

Introduction

De nos jours, les villes sont confrontées à de multiples enjeux tels que l'urbanisation croissante, la pollution ou la congestion. Une étape fondamentale pour faire face à ces défis consiste à caractériser avec précision la façon dont les gens se déplacent dans la ville. Cette connaissance fondamentale peut être utilisée pour améliorer et optimiser le système de transport urbain.

Jusqu'à aujourd'hui, les autorités en charge du transport utilisent des enquêtes de déplacements pour recueillir des informations sur la mobilité des populations. Toutefois, ces méthodes présentent de nombreux inconvénients. Les enquêtes sont coûteuses à réaliser, deviennent rapidement obsolètes, sont inévitablement basées sur des échantillons relativement petits de la population et ne peuvent pas capturer de manière dynamique la mobilité des personnes. Au cours des dernières décennies, avec la digitalisation de la société, l'émergence de nouvelles sources de données telles que les données billettiques, les données GPS, les données géolocalisées issues des réseaux sociaux ou encore les données de téléphonie mobile a attiré l'attention des chercheurs et des acteurs opérationnels pour l'estimation de la mobilité. Les chercheurs ont commencé à étudier l'utilisation de ces données dans le domaine du transport et particulièrement de la mobilité, où elles pourraient permettre des analyses à une échelle sans précédent (tant au niveau spatial que temporel) par rapport aux enquêtes traditionnelles. Parmi les sources de données émergentes, les données de téléphonie mobile sont l'une des sources les plus prometteuses. Ce type de données présente en effet une combinaison unique de propriétés souhaitables : (i) elles offrent un taux de pénétration sans précédent, car disponibles pour l'ensemble des abonnés d'un fournisseur de réseau ; (ii) elles sont enregistrées en continu sur de longues périodes, ce qui permet des études longitudinales fines sur des mois voir des années~; et (iii) elles sont déjà collectées passivement par l'opérateur à des fins de facturation, ce qui en fait une source de données peu coûteuse pour une utilisation ultérieure. Cependant, il n'existe que peu de connaissances sur l'utilisation des données de téléphonie mobile à large échelle pour analyser la mobilité des populations en milieu urbain. Cette thèse vise à apporter des contributions à ce sujet.

Dans cette thèse, nous montrons qu'en dépit des limitations bien connus des données de téléphonie mobile (données éparpillées dans le temps, ayant une large incertitude spatiale et soumises à des phénomènes d'oscillation récurrents), ces données contiennent de riches informations spatio-temporelles qui peuvent être utilisées pour diverses applications : la construction de matrices Origine Destination, l'analyse de densité de population ou encore l'inférence des chemins populaires du réseau de transport. Par ailleurs, nous présentons une nouvelle approche TRANSIT (TRAjectory inference from Network Signaling daTa) qui porte sur l'analyse de données de téléphonie mobile à grande échelle. D'une part, notre approche est capable de distinguer les sessions mobiles des sessions statiques pour un utilisateur donné. Cela permet de reconstruire la demande de déplacement à une échelle spatio-temporelle fine. D'autre part, TRANSIT est capable de réduire fortement l'erreur spatiale des trajectoires de téléphonie mobile. Ainsi, notre approche arrive à surmonter les principales limitations déjà mentionnées. TRANSIT exploite la récurrence de la mobilité humaine : le fait qu'un individu effectuant des déplacements réguliers entre deux zones va généralement prendre toujours le même chemin. Cela crée une redondance dans les informations de mobilité que notre approche TRANSIT utilise pour augmenter la précision spatio-temporelle des trajectoires. Afin d'exploiter pleinement le potentiel des données de téléphonie mobile, nous avons également développé une approche de map-matching qui peut être couplée à TRANSIT. Cette dernière permet de retrouver le chemin suivi par un utilisateur dans le système de transport multimodal à partir des trajectoires reconstruites par TRANSIT. Pour démontrer le potentiel des approches développées dans cette thèse, nous avons construit

quelques applications qui sont rendus possible par nos travaux : l'analyse de la mobilité lors d'évènements atypiques à une échelle spatio-temporelle fine, l'analyse des trajectoires des véhicules roulant sur le périphérique de Paris ou encore un modèle épidémiologique pour l'étude de la propagation du COVID-19. En conclusion, les résultats de cette thèse démontrent que l'analyse de la mobilité urbaine à large échelle est possible avec les données de téléphonie mobile. Cela ouvre la voie à de nouvelles applications qui pourraient être étudiées par la suite. Aussi, afin de rendre les données des téléphones mobiles totalement opérationnelles, deux problématiques principales doivent être traitées : le respect de la vie privé des utilisateurs et les biais que peuvent contenir les résultats issus des données des téléphones mobiles.

Résumé chapitre 2

Ce chapitre présente une approche permettant de traiter les données de signalisation afin d'estimer avec précision les profils de demande temporels des déplacements dans une ville et d'en extraire les profils récurrents. Cette approche repose sur des étapes préliminaires de prétraitement et de filtrage des données afin de ne conserver que les données utiles à l'extraction d'informations pertinentes sur la mobilité. Ensuite, en analysant les données de signalisation de 2 millions d'utilisateurs (se déplaçant avec leurs téléphones portables), nous montrons que les données de signalisation sont capables d'extraire de manière robuste les déplacements des résidents et d'estimer la distribution horaire des déplacements dans la région étudiée. Ceci sous réserve que les biais spatio-temporels des données de signalisation soient correctement détectés et supprimés. Ensuite, grâce à une approche de clustering des profils temporels de demande émise par chaque zone et en utilisant les données officielles d'occupation des sols, nous sommes capables d'extraire et interpréter de riches « patterns » de mobilité sur la zone étudiée. Dans cette étude, les résultats ont été obtenus en explorant les données de signalisation couvrant un large territoire d'environ 44.000km² comprenant différents profils de zones socio-démographiques et économiques.

L'évaluation effectuée sur les dimensions temporelles et spatiales montre que les profils de demande obtenus avec les données de signalisation sont fortement corrélés à ceux obtenus à partir des données d'enquête sur les voyages, avec des coefficients de corrélation supérieurs à 0,9. Cela confirme que les données de signalisation peuvent être exploitées efficacement comme un bon proxy pour les estimations de la mobilité de la population. Ainsi, ces données devraient être reconnues comme une source de données précieuse et fiable pour estimer la mobilité à large échelle, en particulier dans le cas de territoires où des données précises sur la mobilité ne sont pas disponibles ou sont difficiles à collecter par le biais d'enquêtes ou de sondes et capteurs de trafic dédiés.

De plus, nous avons pu identifier des corrélations significatives entre les modèles dynamiques basés sur les téléphones mobiles et les données dites « land-use ». Les zones urbaines très denses sont caractérisées par un pic de demande élevé l'après-midi, tandis que les zones à faible densité présentent un pic plutôt élevé le matin. Par ailleurs, dans les zones rurales et mixtes/suburbaines, les données de signalisation présentent des flux de déplacements nettement plus élevés et des patterns de mobilité plus raisonnables que les données d'enquête.

Résumé chapitre 3

Dans ce chapitre, nous avons présenté TRANSIT, une approche permettant de classer les phases mobiles et statiques d'une trace de téléphonie mobile correspondant au déplacement en milieu urbain d'une personne ayant un téléphone portable. De plus, l'approche permet de reconstruire des trajectoires de mobilité humaine individuelle à une échelle de granularité spatio-temporelle fine. TRANSIT améliore l'état de l'art en matière d'inférence de trajectoire de mobilité centrée sur l'humain en tirant parti d'une heuristique dédiée, de la consolidation de la localisation de l'activité statique et de l'amélioration de la trajectoire via le regroupement spatial pour : (a) extraire des informations utiles en tirant profit du taux d'échantillonnage plus élevé auquel les événements de communication sont collectés dans NSD comparé aux autres types de données de téléphonie mobile; (b) effectuer une détection et une suppression efficaces des oscillations ; (c) capturer la nature répétitive de la mobilité individuelle dans la dimension temporelle. Cette combinaison de caractéristiques uniques permet d'obtenir une meilleure classification des sessions mobiles et statiques, ainsi qu'une précision accrue des trajectoires reconstruites.

Nous validons TRANSIT à l'aide de trajectoires GPS considérées comme vérité terrain et collectées par un petit groupe de volontaires, et montrons que notre approche atteint une précision de 80% et un rappel de 96% dans l'identification des périodes de mouvement, ainsi qu'une précision spatiale moyenne de 190 m dans l'estimation des trajectoires. Les comparaisons avec des outils antérieurs pour effectuer les mêmes traitements que TRANSIT à partir de données de téléphones mobiles montrent également des gains de l'ordre de 50-70%.

Nous avons appliqué TRANSIT à large échelle, c'est à dire à l'ensemble de la base d'abonnés d'un grand opérateur de réseau dans deux grandes villes de France, Paris et Lyon. Cela permet de valider TRANSIT à grande échelle en montrant que le profil de demande qu'il permet de reconstruire atteint une corrélation de Pearson de 0,89\$ avec ceux obtenus grâce aux. Un tel résultat permet à TRANSIT de surpasser les approches de l'état de l'art, y compris celle étudiée au chapitre 2. De plus, sur le jeu de données à grande échelle, TRANSIT est capable d'identifier 480 millions de trajectoires de plus de 10 millions d'individus pendant une période de trois mois en 2019 et ainsi d'améliorer considérablement la précision de 100 millions d'entre elles.

Nous tirons parti des résultats agrégés de TRANSIT pour mener des explorations préliminaires comme base de futures applications rendues possible par TRANSIT : (a) l'analyse de la fraction des déplacements utilisant les transports publics par rapport aux autres modes de transport, (b) l'analyse des trajets domicile-travail à l'échelle métropolitaine. À notre connaissance, nous sommes les premiers à utiliser la donnée de signalisation pour effectuer une analyse de la mobilité à une si grande échelle.

Résumé chapitre 4

Dans ce chapitre, nous avons étudié le potentiel des données de signalisation du réseau pour fournir des informations spatio-temporelles à une granularité fine afin de reconstituer la mobilité des utilisateurs. Ainsi, nous avons développé un algorithme de map-matching basé sur les chaînes de Markov pour mettre en correspondance les trajectoires cellulaires éparses et bruitées avec le réseau de transport multimodal sous-jacent. L'approche de map-matching est appliquée à la sortie de TRANSIT (le cadre a déjà été présenté en détail dans le chapitre 3).

Pour valider notre approche, nous avons réalisé une étude de cas, relative à la ville de Lyon, en exploitant à la fois des traces de signalisation réelles collectées par un opérateur de téléphonie mobile (Orange) et des données GPS collectées via une application de téléphone mobile. Ces données ont été utilisées pour effectuer une validation microscopique, visant à la fois à déterminer les paramètres optimaux de l'étape de map-matching basée sur les chaînes de Markov et à montrer la capacité de notre approche à map-matcher avec précision les trajectoires de signalisation sur un réseau de transport composé de plusieurs modes de transport.

Nous avons également démontré l'importance d'avoir une connaissance a priori du mode de transport avant d'appliquer le processus de map-matching pour améliorer la performance de ce dernier.

Ensuite, à l'aide d'une simple inférence de mode de transport, nous avons démontré la possibilité de récupérer les chemins populaires par mode de transport pour des paires OD multiples. Nous soulignons le fait que, en s'appuyant sur notre approche et sur les données de signalisation du réseau, une telle connaissance peut être fournie à très grande échelle (un pays entier), avec une description temporelle (les chemins populaires peuvent être différents à des moments donnés de la journée ou pendant les week-ends), et à une résolution spatiale beaucoup plus élevée (couvrant également les zones périphériques, ou les régions rarement observées via les données GPS) que celle fournie par les simples planificateurs d'itinéraires traditionnels, prouvant ainsi l'utilité de notre approche et l'intérêt de l'étude de cas analysée.

Les directions de travaux futurs qui ne sont pas abordées dans cette thèse incluent des améliorations sur les paramètres des HMM. Pour la probabilité initiale, au lieu d'avoir tous les états équiprobables, la probabilité pourrait être pondérée par le niveau de confiance qu'un utilisateur est dans un mode de transport particulier. Ce niveau de confiance pourrait être estimé sur la base des informations relatives au trajet (temps de trajet, vitesse, etc.). En outre, la matrice de transition pourrait être dynamique et estimée sur la base du temps de trajet en fonction des conditions réelles de circulation.

D'autres directions de travail incluent l'amélioration de la technique d'inférence du mode de transport. En effet, au lieu d'utiliser uniquement la vitesse, de multiples caractéristiques pourraient être utilisées pour l'inférence, telles que : la probabilité renvoyée par l'algorithme de Viterbi, l'heure de début/durée du trajet et la durée totale de l'activité statique au cours du trajet. Une analyse détaillée des résultats découlant d'une application de notre solution à l'échelle d'un pays pourrait être explorée dans les orientations futures.

Résumé chapitre 5

Dans ce chapitre, nous avons étudié le problème de l'extraction des comportements récurrents et structurants de mobilité à l'échelle agrégée. Nous avons étudié le cas la mobilité quotidienne dans la ville de Lyon. Nous modélisons la mobilité au sein de la ville par une représentation mathématique agnostique : un tenseur dit de mobilité. Ce tenseur permet de représenter la mobilité quotidienne à une échelle agrégée selon plusieurs dimensions. Ici, nous avons sélectionné trois dimensions : le temps, l'origine et la destination. Grâce à l'approche TRANSIT du chapitre 3, nous sommes en mesure de caractériser ce tenseur à une granularité spatio-temporelle fine. En particulier, la segmentation spatiale utilisée est une grille spatiale composée de 625 carrés, chaque carré ayant la dimension de 800m par 800m. Une telle segmentation basée sur les carrés est significativement plus fine que la segmentation spatiale administrative habituelle (comme celle utilisée dans le chapitre 2).

Pour extraire la dynamique urbaine de la ville à partir du tenseur de mobilité quotidienne, nous nous appuyons sur une approche de décomposition non négative de Tucker capable de décomposer le tenseur de mobilité en un tenseur plus petit avec des dimensions prédéterminées (le tenseur de base, en anglais « core tensor »), multiplié par des matrices de facteurs (en anglais « factor matrices »). Les matrices de facteurs sont capables de capturer la dynamique urbaine latente de chaque dimension (origine, destination et temps) tandis que le tenseur central saisit les interactions multi-dimensionnelle entre les matrices de facteurs.

L'approche mise en œuvre consiste en une décomposition de Tucker non négative qui est couplée à des approches de régularisation. La méthode d'optimisation utilisée est le gradient proximal alternatif. Les résultats contenus dans ce chapitre permettent d'atténuer le rôle de la régularisation et de l'intégration de facteurs urbains pour améliorer la décomposition par rapport au résultat de la littérature. En outre, nous démontrons la capacité de l'approche en conjonction avec le de mobilité fine calculé à partir des sorties TRANSIT de capturer la dynamique urbaine latente de la ville. Enfin, nous effectuons une analyse comparative de l'état de l'art des approches de décomposition tensorielle. Bien que l'erreur RMSE soit plus élevée par rapport aux autres méthodes, l'approche étudiée comporte beaucoup moins de paramètres. Nous avons également montré que l'approche est robuste à l'échantillonnage du tenseur de mobilité quotidienne.

Des travaux futurs pourraient explorer la possibilité d'ajouter plus de dimensions dans le tenseur de mobilité, comme le mode de transport, afin d'améliorer la qualité des patterns de mobilité urbain extraits. De plus, le tenseur de mobilité pourrait être construit à partir de sources de données multiples en utilisant une approche de fusion de données appropriée. Cela permettrait d'obtenir une représentation plus précise de la mobilité urbaine.

Résumé chapitre 6

Dans la littérature, les données des téléphones mobiles ont été exploitées pour un grand nombre d'applications telles que la modélisation des lois physiques qui régissent la mobilité humaine à l'échelle individuelle et agrégée, l'estimation de la demande à l'échelle du réseau ou l'estimation de la densité de population (recensement de population).

En surmontant les principales limitations des trajectoires des téléphones portables, à savoir l'incertitude spatiale, l'hétérogénéité temporelle et les phénomènes d'oscillation, nous pouvons fortement améliorer l'usage qui peut être fait des données des téléphones portables. Pour démontrer ce potentiel, nous développons, dans ce chapitre, plusieurs applications. Ces dernières comprennent l'analyse fine des émissions/attractions en déplacement lors d'événements anormaux, l'analyse des trajectoires individuelles des véhicules passant par certaines sections du périphérique de Paris et un modèle épidémiologique basé sur la mobilité pour modéliser la propagation du COVID-19. Les approches développées ouvrent la voie à de nouvelles applications qui rendent possible l'analyse de la mobilité humaine à une granularité spatio-temporelle fine et dans des contextes multimodaux. Ces applications pourraient être développées dans des travaux futurs et les directions futures d'amélioration de notre propre travail ont été discutées dans les différents chapitres de cette thèse.

En plus des nombreuses opportunités qu'elles amènent, les approches basées sur les données de signalisation sont confrontées à de nombreux défis. D'une part, les estimations produites par l'analyse des données de signalisation comportent des biais spatiaux et temporels qu'il faut être en mesure d'identifier voire supprimer pour obtenir des résultats pertinents. L'identification et la suppression de ces biais est un sujet peu abordé dans la littérature scientifique mais nécessaire pour que ces données puissent être considérées comme fiables d'un point opérationnel. D'autre part, il est également nécessaire de développer des approches d'anonymisation de l'ensemble des données de signalisation (ainsi que le résultat de leur agrégation). Cela permettrait de protéger la vie privée de l'utilisateur et de faciliter le partage de ces données au sein des communautés de chercheurs et de praticiens.

Conclusion

Cette thèse a permis le développement d'approches et méthodologie dans le domaine de l'estimation de la mobilité urbaine à grande échelle. Dans ce cadre elle a permis de répondre à plusieurs questions de recherche.

1^{ère} question de recherche : dans quelle mesure les données de signalisation sont adaptées pour l'analyse de la mobilité ?

Sur la base des travaux de Fekih et al. Portant sur les données de signalisation de 2 millions d'utilisateurs de téléphones portables, nous montrons que les données de signalisation est une source robuste pour extraire les déplacements des résidents et estimer la distribution horaire des déplacements dans toute la région lyonnaise, à condition que les biais spatio-temporels de ces données soient correctement détectés et supprimés. Après notre procédure de débiaisage, les profils de demande obtenus avec les données de signalisation sont fortement corrélés avec ceux obtenus avec les enquêtes de déplacement (considéré comme vérité terrain. Ensuite, en couplant une approche de clustering hiérarchique appliquée aux profils de demande de déplacements avec des données dites « land-use », nous sommes capables d'extraire de la connaissance riche sur les comportements récurrents de mobilité à l'échelle d'une ville.

2^{ème} question de recherche : la nature répétitive de la mobilité humaine peut-elle être utilisée pour améliorer dans l'espace et dans le temps les trajectoires humaines telles qu'observées par le biais des données de téléphonie mobile ?

La nature répétitive de la mobilité humaine a été largement démontrée dans la littérature. Sur la base de cette propriété et en tirant parti de l'effet d'oscillation comme de la triangulation, nous concevons TRANSIT, une approche capable d'améliorer la granularité temporelle des trajectoires des téléphones mobiles et de réduire l'incertitude spatiale. TRANSIT surpasse les autres approches de la littérature sur un ensemble de données de vérité terrain contenant à la fois des données de téléphone mobile et de GPS pour un ensemble de volontaires dans la région métropolitaine de Lyon. TRANSIT atteint une précision spatiale moyenne de 190m. Outre ces résultats satisfaisants, TRANSIT peut également être utilisé pour l'estimation de la mobilité humaine à grande échelle.

3^{ème} question de recherche : comment estimer des informations de mobilité à une granularité fine comme le chemin parcouru sur un réseau de transport multimodal, à partir des trajectoires des téléphones portables ?

Pour répondre à cette question, nous avons d'abord étudié le problème de map-matching des traces de téléphonie mobile avec le réseau de transport. Pour mettre en correspondance les trajectoires des téléphones mobiles avec le réseau de transport multimodal, nous développons une approche basée sur les chaînes de Markov qui permet d'obtenir une forte précision spatiale. Cette performance peut être améliorée de manière significative en utilisant TRANSIT avant d'appliquer le map-matching. Nous utilisons les approches utilisées dans le cadre d'applications comme l'inférence de trajets populaires par mode de transport dans la ville de Lyon pour des paires d'origines/destinations données.

4^{ème} question de recherche : comment pouvons-nous inférer les comportements récurrents de mobilité au niveau agrégé selon les principales dimensions qui caractérisent la mobilité humaine ?

Dans le cadre de ce chapitre : nous envisageons d'analyser la mobilité en trois dimensions : origine des déplacements (espace), destination des déplacements (espace) et date de départ du déplacement (temps). Nous construisons un tenseur de mobilité quotidienne visant à capturer les flux selon ces trois dimensions dans la ville de Lyon. Le tenseur de mobilité a été déduit en s'appuyant sur l'agrégation dans l'espace et dans le temps des résultats de TRANSIT. La faible erreur spatiale de TRANSIT nous permet d'utiliser une segmentation spatiale fine pour les dimensions origine et destination : la région lyonnaise a été divisée en 625 zones, chaque zone étant un carré de 800 m de côté. En se basant sur des approches consolidées de la littérature, nous appliquons une décomposition de Tucker au tenseur de mobilité quotidienne. Nous avons ensuite analysé la décomposition résultante et montré que l'approche était capable d'extraire des « patterns » temporels et spatiaux (origine et destination) à une granularité fine. L'approche permet également de capturer des dépendances spatio-temporelles complexes entre ces patterns.

5^{ème} question de recherche : quels types d'applications sont rendus possibles par notre approche lorsqu'elle est appliquée aux données des téléphones mobiles ?

Les approches développées dans cette thèse décuplent le potentiel des données de signalisation. D'une part, nous développons TRANSIT capable de réaliser la segmentation de trajectoire ainsi que l'augmentation spatio-temporelle des trajectoires mieux que les approches actuelles de l'état de l'art. Ce dernier point permet d'améliorer les études/applications déjà existantes liées à l'inférence de la demande de déplacement à grande échelle dans un environnement urbain avec des données de téléphone mobile. De plus, elle permet de nouveaux types d'études de mobilité, y compris l'analyse fine des trajectoires, qui étaient auparavant réservées aux données GPS. En outre, notre approche map-matching multimodal ouvre la voie à l'analyse de la multimodalité à grande échelle et en milieu urbain à l'aide de données de téléphonie mobile. Enfin, nous tirons parti des approches développées dans cette thèse pour alimenter une étude de cas qui consiste en un modèle épidémiologique pour étudier la propagation du COVID-19.