TIKIWIKI

Home

FICHE DE PRÉSENTATION

Ceci est un commentaire exemple. Faites moins long.

QUEL ARTICLE

Algorithms for Characterization and Trend Detection in Spatial Databases
Ester et al., KDD 1998 http://www.lri.fr/~sebag/Examens/Ester_KDD98.pdf

RÉSUMÉ

cet article adapte la recherche de motifs fréquents de type Apriori (Agrawal) dans le cadre d'une base de données spatiales, sous trois angles:

- Il s'interesse aux collections d'objets que l'on peut construire en suivant une relation de voisinage, eventuellement sous contraintes (direction)
- Il s'interesse aux attributs verifiés par ces collections d'objets, dont la valeur differe de la valeur moyenne sur la base de données.
- Il s'intéresse aux tendances (la valeur d'un attribut croit quand on s'éloigne d'une collection d'objets).

INTRODUCTION

Les auteurs se situent par rapport à l'état de l'art, notamment

- l'exploitation de hiérarchie de concepts (Lu et al. 93)
- la construction des attributs spatiaux afferents aux objets (Koperski et al 95)
- la caractérisation des "items spatiaux fréquents" (les polygones) dans le voisinage d'une majorité de points/clusters de points (Ng, Knorr et Ng 96).

Ils argumentent leur approche comme fondée sur des relations de voisinages (sous forme de primitives de la base de données, par opposition à des voisinages pré-compilés).

SECTION 2

Les auteurs reprennent leur travail précédent (Ester et al. 97),
introduisant 3 types de primitives spatiales :

- topologique (adjacence ds un graphe),
- distance (en km),
- direction (sud-ouest).

Sur ces primitives, sont construites trois relations:

- Voisins(objet O, graphe, contrainte) : tous les objets connectés à O ds le graphe, vérifiant contrainte.
- Chemins(Objets) : crée tous les chemins de longueur 1 ayant un objet de Objets come point de départ
- Extension(Chemin, nb, contrainte) : étend chaque chemin C de Chemin par un chemin C' (ayant pour point de départ le point d'arrivée de C) de longueur au plus nb, et tel que les objets sur C' vérifient contrainte.

Cette section introduit enfin trois types de contraintes (direction)
utiles sur les chemins:

- star-like : aucune contrainte de direction
- variable-starlike: le chemin C' doit être ds la même direction (e.g. nord, nord est ou nord-ouest) que le premier pas de C (e.g. nord),
- vertical starlike : du meme esprit.

Cette section est un peu confuse. Les auteurs veulent montrer la richesse
du langage, et sacrifient un peu la clarté des définitions.

SECTION 3

Etant donné un ensemble d'objets initiaux (targets), on considère
les voisinages croissants de ces objets,

V1 = union Voisin(objet, graphe), pour objet in targets

V2 = union Voisin (objet, graphe), pour objet in V1,

etc.

Sur le voisinage V_n, on retient les propriétés (e.g., présence de
montagne), dont la fréquence sur V_n est suffisamment différente
de la fréquence sur la base de données entière.

Les auteurs notent l'importance du paramètre n, qui conditionne ce qu'on va
trouver.

Remarque. Evidemment, comme la contrainte: "avoir une fréquence différente de la fréquence moyenne"
n'est pas monotone (à la différence du critère classique de support),
on ne voit pas du tout la même chose, selon le n considéré.

SECTION 4

Le but est ici, pour tout attribut a, de construire les "gradients" de a autour d'un objet donné : la valeur de a globalement diminue, augmente, ou fait n'importe quoi, au fur et à mesure qu'on s'éloigne de a.

Le gradient ne peut pas être calculé de manière plus précise, en filtrant les voisinages par la direction (e.g., nord).

Remarque: si au lieu de s'intéresser à la régression : valeur = f(distance) en filtrant les chemins par la direction, ils considéraient la régression : valeur = f(distance, direction) sur tous les chemins, ils auraient tout à la fois...

SECTION 5. EVALUATION

Un DBMS est un système de gestion de base de données (DataBase Management System). L'intérêt de l'approche formelle (primitives de DBMS) est argumenté de la facon suivante : d'un côté l'approche est portable; de l'autre, on souffre du surcoût de la BD.

Ici, je ne comprends pas très bien. Une BD devrait permettre un accès rapide aux données ?

L'expérimentation concerne environ 7,000 objets, décrits par 52 attributs et une description spatiale (polygone).

Les auteurs constatent que les performances ne sont pas liées à la taille du graphe, mais à sa connectivité (nombre moyen k de voisin par objet), et au fait que les contraintes considérées sont plus ou moins filtrantes. Ce qui parait évident: l'explosion des objets candidats est en k^n, si n est la longueur des chemins de voisinage.

CONCLUSION

Les auteurs évoquent en perspective la construction de règles discriminantes, et la jonction avec l'apprentissage relationel - "such as decision trees" (?).

MON AVIS

La validation est soigneusement expliquée, on voit qu'on pourrait poser des questions intéressantes au système (caractériser les zones non riches où le chomage est bas...)

La base de données peut paraitre un peu petite ? 7,000 objets...

Ils devraient mentionner une visualisation des règles : utiliser un code de couleur pour représenter les valeurs, et pour chaque attribut, montrer la valeur moyenne de l'attribut dans les voisinages des objets targets.

Mais alors, pourquoi ne pas avoir une carte fixe pour chacun des attributs ? Parce que les voisinages dépendent des contraintes retenues.

En résumé, cet article est intéressant, il présente quelque chose d'utile et de bien fondé; je ne suis pas tout à fait persuadé que l'approche passe à l'échelle.

Contributors to this page: lopes .
Page last modified on Monday 27 of February, 2006 [16:40:55] by lopes.

The original document is available at
http://tao.lri.fr/tiki-index.php?page=Exemple%20de%20fiche%20de%20pr%C3%A9sentation