifier chaque épisode de la trajectoire afin de déterminer l'activité effectuée. Dans cet article, la méthode proposée permet d'identifier si la personne est debout, couchée, assise ou entrain de marcher. Lorsque la personne marche, la solution proposée permet de déterminer si la personne effectue une marche vers l'avant, une montée des escaliers, une descente des escaliers vers l'avant ou bien une descente d'escalier vers l'arrière (les pompiers utilisent souvent la descente d'escalier vers l'arrière, car elle est plus sécurisée).

Dans [Klotz et al., 2018], les auteurs enrichissent des trajectoires issues de voitures. En utilisant les informations fournies par le réseau CAN du véhicule, les auteurs cherchent à étiqueter les différents sous-segments de trajectoires en fonction de valeurs émises par les capteurs présents dans le véhicule. Les auteurs proposent également une ontologie issue de la fusion de plusieurs ontologies issues d'une part du domaine de l'automobile, et d'autres issues du domaine des trajectoires sémantiques. Dans cet article, l'utilisation d'ontologies permet surtout de lier les données extraites avec les connaissances issues de ces différentes ontologies. La méthode proposée permet ainsi d'extraire des segments du type "angle de braquage des roues croissant", "freinage brusque" mais permet également d'annoter globalement une trajectoire "conduite douce", par exemple. L'objectif de la méthode proposée à long terme est de pouvoir créer des requêtes reconnues (ontologies) sur des trajectoires. Cependant, cette méthode est très spécifique et il est nécessaire d'adapter les valeurs suivant l'objet mobile étudié (qu'est-ce qu'un freinage fort ? qu'est-ce qu'une conduite souple ? etc.)

2. `http://twitter.com`
3. `https://foursquare.com/`

Comme présenté ci-dessus, de nombreuses méthodes existent dans la littérature afin de monter en sémantique les trajectoires. Cependant, il n'existe pas à notre connaissance de méthode permettant une montée en sémantique s'appuyant sur les contraintes métiers. En effet, pour des véhicules professionnels (poste, secours, police, entretien, etc.) il est indispensable d'intégrer au processus de montée en sémantique un modèle ou des règles métiers. Qui plus est, ces véhicules sont équipés de capteurs spécifiques qui renvoient des informations riches encore peu intégrées actuellement dans ces processus. L'objectif étant en fonction du métier et des valeurs captées de déduire les activités spécifiques à ces véhicules.

2.4 Analyse statistique et fouille de trajectoires

Une fois les trajectoires collectées, nettoyées et montées en sémantique, la dernière étape consiste à les analyser. L'analyse de trajectoires sémantiques permet une compréhension fine des comportements des objets mobiles dans l'environnement dans lequel ils évoluent. Plusieurs états de l'art généraux récents sur l'analyse des trajectoires sont disponibles dans la littérature [Yuan et al., 2017, Mazimpaka and Timpf, 2016, Zheng, 2015]. Ces états de l'art sont pertinents et montrent l'intérêt de l'analyse de trajectoires. Néanmoins, l'analyse porte essentiellement sur des trajectoires peu ou pas enrichies. La dimension sémantique des trajectoires est alors obtenue à l'issue du processus d'analyse et de fouille.

Nous avons identifié 2 familles de processus d'analyse de trajectoires enrichies. La première étant l'analyse statistique et la seconde la fouille de données.

2.4.1 Analyse statistique

L'analyse statistique est le premier type d'analyse que nous avons retenu. En effet, plusieurs indicateurs statistiques peuvent être pertinents lors de l'étude de trajectoires sémantiques [Wang et al., 2009, Su et al., 2014].

2.4.1.1 Analyse statistique classique

Tout d'abord, des indicateurs classiques (nombre, moyenne, maximum, minimum ...) peuvent être extraits sur différentes propriétés (longueur de la trajectoire, vitesse, vélocité, orientation, durée, ...) afin de répondre à des questions basiques par exemple : Combien de trajectoires composent le jeu de données ? Quelle est la durée moyenne de ces trajectoires ? Quelle est la longueur moyenne de ces trajectoires ? Quelle est la vitesse moyenne de déplacement d'un objet mobile ?

D'autres indicateurs avancés peuvent aussi être extraits à partir des premiers (dispersion, corrélation ...). Ces mesures statistiques sont souvent associées à des outils de géo-visualisation et des systèmes SOLAP (Spatial Online Analytical Processing [Bimonte et al., 2020].

L'utilisation de trajectoires sémantiques permet une analyse statistique plus fine de celles-ci. Bien que les méthodes et les indicateurs soient identiques, les questions posées peuvent être plus précises :

— Quelle est la température des oiseaux migrateurs lors de déplacements sous la pluie ? [Bogorny et al., 2014]

— Quelle est la vitesse moyenne des randonneurs en fonction de la pente en montagne ? [Kerouanton, 2020]

— Quelle est la vitesse moyenne de déplacement des herbivores lorsqu'ils broutent ? [Buard, 2013]

Dans le cadre de cette thèse, des études statistiques similaires sont pertinentes, car elles pourraient permettre de calculer, entre autres, le temps d'attente moyen aux urgences en semaine et le week-end, ainsi que la durée moyenne de déplacement d'une ambulance entre la caserne et le lieu d'intervention.

2.4.1.2 Analyse statistique orientée graphe

Une fois l'enrichissement contextuel effectué (comme vu dans la section 2.3.2), les trajectoires peuvent être représentées comme :

— Des suites de lieux d'arrêt (POI)

— Des suites d'éléments de graphes (arêtes, noeuds)

Pour cette dernière représentation, dans le cadre d'objets mobiles se déplaçant sur un réseau routier, les arêtes sont les tronçons empruntés par l'objet mobile et les noeuds les intersections (carrefours...) ou lieux d'arrêts (parkings...).

Des statistiques peuvent alors être déduites de l'utilisation de ces graphes. Par exemple :

— Vitesse moyenne sur un certain tronçon

— Nombre de trajectoires ayant emprunté un certain tronçon

— Chemin le plus suivi entre deux lieux d'arrêts

Ces statistiques sont fréquemment utilisées dans le domaine du transport et des mobilités.

Il est aussi possible de monter en abstraction et de travailler au niveau des types génériques de déplacements. Par exemple, [Schneider et al., 2013] propose d'extraire des chaînes de Markov de mobilité quotidienne résumant un peu plus de 39.000 journées de déplacements quotidiens. Le motif ayant l'ID numéro 2 de la figure

2.10 résume jusqu'à 40% des déplacements collectés selon le jeu de données étudié. Il résume les déplacements d'individus effectuant uniquement un aller-retour entre deux lieux dans une journée. Les 17 motifs présentés dans cette figure schématisent plus de 90% des motifs extraits.

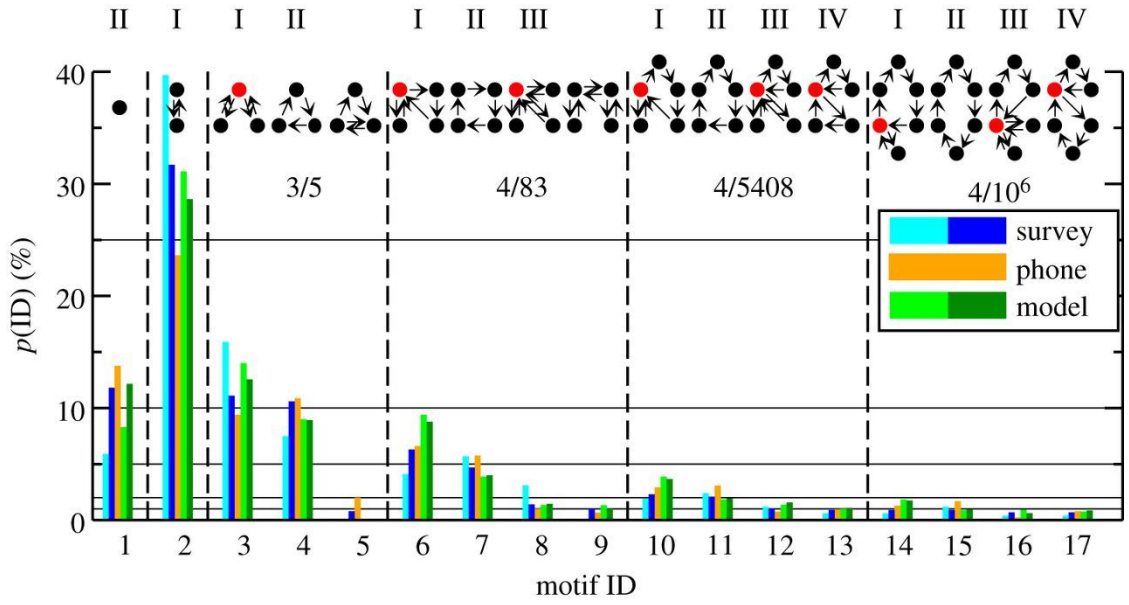


FIGURE 2.10 – Exemple des motifs principaux (daily mobility pattern) et de leur fréquence d'observation en fonction du jeu de données étudié, présenté dans [Schneider et al., 2013]

Les processus d'analyse de fouille de données ont aussi été adaptés pour extraire d'avantage d'informations issues du comportement des objets mobiles étudiés. La section suivante présente l'analyse de trajectoire au travers d'un processus de fouille de données.

2.4.2 Fouille de données

La figure 2.11 résume le processus générique de fouille des trajectoires . Ce schéma s'inspire des schémas présentés dans [Mazimpaka and Timpf, 2016, Li et al., 2011]. Différentes trajectoires d'objets mobiles (piétons, vélos, animaux, etc...) peuvent être analysées. Ces trajectoires peuvent être montées en sémantique. En fonction de l'objectif, différents besoins applicatifs sont identifiés. Par exemple, si l'étude porte sur les déplacements d'individus dans une ville, les questions suivantes sont pertinentes.

— "Q1 - Quel est le mode de transport d'un individu ?"

— "Q2 - Quels sont les lieux les plus fréquentés à certaines heures de la journée ?"

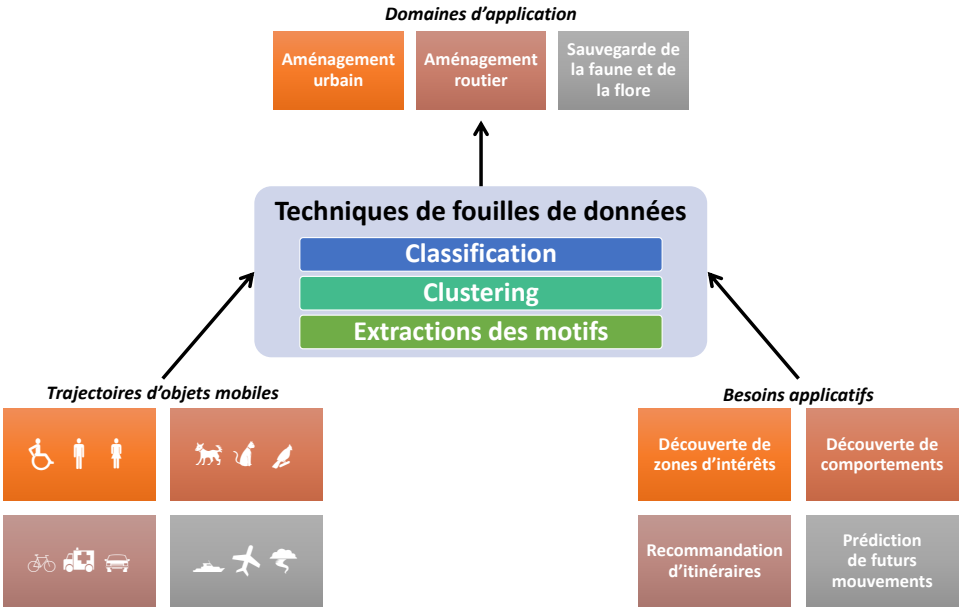


FIGURE 2.11 – Processus générique de fouille de trajectoires

— "Q3 - Quel est le trajet classique d'un piéton pour se rendre de la gare au centre-ville ?"

Ces différents questionnements vont déterminer la technique de fouille de données à employer. A l'aide de ces questionnements, nous allons présenter les 3 principales familles de fouille de trajectoires.

2.4.2.1 Classification

L'objectif de la classification est l'assignation d'objets dans des classes prédéfinies. L'idée générale est de s'appuyer sur un modèle (réseau de neurones, arbre de décision, règles d'inférence ...) pour classer de nouveaux objets [Junior et al., 2017, Saini et al., 2020]. Dans le cadre de classification de trajectoires, l'objectif peut être :

— Déterminer le mode de transport : Q1

— Déterminer le type d'objet mobile (animal, piéton, ambulance ...)

— Déterminer le niveau de danger d'un ouragan

Pour obtenir un modèle pour la classification, trois types de méthodes existent. La première méthode la plus simple étant la définition de règles d'inférence. Par exemple, pour évaluer le niveau de dangerosité d'un ouragan, les experts métiers ont défini des règles basées principalement sur la vitesse du vent en son centre et son diamètre. De manière assez similaire, un arbre de décision est aussi une

modélisation adéquate. Les feuilles de l'arbre représentent alors les différentes classes. Les nœuds intermédiaires sont des tests qui, en fonction des valeurs des propriétés de la trajectoire, l'orientent vers un ensemble de classes potentielles.

Des méthodes plus complexes de techniques d'apprentissage supervisé – *Support Vector Machines* (SVM), réseaux de neurones – permettent de classifier les trajectoires. Néanmoins, afin d'obtenir des résultats pertinents, ces méthodes nécessitent des jeux de données conséquents pour l'apprentissage du modèle. Dans [Lee et al., 2008], les auteurs proposent un algorithme de classification de trajectoires nommé TraClass. Des algorithmes de clustering sont d'abord appliqués, l'un basé sur les régions (des segments dans la même région et de la même classe) et l'autre sur les segments de trajectoires. TraClass convertit ensuite chaque trajectoire en un vecteur de caractéristiques, où chaque entrée de vecteur est soit un cluster basé-région, un cluster basé-trajectoire ou une caractéristique numérique (vitesse, etc.). Le classificateur est construit en utilisant un SVM qui est une généralisation des classificateurs linéaires.

[Bolbol et al., 2012] propose d'utiliser les propriétés de segments de trajectoires. Les trajectoires sont d'abord segmentées en un nombre prédéfini de sous trajectoires. Ensuite, en utilisant une fenêtre glissante de 4 points, on calcule pour chaque sous trajectoires différentes propriétés (accélération moyenne, vitesse moyenne, vélocité, etc.). Ces propriétés sont ensuite utilisées pour créer un classifieur de type SVM.

Ces méthodes complexes d'apprentissage supervisé nécessitent un grand volume de données préalablement étiquetées. Le coût de l'étiquetage par un expert étant élevé, [Junior et al., 2017] propose une méthode d'apprentissage active pour sélectionner le meilleur sous-ensemble de trajectoires à étiqueter.

2.4.2.2 Clustering

Le clustering est une tâche typique de la fouille de données, qui consiste à regrouper ensemble des objets similaires. Dans le cas d'objets mobiles, plusieurs types de clusters sont possibles. Dans le cadre de clustering de positions, des zones d'intérêt denses peuvent être découvertes [Ester et al., 1996, Khan et al., 2014] : Q2.

De même, dans le cadre de clustering de trajectoires, des ensembles de trajectoires semblables sont identifiés. Ces clusters sont obtenus à l'aide de distances entre trajectoires et d'algorithmes classiques de clustering [Giannotti et al., 2011].

Finalement, des clusters de segments de trajectoires sont aussi extraits. L'algorithme TraClus proposé dans [J-G Lee, 2007] commence par diviser les trajectoires en segments, en identifiant des points caractéristiques qui sont les points où le comportement d'une trajectoire change rapidement. Par la suite il relie les points caractéristiques consécutifs afin de former des segments. Dans la deuxième étape, TraClus applique l'algorithme DBSCAN sur les segments obtenus précédemment afin de pouvoir identifier les sous-trajectoires communes et créer les clusters. Enfin,

une fois que tous les clusters sont calculés, des suites de segments représentatifs sont générées pour chaque cluster à l'aide de segments moyens.
D'autres méthodes prennent en compte la dimension temporelle des trajectoires. C'est le cas de l'algorithme CONVOY [Jeung et al., 2010], qui applique un algorithme de clustering à chaque instant t. Il relie ensuite les différents clusters obtenus aux différents instants afin d'identifier les objets mobiles qui font route ensemble. Pour plus de détails sur le clustering de trajectoires, vous pouvez consulter l'état de l'art récent proposé dans [Bian et al., 2018].

2.4.2.3 Extraction de motifs

Les méthodes de classification et de clustering sont employées pour définir des groupes de données similaires. Les méthodes de fouille de motifs permettent de synthétiser ces groupes grâce à l'extraction d'éléments récurrents. Par exemple la réponse à la question Q3 : Quel est le trajet classique d'un piéton pour se rendre de la gare au centre-ville ? est une trajectoire synthétique définie à partir de l'ensemble des trajectoires des piétons entre la gare et le centre-ville.

Cette synthèse peut être représentée sous la forme :

— d'une suite ordonnée temporellement de positions [Etienne, 2011]

— d'une suite ordonnée temporellement d'étiquettes sémantiques [Giannotti et al., 2007, Moreau et al., 2020b, Moreau et al., 2020a]

La découverte de motifs de comportements est intéressante dans l'analyse des données et peut également aider à la prédiction de mouvements, la recommandation d'itinéraires et à la détection d'évènements anormaux (outliers). Ces méthodes de fouille de motifs peuvent être classées en trois catégories : (i) motifs périodiques, (ii) motifs fréquents, (iii) motifs de groupes.

Un motif périodique est un motif issu d'un seul objet mobile qui se répète sur une période donnée. Par exemple, un individu allant au travail et qui rentre chez lui chaque jour, un animal qui recherche quotidiennement de la nourriture. Le problème de la fouille de motifs périodiques est la nécessité de connaître la période à laquelle l'évènement se répète [Mamoulis et al., 2004, Cao et al., 2007]. L'algorithme Periodica, présenté dans [Li et al., 2010], ne nécessite pas de connaître à l'avance cette période. La première étape consiste à chercher toutes les zones fréquemment visitées. Ensuite pour chaque zone, l'algorithme trouve toutes les périodes possibles en transformant la séquence de mouvement dans chaque zone en une séquence binaire (la $i$ème valeur de la séquence est égale à 1 lorsque l'objet est dans cette zone à l'instant i). Et enfin, il applique une transformation de Fourier et une auto-corrélation circulaire pour déterminer la valeur exacte de la période. Les segments de trajectoires correspondant à la période et aux zones trouvées sont ensuite extraits pour créer des motifs périodiques.

Un motif fréquent est un motif issu de plusieurs objets mobiles qui ont effectué les mêmes déplacements à un instant différent. Par exemple, des cyclistes qui empruntent la même portion d'une piste cyclable, à des horaires différents. Dans la littérature, beaucoup de méthodes existent afin de découvrir des motifs fréquents dans les trajectoires. L'idée première est d'appliquer une méthode de clustering sur des sous-segments de trajectoires sans prendre en compte la dimension temporelle de celles-ci. Les trajectoires étant dans le même cluster peuvent être employées pour définir un motif fréquent. Cependant, cette méthode est difficilement applicable lorsque le volume de données devient trop important. Dans [Giannotti et al., 2007], les auteurs proposent une méthode nommée T-pattern. Cette méthode se décompose en deux étapes. La première est la découverte de zones d'intérêt dans les trajectoires. Les trajectoires sont ensuite reconstruites comme étant des successions de visites de ces zones d'intérêts. Ainsi, des motifs fréquents sont extraits dès lors que des trajectoires comportent des séquences semblables. Les motifs extraits sont des suites ordonnées temporellement d'étiquettes sémantiques. A l'inverse, [Etienne et al., 2016] propose une méthode permettant l'extraction de motif représentés sous la forme d'une trajectoire médiane et d'un boxplot spatio-temporel.

Un motif de groupe est un motif issu de plusieurs objets mobiles qui ont effectué les mêmes déplacements au même instant. Par exemple, des oiseaux qui se déplacent ensemble lors d'une migration, des usagers du métro empruntant une même rame. L'algorithme CONVOY, vu précédemment, peut être assimilé à un algorithme de fouille de motif de groupe. En effet, les positions issues des trajectoires étant regroupées à chaque instant t, si des trajectoires se trouvent dans le même cluster, cela signifie que les objets mobiles ont suivi le même chemin au même instant. En synthétisant les trajectoires d'un même cluster, il est donc possible d'extraire une trajectoire médiane résumant le motif de groupe.

2.4.3 Discussion

Actuellement, les outils d'analyses statistiques ou orientés fouille de données sont déjà très riches. Cependant, ils sont peu appliqués à des spécificités métiers. L'objectif de cette thèse en termes d'analyse n'est donc pas de développer de nouveaux outils mais d'adapter les outils existants afin de répondre aux besoins des experts. Par exemple, en ce qui concerne l'analyse des déplacements des véhicules d'urgence à l'aide de Daily Mobility Patterns, deux adaptations sont nécessaires. Premièrement, nous devons passer de l'étude des activités quotidiennes à l'analyse des activités après chaque sortie d'une ambulance de la caserne. De même, les lieux d'arrêt doivent être regroupés en fonction du type métiers (lieu d'intervention, hôpitaux, lieu logistique, casernes...).

2.5 Conclusion

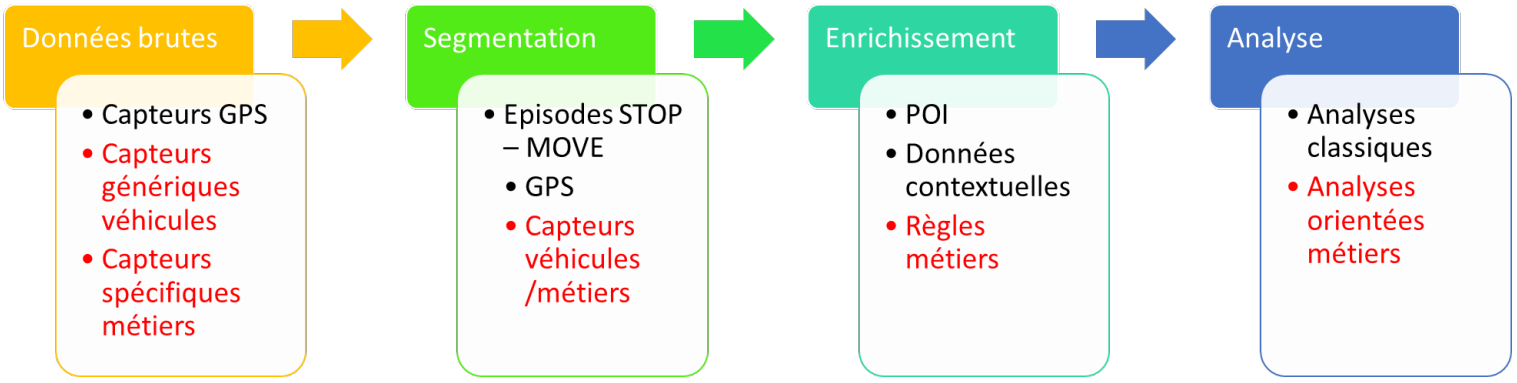


FIGURE 2.12 – Schéma du processus de traitement des données brutes jusqu'à l'analyse de trajectoires sémantiques

Dans cet état de l'art, nous avons vu qu'il était possible de passer de trajectoires brutes à l'analyse de trajectoires segmentées et riches sémantiquement. La figure 2.12 représente ce processus. Dans cette figure, les données, connaissances et analyses dont l'utilisation dans la littérature est commune sont en noir sur le schéma. L'originalité de cette thèse est d'intégrer la dimension métier aux trajectoires sémantiques en s'appuyant sur les capteurs spécifiques aux véhicules métiers. Dans le cadre de cette thèse, nous allons enrichir le processus classique. Les données, connaissances et analyses supplémentaires nécessaires que nous utiliserons sont en rouge. Pour les données brutes, nous utiliserons en plus, des données issues de multiples capteurs. La prise en compte de données issues de multiples sources implique la création d'un nouveau modèle pour les trajectoires sémantiques. Pour améliorer la phase de segmentation, ces données supplémentaires seront employées pour détecter plus précisément les épisodes de STOP et de MOVE. De même, pour la phase de montée en sémantique, des règles métiers fines, obtenues avec l'aide des professionnels

du métier seront intégrées. Finalement, les trajectoires sémantiques ainsi obtenues pourront aussi être analysées en tenant compte des attentes inhérentes à chaque métier. Nous proposons donc dans cette thèse d'intégrer des données issues de multiples capteurs afin d'extraire des trajectoires sémantiques plus fines orientées métiers à l'aide de règles issues d'experts du domaine. Le processus que nous présentons est un processus générique. Il doit pouvoir être adapté à différents métiers (secours, maintien de l'ordre, gestion du domaine public routier, collecte des déchets, etc.). Dans cette thèse, ce processus est expérimenté à l'aide de données issues des véhicules d'urgence et de secours. Le chapitre suivant présente notre proposition de modèle pour les trajectoires sémantiques enrichies à l'aide de données multi-capteurs.

Chapitre 3

Principes et modèle générique de l'extraction de trajectoires sémantiques à partir de données multi-capteurs

3.1 Introduction

Dans ce chapitre, nous introduisons le processus qui nous permet d'extraire des trajectoires sémantiques fortes à partir des données brutes. Ce processus s'inscrit dans une démarche de fouille de trajectoires [Zheng, 2015].

En effet, nous avons vu dans l'état de l'art (cf. 2.4) que la fouille de trajectoires s'appuie régulièrement sur la montée en sémantique de type STOP-MOVE, où les STOP peuvent être associés à des points d'intérêt (Point Of Interest - POI). Le fait de passer d'une **trajectoire brute** – une séquence de données GPS historisées – à une séquence de successions de STOP et MOVE, est appelé montée en sémantique, et les trajectoires obtenues sont des trajectoires sémantiques (cf. 2.3). Nous considérons que les sémantiques des STOP et des MOVE sont in fine assez peu informatives. Nous appelons cette démarche **montée en sémantique faible** et les trajectoires ainsi obtenues, **trajectoires sémantiques faibles**.

Nous proposons dans ce travail de thèse de monter à une sémantique plus forte en utilisant des informations issues de capteurs embarqués tout au long du processus.

En effet, dans notre contexte, nous disposons de véhicules équipés de plusieurs capteurs. Aussi, les trajectoires que nous pouvons acquérir sont des séquences de données GPS et de différents capteurs historisés. Nous les appelons **trajectoires brutes étendues**.



FIGURE 3.1 – Schéma de la montée en sémantique forte

À l'aide de ces trajectoires brutes étendues, nous pouvons 1) définir de nouvelles méthodes de segmentation de ces trajectoires brutes étendues, 2) monter ces sous-trajectoires en sémantiques faibles de type STOP ou MOVE, et 3) associer à ces STOP ou MOVE des sémantiques métiers. Nous appelons ce processus, dont les grandes étapes sont rappelées dans la figure 3.1, la **montée en sémantique forte**. Les trajectoires ainsi obtenues sont **des trajectoires sémantiques fortes**.

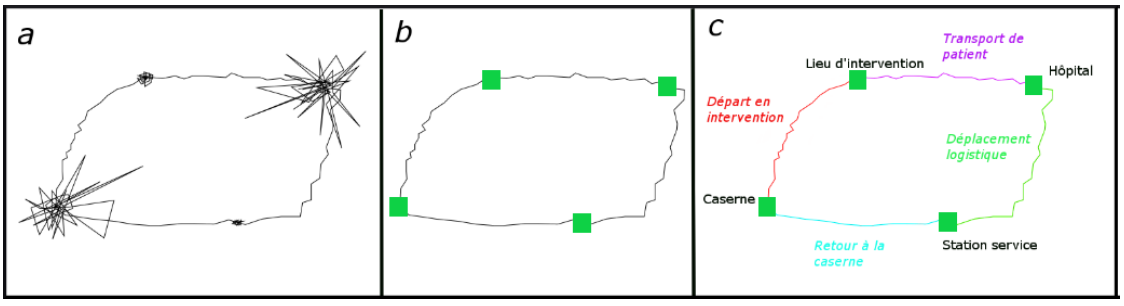


FIGURE 3.2 – Illustration du principe de montée en sémantique forte : Trajectoire brute étendue (a), Extraction des lieux d'arrêts(b), Étiquetage des STOP et des MOVE (c)

La figure 3.2 donne un exemple des résultats des étapes sur un cas fictif.

Ainsi, la première étape de la montée en sémantique forte est l'acquisition des données issues des différentes sources. Les véhicules sont équipés de capteurs GPS classiques, de capteurs génériques liés au véhicule (frein à main, vitesse à l'ordinateur de bord, clignotants...) ainsi que de capteurs liés à l'activité du véhicule (compteur de taxi, sirène d'une ambulance, charge d'une benne à ordure...). Le système de récoltes et les capteurs pouvant être utilisés dans les ambulances à l'heure actuelle sont présentés dans la section 4.3. L'acquisition de données consiste à importer les données dans une base de données géographiques et à y traiter les problèmes de qualités (cf. 4.4).

La deuxième étape est la montée en sémantique faible. Les trajectoires brutes étendues sont segmentées. Les STOP et MOVE sont identifiés. Ensuite, les STOP associés à des lieux correspondant possiblement à des arrêts liés au métier (caserne,

hôpital, station essence, etc.) identifiés à l'aide de données extérieures (objets géographiques d'OpenStreetMap (OSM), liste de points d'intérêt métier, etc.). Nous obtenons ainsi des trajectoires sémantiques faibles.

Enfin la dernière étape est la montée en sémantique forte. À l'aide d'informations métier, les activités des STOP et celles des MOVE sont déduites (objectifs du déplacement ou de l'arrêt). Cette étape finale nous permet de produire des trajectoires sémantiques fortes. Dans cet étape, en fonction des processus métiers étudiés, il peut être nécessaire de segmenter les MOVE à nouveau pour identifier de possibles reroutages. Par exemple, un taxi rentrant d'une course peut-être redirigé vers une nouvelle prise en charge d'un client, une ambulance une fois que le patient a été déposé à l'hôpital peut à tout moment être reroutée sur une nouvelle intervention ou à l'inverse, une intervention peut être annulée et le véhicule rentre alors dans sa caserne d'affectation.

Dans ce chapitre, nous présentons tant d'un point de vue formel que UML les différentes notions impliquées dans le processus de montée en sémantique. Ainsi, dans la section suivante nous abordons les notions de trajectoire brute et de trajectoire brute étendue. Les notions de trajectoire sémantique faible puis de trajectoire sémantique forte sont introduites dans la section 3.3. La section 3.4 présente l'ensemble du processus permettant l'extraction des trajectoires sémantiques.

3.2 Trajectoire brute et trajectoire brute étendue

3.2.1 Trajectoire brute

Définition 1 (Trajectoire Brute) *Une trajectoire brute* $T = \langle e_1^-, ..., e_n^- \rangle$ *est définie, pour tout* $i \in [\![1, n]\!]$ *telle que :*

- $e_i^- = (t_i, p_i)$, e_i^- *est une position*
- $p_i = (x_i, y_i) \in \mathbb{R}^2$ *est un point issu d'un système de positionnement par satellites (GPS, GALILEO...)*
- *une estampille temporelle* $1 \leq t_i \leq t_{i+1}$.

Exemple 1 *L'exemple suivant est illustré dans la figure 3.3. Pour une trajectoire brute* $T = \langle e_1^-, e_2^-, e_3^-, e_4^-, e_5^- \rangle$ *on a :*

- $e_1^- = (t_1, p_1) = (1, (0, 0))$
- $e_2^- = (t_2, p_2) = (2, (0, 0))$
- $e_3^- = (t_3, p_3) = (3, (0, 0))$
- $e_4^- = (t_4, p_4) = (4, (1, 2))$
- $e_5^- = (t_5, p_5) = (5, (3, 4))$

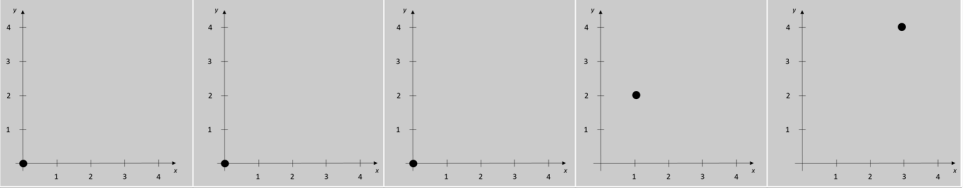
Elément	e_1^-	e_2^-	e_3^-	e_4^-	e_5^-
T_i	1	2	3	4	5
Position	(0,0)	(0,0)	(0,0)	(1,2)	(3,4)
Représentation					

FIGURE 3.3 – Représentation de la trajectoire brute de l'exemple 1

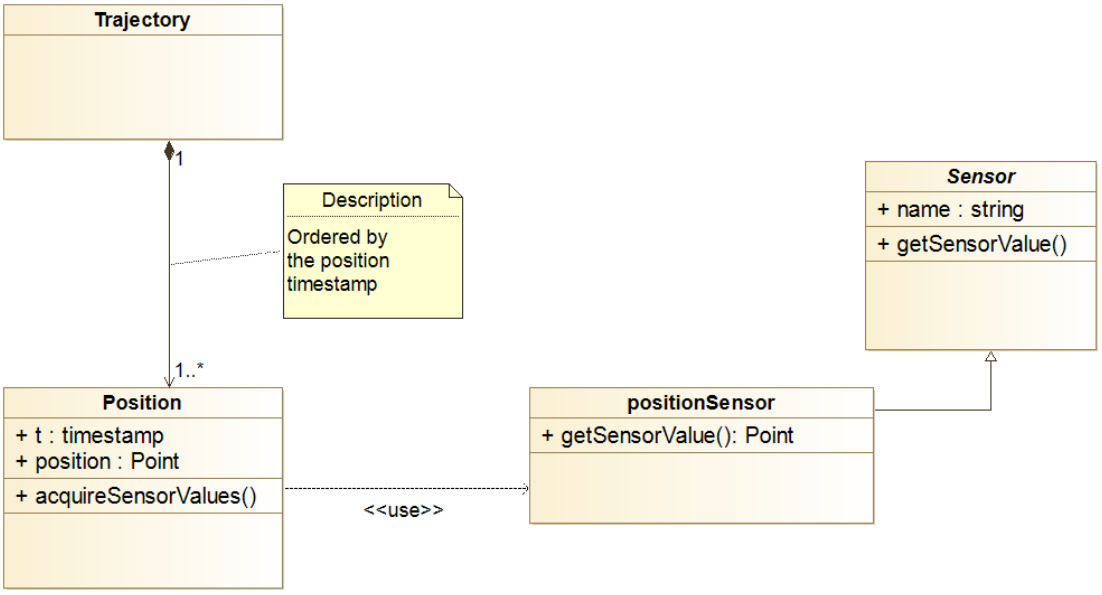


FIGURE 3.4 – Modélisation d'une trajectoire brute

Ainsi, une trajectoire brute peut être modélisée par une composition d'un ensemble de positions ordonnées temporellement. Chaque position ayant une estampille temporelle ainsi que des coordonnées issues d'un capteur GPS obtenues au moment déterminé par l'estampille temporelle. Le modèle est alors proposé dans la figure 3.4.

Cependant, les objets mobiles étudiés sont équipés, en plus du capteur GPS classique, de capteurs génériques qui renseignent l'état et l'utilisation du véhicule au moment t_i (frein à main, vitesse à l'ordinateur de bord, clignotants...) ainsi que de capteurs liés à l'activité du véhicule (compteur de taxi, sirène d'une ambulance, charge d'une benne à ordure...).

3.2.2 Trajectoire brute étendue

Définition 2 (Trajectoire brute étendue) *Une trajectoire brute étendue* $T^+ = \langle e_1^+, ..., e_n^+ \rangle$*, est définie à partir d'une trajectoire brute* T *où, pour chaque élément* $e_i^- \in T$*, un ensemble de valeurs de capteurs* $D_i = (B_i, C_i)$ *est associé tel que* $e_i^+ = (t_i, p_i, D_i)$*. Ces capteurs peuvent être binaires* B_i *tel que* $B_i = (b_1, ..., b_p) \in \{0,1\}^p$ *ou numériques* C_i *tel que* $C_i = (c_1, ..., c_m) \in \mathbb{R}^m$*.*

Exemple 2 *L'exemple suivant est illustré dans la figure 3.5. Par exemple, nous pouvons disposer de l'ensemble* S *de 5 capteurs :*

- *Capteurs binaires*
  - $b_1 =$ *Gyrophares*
  - $b_2 =$ *Sirène*
  - $b_3 =$ *Frein à main*
- *Capteurs numériques*
  - $c_1 =$ *Vitesse compteur*
  - $c_2 =$ *Pression sur la pédale d'accélérateur*

La trajectoire brute étendue T^+*, issue de* T *est définie telle que* $T^+ = \langle e_1^+, e_2^+, e_3^+, e_4^+, e_5^+ \rangle$

- $e_1^+ = (t_1, p_1, B_1, C_1) = (1, (0,0), (0,0,1), (0,0))$ *traduit qu'à l'instant 1, le véhicule est en position* $(0,0)$*, les gyrophares sont éteints, la sirène également, le frein à main est tiré, le véhicule ne roule pas, et la pédale d'accélérateur n'est pas enfoncée.*
- $e_2^+ = (t_2, p_2, B_2, C_2) = (2, (0,0), (0,0,1), (0,0))$ *traduit qu'à l'instant 2, le véhicule est dans le même état qu'à* t_1*.*
- $e_3^+ = (t_3, p_3, B_3, C_3) = (3, (0,0), (0,0,0), (0,10))$ *traduit qu'à l'instant 3, le véhicule est en position* $(0,0)$*, les gyrophares sont éteints, la sirène également, le frein à main est relâché, le véhicule ne roule pas, mais la pédale d'accélérateur est légèrement enfoncée (10%).*
- $e_4^+ = (t_4, p_4, B_4, C_4) = (4, (1,2), (1,1,0), (46,45))$ *traduit qu'à l'instant 4, le véhicule est en position* $(1,2)$*, les gyrophares sont allumés, la sirène également, le frein à main est relâché, le véhicule roule à 46 km/h et l'accélérateur est enfoncé à 45%.*
- $e_5^+ = (t_5, p_5, B_5, C_5) = (5, (3,4), (1,0,0), (75,100))$ *traduit qu'à l'instant 5, le véhicule est en position* $(3,4)$*, les gyrophares sont allumés, la sirène est éteinte, le frein à main est relâché, le véhicule roule à 75 km/h et l'accélérateur est enfoncé à 100%.*

Ainsi, une trajectoire brute étendue peut être modélisée par une composition d'un ensemble de positions étendues ordonnées temporellement. Chaque position ayant une estampille temporelle ainsi que des coordonnées issues d'un capteur GPS et les

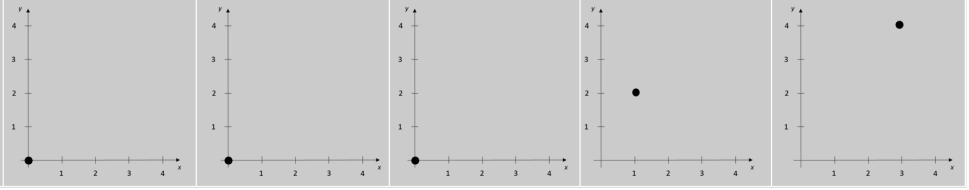
Elément	e_1^+	e_2^+	e_3^+	e_4^+	e_5^+
T_i	1	2	3	4	5
Position	(0,0)	(0,0)	(0,0)	(1,2)	(3,4)
b_1 b_2 b_3	0 0 1	0 0 1	0 0 0	1 1 0	1 0 0
c_1 c_2	0 0	0 0	0 10	46 45	75 100
Représentation					

FIGURE 3.5 – Représentation de la trajectoire brute étendue de l'exemple 2

valeurs des différents capteurs binaires ou non obtenues au moment déterminé par l'estampille temporelle. Il est possible d'ajouter également des capteurs de données ordinales et qualitatives, cependant, afin de ne pas surcharger le modèle, seules les données binaires et numériques continues sont affichées. Le modèle est alors proposé dans la figure 3.6.

Sur ces éléments de trajectoire brute étendue, un processus de pré-traitement des données est effectué afin d'identifier les données spatialement problématiques, de les étiqueter comme aberrantes et de corriger d'autres possibles erreurs. Une fois l'ensemble de ces informations réunies, nous pouvons engager le processus de montée en sémantique faible.

3.3 Trajectoire sémantique

Usuellement, la montée en sémantique faible cherche d'abord à permettre l'identification des arrêts et des déplacements dans la trajectoire observée. Si en analyse de traces GPS sans autre information, des méthodes classiques de segmentation lorsque les points sont récoltés régulièrement est l'utilisation plus ou moins dérivée de la méthode de clustering DBSCAN [Kami et al., 2010, Zimmermann et al., 2009, Andrienko et al., 2013], dans notre contexte, d'autres éléments peuvent permettre d'améliorer la segmentation à l'instar de l'utilisation du frein à main ou de la vitesse. Une discussion sur les différentes approches de segmentation et des choix réalisés est présentée dans la section 5.3.1.3.

Ainsi, la segmentation a pour objectif de construire des sous-trajectoires pouvant ensuite correspondre aux arrêts et déplacements que l'on cherche à identifier.

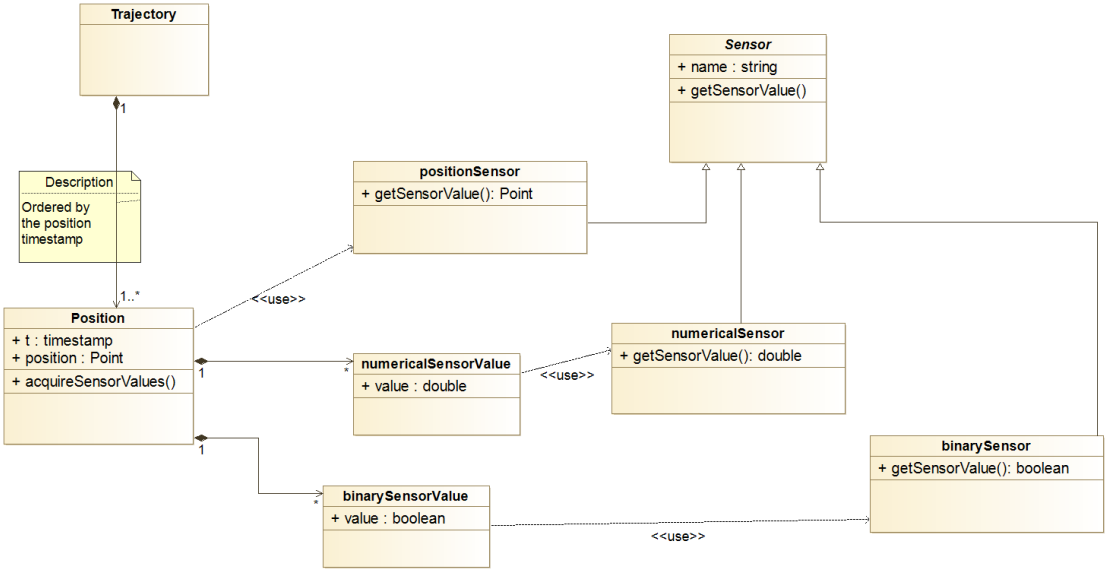


FIGURE 3.6 – Modélisation d'une trajectoire brute étendue

3.3.1 Sous-trajectoire

Pour pouvoir construire une trajectoire sémantique (même faible) à partir de T^+, il faut segmenter cette trajectoire en sous-trajectoires auxquelles nous ajoutons une sémantique sur l'action en cours. Une sous-trajectoire est un séquence d'éléments consécutifs de T.

Définition 3 *Sous-trajectoire T_i est une sous trajectoire de T si et seulement si :*

- *T_i est une trajectoire telle que $T_i = \langle e_{i_1}, ..., e_{i_k} \rangle$, e_{i_j} est une position d'une trajectoire brute ou étendue.*
- *Tous les éléments e_{i_j} de T_i sont des éléments de la trajectoire T : $\{e_{i_j}\}_{j \in [1,k]} \subset \{e_l\}_{l \in [1,n]}$*
- *Tout éléments consécutifs dans T_i le sont aussi dans T, c'est-à-dire qu'il n'existe pas d'éléments de T se situant entre deux éléments de T_i n'appartenant pas à T_i :*

$$\nexists e_j \in T \text{ et } e_j \notin T_i \text{ avec } Rang(T, e_{i_1}) \leq Rang(T, e_j) \leq Rang(T, e_{i_k})$$

Avec $Rang(T, e)$ une fonction indiquant l'indice (ou rang) de la position d'un élément e dans une trajectoire T. Par construction, la valeur retournée appartient aux entiers naturels positifs si l'élément appartient à la trajectoire et vaut -1 sinon. Le rang est déterminé par l'ordre temporel (chronologique) des éléments dans T.

Nous noterons la relation T_i est une sous-trajectoire de T par :

- $T_i \sqsubseteq T$ Si T_i peut être égale à T,

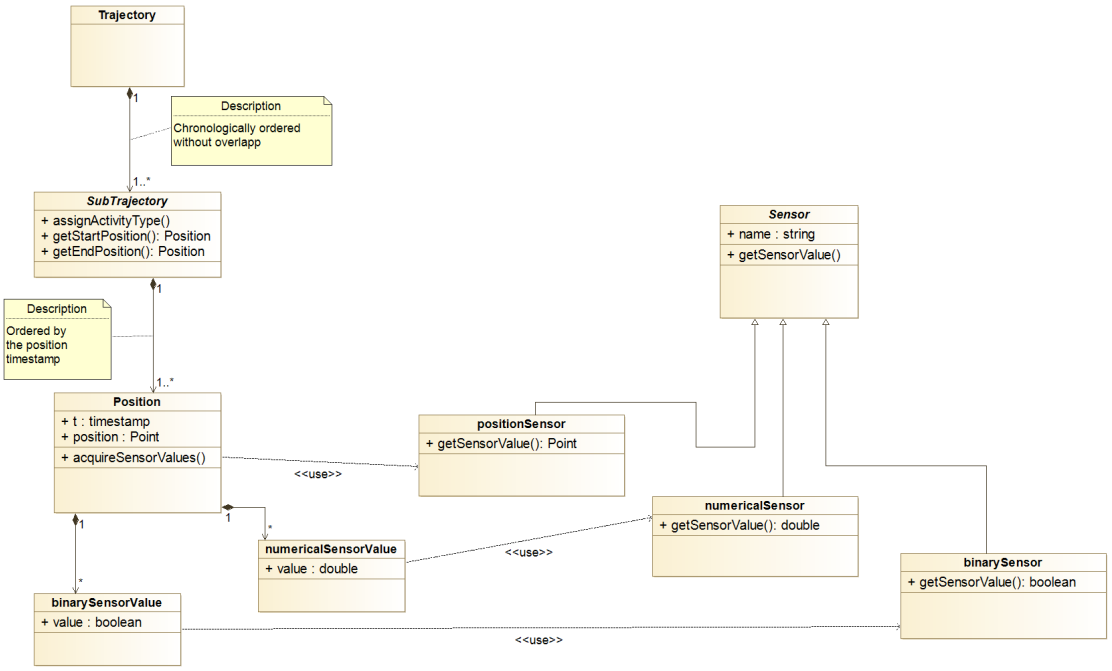


FIGURE 3.7 – Modélisation d'une trajectoire brute étendue avec des sous-trajectoires

— et $T_i \sqsubset T$ si T_i ne peut pas être égale à T.

Ainsi, d'une manière pratique, une sous-trajectoire T_i de T peut être définie comme :

— $T_i = \langle e_{i_1}, ..., e_{i_k} \rangle$

— $\{e_{i_j}\}_{j \in [1,k]} \subset \{e_l\}_{l \in [1,n]}$

— $Rang(T, e_{i_k}) - Rang(T, e_{i_1}) = k - 1$

Avec T_i pouvant être une trajectoire brute ou bute étendue. Formellement, une sous-trajectoire peut ne contenir qu'un seul élément, même si dans la pratique, la question du filtrage de ces sous-trajectoires peut se poser en fonction du contexte.

D'un point de vue UML, nous aurions pu modéliser la notion de trajectoire et de sous trajectoire à l'aide du modèle composant/composite, mais il ne nous semblait pas forcément naturel d'avoir des positions en feuille. Aussi, nous avons choisi de ne pas traiter l'aspect héritage entre sous-trajectoire et trajectoire. Le modèle sous-jacent est proposé dans la figure 3.7.

Ainsi, T^+ est segmentée en k sous-trajectoires consécutives et non chevauchantes telle que $T^+ = \langle T_1^+, ..., T_k^+ \rangle$. T_i^+ est par nature une séquence telle que $T_i^+ = \langle e_{i_1}^+, ..., e_{i_z}^+ \rangle$.

L'idée est de pouvoir associer à chacune de ces sous-trajectoires T_i^+ une sémantique de type STOP ou MOVE. Les STOP peuvent aussi être associés à des POI. La résultante est appelée trajectoire sémantique faible.

3.3.2 Trajectoire sémantique faible

Définition 4 (Trajectoire sémantique faible) *Une trajectoire sémantique faible* $\Sigma^- = \langle s_1^-, ..., s_p^- \rangle$ *est définie comme étant une suite ordonnée d'items* $s_i^- = (sm_i, T_i^+)$ *tels que :*

- $sm_i \in \{\text{"MOVE"}\} \cup \{(\text{"STOP"}, POI^*)\}$
- $T_i^+ \sqsubseteq T^+$
- $T = \langle T_i \rangle_{i \in [1,k]}$
- $\forall (i, i'), i \neq i', T_i^+ \sqsubseteq T^+$ *et* $T_{i'}^+ \sqsubseteq T^+$ *alors* $\{e_i^+\} \cap \{e_{i'}^+\} = \emptyset$

Soit $arret(T)$ une fonction booléenne indiquant si une sous-trajectoire donnée forme un arrêt significatif ou non. Elle peut donc potentiellement servir de filtrage pour la suppression des arrêts non pertinents (e.g. micro-arrets). Cette fonction, est dans le cadre classique de l'analyse de traces GPS, uniquement définie sur les données GPS. Dans notre contexte, si un ensemble de capteurs $\{D^a\}$ inclus dans D est utilisé, elle correspond à une agrégation des fonctions $arret_{d_l^a}$, $l \in [1, \|D^a\|]$, pour chacun de ces capteurs.

$$arret(T) = Ag(arret_{d_l^a}), l \in [1, \|D^a\|] \tag{3.1}$$

L'idée est simple : on agrège des tests de caractérisation de l'arrêt par capteur. Par exemple, si la fonction $arret_{freinMain}$ pour le frein à main renvoie vrai, le résultat du st-dbscan sur le GPS indique faux, $arret_{accelerateur}$ dit faux, mais $arret_{moteur}$ dit vrai, alors on agrège ces données par une fonction (qui peut être un moteur de raisonnement) qui renverra potentiellement vrai à l'issue d'une décision multicritères. Des contraintes métiers peuvent être ajoutées dans la fonction d'agrégation (e.g. durée minimale d'un arrêt).

La construction d'une trajectoire sémantique faible est obtenue par l'utilisation de la fonction *arret* sur chacune des sous-trajectoires : si la fonction vaut vrai, alors c'est un STOP, sinon c'est un MOVE.

Exemple 3 *L'exemple suivant est illustré dans la figure 3.8.*

Soit un ensemble de 2 POI :

- *caserne1*
  - *nom = Caserne des sapeurs pompiers de Tours*
  - *emprise = emprise spatiale de la caserne*
  - *adresse = adresse postale de la caserne*
  - *type = caserne*
- *hopital1*
  - *nom = CHU Trousseau*

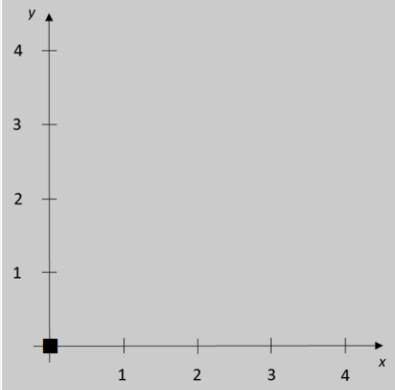
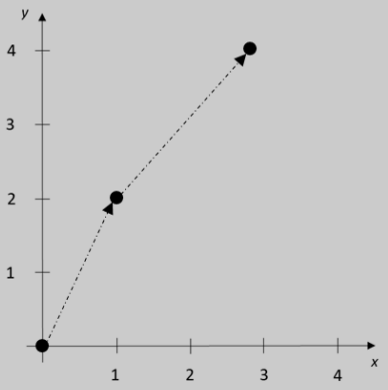
Elément	s_1^-	s_2^-
sm_i	(STOP,"caserne1")	(MOVE)
T_i^+	(e_1^+, e_2^+)	(e_3^+, e_4^+, e_5^+)
Représentation		

FIGURE 3.8 – Représentation de la trajectoire sémantique faible de l'exemple 3

— *emprise = emprise spatiale de l'hôpital*

— *adresse = adresse postale de l'hôpital*

— *type = hopital*

La trajectoire sémantique faible Σ^-*, issue de* T^+ *est définie telle que* $\Sigma^- = \langle s_1^-, s_2^- \rangle$ *avec :*

— $s_1^- = (sm_1, T_1^+)$ *avec*

— $sm_1 = (STOP, "caserne1")$

— $T_1^+ = \langle e_1^+, e_2^+ \rangle$

traduit que le véhicule a effectué un STOP dans le POI "Caserne1" durant la période de la sous-trajectoire T_1^+

— $s_2^- = (sm_2, T_2^+)$ *avec*

— $sm2 = (MOVE)$

— $T_2^+ = \langle e_3^+, e_4^+, e_5^+ \rangle$

traduit que le véhicule a effectué un MOVE durant la période de la sous-trajectoire T_2^+

Dans l'exemple précédent, e_3^+ est considéré dans le MOVE à l'issu de notre exemple de segmentation, car la pédale d'accélérateur est enfoncée et le frein à main relâché. Une autre segmentation est tout à fait possible. Le choix de la segmentation

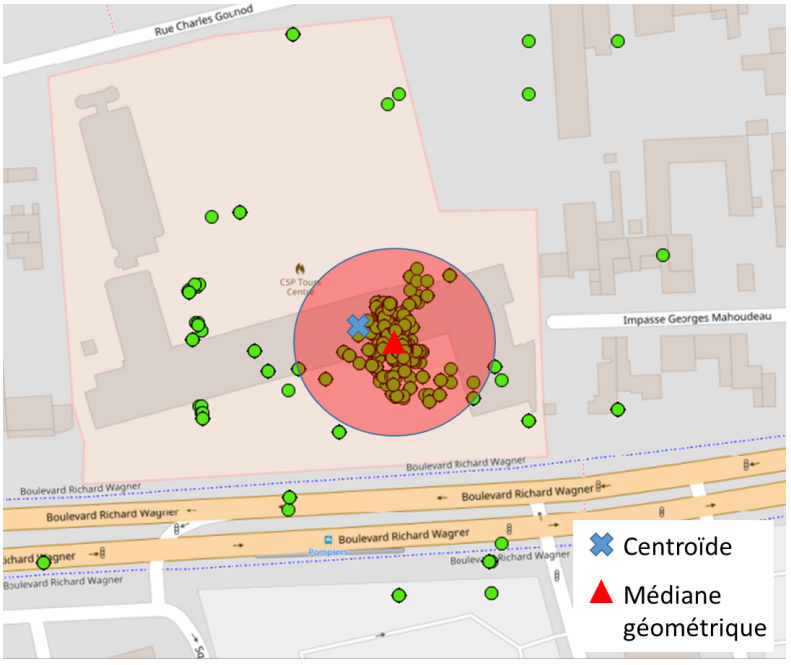


FIGURE 3.9 – Exemple de représentation d'un STOP

a donc une incidence sur les positions pivots (de début et fin de sous-trajectoires). Aussi, certains indicateurs statistiques issus de l'analyse de ces sous-trajectoires peuvent donc changer.

Par ailleurs, lorsqu'un objet mobile est à l'arrêt, il ne devrait occuper normalement qu'une seule et une unique localisation durant toute la période d'arrêt. Cependant, du fait du bruit lors des captations des données GPS, cela est très rarement le cas. Dans le but de respecter la sémantique de l'arrêt, une fois que la période d'arrêt est déterminée, nous lui affectons une position étendue représentative et robuste de l'ensemble des positions acquises durant cette période ainsi que deux propriétés pour modéliser la qualité de la représentation. Statistiquement, la médiane est une valeur robuste et représentative d'un échantillon.

La médiane géométrique proposée par [Weiszfeld and Plastria, 2009] a été retenue. Cette dernière est moins sensible au bruit qu'un centroïde. Elle a donc été choisie pour représenter l'emplacement d'un STOP. Dans notre approche, elle est calculée uniquement sur les positions GPS non étiquetées comme aberrantes par le processus de pré-traitement. [Small, 1990] présente un état de l'art sur la représentation de médianes multidimensionnelles. En plus de la médiane géométrique, deux propriétés sont calculées pour représenter la qualité de la représentation. Cette représentation sur un cas réel de STOP est présentée dans la figure 3.9.

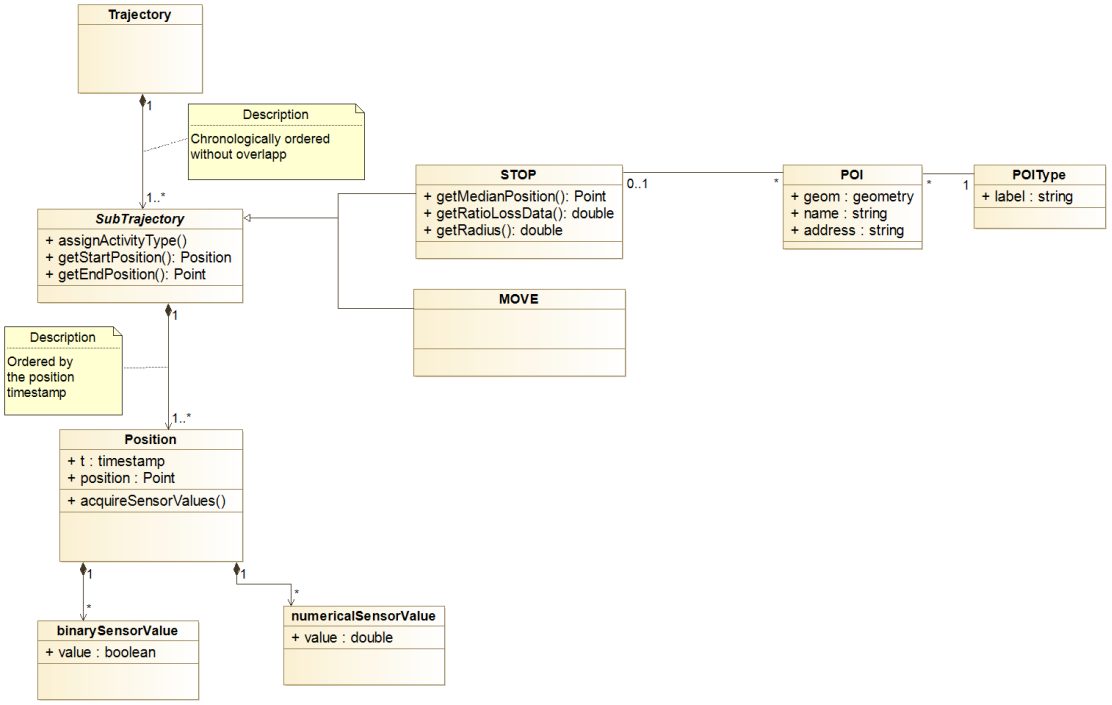


FIGURE 3.10 – Extrait du diagramme de classe d'une trajectoire sémantique faible

La première propriété est le rayon du cercle R_{med}, en rouge sur la figure 3.9, centré sur la médiane géométrique (triangle rouge de la même figure). Elle décrit la dispersion des valeurs captées. Ce cercle contient au moins 95% des points non considérés comme aberrants par le processus de pré-traitement. Ce seuil a été retenu par analogie à la précision des GPS [Chin et al., 1987, William et al., 2014].

La seconde propriété est le pourcentage de données aberrantes ($\%_{DP}$) qui constitue un STOP.

$$\%_{DP} = \frac{\text{nombre de positions aberrantes}}{\text{nombre de positions total}} \tag{3.2}$$

Un extrait du diagramme de classe UML correspondant à la modélisation des trajectoires sémantiques faibles est proposé dans la figure 3.10. Ici, la sous-trajectoire devient une classe abstraite, et seules les classes STOP et MOVE héritées de SubTrajectory peuvent donc être instanciées. Cela permet de ne considérer une trajectoire que comme une séquence ordonnée temporellement d'objets STOP et d'objets MOVE. Les POI sont associés au STOP. À chaque POI est associé un type.

3.3.3 Trajectoire sémantique forte

Définition 5 (Trajectoire sémantique forte) *Une trajectoire sémantique forte* $\Sigma = \langle s_1, ..., s_p \rangle$ *est définie à partir d'une trajectoire sémantique faible* Σ^- *où, pour chaque item* $s_i^- \in \Sigma^-$*, une activité est associée tel que* $s_i = (s_i^-, a_i)$ *avec* $a_i \in A$ *où* A *est un ensemble d'activités possible pour l'objet mobile.*

A peut être une ontologie ou une liste d'activités ayant un sens précis au niveau du métier pour l'objet mobile étudié. Le choix de l'ontologie permet d'avoir des relations sémantiques plus complexes entre les activités.

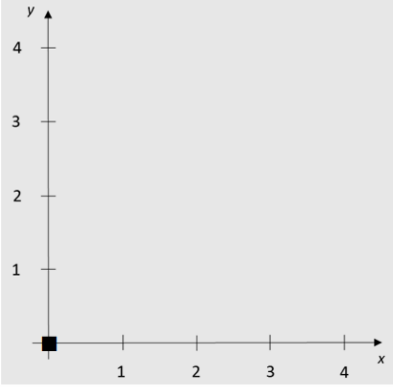
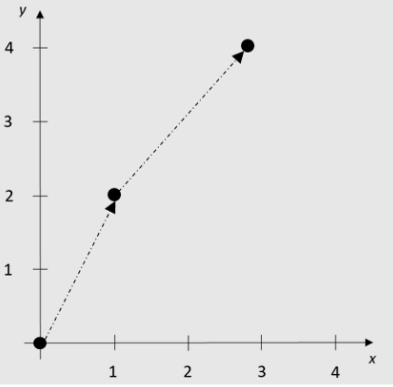
Elément	s_1^+	s_2^+
sm_i	(STOP,"caserne1")	(MOVE)
T_i^+	(e_1^+, e_2^+)	(e_3^+, e_4^+, e_5^+)
Activité	Attente caserne	Départ vers intervention
Représentation		

FIGURE 3.11 – Représentation de la trajectoire sémantique forte de l'exemple 4

Exemple 4 *L'exemple suivant est illustré dans la figure 3.11. Soit un ensemble A de 3 activités :*

- *Attente caserne*
- *Départ vers intervention*
- *Attente hôpital avant prise en charge patient*

La trajectoire sémantique forte Σ *issue de* Σ^- *est défini telle que* $\Sigma = \langle s_1^+, s_2^+ \rangle$ *avec :*

- $s_1^+ = (s_1^-,$ *"Attente caserne"*$)$
- $s_2^+ = (s_1^-,$ *"Départ vers intervention"*$)$

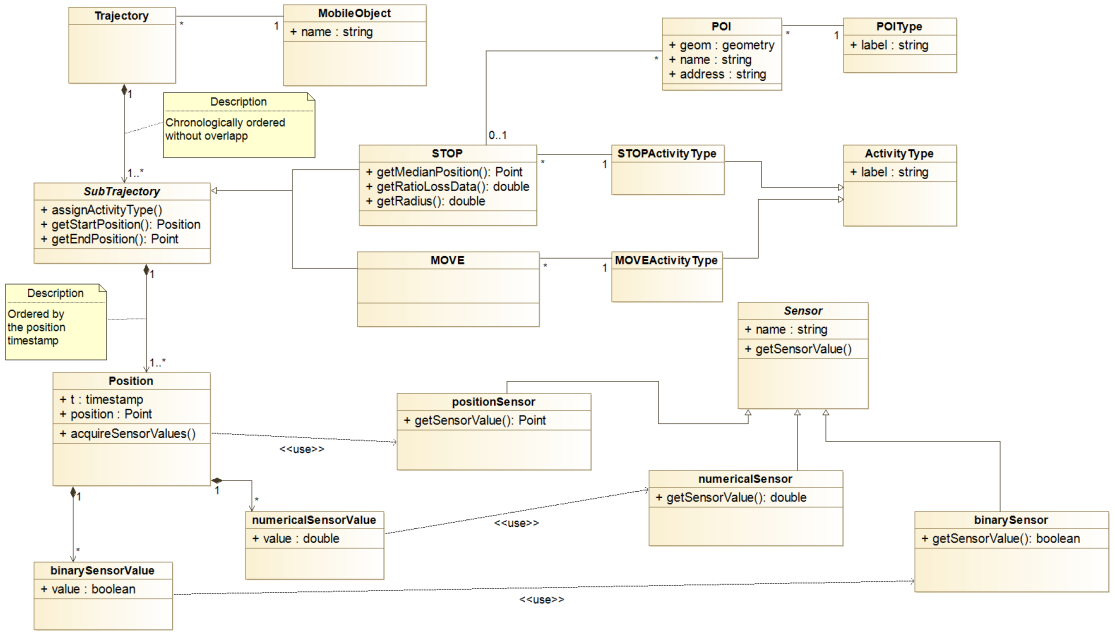


FIGURE 3.12 – Diagramme de classes UML des trajectoires sémantiques fortes des véhicules

Le modèle complet est présenté dans la figure 3.12. L'association des sémantiques aux différentes sous-trajectoires se fait à l'aide de la méthode assignActivityType. Cette assignation, non implémentée dans la classe abstraite SubTrajectory, est implémentée dans les classes héritées STOP et MOVE en lien avec respectivement les classes STOPActivityType et MOVEActivityType. Ces deux classes héritent de la classe abstraite ActivityType. D'autres choix de modélisation sont sans aucun doute possibles. Cependant, cette modélisation offre la capacité d'être suffisamment générique pour être exploitée/adaptée dans d'autres contexte que le notre. Par ailleurs, elle autorise d'avoir plusieurs STOP consécutifs non séparés par des MOVE, ainsi que plusieurs MOVE consécutifs non séparés par des STOP. Cela permet pour les STOP, de, par exemple, pouvoir dissocier durant un STOP la phase pure d'attente avec une phase de logistique à l'instar d'une vérification du matériel. La succession de MOVE permet de pouvoir modéliser de possibles reroutages qui, dans notre contexte, peuvent être la mobilisation d'une ambulance vers une nouvelle intervention lors de son trajet retour vers la caserne.

Nous nous sommes concentrés sur la modélisation des trajectoires sémantiques à partir d'une trajectoire brute. Comme plusieurs véhicules sont en jeu, et que chaque véhicule est amené à fournir plusieurs trajectoires différentes, mais que chaque trajectoire ne concerne qu'un seul et unique véhicule, une liaison entre la classe trajectoire et la classe objet mobile peut être ajoutée.

3.4 Processus opérationnel



FIGURE 3.13 – Schéma étendu de la montée en sémantique forte

Le pipeline général de la montée en sémantique est schématisé dans la figure 3.13. Trois phases se succèdent, une première consiste à la préparation des trajectoires brutes étendues et des lieux d'intérêt propres au métier étudié, une deuxième porte sur la montée en sémantique faible basée sur la première phase, et la dernière permet la montée en sémantique forte des trajectoires sémantiques faibles obtenues à l'issue de la précédente étape.

Ainsi, afin d'extraire nos trajectoires sémantiques fortes, la première étape de la montée en sémantique porte sur le pré-traitement des données.

3.4.1 Pré-traitement : extraction des trajectoires brutes étendues et de la liste des POI

L'étape de pré-traitement, appelée "Traitement et qualité des données brutes", correspond à ces étapes :

— Extraction des POI en rapport avec l'activité de l'objet mobile

— Extraction des trajectoires brutes étendues

— Identification et étiquetage des données aberrantes

L'étape d'extraction des POI est indépendante des étapes d'extraction des trajectoires brutes étendues et d'identification et d'étiquetage des données aberrantes. Par contre, ces deux dernières doivent s'exécuter de manière consécutive. Ce processus de traitement est détaillé dans la figure 3.14.

3.4.1.1 Extraction des POI

Tout d'abord, pour pouvoir monter nos trajectoires en sémantique faible, il est nécessaire de récupérer une liste de points d'intérêts (POI) métier, avec leur type et leur représentation spatiale, pouvant être associés aux différents STOP de l'objet mobile étudié.

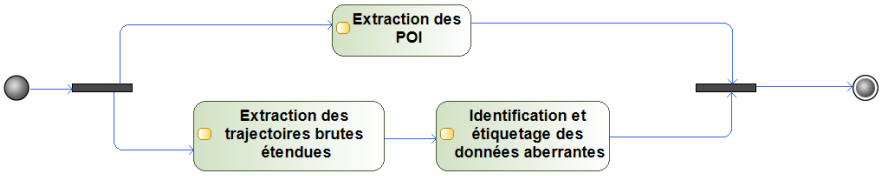


FIGURE 3.14 – Processus de pré-traitement

Cette phase consiste à identifier le lieu d'intérêt à l'aide de connaissances externes. Des sources ouvertes telles qu'OpenStreetMap (OSM) ou professionnelles (BD TOPO ® IGN) peuvent être employées à cet effet. Dans notre cadre, nous avons utilisé les données OSM du fait de leur gratuité d'utilisation, de la liberté d'accès et de leur bonne qualité globale [Mooney and Minghini, 2017].

Pour cette extraction, il faut définir un ensemble de lieux potentiellement intéressants. Les objets géographiques d'OSM sont caractérisés par des couples clé et valeur. Le nom de la clé étant la catégorie de l'objet. Dans notre cas, la clé amenity [1], qui regroupe différents types d'installations communautaires (toilettes, téléphones, bâtiments...) est la plus utilisée. Les clés sont associées à des valeurs qui correspondent à des sous-catégories. Par exemple, la valeur taxi de la clé amenity identifie une station de taxis. Pour OSM, un ensemble d'objets géographiques pertinents peut être ainsi extrait et localisé à l'aide d'une liste de couples clé-valeur prédéfinis. Des POI externes sont ainsi extraits et localisés.

Un processus d'appariement est alors lancé afin d'associer les STOP à des POI externes. [Beeri et al., 2004] décrit un grand nombre de processus possibles d'appariement entre objets spatiaux issus de deux jeux de données différents. Comme nos STOP sont géographiquement modélisés via leur médiane géométrique, les appariements à mettre en oeuvre, selon la forme des POI extraits, sont de types : point-point, point-ligne, point-polygone, etc.

3.4.1.2 Extraction des trajectoires brutes étendues

Parallèlement à l'extraction des POI, nos véhicules sont équipés d'un système d'ensemble de capteurs. Nous récupérons à chaque timestamp (en général 1 par seconde), les données issues des différents capteurs mis en place. L'horloge est celle du boîtier assurant la remontée de la récolte afin d'éviter d'avoir des problèmes d'alignements temporels entre les différents capteurs.

Pour les capteurs ayant une fréquence différente de notre méthode de relevé (60Hz), le matériel propriétaire agrège les différentes valeurs captées, pour ne récupérer qu'une valeur représentative pour chaque timestamp. Cette étape forme une sorte de boîte noire que nous ne pouvons modifier.

1. http ://wiki.openstreetmap.org/wiki/Key :amenity



FIGURE 3.15 – Processus de montée en sémantique faible

Nous enregistrons à la volée les messages sous la forme suivante :

$$< date, heure, latitude, longitude, b1, ..., c1, .. > .$$

En concaténant ces messages lors de leur stockage, nous obtenons nos trajectoires brutes étendues pour chacun des véhicules.

3.4.1.3 Identification et étiquetage des données aberrantes

Une fois nos données acquises, une phase de traitement des problèmes de qualité des données est nécessaire. Pour cela, un ensemble de processus doit être mis en œuvre pour possiblement corriger certains problèmes mineurs dans les données, mais principalement pour identifier les données spatialement problématiques qui seront identifiées et non supprimées comme aberrantes.

Les données considérées comme aberrantes ne sont pas supprimées car les autres informations captées au même instant ne sont pas forcément problématiques. Par contre, étiqueter les données aberrantes permet de ne pas en tenir compte dans les indicateurs purement spatiaux. Ces traitements des problèmes de qualité sont dépendants des données étudiées et de leur systèmes d'acquisition. Le chapitre 4 présente à la fois notre cas d'étude et les pré-traitements mis en oeuvre.

3.4.2 Montée en sémantique faible

Le processus de montée en sémantique faible est décomposé en 3 étapes. Dans un premier temps, il faut segmenter les trajectoires brutes étendues en sous trajectoires. Ensuite, les STOP et MOVE sont extraits à l'aide de la fonction *arret*. Les micro-arrêts et micro-déplacements sont fusionnés avec les deux sous-trajectoires les encadrant, afin d'éviter d'avoir des déplacements ou des arrêts qui n'en sont pas réellement (feux rouges, erreurs de capteurs, etc.). Et enfin, la dernière phase est l'appariement entre les STOP et les POI. La figure 3.15 présente le diagramme d'activité de la montée en sémantique faible.

Les objets sortis de l'étape de montée en sémantique faible sont donc des sous-trajectoires ayant des sémantiques qui peuvent être soit des MOVE, soit des STOP.

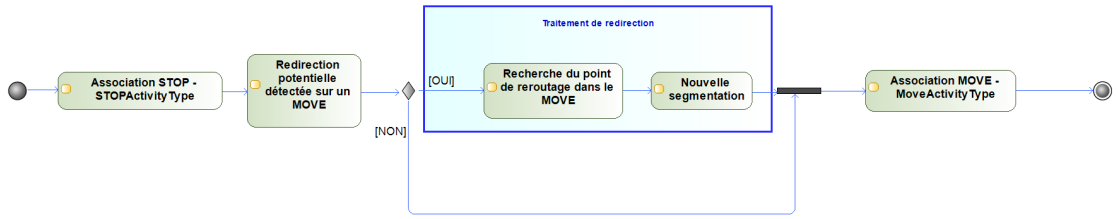


FIGURE 3.16 – Diagramme d'activités de la montée en sémantique forte à partir d'une trajectoire sémantique faible.

3.4.3 Montée en sémantique forte

La suite du processus est la montée en sémantique forte d'une sous trajectoire exprimée en sémantique faible. Cette montée en sémantique forte s'effectue en trois phases (cf. 3.16).

La première phase est tournée vers l'étiquetage des STOP. En fonction de la méthode de raisonnement utilisée, une activité parmi les activités de type STOPActivityType est choisie et associée au STOP.

La seconde phase concerne la détection de reroutage potentiel durant les MOVE. En effet, une fois la montée en sémantique des STOP effectuée, il est possible de déterminer si le véhicule a effectué deux activités différentes durant un même MOVE. Par exemple, un véhicule de secours partant d'un hôpital retourne normalement vers sa caserne d'affectation. Cependant, durant son retour, il peut avoir à changer de destination pour repartir en intervention. Un cas comme celui-ci est ensuite analysé plus précisément afin de détecter le moment ou le véhicule à été rerouté. Si le reroutage est avéré, la sous trajectoire est segmentée en deux sous trajectoires de type MOVE. Ces deux sous trajectoires sont alors traitées dans l'étape suivante.

La dernière étape de ce processus est l'étiquetage des MOVE. En entrée de ce processus, nous avons des sous trajectoires de type STOP ou MOVE, les STOP étant montés en sémantique forte, et les MOVE en sémantique faible. Comme pour le processus d'étiquetage des STOP, les étapes sont le choix de l'activité puis l'affectation de cette dernière au MOVE. L'étape d'association est dépendante de la méthode de raisonnement utilisée pour choisir quelle activité de type MOVEActivityType doit être associé au MOVE en question.

L'intégralité de ce processus de montée en sémantique forte à partir de l'acquisition des données GPS est schématisée dans le diagramme d'activité présenté dans la figure 3.17.

À l'issu de ces traitements, nous obtenons donc une trajectoire sémantique forte pour chaque trajectoire analysée. Ainsi, chaque trajectoire issue de chaque véhicule peut être montée en sémantique forte.

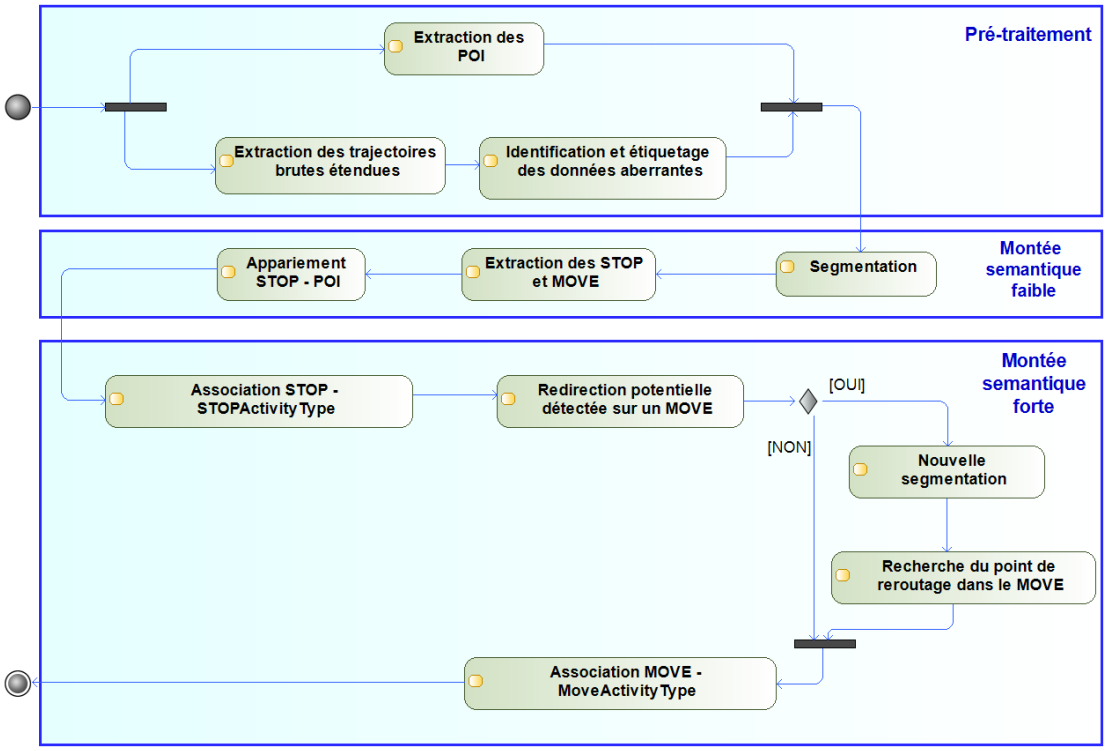


FIGURE 3.17 – Diagramme d'activités du processus global de la montée en sémantique forte à partir des données d'objets mobiles connectés

3.5 Discussion

Le modèle présenté autorise une segmentation plus fine des trajectoires sémantiques. Contrairement aux modèles classiques, [Parent et al., 2013], où seuls les modes de déplacement sont ajoutés dans le cas d'un MOVE et un POI dans le cas d'un STOP, notre modèle, dans une démarche d'extraction de comportements, permet une analyse sémantique plus fine des trajectoires en permettant l'ajout d'activités sur les déplacements ainsi que sur les arrêts.

Par exemple, les véhicules spécifiques ont, dans le cadre de véhicules de secours, des types de déplacement qui dépendent du métier : transfert rapide d'un patient, transfert délicat, déplacement logistique. Également, un STOP peut être associé à une zone d'arrêts fréquents.

La richesse des capteurs permet aussi la détection de changement de comportement dans un même type de trajectoire sémantique (STOP ou MOVE). C'est pourquoi il est possible dans ce modèle de se faire succéder deux MOVE ou deux STOP.

Par exemple, dans le cas d'un STOP à l'hôpital,lors de l'étude de trajectoires pour un véhicule de secours, il est possible via l'étude de différents capteurs (porte arrière, descente du brancard ...) de diviser une sous-trajectoire sémantique faible de type STOP en plusieurs sous-trajectoires sémantiques fortes également de type STOP, mais avec deux activités différentes : la première serait "Attente hôpital avant descente du patient" et la seconde "Attente hôpital après descente du patient".

Dans le cas d'un MOVE également, il est possible d'avoir deux activités distinctes pour un même épisode. Par exemple, un véhicule de retour à la caserne après avoir déposé un patient à l'hôpital peut être appelé par radio pour être redirigé vers un lieu d'intervention. Là encore, l'analyse des capteurs permet de détecter ce changement de comportement (utilisation soudaine et continue des gyrophares, augmentation de la vitesse du véhicule par rapport aux limites légales ...). Comme dans le cas précédent, une sous trajectoire brute de type MOVE pourra ainsi être divisée en deux sous trajectoires sémantiques de type MOVE, mais avec des activités différentes.

Ainsi, le modèle présenté s'inscrit dans un objectif de fouille de trajectoires d'objets mobile. L'approche présentée dans ce chapitre est suffisamment générique pour être appliquée à différents types d'objets mobiles. Par exemple, ambulances, véhicule de livraison, véhicule de collectes d'ordures, etc.

En effet, les valeurs de capteurs associées à l'activité de l'objet mobile sont prises en compte dans ce modèle et permettent de déterminer plus précisément l'activité de l'objet mobile durant un STOP ou un MOVE.

Également, la liste d'activité A, étant soit une ontologie soit une liste d'activités métier, définie suivant l'objet d'étude, permet d'étudier plusieurs types d'objets mobiles. Cependant, la montée en sémantique des trajectoires est fortement contrainte par la liste des activités définies (ou de l'ontologie utilisée), ainsi que de la méthode choisie pour associer une activité à une sous trajectoire.

D'autre part, l'utilisation de multi-capteurs implique de devoir intégrer l'utilisation de ces données dans l'association d'une activité à une sous trajectoire.

3.6 Conclusion

Nous avons présenté dans ce chapitre une extension du modèle de [Parent et al., 2013]. Le modèle ainsi qu'un processus formel du passage d'une trajectoire brute à une trajectoire sémantique à été présenté au travers de différentes étapes. Chaque étape est détaillée et illustrée à travers différents exemples.

Dans les modèles classiques, seules les coordonnées GPS sont utilisées pour construire des trajectoires sémantiques. Des POI sont associés et les modes de déplacements sont déduits des traces GPS. La première étape de notre processus est l'ajout de données issues de capteurs placés sur l'objet mobile étudié. Puis, une fonction prenant en compte les capteurs disponibles est utilisée pour segmenter les trajectoires en sous trajectoires. Chaque sous trajectoire est ensuite qualifiée sémantiquement pour déterminer si c'est un STOP ou un MOVE. Les STOP sont modélisés avec des propriétés permettant de déterminer la précision de la localisation de ceux-ci. Ensuite, à l'aide d'un système de raisonnement, chaque sous trajectoire est associée sémantiquement à une activité issue d'une ontologie ou d'une liste activités définie par des experts métiers de l'objet mobile étudié.

Dans ce mémoire, le processus de montée en sémantique, pour notre problématique métier sur les trajectoires d'ambulances connectées, est illustré dans le chapitre 5 pour des trajectoires brutes étendues contenant les données GPS, temporelles ainsi que l'état du frein à main. Le chapitre 6 en présente une mise en place sur un plus grand nombre de capteurs. Ces montées en sémantique sont dépendantes du cas d'étude et des pré-traitements présentés dans le chapitre suivant.

Chapitre 4

Cas d'étude et phase de pré-traitement

4.1 Introduction

Dans cette thèse, notre problématique métier consiste en l'identification des différentes interventions effectuées par les ambulances connectées des pompiers et leur caractérisation sémantique, ce qui s'inscrit dans notre problématique scientifique autour de la montée en sémantique de données de trajectoires issues de capteurs embarqués dans les véhicules.

À cette fin de montée en sémantique, comprendre les logiques métiers sous-jacentes aux déplacements et arrêts des ambulances est fondamental et est l'objet de la première partie de ce chapitre.

Notre étude s'appuie sur une campagne de récolte de données effectuées du 31 décembre 2016 au 17 janvier 2017 sur un ensemble d'ambulances connectées des sapeurs-pompiers d'Indre-et-Loire. Cette récolte est effectuée à l'aide d'un système propriétaire qui a été développé dans un objectif de visualisation des déplacements. Ce système offre de nombreuses possibilités en termes de capteurs notamment, mais présente aussi des limites importantes qui seront présentées dans ce chapitre. Les capteurs récoltés sont riches d'information pour des processus de montée en sémantique, notamment ceux présentés dans les deux prochains chapitres. Cependant, il faut pour cela aussi tenir compte des limites du système d'acquisition.

Conformément au protocole établi dans le chapitre 3.1 pour les processus de montée en sémantique présentés dans les prochains chapitres, afin de tenir compte de ces limites d'une part, mais aussi de permettre de trouver les lieux d'intérêt d'autre part, nous détaillerons dans la dernière partie de ce chapitre la phase de pré-traitement des données. Pour cela, un processus de pré-traitement générique est proposé, puis son adaptation sur les véhicules de notre étude ainsi que les résultats obtenus seront présentés.

Symbole	Sémantique
Symbologie des types de STOP	
	Prise en charge du patient sur lieu d'intervention
	Prise en charge du patient à l'hôpital
	Activité de Logistique
	Attente caserne
Symbologie des types de MOVE	
	Départ vers lieu d'intervention
	Transport vers hôpital
	Retour caserne
	Trajet logistique

TABLE 4.1 – Symbologie des STOP et MOVE métiers

Ce chapitre rappelle donc le métier de notre étude, ainsi que les structures des différentes trajectoires que nous désirons extraire, le système permettant l'acquisition des données et enfin la phase de pré-traitement sur les données récoltées.

4.2 Déplacements des ambulances

Dans cette section, nous présenterons les différents déplacements et arrêts typiques effectués lors d'une tournée d'une ambulance. Nous considérons une tournée comme étant le parcours d'une ambulance partant d'une caserne jusqu'à son retour à celle-ci. La symbologie utilisée dans les figures illustrant les motifs des tournées des ambulances présentées ci-après est donnée dans le tableau 4.1.

Afin de pouvoir extraire des trajectoires pertinentes en relation avec le métier, nous avons étudié le fonctionnement d'une ambulance de pompiers. Plusieurs réunions avec les experts métiers nous ont permis d'extraire le fonctionnement propre aux urgences des pompiers, ainsi que des règles et habitudes propres à leur métier.

Tout d'abord, chaque département possède un centre de traitement et de régulation de l'alerte (CETRA). Ce centre réceptionne les appels passés aux secours (SAMU, pompiers...). Ce centre redirige ensuite, selon la nécessité, vers le SAMU, des ambulanciers privés, ou les pompiers. Les ambulances de pompiers ne reçoivent donc pas directement d'appels, mais sont « engagées » sur des interventions.

Chaque caserne de pompiers possède une ou plusieurs ambulances, suivant la taille du secteur de couverture de celle-ci. Cependant, une ambulance n'est affectée qu'a une seule caserne au même instant. C'est pourquoi lorsqu'un appel est reçu, l'opérateur qui engage les moyens sélectionne la caserne la plus proche du lieu d'intervention ayant une ambulance disponible (dans le cas où une ambulance est requise sur l'intervention). Le déclenchement d'un appel est un processus rapide, en général moins de deux minutes entre la réception de l'appel au 112 et l'engagement d'un véhicule de secours.

De ces faits, nous pouvons décrire une règle métier importante. Une ambulance de pompier stationne usuellement dans une caserne, dans l'attente d'être engagée sur une intervention.

Lorsque l'ambulance est engagée, elle se rend directement sur le lieu d'intervention, afin de porter secours et assistance à une victime potentielle. Dès le départ de la caserne, via un appareil spécifique, un statut est affecté manuellement par les pompiers à l'ambulance (départ vers intervention), ce qui a pour effet de la rendre indisponible au niveau du CETRA. Une fois arrivés sur les lieux, les pompiers indiquent via l'appareil qu'ils sont arrivés sur le lieu d'intervention. Dès lors, ils prennent en charge la victime, et effectuent des tests de vérification (prise de tension, saturation, rythme cardiaque...) afin d'évaluer l'état du patient. Suivant les conditions, ces tests peuvent se faire directement dans l'ambulance (c'est souvent le cas lors des accidents sur la voie publique), ou au domicile de la victime. Ces informations sont ensuite transmises à un médecin du SAMU qui juge, alors, de la nécessité ou non de transférer le patient à l'hôpital.

Si la victime ne nécessite pas de prise en charge, le statut du véhicule est défini sur disponible et celui-ci rentre dans sa caserne d'affectation. La figure 4.1 illustre une tournée la plus simple d'une ambulance, à savoir, le cas où le patient n'est pas transporté à l'hôpital, dans ce cas, le véhicule part donc de la caserne vers le lieu d'intervention, puis, une fois le patient pris en charge, retourne à la caserne :



FIGURE 4.1 – Tournée typique 1 d'une ambulance pompier

Si la victime doit être transportée à l'hôpital, elle est installée sur un brancard, puis amenée dans l'ambulance (ou bien elle s'y installe d'elle même si son état le lui permet). Lors du transport, les pompiers indiquent via un bouton qu'ils transportent le patient à l'hôpital.

Lors de l'arrivée à l'hôpital, les pompiers indiquent via un bouton spécifique qu'ils y sont arrivés, et un processus strict de prise en charge est respecté. En effet, chaque patient transféré aux urgences doit être inscrit à l'accueil. C'est pourquoi, les pompiers doivent tout d'abord inscrire le patient lors de son arrivée aux urgences, puis ensuite attendre sa prise en charge par une infirmière du service des urgences. Une fois le patient pris en charge, les pompiers sont, à nouveau, disponibles.

Lors de cette attente, il est de coutume qu'un des pompiers prépare le véhicule pour sa prochaine intervention (désinfection des zones en contact avec le patient, changement des draps du brancard...). Également, lors de cette attente, il est possible que l'ambulance soit déplacée, dans le cas où le sas des urgences soit rempli. L'ambulance est alors déplacée puis garée à l'extérieur du sas, en attendant la prise en charge du patient.

Lorsque le patient a été pris en charge, deux cas sont possibles. Le premier, si la cellule sanitaire à besoin d'être désinfectée ou nettoyée entièrement, ou si le véhicule a besoin d'être réarmé (rechargé en oxygène, compresses, ou tout autre matériel médical), il se dirige alors vers le CETRA, qui dispose d'un sas spécifique pour accueillir les ambulances à cet effet. Cependant, la caserne de notre étude étant une très grosse caserne, celle-ci dispose aussi des équipements nécessaires pour réarmer le véhicule. Si le véhicule n'a pas besoin d'être désinfecté, ou, une fois que celui-ci l'a été, le statut du véhicule est alors défini sur disponible.



FIGURE 4.2 – Tournée typique 2 d'une ambulance pompier avec prise en charge à l'hôpital

La figure 4.2 présente une trajectoire d'une tournée typique 2 d'une ambulance de pompier, où le véhicule qui stationnait en caserne est engagé sur une intervention. Une fois arrivé sur le lieu d'intervention, le patient est pris en charge, puis emmené à l'hôpital. Une fois le patient pris en charge par le service des urgences, le véhicule rentre à sa caserne d'affectation.

Également, les véhicules pouvant être appelés à toute heure, à une distance parfois éloignée (suivant la disponibilité des ambulances aux alentours), le plein de carburant est toujours effectué lorsque la charge descend en dessous de 75% de la capacité maximum du réservoir. Les pleins de carburant sont effectués lorsque le véhicule est disponible, la plupart du temps sur un trajet retour vers la caserne. Les véhicules étant équipés de carte carburant, c'est souvent la même station-service

Tournée typique 3



Tournée typique 4



FIGURE 4.3 – Tournées typiques 3 et 4 d'une ambulance pompier avec un arrêt logistique avant le retour à la caserne

(parfois deux suivant la typologie du secteur) qui est utilisée. La figure 4.3 présente deux tournées similaires à celles des figures 4.1 et 4.2, mais où l'ambulance effectue un arrêt logistique avant de rentrer à la caserne.

Le véhicule peut être engagé sur une intervention dès qu'il a le statut disponible, même si celui-ci ne se trouve pas dans sa caserne. Aussi, le véhicule peut être rerouté à tout moment de son trajet de retour d'intervention vers une nouvelle intervention. Il peut également être annulé, c'est-à-dire que son intervention n'est plus requise alors même qu'il se dirigeait vers le lieu d'intervention.



FIGURE 4.4 – Tournée fictive avec reroutage et arrêt logistique

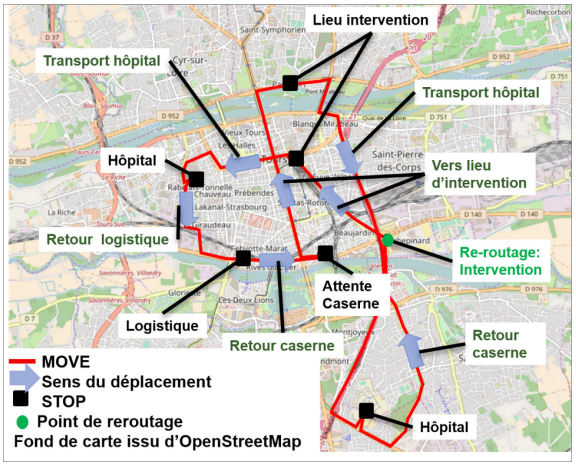


FIGURE 4.5 – Représentation cartographique d'une tournée fictive avec reroutage et arrêt logistique

La figure 4.4 présente le chaînage d'une tournée fictive plus complexe dont la visualisation cartographique est présentée dans la figure 4.5. Tout d'abord, le véhicule est appelé sur un lieu d'intervention. Une fois le patient pris en charge, celui-ci est ensuite transporté à l'hôpital. Puis, lors du chemin du retour, le véhicule est redirigé vers un autre lieu d'intervention. Le patient est pris en charge sur le lieu d'intervention puis amené à l'hôpital. Une fois le patient pris en charge par le service des urgences, le véhicule retourne à la caserne. Il effectue cependant un arrêt logistique avant de terminer sa tournée.

Nous avons présenté dans cette section des processus métier propres aux déplacements des ambulances des pompiers. De cette étude nous avons identifié des déplacements ayant des sémantiques particulières ainsi que des patrons de tournées des ambulances. Cela est important dans notre objectif de montée en sémantique sur les données de notre cas d'étude. En effet, une montée en sémantique est forcément dépendante du domaine d'étude, même si les processus sont génériques.

La section suivante présente le cas d'étude sur lequel nous avons travaillé dans cette thèse. De la présentation du contexte général, au système embarqué dans les ambulances jusqu'aux données finalement récoltées.

4.3 Récolte des données d'étude

Notre étude est menée sur des données issues d'ambulances connectées de pompiers. L'extraction de trajectoires est réalisée à partir d'un système d'acquisition de données issues de capteurs. Les capteurs peuvent renvoyer des données numériques ou des booléens. Chaque système est caractérisé par des méta-données portant sur la qualité. Elles dépendent des objectifs du système.

Deux méta-données sont particulièrement importantes :

— La fréquence de rafraîchissement des capteurs.

— La précision de la mesure renvoyée. Pour le GPS, une précision de 5 m, signifie que l'écart entre la position mesurée et la position exacte est, inférieur à 5m, pour 95% du temps et en zone dégagée.

4.3.1 Système d'acquisition

Dans cette thèse, nous avons utilisé le système de la société canadienne PRAN systems [1]. Ce système était celui qui a été installé sur 5 ambulances du S.D.I.S. 37 quelques mois avant le début de cette thèse. Cette solution a été retenue pour environ 100 000 ambulances à travers le monde.

1. `https://pransystems.com/fr/`

4.3.1.1 Présentation

PRAN est un système complet de visualisation temps réel et de contrôles électroniques et technologies de multiplexage CAN Bus pour véhicules d'urgence, transport adapté et véhicules spéciaux. Son objectif est donc de visualiser en temps réel sur un écran de contrôle la position d'une flotte de véhicules et de faire remonter des valeurs aberrantes issues d'un des nombreux capteurs. Il récupère les données issues du GPS et de deux bus de données CAN (Controller Area Network) :

— le bus générique aux véhicules de base,

— le bus spécifique aux véhicules d'urgence.

Les données des capteurs remontés sont détaillées dans la section suivante. Ces données sont stockées localement. Bien que, potentiellement, elles puissent être renvoyées, via un réseau de téléphonie mobile, cela n'a pas été le cas dans le cadre de nos travaux. Ce système est une boîte noire, il n'est pas possible de modifier les données renvoyées et ses modes de filtrage. Par exemple, afin d'économiser le volume de données à transmettre, le boîtier envoie uniquement les valeurs maximales sur les données de type numérique sur un pas de temps prédéfini. Cette politique est adaptée à des systèmes de contrôles et de visualisation temps réel, mais n'est pas la plus pertinente pour des systèmes d'analyse des activités. Nous avons donc constaté plusieurs facteurs qui dégradent la qualité des données à analyser.

4.3.1.2 Limites

Après analyse des premières données récoltées, nous listons ici les facteurs de dégradation principaux.

Les quatre premiers facteurs sont génériques au système GPS :

— Effet de pelote :

  La position à l'arrêt d'un véhicule n'est pas unique. Cette position recalculée par le système GPS varie constamment. L'amplitude de la variation est entre autres fonction de l'environnement (présence d'obstacles), du nombre de satellites visibles et des différents phénomènes physiques qui viennent perturber la propagation du signal.

  La figure 4.6 montre un exemple de pelote de positions d'un objet stationné en un point fixe à l'intérieur d'un bâtiment.

— Effet de rebond (*multipath*) :
  En milieu urbain avec des immeubles de plusieurs étages ou dans des rues étroites, le signal va rebondir sur ces obstacles et la précision de la position est dégradée. Un véhicule garé à droite de la chaussée peut être positionné à gauche par exemple. Cet effet est illustré dans la figure 4.7.

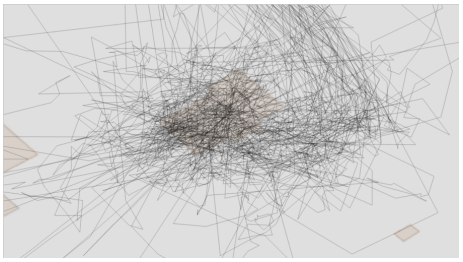


FIGURE 4.6 – Exemple de l'effet pelote résultant de données acquises par un capteur GPS à l'intérieur d'un bâtiment.

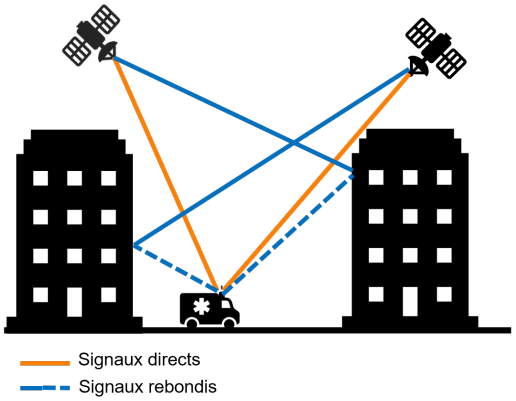


FIGURE 4.7 – Illustration de l'effet de rebond entre des bâtiments

— Perte de signal en indoor :
Quand les véhicules se trouvent sous une structure, le signal est très fortement perturbé, voire perdu. Dans ce cas, la position du véhicule est perdue (no data) ou avec une précision très dégradée qui rend la localisation impossible. Dans notre étude, ces cas sont fréquents : hangar de stationnement d'une caserne, sas couvert d'un hôpital, tunnel, etc.

— Pas de positions renvoyées après une mise en veille :
Après une mise en veille du système, l'acquisition des signaux GPS issue de plusieurs satellites et du calcul de la position ne sont pas immédiats. Le paramètre Time to first fix (TTFF) mesure cette durée.

De plus amples informations sur ces sources d'erreurs peuvent être obtenues par exemple dans [Karaim et al., 2018].

Les trois facteurs suivants sont spécifiques au système PRAN :

— Effet de crénelage :
Le système PRAN utilisant des filtres MAX pour l'ensemble des valeurs renvoyées dont la latitude et la longitude. Des effets de crénelage en résultent et sont permanents. Sur la figure 4.8 nous pouvons constater qu'ils sont particulièrement visibles sur l'axe Nord-Est Sud-Ouest. Cet effet n'est pas handicapant

pour des systèmes qui visualisent des points sur une carte à l'échelle d'une agglomération. Malheureusement, pour des systèmes d'analyse du comportement de véhicule, il rend très difficile les calculs de vitesse GPS, de map matching (système qui permet d'apparier des suites de position à des tronçons de route). Cet effet est aussi très néfaste pour les valeurs d'autres capteurs. Par exemple, il est impossible de calculer la différence entre une valeur minimale et une valeur maximale sur un intervalle de temps. Cette dernière opération est indispensable pour calculer la rugosité d'une route ou les secousses engendrées par un dos-d'âne à partir des valeurs des accéléromètres. Qui plus est, ces indicateurs sont fondamentaux pour le transport de patients très fragiles. La figure 4.8 illustre cet effet de crénelage sur les données récoltées.

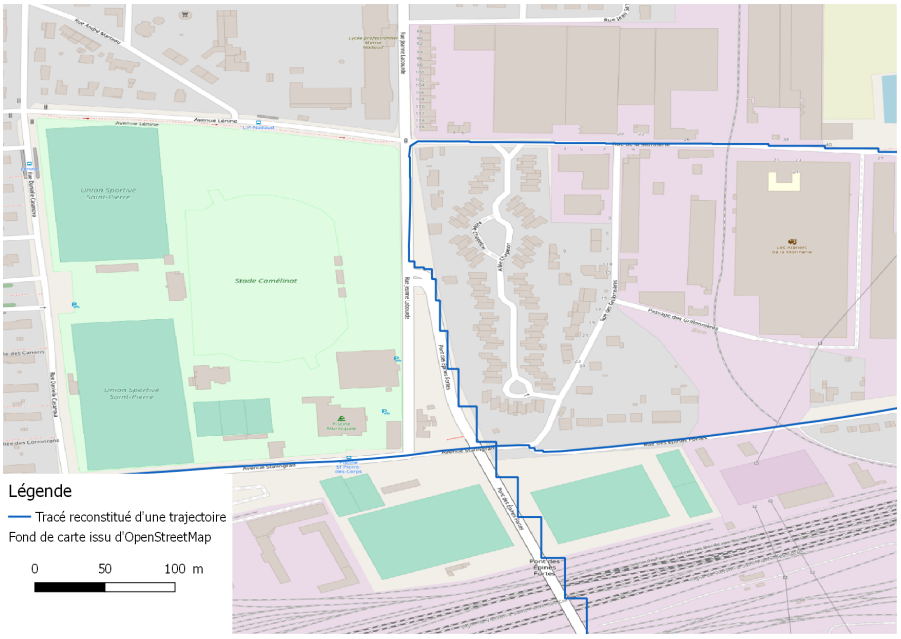


FIGURE 4.8 – Illustration de l'effet de crénelage sur les données émises par le système

— Effet de répétition :
Les valeurs numériques stockées ou renvoyées par le système PRAN ne sont pas toujours rafraîchies. Le système étant une boîte noire et la société n'ayant pas communiqué sur ce phénomène, il nous est impossible de connaître les critères de rafraîchissement. Par contre pour les valeurs booléennes (ouverture de portière, gyrophare allumé...), le changement de valeur est bien envoyé au bon moment (sous 1 seconde).

— Coupure de système veille :
Le système est programmé pour enregistrer les données récoltées chaque seconde. Cependant, l'enregistrement est interrompu lorsque le véhicule est arrêté (contact coupé) pendant plus de 30 minutes dans une caserne (la 1ère

seconde après la 30ème minute n'est pas stockée). Le système réenregistrera à nouveau uniquement si l'état du véhicule change (contact, mise en charge du véhicule). Les données récoltées sont donc considérées comme identiques tout le long de cette interruption.

Les modifications apportées sur le système qui équipe les ambulances sont propriétaires. Il est donc impossible de modifier le comportement du système tel que le mode qui interrompt l'enregistrement à partir de la 30ème minute.

Enfin, les données sont stockées directement dans les véhicules, et il faut, pour l'instant, accéder physiquement au véhicule pour récupérer les données à l'aide d'une clé USB avant qu'elles ne soient écrasées pour des raisons de l'espace mémoire disponible dans le système PRAN des véhicules.

4.3.2 Capteurs disponibles

Cette section présente la liste des capteurs disponibles sur le système propriétaire utilisé ainsi que leur type (binaire ou numérique). Les capteurs sont répartis dans trois catégories.

La première catégorie regroupe les capteurs dits de position (GPS, accéléromètres). Ces capteurs ont étés ajoutés dans le système embarqué, dans un objectif de suivi de flotte. Tous ces capteurs sont de types numériques. Les capteurs de positions sont au nombre de deux : le GPS et l'accéléromètre trois axes. Ils renvoient 5 valeurs numériques : deux pour le GPS et trois pour l'accélérateur 3 axes.

- GPS
  - Longitude exprimée dans le système de coordonnées WGS84
  - Latitude exprimée dans le système de coordonnées WGS84
- Accéléromètre [2] 3 axes
  - Accélération selon l'axe des x
  - Accélération selon l'axe des y
  - Accélération selon l'axe des z

La deuxième catégorie rassemble les capteurs issus du bus CAN générique du véhicule de base, *i.e.* communs aux voitures en général. Le bus CAN est un bus pour les réseaux de capteurs embarqués fortement déployé dans les voitures [iso, 2015, iso, 2016]. Il facilite l'intégration de multiples capteurs. Un état de l'art de son utilisation/adaptation par les chercheurs et constructeurs automobiles est présenté dans [Avatefipour and Malik, 2018]. La liste présentée ci-après n'est pas exhaustive, mais présente l'ensemble des capteurs dont nous pouvons récupérer les valeurs.

Capteurs gérés par le bus CAN communs aux voitures :

2. Mesurant les accélérations en g, avec un g = 9,806 m/s

- Numériques :
  - Vitesse : vitesse en km/h telle qu'affichée sur le compteur du véhicule
  - Voltage châssis : voltage de la batterie d'origine du véhicule, en volts
  - Vitesse de rotation : vitesse de rotation du moteur en tours par minute
  - Accélérateur : mesure de pression sur la pédale d'accélérateur (0 : pédale relâchée, 120 : pédale enfoncée)
  - Odomètre : compteur de kilomètres depuis l'installation du système embarqué
  - Consommation carburant : consommation instantanée du véhicule en litres par heure
- Binaires :
  - Clignotant droit : 1 si allumé, 0 sinon
  - Clignotant gauche : 1 si allumé, 0 sinon
  - Pédale de frein : 1 si la pédale est enfoncée, 0 si elle ne l'est pas
  - Veilleuses : 1 si allumées, 0 sinon
  - Feux de croisement : 1 si allumés, 0 sinon
  - Feux de route : 1 si allumés, 0 sinon
  - Marche arrière : 1 si la marche arrière est activée, 0 sinon
  - Témoin problème moteur : 1 si le témoin d'avarie moteur est allumé sur le tableau de bord, 0 s'il est éteint
  - Contact moteur : 1 si le véhicule à le contact (véhicule pas nécessairement démarré), 0 sinon
  - Porte conducteur : 1 si la portière est ouverte, 0 sinon
  - Porte passager : 1 si la portière est ouverte, 0 sinon
  - Porte arrière : 1 si la portière est ouverte, 0 sinon
  - Frein à main : 1 si le frein à main est tiré, 0 sinon
  - Ceinture de sécurité conducteur oubliée : 1 si le véhicule roule et que le conducteur n'a pas sa ceinture, 0 sinon
  - Airbag : 1 si le témoin d'avarie d'airbag est visible sur le tableau de bord, 0 sinon

La dernière catégorie regroupe les capteurs d'éléments spécifiques aux ambulances (sirène, gyrophares...) dont le recueil des données est possible via le bus CAN proposé par le système PRAN. Ce dernier stocke et transmet l'état des différentes équipements qui lui sont directement connectés. Ces capteurs sont de type binaire ou numérique.

Capteurs gérés par un bus CAN propre au système PRAN et spécifique aux ambulances :

— Binaires
  — Feux de pénétration (feux de calandre) : 1 si les feux sont allumés 0, sinon
  — Gyrophares : 1 si les gyrophares sont allumés 0, sinon
  — Sirène : 1 si la sirène est allumée 0, sinon
  — Chauffage cellule sanitaire : 1 si le chauffage est allumé 0, sinon
  — Mise en charge du véhicule : 1 si le véhicule est branché à une prise de courant 0, sinon
— Numériques
  — Voltage cellule sanitaire : Voltage de la batterie de la cellule sanitaire, en volts.

Ainsi, au final, même si le système embarqué PRAN ne fournit pas les informations de l'ensemble des capteurs présents sur le véhicule, il fournit des informations sur de nombreux capteurs. La section suivante présente l'acquisition des données servant de support aux différents expérimentations de ce mémoire.

4.3.3 Acquisition des données

Notre étude se base sur 18 jours de données issues de 2 ambulances connectées (Véhicule de Secours et d'Assistance aux Victimes - VSAV) de la caserne des Sapeurs-Pompiers de Tours-centre (37). Le fichier de log du premier véhicule comporte 1 300 643 lignes de données, tandis que le deuxième comporte 1 298 929 lignes. Au total, 2 599 572 lignes de données ont étés récoltés pour cette étude. La figure 4.9 représente l'étendue de ces données sur la ville de Tours. Nous avons récupéré les données dans des fichiers CSV (1 fichier par ambulance).

Un extrait du fichier de CSV extrait d'une ambulance est présenté dans la figure 4.10. Les champs présentés dans cette figure sont la date, l'heure, la longitude et la latitude (représentant la position GPS de l'ambulance) ainsi que le capteur du frein à main (PARK_BRAKE) où 1 signifie activé, 0 correspond à relâché.

Chaque fichier extrait d'une ambulance est inséré dans notre base de données. Dans cette thèse, PostGIS [3], extension de la base de données PostgreSQL [4], est utilisé comme moteur de base de données. Lors de l'insertion dans la base, les champs *date* et *time* sont convertis en une unique colonne de type *timestamp*. Enfin, l'identifiant du véhicule est ajouté pour permettre l'identification de chaque véhicule.

Cette section présenta la typologie de notre étude, le système utilisé ainsi que le volume de données récoltées. Comme nous l'avons vu, le système utilisé est un système propriétaire fermé, de plus, certains problèmes, inhérents ou non au système utilisé ont étés présentés.

3. https ://postgis.net/
4. https ://www.postgresql.org/

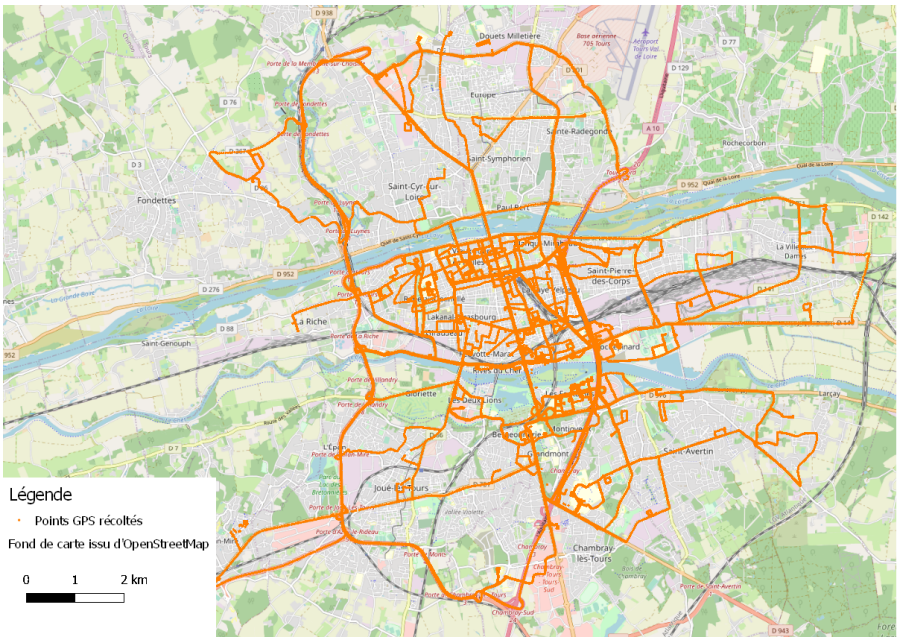


FIGURE 4.9 – Représentation géographiques des données collectées

	A	B	C	D	E
1	Date	Time	LONGITUDE	LATITUDE	PARK_BRAKE
2	30/10/2016	00:47:54	0.7000200	47.3746700	1
3	30/10/2016	00:47:55	0.7000200	47.3746700	1

FIGURE 4.10 – Extrait du format des données collectées

La prochaine section présente la méthode utilisée pour résoudre ses problèmes, et plus généralement la préparation des données avant la montée en sémantique faible.

4.4 Pré-traitement des données d'étude

Cette section présente les pré-traitements effectués sur les données collectées, ainsi que la méthode utilisée pour collecter les POI, deux étapes fondamentales pour pouvoir appliquer le processus de montée en sémantique. Le traitement de ces données a été réalisé en mode hors ligne : fichiers récupérés sur les ambulances à l'aide d'une clé USB, et POI extraits d'OpenStreetMap (dernier extrait en date du 5 mai 2020).

Comme décrit dans la section précédente, nos données souffrent de plusieurs problèmes, et il s'agit ici de présenter les approches utilisées pour les identifier et, selon les cas, les traiter.

Dans ce processus de pré-traitement, une première phase manuelle consiste à corriger les défauts constatés dans les estampilles temporelles. Ensuite, différents traitements seront appliqués pour détecter les différents positions géographiques présentant des problèmes liés aux systèmes d'acquisition (cf.4.3.1.2).

4.4.1 Correction des étiquettes temporelles

Lors des premières études sur les données, nous nous sommes aperçus d'un dysfonctionnement du système au niveau temporel. En effet, pendant certaines périodes, les données ont étés recueillies avec un mauvais *timestamp*. Bien que mineures, ces incohérences sur les étiquettes temporelles peuvent avoir un lourd impact sur les analyses et le processus de montée en sémantique.

	A	B	S	T	AE
1	Date	Time	LONGITUDE(°)	LATITUDE(°)	PARK_BRAKE
5	31/12/2016	16:46:58	0.7004936	47.3743008	1
6	31/12/2016	16:46:59	0.7004936	47.3743008	1
7	31/12/2016	16:47:00	0.7004936	47.3743008	1
8	01/01/2017	16:46:59	0.7004936	47.3743008	1
9	01/01/2017	16:47:00	0.7004936	47.3743008	1
10	01/01/2017	16:47:01	0.7004936	47.3743008	1
11	01/01/2017	16:47:02	0.7004936	47.3743008	1
33	01/01/2017	16:47:24	0.7004936	47.3743008	0
34	01/01/2017	16:47:25	0.7004936	47.3743041	0
35	01/01/2017	16:47:26	0.7004856	47.3743041	0
36	31/12/2016	16:47:27	0.7004856	47.3743277	0
37	31/12/2016	16:47:28	0.7004357	47.3743277	0

FIGURE 4.11 – Extrait d'un fichier issu d'une ambulance montrant un problème sur l'étiquetage temporel

En effet, dans un bloc cohérent de données collectées consécutivement, nous avons observé qu'une partie continue d'entre elles avaient des *timestamp* antérieurs de 24h aux données les encadrant dans le bloc (cf. figure 4.11). Nous ne connaissons pas la raison de cette défaillance sur les données. D'autant plus que nous trouvions par la suite des données ayant les mêmes estampilles temporelles incluses de manières cohérentes.

Nous avons corrigé ces estampilles temporelles pour nous assurer de la continuité des données acquises. Dans les données collectées, 3 périodes de défaillances ont été observées, pour un total d'éléments corrigés inférieur à 1500. Une fois ces erreurs corrigées, il est possible d'identifier les positions GPS problématiques.

4.4.2 Identification des positions GPS problématiques

Nous cherchons ici, à identifier les possibles valeurs GPS anormales fournies par le système PRAN selon les différents problèmes identifiés dans la section 4.3.1.2. L'objectif est de détecter les cas anormaux et non de supprimer (filtrer) les données ayant une anomalie spatiale, car si l'information GPS n'est pas forcément juste, les autres données captées (à minima la date, et ici, le frein à main) le sont vraisemblablement car la probabilité que tous les capteurs soient problématiques en même temps diminue avec le nombre de capteurs, d'autant plus que nous avons corrigé les problèmes de date. Dans la suite de ce document, ces données anormales sont appelées données aberrantes (ou outliers).

Nous avons identifié trois principaux types de cas pour lesquels les données géographiques pourraient être considérées comme aberrantes : les données nulles, les données caduques et les données GPS montrant des vitesses aberrantes.

Nous présenterons tout d'abord notre approche pour identifier ces principaux types de cas. Puis, dans un deuxième temps, nous présenterons les différents paramétrages utilisés ainsi que les résultats obtenus.

4.4.2.1 Détection des données GPS aberrantes

Nous introduisons, ici, les méthodes utilisées pour identifier les données aberrantes dans nos jeux de données. Les données ainsi identifiées sont étiquetées de manière spécifique en fonction de leur type : e.null, e.speed, e.lapsed.

Les données nulles Comme indiqué précédemment, lorsque le véhicule ne parvient pas à déterminer sa position, le système enregistre, par défaut, comme valeur GPS la position (0,0) qui ne correspondent nullement à notre aire géographique d'étude. Également, lorsque le véhicule est arrêté, contact coupé, pendant plus de 30 minutes, le système est désactivé. Lors de sa remise en fonction, le capteur GPS met alors un certain temps pour se synchroniser à nouveau (TTFF). Les données extraites comportent donc plusieurs lignes où les positions GPS ont pour valeur (0,0). Ces données ne doivent pas être prises en compte dans les différents calculs ou études effectuées sur les données. Elles doivent donc être marquées, afin d'être considérées différemment des autres données GPS récoltées.

Algorithm 1 Étiquetage des données GPS nulles

Require: e // un élément de trajectoire
  $e.null \leftarrow False$
  if $e.p.longitude = 0$ **and** $e.p.latitude = 0$ **then**
    $e.null \leftarrow True$
  end if

L'algorithme 1 présente la méthode utilisée pour définir une donnée nulle. Chaque élément de trajectoire où la longitude et la latitude sont nulles est considéré comme *null*.

Les données caduques Lors de l'étude des données acquises, nous nous sommes aperçus que de nombreuses données de position GPS étaient des répétitions exactes des données les précédant immédiatement. Au vu du nombre de chiffres après la virgule, ce phénomène est improbable. Nous supposons que le système ne transmet de nouvelles coordonnées que lorsqu'il estime que le changement est suffisamment significatif. Bien que ces données aient de fortes chances d'être proches de la réalité, il n'est pourtant pas pertinent de conserver les informations lorsque la durée de répétition est longue. En effet, cela pourrait aussi traduire dans ce cas un bogue du système PRAN. Ne pas considérer ces données comme aberrantes peut impacter fortement le calcul des médianes géométriques en donnant un poids beaucoup trop important aux coordonnées ainsi obtenues. Les données issues de ces longues répétitions sont considérées comme caduques. Toutes les données répétées ne sont pas caduques, toutes les données caduques sont de fait des données répétées. Il s'agit donc dans un premier temps d'identifier les données répétées.

Algorithm 2 Étiquetage des données GPS répétées

Require: e_i // un élément de trajectoire
Require: e_{i+1} // un élément de trajectoire suivant e_i
  if $e_i.p = e_{i+1}.p$ **then**
    $e_{i+1}.repeat \leftarrow True$
  end if

L'algorithme 2 présente la méthode utilisée pour définir les données répétées. Pour chaque paire d'éléments (e_i, e_{i+1}) d'éléments successifs de trajectoire, si $e_i.p = e_{i+1}.p$ alors e_{i+1} est considéré comme répété.

Comme vu précédemment, tous les éléments répétés ne seront pas considérés comme aberrant. Lors de la détection d'une sous-trajectoire constituée d'une seule position géographique, si sa longueur dépasse un seuil s_{lapsed} défini, alors tous les éléments de cette séquence situés après avoir atteint ce seuil sont considérés comme caduques.

L'algorithme 3 détaille le processus d'étiquetage des données répétées en données aberrantes.

Les vitesses aberrantes Dans le système propriétaire utilisé dans les ambulances, il n'est pas possible d'obtenir les indices de qualités usuels d'un système GNSS (Nombre de satellites connectés, VDOP, HDOP...). Cependant, afin de pouvoir déterminer d'avantages de positions aberrantes, il est possible de calculer la vitesse

Algorithm 3 Étiquetage des données répétées en données aberrantes

Require: K // une sous trajectoire constituée d'un élément répété $K = \langle e_1, ..., e_n \rangle$
Require: s_{lapsed} // seuil à partir duquel les répétitions deviennent anormales
  for $i = s_{lapsed}; i \leq n; i++$ **do**
    $e_i.lapsed \leftarrow True$
  end for

entre deux coordonnées GPS successives et la comparer aux limites atteignables par le véhicule. Si la vitesse calculée est excessive au regard des capacités du véhicule, cela traduit un défaut du système de positionnement. Les données GPS où la vitesse est excessive ne doivent donc pas être considérées.

Algorithm 4 Calcul de la vitesse GPS entre deux points

Require: e_i // un élément de trajectoire
Require: e_{i+1} // un élément de trajectoire suivant e_i
Require: $dist(e_1, e_2)$ // une fonction renvoyant la distance en mètres entre deux points e_1 et e_2
  $duree \leftarrow e_{i+1}.t - e_i.t$ // durée en secondes
  $distance \leftarrow dist(e_i, e_{i+1})$ // distance en mètres
  $speedGPS \leftarrow distance/duree * 3600$ // vitesse en km/h
  $e_{i+1}.speed \leftarrow speed$

L'algorithme 4 présente la méthode de calcul de la vitesse GPS entre deux points. Ainsi, si la vitesse calculée $e.speedGPS$ est supérieure à un seuil V_{max}^{normal}, alors la donnée est considérée comme aberrante.

Les méthodes présentées permettent l'identification de plusieurs types de données aberrantes. L'utilisation de seuils permet l'adaptation au contexte ainsi qu'aux objets mobiles étudiés. La section suivante présente les résultats obtenus dans le cadre de notre étude.

4.4.2.2 Résultats

Dans cette section, nous présentons les seuils que nous avons retenus pour chaque méthode, ainsi que les résultats obtenus sur les données de notre étude.

Données nulles Comme vu dans la section précédente, les données nulles sont automatiquement étiquetées en tant que données aberrantes. Dans cette étude, 5 719 éléments de trajectoires ont étés classés comme nuls, et donc, comme données aberrantes, ce qui représente 0.1% de l'ensemble des éléments de trajectoires.

Données répétées et données caduques Afin de déterminer le seuil à partir duquel, dans une sous trajectoire d'un élément répété, les données répétées sont considérées comme caduques, nous avons étudié ces cas de répétitions.

Nous avons tout d'abord extrait les suites d'éléments répétés, afin d'étudier leur durée. Au total, 66766 suites d'éléments répétés ont été extraites, qui représentent 2 382 572 éléments de trajectoire, c'est-à-dire 91.7% de l'ensemble des éléments de trajectoires.

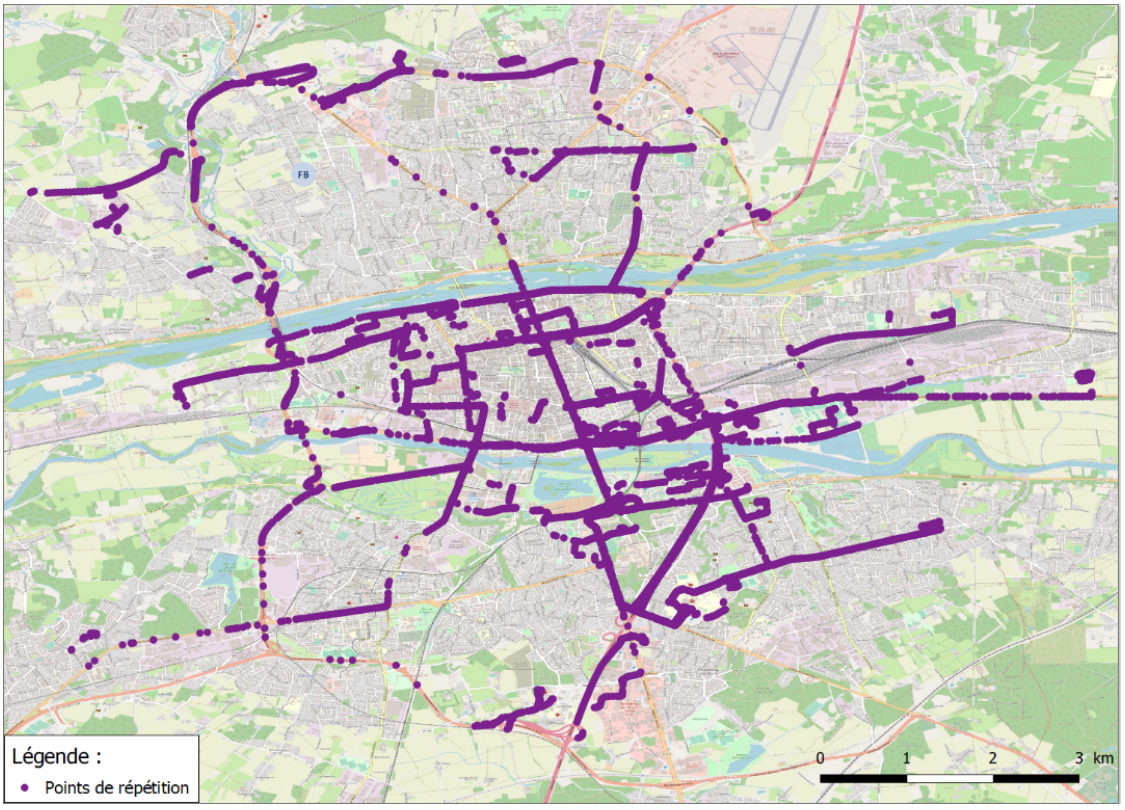


FIGURE 4.12 – Répartition géographique des données répétées pour v1

La répartition des positions géographiques de ces éléments répétés pour V1 est présentée dans la carte de la figure 4.12. Nous pouvons nous apercevoir que les répétitions sont réparties sur l'ensemble de la zone géographique étudiée.

Nous avons étudié, ensuite, la distribution des longueurs (en nombre d'éléments de trajectoire) des périodes de répétitions, présentée dans la figure 4.13. La plupart des périodes sont de longueurs inférieures à 10 éléments de trajectoires.

Plus particulièrement, comme le montre la figure 4.14, elles sont de longueur inférieure à 3, ce qui nous semble une période raisonnable avant de mettre en cause la valeur de la donnée fournie, c'est pourquoi nous avons choisi la valeur 3 comme seuil s_{lapsed} de la longueur des répétitions.

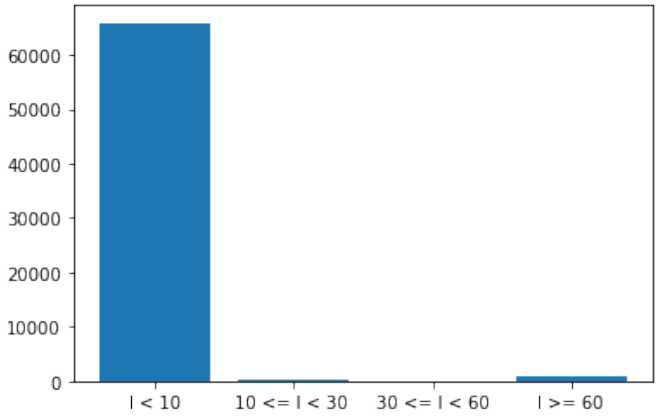


FIGURE 4.13 – Répartition des longueurs des périodes de répétitions

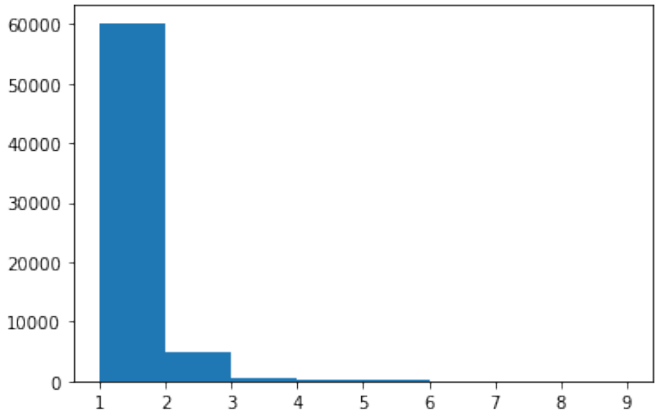


FIGURE 4.14 – Répartition des longueurs des périodes de répétitions inférieures à 10

La figure 4.15 présente la répartition géographique de ces données caduques pour le véhicule V1. Nous remarquons que la répartition des différents points de répétitions caduques recouvre moins le territoire.

Cas des vitesses aberrantes Par expérience, les ambulances n'atteignent que rarement une vitesse supérieure à 140km/h du fait de sa charge (environ 3,5t pour un véhicule équipé). Nous avons cependant ajouté une marge de 25% afin de ne pas étiqueter de vraies données en tant que données aberrantes : Un trajet comprenant une forte pente sur la route, par exemple, justifierait une vitesse dépassant les 140km/h.

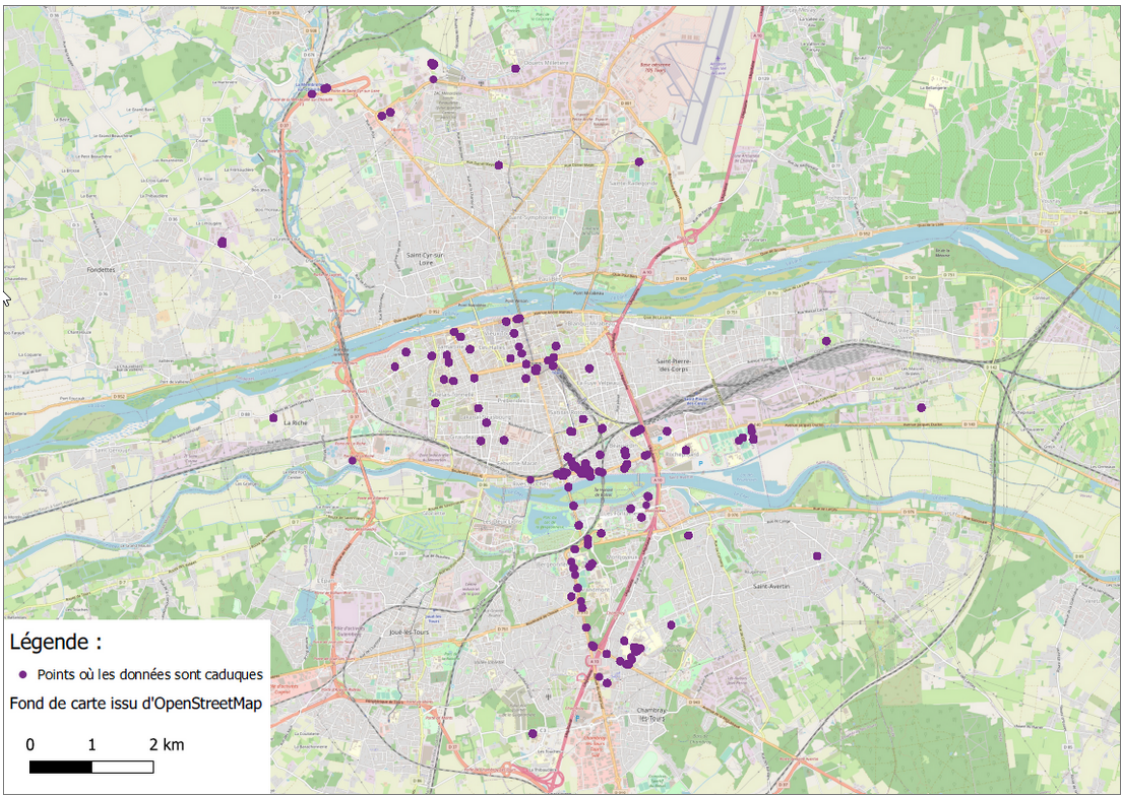


FIGURE 4.15 – Répartition géographique des données caduques pour V1

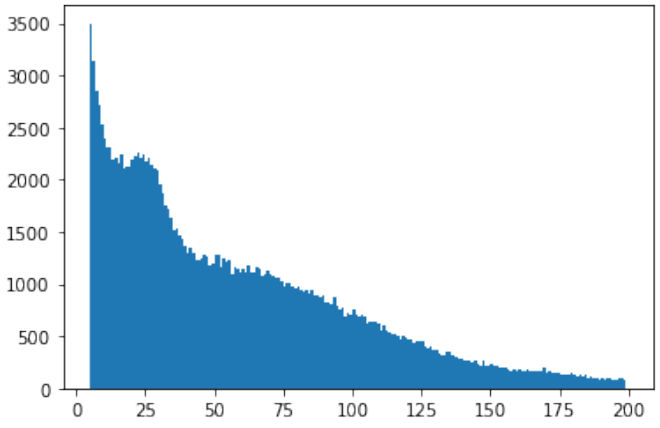


FIGURE 4.16 – Répartition des vitesses calculées inter-éléments de trajectoire supérieures à 5km/h

La vitesse normale maximale V_{max}^{normal} est donc définie à 175km/h. Tout élément pour lequel la vitesse calculée, vis-à-vis de l'élément qui le précède, est supérieure à ce seuil est donc considéré comme aberrant. Comme le montre la figure 4.16, le volume de points dépassant ce seuil est faible comparativement à l'ensemble.

Type d'aberration	Véhicule 1	Véhicule 2	Total
Valeur nulle	2 338	3 381	5 719
Vitesse aberrante	1 533	3 719	5 252
Valeur caduque	1 221 151	1 086 357	2 307 508
Total	1 225 022	1 093 457	2 318 479
Nombre total de données	1 300 643	1 298 929	2 599 572

TABLE 4.2 – Volume des éléments étiquetés comme aberrant selon leurs types et véhicule de provenance

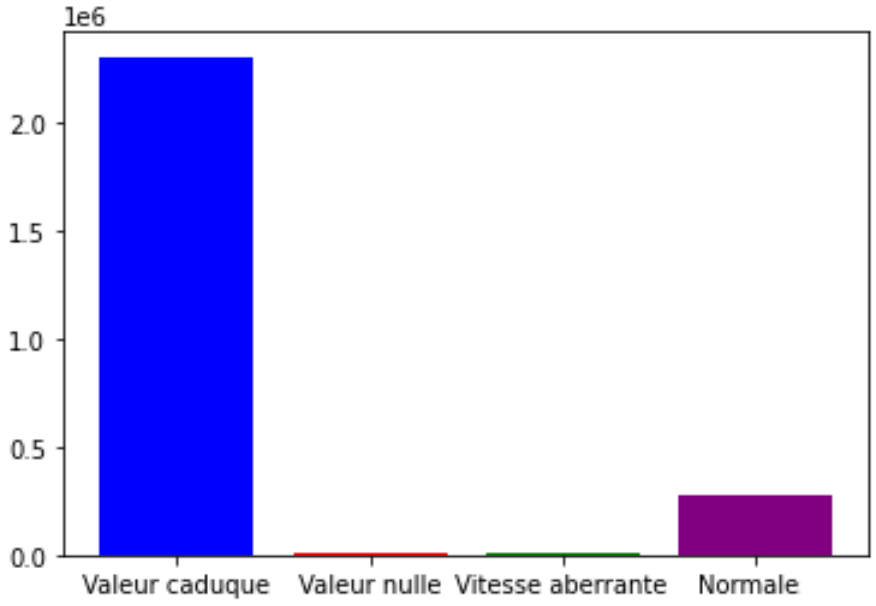


FIGURE 4.17 – Répartition des données aberrantes au regard des données normales pour les deux ambulances

Les résultats, présentés dans le tableau 4.2 et illustrés dans la figure 4.17, indiquent un très fort volume de données répétées. Cependant, ce résultat semble logique parce qu'un véhicule passe la majorité de son temps en stationnement et que le système réplique les données s'il considère que les variations sont faibles.

Conformément aux indications données précédemment, nous n'avons pas étiqueté toutes les répétitions comme aberrantes. Ainsi, parmi nos données normales nous avons toujours des données répétées, comme le montrent la figure 4.18 et le tableau 4.3. Ainsi, dans les données géographiques considérées, nous avons bien une majorité de données qui ne sont pas des répétitions. Désormais, dans la suite de ce manuscrit, les données répétées, mais normales, sont considérées comme répétées, tandis que les données dont la répétition est aberrante sont considérées comme caduques.

Données normales	Véhicule 1	Véhicule 2	Total
Répétées	20 343	54 721	75 064
Non répétées	55 278	150 751	206 029
Total	75 621	205 472	281 093

TABLE 4.3 – Volume des éléments normaux selon leur possible répétition

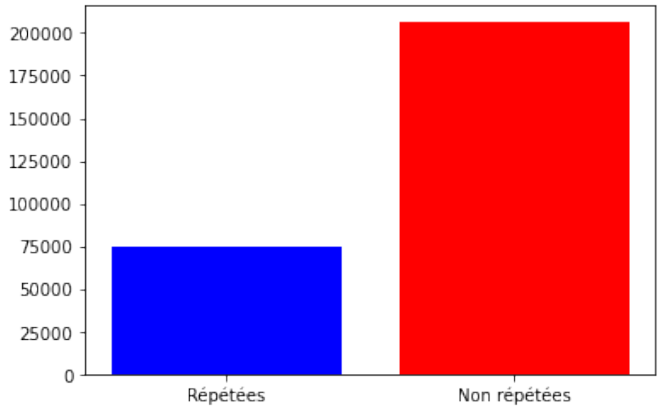


FIGURE 4.18 – Répartition des données normales au regard de leurs possibles répétitions pour les deux ambulances

Une fois ces pré-traitements effectués, on dispose maintenant d'une trajectoire brute étendue (cf. 3.2.2) par véhicule.

Cependant, dans nos processus de montée en sémantique et conformément à la littérature (cf. 2.3.2) et au chapitre précédent, il est nécessaire, lors de la montée en sémantique faible, de pouvoir associer aux STOP que nous obtiendrons des lieux d'intérêt ayant un sens vis-à-vis des processus métier présentés dans la section 4.2.

4.4.3 Extraction des lieux d'intérêt

Dans notre volonté de monter en sémantique faible puis forte, les POI sont des éléments importants à associer aux STOP qui seront identifiés.

Les POI, dans un contexte de montée en sémantique, ont pour objectif d'aider à qualifier l'activité de l'objet mobile étudié. C'est pourquoi les catégories de POI retenues doivent être en relation avec le métier des ambulances. À ce titre, trois catégories principales ont été retenues.

Les casernes de pompiers, et **les hôpitaux** sont deux catégories directement liées au métier des ambulances. La troisième catégorie concerne les lieux liés à la **logistique**, en effet, après de premières études sur les données, nous avons détecté

certains STOP à des endroits récurrents. Après une analyse de ceux-ci, il s'est avéré qu'ils étaient de plusieurs types, mais toujours en rapport avec l'activité du véhicule (entretiens de la mécanique du véhicule, plein de carburant,etc.).

Nous avons extrait des POI issus d'OpenStreetMap [5]. Les données d'OpenStreetMap ont l'avantage d'être ouvertes, facilement accessibles et libres d'utilisation. De plus, cette base de données est alimentée par différentes sources aussi bien officielles (Cadastre de la Direction Générale des Impôts..) que par des citoyens à l'aide d'un processus collaboratif étudié dans [Truong et al., 2019]. Différentes études scientifiques s'intéressent à ces données et à leur qualité, à l'instar de [Girres and Touya, 2010] ou [Raifer et al., 2020], montrant que c'est une source de données transparentes plutôt bien mise à jour et assez complète pour la France, notamment pour les Hôpitaux. Néanmoins des progrès sont souhaitables pour rejoindre le niveau de l'Allemagne par exemple.

Afin d'extraire ces POI, nous avons utilisé un outil d'exploration de données nommé *Overpass turbo* [6]. Cet outil permet la recherche et l'extraction de données issues d'OpenStreetMap à l'aide de requêtes. Les requêtes sont du type $cle = valeur$ et portent sur une emprise spatiale qui pour nous a été centrée sur la ville de Tours. Nous avons extrait trois types de POI : les casernes de pompiers, les hôpitaux ainsi que les lieux considérés comme logistique (station-service, garage du SDIS37). Les objets extraits peuvent être de type polygone, ou simplement des points.

Pour ces requêtes, la clé utilisée est la clé *amenity* qui pourrait se traduire en français par "infrastructure". La requête $amenity = fire_station$ est utilisée pour extraire les casernes de pompiers. Sur la ville de Tours et ses environs, 16 casernes ont étés extraites.

De la même manière, la requête $amenity = hospital$ a été utilisée pour extraire les hôpitaux et a renvoyé 13 éléments.

Pour extraire les POI de type logistique, nous avons tout d'abord extrait les stations-services, via la requête $amenity = fuel$, qui a retourné 63 éléments. Puis, nous avons ajouté manuellement une entité correspondant au garage du SDIS 37, où les véhicules sont emmenés pour les différents entretiens, car ce garage n'était pas directement extractible via une requête.

La figure 4.19 présente une carte positionnant, à l'aide de points du fait de l'échelle, les POI extraits à l'aide de l'outil. Dans cette figure, nous pouvons voir que les hôpitaux, se trouvent plutôt dans la périphérie de la ville, facilement accessibles par de grands axes. Les casernes de pompiers sont nombreuses et couvrent relativement bien le territoire. Les véhicules étudiés dans cette thèse sont issus de la caserne des pompiers de Tours centre, située au centre de la carte.

5. `https://www.openstreetmap.fr/`
6. Outil disponible depuis `https://overpass-turbo.eu/`

Pour les POI de type logistique, ils sont uniformément répartis sur la carte. Cette disposition semble cohérente car les stations-services sont relativement nombreuses dans les grandes villes. Cependant, nous pouvons noter que le POI logistique situé à l'ouest est le POI que nous avons ajouté manuellement, i.e. le garage du SDIS 37 à Fondettes. De plus, le site logistique à l'est de la caserne de Tours centre est la station-service où généralement les véhicules sont amenés à faire le plein de carburant. En effet, dans un schéma classique, les véhicules n'effectuent pas de détour pour faire le plein de carburant, et se retrouvent donc de facto à toujours aller dans la même station-service.

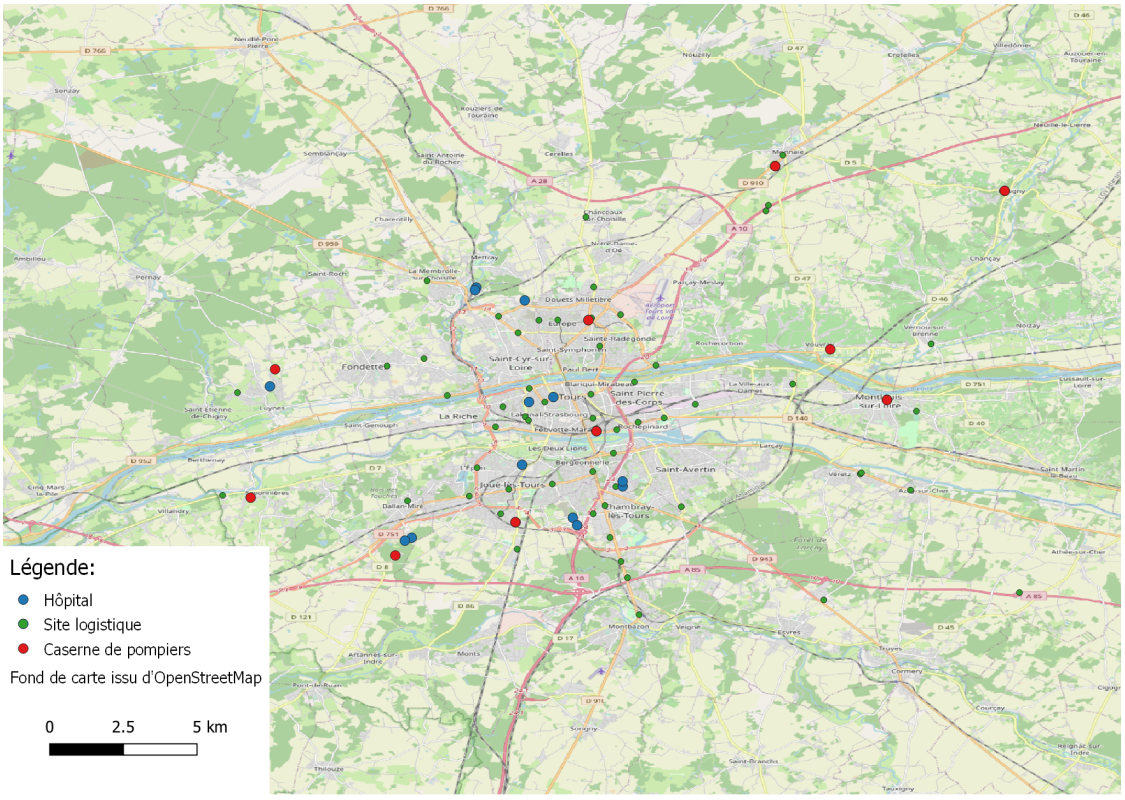


FIGURE 4.19 – Positions des POI issus de l'extraction à partir d'OpenStreetMap

Au total, pour notre zone d'étude, nous obtenons une liste de 92 POI répartis selon les trois catégories décrites préalablement.

4.5 Conclusion

Ce chapitre nous a permis de présenter les configurations types des déplacements métiers des ambulances ainsi que les processus associés. L'objectif de cette thèse porte sur l'étude et l'analyse des séquences d'activités effectuées par les ambulances.

Afin de pouvoir définir ces processus, un cas d'étude basé sur la récolte de 15 jours de données issues de deux ambulances a été choisi. La pré-étude du système d'acquisition et des données acquises a fait apparaître la possibilité d'utiliser de nombreux capteurs, mais aussi des problèmes importants de qualité de données. Aussi, et conformément à la méthodologie présentée dans le chapitre précédent, une phase de pré-traitement des données, exposée dans la section précédente, a été nécessaire.

Cette étape de pré-traitement nous a permis principalement, d'une part, d'étiqueter les données ayant des positions géographiques aberrantes, et, d'autre part, d'extraire d'OSM un certain nombre de lieux d'intérêt métier. Ces deux informations sont capitales pour la mise en œuvre des processus de montée en sémantique proposés dans les deux chapitres suivants.

Chapitre 5

Extraction des trajectoires sémantiques fortes à l'aide de données GPS et frein à main

5.1 Introduction

Dans l'objectif de chercher à identifier et à quantifier les motifs de tournées typiques dans les déplacements d'une flotte d'ambulances connectées, nous avons besoin d'analyser leurs trajectoires afin d'identifier ces tournées.

Pour répondre à ce besoin, nous devons dans un premier temps monter en sémantique forte les trajectoires récoltées pour cette flotte d'ambulances. Une fois monté en sémantique forte, nous analyserons les motifs présents dans les suites d'activités que ces trajectoires comportent.

Comme indiqué dans le chapitre 3, afin de monter en sémantique forte, une première étape de montée en sémantique faible est nécessaire. Afin de permettre cette étape, nous proposons dans ce chapitre de nous focaliser sur les données de trajectoire issues des données GPS et frein à main. Cela s'inscrit dans l'hypothèse au centre de cette thèse "les capteurs embarqués [ici le frein à main] peuvent faciliter et améliorer la montée en sémantique".

Les données issues du récepteur GPS sont en effet un élément clé afin de monter en sémantique faible, mais la segmentation uniquement basée sur celles-ci est assez coûteuse et son intérêt peut être fortement limité par la qualité des données. Or, dans ce mémoire, les données récoltées souffrent de nombreux problèmes, comme indiqué dans le chapitre 4. Aussi, afin de faciliter la montée en sémantique faible, les données du capteur du frein à main ont une signification forte. En effet, en utilisation normale, lorsque le frein à main est tiré, le véhicule est à l'arrêt. Donc, si

le véhicule fait un arrêt significatif, alors le frein à main est tiré. Ainsi, l'utilisation des données du capteur sur le frein à main doit nous permettre d'identifier les STOP et les MOVE.

Cependant, cette association de capteur ne permet pas la sous-segmentation des STOP pour différencier par exemple des phases de test et de logistique, d'une phase d'attente, lors d'un STOP à la caserne. Il en va de même pour les MOVE, où nous ne pourrons pas identifier deux sous MOVE consécutifs pouvant résulter d'un reroutage.

Nous présentons dans ce chapitre les choix méthodologiques pour la montée en sémantique forte, en partant de ceux faits dans la constitution des trajectoires brutes étendues et en passant par la montée en sémantique faible. Les choix pratiques ainsi que les résultats obtenus pour notre cas d'étude (cf. 4) sont présentés pour chaque étape. Nous analysons aussi les motifs de tournées obtenus et discutons de la qualité de notre démarche.

5.2 Construction des trajectoires brutes étendues

5.2.1 Choix des capteurs

Nous présentons dans cette section, les capteurs choisis pour construire nos trajectoires brutes étendues.

Capteur GPS Notre étude portant sur l'étude de trajectoires ainsi que leur montée en sémantique, il est donc normal d'utiliser le capteur GPS. Celui-ci permet de localiser spatialement les sous-trajectoires (STOP, MOVE) et de les lier, par exemple, à des POI. Par contre, les données GPS ne seront pas utilisées pour segmenter les trajectoires. Comme indiqué dans la section 4.3, nous ne disposons pas des méta-données de qualité liées au capteur GPS (Vdop,Hdop, nombre de satellites ...), ces données ne sont donc pas assez fiables pour détecter les mouvements.

État du frein à main Dans un véhicule, le frein à main permet de maintenir un frein sur les roues arrières du véhicule lorsque, par exemple, celui-ci est en stationnement. Sur tous les véhicules, un voyant rouge apparaît sur le tableau de bord lorsque le frein à main est tiré. Dans les véhicules de notre étude, une alarme sonore retentit lorsque le frein à main est tiré et que le véhicule roule. C'est pourquoi le capteur $freinmain$ semble être un capteur pertinent pour la segmentation des trajectoires et la détection des STOP. C'est un capteur qui renvoie une valeur booléenne. Cette dernière vaut 1 lorsque le frein à main est tiré, 0 sinon.

5.2.2 Trajectoires brutes étendues

À l'aide des capteurs choisis, conformément à la définition 2 présentée dans la section 3.2.2, nos trajectoires brutes étendues sont des séquences d'éléments $e_i^+ \in T^+$ ayant une position GPS, une étiquette temporelle, ainsi qu'un ensemble de capteurs $D = (B, C)$.

Comme le capteur $freinMain$ a été retenu comme unique capteur complémentaire au GPS dans ce chapitre, un élément de trajectoire e_i^+ est un objet Position défini comme :

$$e_i^+ . \begin{cases} t & \rightarrow \text{ timestamp} \\ p & \rightarrow \text{ localisation à l'instant } t \\ freinMain & \rightarrow \text{ état du frein à main à l'instant } t \end{cases} \tag{5.1}$$

La section suivante présente de premières analyses sur les éléments de trajectoires brutes étendues acquis.

5.2.3 Analyse des éléments de trajectoires acquis

La répartition des valeurs aberrantes dans nos trajectoires étant indiquée dans la section 4.4.2.2, nous mettons ici le focus sur les données issues du capteur du frein à main.

La répartition des informations sur le frein à main est la suivante. Pour le premier véhicule, 1 176 422 éléments de trajectoires ont pour valeur $freinMain == 1$ et 124 221 où $freinMain == 0$. Pour le second, 993 650 éléments ont pour valeur $freinMain == 1$ et 305 279 où $freinMain == 0$. Au total, sur les 2 599 572 éléments de trajectoires étendues récoltés, 83% des éléments ont le frein à main tiré ($freinMain == 1$) et 17% où il ne l'est pas. Ces résultats sont parfaitement logiques car les ambulances de pompiers sont le plus souvent stationnées (et donc logiquement le frein à main tiré).

La section suivante présente la montée en sémantique faible faite à partir des trajectoires brutes étendues obtenues.

5.3 Montée en sémantique faible des trajectoires brutes étendues

Cette section présente l'étape de montée en sémantique faible.

À ce stade, nous disposons de trajectoires brutes étendues (une liste de tuples position, capteurs). Conformément au processus de montée en sémantique faible présenté dans le chapitre 3, nous devons commencer par segmenter nos trajectoires brutes étendues. L'information sur le frein à main est considérée pour extraire les

sous-trajectoires. Puis, à l'aide d'une fonction *arret*, nous qualifierons les différentes sous-trajectoires en STOP ou MOVE. Enfin, la dernière étape est l'assignation des STOP avec des POI. Nous présenterons, pour chaque étape, la méthode utilisée puis l'expérimentation sur les données collectées.

5.3.1 Segmentation en sous-trajectoires

L'étape suivante est le choix de la fonction de segmentation. Dans la littérature, plusieurs problèmes, dont certains présentés dans la suite, rendent la détection de STOP (lieu et durée) complexe à l'aide de données GPS classiques [Cain III et al., 2005, Duran and Earleywine, 2013, Knight and Wang, 2009, Ivanović et al., 2016, Ordonez et al., 2011].

5.3.1.1 Problèmes rencontrés

Tout d'abord l'imprécision des capteurs GPS doit être prise en compte. Le signal est très perturbé aux abords de bâtiments ou dans des canyons urbains. Ces perturbations entraînent l'apparition de données aberrantes (rebonds, pelotes), voire d'une perte de signal totale si l'objet mobile se trouve à l'intérieur d'un bâtiment. L'analyse lorsque le véhicule stationne en garage fermé ou traverse des tunnels est donc problématique (cf. 4.3).

Qui plus est, afin d'extraire des informations propres aux véhicules et à leurs activités, il faut distinguer les micro-arrêts des arrêts. Pour cela, les contraintes usuelles de circulation (feux rouges, passages piétons) doivent être prises en compte.

5.3.1.2 Méthode de détection basée sur le frein à main

Le capteur de frein à main offre une nouvelle opportunité pour définir les STOP. En effet, pour déterminer l'arrêt d'un véhicule, le capteur de frein à main est pertinent. Une fonction simple est donc définie à l'aide de ce capteur. Le moment où le frein à main du véhicule est serré ou desserré permet de définir une segmentation de trajectoire plus fiable qu'avec un capteur GPS.

Finalement, si un changement d'état du capteur de frein à main est détecté entre T_i et T_{i+1}, alors la trajectoire est segmentée. T_i est inclus dans la première sous-trajectoire et forme l'élément de fin de cette sous-trajectoire. T_{i+1} est l'élément de début de la deuxième sous-trajectoire. L'un des risques de cette approche est qu'un véhicule à l'arrêt mais avec le frein à main relâché est considéré en mouvement.

L'algorithme 5 présente le processus de segmentation des trajectoires en segments (sous-trajectoires) avec l'utilisation du frein à main. Considérons une première sous-trajectoire vide qui sera le segment courant. Chaque élément de la trajectoire à segmenter est ajouté au segment courant. Si on détecte un changement d'état du frein

à main entre l'élément en cours de traitement et le précédent, alors nous stockons la sous-trajectoire courante et créons un nouveau segment vide dans lequel est inséré l'élément en cours. La dernière sous-trajectoire courante obtenue est à son tour stockée. Ce processus est réalisable en ligne, mais, dans le cadre de notre travail, est effectué hors ligne.

Algorithm 5 Segmentation d'une trajectoire brute étendue

Require: T trajectoire $T = \langle e_1^+, ..., e_n^+ \rangle$
  $i \leftarrow 1$ //iterateur
  $L \leftarrow []$ //liste des sous-trajectoires à retourner
  $st \leftarrow []$ //sous-trajectoire en cours
  while $i < n$) **do**
    st.Push(e_i^+)
    if ($e_i^+.freinMain \neq e_{i+1}^+.freinMain$) **then**
      L.Push(st)
      $st \leftarrow []$
    end if
    $i \leftarrow i + 1$
  end while
  st.Push(e_n^+)
  L.Push(st)
  return L

Comme ce choix diffère de l'état de l'art, nous le discutons ci-après.

5.3.1.3 Discussion par rapport à l'état de l'art

Premièrement, il est à noter que la faible complexité en $\mathcal{O}(n)$ de notre segmentation des trajectoires par détection de l'usage du frein à main, est un vrai avantage pour le traitement de gros volumes de données en comparaison avec les complexités maximale et moyenne des approches de la littérature telles que de DBSCAN [Ester et al., 1996] – et de ses variantes ST-DBSCAN [Birant and Kut, 2007], T-DBSCAN [Chen et al., 2014] ou encore celle proposée par Luo et al. [Luo et al., 2017] – qui sont respectivement en ($\mathcal{O}(n^2)$) et $\mathcal{O}(n * \log n)$.

Ainsi, avec plus de 1,3 millions de lignes pour chacun des fichiers, nous n'avons pas pu obtenir de résultats pour la méthode T-DBSCAN [Chen et al., 2014] optimisée avec les critères de [Luo et al., 2017] en 24h de calcul en dédiant un cœur de calcul de type core i7 7600 couplé à 16 GB de ram en python. Le coût de chargement mémoire sur R étant trop lourd, pour notre configuration, nous n'avons pas testé avec le module usuellement utilisé dans la segmentation de trajectoire.

Ce faible coût combinatoire est une motivation importante du choix fait, dans ce chapitre, dans l'objectif de tendre vers le traitement à la volée des informations collectées des ambulances. Cela amène à un gain en capacité de calcul pour l'exploitation sur de flottes plus importantes. D'autant plus que les possibles bruits et répétitions dans nos données impactent les choix des définitions de seuil, qui deviennent alors assez complexes.

Par ailleurs, pour répondre aux problèmes introduits précédemment (cf. 5.3.1.1), [Etienne et al., 2012] propose de définir des zones géométriques d'entrée et de sortie pour déterminer le passage d'un objet mobile et ainsi le début ou la fin d'un arrêt. Cependant, cette méthode implique de devoir connaître à l'avance les frontières géographiques d'entrée et de sortie pour chaque zone.

Cette méthode à été appliquée sur les données récoltées. Cette technique fonctionne lorsqu'on a un effet de "porte" d'entrée/sortie à l'instar du portail de la caserne. Des zones d'entrée et sorties ont été définies manuellement au niveau de la caserne des pompiers, afin de détecter les départs et arrivées des véhicules. Ces polygones ont été définis de manière manuelle et couvrent uniquement la zone d'entrée/sortie des départs en intervention des ambulances.

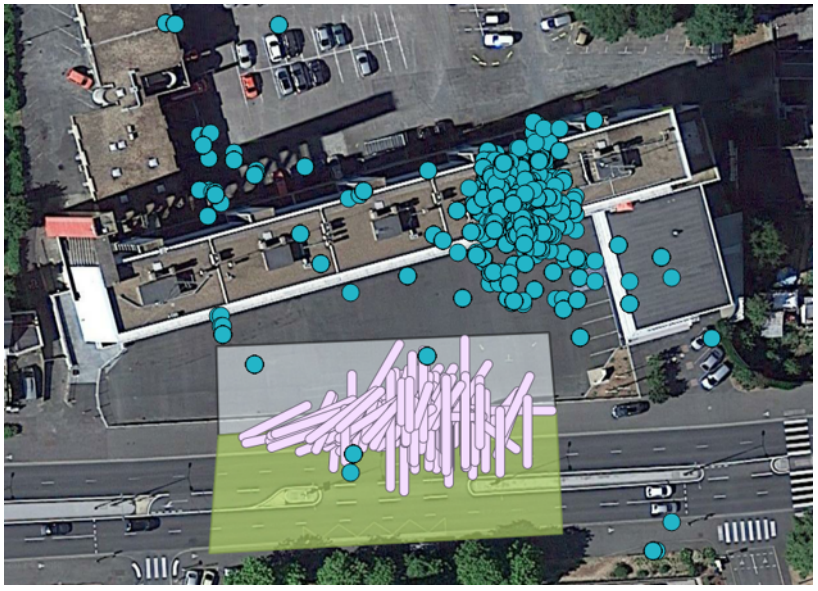


FIGURE 5.1 – Segments détectés comme entrées / sorties de la caserne des pompiers

La figure 5.1 montre les résultats obtenus avec cette méthode ainsi que la manière dont elle est appliquée, la porte de la caserne. Un véhicule présent dans le polygone du dessus à T_i et dans le polygone du dessous à T_{i+1} est considéré comme sortant de la caserne (donc à l'arrêt avant cet instant). À l'inverse, un véhicule présent dans le polygone du dessous à T_i et dans le polygone du dessous à T_{i+1} est considéré comme entrant dans la caserne (et donc à l'arrêt à partir de T_{i+2}). Ces mouvements d'entrée-sortie sont représentés par des segments roses dans la figure.

Cependant, même lors de la sortie d'un bâtiment, la qualité des données est liée au temps dont le capteur GPS a besoin pour se synchroniser avec les satellites (TTF - Time To Fix). Qui plus est, les effets de pelote et les pertes de signal entraînent des sorties virtuelles intempestives de l'objet mobile. Cette méthode efficace en zone ouverte est difficilement transposable en zone urbaine. La figure 5.2 montre l'ensemble des données collectées (traces vertes) dans la zone, où l'on peut voir que certaines trajectoires ne passent pas par les zones définies précédemment, à cause des contraintes énoncées.

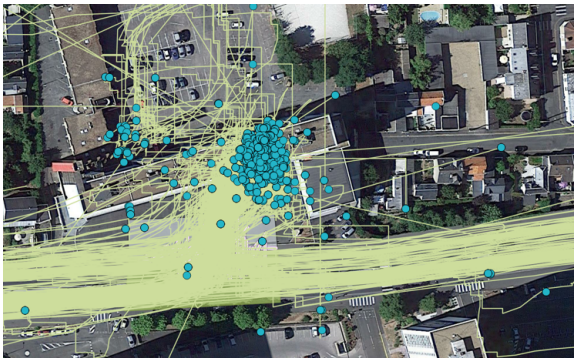


FIGURE 5.2 – Représentation de l'ensemble des données collectées dans la zone

Enfin, cette méthode est applicable pour la caserne, les hôpitaux et certains services à proximité (ex. station-service) mais nous ne pouvons pas l'appliquer à d'autres sites, notamment les lieux d'intervention.

Ainsi, ces limites nous confortent dans le choix fait de notre méthode de segmentation.

5.3.1.4 Analyse des sous-trajectoires identifiées

Comme vu dans la section précédente, la première extraction des sous-trajectoires avec le frein à main permet d'identifier des périodes d'arrêts potentiels dans les trajectoires. La figure 5.3 représente la répartition géographique des sous trajectoires de type arrêts potentiels (période où le frein à main est serré).

Au total, nous avons obtenu 1609 sous-trajectoires pour les deux véhicules : 489 pour le véhicule 1 et 1120 pour le véhicule 2. Nous obtenons plus précisément 805 potentiels arrêts (frein à main tiré), et 804 potentiels déplacements. Cela représente, pour le véhicule 1, 245 arrêts potentiels et 244 déplacements potentiels. Tandis que pour le second véhicule, 560 pour les deux types de sous-trajectoires.

Le tableau 5.1 présente la répartition de la durée de ces sous-trajectoires selon qu'ils soient des arrêts potentiels ou déplacements potentiels.

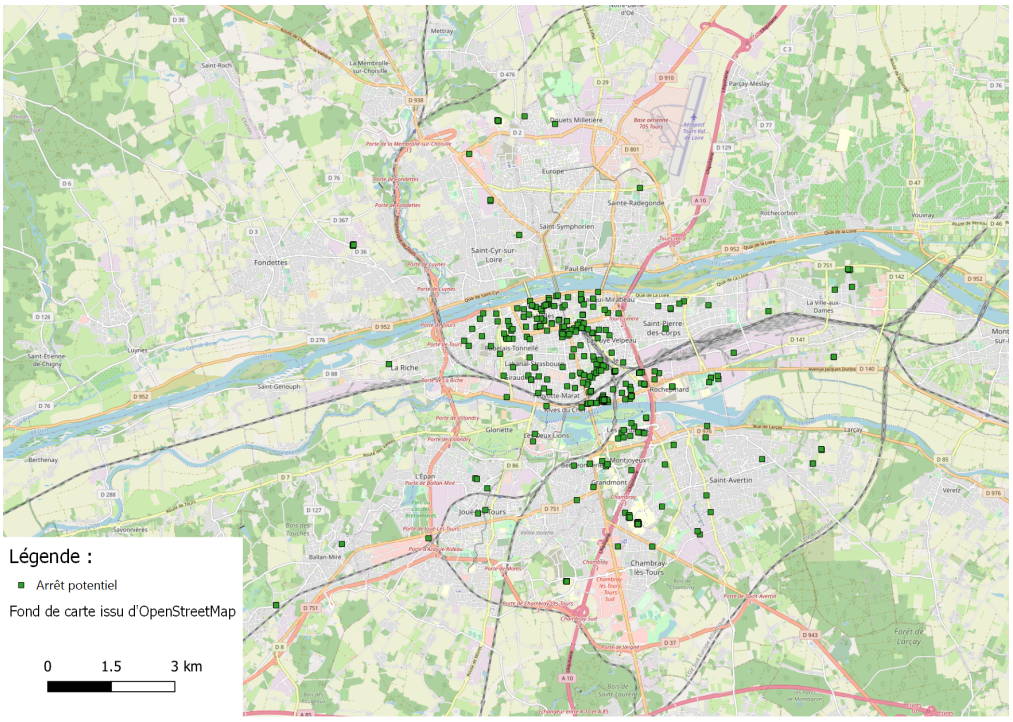


FIGURE 5.3 – Représentation des potentiels arrêts sur une carte OSM

Nous pouvons remarquer des durées maximales longues pour les arrêts et déplacements potentiels. Cependant, les durées médianes et moyennes étant relativement faibles, on peut en déduire que la majorité des sous-trajectoires extraites ont des durées inférieures à 14 minutes (840 secondes) et que de longues sous-trajectoires font augmenter significativement la durée moyenne.

Pour information, la répartition des données aberrantes par arrêts/déplacements potentiels est donnée dans le tableau 5.2. Cette première étude des données aberrantes montre quelques résultats intéressants. Tout d'abord, les données nulles semblent partagées entre les arrêts et les déplacements potentiels. Les vitesses aberrantes se trouvent en grande majorité lorsque le véhicule est en déplacement potentiel, alors

	Arrêts potentiels	Déplacements potentiels
Nombre	805	804
Durée minimum	1	1
Durée moyenne	00 :45 :04	00 :09 :31
Durée médiane	00 :14 :00	00 :05 :42
Durée maximum	28 :53 :24	07 :49 :18

TABLE 5.1 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires extraites

que les données caduques se trouvent pour plus de 93% dans des arrêts potentiels. Ces premiers résultats semblent cohérents. En effet, le système ayant tendance à répéter les informations lorsque celles-ci ne changent pas suffisamment. Il est donc normal que la plupart des données caduques se trouvent dans des arrêts potentiels.

Une fois les sous trajectoires extraites, la prochaine section décrit le processus d'extraction des STOP et des MOVE.

5.3.2 Extraction des STOP et des MOVE

Cette section présente l'extraction des STOP et des MOVE à partir des sous-trajectoires récoltées. Nous proposerons tout d'abord une méthode pour la fonction *arret*, en rapport avec les capteurs choisis dans ce chapitre. Cette fonction permettra de filtrer les micro-arrêts, puis nous présenterons également la méthode de filtrage des micro-déplacements. Nous détaillerons ensuite les choix fait concernant les seuils des méthodes précédemment citées. Puis finalement nous présenterons les résultats de cette extraction.

5.3.2.1 Méthode proposée pour la fonction arret

La fonction *arret* à pour objectif de déterminer si une sous trajectoire est de type STOP ou MOVE. Comme indiqué précédemment, la segmentation étant basée sur le changement d'état du capteur de frein à main, si son état est 1 (frein à main tiré), alors la sous trajectoire est de type STOP, sinon, de type MOVE.

Comme indiqué dans le chapitre précédent, la fonction $arret_{freinMain}(T_i^+)$ est une fonction booléenne, qui renvoie vrai si le frein à main est tiré durant une sous trajectoire T_i^+ et faux sinon, et qui est définie de la façon suivante :

Soit T_i^+ une sous trajectoire.

$$arret_{freinMain}(T_i^+) = true \iff \forall e_i^+ \in T_i^+, e_i^+.freinMain = true \tag{5.2}$$

Il est aussi nécessaire de filtrer les micros-arrêts. Pour cela, un seuil temporaire issu d'une concertation avec des experts métiers est défini. Ce seuil, noté D_{min}^{arret}, indique la durée minimale pour laquelle un arrêt est significatif en relation avec son activité. Ce seuil est dépendant du cadre applicatif : les valeurs pour des déplacements d'un taxi, d'un cycliste ou d'un animal ne seront potentiellement pas identiques.

Soit $T_i^+.D$ la durée de T_i^+ en secondes.

Alors, afin de prendre en considération l'ensemble des critères avec la durée minimale d'un arrêt, la fonction *arret* est définie comme :

$$arret(T_i^+) = arret_{freinMain}(T_i^+) \wedge (T_i^+.D > D_{min}^{arret}) \tag{5.3}$$

	Arrêts potentiels	Déplacements potentiels
Nombre total de données	2 170 072	429 500
Nombre de données nulles	2 932	2 787
Pourcentage de données nulles relatif au nombre total de données nulles	$\frac{2932}{2932+2787}$ =51%	$\frac{2787}{2932+2787}$ =49%
Pourcentage de données nulles relatif au nombre total de données par type de sous-trajectoires	$\frac{2932}{2170072}$ =0,1%	$\frac{2787}{429500}$ =0,6%
Nombre de données caduques	2 162 475	145 035
Pourcentage de données caduques relatif au nombre total de données caduques	93,7%	6,3%
Pourcentage de données caduques relatif au nombre total de données par type de sous-trajectoires	99,6%	33,8%
Nombre de données vitesse aberrante	87	5 165
Pourcentage de données vitesse aberrante relatif au nombre total de données vitesse aberrante	2%	98%
Pourcentage de données vitesse aberrante relatif au nombre total de données par type de sous-trajectoires	<0,1%	1,2%
Nombre de données aberrantes	2 165 494	152 987
Pourcentage de données aberrantes relatif au nombre total de données aberrantes	93%	7%
Pourcentage de données aberrantes relatif au nombre total de données par type de sous-trajectoires	99,8%	35,6%

TABLE 5.2 – Répartition des données aberrantes selon les types de sous-trajectoires

Finalement, les sous-trajectoires durant plus de D_{min}^{arret} secondes et où le frein à main est tiré sont considérés comme des STOP, sinon, elles sont concaténées aux sous-trajectoires précédentes et suivantes. La sous-trajectoire résultante de cette concaténation sera logiquement un MOVE (cf. équation 5.4).

$$!(arret(T_i^+)) \wedge arret_{freinMain}(T_i^+) \rightarrow concat(T_{i-1}^+, T_i^+, T_{i+1}^+) \tag{5.4}$$

Ainsi, les micro-arrêts comme les arrêts à un carrefour, feux rouges, démarrage en côte ne sont pas retenus. Ce temps est un paramètre à définir suivant l'activité du type de véhicules étudiés. Pour des véhicules ayant des activités courtes, il faudrait aussi prendre en compte le contexte (présence de feux de circulation, etc.). Par exemple, un véhicule de distribution du courrier réalise de nombreux arrêts relativement courts, l'utilisation exclusive de ce seuil serait donc insuffisante. Dans ce cas, des valeurs supplémentaires issues de capteurs devraient être aussi employées. Par exemple, le capteur d'ouverture du coffre où se trouve le courrier est pertinent pour détecter les STOP des véhicules de distribution. C'est dans cet objectif que le chapitre suivant étend cette démarche de segmentation de sous-trajectoires avec d'avantages de capteurs.

Par ailleurs, comme vue dans la section 3.3.2, la médiane géométrique est choisie afin de représenter l'ensemble des positions acquises durant une sous-trajectoire de type STOP. Cette médiane est en fait le point qui minimise la somme des distances à tous les autres points. Dans notre cas, pour chaque sous-trajectoire T_i^+ de type STOP, nous ne considérons, pour le calcul de la médiane géométrique que les éléments de trajectoire e_j^+ lui appartenant et qui ne sont pas des outliers spatiaux. Soit $T_i^+.medGeo$, la médiane géométrique des positions de T_i^+ dont la formule est donnée dans l'équation 5.5.

$$T_i^+.medGeo = \underset{(x,y)\in\mathcal{R}^2}{\arg\min} \left(\sum_{e_j^+ \in T_i^+ \wedge !outlier(e_j^+.p)} Dist(e_j^+.p, (x,y) \right) \tag{5.5}$$

$outlier(e_j^+.p)$ est une fonction booléenne qui indique si l'élément passé en paramètre est considéré comme aberrant (caduque, vitesse ou null) ou non.

Dans les cas classiques, c'est une distance euclidienne entre les points qui est utilisée. Cependant, l'utilisation d'une distance euclidienne avec des coordonnées GPS n'est pas correcte. En effet, la terre étant une sphère et non pas une surface plate, la distance entre deux points A et B n'est pas la distance minimale entre ces deux points, mais bien la distance de l'arc de cercle qui sépare ces deux points aussi appelée distance de Haversine.

Soit e_i^+ et e_j^+ deux éléments de trajectoire dont les coordonnées GPS de leur position sont exprimées en degré décimal selon WGS84. Pour tout élément de trajectoire e_i^+, soit ϕ_i la latitude de sa position $e_i^+.p$ exprimée en radian, et ψ_i la longitude de

sa position $e_i^+.p$ exprimée en radian. La distance entre les positions $e_i^+.p$ et $e_j^+.p$ est définie, à l'aide de la simplification de la distance de Haversine par la loi sphérique des cosinus, comme suit :

$$Dist(e_i^+.p, e_j^+.p) = \arccos(\sin(\phi_i) * \sin(\phi_j) + \cos(\phi_i) * \cos(\phi_j) * \cos(\psi_i - \psi_j)) * r_{Terre} \tag{5.6}$$

où r_{Terre} est le rayon conventionnel de la Terre valant 6371km [1].

Ainsi, pour avoir la distance entre deux points en mètres, on a :

$$r_{Terre} = 6371000 \tag{5.7}$$

Afin de déterminer la médiane géométrique, nous avons utilisé l'optimisation de l'algorithme présenté dans [Weiszfeld and Plastria, 2009] qui cherche un optimum local (et non l'optimum absolu). C'est un algorithme itératif, qui considère un seuil, et si par rapport à ce seuil le test de nouvelles valeurs n'est pas nécessaire, l'algorithme s'arrête. Afin de limiter le temps de recherche, l'algorithme peut prendre aussi en paramètre un nombre maximum d'itérations possibles ($nbIter_{Weiszfeld}$).

Enfin, à l'instar des micro-arrêts, nous souhaitons réduire les micro-déplacements, car ceux-ci peuvent être la conséquence de problèmes sur le capteur, de manipulation du véhicule sans volonté de mouvement, ou de tous petits mouvements n'influant pas l'affectation des STOP au POI. Aussi, pour différencier les micro-déplacements restreints des micro-déplacements à conserver, nous avons considéré deux seuils. Le premier, D_{min}^{dep} permet à l'instar de D_{min}^{arret} de ne considérer comme possible micro-déplacements à supprimer que les micro-déplacements ayant une durée inférieure à ce seuil. Le second, $dist_{min}^{dep}$ permet de dire que pour qu'un micro-déplacement, d'une durée inférieure à D_{min}^{dep}, soit conservé, il faut que la distance séparant les arrêts immédiatement précédant et suivant le micro-déplacement soit supérieure à $dist_{min}^{dep}$. Sinon, ils sont concaténés aux sous-trajectoires précédentes et suivantes. La sous-trajectoire résultante de cette concaténation sera logiquement un STOP. L'opération est présentée dans l'équation 5.8.

$$\begin{aligned} &!(arret_{freinMain}(T_i^+)) \wedge (T_i^+.D < D_{min}^{dep}) \\ \wedge &(dist(T_{i-1}^+.medGeom, T_{i+1}^+.medGeom) < dist_{min}^{dep}) \\ &\rightarrow concat(T_{i-1}^+, T_i^+, T_{i+1}^+) \end{aligned} \tag{5.8}$$

Lors de la concaténation, seule la médiane géométrique de la sous-trajectoire précédente ou suivante durant le plus longtemps est conservée. En effet, ayant fusionné les micro-déplacements entre deux STOP, les points constituants le STOP nouvellement fusionné proviennent donc à la fois d'un déplacement, et de deux arrêts. Il nous a paru pertinent d'une part de ne pas prendre en considération les données

1. `https://solarsystem.nasa.gov/planets/earth/by-the-numbers/`

provenant de ce micro-déplacement, et d'autre part, de ne conserver que la médiane géométrique de l'arrêt ayant duré le plus longtemps. En effet, l'arrêt ayant duré le plus longtemps semble plus à même de représenter le STOP final après la fusion.

Dans cette section nous avons défini les différentes méthodes pour extraire les STOP et les MOVE. La section suivante présente et argumente les choix que nous avons faits concernant ces différents seuils pour notre cas d'étude.

5.3.2.2 Choix des seuils pour le cas d'étude

Pour le filtrage des micro-arrêts, la limite de temps (D_{min}^{arret}) a été définie à 120 secondes. En effet, l'activité la plus courte d'une ambulance selon les experts est une activité logistique (faire le plein du véhicule) durant au minimum deux minutes. Les véhicules de secours étant lourds, il est très fréquent que ceux-ci effectuent des arrêts dans la circulation (feux rouges, stops, *etc.*) avec le frein à main, lorsque la pente est trop forte.

Pour le filtrage des micro-déplacements, la limite de temps (D_{min}^{dep}) a été définie à 15 secondes. Nous avons étudié la répartition de la durée des déplacements. Considérant qu'on nous cherchons à réduire les micro-déplacements, c'est à dire des déplacements d'une très courte durée, nous avons, dans la figure 5.4, fais un focus sur la répartition des durées des déplacements des deux ambulances lorsqu'ils sont inférieurs à 30 secondes. On peut clairement distinguer une forte décroissance après 15 secondes, ce qui motive la sélection de ce seuil. Ce seuil semble d'ailleurs cohérent. Les déplacements de moins de 15 secondes ne semble pas avoir d'explications liées à l'activité en cours. Cependant, le seuil de durée n'est pas entièrement suffisant. En effet, il peut arriver par exemple, comme c'est le cas aux urgences que le véhicule ait besoin de se déplacer pour stationner à un autre endroit. Le seuil de distance minimale de déplacement ($dist_m^{dep}in$) a donc été défini à 30 mètres, en dessous, il ne semble pas pertinent de prendre en considération ce micro-déplacement, notamment à cause de l'imprécision du capteur GPS.

5.3.2.3 STOP et MOVE détectés

Pour rappel, via notre segmentation, nous avons identifié 1609 sous-trajectoires pour les deux véhicules. Une fois la fonction arrêt avec le traitement des micro-arrêts et des micro-déplacements, nous obtenons 1251 nouvelles sous-trajectoires issus des concaténations des potentiels arrêts et déplacements, pour ne plus avoir les micro-arrêts et micro-déplacements non pertinents. A l'aide des 1251 sous-trajectoires, 626 STOP ont été extraits. Concernant les positions étendues représentatives de ces STOP (voir la section 3.3.2 sur les trajectoires sémantiques faibles) :

— 14 STOP, représentant des séquences où le frein à main est tiré et où il y a une perte totale du signal GPS, n'ont pu être représentés géographiquement.

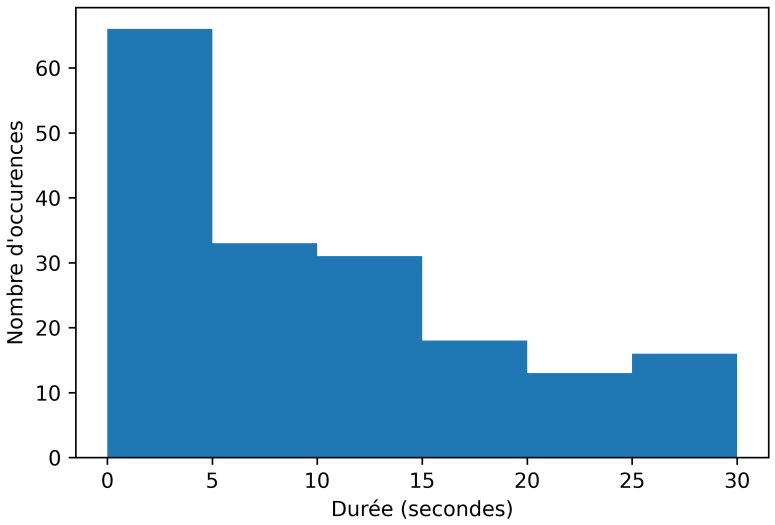


FIGURE 5.4 – Représentation de la durée en secondes des déplacements des véhicules d'une durée inférieure à 30s.

— 612 STOP ont pu être représentés, en deux catégories, en fonction de leur qualité spatiale, évaluée à l'aide de R_{med} (le rayon du cercle centré sur la médiane géométrique contenant au moins 95% des éléments de trajectoires non aberrants qui constituent le STOP) et de $\%_{DP}$ (le pourcentage d'éléments de trajectoires aberrants dans le STOP, eq. 3.2) :
  — 560 STOP où R_{med} est compris entre 0 et 5 mètres, ce qui traduit une très bonne précision de leur position spatiale.
  — 52 STOP où R_{med} est supérieur à 5 mètres, et ou le taux de données perdues ($\%_{DP}$) est assez élevé, ce qui traduit une mauvaise réception du signal lors du STOP.

Ce processus nous a aussi permis d'extraire 625 MOVE au total.

Le tableau 5.3 présente la répartition de la durée de ces sous-trajectoires extraites selon qu'elles soient de type MOVE ou STOP.

Comme pour lors de l'extraction précédente, les durées moyennes sont très supérieures aux durées médianes, ce qui nous indique que des STOP ou des MOVE de très longues durées perturbent le calcul des moyennes. Par ailleurs, la durée maximum pour un MOVE, de près de 8 heures semble aberrante. L'étude de ce MOVE montre que ce cas correspond à une activité essentiellement de mise en réserve du véhicule à la caserne. Durant toute cette période, le frein à main n'a, d'après les données, jamais été tiré. Le véhicule part ensuite en déplacement et termine par un arrêt en dehors de la caserne. La durée minimum d'un STOP prend forcément la valeur définie dans la fonction *arret*, alors que la durée minimum d'un MOVE est de 3 secondes, ce qui

	STOP	MOVE
Nombre	626	625
Durée minimum	00 :02 :00	00 :00 :03
Durée moyenne	00 :54 :37	00 :12 :03
Durée médiane	00 :17 :49	00 :07 :41
Durée maximum	29 :24 :48	07 :52 :23

TABLE 5.3 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires STOP ou MOVE extraites

laisse penser que le véhicule était en déplacement rapide lors de celui-ci (le véhicule a parcouru une distance de plus de 30 mètres en 3 secondes($dist_m^{dep}in$ étant fixé à 30 mètres dans notre étude)

La répartition des données étiquetées comme aberrantes entre les MOVE et les STOP est donnée dans le tableau 5.4. Ce tableau de répartition est sensiblement le même que le précédent. Cependant, on peut remarquer qu'environ 700 données nulles se trouvaient dans des micro-arrêts et se retrouvent donc après fusion dans des MOVE, portant la répartition des données nulles à 60% dans des sous-trajectoires de type MOVE.

Dans notre volonté de montée en sémantique, il nous faut tenter au maximum possible d'associer à nos STOP des lieux d'intérêt métier.

5.3.3 Appariement STOP - POI

Dans cette étape, il s'agit d'appareiller des STOP avec des POI. En effet, la dernière étape de la montée en sémantique faible consiste à tenter d'associer les 626 STOP de l'étape précédente avec les différents POI récupérés dans les données OpenStreetMap présentés dans la sous-section 4.4.3. Pour rappel, nous disposons d'une liste de 92 POI répartis dans 3 catégories différentes. Cette étape présente l'appariement entre les STOP et les POI extraits des étapes précédentes.

5.3.3.1 Méthode

Actuellement, un processus simple est employé pour apparier un STOP a un POI. Un seuil maximum de distance $dist_{max}^{POI}$ est défini. Si le POI le plus proche est à une distance supérieure à ce seuil, alors le lieu n'est pas apparié à un POI. La figure 5.5 représente le résultat de l'appariement entre des STOP (cercles bleus) et des POI externes (étoiles rouges). Les traits noirs définissent les appariements.

Soit $\mathcal{P}$ l'ensemble des POIs $\{poi_1, ..., poi_l\}$, l étant le nombre d'éléments dans $\mathcal{P}$. Soit s un STOP.

	STOP	MOVE
Nombre total de données	2 051 589	452 107
Nombre de données nulles	2 244	3 260
Pourcentage de données nulles relatif au nombre total de données nulles	40,8%	59,2%
Pourcentage de données nulles relatif au nombre total de données par type de sous-trajectoires	0,11%	0,77%
Nombre de données caduques	2 038 333	141 651
Pourcentage de données caduques relatif au nombre total de données caduques	93,5%	6,5%
Pourcentage de données caduques relatif au nombre total de données par type de sous-trajectoires	99,7%	33,6%
Nombre de données vitesse aberrante	78	5 113
Pourcentage de données vitesse aberrante relatif au nombre total de données vitesses aberrantes	1,5%	98,5%
Pourcentage de données vitesse aberrante relatif au nombre total de données par type de sous-trajectoires	<0,1%	1,2%
Nombre de données aberrantes	2 040 655	150 024
Pourcentage de données aberrantes relatif au nombre total de données	93,2%	6,8%
Pourcentage de données aberrantes relatif au nombre total de données par type de sous-trajectoires	99,8%	35,6%

TABLE 5.4 – Répartition des données outliers selon la sémantique faible des sous-trajectoires

Nous définissons $Dist(s, poi)$ comme la distance entre la médiane géométrique du STOP s calculée précédemment et le point, appartenant à l'emprise spatiale du POI poi, le plus proche de cette médiane géométrique. La fonction ST_Distance détermine en utilisant la formule 5.6, le point le plus proche entre deux emprises spatiales géographique et renvoie la distance entre ces points.

$$Dist(s, poi) = ST_Distance(s.getMedianPosition(), poi.geom) \tag{5.9}$$

Soit $\mathcal{P}_s$ un ensemble de POI proches de s (l'ensemble des POI se trouvant à une distance de s inférieure à un seuil $dist_{max}^{POI}$)

$$\mathcal{P}_s = \{poi \in \mathcal{P} | Dist(s, poi) < dist_{max}^{POI}\} \tag{5.10}$$

Dans le cas d'une ambiguïté – s'il existe plus d'un élément dans $\mathcal{P}_s$ – le POI le plus proche de s est choisi. La fonction $IdentifyPOI$ prend en paramètre un stop s ainsi qu'un ensemble de POI ($\mathcal{P}_s$, et retourne l'élément (POI) de $\mathcal{P}_s$ le plus proche de s.

$$IdentifyPOI(s, \mathcal{P}_s) = argmin_{poi \in \mathcal{P}_s}(Dist(s, poi)) \tag{5.11}$$

Si $\mathcal{P}_s$ est vide, alors aucun POI n'est associé au STOP.

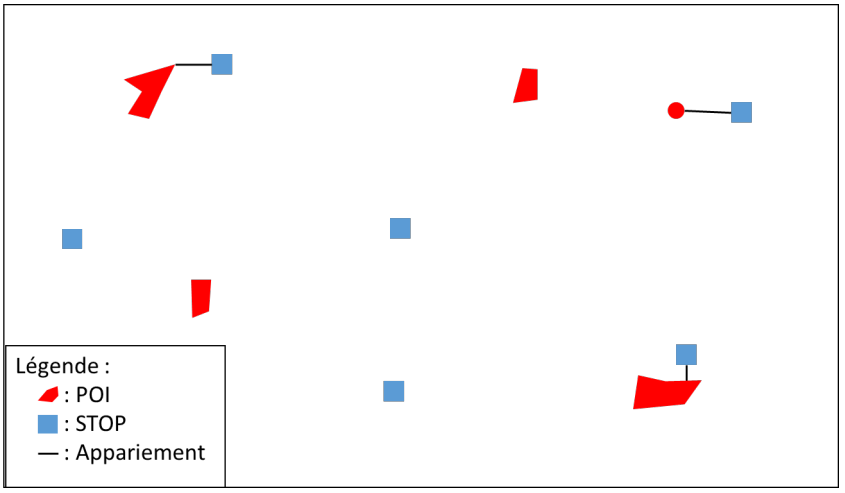


FIGURE 5.5 – Appariement des STOP avec des POI

Ce processus d'appariement est possible car il y a peu d'ambiguïté spatiale : les POI sont assez distants les uns des autres. Dans le cas où les POI seraient plus proches et pourraient poser des problèmes d'ambiguïté spatiale, différentes techniques d'appariement, e.g. [Beeri et al., 2004, Mustière and Devogele, 2008, Olteanu-Raimond et al., 2015], pourraient être utilisées pour répondre à cette problématique. Cependant, dans notre contexte, l'appariement proposé précédemment s'avère suffisant.

5.3.3.2 Appariements obtenus

Pour qu'un POI soit associé à un STOP, il faut que la distance les séparant ne soit pas supérieure à un seuil $dist_{max}^{POI}$. En effet, une distance excessive entre les deux éléments indiquerait que le STOP n'est pas lié au POI en question.

Dans cette optique, nous avons choisi un seuil cohérent avec les données collectées. Les POI collectés peuvent être de deux types géométriques, des polygones représentant les surfaces d'une part, et des points, d'autre part. Ces types sont dépendant des données OpenStreetMap. Lorsque le POI est représenté par un point, il est en général le centroïde d'une forme plus complexe ou le point d'entrée de cette zone (cf. [Feliachi, 2017]). Ainsi, La distance entre l'emplacement sur le parking où l'ambulance se gare et ce point peut rapidement atteindre plusieurs dizaines de mètres.

Aussi, afin de prendre en considération l'emprise réelle des bâtiments, l'imprécision du GPS et la distance éventuelle au lieu de stationnement associé au POI, nous avons défini une distance maximale $dist_{max}^{POI}$ de 50 mètres.

La figure 5.6 représente les STOP ayant été appareillés avec la caserne des sapeurs-pompiers de Tours centre. On distingue que deux STOP au nord et à l'ouest de la caserne se situent dans des zones d'habitation, et peuvent possiblement être des lieux d'intervention. Les points au sud de la caserne sont dus à l'imprécision du GPS. La figure 5.6 représente les STOP ayant été appareillés avec l'hôpital Trousseau on y distingue deux zones différentes, qui correspondent d'une part au sas des urgences de l'hôpital à l'est et le CETRA à l'ouest. Enfin ceux ayant été appareillés avec la station-service proche de la caserne, régulièrement utilisée par les sapeurs-pompiers, figure 5.8. Cette station service étant peu couverte, l'emplacement des STOP est précis et les points sont donc regroupés.

Néanmoins, il n'est pas possible d'associer tous les STOP à des POI. En effet, les lieux d'intervention se trouvent parfois dans les lieux quelconques, sans proximité avec un POI.

Sur notre jeu de données :

- 363 STOP ont été liés à des POI :
  - 203 STOP liés à des casernes
  - 140 STOP liés à des hôpitaux
  - 20 STOP liés à des lieux logistiques
- 263 n'ont pas été liés à un POI dont 14 car ils n'ont pas de représentation spatiale.

Ainsi sur les 626 STOP, plus de 57% des STOP sont effectués sur des lieux d'intérêt pré-identifiés. Les 43% restants sont potentiellement des lieux d'intervention, comme nous pourrons le voir dans la section suivante.

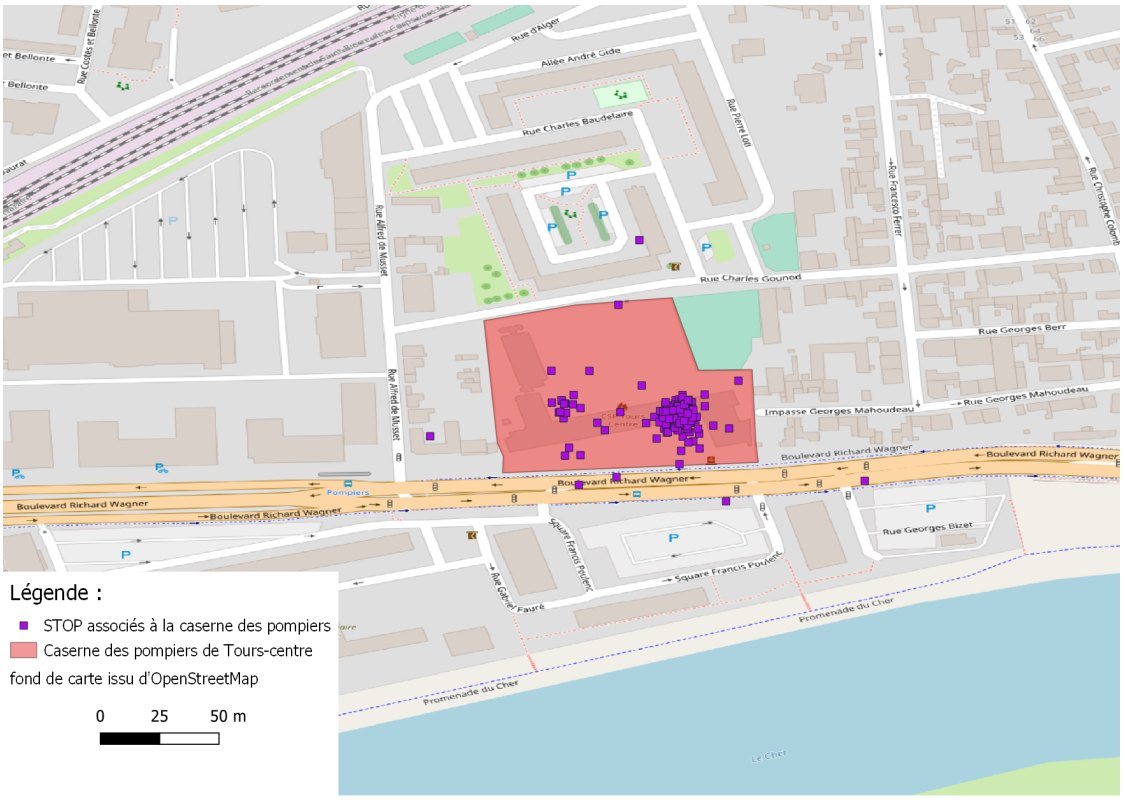


FIGURE 5.6 – Représentation des STOP appareillés avec la caserne des sapeurs-pompiers de Tours centre

Pour les STOP n'ayant pas de POI associés, s'ils n'ont pas de médiane géométrique calculée du fait de valeurs spatiales aberrantes le *poi.poitype.label* "Non positionnable" leur est associé, sinon nous leur affectons le *poi.poitype.label* "Indéterminé".

Le tableau 5.5 présente la répartition des STOP ayant un positionnement spatial de faible qualité dans les différents types de POI identifiés.

Il semble normal que pour les POI logistique (station-service, parking du SDIS 37) il n'y ait pas de POI de faible qualité car les arrêts sont faits dans des espaces ouverts et dégagés. Pour ce qui est des hôpitaux et de la caserne, comme une partie des arrêts se font dans des lieux couverts avec possibles pertes du signal, il est aussi logique d'avoir plus de cas problématiques. Pour les arrêts n'ayant pas de POI associés (indéterminés), de nombreux facteurs peuvent entrer en considération notamment l'effet rebond qui peut perturber le signal GPS en milieu urbain dense.

À ce stade, nous disposons donc de sous-trajectoires montées en sémantique faible. Elles peuvent être de type MOVE ou bien STOP. Les STOP pouvant être appariés avec des POI. La prochaine section présente le processus de montée en sémantique forte.

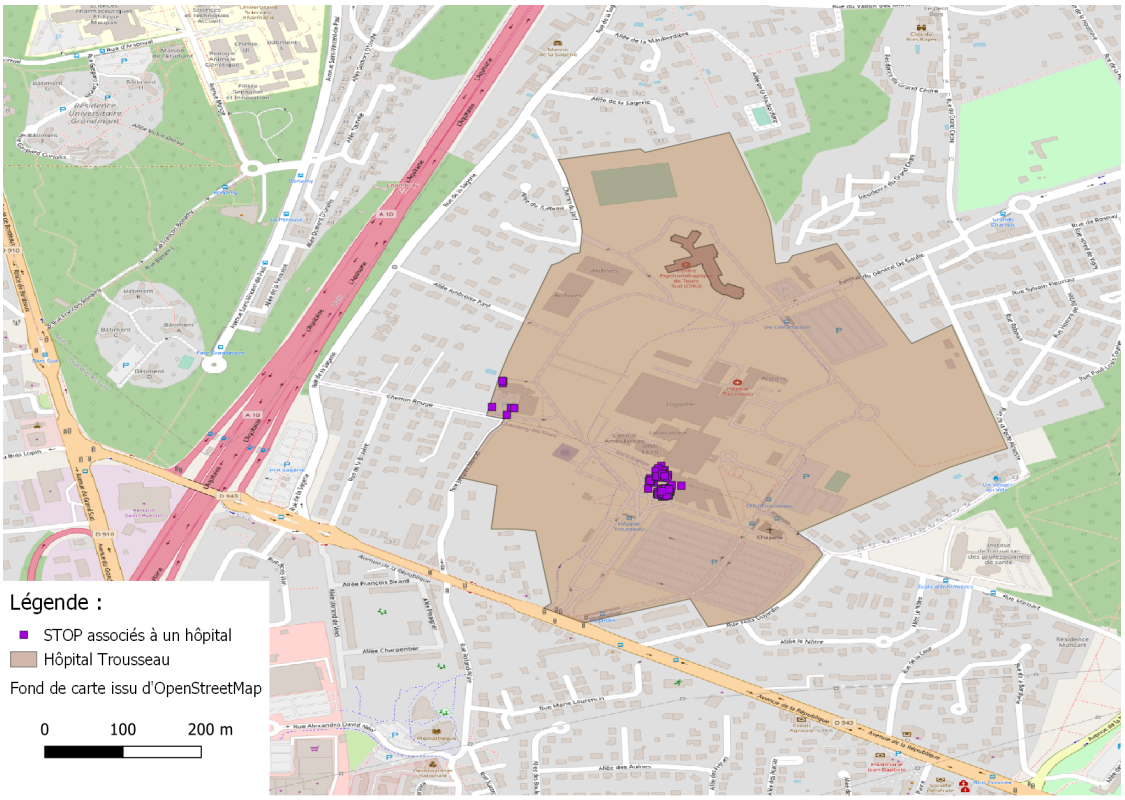


FIGURE 5.7 – Représentation des STOP appareillés avec l'hôpital Trousseau

5.4 Montée en sémantique forte

La dernière étape du processus de montée en sémantique est l'affectation d'une activité à une sous-trajectoire.

Pour pouvoir monter en sémantique forte, nous avons besoin, comme indiqué dans la section 3.3.3 d'une ontologie métier ou à minima d'une liste d'activités métiers. Dans notre cas, une liste est construite en fonction des activités principales recensées des ambulances lors de leurs déplacements.

À partir de cette liste, nous associons à chaque sous-trajectoires de sémantique faible obtenues un des labels de ladite liste.

5.4.1 Construction de la liste des activités

La liste des activités a été établie sur la base du fonctionnement classique d'une ambulance.

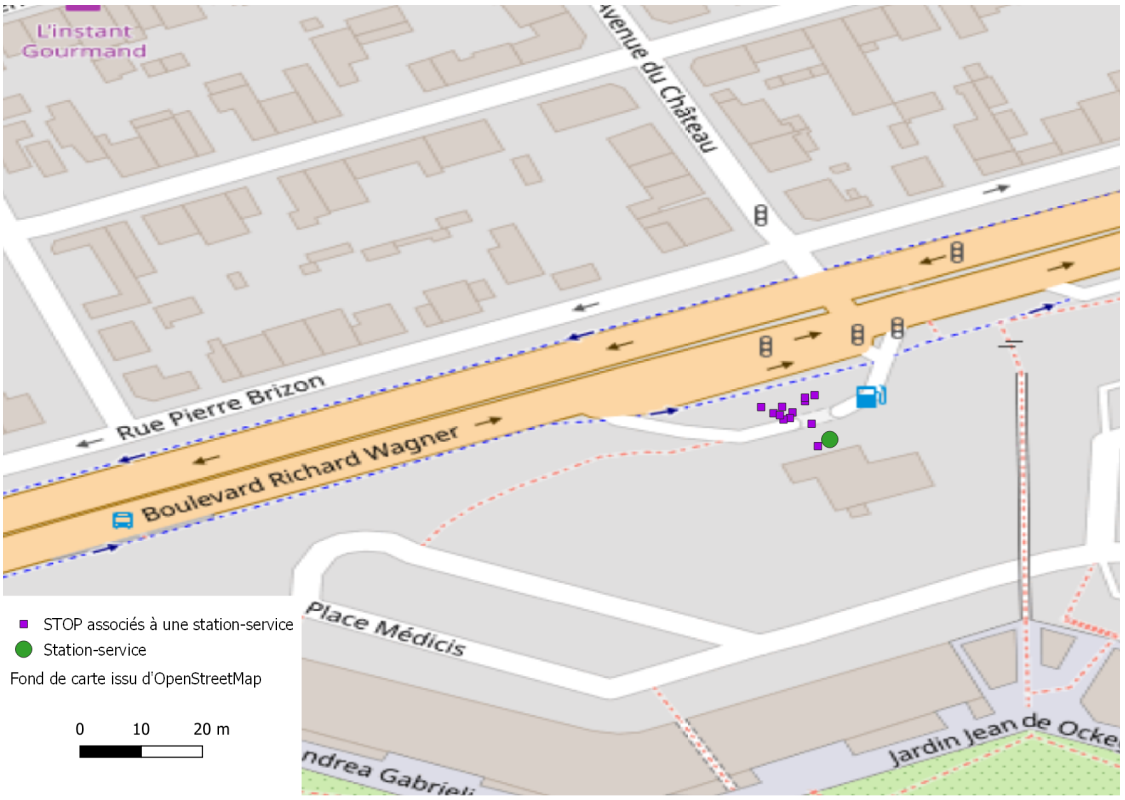


FIGURE 5.8 – Représentation des STOP appareillés avec une station-service

Pour rappel, comme vu dans la section 4.2, les ambulances de pompiers restent en général stationnées dans leurs caserne d'affectation, prête à intervenir (STOPActivityType : "Attente caserne"). Lorsqu'un appel est reçu, l'ambulance se dirige rapidement vers le lieu indiqué par l'appelant (MoveActivityType : "Départ vers intervention").

Arrivés sur les lieux, les pompiers s'occupent du patient, afin de juger de son état, et prodiguer les premiers soins si nécessaire (STOPActivityType : "Prise en charge du patient sur lieu d'intervention"). Ensuite, si son état le nécessite, le patient est transporté à l'hôpital (MoveActivityType : "Transport vers hôpital"). Lors de l'arrivée à l'hôpital, les pompiers font enregistrer l'arrivée du patient et effectuent le transfert entre le véhicule et le service des urgences pour qu'il soit pris en charge (STOPActivityType : "Prise en charge du patient à l'hôpital").

Une fois le patient pris en charge, ou lorsque son état n'a pas nécessité un transport à l'hôpital, l'ambulance peut retourner à sa caserne d'affectation (MoveActivityType : "Retour caserne").

Cependant, le véhicule peut, sur le chemin du retour, effectuer des arrêts afin de faire des activités considérées comme un STOPActivityType de type "Logistique", à l'instar de faire le plein de carburant, nettoyer la cellule de l'ambulance, ou recharger

Type de POI associés	Nombre de STOP ayant un $R_{med} \leq 5$	Nombre de STOP ayant un $R_{med} > 5$
Caserne	188	15
Hôpital	130	10
Logistique	20	0
Indéterminé	222	27

TABLE 5.5 – Répartition des STOP selon la qualité de leur positionnement et des types de POI associés

le matériel médical tel que des bouteilles d'oxygène dans un lieu près de l'hôpital prévu à cet effet. Le trajet jusqu'à ces lieux (station-service, station sanitaire) est considéré comme un trajet logistique dans la liste des MoveActivityType. Une fois l'activité logistique terminée, le véhicule retourne à sa caserne d'affectation (MoveActivityType : "Retour caserne"). Le véhicule restera ensuite en attente jusqu'au prochain appel reçu (STOPActivityType : "Attente caserne").

Dans ce chapitre, nous traitons pas la question des reroutages. Ainsi, si le véhicule est appelé pour une nouvelle intervention lors de son retour à la caserne, l'ensemble de ce déplacement est considéré comme un trajet vers lieu d'intervention (MoveActivityType : "Départ vers intervention").

La liste globale des activités considérées recense toutes les activités possibles d'une ambulance. Elle est décomposée en deux sous-listes. La liste des STOPActivityType regroupe les activités potentielles lors d'un STOP. La liste des MOVEActivityType regroupe les activités potentielles lors d'un MOVE.

5.4.1.1 Liste des STOPActivityType

Dans cette thèse, en concertation avec les experts métiers, seulement 5 types d'activités sont considérées pour les STOP :

— Prise en charge du patient sur lieu d'intervention .

— Prise en charge du patient à l'hôpital .

— Logistique .

— Attente caserne .

— Arrêt indéterminé .

Les arrêts indéterminés ont été ajoutés afin de prendre en considération des possibles arrêts n'entrant pas dans les parcours typiques présentés dans la section 4.2.

À ces activités, du fait de la présence de données spatialement aberrantes, nous avons ajouté l'activité :

— Arrêt non positionnable .

5.4.1.2 Liste des MOVEActivityType

Pour cette liste de STOPActivityType, la liste des MOVEActivityType considérées est la suivante :

— Départ vers intervention : 
Déplacement lorsque l'ambulance se dirige vers un lieu d'intervention pour prendre en charge un patient.

— Transport vers hôpital :
Déplacement lorsque l'ambulance transporte un patient du lieu d'intervention jusqu'à un hôpital.

— Trajet logistique :
Déplacement qui a pour objectif d'arriver à un POI de type logistique.

— Retour caserne :
Déplacement qui à pour objectif d'arriver à la caserne.

— Trajet indéterminé : Déplacement dont l'activité ne peut être déterminée.

Les trajets indéterminés peuvent être de trois types :

— Trajet indéterminé vers Hôpital :
Déplacement dont on ne connaît pas le début, mais qui arrive dans un hôpital.

— Trajet indéterminé vers arrêt indéterminé :
Déplacement vers un lieu indéterminé.

— Trajet indéterminé vers arrêt non positionnable :
Déplacement vers un lieu non positionnable géographiquement.

Une fois les listes d'activités déterminées, il faut ensuite associer ces activités aux MOVE et STOP précédemment extraits. La section suivante présente la méthode d'association entre les activités et les STOP et MOVE.

5.4.2 Association des sous-trajectoires sémantiques faibles et activités

Chaque sous-trajectoire sémantique faible (STOP et MOVE) doit être associée à une seule activité dans le cadre de la montée en sémantique forte.

Pour associer une activité à un STOP, un ensemble de règles d'inférence métiers est défini comme le propose [Browne, 2009]. À savoir, quand les conditions sont réunies, alors un cas précis est appliqué. Le modèle utilisé ne contient pas de règles contradictoires.

Les règles sont stockées sous forme de conditions, pour déterminer si une règle peut être appliquée à une sous trajectoire, pour chaque règle, chaque condition est vérifiée. Si une condition n'est pas vérifiée, le processus ne retient pas l'activité liée à cette règle et passe à la règle suivante. Si toutes les conditions de la règle sont vérifiées, alors l'activité est associée à la sous trajectoire.

Les règles n'étant pas identiques ni contradictoires, l'ordre de leur passage dans l'algorithme n'a pas d'importance. Chaque condition étant déterminante, l'ordre de vérification de celles-ci n'a pas d'incidence sur le résultat. Cependant, l'ordre d'exécution des algorithmes 6 et 7 est important, car le second requiert l'exécution au préalable du premier.

Nous présentons dans un premier temps l'association des STOP aux STOPActivityType puis celle des MOVE aux MOVEActivityType.

5.4.2.1 Association des STOP

Comme indiqué précédemment, dans ce chapitre, 6 activités sont identifiées pour les STOP. Des règles métiers simples sont utilisées. Après concertation avec des experts métiers, les règles suivantes ont été obtenues :

- Si STOP n'a pas de médiane géométrique, alors l'activité est "Arrêt non positionnable"
- Si STOP et POI de type hôpital alors l'activité est "Prise en charge du patient à l'hôpital"
- Si STOP et POI de type logistique alors l'activité est "logistique"
- Si STOP et POI de type caserne alors l'activité est "Attente caserne"
- Si STOP de plus de D_{min}^{inter} minutes et pas de POI associé alors l'activité est "Prise en charge du patient sur lieu d'intervention"
- Si aucune activité n'a été définie, alors l'activité est "Arrêt indéterminé"

L'algorithme 6 présente la méthode d'association des activités aux STOP.

Une fois les STOP étiquetés, l'étape restante de la montée en sémantique est labellisation des MOVE.

5.4.2.2 Association des MOVE

Les activités associées aux MOVE sont également définies avec des experts métiers à l'aide de règles simples.

Algorithm 6 Assignation d'une activité à un STOP

Require: s //STOP étudié
  $duree \leftarrow (s.getEndPosition().t - s.getStartPosition().t)$ //durée du STOP
  switch ($s.poi.poitype.label$)
    case "hopital"**:**
      $s.STOPActivityType.label \leftarrow$ "Prise en charge du patient à l'hôpital"
    case "logistique"**:**
      $s.STOPActivityType.label \leftarrow$ "Logistique"
    case "caserne"**:**
      $s.STOPActivityType.label \leftarrow$ "Attente caserne"
    case "Non positionnable"**:**
      $s.STOPActivityType.label \leftarrow$ "Arrêt non positionnable"
    default:
      //cas où $s.poi.poitype.label$=="Indéterminé"
      if $duree > D_m^{inter} in$ **then**
        $s.STOPActivityType.label \leftarrow$ "Prise en charge du patient sur lieu d'intervention"
      else
        $s.STOPActivityType.label \leftarrow$ "Arrêt indéterminé"
      end if
  end switch

— Si activité précédente était "Prise en charge d'un patient sur le lieu d'intervention" et activité suivante est prise en charge du patient à l'hôpital, alors l'activité est "Transport vers hôpital"

— Si activité suivante est "Prise en charge du patient sur lieu d'intervention" alors l'activité est "Départ vers intervention"

— Si activité suivante est "Logistique", alors l'activité est "Trajet logistique"

— Si activité suivante est "Attente caserne", alors l'activité est "Retour caserne"

Pour rappel, dans notre approche basée frein à main et GPS, nous ne traitons pas le problème du reroutage car il est difficile d'identifier simplement avec les positions GPS et du frein à main les possibles reroutage. En effet, il faudrait pour cela développer et mettre en oeuvre des algorithmes d'apprentissage d'identification de possibles changements (concept drift detection/concept shift detection [Gama et al., 2004, Krempl et al., 2011, Sethi and Kantardzic, 2017], ou identification des anomalies [Wu et al., 2019]) dans les données GPS. Ces techniques sont coûteuses et ne sont l'objet de cette thèse. Par contre, l'apport d'informations venant des autres capteurs peuvent potentiellement aider, comme nous le verrons dans le chapitre 6.

Nos MOVE sont donc précédés et suivis de STOP, et nos STOP sont eux encadrés par des MOVE, à l'exception des STOP et MOVE en début et fin de trajectoire.

L'algorithme 7 présente la méthode d'association des MOVEActivityType aux MOVE.

Algorithm 7 Assignation d'une activité à un MOVE

Require: m //le MOVE étudié.
Require: s_p //le STOP précédent le MOVE.
Require: s_n //le STOP suivant le MOVE.
  $A_{s_p} \leftarrow s_p.STOPActivityType.label$ //l'activité liée à s_p.
  $A_{s_n} \leftarrow s_n.STOPActivityType.label$ //l'activité liée à s_n.
  switch (A_{s_n})
    case "Prise en charge du patient à l'hôpital"**:**
      if A_{s_p} == "Prise en charge du patient sur lieu d'intervention" **then**
        $m.MOVEActivityType.label \leftarrow$ "Transport vers hôpital"
      else
        $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers Hôpital"
      end if
    case "Logistique"**:**
      $m.MOVEActivityType.label \leftarrow$ "Logistique"
    case "Prise en charge du patient sur lieu d'intervention"**:**
      $m.MOVEActivityType.label \leftarrow$ "Départ vers intervention"
    case "Attente caserne"**:**
      $m.MOVEActivityType.label \leftarrow$ "Retour caserne"
    case "Arrêt non positionnable"**:**
      $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt non positionnable"
    default:
      if s_n.empty **then**
        $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt non positionnable"
      else
        $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt indéterminé"
      end if
  end switch

Il n'existe pas de règles contradictoires. Une activité "Trajet indéterminé" est associée avec le MOVE si aucune règle principale ne correspond.

5.4.3 Affectation des activités

5.4.3.1 Affectation des activités aux STOP

Dans le cas d'ambulance de pompiers, et après concertation avec les experts métiers, nous avons défini D_{min}^{inter} tel que $D_{min}^{inter} = 600$ *secondes*. En effet, lors de la plupart des interventions, il y a une prise de contact avec la victime (ou ses proches, ou les témoins). De plus, les pompiers font un bilan de la victime et le transmettent ensuite au médecin régulateur du SAMU. C'est pourquoi le seuil de dix minutes parait pertinent. Cependant, nous ne pouvons pas exclure des interventions beaucoup plus rapides lors de cas atypiques. Par exemple, lorsque qu'une ambulance est appelée en renfort pour transporter un autre patient, et que dans ce cas, la partie bilan et prise de contact a déjà été effectuée par les pompiers déjà sur place.

Les règles définissant les STOP n'utilisent pas d'informations sémantiques des MOVE précédents ou suivants. Les résultats suivants sont obtenus :

— 154 STOP de type "Prise en charge du patient sur lieu d'intervention"

— 140 STOP de type "Prise en charge du patient à l'hôpital"

— 20 STOP de type "Logistique"

— 203 STOP de type "Attente caserne"

— 95 STOP n'ont pu être liés à des activités et sont donc de type "Arrêt indéterminé"

— 14 STOP n'ont pu être liés à des activités car ils n'étaient pas positionnables.

Après présentation à des experts métiers, ces résultats semblent pertinents. En effet, lors d'une intervention, le transport du patient à l'hôpital n'est pas systématique, et il arrive régulièrement que l'intervention des sapeurs-pompiers suffise à régler l'incident.

Une lecture rapide des résultats voudrait qu'a priori plus de 90% des STOP "Prise en charges patient sur le lieu d'intervention" ont donné lieu à une prise en charge à l'hôpital. Cependant, bien qu'ayant été discutée avec les experts métiers, nous pouvons aussi questionner la pertinence de la règle pour l'identification des prises en charge patient sur le lieu d'intervention. En effet, la durée relativement réduite (10 minutes) pour considérer que l'activité est une prise en charge d'un patient sur le lieu d'intervention – choisie afin de ne pas exclure des possibles interventions courtes – est sans doute trop réduite pour éviter de considérer des activités annexes qui auraient dû être "Arrêt indéterminé" en "Prise en charge patient sur le lieu d'intervention".

De plus, vu qu'il y a plus de STOP "Prise en charge du patient sur le lieu d'intervention" que de STOP "Attente caserne" cela incite à penser que plusieurs trajets ont dû être reroutés (une ambulance peut être rappelée sur une intervention dès la sortie de l'hôpital).

Nombre	203	140	154	20	95	14
Durée minimum	00 :02 :17	00 :02 :10	00 :10 :01	00 :02 :22	00 :02 :00	00 :02 :08
Durée moyenne	02 :03 :40	00 :17 :03	00 :23 :59	00 :09 :55	00 :06 :14	02 :37 :54
Durée médiane	00 :49 :41	00 :16 :00	00 :17 :48	00 :03 :29	00 :06 :20	00 :30 :00
Durée maximum	29 :24 :48	00 :47 :50	02 :47 :38	00 :54 :10	00 :10 :00	13 :00 :17

TABLE 5.6 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des STOP selon l'activité

Le nombre des STOP "Logistique" identifiés est également compréhensible. Dans une démarche préventive de panne de carburant, il est d'usage, chez les sapeurs-pompiers de Touraine de faire le plein de carburant à un taux défini ; un véhicule ne devrait pas repartir en intervention si le volume de carburant est inférieur à 75%. Les STOP non liés à des activités ont été analysés manuellement. Pour la plupart, il s'agit d'activités externes aux activités classiques.

Le tableau 5.6 présente la répartition des durées des STOP, en secondes, en fonction de l'activité associée. Ces résultats confirment que les STOP de type "Attente caserne" ont une durée médiane plus haute que les autres types d'activités. Nous pouvons également constater que les STOP de type "Prise en charge du patient sur lieu d'intervention" et "Prise en charge du patient à l'hôpital" durent plus longtemps que les autres types d'activités positionnables. Cela est logique car une intervention prend nécessairement plus de temps que de faire le plein de carburant d'un véhicule. On s'aperçoit aussi que les STOP non positionnables ont des caractéristiques de durées proches des STOP de type "Attente caserne" même si la durée médiane est intermédiaire. Comme il n'y a que 14 STOP de ce type, il n'est pas possible d'obtenir des statistiques robustes qui permettraient de les réaffecter.

Le tableau 5.7 présente la répartition des données aberrantes, en fonction de l'activité associée. On distingue que la plupart des données nulles (66,9%) se trouvent lorsque le véhicule est en attente à la caserne. Cependant, c'est lors d'une activité logistique que le taux de données nulles est le plus élevé (0.7%). Les activités étant liées à des STOP, il est normal que la plupart des données constituant ces STOP soient considérées comme caduques. Pour tous les types de STOP, plus de 98% des données sont aberrantes. Cependant, les problématiques possiblement dues à l'effet pelote se trouvent extrêmement réduites (le taux de données ayant une vitesse aberrante est inférieur à 0,1%) du fait, probablement, du fort taux de données répétées.

Les STOP étant maintenant liés à des activités, il est possible de passer à l'étape suivante du processus et enrichir la sémantique des MOVE.

						
Nombre total de données	1 506 432	143 345	221 637	11 919	35 616	132 640
Nombre de données nulles	1 501	105	181	82	65	310
Pourcentage relatif au nombre total de données nulles	66,9%	4,7%	8,1%	3,7%	2,9%	13,8%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	0,1%	0,1%	0,1%	0,7%	0,2%	0,2%
Nombre de données caduques	1 503 588	142 462	212 774	11 737	34 428	132 344
Pourcentage relatif au nombre total de données caduques	73,8%	7%	10,5%	0,6%	1,7%	6,5%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	99,8%	99,3%	99,4%	98,3%	98,1%	99,8%
Nombre de vitesse aberrante	37	2	36	2	1	0
Pourcentage relatif au nombre total de vitesse aberrante	47,4%	2,6%	46,2%	2,6%	1,3%	0%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	<0,1%	<0,1%	<0,1%	<0,1%	<0,1%	0%
Nombre de données aberrantes	1 505 126	142 569	213 991	11 821	34 494	132 654
Pourcentage relatif au nombre total de données aberrantes	73,8%	7%	10,5%	0,6%	1,7%	6,5%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	99,9%	99,3%	99,5%	99%	98,3%	100%

TABLE 5.7 – Répartition des données aberrantes selon les types d'activités des STOP

Nombre	202	154	73	20	67	95	14
Durée min.	00 :00 :03	00 :00 :03	00 :03 :08	00 :00 :15	00 :00 :09	00 :00 :03	00 :00 :03
Durée moy.	00 :12 :07	00 :13 :40	00 :10 :12	00 :09 :53	00 :06 :24	00 :15 :12	00 :11 :33
Durée méd.	00 :08 :25	00 :05 :40	00 :10 :06	00 :06 :00	00 :06 :43	00 :05 :24	00 :09 :33
Durée max.	07 :35 :23	05 :48 :30	00 :18 :52	00 :51 :12	00 :44 :30	07 :52 :23	01 :01 :11

TABLE 5.8 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des MOVE selon l'activité

5.4.3.2 Affectation des activités aux MOVE

Comme vu précédemment, les règles pour attribuer une activité à un MOVE utilisent les activités des STOP précédent et suivant.

À l'aide des différentes règles métier définies dans la section 5.4.1.2, les résultats suivants ont été obtenus :

- 73 "Transports vers hôpital"
- 154 "Départs vers intervention"
- 20 "Trajets logistique"
- 202 "Retours caserne"
- 176 "Trajets indéterminés" :
  - 14 "Trajet indéterminé - vers arrêt non positionnable"
  - 67 "Trajet indéterminé - vers Hôpital"
  - 95 "Trajet indéterminé - vers arrêt indéterminé"

Seuls 28% des MOVE et 17% des STOP n'ont pas pu monté en sémantique forte. Cela montre donc que la montée en sémantique à l'aide des informations disponibles avec les limites de nos capteurs permet d'identifier des mouvements types des ambulances de pompier, et que dans leur immense majorité, ceux-ci semblent motivés par les besoins métiers.

La figure 5.9 présente les MOVE associés à des "Retour caserne". Par contrainte de confidentialité, seuls les retours à partir de l'hôpital sont affichés. Les autres trajectoires ne sont pas représentées, car liées à des lieux d'intervention. On remarque que les chemins empruntés pour les retours des hôpitaux sont quasiment identiques. Quelques trajets inhabituels sont également identifiés. Ils sont représentés sur la figure par des lignes en tirets sur la carte (axe principal coupé, détour volontaire ...).

Le tableau 5.8 présente la répartition des durées des MOVE, en secondes, en fonction de l'activité associée. Nous observons de très longues durées maximales. Ces MOVE sont tous, après étude, liés à la caserne. En effet, le déplacement de

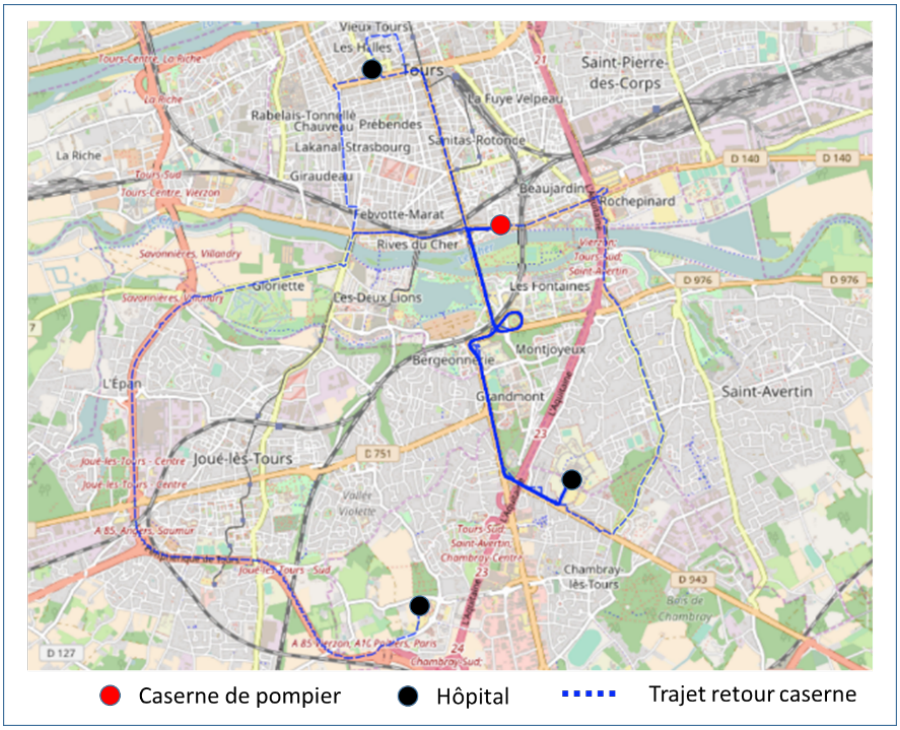


FIGURE 5.9 – Représentation des MOVE liés à une activité "Retour à la caserne" partant d'un hôpital

durée maximale de type "Trajet indéterminé" correspond au passage du véhicule en remise à la caserne. En effet, il arrive régulièrement que le véhicule soit remisé, et dans ce cas, la procédure veut que ce soit sans le frein à main. Pour les autres types d'activités, les durées maximales de plus de 30 minutes peuvent s'expliquer par la fusion des micro-arrêts ainsi que le trafic urbain en situation non urgente.

Le tableau 5.9 présente la répartition des données aberrantes, en fonction de l'activité associée aux MOVE. On distingue que plus de 60% des données aberrantes se trouvent lors de MOVE liés au départ ou à l'arrivée à la caserne, ce qui représente plus de 30% des données récoltées lors de ces déplacements. Il n'est pas étonnant d'observer que les trajets vers un lieu indéterminé ou vers un lieu non positionnable portent aussi un fort taux de données aberrantes ($> 39\%$). Nous constatons également que plus de 60% des données de vitesse aberrante ont lieu lors d'activités liées à une intervention (Départ vers intervention ou transport vers l'hôpital). Nous remarquons que 21,3% des données de type vitesse aberrante se trouvent sur des MOVE de type "Indéterminé vers hôpital" pouvant nous faire penser que certains d'entre eux sont des transports vers l'hôpital.

La section suivante présentera une analyse des résultats obtenus avec uniquement la prise en compte des capteurs GPS et frein à main.

							
Nombre total de données	146 911	126 412	44 733	11 873	25 787	86 681	9 710
Nombre de données nulles	895	1 720	0	27	42	569	7
Pourcentage relatif au nombre total de données nulles	27,5%	52,8%	0%	0,8%	1,3%	17,5%	0,2%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	0,63%	1,6%	0%	0,3%	0,2%	0,7%	0,1%
Nombre de données caduques	43 743	44 188	553	252	629	48 674	3 612
Pourcentage relatif au nombre total de données caduques	30,9%	31,2%	0,4%	0,2%	0,4%	34,4%	2,5%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	30,9%	41,1%	1,3%	2,8%	2,5%	51,1%	37,2%
% Nombre de vitesse aberrante	306	1 321	1 786	87	1 092	349	172
Pourcentage relatif au nombre total de vitesse aberrante	6%	25,8%	34,9%	1,7%	21,3%	6,8%	3,4%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	0,2%	1,2%	4,1%	1%	4,3%	0,4%	1,8%
Nombre de données aberrantes	44 944	47 229	2 339	366	1 763	49 592	3 791
Pourcentage relatif au nombre total données aberrantes	30%	31,5%	1,6%	0,2%	1,2%	33%	2,5%
Pourcentage relatif au nombre total de données par type de sous-trajectoires	31,7%	44%	5,4%	4,1%	6,9%	58,2%	39,1%

TABLE 5.9 – Répartition des données aberrantes selon les types d'activités des MOVE

5.5 Analyse des trajectoires sémantiques obtenues

Une fois la montée en sémantique forte effectuée, nous nous intéressons à l'analyse de ces extractions. La première partie de cette section présente une méthode pour identifier des arrêts fréquents à partir de STOP montés en sémantique faible. La deuxième partie de cette section présente l'analyse de tournée des véhicules étudiés. C'est à dire l'extraction des différentes activités effectuées par les ambulances de la sortie de la caserne jusqu'à leur retour.

5.5.1 Identification des arrêts fréquents

Une fois les STOP localisés, nous pouvons les associer à des lieux d'arrêt fréquents. Pour trouver les arrêts fréquents des véhicules, un clustering de type DBSCAN des STOP est effectué en fonction de leur position géographique médiane. Le fonctionnement de cette méthode a été détaillé dans la section 2.2.3.

Dans notre étude, la valeur d'ϵ correspond à 30 mètres. En effet, nous recherchons ici des lieux d'arrêts fréquents et il nous semble pertinent de considérer deux arrêts distants de moins de 30 mètres comme étant liés à un même lieu. Pour le volume de données récolté (15 jours par véhicule), les experts métiers considèrent que 3 arrêts distincts au même endroit constituent un lieu d'arrêt fréquent. Aussi, le nombre $MinPts$ a été fixé à 3.

Nous obtenons comme résultat l'association de 367 STOP, répartis dans 15 zones (Clusters) différentes. La première zone regroupe 198 STOP. Elle est positionnée près de la caserne de pompiers, lieu de stationnement habituel des véhicules étudiés. Cette zone est visible sur la figure 5.10. Des ajustements des paramètres du clustering permettraient de distinguer deux zones d'arrêt pour cette caserne (lorsque le véhicule se gare dans la cour principale ou dans la cours arrière).

Dans la figure 5.11(b), la zone regroupant 12 STOP est affichée, ces derniers sont associés à un POI de type logistique, une station-service.

L'analyse des clusters présents sur l'empreinte spatiale de l'hôpital Trousseau (voir figure 5.12) montre qu'il est possible de distinguer 3 zones différentes pour un même POI. On observe qu'en réalité, les clusters en bleu et rouge correspondent au CETRA (CEntre de Traitement et de Régulation de l'Alerte), et donc à une activité plutôt logistique, tandis que le marron correspond bien à l'hôpital Trousseau. De plus, pour le cluster marron, nous pouvons distinguer visuellement deux zones, qu'un autre paramétrage de DBSCAN permettrait sans doute de différencier. La première zone au Nord correspond à des STOP au niveau du parking de stationnement alors que la zone au Sud se situe au niveau du SAS de dépose de patients aux urgences.

Le tableau 5.10 présente la répartition des STOP en fonction de leur appartenance à un cluster et de l'activité affectée à ces STOP. Seuls les STOP représentables spatialement sont traités. Mis à part quelques éléments individuels séparés, tous les

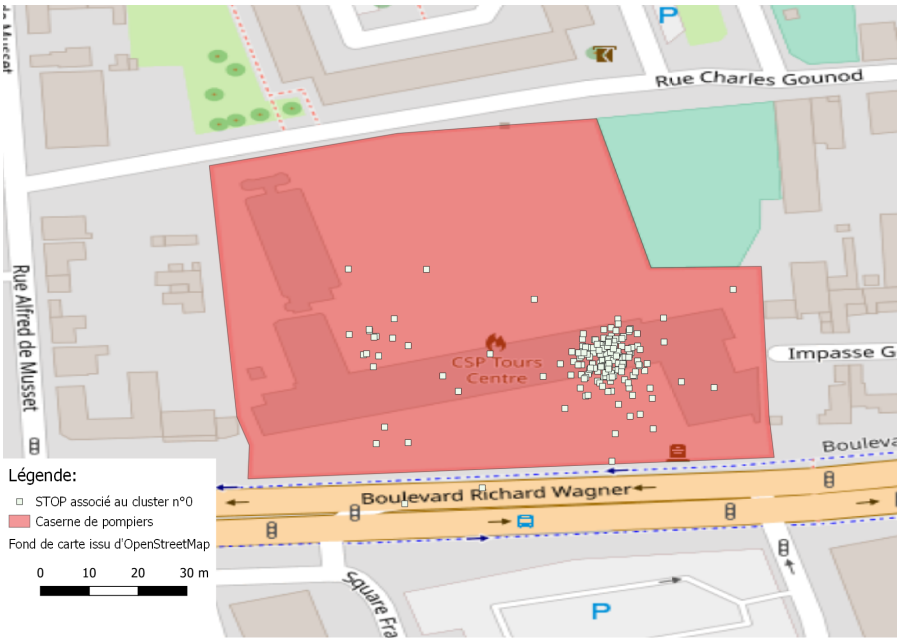


FIGURE 5.10 – Cluster regroupant des STOP près de la caserne de pompiers

STOP associés à des POI sont généralement regroupés dans des clusters homogènes sauf pour le cluster 6. Par exemple, les STOP de type "Attente caserne" se trouvent tous dans le même cluster.

Les STOP présents dans le cluster 6 représentent les STOP au SDIS 37. Cependant, nous pouvons voir que pour deux des STOP, l'association avec le POI a échoué pour des raisons d'emprise spatiale du POI. Pour les STOP non associés à des POI, nous retrouvons souvent des STOP de type "Prise en charge d'un patient sur le lieu d'intervention" regroupé avec des STOP de type "Arrêt indéterminé". Cela peut potentiellement indiquer des lieux réguliers d'interventions.

279 STOP (plus de 40%) ne sont pas associés à un lieu d'arrêt fréquent par l'algorithme de clustering. Ce nombre important est logique. Une ambulance effectue généralement des interventions sur des lieux d'incident très divers. De plus, le transport du patient (et donc l'arrêt à l'hôpital) n'est obligatoire que si son état le nécessite.

Cette recherche de points d'arrêts fréquents nous a aussi permis de trouver des lieux d'intérêts non répertoriés (Pas de POI lié par exemple) autour de la gare ou de la place Plumereau, cependant, il ne nous est pas possible de les montrer dans cette thèse, car étant possiblement liés à des interventions auprès de patient. En effet, pour des raisons d'anonymat, les données traitant des interventions (zone d'intervention fréquente ou répartition géographique) ne sont pas illustrées dans le manuscrit.

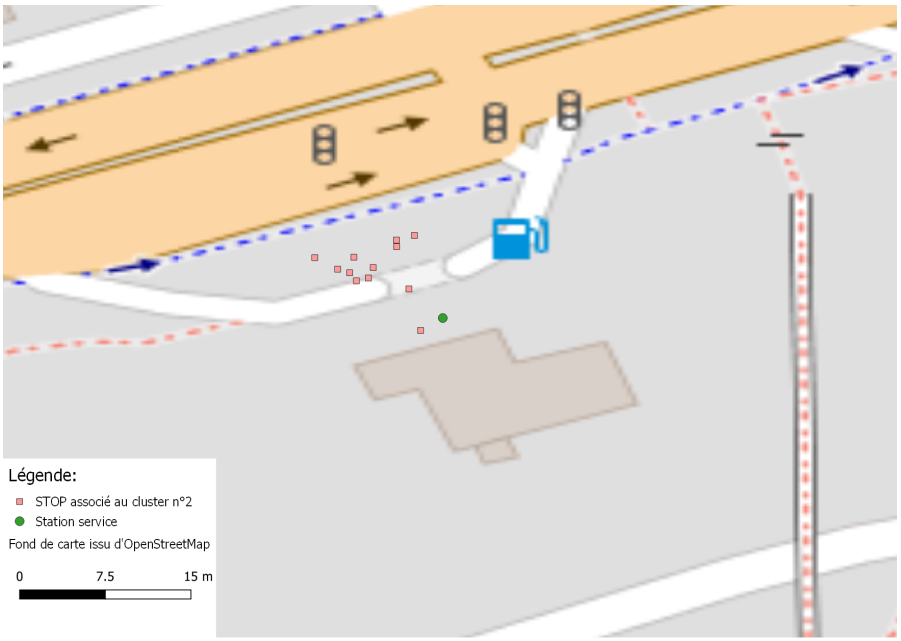


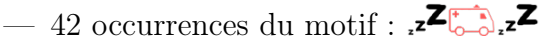
FIGURE 5.11 – Cluster regroupant des STOP près d'une station-service

5.5.2 Analyse des tournées

Afin de procéder à différentes analyses, nous avons décidé d'extraire les parcours partant de la caserne et revenant à la caserne. Nous appelons ces parcours des tournées. Comme indiqué dans la section 4.2, ces tournées ont un sens car un véhicule de pompiers a toujours une caserne d'affectation, et donc part de sa caserne pour y revenir. Une tournée correspond à la liste des activités effectuées par le véhicule, de la sortie de la caserne jusqu'à son retour. Nous ne considérons que les suites d'activités complètes : Une tournée commence et finit toujours par une attente caserne.

Au total sur notre étude, nous avons extraits 201 tournées – i.e. 201 séquences du type "Attente caserne" → ... → "Attente caserne" – dont 70 pour le premier véhicule et 131 pour le deuxième.

Une analyse fréquentielle de ces tournées nous a permis de repérer des motifs typiques de séquences de sous-trajectoires des ambulances, à l'instar de :

— 42 occurrences du motif : 

— 28 occurrences du motif correspondant à la tournée typique 1 (cf. Figure 4.1) : 

— 25 occurrences du motif correspondant à la tournée typique 2 (cf. Figure 4.2). 

— 11 séquences similaires aux deux précédentes avec un "Arrêt logistique" sur le chemin du retour, i.e. tournées typiques 3 et 4, cf. Figure 4.3.

— 9 occurrences du motif : 

Type de STOP					
Cluster 0	198				
Cluster 1			107		
Cluster 2				12	
Cluster 3			4		
Cluster 4		3			
Cluster 5			5		
Cluster 6		2		2	
Cluster 7			5		
Cluster 8		3			1
Cluster 9		3			1
Cluster 10		3			
Cluster 11			8		
Cluster 12					3
Cluster 13	0	1			3
Cluster 14	0		3		
En dehors d'un cluster	5	139	8	6	87

TABLE 5.10 – Répartition des STOP en fonction de leur cluster et de l'activité associée

Type de STOP					
Nombre de tournées	19	91	14	113	69
Pourcentage de tournées	9,45%	45,27%	6,97%	56,22%	34,33%

TABLE 5.11 – Nombre de tournées contenant au moins une fois les différents types de STOP

Type de MOVE						
Nombre de tournées	65	49	113	14	69	19
Pourcentage de tournées	32,34%	24,38%	56,22%	6,97%	34,33%	9,45%

TABLE 5.12 – Nombre de tournées contenant au moins une fois les différents types de MOVE

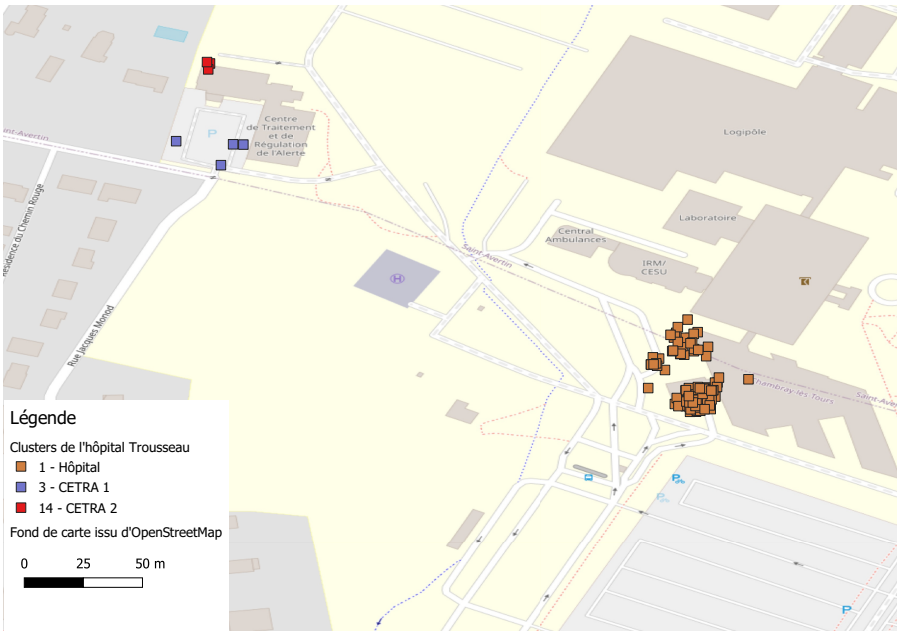


FIGURE 5.12 – Clusters regroupant des STOP près de l'hôpital Trousseau

— 17 occurrences du motif : zZ?,?zZ

— 25 séquences similaires aux deux premières avec un "Arrêt indéterminé" sur le chemin du retour

— 14 séquences avec de fortes présomptions d'un reroutage.

Nous pouvons donc constater que la montée en sémantique forte des trajectoires permet d'extraire des motifs plus riches que la montée en sémantique faible même associée aux POI, ne serait-ce que pour la prise en considération des interventions. Nous retrouvons dans cette étude les motifs typiques 1, 2, 3 et 4 en nombre relativement important. Nous trouvons aussi d'autres motifs de tournées où la dépose de patient se ferait en deux étapes, ainsi que, de nombreux cas de départ en intervention annulés (42) ou de déplacements indéterminés. Cela semble, en soi, indiquer des limites dans notre système de détection STOP/MOVE et de leur étiquetage que nous discuterons dans la prochaine section. Par ailleurs, nous utilisons cette première approche d'extraction de tournées et d'analyse fréquentielle des motifs dans ce chapitre et le suivant pour analyser simplement la qualité des tournées extraites. D'autres approches d'extraction de motifs fréquents plus actuelles (voir [Giacometti et al., 2014]) pourraient enrichir ces analyses avec possiblement une analyse de motifs à l'intérieur des tournées. Nous pourrions aussi les combiner avec une exploitation de la spatiotemporalité [Li, 2014] pour obtenir des tournées fréquentes spatialement contextualisées. Cela sera vraisemblablement très informatif dans un but d'entrepôt de données.

5.6 Discussion

Cette première approche, basée sur l'extraction de trajectoires sémantiques fortes à l'aide des données GPS et des informations de frein à main, a des limites en termes de qualité des données (5.6.1) et du fait du choix des capteurs utilisés et de l'exploitation de règles métiers explicites (5.6.2).

5.6.1 Qualité des trajectoires obtenues

Les MOVE classés comme "Trajet indéterminé" sont pour la plupart des déplacements courts (changement de place dans le sas des urgences, prise de renseignements pour trouver une adresse, déplacement entre des lieux non identifiés éloignés géographiquement,...).

Par ailleurs, certains MOVE, pourtant non identifiés comme "Trajet indéterminé" sont difficilement explicables. Par exemple, nous avons 42 cas de MOVE de type "Retour caserne" partant et revenant à la caserne sans STOP. Il s'agit possiblement d'un problème dans la détection des arrêts, ou d'une intervention qui a été subitement annulée. Un travail d'analyse plus fin des trajectoires de ce type pourrait être réalisé pour mieux en comprendre les tenants et aboutissants. Ainsi, il est important de pouvoir mettre en place une procédure de validation de nos résultats.

La validation de ces premiers résultats est un processus complexe.

En effet, il n'existe pas à l'heure actuelle de vérité terrain. Les pompiers disposent uniquement d'une liste des interventions. Ces données sont collectées à chaque appel déclenchant une intervention, puis enrichies sur la base des informations recueillies par l'opérateur du centre de secours (adresse géocodée, nature de l'intervention) et sur déclarations des pompiers (type d'activité, estampille temporelle). Par ailleurs, il peut y avoir des cas limites comme rouler avec le frein à main, cas qui faussent nos résultats.

Qui plus est, des différences conceptuelles existent. Par exemple, la fin d'une mission, dans notre méthode, est le retour du véhicule dans sa caserne d'affectation. Pour les pompiers, c'est lorsque le véhicule est à nouveau opérationnel (nettoyage, réarmement). De même, la sémantique des interventions des pompiers est plus concise. Six estampilles temporelles sont uniquement présentes dans celle liste (heure de départ caserne, heure d'arrivée sur le lieu d'intervention, heure de départ du lieu d'intervention et heure de retour caserne, heure d'arrivée et de départ de l'hôpital).

Ce processus déclaratif entièrement manuel souffre donc d'inexactitudes. Néanmoins, une comparaison de notre extraction et de cette liste a été effectuée. Pour des raisons éthiques évidentes, seul un pompier est habilité à manipuler la liste des interventions. C'est donc un responsable du SDIS 37 qui a analysé les interventions détectées par notre processus pour deux des quatre véhicules de la caserne des pom-

piers de Tours centre. Les résultats de ces analyses indiquent qu'une relative partie des interventions (plus de 40%) ont effectivement étés détectées par notre système. Une détection par le système est considérée comme correcte si les estampilles temporelles déclaratives et détectées (heure de départ caserne, heure d'arrivée sur le lieu d'intervention, heure de départ du lieu d'intervention et heure de retour caserne) se trouvent dans un intervalle de plus ou moins cinq minutes.

Pour les 60% restants, une partie s'explique par le fait que la caserne où les véhicules étaient stationnés dispose de 4 ambulances dont seulement 2 sont connectées. Bien que ces 2 ambulances étaient celles qui intervenaient le plus souvent (chaque ambulance à un numéro d'affectation qui détermine son ordre de départ, pour chaque demande d'intervention reçue, le véhicule 1 est demandé, si il n'est pas disponible, le véhicule 2 est demandé, etc. ; Les deux véhicules étudiés étaient respectivement numéro 1 et numéro 2), beaucoup d'interventions ont étés effectuées par des véhicules qui ne faisaient pas partis de l'étude. Malheureusement, il n'était pas possible pour les services du SDIS d'extraire une liste des interventions uniquement pour les véhicules étudiés.

Il faudrait ensuite pouvoir mesurer la qualité des informations déclarées par le système utilisé par les pompiers (inexactitude, incomplétude). En effet, plusieurs erreurs présentes sur une même déclaration entraînent un non appariement avec une intervention extraite automatiquement. Par exemple, lors d'interventions, les pompiers ont une charge cognitive importante, il arrive donc que la déclaration manuelle de la mission par les pompiers soit décalée dans le temps, ou même erronée en cas d'erreur de manipulation du boîtier déclaratif.

De même, les points évoqués dans la section suivante entraînent également des erreurs dans la détection d'activités.

5.6.2 Limites du processus

Le processus utilisé dans ce chapitre donne des résultats prometteurs. Cependant, des limites aux différentes étapes sont présentées dans les sections suivantes.

5.6.2.1 Détection des STOP

La méthode actuelle de détection des STOP utilise uniquement les données issues du capteur de frein à main pour déterminer si une sous trajectoire est de type STOP ou MOVE, le GPS étant utilisé uniquement pour modéliser l'emprise spatiale du STOP.

Cependant, l'utilisation du capteur GPS et du frein à main s'avère, dans certains cas, insuffisante pour extraire de manière précise les STOP, notamment quand le frein à main n'a pas été serré pendant un arrêt, ou, pas relâché entièrement pendant un déplacement. Il arrive parfois que les véhicules soient stationnés sans frein à

main, mais avec des cales, pour éviter l'usure prématurée des véhicules par exemple. Comme indiqué dans le tableau 5.3, il existe un MOVE qui fait 28 343 secondes (soit environ 7h52), ce qui semble problématique. Ainsi, une défaillance du capteur peut aussi totalement rendre inopérante la méthode de détection utilisée.

L'utilisation de plus de capteurs (e.g, capteur de contact) dans la fonction *arret* permettrait de réduire les problèmes de robustesse et de qualité des capteurs d'une part, et de potentiellement s'affranchir d'une durée minimale de l'autre.

5.6.2.2 Labellisation des activités

À l'aide du frein à main et du GPS nous pouvons identifier des activités ayant une sémantique assez simple. Cependant, certaines activités peuvent avoir été mal identifiées à l'instar de la prise en charge d'un patient, ou du trajet indéterminé.

En effet, selon notre approche, si un véhicule s'arrête sur un lieu logistique (e.g. station essence) alors l'arrêt sera automatiquement identifié comme logistique. Or, il se peut que l'arrêt soit en réalité effectué pour prendre en charge un patient suite à un malaise dans une station essence.

Dans ce cas, l'utilisation de capteurs complémentaires peut s'avérer fort utile. En utilisant le capteur de la jauge de carburant, le STOP pourrait suivre la règle métier suivante : SI un véhicule effectue un arrêt dans une station-service ET, qu'entre le début de l'arrêt et sa fin, le capteur de la jauge de carburant montre une augmentation du volume de carburant, alors ce STOP est une activité logistique.

De même, en utilisant un capteur sur les portes arrières ouvertes et un autre sur la présence du brancard à l'intérieur du véhicule ou non et un dernier sur l'activité du gyrophare, on pourrait identifier qu'il s'agit alors d'une prise en charge patient par la règle métier suivante : SI STOP n'est ni dans la caserne ni à l'hôpital, ET porte arrière ouverte ET absence du brancard dans le véhicule pendant une sous-période ET gyrophares en fonctionnement, alors l'activité sera prise en charge du patient sur le lieu d'intervention.

5.6.2.3 Segmentation

Dans la mise en œuvre proposée dans ce chapitre, il n'est pas possible d'avoir un STOP directement suivi par un autre STOP, ou un MOVE directement suivi par un autre MOVE. Cependant, comme indiqué dans la section 3.4.3, il peut être pertinent de différencier des changements d'activités dans un STOP en le divisant en autant de STOP que de STOPActivityType présentes dans le STOP de départ. Et il en va de même pour les MOVE. Pour cela, l'utilisation de capteurs supplémentaires pourrait permettre une segmentation plus fine.

En effet, pour les STOP, il serait possible, par exemple, à l'aide du capteur de rechargement électrique du véhicule (les véhicules sont équipés de batteries auxiliaires rechargeables pour la cellule sanitaire), de différencier deux activités pour un STOP lorsque que celui-ci est à la caserne : en charge, puis en attente lorsque la batterie est chargée et la prise débranchée. Deux STOP sont alors distingués, un STOP associé à l'activité en charge, puis un autre associé à attente caserne.

De plus, toujours en complétant les données disponibles avec d'autres capteurs, il serait possible de détecter des changements d'activité pendant un MOVE. Dans le cas d'un véhicule d'urgence, celui-ci peut être rerouté pour répondre à une intervention alors qu'il rentrait initialement à la caserne. Le point de reroutage pourrait être détecté par des changements importants et durables des valeurs des capteurs (augmentation de la vitesse et passage à une vitesse supérieure à la limite légale, allumage des gyrophares, etc.).

5.7 Conclusion

Dans ce chapitre, la mise en œuvre de l'extension au modèle de trajectoires sémantiques classiques pour l'extraction de trajectoires sémantiques fortes à l'aide des capteurs GPS et de frein à main a été présentée. Elle est effectuée sur les données des ambulances.

Lors de ce processus, les STOP sont associés à des POI ainsi qu'à des activités spécifiques aux véhicules étudiés. Nous pouvons aussi, à l'issue du processus déterminer des lieux d'arrêts fréquents. De même, les MOVE sont associés à des activités spécifiques. Les résultats de cette première expérimentation sont encourageants. Ils montrent en effet l'intérêt d'extraire automatiquement les interventions pour éviter les erreurs déclaratives.

Ce processus s'affranchit ainsi d'un grand nombre de problèmes liés aux données GPS classiques, notamment le bruit, l'imprécision ou les pertes de signal. Qui plus est, cette extraction de connaissances permet la définition de nouvelles variables pour l'analyse de données. Par exemple, il est ainsi envisageable de déterminer automatiquement le temps moyen d'immobilisation d'un véhicule pour chaque hôpital, dans un créneau horaire spécifique. De même, elles pourront améliorer l'estimation du temps d'arrivée d'un véhicule de secours sur le lieu d'un accident.

Plus globalement, elle facilite la visualisation et l'analyse des déplacements d'objets mobiles spécifiques à l'échelle d'une ville ou d'un département (point d'intérêts des véhicules, voies de circulation préférées suivant l'activité pratiquée par le véhicule, etc.).

Parallèlement, le caractère sensible de ces données liées à la santé nous oblige à intégrer la dimension éthique lors de l'extraction et l'exploitation de ces connaissances.

La solution proposée est aussi suffisamment générique pour être étendue à d'autres types de véhicules professionnels (taxi, ramassage d'ordures...) ou d'autres zones d'activités.

Par ailleurs, les limites de cette première expérimentation ont été discutées en terme de qualité des données initiales et des résultats, d'étiquetage potentiellement erroné, ainsi que de finesse de la segmentation.

Au final, cette mise en œuvre et ses limites illustrent l'intérêt de disposer de multiples capteurs afin de compléter ou de corriger les informations issues d'un capteur GPS et du frein à main. Cela est l'objet du prochain chapitre.

Chapitre 6

Extraction des trajectoires sémantiques fortes à l'aide de données issues des IOT

6.1 Introduction

Dans ce chapitre, l'étude portera sur une montée en sémantique de trajectoires, avec l'utilisation des données issues de capteurs IoT. IoT est l'acronyme de Internet Of Things, l'internet des objets en français. L'internet des objets est un système où des objets physiques sont connectés à internet. Ces objets (ou systèmes) sont capables de créer et transmettre des données afin de créer une valeur ajoutée à travers divers services (analyses de données, aide à l'utilisation, etc...). L'évolution de la technologie et la miniaturisation de celle-ci entraîne une réduction des coûts de ces différents capteurs. Selon un rapport de GSMA Intelligence [1], en 2016, il y a eu 6,3 milliards de connexions d'objets IoT à internet, les spécialistes en prévoient 25 milliards d'ici 2025. Ce nombre de connections seront pour plus de la moitié (environ 14 milliards) liées à des entreprises ou des secteurs spécifiques. Tandis que les connexions grand publics (maisons et objets personnels connectés principalement) représenteront un peu plus de 11 milliards de connexions. Pour les particuliers, l'IoT se concentre principalement dans l'utilisation d'objets de la vie quotidienne connectés (réfrigérateur, brosse à dents, porte de garage, chauffage, volets roulants, ...). Pour les entreprises, il est possible d'optimiser et contrôler les différents processus de fabrication et de livraison. L'IoT est également très utilisé dans le domaine des véhicules de transport (particuliers ou professionnels). Dans les voitures de particuliers, l'utilisation de l'IoT (de la mise en place de multiples capteurs, à l'analyse des données émises sur les véhicules du monde entier) permet la mise en place de

1. https ://data.gsmaintelligence.com/research/research/research-2019/iot-connections-forecast-the-rise-of-enterprise

systèmes d'assistance à la conduite, afin de surveiller par exemple la consommation de carburant, l'état des pièces d'usures, les distances de sécurité par rapport au véhicule précédant, ralentissements prévus sur l'itinéraire, etc... Pour les véhicules professionnels, l'ajout de capteurs IoT permet de produire des analyses ou études relatives aux spécifications du métiers. Prenons en exemple un camion chargé de la collecte de déchets recyclables dans une ville. L'exploitation de systèmes IoT permet de simplifier le processus de tournée de ce camion. Les points de collecte connectés envoient leurs taux de remplissage en direct. Après une analyse de la position et du taux de remplissage de chaque benne, le camion est en mesure de préparer une tournée adaptée à la collecte en fonction de l'espace restant disponible dans sa propre benne, et les différents volumes à collecter. Enfin, l'analyse a posteriori de ces données permet d'adapter le nombre de tournées nécessaires par semaine, ou même le volume des conteneurs placés dans la ville.

Dans le chapitre précédent, l'approche utilisée a permit une première extraction des STOP et des MOVE cohérents avec le métier de l'objet mobile étudié. L'extraction de ces éléments à rendu possible la montée en sémantique forte des trajectoires de ces ambulances décrivant leurs différentes activités (départ vers intervention, transport à l'hôpital,...).

Des motifs de trajectoires pertinents ont été extraits des résultats de cette montée en sémantique forte. Également, il a été possible de déduire des reroutages potentiels à l'aide des règles métiers et des comportements observés. Cependant, ils n'ont pu être délimités temporellement.

Du fait de l'utilisation d'un nombre réduit de capteurs, l'approche précédente comportait certaines limites. Tout d'abord, la succession de plusieurs STOP ou de plusieurs MOVE n'était pas possible. Les capteurs utilisés ne permettaient en effet pas de déterminer plusieurs activités dans un même segment de trajectoire.

La méthode de segmentation des trajectoires souffrait également du manque de capteurs. Tout d'abord le fait de ne disposer que du capteur de frein à main rendait la méthode dépendante de la qualité du capteur. Par exemple, si le capteur était défaillant, alors les STOP n'étaient simplement pas détectés.

Les résultats du chapitre précédent tendent à poser de nouvelles problématiques :

— Est-il possible de caractériser les possibles successions d'activités dans les STOP et les MOVE, tel que les reroutages ?

— L'ajout de capteurs embarqués dans nos véhicules peut-il permettre une montée en sémantique plus riche ?

— Quels capteurs embarqués peuvent permettre cet enrichissement, et dans quelles mesures ? Pour quels types d'activités ?

— Comment modéliser les règles métiers d'association des activités avec l'ajout de ces capteurs ?

Afin de répondre à ces problématiques, nous proposons dans ce chapitre d'identifier les capteurs pertinents dans la montée en sémantique de trajectoires. Les capteurs sont identifiés selon trois critères :

— Capteurs permettant une meilleure segmentation des trajectoires

— Capteurs permettant une sous-segmentation d'une même sous-trajectoire

— Capteurs permettant une meilleure association des activités

Une fois les capteurs identifiés, de nouvelles règles métiers doivent être déduites, et une nouvelle méthode d'association des activités sera définie à partir ces règles.

Dans ce chapitre, les choix de capteurs seront tout d'abord présentés afin de construire les trajectoires brutes étendues. Chaque capteur sera présenté afin de comprendre son impact sur la montée en sémantique forte. Ensuite, des capteurs seront choisis pour la segmentation, puis la méthode sera détaillée.

6.2 Construction des trajectoires brutes étendues

6.2.1 Capteurs candidats

Nous présentons dans cette section, les capteurs pouvant servir à la construction de nos trajectoires brutes étendues. Nos choix dans les capteurs récoltés parmi l'ensemble des capteurs possibles (cf. A.1) sont motivés par rapport à la segmentation des trajectoires, l'identification des STOP et des MOVE et celle des activités qui leur sont associées.

Plusieurs capteurs peuvent être utilisés pour la segmentation de trajectoire. Tout d'abord, comme pour le chapitre précédent, le capteur $freinMain$ ainsi que le *gps*. Cependant, d'autres capteurs sont également pertinents. Le capteur *vitesse*, qui indique la vitesse affichée au compteur permet de définir lorsque le véhicule est en mouvement ou non. De la même manière, le capteur $contactMoteur$ et le capteur $vitesseMoteur$ permettent de définir si le contact est allumé et si le moteur est tournant et donc, de s'assurer que le véhicule est ou non en mouvement.

Afin de détecter les différentes activités plusieurs capteurs peuvent être sélectionnés.

6.2.2 Choix des capteurs

État du frein à main Comme mis en évidence dans le chapitre précédent, le $freinMain$ continue d'être pertinent pour la segmentation des trajectoires et la détection des STOP. C'est un capteur booléen.

GPS Tout comme le $freinMain$, le *gps* est un capteur pertinent non pas pour la segmentation mais pour la représentation spatiale des STOP. Également, il est utilisé pour calculer la distance entre deux STOP, afin de savoir si un micro-MOVE doit être ou non fusionné.

Vitesse affichée au compteur Le capteur $vitesseCompteur$ est un capteur numérique qui renvoie la vitesse mesurée indiquée sur le tableau de bord du véhicule. Ce capteur a été choisi car il s'avère pertinent dans plusieurs situations. Tout d'abord, il permet de renforcer la segmentation des trajectoires : si la vitesse est très faible, il y a de forte chance que le véhicule soit arrêté, si elle est plus élevée, nous devrions normalement être sur un déplacement. Ensuite, son étude permet de déterminer le comportement d'une ambulance : on sait par exemple que la vitesse moyenne d'un véhicule d'urgence augmente lors d'un déplacement vers une intervention. Il pourrait aussi être également utile pour détecter un reroutage, par exemple lors d'un retour à la caserne, la vitesse moyenne est relativement faible, puis la vitesse, entre autre, augmente lors de l'appel vers un nouveau lieu d'intervention.

Contact moteur Le capteur binaire $contactMoteur$ permet de connaître l'état du contact du moteur (1 si le contact est allumé, mais pas nécessairement démarré, 0 sinon). Ce capteur semble pertinent dans la segmentation des trajectoires brutes étendues et dans la détection de STOP. En effet, si le contact est coupé, on peut dire que le véhicule est à l'arrêt. Par contre, le véhicule peut aussi être à l'arrêt avec le contact allumé, car, par exemple, les véhicules d'urgences restent démarrés sur les lieux d'intervention. Il peut aussi permettre la sous-segmentation de STOP, notamment dans des activités logistiques, à l'instar des tests de maintenance effectués, au moins une fois par jour, sur les véhicules à la caserne.

État du branchement à l'alimentation secteur Le capteur binaire $enCharge$ indique si le véhicule est branché à une prise de courant pour rechargement de la cellule sanitaire. Il doit nous permettre de sous-segmenter les STOP ayant lieu dans une caserne, seul endroit ou se trouvent les prises de rechargement compatibles, en, par exemple "Attente classique" ou "Rechargement du véhicule".

État de la sirène Le capteur binaire *sirene* indique si la sirène est allumée ou non. La sirène est utilisée ponctuellement lors d'un départ en intervention ou lors du transport à l'hôpital. Elle est destinée à avertir les usagers de la route qu'un véhicule d'intervention est proche. Elle est principalement utilisée prés des carrefours de circulation, ou lorsque le véhicule d'intervention doit franchir des bouchons. Dans notre cas, l'utilisation de ce capteur qui indique si la sirène est activée ou non est très pertinent. En effet, il peut être utile pour détecter lorsque le véhicule se dirige vers un patient ou vers un hôpital avec un patient à bord du véhicule.

État des gyrophares Le capteur binaire *gyrophare* indique si les gyrophares sont allumés ou non. Les gyrophares sont des signaux lumineux présents sur le véhicule. Pour une question de responsabilité, un véhicule d'urgence n'a la priorité que si ses gyrophares sont en fonctionnement (bien que cela n'évite pas les oublis). Dans notre cas, ce capteur est donc très pertinent pour identifier les départs en intervention ainsi que les transports à l'hôpital.

État de la porte arrière Le capteur binaire *porteArriere* indique si la porte arrière est ouverte ou non. Sur un véhicule de secours tel que les ambulances de notre étude, il est impossible de manipuler le brancard sans ouvrir la porte arrière du véhicule. En effet, le chargement ou déchargement du brancard se fait uniquement par la porte arrière. C'est pourquoi ce capteur est pertinent afin de pouvoir identifier les prises en charge de patient sur lieu d'intervention ou à l'hôpital.

6.2.3 Trajectoires brutes étendues

À l'aide des capteurs choisis, conformément à la définition 2 présentée dans la section 3.2.2, nos trajectoires brutes étendues sont des séquences d'éléments $e_i^+ \in T^+$ ayant une position GPS, une étiquette temporelle, ainsi qu'un ensemble de capteurs $D = (B, C)$.

Au vu des capteurs choisis, nous avons :

— les capteurs binaires :

$$B = \{freinMain, sirene, contactMoteur, gyrophare, porteArriere\}$$

— le capteur numérique :

$$C = \{vitesseCompteur\}$$

Ainsi, un élément de trajectoire e_i^+ est un objet Position défini comme :

$$e_i^+ . \begin{cases} t & \rightarrow \text{ timestamp} \\ p & \rightarrow \text{ localisation à l'instant } t \\ freinMain & \rightarrow \text{ état du frein à main à l'instant } t \\ sirene & \rightarrow \text{ état de la sirène à l'instant } t \\ contactMoteur & \rightarrow \text{ état du moteur à l'instant } t \\ gyrophare & \rightarrow \text{ état des gyrophares à l'instant } t \\ porteArriere & \rightarrow \text{ état de la porte arrière à l'instant } t \\ vitesseCompteur & \rightarrow \text{ état de la vitesse du compteur à l'instant } t \end{cases} \tag{6.1}$$

6.2.4 Analyse des éléments de trajectoires obtenus

Les données traitées ici étant les mêmes que dans le chapitre précédent, nous mettons ici le focus sur la répartition des informations des capteurs nouvellement ajoutés.

Gyrophares La répartition des informations sur les gyrophares est la suivante. Pour le total des données récoltées sur les deux véhicules, 352 731 éléments de trajectoires ont pour valeur $gyrophare == 1$ et 2 246 841 où $gyrophare == 0$. Nous avons donc environ 14% des données récoltées ou le gyrophare est allumé. Ce chiffre parait logique, en effet, une ambulance passe une grosse partie de son temps à la caserne, et seules les activités de départ vers intervention et de transport patient nécessitent l'utilisation des gyrophares.

Sirène Le capteur de l'état de la sirène est un capteur particulier. En effet, l'activation de la sirène dans le véhicule entraîne l'activation des gyrophares. Dans notre jeu de données, 18 269 éléments de trajectoires ont pour valeur $sirene == 1$, ces données sont à mettre en perspective avec le nombre d'éléments de trajectoires ou les gyrophares sont allumés (352 731). Nous constatons donc une activation de la sirène de l'ordre de 5%. La encore, ce chiffre se justifie par le fait que l'utilisation de la sirène est très ponctuelle. En effet, elle n'est utilisée que lorsque les véhicules de secours prennent la priorité (Feux rouges, carrefour), où pour annoncer leur présence (bouchons). Ainsi, ce capteur, bien qu'intéressant d'un point de vue analyse statistique (étude de zones dangereuses, circulation difficile, ...), ne sera pas directement utilisé pour la montée en sémantique.

Porte arrière La répartition des informations sur le capteur de porte arrière est la suivante. Pour le total de données récoltées, 51 463 éléments de trajectoires ont pour valeur $porteArriere == 1$ et 2 548 109 où $porteArriere == 0$, soit un peu moins de 2% des données. Ce pourcentage est faible, en effet, l'ouverture des portes arrières sur les véhicules de secours ne se produisent en général que lors de la montée / descente d'un patient dans le véhicule. Cependant, ne disposant pas des spécifications, et malgré l'étude de la répartition des séquences d'ouvertures de portes, nous ne pouvons dire avec certitude que ce capteur correspond uniquement à la porte arrière ou également à la porte arrière latérale de l'ambulance. Ce capteur ne sera donc pas utilisé pour la montée en sémantique.

La section suivante présente la montée en sémantique faible faite à partir des trajectoires brutes étendues obtenues.

6.3 Montée en sémantique faible des trajectoires des trajectoires brutes étendues

Cette section présente l'étape de montée en sémantique faible à l'aide de multi-capteurs.

À ce stade, nous disposons de trajectoires brutes étendues à l'aide des capteurs choisis dans la section précédente. Conformément au processus de montée en sémantique faible présenté dans le chapitre 3, nous devons commencer par segmenter nos trajectoires brutes étendues. Dans cette section, nous présenterons la nouvelle méthode de segmentation des trajectoires puis nous qualifierons les différentes sous-trajectoires en STOP ou MOVE. Enfin, la dernière étape est l'assignation des STOP aux POI. Pour chaque étape, nous présenterons la méthode utilisée puis l'expérimentation sur les données collectées.

6.3.1 Segmentation en sous-trajectoires : choix de la méthode

6.3.1.1 Méthode de segmentation

Comme vu dans la section précédente, plusieurs capteurs ont étés identifiés comme pouvant servir à la segmentation en sous-trajectoires des trajectoires brutes étendues.

Comme dans le chapitre 5, le capteur $freinMain$ est à nouveau utilisé. Cependant, son utilisation seule n'est pas satisfaisante. Il arrive parfois que le frein à main soit mal relâché lors d'un trajet, rendant parfois la segmentation incorrecte. Un autre capteur identifié est le capteur $vitesseCompteur$. Une vitesse positive indique un potentiel mouvement du véhicule et donc une vitesse nulle un possible arrêt, cependant, la sensibilité de ce capteur ne permet pas son usage seul.

Enfin, le capteur $contactMoteur$, peut également servir à la segmentation. En effet, un véhicule ayant le contact coupé n'est pas censé être en mouvement. Cependant, un véhicule ayant le contact allumé n'est pas nécessairement en mouvement.

Ces trois capteurs sont donc pertinents et complémentaires pour la segmentation en sous-trajectoires.

Ces valeurs de capteurs doivent être combinées pour déterminer à quel instant la trajectoire doit être segmentée.

Dans l'approche basée sur le GPS et le frein à main du précédent chapitre, la détection du changement pour la construction des segments, présentée dans la section 6.3.1, repose sur le changement d'état du frein à main entre t et $t+1$.

Dans l'approche présentée ici, nous souhaitons faire reposer la détection d'un possible changement pour la segmentation sur une caractérisation prenant en considération les différents capteurs choisis.

L'algorithme 8 présente la segmentation définie pour l'approche courante.

La modification, vis-à-vis de l'algorithme 5 utilisé dans le chapitre précédent, porte sur l'introduction de la fonction booléenne $etatElementArret$ utilisée afin d'agréger les données de ces trois capteurs. Cette fonction indique si l'état traité correspond à un possible arrêt. Si les valeurs de retour de cette fonction pour les états courant et suivant sont différentes, alors il y a changement.

Algorithm 8 Segmentation d'une trajectoire brute étendue

Require: T trajectoire $T = \langle e_1, ..., e_n \rangle$
  $i \leftarrow 1$ //iterateur
  $L \leftarrow [\,]$ //liste des sous-trajectoires à retourner
  $st \leftarrow [\,]$ //sous-trajectoire en cours
  while $i < n$) **do**
    st.Push(e_i^+)
    if $etatElementArret(e_i^+) \neq etatElementArret(e_{i+1}^+)$ **then**
      L.Push(st)
      $st \leftarrow [\,]$
    end if
    $i \leftarrow i + 1$
  end while
  st.Push(e_n)
  L.Push(st)
  return L

Cette fonction booléenne pour un état donné est définie à l'aide d'un principe simple : si au moins deux capteurs sur trois indiquent un arrêt potentiel pour cet état alors la valeur de vérité de la relation pour cet état vaut vrai, sinon faux. L'algorithme 9 présente ce principe.

Algorithm 9 etatElementArret : fonction d'identification d'un élément de trajectoire en possible arrêt

Require: eC // élément courant

  //État courant de la vitesse compteur
  $eCVC \leftarrow (eCourant.vitesseCompteur < Vmin)$
  //État courant du frein à main
  $eCFM \leftarrow eCourant.freinMain$
  //État courant du moteur
  $eCMD \leftarrow eCourant.contactMoteur$

  //État courant est il a l'arrêt
  $eC = (eCVC \wedge eCFM) \vee (eCFM \wedge NOT(eCMD)) \vee (eCVC \wedge NOT(eCMD))$

  return eC

L'algorithme ainsi défini permet de segmenter des trajectoires brutes étendues en sous-trajectoires.

	Arrêts potentiels	Déplacements potentiels
Nombre	797	796
Durée minimum	00 :00 :01	00 :00 :01
Durée moyenne	00 :55 :48	00 :06 :10
Durée médiane	00 :14 :29	00 :05 :34
Durée maximum	29 :52 :29	00 :31 :00

TABLE 6.1 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires extraites

6.3.1.2 Analyse des sous-trajectoires identifiées

Comme vu dans la section précédente, la première extraction des sous-trajectoires avec la fonction *etatElementArret* permet d'identifier des périodes d'arrêts potentiels dans les trajectoires en prenant en comptes plus de capteurs que dans le chapitre précédent.

Au total après cette extraction, nous avons obtenu 1 593 sous-trajectoires pour les deux véhicules : 479 pour le véhicule 1 et 1114 pour le véhicule 2. Nous obtenons plus précisément 797 potentiels arrêts, et 796 potentiels déplacements. Cela représente, pour le véhicule 1, 240 arrêts potentiels et 239 déplacements potentiels. Tandis que pour le second véhicule, 557 pour les deux types de sous-trajectoires. Nous remarquons que nous avons extrait moins de sous-trajectoires. Pour rappel, la méthode avec le frein à main nous avait permis d'extraire 1609 sous-trajectoires. Ce résultat s'explique par le faire que la méthode de segmentation utilise plus de capteurs. En effet, la méthode précédente pouvait par exemple détecter deux arrêts potentiels pour une même période si le frein à main était relâché pendant quelques instants alors même que le véhicule était pourtant à l'arrêt.

Le tableau 6.1 présente la répartition de la durée de ces sous-trajectoires selon qu'ils soient des arrêts potentiels ou déplacements potentiels.

Les résultats obtenus sont à mettre en comparaison avec les résultats obtenus dans le chapitre précédent. Tout d'abord, nous pouvons voir que la durée maximale des déplacements potentiels est tout à fait cohérente (environ 30 minutes) alors qu'elle était précédemment de plusieurs heures. Également, la durée moyenne de déplacement à fortement baissé, passant de presque 10 minutes à 6 minutes et 10 secondes. Ces constatations confirment l'intérêt de l'utilisation de multiples capteurs pour segmenter nos trajectoires sémantiques.

Cette nouvelle méthode de segmentation ayant permit de supprimer les cas de mauvaise détection d'arrêts potentiels constatés lors de la segmentation des trajectoires dans le chapitre précédent, mais pas d'en changer le contenu (les données ou le véhicule est considéré en arrêt ou en déplacement ne changent pas, seul les périodes de détection changent), le tableau de répartition des outliers ne sera pas présenté une nouvelle fois.

	Gyrophares allumés	Gyrophares éteins	Total
Nombre de données lors d'un arrêt potentiel	217 128	2 094 481	2 311 609
Pourcentage de données relatif aux arrêts potentiels	9,4	90,6	100
Nombre de données lors d'un déplacement potentiel	135 603	152 360	287 963
Pourcentage de données relatif aux déplacements potentiels	47,1	52,9	100
Nombre de données total	352 731	2 246 841	2 599 572
Pourcentage de données relatif au nombre total de données	13,6	86,4	100

TABLE 6.2 – Nombre de données où les gyrophares sont activés suivant les arrêts ou déplacements potentiels

Par ailleurs, pour ce qui concerne le capteur des gyrophares et sa qualité, le tableau 6.2 représente la répartition de l'allumage des gyrophares, selon que le véhicule est en arrêt ou en déplacement potentiel. On peut voir que les gyrophares sont allumés moins de 10% du temps lors d'un arrêt potentiel. Ces chiffres semblent logiques. En effet, le véhicule ne conserve les gyrophares lors d'un arrêt uniquement sur le lieu d'intervention, et parfois lors de la prise en charge du patient à l'hôpital (si le conducteur oublie de désactiver les gyrophares en entrant aux urgences, comme il est coutume de le faire pour ne pas éblouir le personnel). Ainsi, ce capteur fournit des informations qui semblent de qualité suffisante pour être exploitées par la suite.

6.3.2 Extraction des STOP et MOVE

6.3.2.1 Choix de la fonction arrêt

La fonction *arret* a pour objectif de déterminer si une sous trajectoire est de type STOP ou MOVE.

La segmentation étant basée sur le changement d'états définis à l'aide de trois capteurs. Si les valeurs de la fonction *etatElementArret* pour les éléments d'une sous-trajectoire indiquent un possible arrêt, alors cette sous-trajectoire est de type STOP, sinon, de type MOVE.

Soit $capteursArret = \{freinMain, contactMoteur, vitesseCompteur\}$ l'ensemble des trois capteurs, alors :

$$arret_{capteursArret}(T_i^+) = true \iff \forall e_i^+ \in T_i^+, etatElementArret(e_i^+) = true \quad (6.2)$$

À l'instar de l'approche présentée dans le chapitre précédent, il est nécessaire de filtrer les micros-arrêts. Nous utilisons pour cela les mêmes principes que dans 5.3.2.1, et plus particulièrement l'utilisation de la D_{min}^{arret}. La fonction *arret* est donc définie comme :

$$arret(T_i^+) = arret_{capteursArret}(T_i^+) \wedge (D_{T_i^+} > D_{min}^{arret}) \quad (6.3)$$

Finalement, les sous trajectoires durant plus de D_{min}^{arret} secondes et où le frein à main est tiré sont considérées comme des STOP, sinon, elles sont concaténées aux sous-trajectoires précédentes et suivantes. La sous-trajectoire résultante de cette concaténation sera logiquement un MOVE.

$$!(arret(T_i^+)) \wedge arret_{capteursArret}(T_i^+) \rightarrow concat(T_{i-1}^+, T_i^+, T_{i+1}^+) \quad (6.4)$$

Les STOP durent donc nécessairement plus de D_{min}^{arret}. Pour le traitement des micro-déplacements, nous utiliserons la même approche que celle présentée dans le chapitre précédent. L'équation suivante (6.5) présente ce traitement.

$$\begin{aligned} &!(arret_{capteursArret}(T_i^+)) \wedge (T_i^+.D < D_{min}^{dep}) \\ &\wedge(dist(T_{i-1}^+.medGeom, T_{i+1}^+.medGeom) < dist_{min}^{dep}) \\ &\qquad \rightarrow concat(T_{i-1}^+, T_i^+, T_{i+1}^+) \end{aligned} \quad (6.5)$$

Par ailleurs, l'association entre STOP et POI suit la méthode présentée dans la section 5.3.3 du chapitre précédent.

À ce stade, nous disposons donc de sous-trajectoires brutes étendues, de type STOP ou MOVE. Une sous trajectoire brute étendue est donc associée à :

— Une étiquette sémantique : $sm_i \in \{\text{"MOVE"}\} \cup \{(\text{"STOP"}, POI^*)\}$.

— Une liste d'éléments de trajectoires (positions) tels que présentés dans l'équation 6.1.

Nous obtenons finalement des trajectoires sémantiques faibles qui forment des séquences successives de STOP et de MOVE sans pour l'instant avoir la possibilité d'avoir deux STOP ou deux MOVE consécutifs.

6.3.2.2 STOP et MOVE détectés

Suite à une étude rapide des premiers résultats, nous avons choisi de conserver les seuils définis dans le chapitre précédent (cf. 5.3.2.2). C'est à dire :

	STOP	MOVE
Nombre	640	639
Durée minimum	00 :02 :00	00 :00 :04
Durée moyenne	01 :06 :45	00 :07 :41
Durée médiane	00 :18 :11	00 :07 :18
Durée maximum	38 :41 :44	00 :44 :30

TABLE 6.3 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des sous-trajectoires STOP ou MOVE extraites

— $D_{min}^{arret} = 120$ secondes
— $dist_{min}^{dep} = 30$ mètres
— $D_{min}^{dep} = 15$ secondes

Nous avons précédemment identifié, via notre segmentation, 1 593 sous-trajectoires pour les deux véhicules. Une fois la fonction arrêt avec le traitement des micro-arrêts et des micro-déplacements, nous obtenons 1 279 nouvelles sous-trajectoires issues des concaténations des potentiels arrêts et déplacements, pour ne plus avoir les micro-arrêts et micro-déplacements non pertinents.

Parmi ces 1 279 sous-trajectoires, 640 d'entre-elles sont considérées comme des STOP. Concernant les positions étendues représentatives de ces STOP (voir la section 3.3.2 sur les trajectoires sémantiques faibles) :

— 6 STOP n'ont pu être représentés géographiquement du fait d'une perte totale du signal GPS.
— 634 STOP ont pu être représentés, en deux catégories, en fonction de leur qualité spatiale, évaluée à l'aide de R_{med} et de $\%_{DP}$:
  — 580 STOP où R_{med} est compris entre 0 et 5 mètres, ce qui traduit une très bonne précision de leur position spatiale.
  — 54 STOP où R_{med} est supérieur à 5 mètres, et ou le taux de données perdues ($\%_{DP}$) est assez élevé, ce qui traduit une mauvaise réception du signal lors du STOP.

En comparaison avec les résultats du chapitre précédent, nous observons une nette diminution du nombre de STOP non positionnables (6 contre 14), une augmentation du nombre de STOP positionnables (634 vs. 612) et de ce fait, du nombre de STOP positionnables de bonne qualité (580 vs. 560) ainsi qu'une légère hausse de ceux de moins bonne qualité (54 vs. 52). La détection des STOP est donc de meilleure qualité que précédemment car l'utilisation de capteurs additionnels permet sans doute d'allonger la durée des STOP et de réduire la durée des MOVE (notamment les très longs), comme l'indique le tableau 6.3. Ce processus nous a aussi permis d'extraire 639 MOVE au total.

Nous n'observons plus de MOVE dont la durée semble aberrante (prés de 8 heures dans l'approche précédente). En effet, la durée maximale d'un MOVE est d'environ 45 minutes, ce qui semble cohérent dans une situation de déplacement non urgent (retour caserne aux heures de pointe, par exemple). La durée maximale d'un STOP a également augmenté mais correspond sans doute à un véhicule mis en réserve. Cette approche nous parait donc tout à fait pertinente vis à vis de nos connaissances métier.

6.3.3 Appariement des STOP et des lieux d'intérêts

Comme vu dans la section 5.5.1, le clustering des STOP permet l'identification de plusieurs zones pour un même POI. C'est le cas pour l'hôpital Trousseau, ou nous pouvons distinguer deux zones d'arrêts différentes (une dans le sas des urgences, et l'autre lorsque le véhicule est stationné après la dépose du patient). Également, nous distinguons deux zones d'arrêts au CETRA, l'une pour la logistique du véhicule (sas de réarmement du CETRA), et l'autre lorsque les pompiers ont des besoins administratifs (parking des bureaux du CETRA). Également, nous avons vu dans le chapitre précédent que le POI associé au SDIS 37 était représenté par un point, ce qui entraînait parfois un non appariement avec le POI lorsque le véhicule effectuait pourtant bien un STOP dans ce centre d'intérêt.

C'est pourquoi dans ce chapitre, nous avons modifiés certains POI. Pour le SDIS 37, nous avons représenté manuellement deux polygones, ayant l'emprise des deux bâtiments du SDIS 37. Pour le POI correspondant à l'hôpital trousseau, celui-ci à été supprimé, puis remplacé par 4 zones distinctes. La première zone correspondant au sas des urgences, la deuxième au parking ou les véhicules d'urgences stationnent en attendant la prise en charge du patient. Les deux dernières zones correspondent pour l'une au sas de ravitaillement du CETRA, et l'autre, au parking administratif du CETRA. Ces zones ont étés déterminées visuellement, après études des clusters obtenus dans la section 5.5.1.

Sur notre jeu de données :

- 390 STOP ont été liés à des POI :
  - 224 STOP liés à des casernes
  - 136 STOP liés à des hôpitaux
  - 30 STOP liés à des lieux logistiques
- 250 n'ont pas été liés à un POI dont 6 car ils n'ont pas de représentation spatiale.

Nous remarquons que nous obtenons l'association d'un plus grand nombre de STOP avec des POI et un moins grand nombre de STOP non associés par rapport au chapitre précédent. La nette augmentation se situe principalement sur les STOP liés à la caserne (sûrement issus de la réduction du nombre de STOP non positionnables)

ainsi que sur les STOP logistiques. Pour ces derniers, l'explication tiens du fait de l'ajout du CETRA comme un POI de type logistique ainsi que de la redéfinition du POI du SDIS 37. La redéfinition du type de POI du CETRA en logistique est sans doute la cause de la réduction du nombre de STOP liés à des POI de type hôpital.

La prochaine section présente le processus de montée en sémantique forte prenant en considération des informations issues d'autres capteurs afin d'enrichir le processus montée en sémantique vu dans le chapitre précédent.

6.4 Montée en sémantique forte

La dernière étape du processus de montée en sémantique est l'affectation d'une activité à une sous-trajectoire.

Comme vu dans le chapitre précédent, nous avons besoin d'une ontologie métier ou à minima d'une liste d'activités métiers. Ensuite, un nouveau processus d'appariement sera définit afin de d'associer une activité aux sous trajectoires.

6.4.1 Construction de la liste d'activités

Ayant plus de capteurs disponibles, la liste des activités issue de la section 5.4.1 peut être affinée. Les activités principales issues d'experts métiers vont être conservées, cependant, l'utilisation de plus de capteurs rend possible l'identification des cas de reroutage, ce qui correspond à un re-découpage possible de MOVE en plusieurs MOVE. À partir de cette nouvelle liste, nous associations à chaque sous-trajectoire de sémantique faible obtenue un des labels de ladite liste.

6.4.1.1 Liste des STOPActivityType

Nous considérons de nouveau les 5 types d'activités de STOP présentés dans le chapitre précédent pour les STOP :

— Prise en charge du patient sur lieu d'intervention

— Prise en charge du patient à l'hôpital

— Logistique

— Attente caserne

— Arrêt indéterminé

Cependant, cette liste d'activités peut être affinée, toujours en concertation avec les experts, de par l'utilisation de capteurs complémentaires. Ainsi, nous cherchons, dans ce chapitre, à étiqueter les STOP à l'aide des activités supplémentaires précisant certaines des précédentes.

Ainsi, l'activité "Prise en charge du patient à l'hôpital" devient :

-> Prise en charge du patient à l'hôpital

-> Attente fin de prise en charge hôpital

-> Attente hôpital

Lors de la prise en charge d'un patient à l'hôpital, le patient est immédiatement déposé aux urgences (STOPActivityType :"Prise en charge du patient à l'hôpital"). Une fois la victime dans le sas des urgences, les pompiers doivent attendre l'inscription de la victime et sa consultation par une infirmière du service. Dans ce cas, le véhicule peut rester dans le sas, ou se déplacer sur le parking à proximité pour laisser le sas des urgences libre (STOPActivityType :"Attente fin de prise en charge hôpital"). Enfin, il arrive également que les pompiers doivent passer à l'hôpital sans nécessairement déposer un patient, dans le cas d'oubli de matériel notamment. Ces activités non liées à la gestion de patients entrent dans la catégorie STOPActivityType :"Attente hôpital".

Les autres activités ne sont pas modifiées par rapport au chapitre précédent. Nous avons donc finalement la liste des activités STOPActivityType suivante :

— Prise en charge du patient sur lieu d'intervention : 

— Prise en charge du patient à l'hôpital : 

— Attente fin de prise en charge hôpital : 

— Attente hôpital : 

— Logistique : 

— Attente caserne : 

— Arrêt indéterminé : 

— Arrêt non positionnable : 

6.4.1.2 Liste des MOVEActivityType

De la même manière que la section précédente, la liste des activités de la section 5.4.1.2 est à nouveau considérée pour les MOVE :

— Départ vers intervention

— Transport vers hôpital

— Trajet logistique

— Retour caserne

— Trajet indéterminé vers Hôpital

— Trajet indéterminé vers arrêt indéterminé

— Trajet indéterminé vers arrêt non positionnable

Cependant, l'identification des cas de reroutage ainsi que les nouvelles activités de STOP à l'hôpital entraînent naturellement de nouvelles activités de MOVE.

— Retour avant redirection

— Départ vers intervention (reroutage)

— Départ vers intervention avant annulation

— Retour caserne après annulation

— Trajet logistique hôpital

— Trajet intra hôpital

— Trajet intra caserne

Les reroutages peuvent être de deux types. Le premier type est l'annulation, lorsque le véhicule est appelé vers un lieu d'intervention, et qu'il est finalement rappelé car la mission est annulé. Dans ce cas, nous obtenons deux activités de MOVE différentes. La première étant le trajet jusqu'à ce que le véhicule soit annulé (MOVEActivityType :"Départ vers intervention avant annulation"), la seconde étant le trajet de retour jusqu'à la caserne après l'annulation (MOVEActivityType :"Retour caserne après annulation"). Dans le cas d'une annulation, les deux activités se suivent. Le deuxième type de reroutage et le renvoi vers intervention, lorsque le véhicule est appelé vers une nouvelle intervention, alors même qu'il n'était pas en attente à la caserne. Les deux types d'activité issues de ce reroutage sont "Retour avant redirection" lorsque le véhicule est sur le retour avant d'être envoyé sur une nouvelle intervention et "Départ vers intervention (reroutage)" lorsque le véhicule se dirige vers une nouvelle intervention. Cependant, il est possible que les deux activités ne se suivent pas. Par exemple, un véhicule peut être rerouté alors qu'il effectue un STOP sur un lieu d'intervention, dans ce cas, l'enchaînement des activités sera (Prise en charge d'un patient sur le lieu d'intervention -> Départ vers intervention (reroutage) -> Prise en charge d'un patient sur le lieu d'intervention). Le premier des deux nouveaux types d'activité de MOVE liés à l'hôpital est "Trajet intra hôpital" lorsque le véhicule effectue un déplacement à l'intérieur de l'hôpital, par exemple lorsque le patient est déposé au sas des urgences, puis que le véhicule libère le sas. Le second type est "Trajet logistique hôpital" lorsque le véhicule effectue un trajet vers l'hôpital dans un but différent de la dépose d'un patient.

Finalement, la liste MOVEActivityType des activités considérées est la suivante :

— Départ vers intervention : 

— Départ vers intervention (reroutage) : 

— Départ vers intervention avant annulation : 

— Transport vers hôpital : 

— Trajet intra hôpital : 

— Trajet logistique hôpital : 

— Trajet logistique : 

— Retour caserne : 

— Trajet intra caserne : 

— Retour avant redirection : 

— Retour caserne après annulation : 

— Trajet indéterminé vers Hôpital : 

— Trajet indéterminé vers arrêt indéterminé : 

— Trajet indéterminé vers arrêt non positionnable : 

Une fois les listes d'activités déterminées, il faut ensuite associer ces activités aux MOVE et STOP précédemment extraits. La section suivante présente la méthode d'association entre les activités et les STOP et MOVE.

6.4.2 Association des sous-trajectoires sémantiques faibles et activités

Chaque sous-trajectoire sémantique faible (STOP et MOVE) doit être associée à une seule activité dans le cadre de la montée en sémantique forte.

Pour associer une activité à un STOP, nous utilisons dans ce chapitre la même méthode que dans la section 5.4.2.

Les règles sont stockées sous forme de conditions, pour déterminer si une règle peut être appliquée à une sous trajectoire, pour chaque règle, chaque condition est vérifiée. Si une condition n'est pas vérifiée, le processus ne retient pas l'activité liée

à cette règle et passe à la règle suivante. Si toutes les conditions de la règle sont vérifiées, alors l'activité est associée à la sous trajectoire.De plus, l'ordre d'exécution des algorithmes 6 et 7 est important, car le second requiert l'exécution au préalable du premier.

Nous présentons dans un premier temps l'association des STOP aux STOPActivityType puis celle des MOVE aux MOVEActivityType.

6.4.2.1 Association des STOP

Comme indiqué précédemment, dans ce chapitre, 8 activités sont identifiées pour les STOP. Des règles métiers sont à nouveau utilisées. Après concertation avec des experts métiers, les règles issues du chapitre précédent ont été complétées avec l'ajout des nouveaux capteurs disponibles et sont les suivantes :

— Si STOP n'a pas de médiane géométrique, alors l'activité est "Arrêt non positionnable"

— Si STOP et POI de type hôpital et que le POI du STOP précédent est également de type hôpital alors l'activité est "Attente fin de prise en charge hôpital"

— Si STOP et POI de type hôpital et gyrophares activés plus de $gyro_{min}^{hopital}$ % du temps durant le MOVE précédent alors l'activité est "Prise en charge du patient à l'hôpital"

— Si STOP et POI de type hôpital alors l'activité est "Attente hôpital"

— Si STOP et POI de type logistique alors l'activité est "Logistique"

— Si STOP et POI de type caserne alors l'activité est "Attente caserne"

— Si STOP et gyrophares activés plus de $gyro_{min}^{inter}$ % du temps durant le STOP et pas de POI associé alors l'activité est "Prise en charge du patient sur lieu d'intervention"

— Si aucune activité n'a été définie, alors l'activité est "Arrêt indéterminé"

L'algorithme 10 présente la méthode d'association des activités aux STOP.

A ce stade, les reroutages n'ont pas encore été détectés, les trajectoires sont donc nécessairement des successions de STOP et de MOVE. Une fois les STOP étiquetés, l'étape restante de la montée en sémantique est labellisation des MOVE.

6.4.2.2 Identification et traitement des reroutages

Avant de déterminer les activités associées aux MOVE, nous détectons s'il y a des reroutages pendant le MOVE.

Comme vu dans la section 4.2, il y a deux types de reroutage. Le premier est l'annulation, et le second la redirection. Pour rappel, une annulation se produit lorsqu'un véhicule initialement appelé vers un lieu d'intervention est finalement rappelé

Algorithm 10 Assignation d'une activité à un STOP

Require: s //STOP étudié
Require: s_p //STOP précédent s
Require: m //MOVE précédent s
Require: $gyro^s$ //pourcentage d'activation des gyrophares pendant s
Require: $gyro^m$ //pourcentage d'activation des gyrophares pendant m
Require: $gyro^m$ //pourcentage d'activation des gyrophares pendant m
Require: $gyro_{min}^{hopital}$ //seuil minimum de pourcentage d'activation des gyrophares pour considerer un transport vers l'hôpital
Require: $gyro_{min}^{inter}$ //seuil minimum de pourcentage d'activation des gyrophares pour considerer un départ vers intervention
switch ($s.poi.poitype.label$)
case "hopital"**:**
if $s_p.poi.poitype.label == "hospital"$ **then**
$s.STOPActivityType.label \leftarrow$ "Attente fin de prise en charge hôpital"
else if $gyro^m >= gyro_{min}^{hopital}$ **then**
$s.STOPActivityType.label \leftarrow$ "Prise en charge du patient à l'hôpital"
else
$s.STOPActivityType.label \leftarrow$ "Attente hôpital"
end if
case "logistique"**:**
$s.STOPActivityType.label \leftarrow$ "Logistique"
case "caserne"**:**
$s.STOPActivityType.label \leftarrow$ "Attente caserne"
case "Non positionnable"**:**
$s.STOPActivityType.label \leftarrow$ "Arrêt non positionnable"
default:
//cas où $s.poi.poitype.label ==$"Indéterminé"
if $gyro^s >= gyro_{min}^{inter}$ **then**
$s.STOPActivityType.label \leftarrow$ "Prise en charge du patient sur lieu d'intervention"
else
$s.STOPActivityType.label \leftarrow$ "Arrêt indéterminé"
end if
end switch

à la caserne. C'est pourquoi, il est possible de détecter ce type d'annulation si un MOVE commence et termine par une activité d'attente caserne. Afin d'éviter les faux positifs, deux contraintes sont ajoutées. La première étant que le gyrophare doit être activé un certain temps $duree^{gyro}_{debut}$ maximum après le début du MOVE et la seconde que les gyrophares aient étés allumés au minimum un certain temps $duree^{gyro}_{min}$ pendant ce MOVE. Ces contraintes permettent d'éviter les cas ou l'activation des gyrophares n'était pas liée à une intervention. L'endroit définit comme point de segmentation (la position ou le MOVE sera segmenté en deux MOVE distincts) est la position ou le gyrophare à été allumé pour la dernière fois pendant ce MOVE. Afin de détecter une annulation, une règle simple est utilisée :

— Si activité du STOP précédent est "Attente caserne" et activité du STOP suivant est "Attente caserne" et le moment de première activation du gyrophare pendant le MOVE est inférieur à un seuil $duree^{gyro}_{debut}$ et durée d'activation du gyrophare pendant le MOVE est supérieure à un seuil $duree^{gyro}_{min}$ alors c'est un reroutage de type "Annulation"

Une redirection se produit lorsque le véhicule, disponible est appelé pour se diriger vers une nouvelle intervention mais que celui-ci n'est pas en attente à la caserne. Afin de détecter les redirections, une règle simple est également utilisée. Si un MOVE commence par une activité différente d'attente caserne et qu'il termine par une activité de type "Prise en charge patient sur lieu d'intervention" alors il y a une redirection. Également, une contrainte est ajoutée pour déterminer si le MOVE doit être segmenté ou pas (en effet, il peut y avoir une redirection directe, par exemple, lorsque le véhicule est rappelé sur une nouvelle intervention alors qu'il effectue toujours un STOP). Cette contrainte est définie par l'instant ou les gyrophares sont réactivés après le début du MOVE. Dans ce cas, il y a une segmentation, sinon, il n'y en a pas. L'endroit définit comme point de segmentation est la position ou le gyrophare à été activé pour la première fois pendant ce MOVE.

— Si activité du STOP précédent est différente de "Attente caserne" et activité du STOP suivant est "Prise en charge patient sur lieu d'intervention" alors c'est un reroutage de type "Redirection"

L'algorithme 11 présente la méthode de détection de reroutage dans les MOVE et leur segmentation.

Ainsi, à l'issue de cette détection, nous avons un ensemble de MOVE ayant une étiquette s'ils sont issus d'un reroutage.

6.4.2.3 Association des MOVE

Les activités associées aux MOVE sont également définies avec des experts métiers à l'aide de règles. Les MOVE ayant été segmentés du fait d'un reroutage sont ici pris en considération pour déterminer leurs activités.

Algorithm 11 Détection de reroutage dans un MOVE et segmentation

Require: s_{i-1}^{+} // STOP précédent le MOVE étudié
Require: s_i^{-} // MOVE étudié avec $s_i^{-} =< e_y^{+}, ..., e_z^{+} >$
Require: s_{i+1}^{+} // STOP suivant le MOVE étudié
Require: $duree^{gyro}$ // Durée d'activation du gyrophare
Require: $duree^{avantGyro}$// Durée avant première activation du gyrophare
  $s_i^{-}.reroutage \leftarrow$ ""
  $M \leftarrow< s_i^{-} >$ //liste des MOVE construits
  //Cas d'une annulation
  if $(s_{i-1}^{+}.STOPActivityType.label ==$ "Attente caserne")
  $\wedge$ $(s_{i+1}^{+}.STOPActivityType.label ==$ "Attente caserne")
  $\wedge$ $(duree^{gyro} > duree_{min}^{gyro}) \wedge (duree^{avantGyro} < duree_{debut}^{gyro})$ **then**
    //Détection du point de reroutage
    $k = z$
    while $k > y \wedge e_k^{+}.gyrophare == 0$ **do**
      $k = k - 1$
    end while
    //segmentation
    $s_i' =< e_y^{+}, ..., e_k^{+} >$ //MOVE avant reroutage annulation
    $s_i'.reroutage \leftarrow$ "Début annulation"
    $s_i'' =< e_{k+1}^{+}, ..., e_z^{+} >$ //MOVE après reroutage annulation
    $s_i''.reroutage \leftarrow$ "Fin annulation"
    $M \leftarrow< s_i', s_i'' >$

    //Cas d'une redirection vers une intervention
  else if $(s_{i-1}^{+}.STOPActivityType.label! =$ "Attente caserne")
  $\wedge(s_{i+1}^{+}.STOPActivityType.label ==$ "Prise en charge patient sur lieu d'intervention")) **then**
    //Détection du point de reroutage
    $k = y$
    while $k < z \wedge e_k^{+}.gyrophare == 0$ **do**
      $k = k + 1$
    end while
    //segmentation
    $s_i' =< e_y^{+}, ..., e_k^{+} >$ //MOVE avant reroutage redirection
    $s_i'.reroutage \leftarrow$ "Début redirection"
    $s_i'' =< e_{k+1}^{+}, ..., e_z^{+} >$ //MOVE après reroutage redirection
    $s_i''.reroutage \leftarrow$ "Fin redirection"
    $M \leftarrow< s_i', s_i'' >$
  end if
  return M

Comme vu précédemment, nous avons une liste de 14 activités potentielles. Les règles suivantes ont été définies :

- Si un reroutage à été détecté et est de type "Début redirection", alors l'activité est "Retour avant redirection"
- Si un reroutage à été détecté et est de type "Fin redirection", alors l'activité est "Départ vers intervention (reroutage)"
- Si un reroutage à été détecté et est de type "Début annulation", alors l'activité est "Départ vers intervention avant annulation"
- Si un reroutage à été détecté et est de type "Fin annulation", alors l'activité est "Retour caserne après annulation"
- Si l'activité suivante est vide ou si elle est de type "Arrêt non positionnable", alors l'activité est "Trajet indéterminé - vers arrêt non positionnable"
- Si l'activité suivante est de type "Attente hôpital", alors l'activité est "Trajet logistique hôpital"
- Si l'activité suivante est de type "Attente fin de prise en charge hôpital", alors l'activité est "Trajet intra hôpital"
- Si l'activité suivante est de type "Prise en charge du patient à l'hôpital" et si l'activité précédente est vide ou différente de "Prise en charge patient sur lieu d'intervention", alors l'activité est "Trajet indéterminé - vers Hôpital"
- Si l'activité suivante est de type "Prise en charge du patient à l'hôpital" et si l'activité précédente est "Prise en charge patient sur lieu d'intervention", alors l'activité est "Transport vers hôpital"
- Si l'activité suivante est de type "Logistique", alors l'activité est "Trajet logistique"
- Si l'activité suivante est de type "Prise en charge patient sur lieu d'intervention" et si l'activité précédente est de type "Attente caserne", alors l'activité est "Départ vers intervention"
- Si l'activité suivante est de type "Prise en charge patient sur lieu d'intervention" et si l'activité précédente est différente de "Attente caserne", alors l'activité est "Départ vers intervention (reroutage)"
- Si l'activité suivante est de type "Attente caserne" et si l'activité précédente est différente de "Attente caserne", alors l'activité est "Retour caserne"
- Si l'activité suivante est de type "Attente caserne" et si l'activité précédente est vide ou de type "Attente caserne", alors l'activité est "Trajet intra caserne"
- Si aucune règle n'est valide, alors l'activité est "Trajet indéterminé - vers arrêt indéterminé"

L'algorithme 12 présente la méthode d'association des MOVEActivityType aux MOVE.

Nous avons donc maintenant l'ensemble de la solution algorithmique pour monter en sémantique forte.

6.4.3 Affectation des activités

6.4.3.1 Affectation des activités aux STOP

Comme vu précédemment, les règles pour attribuer une activité à un MOVE utilisent les activités des STOP précédent et suivant.

Les règles métier présentées dans la section 6.4.1.2, nécessitent de déterminer des seuils. Pour rappel, le premier seuil est $gyro_{min}^{inter}$ est le seuil minimum d'activation des gyrophares que l'on considère pour déterminer si un STOP est de type intervention. Pour cela, nous avons étudié la répartition du pourcentage de la durée d'allumage des gyrophares durant les épisodes de STOP. La figure 6.1 présente la répartition du pourcentage de la durée d'allumages des gyrophares pendant des STOP.

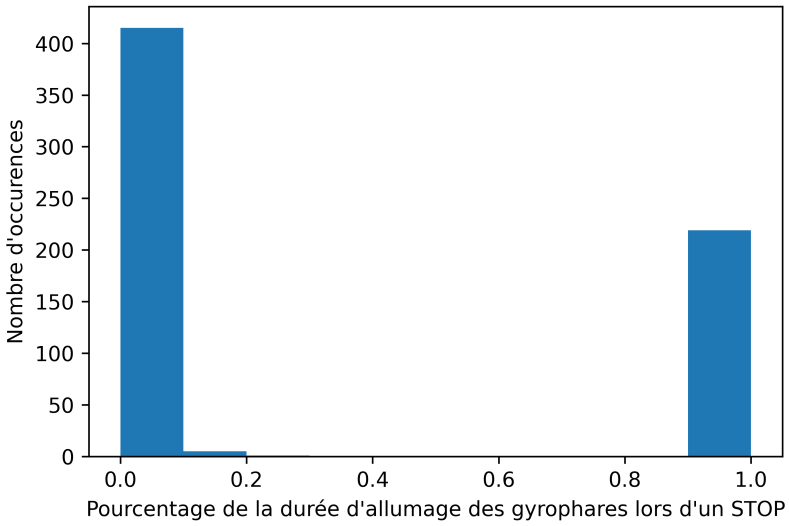


FIGURE 6.1 – Répartition du pourcentage de la durée d'allumages des gyrophares pendant des STOP

Nous pouvons voir que lors des STOP, l'activation des gyrophares est relativement stable. En effet, les gyrophares sont pour la plus grande majorité soit allumé plus de 90% du temps, soit éteins plus de 20% du temps. C'est pourquoi le seuil $gyro_{min}^{inter}$ est fixé à 70%, afin de prendre en compte de potentielles variations des données.

Le deuxième seuil est $gyro_{min}^{hopital}$. Pour rappel, c'est le seuil qui permet de déterminer si un patient est transporté à l'hôpital ou s'il s'agit d'un déplacement logistique vers celui-ci. Pour cela nous avons étudié la répartition du pourcentage de la durée d'allumage des gyrophares durant les épisodes de MOVE. La figure 6.2 présente la répartition du pourcentage de la durée d'allumages des gyrophares pendant des MOVE.

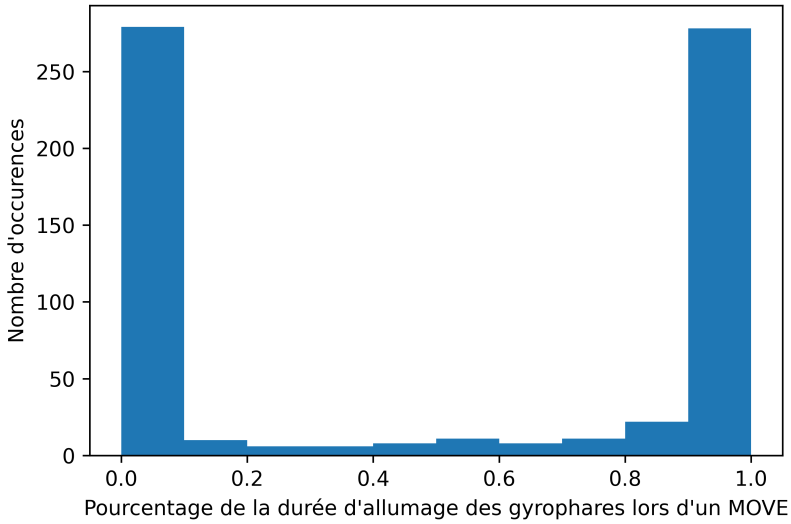


FIGURE 6.2 – Répartition du pourcentage de la durée d'allumages des gyrophares pendant des MOVE

Nous pouvons voir que la répartition de ces pourcentages est moins tranchée que lors des STOP. En effet, il existe quelques dizaines de MOVE ou le ratio d'activation des gyrophares est compris entre 20% et 70%. Cependant, les pompiers ont pour habitude de couper les gyrophares lors de l'arrivée aux urgences, afin de ne pas éblouir les personnes présentes. De plus, l'activation des gyrophares pendant le MOVE permet de penser qu'il s'agit effectivement d'un transport patient. C'est pourquoi le seuil $gyro_{min}^{hopital}$ est fixé à 30%.

Avec les seuils précédemment déterminés, les résultats suivants ont étés obtenus :

— 204 STOP de type "Prise en charge du patient sur lieu d'intervention"

— 111 STOP de type "Prise en charge du patient à l'hôpital"

— 30 STOP de type "Logistique"

— 224 STOP de type "Attente caserne"

— 0 STOP de type "Attente hôpital"

— 25 STOP de type "Attente fin de prise en charge hôpital"

— 40 STOP n'ont pu être liés à des activités et sont donc de type "Arrêt indéterminé"

— 6 STOP n'ont pu être liés à des activités car ils n'étaient pas positionnables.

Le tableau 6.4 présente la répartition des durées des STOP, en secondes, en fonction de l'activité associée. Dans ce tableau, nous pouvons constater plusieurs changements par rapport à la méthode employée dans le chapitre précédent.

Tout d'abord, le changement de la méthode de détection des STOP à permit une meilleure identifications de ceux-ci. En effet, plus de la moitié (14 vs 6) des STOP non positionnables ont finalement pu être positionnés et identifiés. La durée maximum de ces STOP non positionnables à fortement diminué (46 817 vs 7 956). Également, grâce à la prise en compte de plus de capteurs et donc de règles plus précises, plus de 60% des STOP initialement indéterminés ont pu être identifiés (95 vs 40).

Les durées moyenne et médiane des STOP de type "Prise en charge d'un patient sur lieu d'intervention" ont diminué. En effet, les contraintes étant maintenant sur l'utilisation du gyrophare pendant le STOP, la durée minimale de 600 secondes du chapitre précédent n'existe plus et sa disparition fait donc baisser ces chiffres. Le nombre de ces activités à également fortement augmenté (154 vs 204).

Du fait de la modification du POI du CETRA, qui est maintenant de type logistique, la durée maximum de l'activité concernée ("Logistique") a fortement augmenté (3 250 vs 9 596) cependant, cette valeur (environ 2h30) reste cohérente avec une activité de désinfection et nettoyage total d'un véhicule et rechargement de l'équipement de secours.

Nous constatons qu'aucun STOP de type "Attente hôpital" n'a été détecté dans ce jeu de données. Le fait qu'aucun STOP de ce type n'ait été détecté montre les problèmes de la méthode de détection du chapitre précédent, qui nous laissait penser qu'il y avait des déplacements à l'hôpital sans transport de patients.

Nous identifions deux raisons au fait de ne pas détecter de MOVE de ce type. La première étant, comme vu précédemment, la meilleure détection des prises en charge de patients sur lieu d'intervention. Les MOVE sont alors correctement identifiés comme des transport de patient à l'hôpital. La seconde vient du fait que certains déplacement étaient peut-être du à des déplacements au CETRA et non à l'hôpital. Finalement, l'affinement de la détection joint à l'affinement des POI change la lecture que l'on a pu faire dans un premier temps du jeu de données et les nouveaux résultats obtenus montre que ce nouvel approche apporte un réel plus dans la montée en sémantique de trajectoires.

Les STOP étant maintenant liés à des activités, il est possible de passer à l'étape suivante du processus et détecter les reroutages dans les MOVE.

								
Nombre	224	111	25	0	204	30	40	6
Durée minimum	00 :02 :17	00 :03 :36	00 :02 :10	-	00 :02 :25	00 :02 :14	00 :02 :00	00 :13 :26
Durée moyenne	02 :35 :35	00 :18 :49	00 :20 :40	-	00 :17 :09	00 :21 :05	00 :18 :32	01 :04 :11
Durée médiane	01 :19 :20	00 :15 :54	00 :16 :12	-	00 :13 :19	00 :04 :00	00 :09 :58	00 :51 :02
Durée maximum	38 :41 :44	01 :11 :07	01 :15 :43	-	02 :47 :38	02 :39 :56	01 :17 :14	02 :12 :36

TABLE 6.4 – Nombre et durées minimum, moyenne, médiane et maximum, en (heures :minutes :secondes), des STOP selon l'activité

6.4.3.2 Traitement des reroutages durant un MOVE

Afin de détecter les reroutages, les règles présentées dans la section 6.4.2.2 sont utilisées. Pour rappel, deux seuil doivent à nouveau être définis. Le premier seuil $duree^{gyro}_{debut}$ est le temps maximum avant le déclenchement du gyrophare lors d'un MOVE et le second $duree^{gyro}_{min}$ est le temps minimum de déclenchement du gyrophare. Ces seuils permettent la détection de reroutage de type annulation.

La figure 6.3 présente la répartition de la durée avant le premier allumage des gyrophares, si celui-ci à été activé durant le MOVE. Nous constatons que pour plus de 90% des cas, les gyrophares, si ils sont allumés pendant le MOVE le sont dans les 40 premières secondes de celui-ci. Quelques rares cas existent ou l'activation du gyrophare à lieu plus de 60 secondes aprés le début du MOVE. Le seuil $duree^{gyro}_{debut}$ à donc été fixé à 60 secondes.

Le seuil $duree^{gyro}_{min}$ à été fixé en relation avec les contraintes métiers. En effet, ce seuil est présent pour déterminer que l'appuie sur les gyrophares était bien volontaire pour un départ vers intervention. C'est pourquoi, celui-ci à été fixé à 10 secondes. En effet, cela permet de retirer les cas d'appuis involontaires qui ne durent que quelques secondes. La définition d'un seuil plus élevé ne permettrait de toute façon pas de supprimer les oublis de longue durée, et remettrais en cause la détection des cas de reroutage ou le véhicule à été annulé rapidement après son départ.

La mise en oeuvre du processus de détection de reroutage avec les seuils précédemment définis à permit d'obtenir les résultats suivants :

— 8 Reroutages et segmentations de type "Annulation"

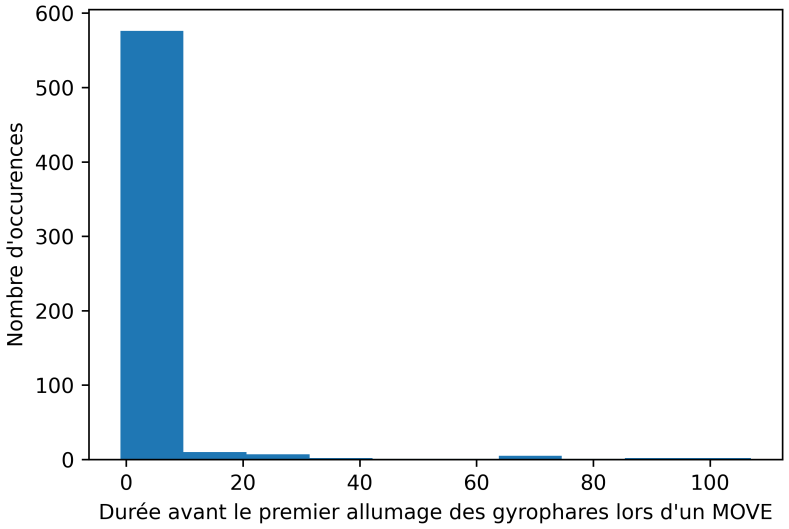


FIGURE 6.3 – Répartition de la durée avant le premier allumage des gyrophares lors d'un MOVE

— 19 Reroutages et segmentations de type "Redirection"

Ces chiffres semblent pertinents sur la période étudiée. En effet, l'annulation reste un cas relativement rare. Concernant les redirections, celles-ci peuvent être plus nombreuses, mais il s'agit ici uniquement des redirections ou le véhicule avait déjà commencé un trajet retour vers la caserne (et donc ou le MOVE à été segmenté). L'étape suivante présente l'assignation d'activités à la liste des MOVE enrichie des trajectoires segmentées.

6.4.3.3 Affectation des activités aux MOVE

Comme vu précédemment, les règles pour attribuer une activité à un MOVE utilisent les activités du STOP précédent et suivant.

Les résultats obtenus, à l'aide des différentes règles métier définies dans la section 6.4.1.2, sont présentés dans le tableau 6.5, pour les cas de reroutage, et le tableau 6.6 pour les autres.

Comme on peut le voir, il y a plus de départ vers intervention (reroutage) que de retour avant redirection, qui est pourtant l'activité liée (un véhicule effectue un retour avant redirection, puis est redirigé et effectue un Départ vers intervention (reroutage)). Cependant, un véhicule peut également être redirigé alors qu'il effectue une activité de STOP. On peut donc déterminer que 32 reroutages de type redirection ont eu lieu sans segmentation du MOVE.

				
Nombre	19	51	8	8
Durée minimum	00 :03	00 :19	00 :25	00 :01
Durée moyenne	07 :13	06 :36	03 :30	02 :50
Durée médiane	06 :17	06 :04	00 :56	01 :40
Durée maximum	19 :13	19 :57	10 :10	09 :55

TABLE 6.5 – Nombre et durées minimum, moyenne, médiane et maximum, en (minutes :secondes), des STOP de type reroutage selon l'activité

										
Nombre	153	182	33	30	102	25	0	9	40	6
Durée minimum	00 :31	01 :32	00 :06	00 :15	03 :09	00 :10	-	05 :32	00 :15	00 :04
Durée moyenne	04 :56	09 :42	01 :55	409	06 :49	01 :29	-	13 :48	09 :21	02 :56
Durée médiane	04 :32	09 :00	01 :27	04 :55	09 :41	00 :26	-	10 :06	09 :05	01 :37
Durée maximum	14 :21	34 :45	16 :56	22 :44	18 :53	15 :59	-	44 :30	24 :52	11 :04

TABLE 6.6 – Nombre et durées minimum, moyenne, médiane et maximum, en (minutes :secondes), des MOVE (Hors reroutage) selon l'activité

Nous pouvons noter également une très forte diminution du nombre de trajets indéterminés de tout types. La nouvelle méthode de détection des STOP ayant entraîné une baisse significative du nombre de STOP indéterminés, il en découle donc moins de trajets du type correspondant.

Les reroutages n'ayant pas été traités lors du précédent chapitre, il n'est pas possible de comparer les résultats obtenus. Cependant, on peut voir, dans le tableau 6.5, que les durées maximum de ces MOVE liés à des reroutages sont très cohérents avec le métier. En effet, les MOVE liés à des annulations ont des durées très faibles. Les annulations ont en effet lieu rapidement après que le véhicule ait été appelé (dans le cas ou un autre véhicule initialement indisponible le devient, et est alors envoyé sur l'intervention), ce qui se traduit par un Départ vers intervention avant annulation et l'activité de retour liée Retour caserne après annulation relativement court.

On note cependant que la durée médiane est plus élevée dans le cas du retour, ce qui peut s'expliquer par le fait que le véhicule n'est plus en situation d'urgence lorsqu'il est annulé. Concernant les redirections, les durées des activités Départ vers intervention (reroutage) sont également peu élevées (quelques minutes pour les durées moyenne et médiane). Cela peut s'expliquer par le fait qu'un véhicule est choisit pour être redirigé généralement parce que sa position géographique est proche du lieu de la nouvelle intervention.

La durée maximum est quand à elle d'une durée plus élevée (prés de 20 minutes) et s'explique par le fait qu'un véhicule peut parfois être appelé sur une nouvelle intervention dès la sortie de l'hôpital, dans une zone très éloignée de l'hôpital mais pourtant couverte par sa caserne d'affectation. En effet, le CETRA, qui attribue les moyens sur les différentes interventions n'a connaissance que de certaines données que transmettent les pompiers (Départ du lieu d'intervention, départ de l'hôpital,...). La durée minimum d'une seconde pour la catégorie des retours caserne après annulation peut s'expliquer par l'oubli d'enlever les gyrophares lorsque l'ordre d'annulation à été reçu.

Nous constatons, dans le tableau 6.6, que les types de MOVE nouvellement ajoutés – hors cas de reroutage – ont des durées cohérentes. En effet, les cas des trajets intra caserne ont des durées moyennes et médiane relativement courte (moins de deux minutes) qui correspondent à un déplacement du véhicule de la cour avant à la cour arrière de la caserne.

Cependant, la durée maximum est surprenante (plus de 15 minutes), et peut s'expliquer par une activité logistique (test du véhicule sur la route par exemple). Également, les durées des déplacements intra hôpital sont très faibles, et correspondent à des déplacements entre le sas et le parking des urgences. Cependant, la durée maximum de prés de 10 minutes semble indiquer départ puis un retour à l'hô-

pital sans que l'on puisse en déterminer, de manière certaine, le motif. Différentes hypothèses peuvent être posées comme par exemple le fait qu'il soit possible que les pompiers retournent à l'hôpital dans le cas ou ils s'aperçoivent de l'oubli de matériel.

Les 3 types de déplacements indéterminés sont bien moins nombreux qu'avec la méthode de détection précédente (176 vs 55). La précédente valeur maximale de plus de 7h n'existe plus, l'affinage de la méthode de détection des STOP à permis d'identifier précisément l'activité du véhicule (un STOP sans le frein à main).

Globalement, l'apport des nouveaux capteurs permet une détection plus fine des STOP et une meilleure description des activités. Les activités détectées sont donc plus nombreuses et plus courtes.

La section suivante présentera une analyse des résultats obtenus avec la prise en compte de multiples capteurs.

6.5 Analyse des trajectoires sémantiques obtenues

Nous pouvons donc désormais mener les mêmes analyses que celles présentées dans la section 5.5 du chapitre précédent.

6.5.1 Identification des arrêts fréquents

En utilisant les mêmes paramètres que ceux de la section 5.5.1, nous nous intéressons ici à l'étude des clusters obtenus. Pour rappel, la méthode de clustering utilisée est DB-SCAN [Ester et al., 1996], et utilise les paramètres suivant :

— ϵ : 30 mètres

— $MinPts$: 3

Le tableau 6.7 représente la répartition des STOP associés à des cluster en fonction de l'activité.

Nous retrouvons encore 14 cluster, tout comme pour le chapitre précédent. Ce chiffre s'explique par le fait que la nouvelle méthode de détection de STOP utilise plus de capteurs, mais leur modélisation spatiale est toujours effectuée à l'aide de données GPS. Ainsi, du fait également de l'utilisation de la médiane géométrique, un STOP même si son étendue dans le temps diffère en fonction de la méthode de segmentation utilisée, sa représentation spatiale reste sensiblement la même. Nous obtenons donc les mêmes zones géographiques, avec un nombre de STOP différent.

239 STOP (moins de 40% des STOP positionnables) ne sont pas associés à un cluster. Cependant, il est logique que les lieux d'interventions ne soient pas toujours associés à un cluster (il faut pour cela que plusieurs interventions aient eu lieu dans des endroits proches).

Type de STOP							
Cluster 0	222						
Cluster 1			87			23	
Cluster 2				13			
Cluster 3				4			
Cluster 4					3		
Cluster 5			4			1	
Cluster 6				4			
Cluster 7			5				
Cluster 8		4					
Cluster 9		2			2		
Cluster 10		3					
Cluster 11			8				
Cluster 12					3		
Cluster 13		3			1		
Cluster 14				3			
En dehors d'un cluster	2	192	7	6	31	1	

TABLE 6.7 – Répartition des STOP en fonction de leur cluster et de l'activité associée

Nous retrouvons dans les mêmes clusters l'attente de la fin de prise en charge à l'hôpital et la prise en charge patient à l'hôpital. Cela semble logique du fait que la zone d'attente est proche géographiquement du sas des urgences où les patients sont pris en charge. Nous remarquons également que les corrections apportées sur le POI du CETRA font que les arrêts logistiques sont regroupés entre eux. Ils apparaissent uniquement dans des clusters entièrement logistiques. Les clusters détectés sont plutôt purs et les activités des STOP détectées sont cohérentes pour un même cluster. Il est cependant étonnant qu'il y ait 8 STOP liés aux hôpitaux non associés aux clusters hôpitaux, et 6 STOP liés à la logistique où il en va de même.

Après étude, il s'avère que les STOP non inclus dans un cluster mais ayant une activité liée à un hôpital sont en fait des interventions se trouvant proches de l'hôpital Bretonneau. La typologie de cet hôpital, situé en plein centre-ville de Tours rend difficile la distinction par l'algorithme entre l'hôpital lui même et une rue adjacente. Cependant, une redéfinition manuelle plus fine de la zone de prise en charge aux urgences permettrait de résoudre cette ambiguïté. Les activités logistiques détectées en dehors d'un cluster sont du fait de STOP ayant eu lieu proche d'un POI de type logistique différent du lieu habituel, les STOP n'ont donc pas étés classés dans un cluster. Dans le cluster 1, à l'hôpital Trousseau, nous constatons que du fait du paramétrage de DBSCAN, les zones d'attente et de prise en charge à l'hôpital sont regroupées dans le même cluster. Un paramétrage différent permettrait de différencier ces deux zones. Cependant, dans le cas du pôle santé Leonard de Vinci (cluster

5), même un paramétrage plus adapté ne permettrait pas de différencier les deux zones de prise en charge et d'attente à l'hôpital. En effet, physiquement ces deux zones sont extrêmement proches, et il est difficile de les séparer sans l'ajout d'une zone définie manuellement.

Nous observons également que 192 lieux d'interventions ne sont pas inclus dans des clusters, ce qui montre bien que ces lieux d'interventions sont dispersés dans la ville.

6.5.2 Analyse des tournées

À nouveau dans ce chapitre, nous procédons à l'extraction des parcours partant de la caserne et revenant à celle-ci.

Au total dans cette étude, nous avons extraits 222 tournées, soit 21 tournées de plus que dans le chapitre précédent. 66 de ces tournées sont associées au véhicule 1, et 156 au véhicule 2. L'étude de fréquence de ces tournées permet l'extraction de motifs typiques de séquences de sous-trajectoires des ambulances.

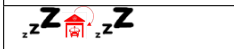
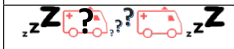
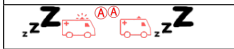
Motif de tournée	Nombre	Durée minimum	Durée médiane	Durée moyenne	Durée maximale
	46	12 :09	23 :06	27 :36	73 :04
	43	30 :26	61 :56	67 :44	166 :48
	13	53 :31	72 :28	76 :48	121 :35
	33	00 :06	01 :27	01 :55	16 :56
	11	18 :26	62 :05	57 :27	98 :02
	8	00 :46	04 :13	06 :20	20 :05
	7	12 :49	40 :59	33 :36	47 :33
	5	37 :14	56 :43	67 :00	130 :39
	5	45 :16	62 :55	74 :51	129 :57
	5	58 :50	63 :46	71 :13	87 :24

TABLE 6.8 – Motifs des tournées les plus fréquents et durées calculées, en (minutes :secondes)

Le tableau 6.8 présente les tournées identifiées les plus redondantes ainsi qu'une analyse de leurs durées. L'étude de l'analyse des durées révèle que la nouvelle méthode de segmentation est pertinente. En effet, il n'existe pas de durées extrêmement longues, comme vues lors du chapitre précédent. Les durées très courtes sont toujours présentes, en effet, une annulation lors d'un départ vers intervention peu surgir des le départ du véhicule d'urgence. Également, un trajet intra caserne peu être extrêmement court (changement de place sur le parking). La durée maximale relevée de

10 008 secondes (prés de 3 heures) est possible dans le cadre d'une intervention suivie d'une prise en charge à l'hôpital. Nous constatons que les durées sont cohérentes avec le nombre d'activité de la tournée. Concernant la fréquence de ces tournées, les plus fréquentes sont les tournées identifiées dans la section 4.2. En effet, en prenant en compte l'ajout de l'activité d'attente de fin de prise en charge à l'hôpital après la prise en charge d'un patient, 51% des tournées sont définies comme étant typiques. 6% des tournées sont liées à des activités de reroutages. 11 tournées n'ont pas d'activités identifiées, pour la plupart, ces activités ont lieu dans des POI qui ne sont pas en rapport direct avec l'activité des véhicules d'urgence. Enfin, une partie non négligeable des tournées détectées sont finalement des trajets intra caserne (dans 15% des cas).

6.5.3 Vers un entrepôt de données multi-dimensionnel

Les analyses précédentes permettent de constater l'efficacité et la cohérence des algorithmes utilisés pour la segmentation des trajectoires et leurs montées en sémantique. Dans cette section, nous construisons un entrepôt de données regroupant les données extraites à différents niveaux de granularité afin de permettre une analyse plus fine. Le but d'un entrepôt de données est de permettre le stockage et l'exploitation de gros volumes de données [Inmon, 1995]. Dans notre cas, cet entrepôt permet à des utilisateurs non experts en analyse de données de disposer d'indicateurs via des tableaux de bord répondants à leurs questions. Nous présenterons le schéma conceptuel en étoile de l'entrepôt de données, ensuite, quelques questions typiques émanant des experts métiers qui exploitent les ambulances.

La figure 6.4 présente le schéma en étoile développé pour notre cas d'étude des STOP et des MOVE. A terme, suivant les analyses voulues, il peut être pertinent de disposer de plusieurs schémas. Dans le notre, les épisodes de STOP et de MOVE sont considérés comme des faits (les objets sur quoi vont porter l'analyse). Les tournées pourraient aussi être considérées comme des faits afin d'étudier plus finement les séquences complètes de celles-ci. Cependant, leurs nombres étant encore plus faibles, les résultats obtenus ne seraient pas pertinents.

Ce schéma simple permet de créer un premier entrepôt de données permettant de répondre à des questions orientées métiers par exemple :

— "Quelle est la durée moyenne d'attente à l'hôpital en fonction des jours de la semaine ?"

— "Quelle est la durée moyenne de transit en intervention (entre le départ de la caserne et l'arrivée sur les lieux) ?"

— "Combien d'interventions donnent lieu à un transport à l'hôpital en fonction du jour et de la nuit ?"

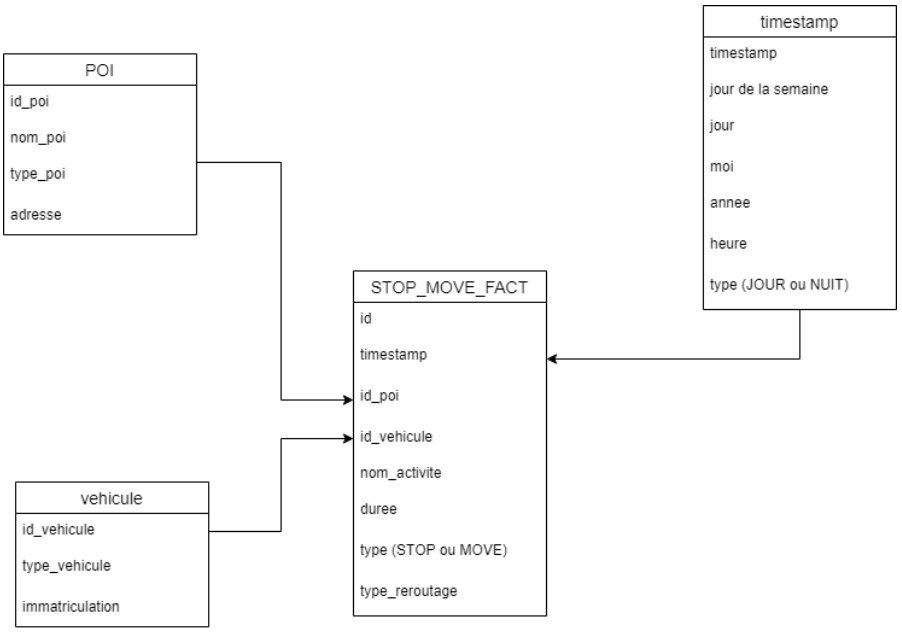


FIGURE 6.4 – Schéma en étoile de la création de l'entrepôt de données

La structure de notre entrepôt nous permet de répondre à ces questions. Cependant, nous ne disposons actuellement que de 222 tournées de deux véhicules sur les 5 de la caserne sur une durée de 15 jours. Ce volume de données n'est pas suffisant pour présenter des résultats représentatifs. Par exemple, pour la durée d'attente à l'hôpital, nous observons une durée de X minutes pour le lundi et X minutes pour le mardi. Pour obtenir des valeurs précises, nous devrions recueillir au minimum 200 instances dans chaque catégorie.

Cette modélisation d'entrepôt ouvre de nouvelles perspectives pour la création de tableaux de bord spécifiques aux métiers des ambulances.

6.6 Discussion

Nous discutons dans cette section de la qualité des trajectoires obtenues, ainsi que de la montée en sémantique de celles-ci. Les résultats obtenus dans cette étude sont très cohérents avec les observations métiers. L'insertion des trajectoires sémantiques dans un entrepôt de données pourrait permettre une analyse fine et précise du métier. Ces analyses de durées ainsi que des tournées typiques pourraient également aider à identifier les points où notre processus de montée en sémantique montre ses limites.

6.6.1 Segmentation

Pour rappel, la méthode de segmentation employée dans ce chapitre utilise trois capteurs afin de déterminer si le véhicule est en arrêt ou en déplacement. Bien que cette méthode de détection s'affranchit du GPS, et donc, des désavantages liés à celui-ci (Effet rebond, pelote, perte de signal indoor,...), restent les problèmes liés à la qualité des capteurs et l'erreur humaine. Si le conducteur ne desserre pas entièrement le frein à main, et qu'il roule à une vitesse inférieure à 5km/h (seuil que nous avons fixé pour déterminer un arrêt), alors l'algorithme considérera que le véhicule est à l'arrêt. L'utilisation d'avantages de capteurs permettrais d'améliorer la précision de cette segmentation. La méthode de segmentation, dans le cadre de l'étude de déplacement de véhicule d'urgence est donc satisfaisante, et, les déplacements très courts non détectés ne sont pas une gêne pour l'analyse des déplacements de ces véhicules.

6.6.2 Appariement

Comme vu lors des différentes étapes d'analyses des résultats, l'appariement des STOP avec des POI pose parfois quelques problèmes. Tout d'abord, bien que beaucoup de POI soient facilement extraits de différentes sources de données (OSM dans notre étude), on ne peut considérer cette liste comme exhaustive. L'activité des STOP étant en partie déterminée par l'association à un POI, ce manque potentiel peut donc entraîner un mauvais étiquetage des activités de STOP. Ensuite, les POI sont représentés soit par des points, soit par des polygones. Parfois, les STOP sont relativement éloignés d'un POI si celui-ci est représenté par un point, alors que sa représentation par un polygone ferait que le STOP serait automatiquement inclus dans celui-ci. Également, certains POI représentés par des polygones ont une emprise très grande dans l'espace. Les STOP alentours sont donc automatiquement appariés à ceux-ci alors qu'ils ne devraient pas. De ce fait, il est difficile de donner un seuil d'appariement valable pour tous les POI. Une des évolutions possible est de changer la méthode d'appariement afin qu'elle prenne en compte plusieurs critères :

— Le mode de représentation du POI (point ou polygone)
— L'emprise spatiale du POI
— Le nombre de POI dans la zone du STOP

Ces différents critères permettraient de non pas affecter nécessairement un POI à un STOP, mais plutôt d'attribuer une probabilité d'appariement avec un POI ou pas d'appariement.

6.6.3 Utilisation de règles métiers

La difficulté de l'utilisation d'un moteur d'inférence classique comme proposé dans ce chapitre est l'acquisition des règles qui le constituent.

Bien qu'il soit assez simple pour un expert de déterminer une liste d'activités simples de ses objets mobiles, il lui est très difficile de déterminer une liste d'activités fines et exhaustives. Une fois cette liste arrêtée, créer les règles permettant l'attribution d'une activité est un processus pouvant être long et fastidieux.

De plus, la méthode de raisonnement utilisée actuellement à des contraintes fortes. Les règles ne doivent pas être contradictoires ni redondantes, rendant l'ajout de règle de plus en plus contraignant.

La description des règles devient extrêmement ardue pour un humain en cas de multiplication des capteurs, même expert du domaine concerné, si les activités listées sont précises (départ en urgence, transport délicat, par exemple) et que les capteurs disponibles sont nombreux (comment traduire par une règle les concepts "rapide" ou "lent"). D'ailleurs, nous avons discuté de cette piste dans [Bisone et al., 2019] en proposant d'utiliser une approche basée sur le raisonnement à base de cas. Dans ce contexte, il se pose la question de la définition des cas prototypiques ainsi que de la mesure de similarité entre les cas. Les approches d'apprentissage plus gourmandes en données à l'instar des réseaux de neurones ou des forêts aléatoires auraient pu être intéressantes si nous disposions de suffisamment de données. C'est d'ailleurs la démarche utilisée dans [Rehrl et al., 2020] sur les données GPS. Cependant, dans cet article, la démarche est purement basée sur la construction d'un jeu d'apprentissage performant (avec un coût d'étiquetage non négligeable). Elle ne prend pas en compte les règles métiers connues. Aussi, il faudrait peut-être privilégier des approches permettant d'intégrer l'humain et ses connaissances au cours de l'apprentissage.

6.7 Conclusion

À moyen terme, de nouvelles règles métiers seront ajoutées pour définir l'activité des STOP. De même, de nouveaux cas prototypiques seront introduits afin de détecter des activités plus fines (transport délicat ou rapide d'un patient à l'hôpital, ...). Ces nouveaux cas devront intégrer les activités très courtes des MOVE comme les déplacements sur un même parking. L'algorithme de segmentation des MOVE devra également être validé pour détecter les cas de reroutage. Enfin, ce processus sera testé avec un jeu de données issu d'autres types de véhicules par exemple des taxis.

À plus long terme, l'objectif est de développer des systèmes de guidage GPS adaptés à des métiers et des véhicules spécifiques. Ils proposeront des itinéraires optimisés en fonction des contraintes du véhicule et de l'objectif du déplacement.

Algorithm 12 Assignation d'une activité à un MOVE

Require: m, s_p, s_n //le MOVE étudié, le STOP précédent le MOVE, le STOP suivant le MOVE.

$A_{s_p} \leftarrow s_p.STOPActivityType.label$ //l'activité liée à s_p.
$A_{s_n} \leftarrow s_n.STOPActivityType.label$ //l'activité liée à s_n.
switch ($m.reroutage$)
  case "Début redirection"**:**
    $m.MOVEActivityType.label \leftarrow$ "Retour avant redirection"
  case "Fin redirection"**:**
    $m.MOVEActivityType.label \leftarrow$ "Départ vers intervention (reroutage)"
  case "Début annulation"**:**
    $m.MOVEActivityType.label \leftarrow$ "Départ vers intervention avant annulation"
  case "Fin annulation"**:**
    $m.MOVEActivityType.label \leftarrow$ "Retour caserne après annulation"
  default:
    switch (A_{s_n})
      case "Attente fin de prise en charge hôpital"**:**
        $m.MOVEActivityType.label \leftarrow$ "Trajet intra hôpital"
      case "Arrêt non positionnable"**:**
        $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt non positionnable"
      case "Attente hôpital"**:**
        $m.MOVEActivityType.label \leftarrow$ "Trajet logistique hôpital"
      case "Logistique"**:**
        $m.MOVEActivityType.label \leftarrow$ "Logistique"
      case "Prise en charge du patient à l'hôpital"**:**
        if A_{s_p} == "Prise en charge du patient sur lieu d'intervention" **then**
          $m.MOVEActivityType.label \leftarrow$ "Transport vers hôpital"
        else
          $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers Hôpital"
        end if
      case "Prise en charge patient sur lieu d'intervention"**:**
        if A_{s_p} =="Attente caserne" **then**
          $m.MOVEActivityType.label \leftarrow$ "Départ vers intervention"
        else
          $m.MOVEActivityType.label \leftarrow$ "Départ vers intervention (reroutage)"
        end if
      case "Attente caserne"**:**
        if A_{s_p} =="Attente caserne" **then**
          $m.MOVEActivityType.label \leftarrow$ "Trajet intra caserne"
        else
          $m.MOVEActivityType.label \leftarrow$ "Retour caserne"
        end if
      default:
        if s_n.empty **then**
          $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt non positionnable"
        else
          $m.MOVEActivityType.label \leftarrow$ "Trajet indéterminé - vers arrêt indéterminé"
        end if
    end switch
end switch

Chapitre 7

Conclusion générale

Ce chapitre résume les contributions de cette thèse et propose des travaux futurs. La figure 7.1 reprend la figure 1.1. Le cadre vert inclut les tâches réalisées et validées durant cette thèse. Les cadres jaunes indiquent que les tâches ont été partiellement traitées, mais des études supplémentaires sont nécessaires. Les cadres rouges indiquent que ces tâches ont été identifiées et donneront lieu à des futurs travaux. Les sections suivantes décrivent ces différentes tâches.

7.1 Rappel des contributions

La contribution majeure de cette thèse est la mise en place d'un processus permettant l'extraction automatique des activités à partir de données brutes issues de véhicules de secours. Pour concevoir ce processus, nous avons dû apporter des solutions à des verrous technologiques et scientifiques.

Le premier verrou technologique porte sur l'acquisition et le traitement des données. Au total, nous avons utilisé 4 capteurs (frein à main, état du moteur, état du gyrophare, vitesse compteur) issus de l'IoT en plus des données classiques provenant du capteur GPS. Ces données sont agrégées à la seconde (la valeur absolue maximum est retenue pour les capteurs ayant une fréquence supérieure à 60Hz). La pertinence des résultats que nous avons obtenus montre que le choix de ces 5 capteurs (GPS + IoT) pourrait permettre de limiter le transfert de données dans un objectif de produire les traitements en ligne.

Nous avons ensuite défini un modèle générique pour la modélisation globale du processus de montée en sémantique de trajectoires. Ce modèle est générique et peut être étendu à d'autres véhicules métiers. Pour peupler la base issue de ce modèle, nous avons du définir un processus de montée en sémantique et résoudre deux verrous. Le premier concerne la segmentation de trajectoire, afin de définir si le véhicule est considéré en mouvement ou non. Ce processus de segmentation a permis de distinguer les arrêts courts (attente à un feu...) d'une part et d'autre part les arrêts

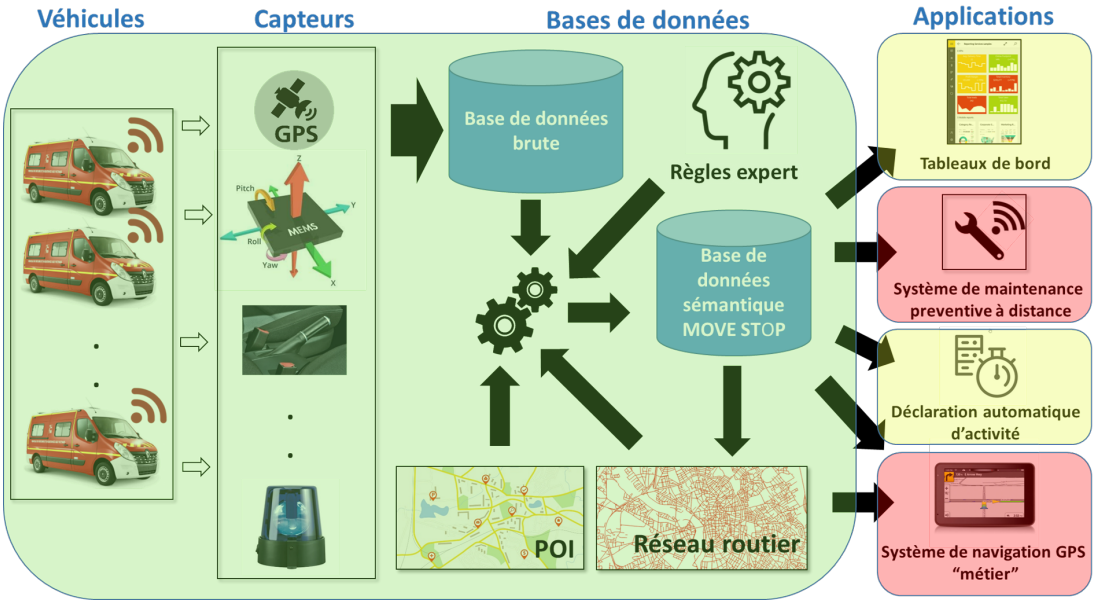


FIGURE 7.1 – Aperçu des contributions et des futurs travaux

liés à l'activité métier d'une ambulance. Nous avons ensuite reliés les STOP extraits à des points d'intérêt (POI) différents (casernes, hôpitaux, cliniques, lieux d'intervention ...) que nous avions préalablement identifiés. Le deuxième verrou concerne l'automatisation de l'enrichissement sémantique. Dans ce but, un système expert basé sur des règles métiers et des données des 6 capteurs identifiés a été défini. Ce processus a été appliqué sur les données collectées durant 15 jours de deux ambulances connectées conçues par Gruau et mises en service au SDIS 37. Nous avons pu extraire automatiquement 222 tournées regroupant 1 306 activités. Ce processus a permis d'extraire des comportements complexes tels que 27 reroutages et des zones d'interventions fréquentes. De même, les STOP et les MOVE ont pu être classifiés finement en fonction de l'activité.

Les travaux présentés dans cette thèse ont donné lieu à un article dans la revue internationale de géomatique [Bisone et al., 2018], une communication dans la conférence francophone SAGEO 2017 [Bisone et al., 2017] ainsi qu'à l'atelier EM-GIS 2019 associé à la conférence internationale SIGSPATIAl2019 [Bisone et al., 2019].

7.2 Travaux futurs

Ces travaux ouvrent aussi de nombreuses perspectives pour de nouvelles applications et peuvent également être adaptés à d'autres contextes.

A court terme, nous devons traiter les différents problèmes de **validation de nos résultats**. Pour cela, nous devons collecter manuellement des données directement dans les véhicules d'intervention afin de confronter cette vérité terrain aux résultats obtenus par notre processus. De même, un passage en temps réel permettra la définition d'une déclaration automatique d'activité des équipes d'intervention et une collecte plus aisée de gros volumes de données.

À moyen terme, la collecte de ces gros volumes de données permettra une meilleure compréhension des activités des ambulances à l'aide de **tableaux de bords** spécifiques co-construit avec des experts du domaines. L'objectif est de les aider à gérer plus finement leurs flottes de véhicules d'intervention, justifier les besoins en véhicules professionnels pour chaque centre et identifier de potentielles difficultés dans le traitement des interventions. De même, une fois la collecte en temps réel mise en place et la qualité de la détection automatique démontrée, un **système de déclaration automatique d'activité** pourra être définit. Ce dernier améliorera le système manuel actuel et déchargera les pompiers de cette activité secondaire.

A plus long terme, lorsque le volume de données sera beaucoup plus important, des systèmes intelligents s'appuyant sur les connaissances extraites doivent pouvoir être développés. Le premier est un **système de navigation GPS métier**. Avant de proposer ce système, l'étude des trajectoires devra encore être affinée et de nouveaux capteurs IoT intégrés dans le processus. Par exemple, l'ajout des accéléromètres dans les données collectées permettra de mieux connaître les propriétés des tronçons routiers empruntées par les ambulances (nid de poule, dos d'âne, virage dangereux...). De même, nous disposerons de vitesses moyennes mesurées réelles pour des activités et des périodes similaires spécifiques aux ambulances. Ces nouvelles connaissances permettront de définir un itinéraire adapté aux véhicules d'urgence en fonction de leur activité. Également, un **système de maintenance préventive** à distance devra être produit. L'étude poussée de l'évolution des données de capteurs, couplée à la mise en place de nouveaux capteurs sur le véhicule permettra de déterminer l'usure des différents composants en fonction de son utilisation réelle et ainsi d'adapter à chaque véhicule son entretien.

Afin de démontrer que les résultats obtenus à Tours sont transposables dans d'autres régions, il serait important d'explorer un nouveau terrain d'étude. Finalement, il existe de nombreux véhicules métiers différents (camion de collecte de déchets, camions de livraison...). Ces véhicules sont équipés de nombreux capteurs et sont suivis à travers des systèmes de gestion de flottes en temps réel. Ce modèle et ce processus de montée en sémantique étant **génériques**, ils devraient être adaptables à ces véhicules.

Cette thèse a donc proposé un modèle et un processus de montée en sémantique génériques et originaux se focalisant sur la détection des activités métiers. Elle offre des perspectives applicatives prometteuses.

Bibliographie

[iso, 2015] (2015). Véhicules routiers — gestionnaire de réseau de communication (can) — partie 1 : Couche liaison de données et signalisation physique.

[iso, 2016] (2016). Véhicules routiers — gestionnaire de réseau de communication (can) — partie 2 : Unité d'accès au support à haute vitesse.

[Al-Mallah, 2018] Al-Mallah, R. (2018). Road Traffic Congestion Analysis Via Connected Vehicles. PhD thesis, École Polytechnique de Montréal.

[Alvares et al., 2007] Alvares, L., Bogorny, V., Kuijpers, B., de Macedo, J., Moelans, B., and Vaisman, A. (2007). A model for enriching trajectories with semantic geographical information. Proc. of the 15th annual ACM international symposium on Advances GIS, 22 :1–8.

[Amici et al., 2014] Amici, R., Bonola, M., Bracciale, L., Rabuffi, A., Loreti, P., and Bianchi, G. (2014). Performance assessment of an epidemic protocol in vanet using real traces. In MoWNet, pages 92–99.

[Andrienko et al., 2013] Andrienko, G. L., Andrienko, N. V., Fuchs, G., Raimond, A.-M. O., Symanzik, J., and Ziemlicki, C. (2013). Extracting semantics of individual places from movement data by analyzing temporal patterns of visits. In COMP@ SIGSPATIAL, pages 9–15.

[Avatefipour and Malik, 2018] Avatefipour, O. and Malik, H. (2018). State-of-the-art survey on in-vehicle network communication (can-bus) security and vulnerabilities. arXiv preprint arXiv :1802.01725.

[Beber et al., 2017] Beber, M. A. et al. (2017). Individual and group activity recognition in moving object trajectories.

[Beber et al., 2016] Beber, M. A., Ferrero, C. A., Fileto, R., and Bogorny, V. (2016). Towards activity recognition in moving object trajectories from twitter data. In GeoInfo, pages 68–79.

[Beeri et al., 2004] Beeri, C., Kanza, Y., Safra, E., and Sagiv, Y. (2004). Object fusion in geographic information systems. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, pages 816–827.

[Bermingham and Lee, 2018] Bermingham, L. and Lee, I. (2018). A probabilistic stop and move classifier for noisy gps trajectories. Data Mining and Knowledge Discovery, 32(6) :1634–1662.

[Bian et al., 2018] Bian, J., Tian, D., Tang, Y., and Tao, D. (2018). A survey on trajectory clustering analysis.

[Bimonte et al., 2020] Bimonte, S., Hassan, A., and Lenain, R. (2020). Une nouvelle méthodologie de géovisualisation pour les entrepôts de données de trajectoires. In Business Intelligence & Big Data : 15ème Edition de la conférence EDA, Montpellier France 2019. BoD-Books on Demand.

[Birant and Kut, 2007] Birant, D. and Kut, A. (2007). St-dbscan : An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering, 60(1) :208–221.

[Bisone et al., 2019] Bisone, F., Devogele, T., and Etienne, L. (2019). From raw sensor data to semantic trajectories. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management, pages 1–4.

[Bisone et al., 2017] Bisone, F., Etienne, L., and Devogele, T. (2017). Extraction automatique de la sémantique des trajectoires. In Spatial Analysis and GEOmatics 2017, Rouen, France. INSA de rouen.

[Bisone et al., 2018] Bisone, F., Etienne, L., and Devogele, T. (2018). Modélisation et extraction de la sémantique des trajectoires à partir de données multicapteurs. Revue Internationale de Géomatique, 28(4) :461–483.

[Bogorny et al., 2014] Bogorny, V., Renso, C., de Aquino, A. R., de Lucca Siqueira, F., and Alvares, L. O. (2014). Constant–a conceptual data model for semantic trajectories of moving objects. Transactions in GIS, 18(1) :66–88.

[Bolbol et al., 2012] Bolbol, A., Cheng, T., Tsapakis, I., and Haworth, J. (2012). Inferring hybrid transportation modes from sparse gps data using a moving window svm classification. Computers, Environment and Urban Systems, 36(6) :526–537.

[Browne, 2009] Browne, P. (2009). JBoss Drools business rules. Packt Publishing Ltd.

[Buard, 2013] Buard, É. (2013). Dynamiques des interactions espèces-espace : mise en relation des pratiques de déplacement des populations d'herbivores et de l'évolution de l'occupation du sol dans le parc de Hwange (Zimbabwe). PhD thesis.

[Cain III et al., 2005] Cain III, J. W., Krausman, P. R., Jansen, B. D., and Morgart, J. R. (2005). Influence of topography and gps fix interval on gps collar performance. Wildlife Society Bulletin, 33(3) :926–934.

[Cao et al., 2007] Cao, H., Mamoulis, N., and Cheung, D. W. (2007). Discovery of periodic patterns in spatiotemporal sequences. IEEE Transactions on Knowledge and Data Engineering, 19(4) :453–467.

[Chen et al., 1996] Chen, M.-S., Han, J., and Yu, P. S. (1996). Data mining : an overview from a database perspective. IEEE Transactions on Knowledge and data Engineering, 8(6) :866–883.

[Chen et al., 2014] Chen, W., Ji, M., and Wang, J. (2014). T-dbscan : A spatiotemporal density clustering for gps trajectory segmentation. International Journal of Online Engineering, 10(6).

[Chin et al., 1987] Chin, G. Y. et al. (1987). Two-dimensional measures of accuracy in navigational systems. Technical report, United States. Dept. of Transportation. Office of Research and Special Programs.

[Douglas and Peucker, 1973] Douglas, D. H. and Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica : the international journal for geographic information and geovisualization, 10(2) :112–122.

[Duran and Earleywine, 2013] Duran, A. and Earleywine, M. (2013). Gps data filtration method for drive cycle analysis applications. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

[Ester et al., 1996] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pages 226–231.

[Etienne, 2011] Etienne, L. (2011). Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime. PhD thesis.

[Etienne et al., 2012] Etienne, L., Devogele, T., and Bouju, A. (2012). Spatio-temporal trajectory analysis of mobile objects following the same itinerary. Advances in Geo-Spatial Information Science, 10 :47–57.

[Etienne et al., 2016] Etienne, L., Devogele, T., Buchin, M., and McArdle, G. (2016). Trajectory box plot : a new pattern to summarize movements. International Journal of Geographical Information Science, 30(5) :835–853.

[Feliachi, 2017] Feliachi, A. (2017). Interconnexion et visualisation de ressources géoréférencées du Web de données à l'aide d'un référentiel topographique de support. PhD thesis, Paris Est.

[Gama et al., 2004] Gama, J., Medas, P., Castillo, G., and Rodrigues, P. (2004). Learning with drift detection. In Brazilian symposium on artificial intelligence, pages 286–295. Springer.

[Genre-Grandpierre and Josselin, 2005] Genre-Grandpierre, C. and Josselin, D. (2005). Des transports à la demande pour répondre aux nouvelles formes de mobilité. le concept de modulobus.

[Giacometti et al., 2014] Giacometti, A., Li, D. H., Marcel, P., and Soulet, A. (2014). 20 years of pattern mining : a bibliometric survey. ACM SIGKDD Explorations Newsletter, 15(1) :41–50.

[Giannotti et al., 2011] Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., and Trasarti, R. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB Journal, 20(5) :695.

[Giannotti et al., 2007] Giannotti, F., Nanni, M., Pinelli, F., and Pedreschi, D. (2007). Trajectory pattern mining. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 330–339. ACM.

[Girres and Touya, 2010] Girres, J. and Touya, G. (2010). Quality assessment of the french openstreetmap dataset. Trans. GIS, 14(4) :435–459.

[Gudmundsson et al., 2009] Gudmundsson, J., Katajainen, J., Merrick, D., Ong, C., and Wolle, T. (2009). Compressing spatio-temporal trajectories. Comput. Geom., 42 :825–841.

[Hägerstrand, 1970] Hägerstrand, T. (1970). What about people in regional science ? In Papers of the Regional Science Association, volume 24.

[Hershberger and Snoeyink, 1992] Hershberger, J. and Snoeyink, J. (1992). Speeding Up the Douglas-Peucker Line-Simpliication Algorithm. Technical report, University of British Columbia.

[Huang et al., 2020] Huang, Z., Qiao, S., Han, N., Yuan, C.-a., Song, X., and Xiao, Y. (2020). Survey on vehicle map matching techniques. CAAI Transactions on Intelligence Technology.

[Inmon, 1995] Inmon, W. H. (1995). What is a data warehouse ? Prism Tech Topic, 1(1) :1–5.

[Ivanović et al., 2016] Ivanović, S. S., Raimond, A.-M. O., Mustière, S., and Devogele, T. (2016). Detection of outliers in crowdsourced gps traces. In 12th international symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

[J-G Lee, 2007] J-G Lee, J Han, K.-Y. W. (2007). Trajectory clustering : a partition and- group framework. in Proceedings of the 2007 ACM SIGMOD international conference on Management of data. ACM , 2007, pp. 593–604.

[Jeung et al., 2010] Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., and Shen, H. T. (2010). Discovery of convoys in trajectory databases. arXiv preprint arXiv :1002.0963.

[Junior et al., 2017] Junior, A. S., Renso, C., and Matwin, S. (2017). Analytic : An active learning system for trajectory classification. IEEE computer graphics and applications, 37(5) :28–39.

[Kami et al., 2010] Kami, N., Enomoto, N., Baba, T., and Yoshikawa, T. (2010). Algorithm for detecting significant locations from raw gps data. In International Conference on Discovery Science, pages 221–235. Springer.

[Karaim et al., 2018] Karaim, M., Elsheikh, M., Noureldin, A., and Rustamov, R. (2018). Gnss error sources. Multifunctional Operation and Application of GPS ; Rustamov, RB, Hashimov, AM, Eds, pages 69–85.

[Kerouanton, 2020] Kerouanton, C. (2020). Analyse spatio-temporelle des mobilités de randonneurs da Theses, Université de Savoie, Chambéry, FRA.

[Khan et al., 2014] Khan, K., Rehman, S. U., Aziz, K., Fong, S., and Sarasvady, S. (2014). Dbscan : Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014), pages 232–238. IEEE.

[Klotz et al., 2018] Klotz, B., Troncy, R., Wilms, D., and Bonnet, C. (2018). Generating semantic trajectories using a car signal ontology. In Companion Proceedings of the The Web Conference 2018, pages 135–138. International World Wide Web Conferences Steering Committee.

[Knight and Wang, 2009] Knight, N. L. and Wang, J. (2009). A comparison of outlier detection procedures and robust estimation methods in gps positioning. The Journal of Navigation, 62(4) :699–709.

[Krempl et al., 2011] Krempl, G., Siddiqui, Z. F., and Spiliopoulou, M. (2011). Online clustering of high-dimensional trajectories under concept drift. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 261–276. Springer.

[Laney, 2001] Laney, D. (2001). 3d data management : Controlling data volume, velocity, and variety. [Online ; accessed 23-July-2019].

[Lee et al., 2008] Lee, J.-G., Han, J., Li, X., and Gonzalez, H. (2008). Traclass : trajectory classification using hierarchical region-based and trajectory-based clustering. Proceedings of the VLDB Endowment, 1(1) :1081–1094.

[Li, 2014] Li, Z. (2014). Spatiotemporal pattern mining : algorithms and applications. In Frequent Pattern Mining, pages 283–306. Springer.

[Li et al., 2010] Li, Z., Ding, B., Han, J., Kays, R., and Nye, P. (2010). Mining periodic behaviors for moving objects. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1099–1108.

[Li et al., 2011] Li, Z., Han, J., Ji, M., Tang, L.-A., Yu, Y., Ding, B., Lee, J.-G., and Kays, R. (2011). Movemine : Mining moving object data for discovery of animal movement patterns. ACM Transactions on Intelligent Systems and Technology (TIST), 2(4) :1–32.

[Longgang and Xiaotian, 2016] Longgang, X. and Xiaotian, S. (2016). Visualization and extraction of trajectory stops based on kernel-density. Acta Geodaetica et Cartographica Sinica, 45(9) :1122.

[Luo et al., 2017] Luo, T., Zheng, X., Xu, G., Fu, K., and Ren, W. (2017). An improved DBSCAN algorithm to detect stops in individual trajectories. ISPRS Int. J. Geo-Information, 6(3) :63.

[Mamoulis et al., 2004] Mamoulis, N., Cao, H., Kollios, G., Hadjieleftheriou, M., Tao, Y., and Cheung, D. W. (2004). Mining, indexing, and querying historical spatiotemporal data. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 236–245.

[Mazimpaka and Timpf, 2016] Mazimpaka, J. D. and Timpf, S. (2016). Trajectory data mining : A review of methods and applications. Journal of Spatial Information Science, 2016(13) :61–99.

[Meratnia and Rolf, 2004] Meratnia, N. and Rolf, A. (2004). Spatiotemporal compression techniques for moving point objects. In International Conference on Extending Database Technology, pages 765–782. Springer.

[Mittelstadt and Floridi, 2016] Mittelstadt, B. D. and Floridi, L. (2016). The ethics of big data : current and foreseeable issues in biomedical contexts. Science and engineering ethics, 22(2) :303–341.

[Mooney and Minghini, 2017] Mooney, P. and Minghini, M. (2017). A review of openstreetmap data. In Citizen Sensor, pages 37–59.

[Moreau et al., 2020a] Moreau, C., Devogele, T., Étienne, L., Peralta, V., and de Runz, C. (2020a). Methodology for mining, discovering and analyzing semantic human mobility behaviors. CoRR, abs/2012.04767.

[Moreau et al., 2020b] Moreau, C., Devogele, T., Peralta, V., and Étienne, L. (2020b). A contextual edit distance for semantic trajectories. In Hung, C., Cerný, T., Shin, D., and Bechini, A., editors, SAC '20 : The 35th ACM/SIGAPP Symposium on Applied Computing, online event, [Brno, Czech Republic], March 30 - April 3, 2020, pages 635–637. ACM.

[Mustière and Devogele, 2008] Mustière, S. and Devogele, T. (2008). Matching networks with different levels of detail. GeoInformatica, 12(4) :435–453.

[Olteanu-Raimond et al., 2015] Olteanu-Raimond, A.-M., Mustiere, S., and Ruas, A. (2015). Knowledge formalization for vector data matching using belief theory. Journal of Spatial Information Science, 2015(10) :21–46.

[Ordonez et al., 2011] Ordonez, C., Martínez, J., Rodríguez-Pérez, J., and Reyes, A. (2011). Detection of outliers in gps measurements by using functional-data analysis. Journal of Surveying Engineering, 137(4) :150–155.

[Paiva Nogueira, 2017] Paiva Nogueira, T. (2017). A Framework for Automatic Annotation of Semantic Trajectories. Theses, Université Grenoble Alpes.

[Parent et al., 2013] Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., et al. (2013). Semantic trajectories modeling and analysis. ACM Computing Surveys (CSUR), 45(4) :1–32.

[Parkes and Thrift, 1980] Parkes, D. and Thrift, N. J. (1980). Times, spaces, and places : a chronogeographic perspective. J. Wiley, Chichester [Eng.] ; New York.

[Perret et al., 2013] Perret, J., Marty, J., and Abouelijane, L. (2013). Posamu : Les défis de la géomatique pour répondre aux besoins de l'urgence. In Journées de la recherche de l'IGN, 25 avril 2013.

[Popa et al., 2015] Popa, I. S., Zeitouni, K., Oria, V., and Kharrat, A. (2015). Spatio-temporal compression of trajectories in road networks. GeoInformatica, 19(1) :117–145.

[Quddus et al., 2007] Quddus, M. A., Ochieng, W. Y., and Noland, R. B. (2007). Current map-matching algorithms for transport applications : State-of-the art and future research directions. Transportation research part c : Emerging technologies, 15(5) :312–328.

[Raifer et al., 2020] Raifer, M., Auer, M., Loos, L., Troilo, R., Kowatsch, F., Visintini, J., and Zipf, A. (2020). Ohsome – openstreetmap data quality analysis. In 3rd International Workshop on Spatial Data Quality, Valletta, Malta.

[Rajaraman and Ullman, 2011] Rajaraman, A. and Ullman, J. D. (2011). Mining of massive datasets. Cambridge University Press.

[Ramon-Gonen and Gelbard, 2017] Ramon-Gonen, R. and Gelbard, R. (2017). Cluster evolution analysis : Identification and detection of similar clusters and migration patterns. Expert Systems with Applications, 83 :363–378.

[Rehrl et al., 2020] Rehrl, K., Gröchenig, S., and Kranzinger, S. (2020). Why did a vehicle stop ? a methodology for detection and classification of stops in vehicle trajectories. International Journal of Geographical Information Science, 34(10) :1953–1979.

[Renaudin et al., 2007] Renaudin, V., Yalak, O., Tomé, P., and Merminod, B. (2007). Indoor navigation of emergency agents. European Journal of Navigation, 5(ARTICLE) :36–45.

[Renso et al., 2013] Renso, C., Spaccapietra, S., and Zimányi, E. (2013). Mobility data. Cambridge University Press.

[Rocha et al., 2010] Rocha, J. A. M., Times, V. C., Oliveira, G., Alvares, L. O., and Bogorny, V. (2010). Db-smot : A direction-based spatio-temporal clustering method. In 2010 5th IEEE international conference intelligent systems, pages 114–119. IEEE.

[Saini et al., 2020] Saini, R., Kumar, P., Roy, P. P., and Pal, U. (2020). Trajectory classification using feature selection by genetic algorithm. In Proceedings of 3rd International Conference on Computer Vision and Image Processing, pages 377–388. Springer.

[Schneider et al., 2013] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., and González, M. C. (2013). Unravelling daily human mobility motifs. Journal of The Royal Society Interface, 10(84) :20130246.

[Sethi and Kantardzic, 2017] Sethi, T. S. and Kantardzic, M. (2017). On the reliable detection of concept drift from streaming unlabeled data. Expert Systems with Applications, 82 :77–99.

[Small, 1990] Small, C. G. (1990). A survey of multidimensional medians. International Statistical Review/Revue Internationale de Statistique, pages 263–277.

[Spaccapietra et al., 2008] Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. Data & knowledge engineering, 65(1) :126–146.

[Stopher et al., 2005] Stopher, P. R., Jiang, Q., FitzGerald, C., et al. (2005). Processing gps data from travel surveys. 2nd international colloqium on the behavioural foundations of integrated land-use and transportation models : frameworks, models and applications, Toronto.

[Su et al., 2014] Su, J., Kurtek, S., Klassen, E., Srivastava, A., et al. (2014). Statistical analysis of trajectories on riemannian manifolds : bird migration, hurricane tracking and video surveillance. The Annals of Applied Statistics, 8(1) :530–552.

[Tran et al., 2011] Tran, L. H., Nguyen, Q. V. H., Do, N. H., and Yan, Z. (2011). Robust and hierarchical stop discovery in sparse and diverse trajectories. Technical report.

[Truong et al., 2019] Truong, Q. T., de Runz, C., and Touya, G. (2019). Analysis of collaboration networks in openstreetmap through weighted social multigraph mining. International Journal of Geographical Information Science, 33(8) :1651–1682.

[US Office of the Department of Defense, 2020] US Office of the Department of Defense (2020). Global positioning system (gps) standard positioning service (sps) performance analysis report. GPS Product Team : Washington, DC, USA.

[Wang et al., 2009] Wang, Y., Zhang, X., and Draxler, R. R. (2009). Trajstat : Gis-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data. Environmental Modelling & Software, 24(8) :938–939.

[Weiszfeld and Plastria, 2009] Weiszfeld, E. and Plastria, F. (2009). On the point for which the sum of the distances to n given points is minimum. Annals of Operations Research, 167(1) :7–41.

[William et al., 2014] William, J. H., NSTB/WAAS, and Airport, A. C. I. (2014). Global positioning system (gps) standard positioning service (sps) performance analysis report. GPS Product Team : Washington, DC, USA.

[Wu et al., 2015] Wu, F., Li, Z., Lee, W.-C., Wang, H., and Huang, Z. (2015). Semantic annotation of mobility data using social media. In Proceedings of the 24th International Conference on World Wide Web, pages 1253–1263.

[Wu et al., 2019] Wu, X., Liao, L., Zou, F., Liu, J., Chen, B., and Zheng, Y. (2019). Ts-dbscan : To detect trajectory anomaly for transportation vehicles. In International Conference on Genetic and Evolutionary Computing, pages 151–160. Springer.

[Yan et al., 2011] Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., and Aberer, K. (2011). Semitri : a framework for semantic annotation of heterogeneous trajectories. In Proceedings of the 14th international conference on extending database technology, pages 259–270.

[Yuan et al., 2017] Yuan, G., Sun, P., Zhao, J., Li, D., and Wang, C. (2017). A review of moving object trajectory clustering algorithms. Artificial Intelligence Review, 47(1) :123–144.

[Zhang et al., 2018] Zhang, D., Ding, M., Yang, D., Liu, Y., Fan, J., and Shen, H. T. (2018). Trajectory simplification : An experimental study and quality analysis. Proc. VLDB Endow., 11(9) :934–946.

[Zhang and Lin, 2019] Zhang, Y. and Lin, Y. (2019). An interactive method for identifying the stay points of the trajectory of moving objects. Journal of Visual Communication and Image Representation, 59 :387–392.

[Zheng, 2015] Zheng, Y. (2015). Trajectory data mining : an overview. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3) :1–41.

[Zheng et al., 2009] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In Proceedings of the 18th international conference on World wide web, pages 791–800.

[Zimmermann et al., 2009] Zimmermann, M., Kirste, T., and Spiliopoulou, M. (2009). Finding stops in error-prone trajectories of moving objects with time-based clustering. In International Conference on Intelligent Interactive Assistance and Mobile Multimedia Computing, pages 275–286. Springer.

Annexes

Annexe A

Capteurs

A.1 capteurs

- Capteurs classiques
  - Longitude
  - Latitude
  - Accéléromètre longitude
  - Accéléromètre latitude
  - Accéléromètre vertical
- Capteurs véhicule (CAN)
  - Vitesse
  - Clignotant droit
  - Clignotant gauche
  - Pédale de frein
  - Veilleuse
  - Feux de croisement
  - Feux de route
  - Marche arrière
  - Voltage châssis
  - Vitesse de rotation du moteur
  - Témoin problème moteur
  - Accélérateur
  - Odomètre
  - Contact moteur
  - Consommation carburant
  - Porte conducteur
  - Porte passager
  - Porte arrière
  - Frein à main
  - Ceinture de sécurité conducteur
  - Airbag
- Capteurs spécifiques ambulance
  - Feux de pénétration
  - Gyrophares
  - Sirène
  - Chauffage cellule sanitaire
  - Voltage cellule sanitaire