



HAL
open science

Extraction automatique de la sémantique des trajectoires

Frédéric Bisone, Laurent Etienne, Thomas Devogele

► **To cite this version:**

Frédéric Bisone, Laurent Etienne, Thomas Devogele. Extraction automatique de la sémantique des trajectoires. Spatial Analysis and GEomatics 2017, INSA de rouen, Nov 2017, Rouen, France. hal-01643365

HAL Id: hal-01643365

<https://hal.science/hal-01643365v1>

Submitted on 21 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extraction automatique de la sémantique de trajectoires

Frédéric Bisone^{1,2}, Laurent Etienne¹, Thomas Devogele¹

1. Laboratoire d'informatique de l'Université François Rabelais
64, avenue Jean Portalis, 37200 Tours France
prenom.nom@univ-tours.fr

2. Petit Picot by GRUAU
ZI n°2 – 18 rue de Prony, 37303 Joué-lès-Tours France
prenom.nom@gruau.com

RESUME. 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. Les objets mobiles sont maintenant équipés de plus en plus de capteurs complémentaires (génériques ou spécifiques à des types de véhicules). Cet article présente une méthode pour extraire la sémantique liée aux déplacements à partir des données brutes issues de l'ensemble de ces capteurs. Ces outils de fouille de trajectoires permettent d'identifier des lieux d'intérêt fréquents et les objectifs de déplacements. Ce processus a été appliqué aux déplacements d'ambulances connectées des sapeurs-pompiers d'Indre et Loire.

ABSTRACT. GPS position are useful to analyse movements of mobile objects. Unfortunately, the outcome can be unsatisfactory due to imprecision and signal lost. Several sensors (generic or specific ones depending on the type of vehicle) are now included into mobile objects. This article describes a data mining process based on data from GPS and other sensors.

This article presents the movement pattern extraction process based on the raw data from these sensors which help identified frequent places of interest and movement patterns. This process was successfully applied to the Indre et Loire fire department and its connected ambulances.

MOTS-CLES : objet mobile, fouille de trajectoires, trajectoires sémantiques, capteurs GPS, analyse de mobilité, véhicule de secours

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

1. Introduction

Depuis plusieurs années, la miniaturisation des capteurs GPS ainsi que la baisse de leur coût de fabrication entraînent leur apparition dans un grand nombre d'objets mobiles (Smartphones, véhicules, balises autonomes...). Notamment, les capteurs GPS font leur apparition dans les véhicules de plus en plus connectés capables de transmettre en temps réel l'état d'un grand nombre de capteurs présents dans ceux-ci.

Par exemple, l'envoi de positions GPS autorise le suivi en temps réel de flottes de véhicules, pour définir des zones de couvertures optimales, annoncer des temps d'arrivée estimée (dépanneuse, livreur.) Tandis que les informations issues d'autres capteurs se focalisent sur l'état du véhicule et ses capacités à fournir un service (niveau de carburant, taux de remplissage d'une benne pour un camion à ordures, etc.). De même, la combinaison des coordonnées GPS et des données issues des capteurs des véhicules permet la recherche de comportements de conduite : éco conduite, etc. Cependant, l'envoi par les véhicules de grandes quantités de données chaque seconde entraîne des problématiques classiques liées au big data. Le flux de données arrivant en continu produit une masse de données hétérogène considérable à traiter.

La détection de comportements est spécifique à chaque type d'objet mobile : voiture individuelle, bus scolaire, ambulance, véhicule de police, taxis, ramassage d'ordures, etc. L'analyse de comportement se révèle pertinente dans le cas d'un ensemble homogène d'objets mobiles. Les véhicules d'urgences ont notamment des comportements liés à leurs caractéristiques. Ils répondent à des contraintes génériques : hauteur, largeur, capacité de franchissement de carrefour, etc. De même, leurs contraintes métiers : urgence vitale, patient traumatisé, etc. entraînent un aménagement spécifique du code de la route qui nécessite l'adaptation des itinéraires classiques. Les systèmes de navigation GPS classiques ne tiennent pas compte de ces spécificités. Par contre, des systèmes d'extraction de connaissance sur de gros volumes de données portant sur les traces collectées d'un ensemble homogène d'objets mobiles devraient autoriser l'identification de ces comportements propres.

Cet article s'intéresse aux problématiques suivantes : Est-il possible d'extraire automatiquement des comportements métier d'un type d'objets mobiles à partir d'un grand nombre de trajectoires sur une agglomération ? Ce processus est-il suffisamment générique pour être transposable à d'autres agglomérations où à d'autres types de véhicules spécifiques ? Quelles sont les informations complémentaires aux données GPS nécessaires pour obtenir des résultats précis ?

Les sections suivantes décrivent les processus génériques d'extraction des connaissances. Tout d'abord la méthode pour extraire les stops et filtrer les micros - arrêts est présentée. Ensuite, la montée en sémantique sur les arrêts puis sur les trajectoires est abordée. Une expérimentation sur le cas d'ambulances de secours d'une caserne de pompiers est ensuite décrite. Une discussion autour de l'apport de ce processus conclut cet article.

2. Processus général

Le processus général qui est décrit dans cet article s'inscrit dans une démarche de fouille de trajectoires (Zheng 2015) il s'appuie sur des concepts autour de la mobilité.

La position est une suite de données longitude, latitude et valeurs des différents capteurs véhicules pour un instant t .

Le stop est une position particulière où l'objet est immobile.

Le micro-arrêt est une suite continue de positions stop de courte durée n'ayant pas de rapport avec l'activité du véhicule (Par exemple, une suite de positions stop à un cédez-le-passage pour un taxi)

Un arrêt est une suite continue d'un ou plusieurs stops en rapport avec l'activité du véhicule. (Par exemple, une suite de stops près d'un hôpital pour un véhicule du SAMU¹).

Un lieu d'intérêt (Point of Interest POI) est un lieu où le véhicule effectue des arrêts fréquents.

Un parcours est une suite ordonnée de positions, ces positions sont reliées 2 à 2 par un segment.

La trajectoire est le parcours d'un véhicule entre deux arrêts.

Le trajet est une boucle parcourue par un véhicule de son lieu de stationnement habituel jusqu'à son retour.

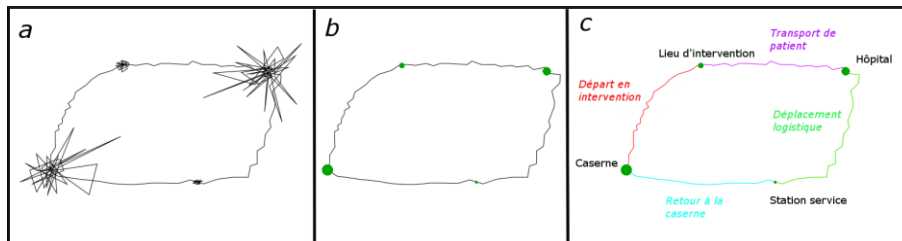


FIGURE 1. Exemple de montée en sémantique : données brutes (a), extraction des lieux d'arrêts(b), labellisation des lieux d'arrêts et des trajectoires(c)

Le schéma de la figure 1 présente les différentes étapes du processus de montée en sémantique jusqu'à l'obtention d'une trajectoire enrichie sémantiquement.

La première étape est l'acquisition des données issues des différents capteurs. Les véhicules sont équipés de GPS classique, de capteurs génériques liés au véhicule (frein à main, vitesse à l'ordinateur de bord, clignotants...) ainsi que de capteurs liés

¹ SAMU – Service d'Aide Médicale Urgente

à l'activité du véhicule (compteur de taxi, sirène d'une ambulance, charge d'une benne à ordures...). Cette étape consiste à importer les données dans une base de données géographique n'est pas détaillée dans cet article. Dans la deuxième étape, à partir de ces données, les lieux où les véhicules s'arrêtent fréquemment sont localisés à l'aide d'une méthode de clustering.

Dans la dernière étape, à l'aide de connaissances extérieures (Objet géographique d'OpenStreetMap (OSM), liste de points d'intérêt métier, etc.) et de règles métiers les lieux sont identifiés et les types de déplacements (objectifs du déplacement) sont déduits.

2.1. Détection des stops et filtrage des micros-arrêts :

2.1.1. Problèmes et solutions classiques rencontrés dans la littérature

Dans la littérature, plusieurs problèmes rendent la détection de stop (lieu et durée) complexe à l'aide de données GPS classiques ((Cain et al. 2005), (Duran, et Earleywine 2012), (Knight et al 2009), (Ivanovic et al 2016), (Ordoñez et al. 2011)).

Tout d'abord l'imprécision des capteurs GPS, et le fait que le signal soit très perturbé aux abords de bâtiments ou dans des canyons urbains 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 épais. L'analyse lorsque le véhicule stationne en garage fermé ou traverse des tunnels est donc problématique. La figure 2 montre les pelotes de positions d'un capteur à l'intérieur d'un bâtiment.



FIGURE 2. *Pelote à l'intérieur d'un bâtiment.*

De plus, même lors de la sortie d'un bâtiment, la qualité des données est liée au temps dont le GPS a besoin pour se synchroniser avec les satellites (TTF - Time To Fix).

Qui plus est, afin d'extraire des informations propres au véhicule et son activité, il faut distinguer les micro-arrêts des arrêts. Pour cela, les contraintes usuelles de circulation (feux rouges, passages piétons) doivent être prises en compte. Dans cet article, seule la fréquence des arrêts est étudiée.

Pour répondre à ces problèmes, (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 fronts géographiques d'entrées et de sorties pour chaque zone. 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.

2.1.2. Apport des capteurs autres que le GPS

Afin de déterminer le lieu d'arrêt d'un véhicule, le capteur de frein à main est pertinent. Le moment où le frein à main du véhicule est serré jusqu'à ce que celui-ci soit desserré permet de définir un arrêt de manière plus fiable que le capteur GPS. En effet, un capteur GPS peut indiquer un déplacement alors que l'objet est immobile (l'effet pelote peut par exemple indiquer une vitesse aberrante).

Une fois un arrêt déterminé, tous les stops de cet arrêt peuvent être remplacés par une position unique. Cette position est soit une position centrale (calculée à l'aide d'un concept proposé par (Small 1990) pour une pelote ou une position estimée en fonction des positions avant et après l'arrêt en cas de perte de signal.

Pour filtrer les micros-arrêts, un seuil temporaire est défini. Seuls les stops détectés durant plus d'un certain temps sont considérés comme des arrêts afin d'éviter les possibles stops, feu rouge, démarrage en côte. 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 est donc insuffisante mais des capteurs auxiliaires pourraient détecter ces stops (ouverture du coffre où se trouve le courrier dans le cas d'un véhicule de distribution.).

2.2. Montée en sémantique sur les arrêts :

2.2.1. Extraction des lieux d'arrêts fréquents ;

Une fois les arrêts localisés, nous pouvons les qualifier entre autres en termes de fréquence et de type de lieux. Pour trouver les arrêts fréquents des véhicules, un clustering des arrêts est effectué en fonction de leur position géographique.

L'algorithme de clustering DB SCAN (Ester et al. 1996) est un processus de partitionnement de données. Il utilise 2 paramètres en entrée : une distance ϵ ainsi qu'un nombre minimum de points MinPts. 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 3).

Pour qu'un point soit inclus dans un cluster, il doit soit être un centre de cluster, soit être couvert par un centre de cluster de rayon ϵ (points B et C de la figure 3). 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 3 sont tous des centres de

cluster qui forment un ensemble plus grand). Les points ne remplissant aucune des deux conditions sont considérés comme du bruit (Point N de la figure 3).

Dans notre étude, nous utiliserons uniquement la position centrale de chaque arrêt, les valeurs temporelles ne sont pas prises en compte avec cet algorithme.

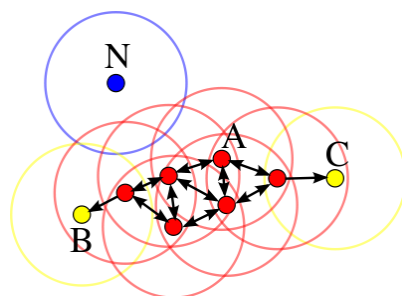


FIGURE 3. Représentation du fonctionnement typique de DB SCAN

La méthode DB SCAN a été choisie, car elle ne nécessite pas de connaître le nombre attendu de clusters, et elle n'oblige pas à associer à un cluster les valeurs isolées. Cependant, elle nécessite de définir manuellement les valeurs ϵ et MinPts. Dans notre étude, la valeur d' ϵ correspond à la valeur de l'imprécision du gps (10 mètres). Pour un mois de données, les experts métiers considèrent que 3 arrêts distincts au même endroit constituent un arrêt fréquent. Aussi, le nombre MinPts a été fixé à 3.

2.2.2. Identification des lieux d'intérêts ;

L'hypothèse que seuls des arrêts fréquents peuvent être des lieux d'intérêt a été posée. 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) sont employées à cet effet. Pour les sources ouvertes, il faut définir un ensemble de lieux potentiellement intéressants. OSM est un projet qui a pour objectif de créer une base de données géographique contributive mondiale qui recense des routes, des voies ferrées, des rivières, des bâtiments... 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 (<http://wiki.openstreetmap.org/wiki/Key:amenity>), 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 taxi. Pour OSM, un ensemble d'objets géographiques pertinents peuvent être ainsi extraits et localisés à l'aide d'une liste de couples clés-valeurs prédéfinis. Des POI externes sont ainsi extraits et localisés.

Un processus d'appariement est alors lancé afin d'associer les lieux d'arrêts fréquents à des POI externes. (Beeri et al. 2004) décrit un grand nombre de

processus d'appariements de points issus de deux jeux de données différents. Actuellement, un processus simple est employé. Il associe chaque centre de cluster à un POI. Le calcul du centre de cluster est une moyenne pondérée des points le constituant. Un seuil maximum de distance D_{max} est défini. Si le POI externe le plus proche est à une distance supérieure à ce seuil, alors le lieu n'est pas apparié à un POI Externe. A l'inverse, plusieurs lieux d'arrêts fréquents peuvent être associés à un même POI Externe. Les véhicules pouvant se garer à différents endroits à proximité d'un POI, cette propriété est cohérente. La figure 4 représente le résultat de l'appariement entre des arrêts récurrents (cercles bleus) et les POI externes (étoiles rouges). Les traits noirs définissent les appariements.

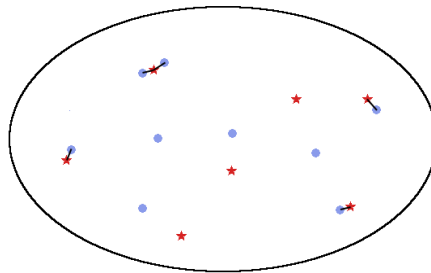


FIGURE 4. Appariement des arrêts et des données externes

Une fois apparié, les propriétés du POI externe sont associées au lieu d'arrêts fréquents pour définir un lieu d'intérêt.

2.3. Montée en sémantique sur les trajectoires :

La dernière étape du processus consiste à qualifier la trajectoire du véhicule, à savoir, qu'elle est l'activité du véhicule lorsqu'il se déplace d'un POI A à un POI B.

2.3.1. Problèmes et solutions classiques rencontrés dans la littérature ;

Dans la plupart des représentations sémantiques de trajectoires, les activités sont associées uniquement aux arrêts. Pour les déplacements, la seule information sémantique est le mode de déplacements. Par exemple, la figure 5 décrit sémantiquement la trajectoire d'un touriste à l'aide du modèle de Parent et al. (2013).

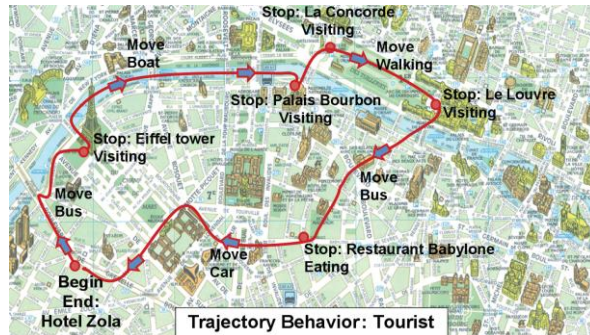


FIGURE 5. *Trajectoire sémantique d'un touriste*

Pour des véhicules spécifiques, l'objectif du déplacement est une information riche sémantiquement qui se trouve être manquante.

2.3.2. *Processus d'enrichissement sémantique des trajectoires ;*

A partir de descriptions sémantiques des lieux de départ et d'arrivée et des valeurs des capteurs, il est souvent possible de déduire l'objectif de chaque déplacement.

En effet si un déplacement part d'un POI de type A et arrive à un POI de type B et que durant ce déplacement les capteurs renvoient des valeurs significatives (Compteur en marche sur un taxi, gyrophare allumé sur une ambulance...) alors nous pouvons en déduire que l'objectif du déplacement est connu. Dans ce but, un ensemble de règles métiers est défini tel que proposent Browne (2009) et Ray (2013), à savoir, quand les conditions sont réunies, alors un cas précis est appliqué. Le modèle utilisé ne contient pas de règles contradictoires.

Par exemple, si un véhicule de secours quitte un lieu quelconque pour rejoindre un lieu de type hôpital et que durant ce trajet un capteur renvoie la valeur gyrophare allumé, une règle métier infère que ce déplacement est un déplacement d'un patient vers un hôpital.

De même, si un déplacement part du lieu classique de stationnement vers une station essence, une règle métier va qualifier ce déplacement de déplacement logistique.

Les valeurs d'un grand nombre de capteurs autorisent même l'extraction de connaissances plus précises. Par exemple, dans le cas d'un transport de patient, si la vitesse est faible et les données des accéléromètres indiquent une conduite très souple alors la sémantique précise de cette trajectoire est : Transport patient à risque où le confort de déplacement prime sur la vitesse.

3. Expérimentation

Ce processus d'extraction a été expérimenté à l'aide des données issues d'ambulances connectées de pompiers. Les véhicules envoient chaque seconde leur position GPS, les valeurs maximales issues d'un gyroscope 3 axes, des valeurs capteurs spécifiques aux ambulances (Sirène, gyrophares, taux de remplissage des bouteilles d'oxygène...) ainsi que 30 valeurs issues du bus CAN (Controller Area Network) du véhicule : Consommation du véhicule, frein à main, accélérateur, etc.

Notre étude préalable se base sur 30 jours de données issues des 3 premières ambulances connectées (Véhicule de Secours et d'Assistance aux Victimes - VSAV) de la caserne des Sapeurs-Pompiers de Tours-centre (37). Ce jeu de données regroupe un peu moins de 2 millions de lignes de données (50 attributs) par mois par véhicule, soit plus de 5 millions au total. La figure 6 représente l'étendue de ces données sur la ville de Tours.

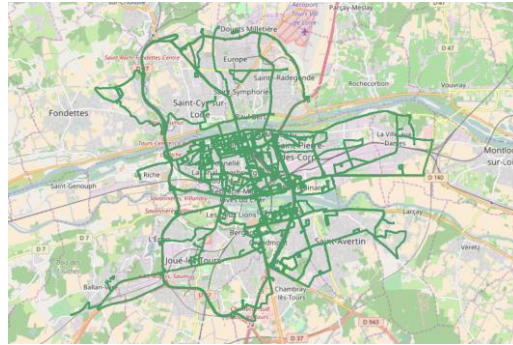


FIGURE 6. Représentation des données collectées

Les ambulances sont programmées de manière à ce qu'elles envoient toujours leurs données toutes les secondes avec une mise en mémoire tampon si le réseau n'est pas disponible, le seul cas où le transfert est interrompu est lorsque le véhicule est arrêté (contact coupé) pendant plus de 30 minutes dans une caserne, alors le transfert sera interrompu jusqu'au redémarrage du véhicule.

Chaque ligne est composée des informations suivantes :

Capteurs classiques

- *Date*
- *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

Le traitement de ces données a été réalisé en mode hors ligne. C'est-à-dire que les données ont été collectées directement sur les ambulances vers un fichier CSV, puis importées dans une base de données PostGIS indexée pour améliorer la vitesse de traitement des requêtes et des algorithmes.

3.1. Localisation des lieux d'arrêts :

L'activité la plus courte d'une ambulance selon les experts est une activité logistique (faire le plein du véhicule) durant au minimum 2 minutes.

Lour le filtrage des micro-arrêts la limite de temps a été défini à 60 secondes. Suite à l'extraction des stops, 667 arrêts et 140 micro-arrêts ont été localisés. La figure 7 représente la répartition des arrêts.

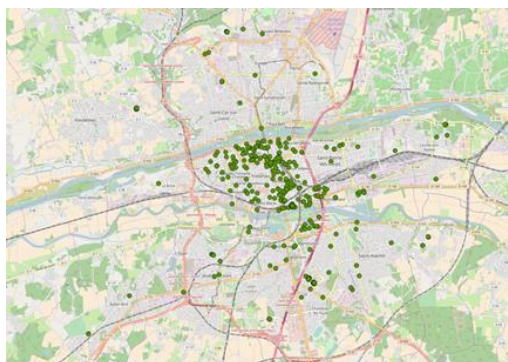


FIGURE 7. Représentation des arrêts sur une carte OSM

Par la suite, le clustering de DB SCAN a été utilisé avec les paramètres suivants, un epsilon de 10 mètres, et un minimum de points de 3 sur les données des 3 véhicules. Considérant que si sur un mois, un véhicule s'arrête plus de 3 fois dans la même zone, c'est un lieu fréquent.

Nous obtenons comme résultat le classement de 388 arrêts, répartis dans 13 clusters différents.

Le cluster numéro 1 regroupe 199 Arrêts. Il est positionné tout près du cluster n°2 qui en compte 14. Ils sont visualisés sur la figure 8(a). Le résultat montre qu'il est possible d'identifier deux zones de stationnement pour un même lieu d'intérêt (Cour avant et arrière de la caserne). Dans la figure 8 (b), le cluster regroupant 13 arrêts est affiché on y reconnaît une station-service.

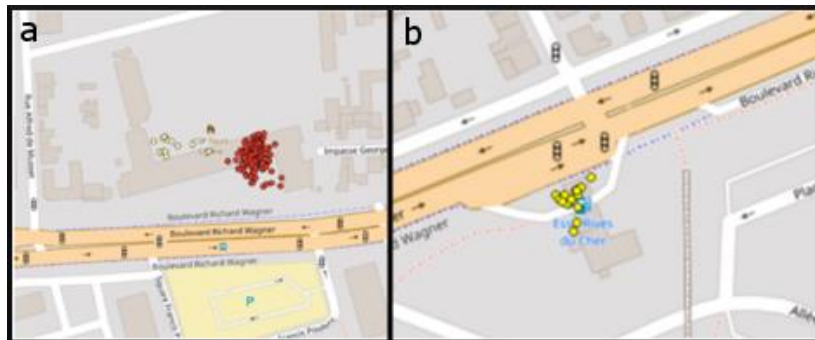


FIGURE 8. Cluster n°1 regroupant les arrêts rouges(a) , le cluster n°2 en jaune clair(a) et le cluster n°3 en jaune (b)

L'algorithme de clustering identifie 279 arrêts (plus de 40%) comme étant des outliers, c'est-à-dire qu'ils sont considérés comme non fréquents et ne remplissent pas les conditions pour former un groupe. Ce nombre important d'arrêts non fréquents est logique. Une ambulance essentiellement pour effectuer des interventions, le transport du patient (et donc l'arrêt à l'hôpital) n'est obligatoire que si son état le nécessite. Les interventions ne sont pas toujours situées au même endroit. Elles ne sont donc pas toutes regroupées dans des clusters. Pour des raisons d'anonymats, ces données ne seront pas illustrées (Clusters liés à des interventions fréquentes ou répartition des lieux d'intervention).

3.2. Localisation des lieux d'intérêts :

Afin d'identifier les lieux d'intérêts, la méthode d'appariement présentée est utilisée.

Préalablement, des POI externes sont extraits d'OSM via un outil nommé Overpass turbo, disponible depuis [<https://overpass-turbo.eu/>] suivant un ensemble de couples clé valeur comme expliquée dans la partie 2.2.2. La figure 9 représente sur une carte les données extraites de l'outil, la recherche a porté sur la clé "amenity" avec pour valeur "fire_station". Les points rouges sont les casernes de pompiers et les jaunes les centres de secours.

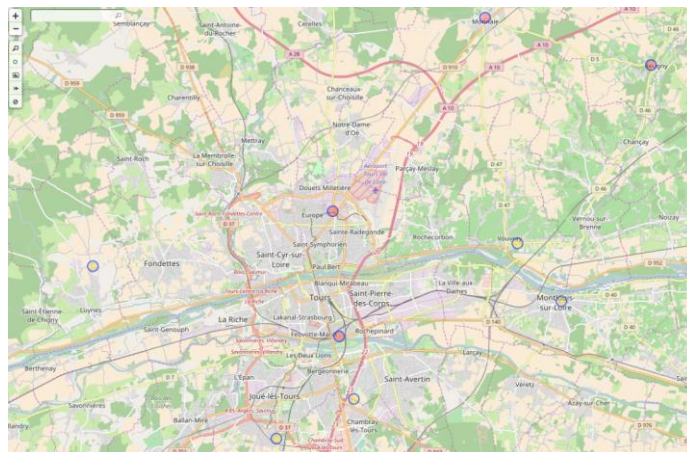


FIGURE 9. Représentation visuelle de l'extraction de données pour les casernes de pompiers issues de <https://overpass-turbo.eu/>

La figure 10 partie (a) représente le résultat de l'appariement des centres du cluster 1 en rouge et 2 en vert de la figure 8 (a). Ils sont appariés au POI le plus proche issu d'OSM (polygone jaune avec une bordure bleu) qui correspond à la Caserne des Sapeurs-Pompiers de Tours Centre. Le couple `amenity="fire_station"` a permis l'extraction de ce POI. La partie (b) représente le centre de cluster 3 en jaune apparié au POI le plus proche (polygone bleu) qui correspond à la station-service "Esso Rives du Cher". Le couple `amenity="fuel"` a permis l'extraction de ce POI.

Dans les deux cas, l'appariement est sans ambiguïté. Les centres de clusters sont inclus dans le polygone du POI trouvé le plus proche, la distance est donc nulle.

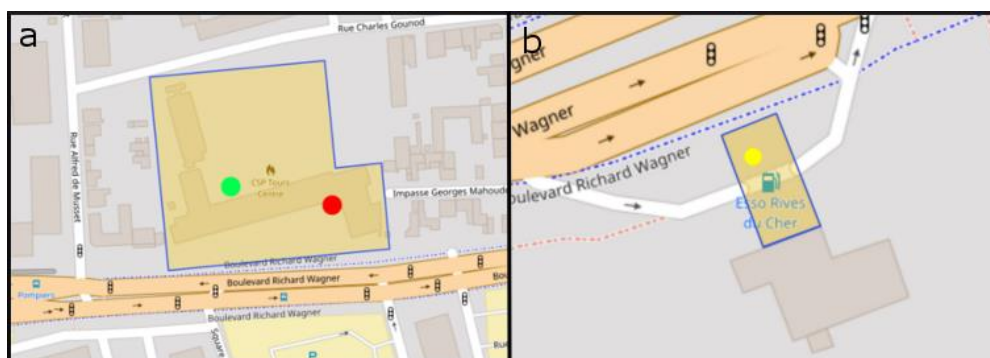


FIGURE 10. Représentation de l'appariement des centres du cluster 1 et 2 et d'une caserne de pompiers (a) et du centre du cluster 3 avec une station-service (b)

Néanmoins, il n'est pas possible d'associer automatiquement un cluster à des lieux informels. Par exemple, pour les ambulances, les lieux associés à l'économie informelle, les activités illégales (vente de drogue et de cigarette au marché noire, prostitution) ainsi que les zones accidentogènes ou de précarité sociale sont localisées, car fréquentes. Ces dernières génèrent en effet des interventions récurrentes aux mêmes lieux. Ces arrêts ne peuvent être identifiés automatiquement s'ils ne sont pas présents dans les sources externes.

3.3. Règles métiers

Des règles métiers doivent maintenant être définies afin d'identifier les objectifs des déplacements. Actuellement les objectifs et les règles associées suivants ont été identifiés :

- Départ en intervention : Trajectoire d'un véhicule avec les gyrophares de la caserne jusqu'à un arrêt.
- Retour à la caserne : Trajectoire d'un véhicule avec pour lieu d'arrêt retour la caserne.
- Transport de patient : Trajectoire d'un véhicule ou le lieu d'arrivée est de type hôpital et ou le gyrophare est globalement allumé pendant le trajet.
- Déplacement logistique Trajectoire d'un véhicule ou le lieu de départ ou d'arrivée est un POI de type logistique.

Un affinage de ces objectifs et donc des règles métiers est dans un deuxième temps possible comme il a été proposé dans la section 2.3.2. De même, des interventions très proches de POI associé à d'autres types d'activités peuvent être discriminées avec l'apport des capteurs du véhicule. Par exemple, une intervention à proximité d'une station-service est identifiée à l'aide de sa durée anormalement longue et des gyrophares allumés.

3.4. Montée en sémantique des trajectoires

Une fois les règles métiers définis et les objectifs identifiés, il est possible d'étiqueter les trajectoires. La figure 11 présente les trajectoires étiquetées comme étant des trajectoires de retour à la caserne (point noir sur la carte) à partir d'un hôpital (points rouges sur la carte). On remarque que les chemins empruntés pour les retours des hôpitaux sont quasiment identiques. On identifie également quelques trajets anormaux, en tirets sur la carte (axe principal coupé, détour volontaire ...)

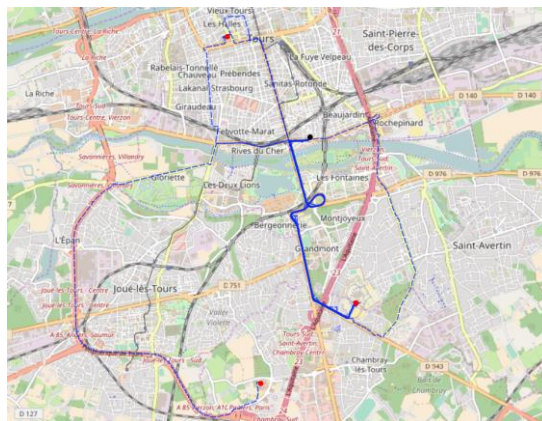


FIGURE 11. Représentation des trajectoires étiquetées retour à la caserne après passage à l'hôpital

4. Conclusion

Dans cet article a été présenté un processus d'enrichissement sémantique de trajectoires. Ce dernier s'appuie sur gros volumes de traces brutes GPS et de valeurs issues d'autres capteurs.

Pour des véhicules spécifiques, l'identification des lieux d'arrêts fréquents et la détection des objectifs des déplacements sont rendues possibles. Une première expérimentation a permis de valider l'approche pour l'extraction et l'identification des lieux d'intérêt fréquent pour les véhicules connectés d'une caserne de pompiers.

Un premier ensemble de règles métiers simples a été défini. Ces règles ont permis d'extraire automatiquement les objectifs des déplacements. La solution proposée est suffisamment générique pour être employée pour d'autres types de véhicules (Taxi, Ramassage d'ordures...) ou d'autres zones d'activités. Cette méthode démontre l'intérêt de disposer de multiples capteurs afin de compléter ou de corriger les informations issues d'un capteur GPS. 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, la montée en sémantique 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. 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.)

A moyen terme, la connaissance extraite doit pouvoir être mise à jour. Pour cela une méthode de clustering plus avancée doit être employée, par exemple Incremental DBSCAN proposée par (Ester,1998). Elle prendra en compte la date des arrêts, en plus de leur fréquence. Des lieux qui étaient fréquents pourront ainsi disparaître s'ils ne sont plus visités. De même, de nouveaux lieux apparaîtront.

A 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.

Ce travail a été financé et soutenu par Petit Picot by GRUAU, ainsi que l'ANRT (Association Nationale de la Recherche et de la Technologie).

Bibliographie et références

- Beeri, C., Kanza, Y., Safra, E., & Sagiv, Y. (2004) Object fusion in geographic information systems. *In Proceedings of the Thirtieth international conference on Very large databases-Volume 30* (pp. 816-827). VLDB Endowment.
- Browne, P. (2009). *JBoss Drools business rules*. Packt Publishing Ltd.
- 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
- Duran, A., Earleywine, M. (2012). GPS Data Filtration Method for Drive Cycle Analysis Application. *In SAE 2012 World Congress, Detroit, Michigan*
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X.(1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *In Kdd (Vol. 96, No. 34, pp. 226-231)*.
- Ester, M., Kriegel, H. P., Sander, J., Wimmer, M., & Xu, X. (1998). Incremental clustering for mining in a data warehousing environment. *In VLDB (Vol. 98, pp. 323-333)*.
- Etienne, L., Devogele, T., & Bouju, A. (2012). Spatio-temporal trajectory analysis of mobile objects following the same itinerary. *Advances in Geo-Spatial Information Science, 10, 47-57*.
- Knight et al. (2009). A comparison of outlier detection procedures and robust estimation methods in GPS positioning. *J. Geodesy 62* (4), 699-709
- Ivanović, S. S., Raimond, A. M. O., Mustière, S., & Devogele, T. (2016) Detection of outliers in crowdsourced GPS trace, *12th international symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 6 pages*
- Ordoñez, 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
- Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., ... & Theodoridis, Y. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR), 45*(4), 42.

Ray, C., Grancher, A., Thibaud, R., & Etienne, L. (2013). Spatio-temporal rule-based analysis of maritime traffic. In *Conference on Ocean & Coastal Observation (OCOSS 2013)*.

Small, C.A. : Survey of Multidimensional Medians. *International Statistical Review* 58(3), 263–277 (1990)

Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 29.