

Université KASDI Merbah - Ouargla

Faculté des Nouvelles Technologies de l'Information et de la Communication

Département d'Informatique et de la Technologie de l'Information



Mémoire présenté pour l'obtention du diplôme de Master

Spécialité : Administration et Sécurité de Réseaux

Présenté par :

Khoukhou Boutaina

**La considération du facteur temporel pour
l'algorithme CPT pour la prédiction et la
recommandation**

Soutenu le : 20/06/2021

Devant le jury composé de :

Dr Amirat Hanane

Encadreur

UKM Ouargla

Dr Harrouz Abdelhakim

Examineur

UKM Ouargla

Dr Benmir Abdelkader

Examineur

UKM Ouargla

L'année universitaire : 2020/2021

Remerciement

On remercie dieu tout puissant de nous avoir donné la santé et la volonté et de terminer ce mémoire.

*Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Dr Amirat Hanane**, on le remercie pour la qualité de son encadrement exceptionnel,, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.*

Nous remercions en particulièrement nos parent qui ont su nous soutenir, nous encourager, nous aider et nous supporter tout au long pertinentes.

Résumé

La prédiction de la localisation future des objets en mouvement est devenue récemment un sujet de recherche important en raison du besoin humain de mouvement continu, ou était appliquée dans plusieurs domaines, dont le plus important est les services basés sur la localisation. Il existe de nombreux modèles pour prédire le lieu, mais la plupart d'entre eux se concentrent sur le contexte spatial et négligent le reste des contextes, bien que cela affecte négativement la précision de la prédiction.

Dans ce mémoire, nous proposons un modèle de prédiction de localisation appelé TCPT. Ce dernier est basé sur un le modèle (Compact Prediction Tree) qui se caractérise comme non perdant car il préserve toutes les données pour faire la prédiction. Il est également tolérant au bruit contrairement aux autres modèles de prédiction. De plus, notre modèle intègre le contexte temporel et l'utilise dans le processus de prévision. Nos résultats expérimentaux sur un jeu de données réel montrent que le modèle TCPT présente des bons résultats par rapport à deux méthodes de prédiction.

Mots Clés : Prédiction de mobilité, Compact Prediction Tree, Localisation futur, Contexte temporel.

Abstract

Predicting the future location of moving objects has become an important topic of research recently due to the human need for continuous movement, as it has been applied in several areas, the most important of which is location-based services. There are many models for predicting the place, but most of them focus on the spatial context and neglect the rest of the contexts, although this negatively affects the accuracy of the prediction.

In this thesis, we propose a location prediction model called TCPT. The latter is based on a new and accurate model (Compact Prediction Tree) which is characterized as lossless as it preserves all data to make prediction. It is also tolerant of noise unlike other forecasting models. In addition, our model incorporates time context and uses it in the forecasting process. Our experimental results on a realistic dataset show that the TCPT model performed well compared to two other prediction methods.

Keywords: Mobility prediction, Compact Prediction Tree, future location, Temporal context.

المخلص

أصبح التنبؤ بالموقع المستقبلي للأشياء المتحركة موضوع بحث مهم مؤخرا نظرا لحاجة الإنسان للحركة المستمرة حيث تم تطبيقه في عدة مجالات من أهمها الخدمات المعتمدة على الموقع. يوجد الكثير من النماذج للتنبؤ بالمكان لكن أغلبها تركز على السياق المكاني وتهمل باقي السياقات على الرغم من أن هذا يؤثر سلبا على دقة التنبؤ .

في هذه الأطروحة نقترح نموذج للتنبؤ بالموقع يسمى TCPT . يعتمد هذا الأخير على نموذج جديد ودقيق (شجرة التنبؤ المدمجة) والذي يتميز بأنه غير خاسر حيث يحافظ على جميع البيانات للقيام بالتنبؤ. كما أنه متسامح مع الضوضاء على عكس نماذج التنبؤ الأخرى بالإضافة إلى ذلك فإن نموذجنا يضم السياق الزمني و يستعمله في عملية التنبؤ. تُظهر نتائجنا التجريبية على مجموعة بيانات واقعية أن نموذج TCPT قدم أداء جيدا مقارنة بأحدث أساليب التنبؤ.

الكلمات المفتاحية : التنبؤ بالتنقل, شجرة التنبؤ المدمجة, الموقع التالي, السياق الزمني.

Table des matières

Remerciement.....	I
Résume	II
Abstract	II
الملخص	III
Table des matières	IV
Liste des figures	VII
Liste des tableaux	VII
Introduction générale.....	VIII

Chapitre 1 : État de l’art

1.1 Introduction	1
1.2 Notions préliminaires	1
1.3 Prédiction de mobilité	2
1.4 Types de prédiction de mobilité	3
4.1 Prédiction de la mobilité humaine.....	3
4.1.1 Régularité de la mobilité humaine.....	3
4.1.2 La prédictibilité de la mobilité humaine.....	5
4.2 Prédiction de la mobilité des objets.....	5
1.5 L’effet de temps dans la prédiction de mobilité	6
1.6 Travaux connexes.....	6
6.1 Taxonomie des modèles de prédiction de localisation	7
6.1.1 Modèles de la fouille de données	7
6.1.2 Modèles probabilistes.....	8
6.1.3 Modèles de la correspondance des trajets (Trip Matching)	9
6.2 La considération de temps ou contexte temporel	10

6.3 Comparaison des travaux	11
1.7 Conclusion.....	13

Chapitre 2 : Le modèle TCPT pour la prédiction de localisation

2.1 Introduction	15
2.2 Notions Préliminaires.....	15
2.3 Le modèle CPT (Compact Prediction Tree).....	16
2.4 Le modèle TCPT (Temporal Compact Prediction Tree).....	18
2.5 Définitions.....	18
2.6 La prédiction de localisation avec TCPT	19
6.1 La phase d'apprentissage.....	20
6.2 La phase de prédiction.....	23
2.7 Architecture de notre système de prédiction basé sur TCPT	24
7.1 Pré-traitement	25
7.2 Prédiction de localisation	25
7.2.1 Apprentissage	25
7.2.2 Prediction	25
2.8 La considération de comportement globale et individuel de mobilité	26
2.9 Conclusion.....	27

Chapitre 3 : Expérimentations

3.1 Introduction	29
3.2 Jeu de données.....	29
3.3 Paramètres d'expérimentation.....	29
3.4 Métrique d'évaluation.....	30
3.5 Expérimentations.....	31
5.1 Modèles évalués	31
5.2 Expérimentation1 (étude de Scalabilité)	32

5.3 Expérimentation 2 (Impact de la variation de la durée de time-slot)	34
3.6 Conclusion.....	35
Conclusion générale	36
Bibliographie.....	37

Liste des figures

Figure 1. Un journal GPS, une trajectoire GPS, un point de séjour et un trajet.....	2
Figure 2. Types de prédiction de mobilité.....	3
Figure 3. Régularité de la mobilité humaine.	4
Figure 4. Exemple de prédiction de mobilité d'un objet en mouvement.....	5
Figure 5. Taxonomie des modèles de prédiction de localisation et de route.....	10
Figure 6. La construction des structures CPT.	17
Figure 7. La phase de prédiction CPT.....	18
Figure 8. Un exemple de processus d'apprentissage TCPT.....	21
Figure 9. Application des stratégies CSF & CBS	23
Figure 10. le conséquent d'une séquence de mobilité	24
Figure 11. Architecture de TCPT	26
Figure 12. Résultats de l'étude de scalabilité.	32
Figure 13. L'impact de la variation de time-slot.....	35

Liste des tableaux

Tableau 1. Littérature pour les études de prédiction de localisation	13
Tableau 2. Séquences de mobilité.	16
Tableau 3. Séquences de mobilité temporelle des trois véhicules	19
Tableau 4. Les séquences de mobilité temporelle.....	20

Introduction générale

Prédire la prochaine étape ou état d'un système est une tâche très importante dans plusieurs domaines. Elle offre un ensemble d'avantages, tels que la réduction du pourcentage de pertes et l'élimination des risques qui peuvent être rencontrés à l'avenir. Elle a été appliquée dans plusieurs domaines notamment dans la prédiction de mobilité des objets mobiles. Cette dernière consiste à prédire la prochaine localisation ou la future série de localisation pour une personne ou un objet en mouvement en fonction de sa précédente mobilité.

La prédiction de la mobilité a montré un grand intérêt dans le domaine de la prédiction des objets mobiles vue le besoin des personnes à des mouvements continus. De ce fait, de nombreuses applications, y compris la prise en charge de services basés sur la localisation (Location based services) tels que la publicité ciblée aux clients susceptibles d'approcher un point (localisation) d'intérêt, des services de météo personnelle et de divertissement, ont été émergés. La prédiction de la mobilité est également essentielle pour améliorer la qualité des systèmes de transport intelligents (Intelligent transportation Systems) pour les véhicules car elle peut être utilisée pour prédire les embouteillages et réduire les accidents et les risques de circulation.

Récemment, les prédictions de la localisation ont attiré l'attention de nombreux chercheurs, où ils s'appuyaient sur la régularité spatiale pour faire des prédictions. Ils ont considéré que le mouvement des personnes ou des véhicules est généralement *régulier*, par exemple, aller travailler tous les jours, et donc, les chemins précédents qui ont été traversés par le personne doivent être considérer et mis en correspondance avec le chemin actuel pour prédire la prochaine localisation où il peut se rendre. Cependant, les habitudes et les préférences des personnes doivent être prises en compte car chacun a ses propres habitudes (par exemple l'un préfère la pratique du sport le soir et l'autre préfère regarder un film), ce qui influe sur la prédiction.

Motivations

Dans la littérature, plusieurs technique et modèles ont été proposés pour la prédiction de localisation notamment : modèles de Markov[1], les réseaux de neurones[2], les modèles statistiques[3], les techniques de la fouille de données[4].

Bien que ces modèles aient montré de bonnes performances, ils présentent généralement des limitations importantes. Par exemple, les modèles de Markov dépendent de l'hypothèse markovienne qui stipule que le état suivant ne dépend que du état actuel(ou un petit nombre d'états) mais que la plupart des états précédents sont négligés. Ces modèles sont entraînés à l'aide de données historiques en résumant les données d'apprentissage. En conséquence, ils ignorent une grande partie des informations contenues dans les données. De plus, la plupart des systèmes de prédiction proposés prennent en compte que le contexte spatial pour faire la prédiction, négligeant le contexte temporel de mobilité. Un contexte temporel comme le jour de la semaine joue un rôle important dans la prédiction où la majorité des gens passent les jours de la semaine dans leurs postes de travail et les jours de week-end pour des localisations de loisir et distraction.

Contribution

Pour remédier les limitations de ces modèles, nous proposons dans ce mémoire un modèle de prédiction de localisation appelé TCPT. Ce modèle repose sur le modèle de prédiction CPT (Compact Prediction Tree)[5]. TCPT se caractérise par les avantages du CPT de fait qu'il n'est pas perdant (lossless model) car il utilise toutes les données d'apprentissage pour faire la prédiction. TCPT étend CPT pour inclure le contexte temporel et le considérer dans le processus de prédiction au lieu de s'appuyer uniquement sur le contexte spatial. TCPT prédit la localisation actuelle en utilisant l'historique de mobilité des objets mobiles qui sont converties en séquences spatio-temporelles où et chaque élément de la séquence représente un lieu et le temps de visite.

Les contributions du ce mémoire se résument comme suit:

1. Nous présentons TCPT comme un modèle de prédiction de localisation qui étend le modèle de prédiction sans perte (CPT) pour inclure le contexte temporel.
2. Suggérer deux modèles : personnel et global pour considérer le comportement de mobilité humaine.
3. Mener un ensemble d'expérimentations pour confirmer l'efficacité du TCPT à l'aide d'un jeu de données réel (Brightkite) et comparer ses performances avec un certain nombre des modèles de prédiction de localisation existants.

Organisation de mémoire

Ce mémoire est organisé comme suit:

* Dans le premier chapitre, nous allons présenter un ensemble de concepts de base liés à la prédiction de la mobilité en général ainsi que les travaux de littérature. Ces travaux ont été classifiés par la technique de prédiction utilisée et comparés par la suite en fonction de certains critères tels que la précision de prédiction et la prise en compte du contexte temporel.

* Dans le deuxième chapitre, nous avons fourni une description détaillée de TCPT pour qu'il considère le contexte temporel.

* Dans le troisième chapitre, nous présentons une étude expérimentale pour 1) tester les performances de TCPT et les comparer avec d'autres modèles de prédiction de la littérature et 2) d'étudier l'effet de l'augmentation de l'intervalle de temps sur le résultat de la prédiction.

* Enfin, nous terminons le mémoire par une conclusion.

Chapitre 1 : État de l'art

1.1 Introduction

La prédiction de la mobilité est un domaine de recherche qui a été appliquée dans plusieurs domaines à des fins différentes. Elle consiste à prédire le mouvement d'un objet mobile en prédisant les futurs endroits où ils pourraient se rendre.

Dans ce chapitre, nous allons présenter un ensemble de définitions concepts liés à la prédiction de la mobilité, puis nous allons aborder ses types que nous avons résumés en prédire le mouvement humain et le mouvement des objets. Ensuite, nous allons montrer l'impact du facteur temps et son importance dans le processus de prédiction. Par la suite, nous allons présenter et classer les travaux pertinents liés à la prédiction de la localisation et de la prochaine route selon la technique utilisée pour la prédiction.

1.2 Notions préliminaires

Définition 1 (Journal GPS)

Un journal GPS est un ensemble P des points GPS tel que $P = \{p_1, p_2, \dots, p_n\}$. Comme illustré dans la partie gauche de la Figure 1, chaque point $p_i \in P$ contient les valeurs de latitude, longitude et temps.

Définition 2 (Trajectoire)

Une trajectoire $Traj = p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$, où $p_i \in P$ et P est l'ensemble des points GPS, est une séquence d'un ensemble de points GPS qu'un objet en mouvement croise au cours de son mouvement.

Définition 3 (Point de séjour)

Un point de séjour représente une région géographique où un utilisateur est resté pendant un certain intervalle de temps. L'extraction d'un point de séjour dépend de deux paramètres d'échelle, un seuil de temps (T_{thre}) et un seuil de distance (D_{thre}).

Définition 4 (Trajet)

Un trajet est une trajectoire GPS commencée et terminée avec un point de séjour; par conséquent, les points de séjour sont des délimiteurs de trajet. Notez que l'origine du conducteur est également considérée comme un point de séjour.

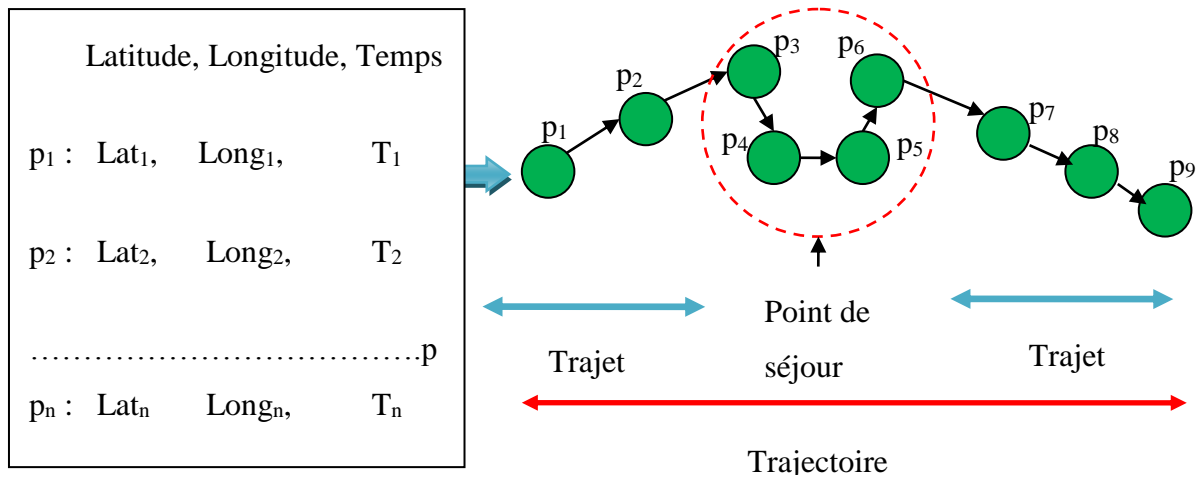


Figure 1. Un journal GPS, une trajectoire GPS, un point de séjour et un trajet.

Définition 5 (Localisation)

Une localisation l_i est une zone géographique où une personne ou un objet mobile se situe ou s'est enregistré (ex. Comme la maison ou l'université).

Définition 6 (Séquence de localisations)

Une séquence de localisations $S = l_i, l_{i+1}, \dots, l_m$ est l'ensemble des localisations qu'un utilisateur u a visités dans l'ordre pendant une période de temps.

1.3 Prédiction de mobilité

La prédiction de la mobilité est un problème de recherche qui a attiré l'attention des chercheurs ces dernières années[6][7][8] en raison de ses vastes applications telles que (1) la prédiction de localisations uniques ou séquentiels, (2) l'étude de comportement de mobilité d'un groupe ou un individu, (3) des applications basées sur les services urbains, etc. La prédiction de la mobilité pourrait être appliquée à n'importe quel domaine tel que les secteurs gouvernementaux, les unités de soins de santé, les domaines scientifiques et autres domaines similaires.

Le terme « prédiction de mobilité » fait référence au processus de suivi des schémas (patterns) de mouvement d'une personne ou d'un objet et de la prédiction de leurs localisations futures sur la base de ces schémas en utilisant une combinaison de techniques telles que les chaînes de Markov, réseaux de neurones, la fouille de données, etc.

1.4 Types de prédiction de mobilité

Comme mentionné précédemment, la prédiction de la mobilité a de nombreuses applications qui peuvent être regroupées en fonction de la tâche de prédiction en prédiction de la mobilité humaine et prédiction de la mobilité des objets.

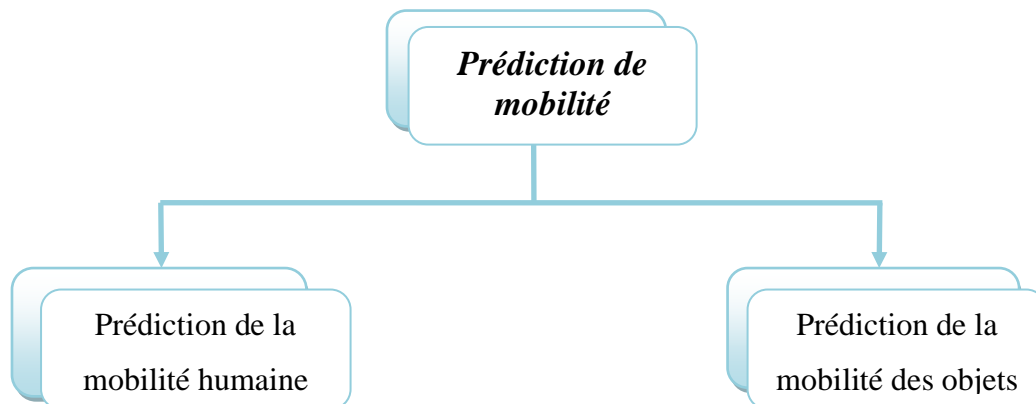


Figure 2. Types de prédiction de mobilité

4.1 Prédiction de la mobilité humaine

La prédiction du mouvement humain a suscité un intérêt croissant de la part de la communauté de recherche et de l'industrie, car de nombreuses recherches ont été consacrées à l'identification des modèles (patterns) de la mobilité humaine et à la prédiction du mouvement humain dans divers domaines.

4.1.1 Régularité de la mobilité humaine

De nombreuses études ont montré qu'il existe de nombreux facteurs qui influent la mobilité humaine, qui peut être divisée en obligations régulières et prévisibles, et en circonstances imprévues.

La majorité des personnes en général exercent certaines activités presque quotidiennement qui sont répétitives (par exemple, aller au travail, se rendre de travail ou emmener l'enfant à l'école.), mais parfois ils changent leur routine et visitent des lieux inconnus (ex. voyager pour aider un parent malade ou aller au cinéma). Par conséquent, nous concluons que la mobilité humaine ne se limite pas à la régularité.

Malgré la diversité apparente des raisons pour lesquelles les individus se déplacent [9][10][11], un grand nombre de travaux a trouvé une régularité et une prévisibilité significatives dans les modèles de mobilité humaine[12][13] (Voir Figure 3) . *Wang et al.* [14]ont proposé un modèle de prédiction de localisation hybride appelé RCH intégrant à la fois la régularité et la conformité de la mobilité humaine, capturant les schémas (patterns) de déplacement réguliers des utilisateurs et leurs visites occasionnelles influencées par d'autres. Dans[15], une méthode simple et efficace a été introduite pour mesurer la régularité des visites d'un individu à un endroit et l'utiliser pour explorer la présence de régularité et de routine dans les données du monde réel. En plus de spécifier la régularité comme un modèle de visite qui se répète avec un laps de temps récurrent (par exemple, semaine par semaine ou jour par jour).*De Lira et al.* [16] ont étudié la tendance des individus mobiles à être réguliers ou irréguliers lors du choix des lieux où effectuer certaines activités où ils ont appelé cette caractéristique la régularité sémantique de l'individu. Ils voulaient pouvoir distinguer les clients réguliers qui ont tendance à se rendre régulièrement au même endroit pour exercer une activité, de ceux qui essaient de nouvelles possibilités d'un point de vue spatial et/ou temporel.

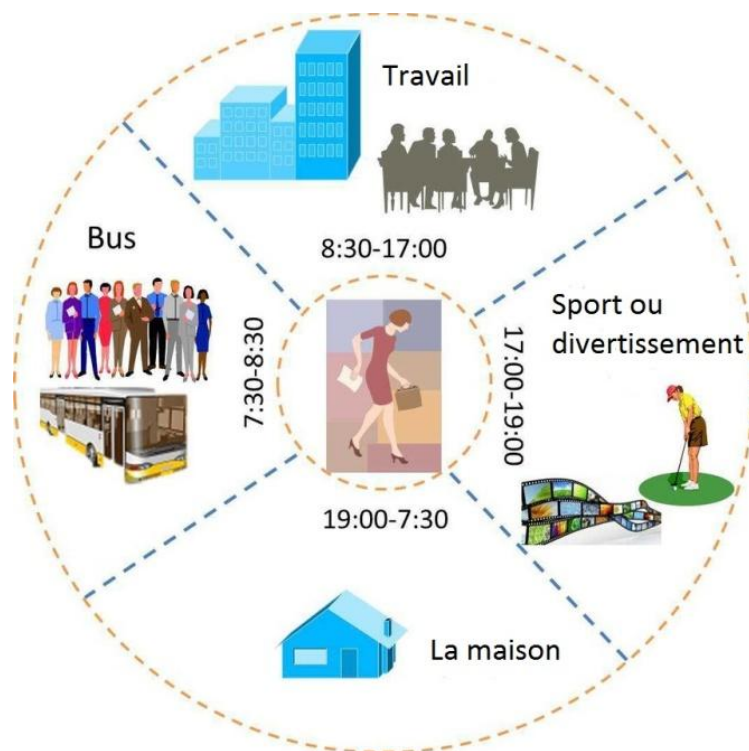


Figure 3. Régularité de la mobilité humaine.

4.1.2 La prédictibilité de la mobilité humaine

La disponibilité croissante des données sur la mobilité humaine, accessibles via différentes sources (ex. appareils intelligents, réseaux sociaux, etc.), a permis de prédire les mouvements des humains. Cette mobilité est en réalité régulière, prévisible et unique dans les espaces temporels et spatiaux, selon un grand nombre d'études [17][18]. Cette régularité et prévisibilité étaient incarnées plus précisément par des mouvements réguliers observables entre quelques lieux souvent visités, comme le domicile et le travail [19]. Selon *Song* et al. [20][10] qui ont proposé deux mécanismes génériques, à savoir l'exploration occasionnelle et le retour préférentiel qui régissent les trajectoires humaines d'un point de vue microscopique, une prédictibilité de potentielle de la mobilité humaine de plus de 93% a été revendiquée.

4.2 Prédiction de la mobilité des objets

La localisation d'un objet en mouvement est une information très appréciée car elle facilite la fourniture des services proactifs avancés tout en aidant le développement des applications mobiles sensibles à la localisation (MLA) (Figure 4). A titre d'exemple, les auteurs dans [21] ont proposé une méthode de suivi des objets dans les réseaux de capteurs via la fouille de données, où ce modèle vise à prédire la localisation suivant de l'objet, ce qui augmente la durée de vie du réseau. De plus, *Balico* et al. [22] ont étudié et analysé des approches proposées pour la localisation, le suivi des cibles et les techniques de prédiction de séries chronologiques qui peuvent être exploitées pour estimer la localisation futur d'un véhicule dans les réseaux ad hoc de véhicules (VANET) plutôt que dans les réseaux mobiles courants.

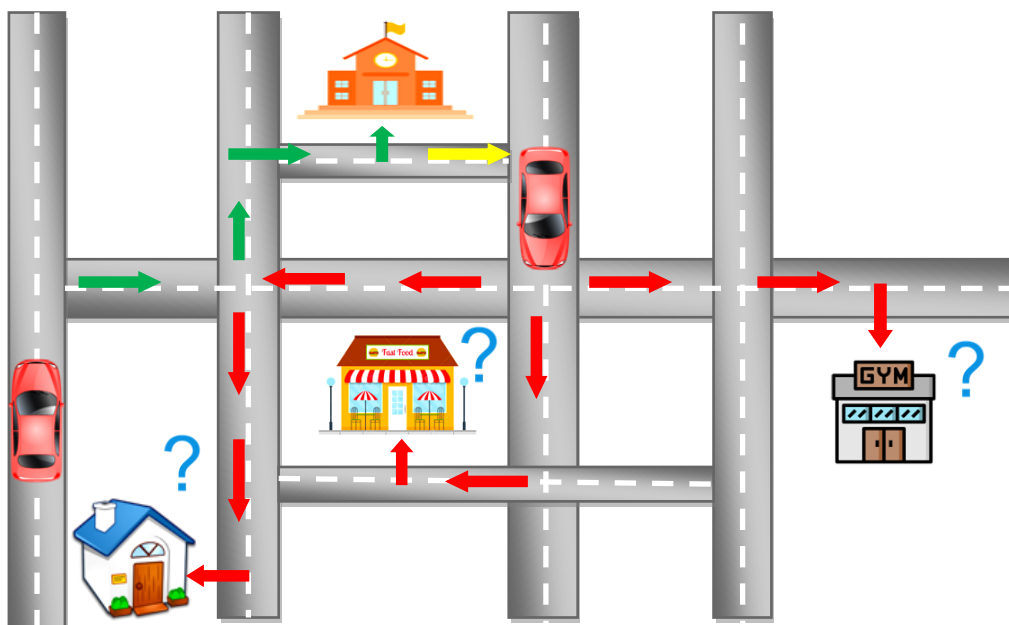


Figure 4. Exemple de prédiction de mobilité d'un objet en mouvement.

1.5 L'effet de temps dans la prédiction de mobilité

L'influence temporelle est importante pour la prédiction de mobilité des humaines où elle agit sous trois aspects : la périodicité, la continuité et la non-uniformité.

Premièrement, en ce qui concerne la périodicité, elle a été introduite pour la première fois dans [23] où *Cho* et al. ont suggéré des données d'enregistrement (check-ins) périodiques pour les LBSNs (Location based social networks). Par exemple, les gens restent toujours dans leurs bureaux et les lieux de travail les jours de travail et vont aux centres commerciaux les week-ends.

Du plus, et concernant la continuité, *Cheng* et al. [24] ont proposé le modèle FPMC-LR (Factorized Personalized Markov Chain) [25] avec contrainte de région locale pour capturer la conséquence, en supposant la forte corrélation entre deux localisations consécutives enregistrées. Par exemple, les utilisateurs peuvent s'amuser dans un manège après avoir mangé au restaurant. Du point de vue géographique, cette tendance à l'enregistrement fréquent indique que le manège et le restaurant sont adjacents et associés.

Ensuite, en ce qui concerne la caractéristique de non-uniformité, *Joe* et al. [26] ont l'introduit en recherchant les caractéristiques temporelles du comportement d'enregistrement des utilisateurs sur les LBSN, ce qui les a aidés à créer un cadre de prédiction et recommandation de localisation avec des effets temporels.

1.6 Travaux connexes

La prédiction de localisation a fait l'objet d'études approfondies ces dernières années où la conception des nouvelles approches à aider d'étendre et d'améliorer la qualité des services des ITS (Intelligent transportation systems) et des LBS (Location-based services).

Dans cette section, nous fournissons tout d'abord une classification des modèles de prédiction selon la technique de prédiction utilisée suivi par une comparaison des modèles les plus importants en concentrant sur quelques critères comme la précision et la considération de contexte temporel.

6.1 Taxonomie des modèles de prédiction de localisation

Cette section présente les modèles de prédiction dans une taxonomie présentée dans Figure 5. Cette dernière classe les modèles en fonction de types de techniques utilisées dans la prédiction en l'une des trois catégories suivantes: (1) modèles de la fouille de données, (2) modèles probabilistes et (3) modèles de correspondance de trajets. Notant que certains modèles combinent plus d'un type de technique.

6.1.1 Modèles de la fouille de données

Les techniques de la fouille de données ont été largement appliquées pour la prédiction de la localisation.

A) La fouille des motifs séquentiels

Les motifs (patterns) séquentiels peuvent être utilisés comme moyen de prédiction pour représenter le comportement d'un objet mobile en une séquence de localisations lors de sa mobilité. Dans [4], *Chen* et al. ont tenté de prédire la destination prévue et la route futur d'une personne. En utilisant des données GPS réelles, les auteurs ont proposé de regrouper les lieux importants d'une personne pouvant s'aller ou se rendre en utilisant l'algorithme de regroupement FBM (Forward Backward Matching). Les motifs séquentiels de mouvement sont ensuite extraits des trajectoires abstraites en utilisant une extension de l'algorithme CRPM (Continuous Route Pattern Mining). Dans un autre travail basé sur les motifs séquentiels, *Merah* et al. [27] ont présenté plusieurs schémas de communication qui pourraient être utilisés pour collecter des chemins de véhicules. Les motifs de mouvement séquentiels les plus fréquents sont ensuite extraits comme les trajets les plus fréquents et utilisés par la suite pour générer des règles de mouvement qui pourraient être utilisées pour prédire les futures routes des véhicules.

B) Les réseaux de neurones

Les réseaux de neurones ont également été appliqués pour la prédiction. *Mikuscak* et al. [2] ont présenté plusieurs algorithmes et méthodes qui ont été utilisés dans les systèmes de transport intelligents (STI). Les auteurs ont proposés une méthode de prédiction des routes basée sur des réseaux de neurones en basant sur l'historique de mobilité d'un véhicule. De plus, *De Brébisson* et al. [28] ont développé un framework pour prédire la destination d'un taxi en

fonction de son emplacement de départ et des méta-informations associées telles que l'heure de départ, l'identifiant du conducteur et les informations sur le client. Un réseau neuronal bidirectionnel récurrent a été appliqué pour présenter le trajet de chaque taxi avec ses métadonnées pertinentes. *Liu et al.*[29] ont étendu les réseaux de neurones en proposant une nouvelle méthode appelée réseaux de neurones récurrents spatiaux temporels (ST-RNN). ST-RNN peut modéliser des contextes temporels et spatiaux locaux dans chaque couche avec des matrices de transition spécifiques au temps pour différents intervalles de temps et des matrices de transition spécifiques à la distance pour différentes distances géographiques.

C) Approche basée sur le regroupement

Le regroupement ou clustering a été largement utilisé pour les tâches de prédiction de la localisation et surtout pour identifier les points de séjour définis comme des emplacements où un conducteur passe une longue période. *Ying et al.*[30] ont proposé une nouvelle technique de prédiction basée sur le clustering pour prédire l'emplacement suivant d'un utilisateur mobile. Aussi, *Chen et al.*[31] ont proposé une approche de prédiction de l'emplacement future en combinant deux modèles de clustering pour regrouper les objets en fonction de leurs emplacements spatiaux et de leurs trajectoires en utilisant une métrique de similarité.

De plus, *Cao et al.*[32] ont proposé un modèle appelé Tra-MM(modèle de Markov en cluster de trajectoires) qui exploite la similitude entre les trajectoires. Tra-MM regroupe d'abord les trajectoires similaires selon une métrique de similarité, puis pour chaque cluster, un modèle de Markov d'ordre variable est utilisé pour la prédiction. Dans [33], *Terroso-Saenz et al.* ont présenté un modèle pour la prédiction des routes basé sur le regroupement de zones de changement de vitesse significatives appelées Prop-Turn. Le modèle proposé adopte une architecture client léger et serveur et propose un nouveau concept d'abstraction des trajets basé sur la détection des régions spatiales où certaines caractéristiques de vitesse changent fréquemment.

6.1.2 Modèles probabilistes

Des modèles probabilistes ont été utilisés pour les problèmes de prédiction des routes. Ces modèles reposent principalement sur l'hypothèse de Markov. Dans [1], *Mathew et al.* ont construit des modèles de Markov cachés (Hidden Markov Model) pour chaque emplacement afin de prédire les emplacements futurs des objets mobiles. De plus, *Gambis et al.* [34] ont

abordé la question de la prédiction de l'emplacement suivant d'un individu sur la base des observations de son comportement de mobilité dans une certaine période de temps et des emplacements récents qu'il a visités. Plus précisément, les auteurs ont étendu un modèle de mobilité appelé Mobility Markov Chain (MMC) pour incorporer les n emplacements précédemment visités pour la prédiction de la localisation suivante. *Chen* et al.[35] ont proposé un système de prédiction nommé (NLPMM) avec une modélisation de Markov pour prédire le prochain emplacement d'échantillonnage auquel un objet en mouvement arrivera avec une séquence de trajectoires donnée. De plus, *Epperlein* et al.[36] ont présenté un algorithme pour prédire la destination, la route, ou plus généralement un modèle d'un voyage en cours en utilisant les chaînes de Markov comme modèles pour les motifs de voyage. *Tiwari* et al. [3] ont conçu une application de prédiction incrémentale des routes basée sur une modélisation de prédiction par correspondance partielle (PPM) des données de mobilité des utilisateurs.

6.1.3 Modèles de la correspondance des trajets (Trip Matching)

La correspondance des trajets (Trip Matching) consiste à trouver le trajet le plus similaire pour un utilisateur en comparant la trajectoire actuelle du conducteur avec ses précédents trajets. Pour mesurer la similitude entre les paires de trajectoires, la métrique de Hausdorff est utilisée dans [37] pour comparer la congruence topologique de deux trajets ainsi que les directions de conduite. Les auteurs dans [37] ont également étendu leur modèle pour prendre en compte le contexte temporel (heure de jour et jour de la semaine) afin de tenir compte de l'homogénéité temporelle des habitudes de conduite de chaque utilisateur.

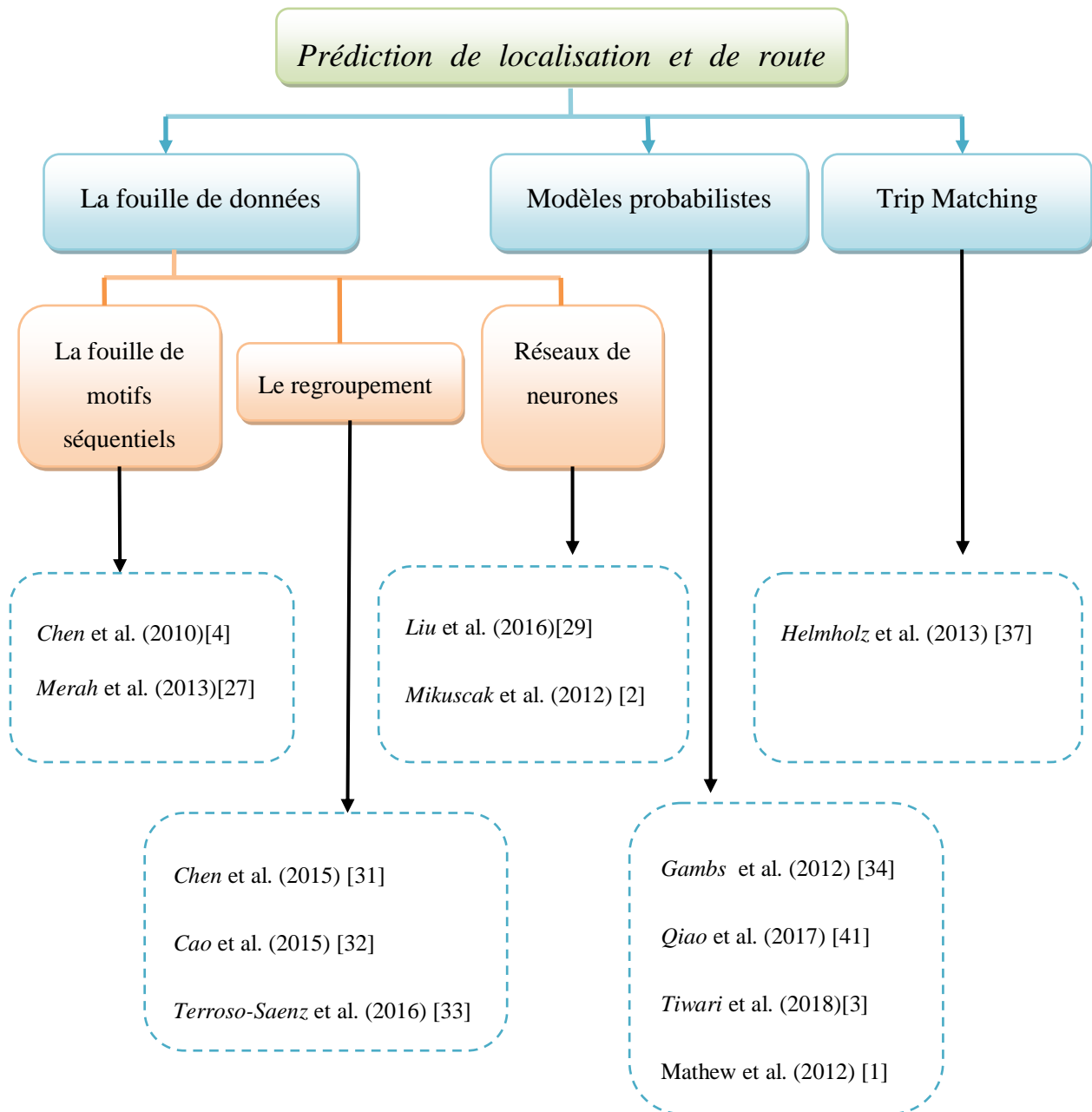


Figure 5. Taxonomie des modèles de prédiction de localisation et de route.

6.2 La considération de temps ou contexte temporel

La majorité des études existantes pour prédire la prochaine localisation ne prennent pas en compte les informations contextuelles. Elles s'appuient que sur l'historique des mouvements ou la localisation spatiale pour faire des prédictions. Cela n'empêche pas les informations contextuelles d'être d'une grande importance, car les applications sensibles au contexte sont capables d'identifier et interagissent avec des situations réelles mieux que des applications qui

ne prennent pas en compte ce type d'informations qui peuvent jouer un rôle dans l'augmentation de la précision de la prédiction. *Bar-David* et al. [38] ont proposé un algorithme de prédiction des futures routes et qui base sur divers types d'informations contextuelles pour prédire l'emplacement futur des véhicules. Les auteurs ont utilisé cinq caractéristiques contextuelles liées à l'environnement de l'objet où à ses données de mouvement actuelles (emplacement actuel, vitesse de l'objet, jour de la semaine, conditions météorologiques, les embouteillages dans la région). Cette étude en réalité nous a motivé à proposer une approche contextuelle dans le but d'essayer d'améliorer la qualité de prédiction.

L'une des informations contextuelles les plus importantes est le contexte temporel. Ce dernier est crucial pour comprendre et détecter les changements d'activité. *Scellato* et al. [39] ont présenté NextPlace, une nouvelle approche de prédiction spatio-temporelle basée sur une analyse non linéaire de la série chronologique des heures de début et de la durée des visites dans des lieux importants.

Scellato et al. [39] ont proposé NextPlace, une nouvelle approche de prédiction spatio-temporelle basée sur une analyse non linéaire de la série chronologique des heures de début et de la durée des visites dans des lieux importants. *Chen* et al. [35] ont proposé le modèle (NLPMM) de prédiction basé sur modélisation de Markov en tenant compte le facteur temps.

6.3 Comparaison des travaux

Dans cette section, nous fournissons un tableau montrant une comparaison des différents algorithmes de prédiction d'itinéraire et de destination examinés ci-dessus. Lorsque ces modèles sont comparés selon certains critères tels que la technique utilisée, le type de données utilisées dans l'évaluation empirique, la précision, en plus de montrer qu'elle a été prise en compte dans le contexte temporel ou non.

Auteurs	Technique utilisée	Les données	Contexte temporel	Précision
Amirat et al. (2019) [40]	Compact prediction Tree (CPT)	*Vrai GPS (conduite) * Ensemble de données synthétiques	Non inclus	35% - 96%
Epperlein et al. (2018) [36]	Modèle de chaîne de Markov	real GPS data	Non inclus	plus de 99%

Tiwari et al. (2018)[3]	prédiction par correspondance partielle (PPM)	Données GPS	Non inclus	-
Qiao et al. (2017)[41]	Modèle de Markov caché (HMM)	Trafic réel du plan de contrôle collecté à partir d'un GSM LTE, UMTS, GPS ou Wi-Fi.	Inclus	plus de 80%
Liu et al. (2016) [29]	méthode de réseaux de neurones récurrents spatiaux temporels (ST-RNN)	*Base de données mondiale sur le terrorisme (GTD) *Ensemble de données Gowalla	Inclus	-
Terroso-Saenz et al. (2016)[33]	Clustering basé sur la densité+ vitesse	Ensemble de données 1 GPS réel (smartphone) Ensemble de données 2 Synthétique	Non inclus	92%
De Brébisson et al. (2017)[28]	Réseaux de neurones artificiels	Données GPS	Non inclus	-
Chen et al. (2015)[31]	Clustering basé sur le temps	GPS réel	Non inclus	60%- 80%
Chen et al. (2014) [35]	Modélisation de Markov (NLPMM)	Ensemble de données d'enregistrement de passage de véhicule réel	Inclus	40% - 50%
Helmholz et al. (2013) [37]	Variation de la métrique de Hausdorff	GPS réel	Non inclus	80%
Merah et al. (2013) [27]	Exploration séquentielle de modèles	Ensemble de données synthétiques	Non inclus	-
Mikuscak et al. (2012) [2]	Réseaux de neurones artificiels à réaction directe	Ensemble de données synthétiques	Non inclus	59.79%- 98.7%
Gambis et al. (2012) [34]	Mobility Markov Chain (MMC)	*L'ensemble de données phonétiques *L'ensemble de données Geolife collecté à l'aide d'appareils compatibles GPS	Non inclus	70% to 95% as soon as n = 2

		*Ensemble de données synthétiques		
Mathew et al. (2012) [1]	Modèle de Markov caché (HMM)	Données GPS	Non inclus	13.85%
Ying et al. (2011)[30]	stratégie de prédiction basée sur les clusters	Ensemble de données de téléphonie mobile collecté par le MIT Media Laboratory	Non inclus	53% - 68%

Tableau 1. Littérature pour les études de prédiction de localisation

1.7 Conclusion

Dans ce chapitre, nous avons présenté un ensemble de définitions et quelques concepts liés à la prédiction de la mobilité, puis nous avons montré l'impact du facteur temps et son importance dans le processus de prédiction.

Ensuite, nous avons présenté un aperçu de certains modèles utilisés pour prédire l'emplacement et le prochain itinéraire. Nous avons ainsi classé ces modèles en fonction de la technique utilisée dans la prédiction et les comparés à travers certains facteurs tels que la précision et en montrant qu'ils considèrent le contexte temporel ou non.

**Chapitre 2 : Le
modèle TCPT pour la
prédiction de
localisation**

2.1 Introduction

La prédiction de la mobilité est un domaine de recherche qui a beaucoup d'attention ces derniers temps. Cela est dû à ses vastes applications telles que la direction des voitures, les appareils d'assistance et les systèmes de recommandation de localisation..

La recherche de prédiction de mobilité s'est concentrée sur la prédiction du prochain emplacement. Par conséquent, la plupart des modèles présentés ne prennent pas en compte les informations contextuelles et sont concentrés uniquement sur le contexte spatial pour faire leurs prédictions, malgré l'importance majeure des autres informations contextuelles (en particulier le contexte de temps) du fait qu'elles influencent la précision de la prédiction.

Dans ce chapitre, nous proposons un système de prédiction de localisation qui adapte le modèle CPT pour qu'il puisse considérer le contexte temporel de mobilité dans la prédiction.

Ce chapitre est organisé comme suit.

Dans la section 1, une description générale sur le modèle CPT de base pour la prédiction des séquences est présentée.

Nous présentons dans la section 2, le modèle TCPT qui étend l'algorithme CPT pour considérer la régularité temporelle de la mobilité humaine.

Dans la section 3, l'architecture générale de notre système basé sur TCPT est présentée.

2.2 Notions Préliminaires

Définition 1 (Localisation ou POI) : Une localisation est une zone géographique, qui a un identifiant unique l_i et a été visitée par un utilisateur u_i dans LBSN (Location Based Social Networks).

Définition 2 (Séquence de mouvement/mobilité) : Une séquence de mobilité $S_m = \langle l_1, l_2, \dots, l_n \rangle$ est une liste de localisations qu'un utilisateur traverse au cours de son déplacement dans une zone géographique. Par exemple le Tableau 2 affiche trois séquences de mobilité pour trois

utilisateurs, montrant que le deuxième utilisateur parcourt une série de localisations classés en commençant par l_1 à l_2, l_3, l_5 jusqu'à l_6 .

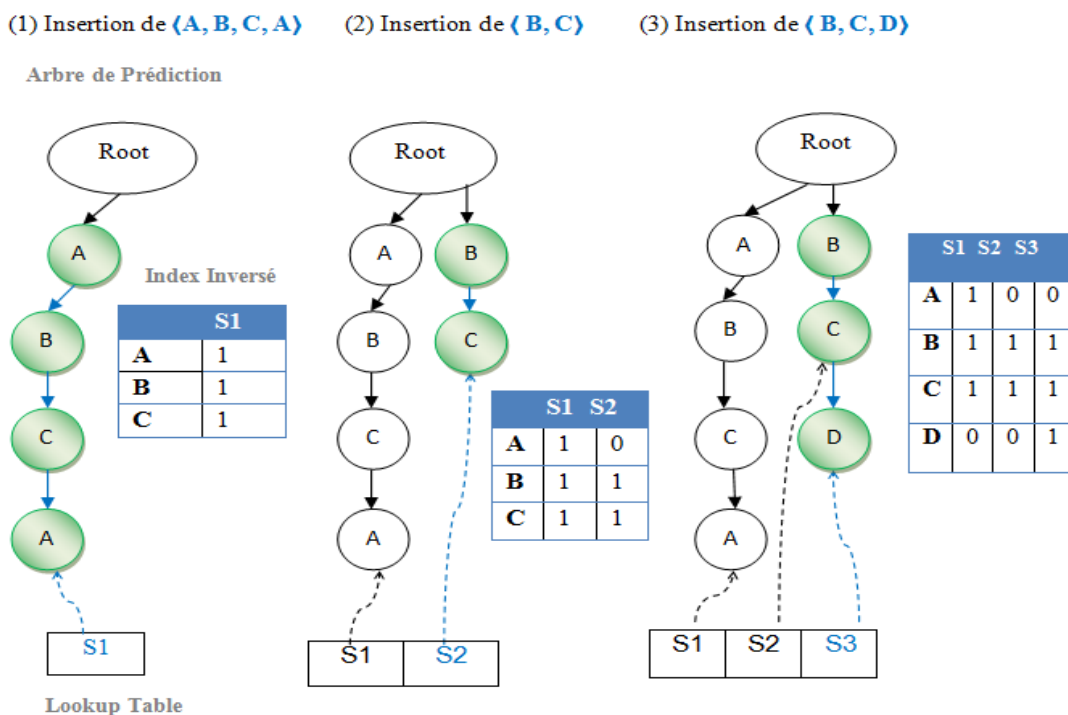
Les utilisateurs	Séquence de mobilité
U_1	$\langle l_1, l_2, l_4, l_6, l_3 \rangle$
U_2	$\langle l_1, l_2, l_3, l_5, l_6 \rangle$
U_3	$\langle l_3, l_4, l_2, l_6, l_5 \rangle$

Tableau 2. Séquences de mobilité.

2.3 Le modèle CPT (Compact Prediction Tree)

Le CPT (Compact Prediction Tree) ou l'arbre de prédiction compacte est un modèle de prédiction récemment proposé par Gueniche et al. dans[5]. CPT se caractérise par sa propriété de la non-perte de données du fait qu'il compresse les données d'apprentissage afin que toutes les informations pertinentes soient disponibles pour chaque prédiction. Ce modèle se compose de deux phases: l'apprentissage et la prédiction. Dans la phase d'apprentissage de CPT, trois structures de données sont construites : (1) un Arbre de Prédiction (AP), (2) un Lookup Table et (3) un Index Inversé (II).

Figure 6 illustre la construction de ces trois structures à l'aide des cinq séquences suivantes : $S_1 = \langle A, B, C, A \rangle$, $S_2 = \langle B, C \rangle$, $S_3 = \langle B, C, D \rangle$, $S_4 = \langle B, C, E, D \rangle$, $S_5 = \langle E, A \rangle$.



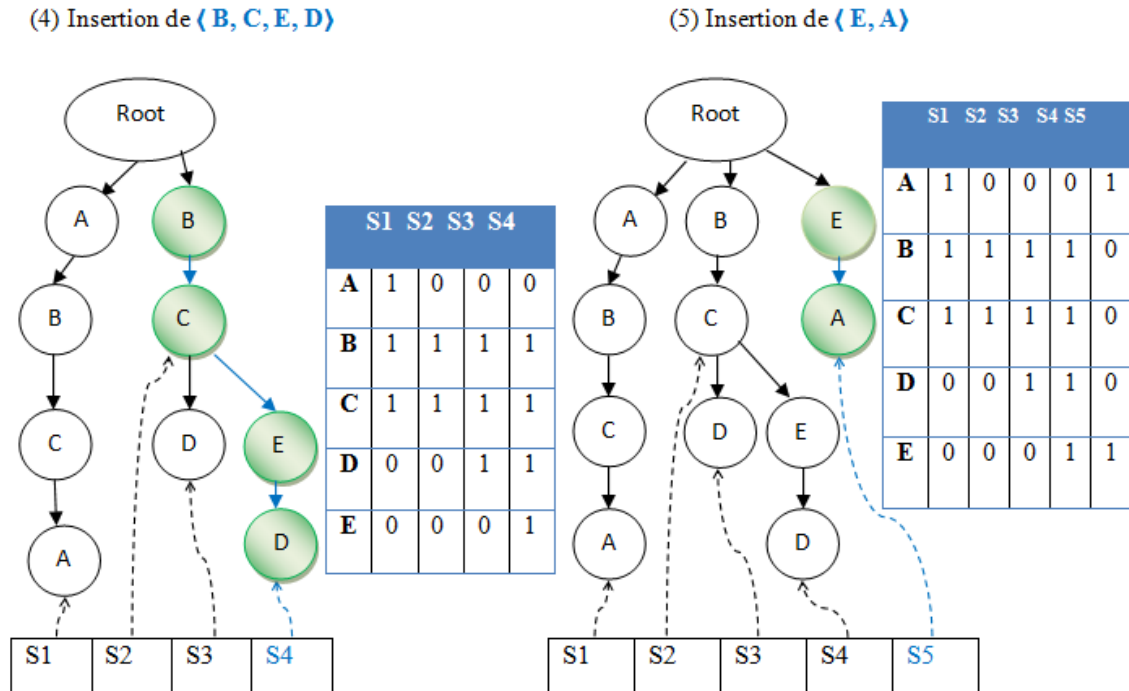


Figure 6. La construction des structures CPT.

Une fois les trois structures sont construites, l'étape de prédiction commence. Pour prédire le symbole (localisation dans notre cas) qui suit une séquence donnée s_i , CPT commence par la recherche de l'ensemble des séquences Y similaires à s_i (Sequence Matching). Après, le conséquent (les symboles juste après les symboles de s_i) de chaque séquence $y \in Y$ similaire à s_i . Ensuite, CPT stocke chaque élément de chacun de ces conséquents dans une structure appelée *CountTable(CT)*. Ce dernier est une table de hachage où la clé représente le conséquent et la valeur associée à cette clé est une valeur score qui estime le nombre de fois que cet conséquent a été traversé après les symboles de s_i . Finalement, CPT renvoie le conséquent avec le score le plus élevé comme localisation prédit. Figure 7 illustre le processus de prédiction du symbole suivant après la séquence $s_i = \langle BC \rangle$ en basant sur le modèle CPT présenté dans Figure 6.

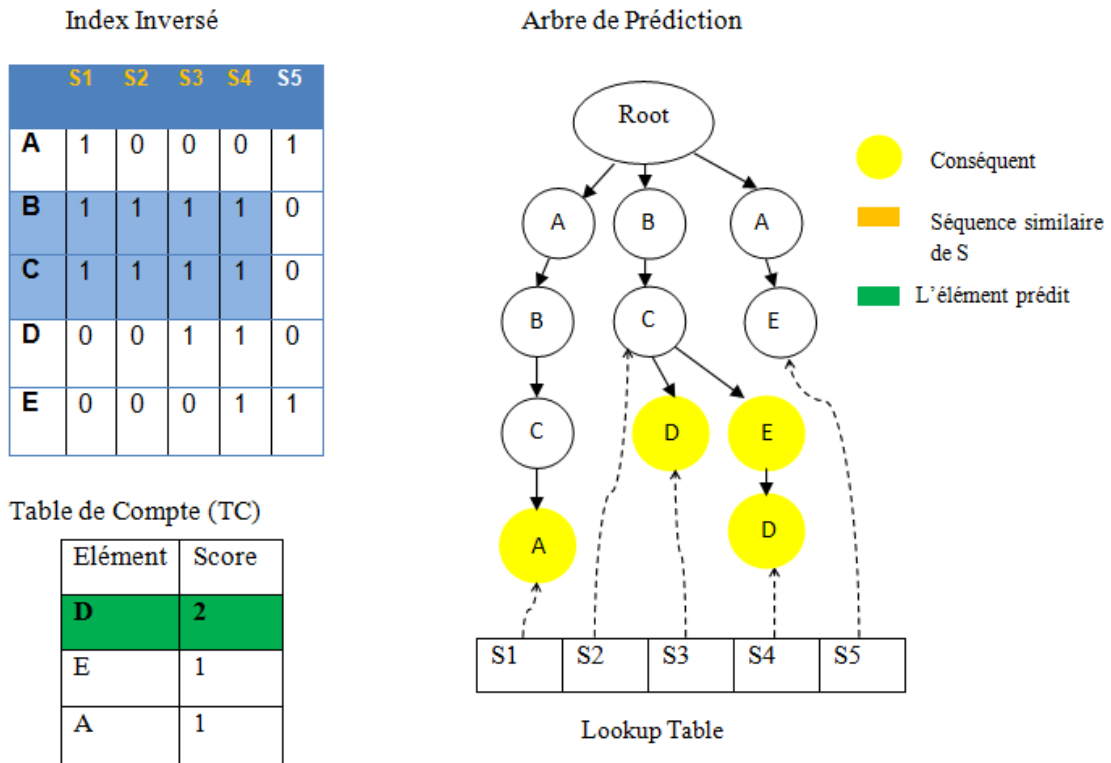


Figure 7. La phase de prédiction CPT.

2.4 Le modèle TCPT (Temporal Compact Prediction Tree)

Le modèle TCPT est l'extension de modèle CPT qui considère la régularité temporelle de la mobilité humaine. TCPT étend localisations et des séquences de mobilité temporelles pour qu'elles contiennent/considèrent le temps de visite de chaque localisation. TCPT donc, prend en entrée des localisations et des séquences de mobilité temporelles et donne en sortie les futures localisations susceptibles d'être visitées par une personne. Dans cette section, nous définissons tout d'abord puis nous présentons une description des processus d'apprentissage et de prédiction de TCPT.

2.5 Définitions

Définition 1 (Visite temporelle, localisation temporelle) : Une visite temporelle est un ensemble de lieux visités par un utilisateur dans le même intervalle de temps. Formellement, une visite temporelle s'écrit $T_{vi} = (\langle t_i \rangle L_i, L_{i+1}, \dots, L_k)$. On note $Loc(T_{vi})$ l'ensemble des localisations visitées en $Temps(T_{vi}) = t_i$ dans T_{vi} . On note également comme localisation temporelle $tl_j = \langle t_j \rangle L_j$ dans T_{vi} si $L_j \in Loc(T_{vi})$ et $t_j = Temps(T_{vi})$.

Définition 2 (décomposition de temps en plage ou time-slot)

La décomposition de temps en time-slot est le processus qui discrétise le temps d'un jour à l'aide de plages horaires, où la journée est divisée en plusieurs plages horaires de longueur égale $taille_time_slot$. Par exemple, la décomposition d'un jour donne $t_1 = [00:00-02:00[$, $t_2 = [02:00-04:00[$, ..., $[22:00-00:00[$ (pour les plages horaires de 2 heures ($taille_time_slot$)).

Définition 3 (Séquence de mobilité temporelle)

La séquence de mobilité temporelle $Sm_t = \langle t_1 \rangle l_1, \langle t_2 \rangle l_2, \dots, \langle t_x \rangle l_y$ où $l_i \in L$ et $t_j \in T$, L et T représentent respectivement l'ensemble des localisations qu'un utilisateur traverse au cours de son déplacement et l'ensemble des time-slots. Où chaque élément de Sm_t représente la localisation traversé par l'utilisateur à un moment précis.

Par exemple, Tableau 3 montre un échantillon de neuf séquences de mobilité générées dans 3 jours par un ensemble $V : \{V_1, V_2, V_3\}$ de trois véhicules. Chaque véhicule représente son propre trajet de jour en une séquence de mobilité temporelle. Par exemple, la voiture V_1 dans jours J_1 a visité la localisation l_1 dans t_1 et l_6 dans t_4 passant par l_3 dans t_3 et l_4 dans t_2 .

ID de véhicule	Jour	Séquence de mobilité
V1	J ₁	$\langle t_1 \rangle l_1, \langle t_2 \rangle l_3, \langle t_3 \rangle l_4, \langle t_4 \rangle l_6$
	J ₂	$\langle t_1 \rangle l_1, \langle t_2 \rangle l_2, \langle t_3 \rangle l_4, \langle t_5 \rangle l_5$
	J ₃	$\langle t_2 \rangle l_3, \langle t_3 \rangle l_5, \langle t_4 \rangle l_6, \langle t_6 \rangle l_2$
V2	J ₁	$\langle t_1 \rangle l_1, \langle t_2 \rangle l_4, \langle t_3 \rangle l_2, \langle t_5 \rangle l_3$
	J ₂	$\langle t_1 \rangle l_3, \langle t_2 \rangle l_5, \langle t_3 \rangle l_4, \langle t_4 \rangle l_2$
	J ₃	$\langle t_2 \rangle l_2, \langle t_3 \rangle l_3, \langle t_4 \rangle l_4, \langle t_6 \rangle l_2$
V3	J ₁	$\langle t_1 \rangle l_1, \langle t_2 \rangle l_3, \langle t_3 \rangle l_4, \langle t_4 \rangle l_2$
	J ₂	$\langle t_1 \rangle l_4, \langle t_3 \rangle l_4, \langle t_4 \rangle l_4, \langle t_5 \rangle l_2$
	J ₃	$\langle t_1 \rangle l_2, \langle t_2 \rangle l_3, \langle t_3 \rangle l_4, \langle t_4 \rangle l_2$

Tableau 3. Séquences de mobilité temporelle des trois véhicules

2.6 La prédiction de localisation avec TCPT

Soit $Sme = \{Sm_1, Sm_2, Sm_3, \dots, Sm_n\}$ un ensemble des séquences de mobilité temporelle. Le problème de prédiction de localisation avec TCPT consiste à prédire la prochaine localisation

avec son temps de parcours $\langle t_{n+1} \rangle_{l_{n+1}}$ d'une séquence temporelle de mobilité donnée $Sm_i = \langle \langle t_1 \rangle_{l_1}, \langle t_2 \rangle_{l_2}, \dots, \langle t_n \rangle_{l_n} \rangle$ en utilisant le modèle de prédiction M construit de l'ensemble Sme .

2.6.1 La phase d'apprentissage

A partir de l'ensemble des séquences de mobilité temporelles, le processus d'apprentissage génère trois structures de données appelées AP, LT, II en considérant les séquences une après l'autre.

Dans l'*Arbre de Prédiction (AP)*, chaque nœud symbolise une localisation et chaque séquence de mobilité est représentée par un chemin partant de la racine de l'arbre et se terminant par la localisation de destination du voyage.

Lookup Table est mise à jour après chaque insertion de séquence de mobilité dans le AP où pour chaque ID de séquence, le LT pointe vers le dernier nœud de la séquence dans le AP.

L'*Index Inversé (II)* est un ensemble de vecteurs binaires qui indique pour chaque localisation l_i la liste des séquences de mobilité contenant l_i . Ainsi, et si la séquence contient la localisation, le bit qui correspond à cette localisation dans le vecteur de la séquence dans le II est mis à 1 et à 0 dans le cas contraire.

La Figure 8 illustre la construction de ces trois structures par l'insertion successive des séquences de mobilité décrites dans Tableau 4 où l'étape (a) représente l'état initial où les structures sont vides, l'étape (b) représente le modèle après l'insertion de la première séquence de mobilité, l'étape (c) représente le modèle après l'insertion de la deuxième séquence de mobilité, et l'étape (d) le modèle après l'insertion de la troisième séquence de mobilité.

ID véhicule	Séquence de mobilité
V1	$\langle \langle t_1 \rangle_{l_1}, \langle t_2 \rangle_{l_3}, \langle t_3 \rangle_{l_4}, \langle t_7 \rangle_{l_2} \rangle$
V2	$\langle \langle t_2 \rangle_{l_3}, \langle t_3 \rangle_{l_4}, \langle t_5 \rangle_{l_7} \rangle$
V3	$\langle \langle t_2 \rangle_{l_3}, \langle t_3 \rangle_{l_4}, \langle t_4 \rangle_{l_5}, \langle t_6 \rangle_{l_1} \rangle$

Tableau 4. Les séquences de mobilité temporelle.

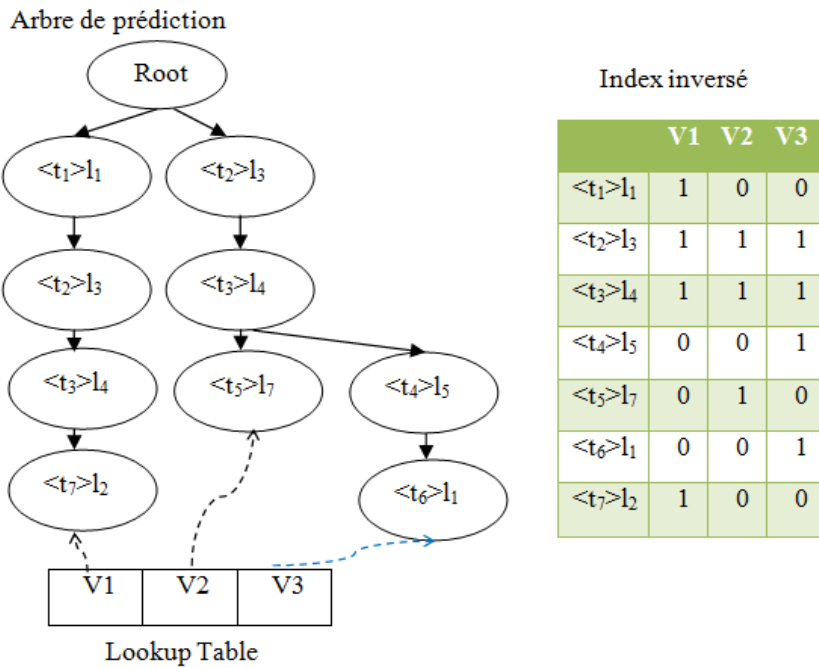
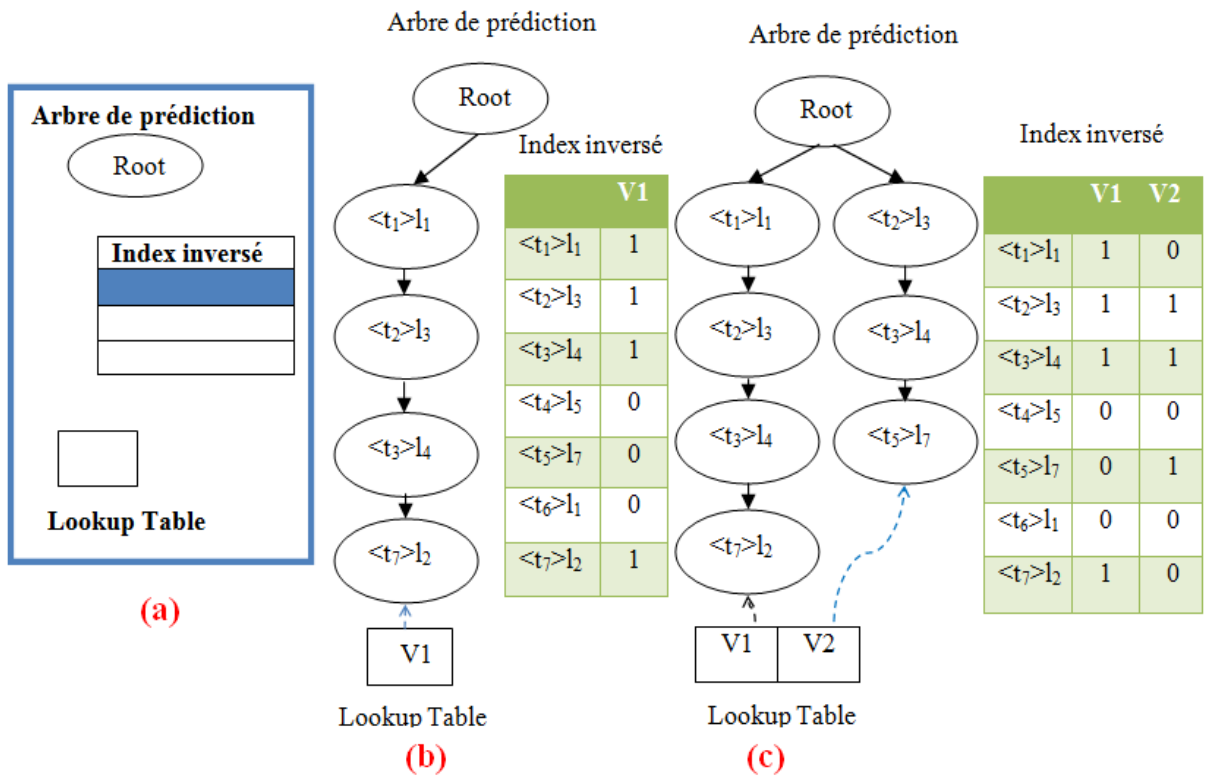


Figure 8. Un exemple de processus d'apprentissage TCPT.

Pour réduire la complexité temporelle et spatiale, le CPT adopte deux stratégies appelées : *compressions des séquences fréquentes* et *compressions des branches simples*.

A. Compressions des sous-séquences fréquentes (CSF)

Certaines répétitions peuvent être identifiées dans les séquences d'apprentissage. Ces répétitions peuvent être nombreuses et fréquentes tout dépend le jeu de données utilisé. La compression des chaînes fréquentes consiste à 1) identifier les sous-chaînes fréquentes des localisations apparaissant dans les séquences d'apprentissage, 2) remplacer chaque sous-séquence fréquente par un symbole individuel stocké dans une nouvelle structure nommée *Dictionnaire des séquences fréquentes (DSF)* et 3) associer chaque symbole à la sous-séquence qu'il remplace.

B. Compressions des Branches Simples (CBS)

La compression des branches simples est une stratégie de compression intuitive et efficace pour réduire la taille de l'arbre de prédiction. Elle consiste à substituer chaque branche simple (définie comme une branche menant à une seule feuille) par un seul nœud représentant la branche entière.

Figure 9 représente l'application de ces stratégies sur l'arbre de prédiction de Figure . En appliquant *CSF*, la sous-chaîne fréquente $\langle \langle t_2 \rangle l_3, \langle t_3 \rangle l_4 \rangle$ est remplacée par un nœud L_1 . L'arbre de prédiction est encore compressé en remplaçant la branche entière $\langle \langle t_1 \rangle l_1, L_1, \langle t_7 \rangle l_2 \rangle$ par un seul nœud en utilisant la stratégie *CBS*.

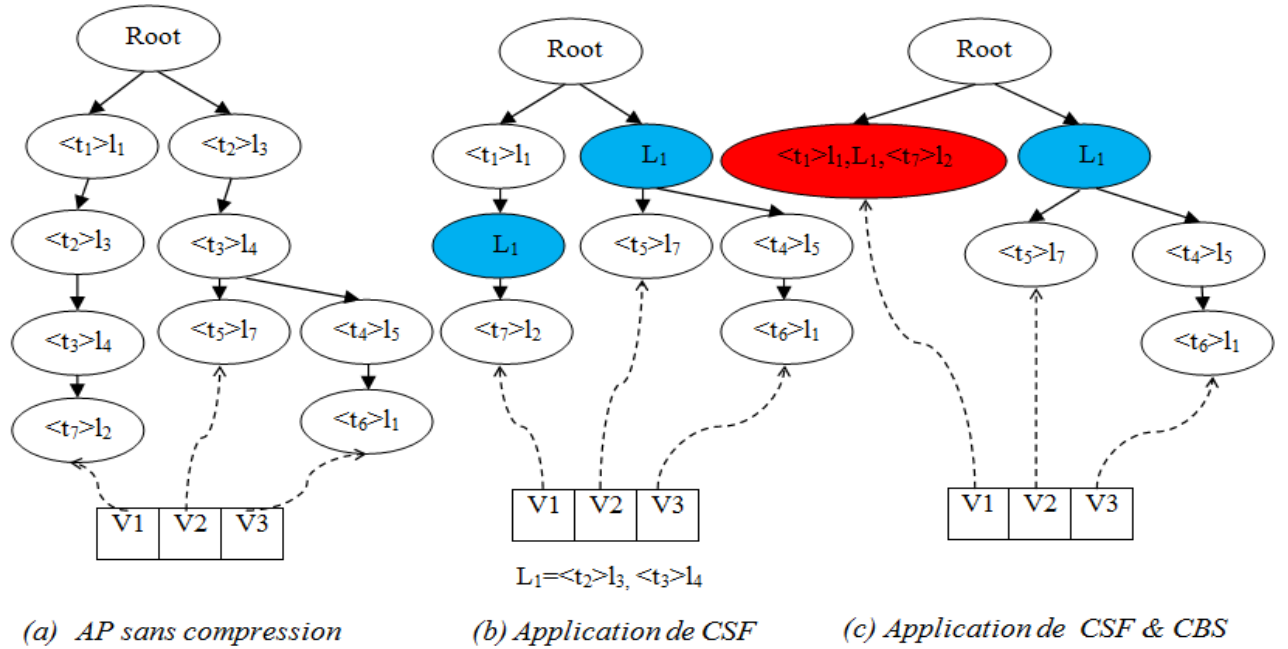


Figure 9. Application des stratégies CSF & CBS

2.6.2 La phase de prédiction

Dans cette phase, les trois structures de données qui ont été construites lors de la phase d'apprentissage sont utilisées. Nous définissons St comme la séquence de mobilité où $St = \langle t_1 \rangle l_1, \langle t_2 \rangle l_2, \dots, \langle t_m \rangle l_m$ qui contient n des localisations avec leur temps de transit. Nous définissons $DLT_j(St)$ comme les derniers localisations traversés à partir de St où $DLT_j(St) = \langle \langle t_{n-j+1} \rangle l_{n-j+1}, \langle t_{n-j+2} \rangle l_{n-j+2}, \dots, \langle t_n \rangle l_n \rangle$ avec $1 < j < n$. Le but du j est de spécifier le nombre de localisations dans la séquence qui doivent être pris en compte lors de la prédiction de la localisation suivant.

Le processus de prédiction de la localisation suivante (future) est effectué en mettant en œuvre un ensemble d'étapes, comme suit :

A. Recherchez des séquences similaires à St

Ce sont des séquences de mobilité qui contiennent toutes les localisations de $DLT_j(St)$ en utilisant II en implémentant l'intersection entre l'ensemble des bits de $DLT_j(St)$. L'ensemble de bits résultant désigne l'ensemble des séquences similaires à St , puis TC est utilisé pour accéder à ces séquences dans AP.

B. La recherche du conséquent de chaque séquence similaire à Sm

En considérant St et $DLTj(St)$, le conséquent d'une séquence similaire à St est défini comme la subséquence qui commence après le dernier élément commun entre $DLTj(St)$ et la séquence similaire jusqu'à la fin de la séquence similaire. La séquence de mobilité similaire à St peut être divisée en trois subséquences : la subséquence de contexte vient en première position suivi du $DLTj(St)$ et en troisième position la subséquence conséquente (comme le montre Figure 5).

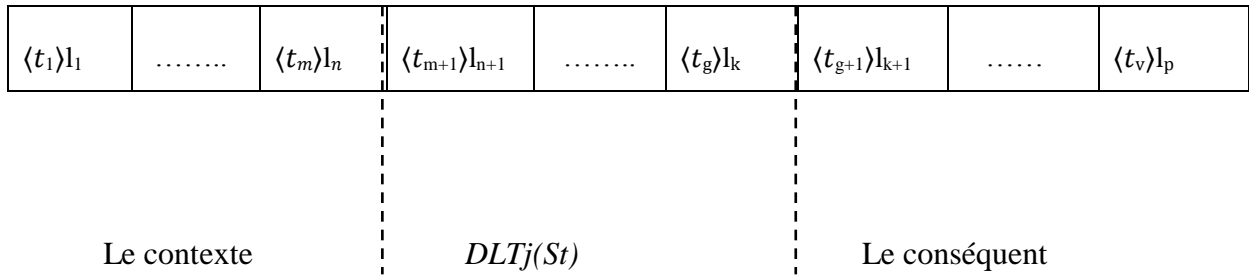


Figure 10. Le conséquent d'une séquence de mobilité

C. Prédire la localisation

Chaque conséquent d'une séquence similaire à St est stocké dans une structure appelée Table de comptage qui est une table de hachage où les localisations sont des clés et le conséquent est associé à la valeur correspondante pour chaque clé. Le score représente le support (fréquence) d'une localisation donnée l_n qui représente le nombre de fois où il apparaît dans le conséquent de séquences de mobilité similaires à St . Si les supports de deux localisations sont égaux, la confiance est utilisée et est définie comme le support (fréquence) de l_n divisé par le nombre total de séquences contenant l_n . Enfin, la localisation ayant le score le plus élevé dans le TC est considéré comme la localisation prédit.

2.7 Architecture de notre système de prédiction basé sur TCPT

Le modèle TCPT constitue le cœur de notre système de prédiction de localisation. Ce dernier adopte un processus qui prend comme entrée les données de mobilité des objets mobiles (ex. GPS, check-in, etc.) et donne en sortie les localisations susceptible d'être visiter par ces objets dans le future. Comme illustré dans Figure 11, le processus de prédiction consiste à 2 principales étapes : prétraitement, prédiction. La phase de prédiction est subdivisé de son tour en deux sous étapes apprentissage et prédiction.

7.1 Pré-traitement

Cette étape vise à préparer les données de mobilités pour la prédiction. A partir des données GPS, les séquences de mobilités temporelles sont générées et cela s'est fait comme suit. Tout d'abord, l'ensemble des points de résidence (séjour ou stay points) sont identifiés en basant sur deux seuils spatiaux et temporels prédéterminés nommées T_{thre} et D_{thre} . Les points extraits sont des localisations qui représentent des zones géographiques où la distance entre ces coordonnées GPS qui ne dépasse pas D_{thre} et la personne (objet mobile de façon générale) a resté pour une période inférieure de T_{thre} dans ces zones. Une fois les stay points sont identifié, les mouvements des personnes sont convertis en séquences de mobilité des localisations en faisant correspondre des coordonnées GPS aux localisations sur la carte géographique dans une un processus nommée Map-matching.

A la fin de cette étape, les séquences générées sont subdivisées en deux sous-ensembles nommés *données d'apprentissage* et *données prédiction* ou de *test*.

7.2 Prédiction de localisation

Pour prédire la prochaine localisation des personnes, deux sous-étapes doivent être effectués : l'apprentissage et la prédiction. Dans cette phase, les séquences générées dans l'étape précédente sont subdivisées en deux sous-ensembles nommés *données d'apprentissage* et *données de prédiction* ou de *test* sont utilisées, l'un est utilisé pour l'apprentissage et l'autre pour la prédiction.

7.2.1 Apprentissage

Cette étape utilise l'ensemble de *données d'apprentissage* pour construire le modèle TCPT (comme illustré dans section 2.6.1). A la fin de cette étape, la prédiction pourra être effectuée.

7.2.2 Prediction

Pour faire de la prédiction de la localisation future, cette étape utilise les *données de prédiction* et les structures de TCPT construites dans la phase d'apprentissage. L'utilisateur de notre modèles système doit définir DLT_j (les dernières localisations traversées) qu'il veut prédire quelle localisation viendra ensuite. Après ça Après cela, Le processus de prédiction est effectué en exécutant la phase de prédiction décrite dans la partie 2.6.2, qui considère la

localisation ayant le score le plus élevé enregistré dans la TC est ensuite sélectionné comme la localisation prédit.

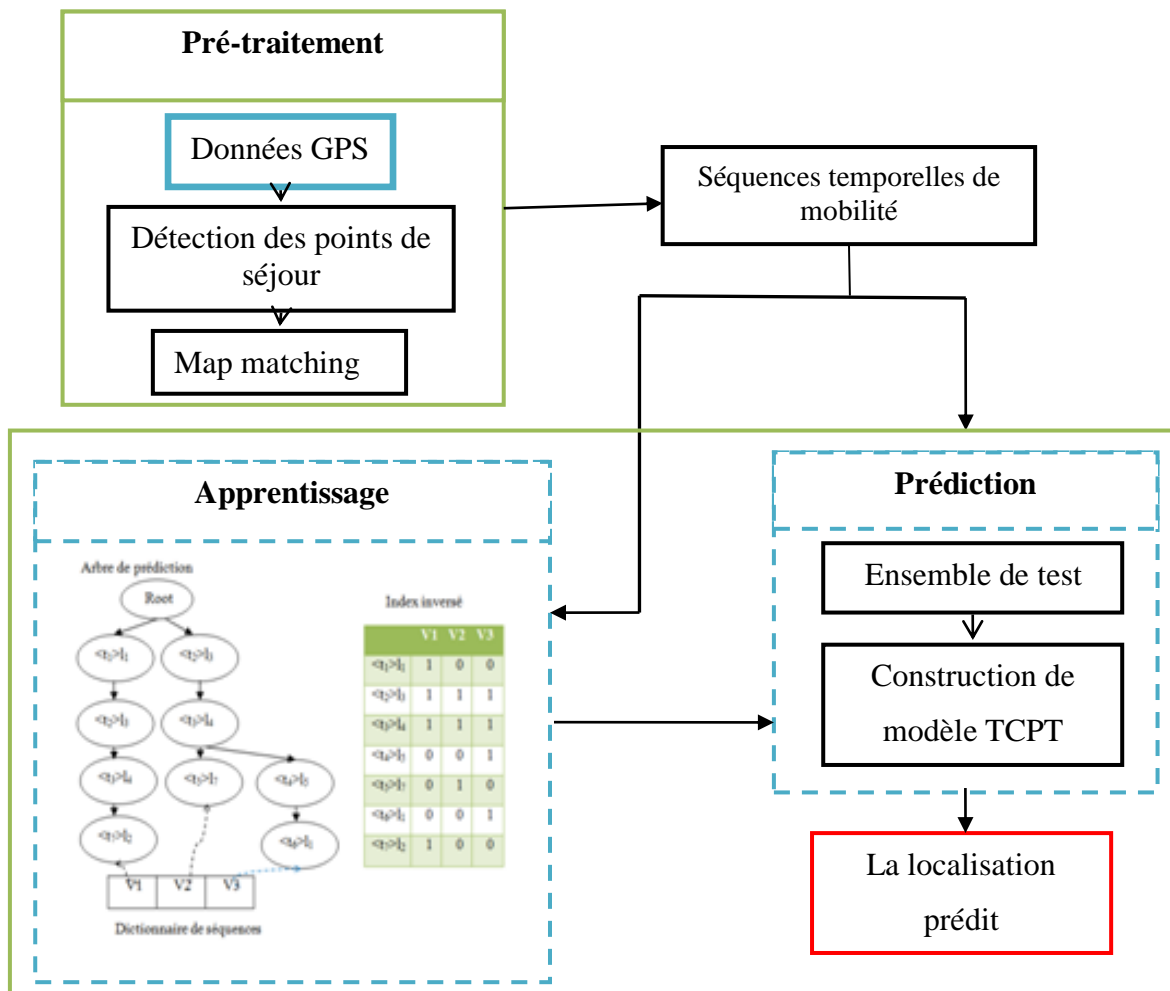


Figure 11. Architecture de TCPT

2.8 La considération de comportement globale et individuel de mobilité

Pour considérer le comportement de mobilité individuel et globale de la mobilité humaine, deux modèles sont dérivé de notre système.

A. Modèle global(MG)

Ce modèle a été suggéré en raison de la similarité dans le mouvement des personnes lors de leur déplacement entre les localisations.

Les gens montrent toujours un certain degré de similitude dans leurs mouvements lorsqu'ils voyagent à partir de lieux. Par exemple, les personnes qui se déplacent d'un emplacement *A* à un emplacement *B* ont tendance à emprunter presque les mêmes itinéraires pour atteindre *B*. Par conséquent, le comportement de mobilité globale doit être pris en compte.

Dans ce cas, notre système de prédiction basé sur notre système de prédiction basé sur TCPT est construit à partir des séquences de mobilité de toutes les personnes. Le modèle global est appliqué aux personnes nouvellement vues où leurs schémas de déplacement ne sont pas encore inclus et définis dans le système de prédiction.

B. Modèle personnel (MP)

Bien qu'il existe une similarité dans le mouvement des personnes, une personne peut également montrer un mouvement personnel spécifique en fonction de ses habitudes ou de ses désirs. En conséquence, le modèle de prédiction devrait se concentrer uniquement sur son comportement individuel en entraînant TCPT sur les données de mobilité de cette personne plutôt que sur les données de toutes les personnes. Par conséquent, une réduction significative en termes de temps et d'espace de prédiction peut être obtenue.

2.9 Conclusion

Dans cette partie, nous présentons notre modèle TCPT pour prédire le prochain emplacement. Ce dernier est basé sur un modèle non perdant appelé CPT (Compact Prediction Tree) et il prend également en considération le contexte temporel, comme nous l'avons expliqué son mécanisme de travail, c'est-à-dire comment il fait des prédictions. De plus, il examine le comportement de la mobilité individuelle et humaine en général à travers deux modèles, un modèle personnel et un modèle général.

Chapitre 3 :

Expérimentations

3.1 Introduction

Ce chapitre présente une étude expérimentale pour évaluer les performances de notre système de prédiction TCPT. Pour mieux étudier son efficacité, les performances de TCPT sont comparées avec d'autres modèles de prédiction notamment : 1^{ier} ordre Markov PPM (prediction with partial matching) [3] et Dependency Graph [42].

Notre modèle a été également comparé avec le modèle CPT de base sans la considération de facteur temporel. Le code source des modèles de prédiction comparés, ainsi que nos modèles, peuvent être téléchargés dans le cadre de la bibliothèque SPMF [43].

Dans ce chapitre, nous procédons comme suit. Tout d'abord, nous allons décrire 1) le jeu de données, 2) les paramètres expérimentaux, puis 3) les métriques d'évaluation pour mesurer les performances de TCPT. Par la suite, nous présentons les expérimentations menées et les résultats obtenus. Nous terminons ce chapitre par une conclusion.

3.2 Jeu de données

Dans cette étude expérimentale, nous avons utilisé le jeu de donnée Brightkite. Brightkite est un fournisseur de services de réseau social basé sur la localisation où les utilisateurs ont partagé leurs emplacements en s'enregistrant. Le réseau est à l'origine dirigé et comprend 58228 nœuds et 214078 arêtes. Il a également un total de 4491143 check-ins dans la période d'avril 2008 à octobre 2010.

Cette dataset a été utilisé dans plusieurs d'autres travaux de la littérature [44] pour la recommandation des localisations.

Dans nos expérimentations, la dataset est préparée puis divisés en ensembles d'apprentissage et de test en appliquant 10-fold cross-validation.

3.3 Paramètres d'expérimentation

Notre environnement d'expérimentation est composé d'un processeur Core Intel CPU 1.70 GHz avec 4 Go de RAM avec Windows7 (32bit). Le modèle proposé a été implémenté en Java en adaptant l'implémentation de CPT disponible dans la bibliothèque SPMF [43].

3.4 Métriques d'évaluation

Pour mesurer les performances de TCPT, quatre mesures d'évaluation ont été utilisées notamment: taux de succès, taux d'échec, couverture et précision globale. Ces métriques sont définies comme suit:

- ❖ **Taux de succès.** Il estime le nombre de séquences où une prédiction réussite a été faite sur le nombre total de prédictions accomplies.

$$Tauxdesuccès = \frac{\text{Nombre de prédiction réussites}}{\text{Nombre de prédictions}}$$

- ❖ **Taux d'échec.** Il détermine le nombre de prédictions incorrectes sur le nombre total de prédictions effectuées.

$$Tauxd'échec = \frac{\text{Nombre de prédiction incorrectes}}{\text{Nombre de prédictions}}$$

- ❖ **Couverture.** Elle estime le nombre de séquences de mobilité de test où des séquences similaires sont trouvées (matching sequences) et utilisées pour la prédiction. Cette métrique est calculée à partir du nombre de prédiction NoMatch qui indique le nombre de séquences où aucune séquence similaire n'a été trouvée, divisé par le nombre total de séquences de test.

$$Couverture = 1 - \frac{\text{Nombre de prédiction NoMatch}}{\text{Nombre de séquences de test}}$$

- ❖ **Précision globale ou accuracy.** Elle est considérée comme un indicateur de performance global qui mesure l'efficacité de la prédiction. Il est défini comme le nombre de localisations prédites avec succès, divisé par le nombre total de séquences de mobilité de test.

$$Précision\ globale = \frac{\text{Nombre de prédiction correctes}}{\text{Nombre de séquences de mobilité de test}}$$

3.5 Expérimentations

5.1 Modèles évalués

Notre modèle a été comparé avec les modèles de prédiction suivant:

- ***CPT (Compact Prediction Tree)***

C'est le modèle sur lequel nous nous basons dans notre travail, mais sans tenir en compte le contexte temporel. L'une de ses caractéristiques est qu'il est sans perte (il peut utiliser toutes les informations des séquences d'apprentissage pour faire une prédiction).

- ***DG (Dependency Graph où Graphe de dépendance)***

DG [42] est un modèle de prédiction qui a été utilisé pour prédire la prochaine page Web qu'un utilisateur visitera en fonction des pages Web précédemment visitées par cet utilisateur et d'autres utilisateurs. Il est représenté par un graphe orienté représentant les dépendances de plusieurs éléments les uns par rapport aux autres. Il se compose d'un ensemble de nœuds et d'arcs dirigés entre eux. où chaque nœud représente un élément et l'arc directionnel relie un nœud A à un nœud B si et seulement si B apparaît dans les éléments de A dans les séquences d'apprentissage.

- ***PPM ou Markov de premier ordre***

Ce modèle est basé sur la propriété Markovienne que les prédictions basent sur les derniers éléments d'une séquence. Il peut être représenté sous la forme d'un graphique où les sous-séquences préfixes sont liées aux sous-séquences suffixes par des arcs sortants ayant des probabilités de transition.

- ***ContextMark1 (Markov de premier ordre avec temps)***

Pour bien étudier les performances de notre modèle en considérant le contexte temporel, nous avons aussi étendu le modèle PPM (Markov de premier ordre) pour qu'il considère l'aspect temps de mobilité.

5.2 Expérimentation1 (étude de Scalabilité)

Dans cette expérience, nous avons étudié la scalabilité de notre modèle et les autres modèles de prédiction en augmentant le nombre de séquences de mobilités considérées de 100 à 1000 séquences. La scalabilité est un facteur important pour la plupart des tâches de prédiction, car la possibilité de mettre à l'échelle un modèle de prédiction peut limiter directement ou indirectement sa précision et sa couverture. Les résultats de cette expérimentation sont présentés dans Figure 12.

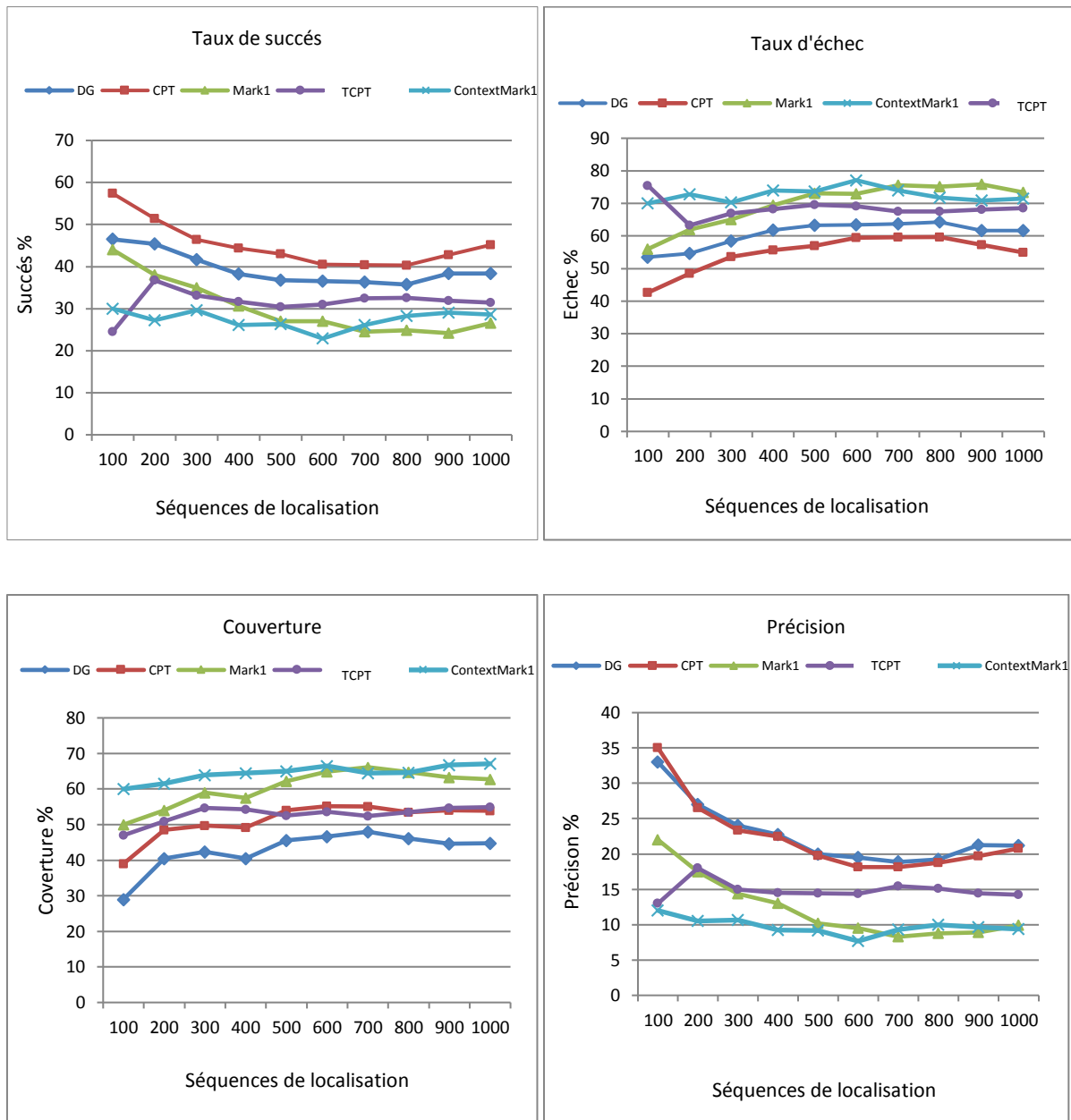


Figure 12. Résultats de l'étude de scalabilité.

A) La comparaison de TCPT avec ContextMark1

A partir des résultats obtenus, nous avons observé que notre modèle TCPT a présenté des performances meilleures par rapport au ContextMarkov1 en termes de toutes les métriques d'évaluation sauf la couverture. Cela peut être dû au fait que TCPT est un modèle sans perte où il utilise toutes les informations de la séquence d'apprentissage pour faire une prédiction à la différence de ContextMarkov1 qui est principalement basé sur l'hypothèse markovienne qui consiste à prendre en compte que la localisation actuelle pour faire de prédiction.

B) La comparaison avec d'autres modèles

Les résultats ont démontré aussi que notre modèle TCPT a présenté généralement des performances inférieures à celles des modèles CPT et DG mais supérieure à celle de Markov1 dans certains cas.

La précision de TCPT était faible par rapport aux autres modèles pour un nombre de séquence entre 100 et 200 séquences, mais elle a augmenté rapidement en augmentant le nombre de séquences en surpassant les performances de ContexteMarkov1.

Quant au taux d'échec, on constate qu'il est important, mais il diminue rapidement avec l'augmentation des séquences, et cela est dû à l'augmentation de la similarité et de la régularité de la mobilité humaine en augmentant le nombre des séquences.

On peut justifier les performances relativement basses de TCPT à plusieurs raisons notamment :

1- Irrégularité partielle de mobilité des humains : malgré l'hypothèse de la régularité spatiale et temporelle de la mobilité humaine, les individus peuvent présenter une irrégularité. Une personne peut changer ses habitudes dans différentes heures de la journée, ou à différents jours de la semaine, ou à différents mois de l'année, car il peut aller à son poste de travail durant la semaine, mais durant le week-end il fréquente généralement des lieux de distraction comme les plages et les stations balnéaires en été plutôt qu'en hiver.

2- Préférences diversifiés des personnes : Chaque personne a ses propres habitudes, par exemple, il y a ceux qui préfèrent se divertir et passer du temps avec des amis le week-end, et certains d'entre eux préfèrent rester à la maison pour se reposer.

3- La sévérité temporelle de TCPT : Les personnes peuvent visiter les mêmes localisations mais à des moments différents. Par exemple, deux personnes A et B visitent toujours le même magasin Z mais pas forcément dans le même temps (A visite Z à 10h et B à 20h). Cela est la principale raison derrière la supériorité des modèles sans temps (DG et CPT) par rapport à TCPT.

En ce qui concerne la couverture, TCPT a obtenu de bons résultats et meilleurs que CPT et DG. La couverture est une mesure importante car elle mesure le nombre de fois qu'un modèle de prédiction puisse faire une prédiction. Plusieurs études ont montré qu'il est préférable que le modèle de prédiction puisse prédire même s'il s'agit d'un échec.

5.3 Expérimentation 2 (Impact de la variation de la durée de time-slot)

Cette expérimentation vise à étudier l'effet de la taille de la plage horaire (time-slot) sur TCPT et ContextMarkov1. Pour faire cela, on a varié la taille des time-slots de 2 à 12 pour diviser le jour en des périodes de longueur spécifiée pour un nombre de séquence de 1000 séquences temporelles de mobilité.

Les résultats présentés dans Figure 13 indiquent que TCPT présente des résultats meilleurs par rapport à ContextMarkov1 dans presque tous les cas.

D'autre part, on a constaté que le taux de succès et la précision atteignent leurs valeurs les plus élevées lorsque les tailles de time-slots sont grandes (entre 10 et 12) pour les deux modèles, où la précision atteint 18% et 10% pour TCPT et ContextMarkov1, respectivement.

Cela peut être dû au fait que l'augmentation de la taille de time slot desserre la sévérité de temps de visite d'une localisation. Une personne peut visiter la même localisation qu'un profil utilisateur dans notre système mais avec une flexibilité dans le temps exact de visite.

De plus, les résultats ont montré une augmentation des valeurs de taux d'échec ont été enregistrées pour des valeurs de time-slot de 6 à 10 heures mais ces valeurs ont été diminuées lorsque la taille des time-slots augmente. En ce qui concerne la couverture, les deux modèles ont obtenu de bonnes valeurs de couverture surtout en augmentant la taille de time slot.

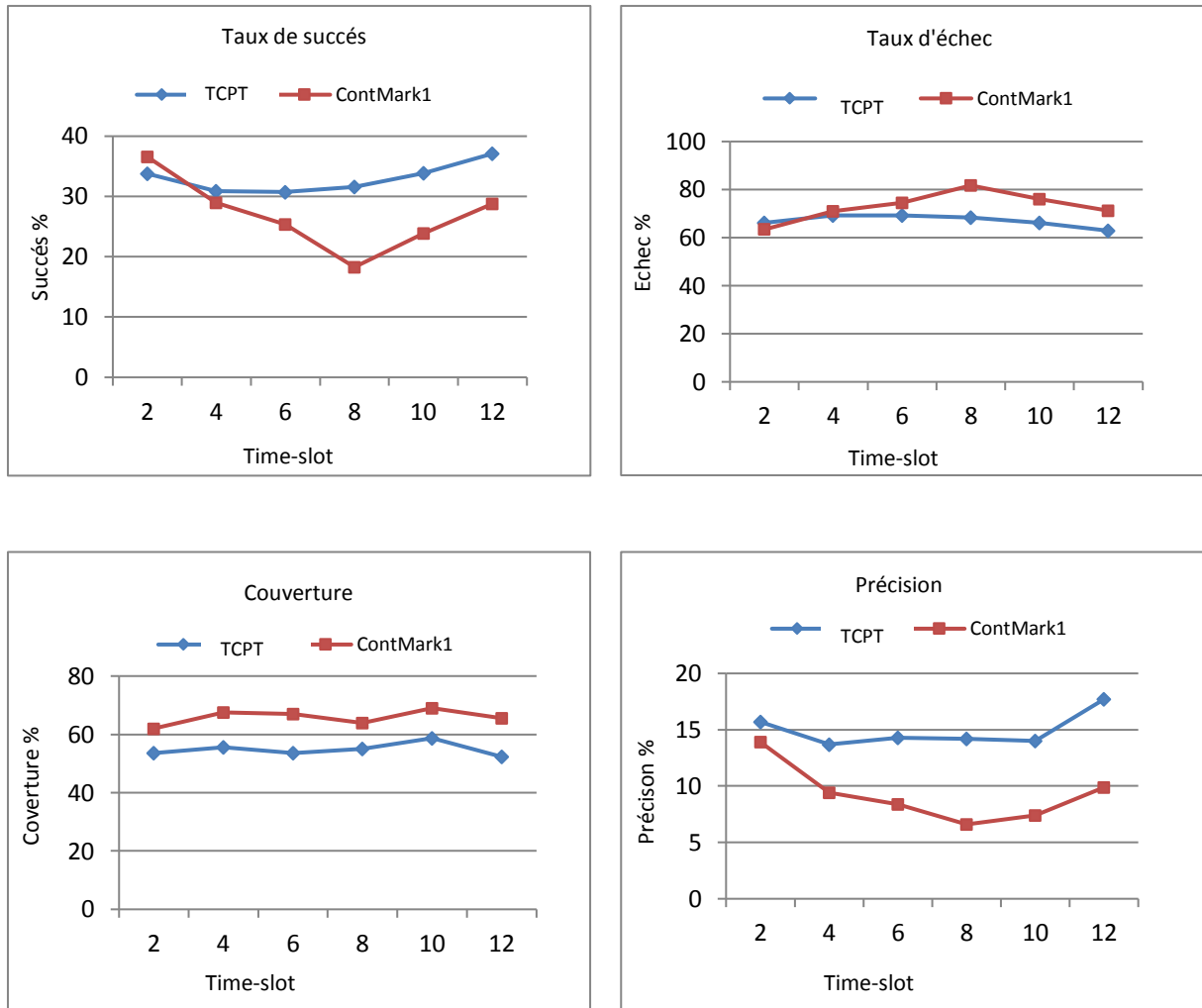


Figure 13. L'impact de la variation de time-slot.

3.6 Conclusion

Dans ce chapitre, nous avons mené un ensemble d'expérimentation afin de comparer les performances de notre modèle (TCPT) avec d'autres modèles de prédiction en utilisant un jeu de données réel (Brightkite).

Les résultats expérimentaux ont montré que les performances de TCPT sont généralement bonnes où elles ont dépassé certains d'entre eux comme ContexteMarkov1 dans plusieurs cas.

À l'avenir, nous souhaitons étendre notre modèle pour inclure d'autres informations contextuelles en plus du contexte temporel, telles que l'âge, le sexe et les informations météorologiques, car cela pourra contribuer à l'augmenter de la précision des prédictions. Nous avons également l'intention de prédire une série d'emplacements à venir au lieu de simplement prédire qu'un seul prochain emplacement.

Conclusion générale

Dans ce mémoire, nous avons proposé un modèle de prédiction de localisation appelé TCPT qui repose sur un le modèle de prédiction sans perte (CPT).CPT a été étendu en TCPT pour inclure le contexte temporel, contrairement à la plupart des modèles, qui sont concernés que par le contexte spatial et négligent généralement l'extension temporelle.

Une étude expérimentale a été réalisée avec un jeu de données réel (Brightkite). Les résultats obtenus ont montré que les performances de TCPT sont relativement bonnes par rapport aux modèles de prédiction en terme de précision, taux de succès, et couverture.

En ce qui concerne les travaux futurs, nous visons à améliorer notre modèle et à étendre sa capacité de prédiction pour qu'il soit capable à prédire plusieurs futures localisations (long term prediction) et non pas juste une localisation (short term prediction). De plus, nous visons à étendre TCPT pour inclure d'autres contextes temporels (par exemple, le jour de la semaine, saison de l'année) et pas seulement cela, mais également d'autres contextes tels que le genre et les conditions météorologiques.

Bibliographie

- [1] W. Mathew, R. Raposo, and B. Martins, “Predicting Future Locations with Hidden Markov Models,” 2012.
- [2] T. Mikluš, M. Gregor, and A. Janota, “Using Neural Networks for Route and Destination Prediction in Intelligent Transport Systems,” pp. 380–387, 2012.
- [3] V. S. Tiwari, A. Arya, and S. Chaturvedi, “Scalable prediction by partial match (PPM) and its application to route prediction,” *Appl. Informatics*, 2018, doi: 10.1186/s40535-018-0051-z.
- [4] L. Chen, M. Lv, and G. Chen, “A system for destination and future route prediction based on trajectory mining,” *Pervasive Mob. Comput.*, vol. 6, no. 6, pp. 657–676, 2010, doi: 10.1016/j.pmcj.2010.08.004.
- [5] T. Gueniche, P. Fournier-Viger, and V. S. Tseng, “Compact prediction tree: A lossless model for accurate sequence prediction,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8347 LNAI, no. PART 2, pp. 177–188, 2013, doi: 10.1007/978-3-642-53917-6_16.
- [6] N. A. Amirrudin, S. H. S. Ariffin, N. N. N. Abd Malik, and N. E. Ghazali, “Mobility Prediction via Markov Model in LTE Femtocell Mobility Prediction via Markov Model in LTE Femtocell,” no. January, 2013.
- [7] S. Parija, P. K. Sahu, S. K. Nanda, and S. S. Singh, “A functional Link Artificial Neural Network for Location Management in Cellular Network,” pp. 1160–164, 2013.
- [8] K. Yap and Y. Chong, “Optimized Access Point Selection with Mobility Prediction Using Hidden Markov Model for Wireless Network,” no. 304, pp. 38–42, 2017.
- [9] F. Simini, M. C. González, A. Maritan, and A. L. Barabási, “A universal model for mobility and migration patterns,” *Nature*, vol. 484, no. 7392, pp. 96–100, 2012, doi: 10.1038/nature10856.
- [10] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, “Limits of predictability in human mobility,” *Science (80-.)*, vol. 327, no. 5968, pp. 1018–1021, 2010, doi: 10.1126/science.1177170.
- [11] P. Hui and J. Crowcroft, “Human mobility models and opportunistic communications system design,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 366, no. 1872, pp. 2005–2016, 2008, doi: 10.1098/rsta.2008.0010.
- [12] J. P. Bagrow and Y. R. Lin, “Mesoscopic structure and social aspects of human mobility,” *PLoS One*, vol. 7, no. 5, pp. 3–8, 2012, doi: 10.1371/journal.pone.0037676.
- [13] S. M. Qin, H. Verkasalo, M. Mohtaschemi, T. Hartonen, and M. Alava, “Patterns, Entropy, and Predictability of Human Mobility and Life,” *PLoS One*, vol. 7, no. 12, 2012, doi: 10.1371/journal.pone.0051353.
- [14] Y. Wang, N. J. Yuan, D. Lian, and L. Xu, “Regularity and Conformity : Location Prediction Using Heterogeneous Mobility Data,” 2015.
- [15] M. J. Williams, R. M. Whitaker, and S. M. Allen, “Measuring Individual Regularity in Human Visiting Patterns,” 2012, doi: 10.1109/SocialCom-PASSAT.2012.93.
- [16] V. M. De Lira and S. Rinzivillo, “Investigating semantic regularity of human mobility lifestyle,” pp. 314–317, 2014.
- [17] M. Verleysen, V. D. Blondel, Y. De Montjoye, and C. A. Hidalgo, “Unique in the Crowd: The privacy bounds of human mobility,” pp. 1–5, 2013, doi: 10.1038/srep01376.
- [18] D. Wang, D. Pedreschi, C. Song, F. Giannotti, A. Barabási, and C. Science, “Human Mobility , Social Ties , and Link Prediction Categories and Subject Descriptors,” pp. 1100–1108.
- [19] Z. Li, R. Kays, and P. Nye, “Mining Periodic Behaviors for Moving Objects,” pp. 1099–1108, 2010.

- [20] C. Song, T. Koren, P. Wang, and A. L. Barabási, “Modelling the scaling properties of human mobility,” *Nat. Phys.*, vol. 6, no. 10, pp. 818–823, 2010, doi: 10.1038/nphys1760.
- [21] W. Liao, “An Object Tracking Scheme for Wireless Sensor Networks Using Data Mining Mechanism,” pp. 526–529, 2012.
- [22] L. N. Balico, A. A. F. Loureiro, E. F. Nakamura, R. S. Barreto, R. W. Pazzi, and H. A. B. F. Oliveira, “Localization Prediction in Vehicular Ad Hoc Networks,” no. c, 2018, doi: 10.1109/COMST.2018.2841901.
- [23] E. Cho, S. A. Myers, and J. Leskovec, “Friendship and mobility: User movement in location-based social networks,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1082–1090, 2011, doi: 10.1145/2020408.2020579.
- [24] C. Cheng, H. Yang, M. R. Lyu, and I. King, “Where you like to go next: Successive point-of-interest recommendation,” *IJCAI Int. Jt. Conf. Artif. Intell.*, no. January, pp. 2605–2611, 2013.
- [25] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, “Factorizing Personalized Markov Chains for Next-Basket Recommendation,” *Proc. 19th Int. Conf. World Wide Web, WWW '10*, pp. 811–820, 2010.
- [26] H. Gao, J. Tang, X. Hu, and H. Liu, “Exploring temporal effects for location recommendation on location-based social networks,” *RecSys 2013 - Proc. 7th ACM Conf. Recomm. Syst.*, pp. 93–100, 2013, doi: 10.1145/2507157.2507182.
- [27] A. Farouk, M. Samer, A. Boukerche, and A. Mammeri, “Vehicular Route Prediction in VANETs,” pp. 788–802, 2013, doi: 10.1007/s11036-013-0459-6.
- [28] A. Auvolat, “Artificial Neural Networks Applied to Taxi Destination Prediction Alexandre de Brébisson Étienne Simon,” no. December, 2017, doi: 10.13140/RG.2.2.20264.26888.
- [29] Q. Liu, S. Wu, L. Wang, and T. Tan, “Predicting the Next Location : A Recurrent Model with Spatial and Temporal Contexts,” pp. 194–200, 2016.
- [30] J. J. Ying, W. Lee, T. Weng, and V. S. Tseng, “Semantic Trajectory Mining for Location Prediction,” pp. 34–43, 2011.
- [31] M. Chen, Y. Liu, and X. Yu, “Predicting Next Locations with Object Clustering and Trajectory Clustering,” vol. 2, pp. 344–356, 2015, doi: 10.1007/978-3-319-18032-8.
- [32] Z. Z. T. Ho, P. Conference, and R. Goebel, *Advances in Knowledge Discovery*. 2015.
- [33] F. Terroso-saenz, M. Valdes-vela, and A. F. Skarmeta-gomez, “Online route prediction based on clustering of meaningful velocity-change areas,” *Data Min. Knowl. Discov.*, 2016, doi: 10.1007/s10618-016-0452-3.
- [34] N. Miguel, “Next Place Prediction using Mobility Markov Chains,” pp. 0–5, 2012.
- [35] M. Chen, Y. Liu, and X. Yu, “NLPMM : A Next Location Predictor with Markov Modeling,” no. 1, pp. 186–197, 2014.
- [36] J. P. Epperlein, J. Monteil, M. Liu, Y. Gu, S. Zhuk, and R. Shorten, “Bayesian classifier for Route prediction with Markov chains,” *2018 21st Int. Conf. Intell. Transp. Syst.*, pp. 677–682, 2020.
- [37] P. Helmholz, E. Ziesmann, and S. Robra-bissantz, “Context-Awareness in the Car: Prediction , Evaluation and Usage of Route Trajectories Context-Awareness in the Car: Prediction , Evaluation and Usage of Route Trajectories,” no. November 2014, 2013, doi: 10.1007/978-3-642-38827-9.
- [38] R. Bar-david and M. Last, “Context-Aware Location Prediction,” pp. 165–185, 2016, doi: 10.1007/978-3-319-29009-6.

- [39] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, “NextPlace: A Spatio-temporal Prediction Framework for Pervasive Systems,” pp. 152–169, 2011.
- [40] H. Amirat, N. Lagraa, P. Fournier-Viger, and Y. Ouinten, “NextRoute: a lossless model for accurate mobility prediction,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 7, pp. 2661–2681, 2020, doi: 10.1007/s12652-019-01327-w.
- [41] Q. Lv and Y. Qiao, “Big Data Driven Hidden Markov Model Based Individual Mobility Prediction at Points of Interest,” vol. 9545, no. c, pp. 1–12, 2016, doi: 10.1109/TVT.2016.2611654.
- [42] N. U. Edakunni, A. Raghunathan, A. Tripathi, J. Handley, and F. Roulland, “Probabilistic Dependency Networks for Prediction and Diagnostics,” no. June, 2015, [Online]. Available: <http://arxiv.org/abs/1508.03130>.
- [43] A. Gomariz, T. Gueniche, and P. Fournier-viger, “SPMF: A Java Open-Source Pattern Mining Library,” vol. 15, pp. 3569–3573, 2014.
- [44] H. Amirat and P. Fournier-viger, “LocRec: Rule-based Successive Location Recommendation in LBSN,” *2018 IEEE Int. Conf. Commun.*, pp. 1–6, 2018.