REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE
DEPARTEMENT D'INFORMATIQUE

MEMOIRE de fin d'étude
Présenté pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique
Filière : Informatique
Spécialité : Systèmes d'Informations Avancés

Par: Layadi Kanza

SUJET

Extraction des motifs séquentiels

Soutenu publiquement le : 01 / 06 /2016 devant le jury composé de :

A. Khettaf	**Université de M'sila**	**Président**
S. Guesmia	**Université de M'sila**	**Rapporteur**
A. Bouda	**Université de M'sila**	**Examinateur**

Promotion : 2015 /20 16

Table Des Matières

INTRODUCTION GENERALE

Avec le développement des outils informatiques, nous assistons ces dernières années à un accroissement considérable de la quantité d'informations stockées dans de grandes bases de données scientifiques, économiques, financières, médicales, etc. et le défi aujourd'hui n'est plus de stocker ces donnée mais d'en extraire de l'information implicite et cachée dans ces données, particulièrement des recherches sur l'**E**xtraction automatique de **C**onnaissances à partir de **D**onnées. Cette discipline est l'intersection des domaines des bases de données, l'intelligence artificielle, et la statistique. L'ECD est décrite comme un processus interactif d'extraction de connaissances à l'aide d'algorithmes de calcul et d'interprétation des résultats, lors d'interactions avec l'expert pour aider à la décision, à partir d'ensemble de méthodes statistiques et algorithmiques sous la terminologie de Data Mining (la fouille de données).

La fouille de données concerne l'étape algorithmiquement difficile de ce processus, qui produit des motifs potentiellement intéressants à partir des données, elle regroupe un certain nombre de tâches, telles que la prédiction, le regroupement par similitude, la classification, la découverte d'associations, etc. L'un des plus importants problèmes de la fouille de données est la recherche de règles d'association. Cette approche, spécialisé dans la gestion de la relation client (GRC) et elle est identifier des corrélations cachées, potentiellement utiles, entre les attributs d'une base de données, il y a plusieurs approches et algorithmes ont été élaborés afin d'extraire les motifs et les règles d'association. Plusieurs types de motifs ont été définis selon le type de corrélation à extraire et selon la nature des données. Et dans ce mémoire on a plus précisément parlé sur les algorithmes d'extraction des motifs séquentiels (des motifs temporels) pour la découverte d'enchaînements fréquents dans les bases de données, avec des contraintes temporelles et l'identification des événements d'individus afin de pouvoir suivre leurs comportements séquentiels au cours du temps.

L'objectif de ce projet est la compréhension du comportement des principaux algorithmes d'extraction de motifs séquentiels en expliquant et illustrant leur fonctionnement l'implémentation des algorithmes pour l'extraction des règles séquentiels. Ce mémoire est structuré en quatre chapitres :

Le premier chapitre comporte une description générale sur la fouille de données et leurs techniques.

Le deuxième chapitre détermine les différents concepts d'extraction des règles d'association et mentionner les algorithmes les plus pratique.

Le troisième chapitre est consacré à quelques algorithmes d'extraction des motifs séquentiels.

Enfin, le quatrième chapitre est consacré à l'environnement logiciel et matériel utilisé ainsi que l'implémentation d'un algorithme.

CONCLUSION GENERALE

Dans le cadre de ce travail de master, nous avons traité le problème de fouille de données (data mining) dans des bases de données. Ce type de problème est présent dans de nombreux domaines d'applications. Au cœur de ce mémoire nous avons présenté les différentes techniques de fouille de données (extraction de Connaissances) et Plus précisément la technique d'extraction des motifs séquentiels, pour mieux cerner la problématique posée, nous avons commencé par la présentation générale des notions concerne le processus d'extraction des connaissances à partir de données et leur sous-processus la fouille de donnée. En deuxième temps, nous avons passé à l'outil d'extraction des motifs fréquents(les règles d'association) on a parlé sur les concepts généraux et les algorithmes utiliser dans ce domaine et aussi nous avons présenté d'une manière générale les algorithmes d'extraction des motifs séquentiels .et puis nous avons implémenté un algorithme parmi ces algorithme. Nous avons choisi le langage java pour écrire et développé notre application, on a utilisé un fichier texte Contient base de données.

Comme perspective à ce travail, les derniers travaux contiennent d'ailleurs de plus en plus de contraintes sur la définition des motifs séquentiels. Initialement la recherche de motifs séquentiels est de considérer que données sont booléennes : un client achète ou n'achète pas un produit et pour minimiser l'espace de recherche, un objet ne peut intervenir qu'une seule fois dans un ensemble d'achats. Par contre, l'intérêt des motifs extraits est très discutable, quand nous obtiendrons des motifs de la forme :'les personnes qui ont acheté trois bouteilles de boisson ont aussi acheté deux fromages', les nombreuses valeurs numériques (acheter 2 ou 3 fromages est difficilement séparable strictement) rendent ces motifs difficiles à extraire et peu informatifs, et dans ce cadre, les experts travaillent pour assouplir ces notions, aussi étendre la théorie de la non-dérivabilité vers d'autres motifs tels que les arbres et les graphes.

Bibliographie

[1] G. Piatetsky-Shapiro, Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from «university» to «business» and «analytics», Data mining and Knowledge Discovery.

[2] Alice Marascu, Extraction de motifs séquentiels dans les flux de données, Docteur en Sciences, France, 2009.

[3]Dhouha Grissa, Étude comportementale des mesures d'intérêt d'extraction de connaissances, doctorat, tunis, 2013.

[4] SASSI Amina, Une approche basée agent pour la fouille de données, magister en informatique, batna, 2013.

[5] René Lefébure et Gilles Venturi, Le Data Mining, Editions Eyrolles, 2001.

[6] Bouchekouf Asma, Perception du comportement de l'apprenant dans un environnement d'apprentissage, annaba, 2013.

[7] Jiawei Han and Micheline Kamber, Data mining concepts and techniques, 2nd edition , Diane Cerra San Francisco.

[8] Mohamed Hatem Haddad, Extraction et Impact des connaissances sur les performances des Systèmes de Recherche d'Information, docteur de l'université joseph fourier, 2002.

[9] Chami Djazia, Une plate forme orientée agent pour le data mining, magister, batna, 2010.

[10] Michael J. A. Berry, Gordon S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, chapitre 01, 2nd Edition, 2004.

[11] Lamiche Chaabane, fusion et fouille de donnees guidees par les connaissances: application a l'analyse d'image, doctorat, biskra, 2013.

[12] Tufféry, S., Data mining et statistique décisionnelle, Editions Technip. (2007).

[13] Benamar houmadi, étude exploratoire d'outils pour le data mining, doctorat, l'université du québec a trois-rivières, 2007.

[14] Guillaume Calas, Études des principaux algorithmes de data mining, Spécialisation Sciences Cognitives et Informatique Avancée, France, 2009.

[15] les arbres de décision
http://www.grappa.univlille3.fr/polys/apprentissage/sortie004.html
Consulté le 12/03/2016.

[16] Kellou Kenza et Mokhtari Abdeldjalil, Réalisation d'une plateforme d'expérimentations et de tests d'algorithmes de data mining, magister,

[17] Les plus proches voisins

http://www.grappa.univlille3.fr/polys/fouille/main.tgz

Consulté le 13/03/2016.

[18] C. Scharff, Méthode des k plus proches voisins, IFI, 2004.

[19] Philippe Preux. « Fouille de données. Notes de cours ». Université de Lille. 31 août 2009.

[20] Mohamed El hadi Benelhadj, entrepôt de données et fouille de données un modèle binaire et arborescent dans le processus de génération des règles d'association, doctorat, constantine

[21] Daniel Rajaonasyfeno, Mesures de qualité des règles d'association : normalisation et caractérisation des bases, doctorat, France, 2007.

[22] Sarra Ayouni, Etude et Extraction de Règles graduelles floues : Définition d'algorithmes efficaces, doctorat, tunis, 2012.

[23] Hassane Hilali, application de la classification textuelle pour l'extraction des règles d'association maximales, université du québec, 2009.

[24] Nicolas Pasquier, Data Mining : algorithmes d'extraction et de réduction des règles d'association dans les bases de données, doctorat, France, 2000.

[25] Allia Mohamed Rachid, BOUADI Tassadit, El MOUTAOUKIL Sami, et KEIRA Mamadou, Fouille de données : Règles séquentielles, master.

[26] Ansaf Salleb, Recherche de motifs fréquents pour l'extraction de règles d'association et de caractérisation, doctorat, Orléans, 2003.

[27] Abdelhak Mansoul, fouille de données biologiques : étude comparative et expérimentation, magister, oran, 2010.

[28] gwenael bothorel, algorithmes automatiques pour la fouille de données visuelle et la visualisation de règles d'association. Application aux données aéronautiques, doctorat, Toulouse, 2014.

[29] Alouan Basma, recherche de partitions floues optimale par la segmentation floue pour la fouille de données quantitatives, magister, boumerdes, 2008.

[30] Rahmani Rabah, découvret d'association sémantique dans les bases de données relationnelles par des méthodes de data mining, magister, tizi-ouzou.

[31] Jérôme Azé, Extraction de connaissances à partir de données numériques et textuelles, doctorat, France, 2003.

[32] Bilal Idiri, Méthodologie d'extraction de connaissances spatio-temporelles par fouille de données pour l'analyse de comportements à risques - Application à la surveillance maritime, doctorat, paris, 2013.

[33] Mickaël Fabrègue, Extraction d'informations synthétiques à partir de données séquentielles Application à l'évaluation de la qualité des rivières, doctorat, strasbourg, 2014.

[34] M'zali hassen, les règles d'association séquentielles, magister, 2006.

[35] Elias Egho, Extraction de motifs séquentiels dans des données séquentielles multidimensionnelles et hétérogènes Une application à l'analyse de trajectoires de patients, doctorat, Lorraine, 2014.

[36] Julien Rbatel, extraction de motifs contextuels : Enjeux et application dans les données séquentielles, France, 2011.

[37] Chedy Raïssi, Extraction de séquences fréquentes : des bases de données statiques aux flots de données, Montpellier, 2008.

[38] Asma Ben Zakour, Extraction des utilisations typiques à partir de données hétérogènes historiées en vue d'optimiser la maintenance d'une flotte de véhicules, doctorat, Bordeaux, 2012

[39] Maguelonne Teisseire, Autour et alentours des motifs séquentiels, doctorat, 2007.

[40] Thabet Slimani, Amor Lazzez, sequential mining: patterns and algorithms analysis, Computer Science, Taif University & LARODEC Lab, Saudia Arabia

[41]Marc Plantevit, Extraction De Motifs Séquentiels Dans Des Données Multidimensionnelles, doctorat, France, 2008.

[42] Hunor albert-lorincz, Contributions aux techniques de Prise de Décision et de Valorisation Financière, doctorat, lyon, 2007.

[43] Manish Gupta, Jiawei Han, Approaches for Pattern Discovery Using Sequential Data Mining, Université de Illinois at Urbana-Champaign, USA.

الملخص

استخراج النماذج المتسلسلة هي تقنية ذات أهمية في مجال استخراج المعلومات من قواعد البيانات، و هي تستعمل في كثير من المجالات وخاصة في مجال تحليل المعلومات في المجمعات الخاصة بالمبيعات.

مشكلة اكتشاف الأنماط المتسلسلة يرتكز على قاعدة بيانات للمعاملات بحيث هده المعاملات هي قائمة سلع متعلقة بالوقت او الزمن. وهدا المجال هو الأكثر صعوبة من مجال استخراج قواعد الترابط . يطبق عدة خوارزميات للحصول على أفضل النتائج بالنسبة إلى زمن التنفيذ وتقليل مساحة مجال البحث

الكلمات المفتاحية: استخراج البيانات، و قواعد البيانات, قواعد الترابط ، و نماذج متسلسلة .

Abstract

The extraction of sequential patterns is a significant challenge for data mining community, they are involved in areas more and more especially in the sales data analysis business organizations , the problem of discovery sequential patterns is given a transactional database where transactions are lists of items with time constraints . It is the most difficult area of discovery of sequential patterns, compared with research association rules. The domain uses multiple algorithm achieved a better result in terms of execution time and minimize the motives search space.

Keywords : data mining, databases , association rules , sequential patterns .

Résumé

L'extraction de motifs séquentiels est un défi important pour la communauté fouille de données, ils se trouvent impliqués dans des domaines de plus en plus nombreux en particulier dans l'analyse de données de vente d'organisations commerciales, Le problème de la découverte des motifs séquentiels consiste, étant donné une base de données transactionnels où les transactions sont des listes d'items avec des contraintes de temps. Il est le plus difficile du domaine de la découverte des motifs séquentiels, comparativement à la recherche des règles d'association. Ce domaine utilise plusieurs algorithme pour obtenue un meilleur résultat en terme de temps d'exécution et minimiser l'espace de recherche des motifs.

Mots clé : la fouille de donnée, les bases de données les règles d'association, les motifs séquentiels.