

Modélisation d'objets mobiles dans un entrepôt de données

Tao Wan, Karine Zeitouni

Laboratoire PRISM, Université de Versailles
45, avenue des Etats-Unis, 78035 Versailles Cedex, France
Tao.Wan@prism.uvsq.fr, Karine.Zeitouni@prism.uvsq.fr
<http://www.prism.uvsq.fr/users/karima/>

Résumé. La gestion d'objets mobiles a connu un regain d'intérêt ces dernières années, particulièrement dans le but de gérer et de prédire la localisation d'objets mobiles. Cependant, il y a peu de recherches sur l'exploitation d'historiques de bases d'objets mobiles. L'extraction de connaissances depuis ce type de données serait pourtant utile aux applications décisionnelles, car elle permettrait l'analyse spatio-temporelle à différentes échelles et l'agrégation flexible passant du niveau micro au niveau macro et vice-versa. La première étape dans ce processus est la mise en œuvre d'un entrepôt d'objets mobiles. Seulement, les modèles d'entrepôts existants ne permettent pas de traiter directement ce type de données complexes. Cet article présente une approche originale pour pallier ce problème. Cette approche offre la puissance de l'algèbre OLAP sur toute combinaison de données classiques, spatiales et/ou temporelles et mobiles. Elle a été validée par un prototype et appliquée à l'analyse de la mobilité urbaine¹. Les résultats de l'expérimentation montrent la validité de l'approche et les tests de performances son efficacité.

1. Introduction

Le développement des technologies mobiles, telles que les téléphones cellulaires et les GPS, a ouvert la voie vers de nouvelles applications exploitant la localisation. En effet, cette connaissance peut permettre de cibler les services offerts – appelés Location Based Service (LBS)- comme de fournir des informations localisées dans la zone du mobile. D'autres applications utilisent des objets mobiles comme le contrôle et la prévision du trafic basés sur les trajectoires de véhicules. Ces applications ont généré de nouveaux problèmes qui ont fait naître ou accéléré la recherche sur la gestion d'objets mobiles.. La plupart des travaux sont axés sur la modélisation d'objet mobiles (Güting et al. 2000, Vazirgiannis et al. 2001) les méthodes d'accès (Pfoser et al. 2000, Saltenis et al. 2000), les requêtes prédictives et l'optimisation des mise à jour (Chon et al. 2002, Tao et al 2002, Jensen 2004).

¹ Ce travail est partiellement financé par le projet HEARTS (Health Effects and Risk of Transport Systems), co-financé par le programme énergie, environnement et développement durable de la Commission Européenne (contrat n°: QLK4-CT-2001-00492). Cet article ne reflète pas nécessairement l'opinion officielle de la Commission Européenne, d'autres institutions de la Communauté Européenne ou de l'OMS. Ni la Commission Européenne, ni l'OMS ne sont responsables de l'usage pouvant être fait du contenu de cet article.

Ces techniques ne s'appliquent pas dans un contexte décisionnel où l'on s'intéresse aux données historiques sur les objets mobiles. Pourtant, l'analyse a posteriori des phénomènes mobiles serait bien utile dans les domaines de la planification du transport, des demandes de services mobiles, etc. Les travaux sur les entrepôts de données spatiaux (Stefanovic et al. 2000, Rao et al. 2003) et spatio-temporels (Marchand et al. 2004) ne considèrent que le cas d'une localisation fixe et pas celui des objets mobiles. Dans (Papadias 2002), les auteurs traitent un modèle multidimensionnel se référant aux objets mobiles où une cellule mesure le nombre de mobiles dans une unité espace-temps. Ils le stockent dans un index, combinaison du R-tree et du B-tree. Ensuite, (Tao et al. 2004) soulève le problème du double comptage des objets mobiles, qui traversent plusieurs unités espace-temps et sont comptés dans plusieurs cellules, qu'ils résolvent par une estimation probabiliste. Leur modèle part de l'hypothèse que les objets mobiles sont agrégés et ne permet pas de représenter d'autres dimensions que l'espace et le temps.

La gestion dans les entrepôts des objets mobiles avec toutes leurs dimensions reste un problème ouvert. Cet article propose une solution originale permettant d'exploration multidimensionnelle des objets mobiles. Il offre la puissance de l'algèbre OLAP sur toute combinaison de données classiques, spatiales et/ou temporelles et mobiles. Cette approche a été validée par un prototype et appliquée à l'analyse de la mobilité urbaine. Après une brève présentation du contexte et la précision des objectifs dans la section 2, la section 3 propose notre approche pour répondre à ces objectifs, suivie par les résultats de l'expérimentation. Nous concluons en soulignant les contributions et projetons les travaux futurs.

2. Contexte et objectifs

Le projet HEARTS (<http://www.euro.who.it/hearts>) a pour objet l'analyse des risques pour la santé liés au transport routier. Ces risques dépendent pour beaucoup de la mobilité – lieux et instants – des individus, en plus d'autres dimensions comme le mode de transport ou l'âge. L'analyse multidimensionnelle et spatio-temporelle du phénomène de la mobilité est essentielle dans ce projet. La mobilité de la population urbaine est décrite par des objets mobiles, extraits d'une enquête nationale dite « ménages et déplacements ».

Il faut pour cela développer un entrepôt de données relatif aux objets mobiles. Celui-ci doit permettre d'explorer toutes les dimensions de la mobilité comme ici les caractéristiques des personnes mobiles (âge, sexe ou catégorie socioprofessionnelle) ou le mode de transport ou le type d'activité ; mais en plus il doit considérer le temps et l'espace (l'horaire et le lieu de passage). Cette exploration doit se faire à différents niveaux de granularité ou à différentes échelles spatiales et temporelles. De plus, les objets mobiles et les objets spatio-temporels sont fortement interdépendants. Dans notre exemple, les trajectoires des individus d'un côté et les mesures localisées de trafic, de pollution et de bruit de l'autre sont clairement liées. L'analyse multidimensionnelle doit donc intégrer hormis les objets mobiles, des faits spatiaux ou spatio-temporels à localisation fixe mais avec des mesures variant dans le temps. Ces différents phénomènes peuvent avoir un référentiel spatial différent et des granularités spatiales et temporelles différentes. Par exemple, le volume de trafic est mesuré sur les sections de routes tandis que le champ de pollution est donné sur une grille.

3. Approche proposée

Tout d'abord, on définit un modèle spatio-temporel d'objets mobiles tout en précisant les hypothèses pour son application. Ensuite, le processus de conception est décrit à la lumière du projet HEARTS. Le but est d'explorer les activités humaines selon les dimensions mobilité spatio-temporelles, catégories socioprofessionnelles, âge etc. Ce processus peut s'appliquer de manière analogue aux applications qui vérifient les hypothèses énoncées.

3.1 Trajectoires contraintes

Formellement, un objet mobile peut être défini par : *ObjetMobile* (*#ID*, *A1*, *A2*, ...*An*, *TR*) où *#ID* est un identifiant, *Ai* sont des attributs et *TR* une trajectoire (Vazirgiannis 2001). Une trajectoire forme une variation continue dans l'espace et dans le temps, parfois représentée en machine par un échantillon de localisations et instants exacts de passage. Or, dans un modèle multidimensionnel, les valeurs des dimensions doivent se limiter à quelques modalités et être largement partagées par les faits représentés. Une représentation exacte de l'espace et du temps ne peut donc former une dimension. D'où l'idée d'une représentation grossière partageant des unités spatiales et temporelles fixées à l'avance. Une trajectoire est alors définie par référence comme un ensemble $T_{Ref} = \{(\#S1, \#T1), \dots\}$ où *#Si* référence l'unité spatiale traversée par *TR* à un instant de l'intervalle *Ti*. A cette représentation grossière, on propose d'associer des mesures sur le taux de recouvrement (*MSi*) de la trajectoire avec l'unité spatiale *Si* et la durée (*MTi*) effective par rapport à l'intervalle *Ti*. Une unité spatiale peut être une cellule d'une grille raster, une zone d'un découpage polygonal ou une polyligne d'un réseau. Très souvent, ces unités de référence existent au préalable et sont parfois organisée en une hiérarchie spatiale. De plus, une trajectoire de mobile est souvent contrainte dans l'espace, par exemple la trajectoire d'un véhicule est contrainte par le réseau linéaire de transport. Le choix de l'unité spatiale de référence est alors naturel. On se place dans l'hypothèse de trajectoires contraintes et d'unités spatiales prédéfinies.

3.2 Schéma multidimensionnel

Le schéma précédent peut être vu comme un schéma relationnel NF2, car *TRef* est non atomique. On peut le ramener en 1ère forme normale relationnelle de la manière suivante : *ObjetMobileNorm* (*#ID*, *A1*, *A2*, ...*An*), *TRefElement* (*#ID*, *#S*, *#T*, *MS*, *MT*). Il suffirait de considérer cette table comme table de faits et les autres comme dimensions, éventuellement décomposées à leur tour en d'autres dimensions, pour aboutir à un schéma en flocons. Appliqué à notre exemple, on obtient le schéma en figure 1. Dans ce schéma, la table de fait est représentée par la table *Mobilité* décrivant le fait qu'un trajet passe par telle localisation à tel instant. Les tables de dimensions sont *Temps* et *Localisation* d'un côté, puis de l'autre, *Trajet*, *Mode*, *Déplacement*, *Motif* (d'activité), *Personne*, *Age* et *Profession*. Certains attributs de dimensions sont donnés à différents niveaux de détail. C'est le cas des zones, du temps, des motifs ou de l'âge par exemple. L'exploration OLAP est particulièrement adaptée et permet un changement interactif de niveau, d'ordre et de combinaison de dimensions.

L'expérimentation a porté sur l'enquête « ménages –déplacements » qui contient un échantillon de 10,354 individus et décrit les activités et déplacement pendant une journée au hasard. La table de faits contient près de 1200 000 tuples. Une personne traverse un nombre de zones variant entre 1 et 108 pendant cette journée. Dans la dimension spatiale, il y a 631 zones définies au niveau le plus fin, pouvant être agrégées en deux autres niveaux prédéfinis.

Modélisation d'objets mobiles dans un entrepôt de données

La table de dimension temps est définie avec trois niveaux : intervalles d'un quart heure, d'une demi-heure et d'une heure. La table de dimension trajet compte environ 50,000 tuples.

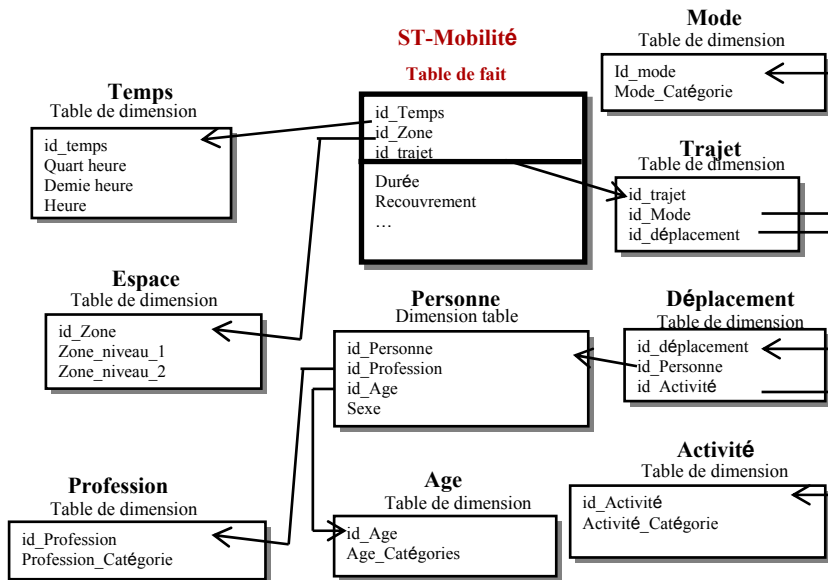


Fig 1 – Aperçu du modèle multidimensionnel dans HEARTS

3.3 Implémentation et exemples de requêtes

Ce schéma a été implémenté sous Microsoft SQL Server 2000. Il permet de générer différents cubes de données, par exemple, donnant le nombre de personnes par zone et par heure afin de pouvoir estimer les lieux et les horaires d'exposition aux risques. Ceci est illustré dans la figure 2 pour certaines heures pour les 67 zones agrégées. Elle montre la variation de déplacement des personnes (élevée vers midi et entre 18 et 19h) et (très basse à minuit). Elle met en évidence la différence de traversées des zones. Par combinaison de la dimension spatiale et des dimensions non spatiales (ici l'âge), l'exemple 3 calcule le nombre de personnes par catégorie d'âge traversant chaque zone.

Ces résultats peuvent être obtenus par des requêtes relationnelles, mais de manière moins flexible et beaucoup moins efficace qu'une approche OLAP. En effet, les requêtes OLAP sont interactives permettant leur utilisation par des non informaticiens. Elles sont aussi optimisées, par exemple, le temps de réponse maximum sur la base décrite plus haut est de 2 mn sous SQL Server 2000 et sur un PC Pentium 2 400 Mhz équipée de 256MB de RAM.

4. Conclusion

Nous avons proposé dans cet article une approche originale de modélisation d'un entrepôt d'objets mobiles pour laquelle il n'existe pas à notre connaissance de travaux équivalents. Le modèle proposé se base sur une représentation simplifiée des trajectoires spatio-temporelles.

Cette représentation fait référence à des unités spatiales (découpage préexistant) et des unités temporelles fixes. Des mesures permettent de pondérer l'occupation de l'espace et du temps.

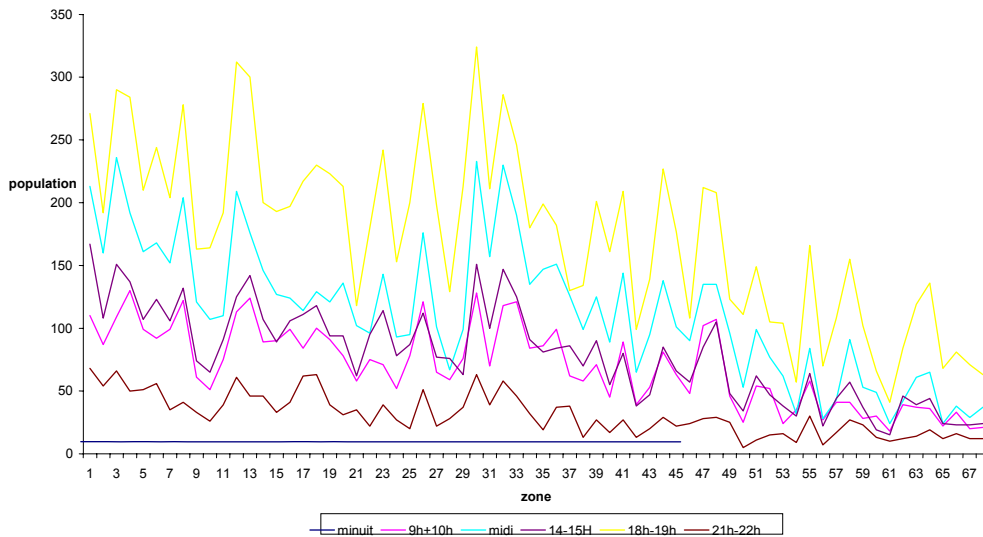


Fig 2 - la population par zone à certaines heures

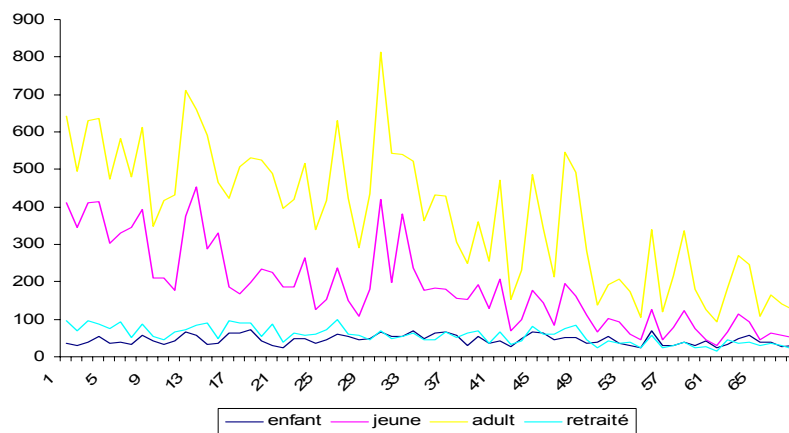


Fig 3 - la population journalière par zone groupée par classe d'âge

Ce modèle présente l'avantage de pouvoir être implémenté avec les outils du marché. SQL Server 2000 a été utilisé dans notre cas. Une application dans l'analyse de la mobilité urbaine a permis la validation de cette approche sur des données réelles.

Cette approche s'est révélée efficace et malgré le volume de données important, donne des performances d'exécution acceptables. Dans nos prochains travaux, nous essaierons d'étudier plus précisément le passage à l'échelle en générant des données synthétiques plus volumineuses. Un autre soucis est de combiner des requêtes de type OLAP et des requêtes spatiales : faut-il un nouvel index spatio-temporel ou peut-on simplement ramener l'accès

aux objets mobiles à l'accès au découpage de référence ? Enfin, notre représentation simplifiée pose le problème de la perte d'information sur les trajectoires. Il est possible d'utiliser un découpage très fin, mais cela engendrerait un volume de données gigantesque et affecterait les performances. La question est jusqu'où peut-on utiliser les modèles multidimensionnels conventionnels ? La représentation de dimensions réellement complexes et de modélisation « abstraite » d'objets mobiles dans une nouvelle forme d'entrepôt serait peut-être la voie de recherche à explorer dans l'avenir.

Références

- Chon H. D., Agrawal D., Abbadi A. E. (2000), Query Processing for Moving Objects with Space-Time Grid Storage Model. *In Proc of Mobile Data Management (MDM)*, pp. 121, 2002.
- Güting R., Böhlen M., Erwig M., Jensen C., Lorentzos N., Schneider M., Vazirgiannis M. Z. (2000), Foundation for Representing and Querying Moving Objects. *In Proc of ACM Transactions on Database Systems (TODS)*, Volume 25, Issue 1, pp. 1-42, 2000.
- Jensen C., Lin D., Ooi B.C. (2004), Query and Update Efficient B+-Tree Based Indexing of Moving Objects. *In Proc of the conference on Very Large Data Bases (VLDB)*, pp. 768-779, 2004.
- Marchand P., Brisebois A., Bédard Y., Edwards G. (2004), Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis, *Journal of the International Society of Photogrammetry and Remote Sensing (ISPRS)*, Volume 59, pp 6-20, 2004.
- Papadias D., Tao Y., Kalnis P., Zhang J. (2002), Indexing Spatio-Temporal Data Warehouses. *In Proc of the IEEE International Conference on Data Engineering (ICDE)*, pp. 166-175, 2002.
- Pfoser D., Jensen C. S., Theodoridis Y. (2000), Novel Approaches in Query Processing for Moving Object Trajectories. *In Proc of the conference on Very Large Data Bases (VLDB)*, pp. 395-406, 2000.
- Rao F., Zhang L., Yn X., Li Y., Chen Y. (2003), Spatial Hierarchy and OLAP-Favored Search in Spatial Data Warehouse. *In Proc of the 6th ACM Workshop on Data Warehousing and OLAP (DOLAP'03)*, pp. 48-55, 2003.
- Saltenis S., Jensen C. S., Leutenegger S. T., Lopez M. A. (2000), Indexing the Positions of Continuously Moving Objects. *In Proc of ACM Conference on Management of Data (SIGMOD)*, pp 331-342, 2000.
- Stefanovic N., Han J., Koperski K., (2000), Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 12(6): 938-958, 2000.
- Tao Y., Kollios G., Considine J., Li F., Papadias D. (2004), Spatio-temporal aggregation using sketches. Spatio-Temporal Aggregation Using Sketches. *In Proc of the 20th IEEE International Conference on Data Engineering (ICDE)*, pp. 214-225, 2004.
- Tao Y., Faloutsos C., Papadias D., Liu B. (2004), Prediction and Indexing of Moving Objects with Unknown Motion Patterns. *In Proc of the ACM Conference on the Management of Data (SIGMOD)*, 2004.
- Vazirgiannis M. et Wolfson O. (2001), A Spatiotemporal Query Language for Moving Objects, *In Proc of the conference on Spatial and Temporal Databases*, Springer Verlag Lecture Notes in Computer Science, No. 2121, 2001.