

Histoires d'entrelacs et d'analyse multidimensionnelle de base de données



~\$ whoami

Sébastien NEDJAR

Geek et Maker depuis ma naissance

Enseignant Chercheur à Aix-Marseille
Université pendant 15 ans

Chercheur Entrepreneur pendant 4 ans

Arrivé au CESI depuis 3 mois

Thèmes de recherches

Bases de données multidimensionnelles

Extraction de connaissances dans les grands ensembles de données

Analyse formelle de concepts et treillis

Fouille de données appliquée aux réseaux de capteurs maillés

Data mining et analyse multidimensionnelle

- Les algorithmes de data mining visent à extraire des connaissances cachées profondément dans les données
- Ces nouvelles connaissances nécessitent une analyse pour en comprendre les causes
- La recherche des causes passe par un processus de navigation
- L'analyse multi-factorielle et multidimensionnelle basée sur la navigation à plusieurs niveaux de granularité est centrale

Problèmes traités

- Analyse multidimensionnelle des renversements de tendance
- Recherche multidimensionnelle d'objets dominants
- Extraction multidimensionnelle d'objets atypiques
- Représentation compacte des connaissances
- Fouille de séquences temporelles

Exemples d'applications



Microélectronique

- En micro-électronique, la fabrication de composants est un processus industriel complexe
- Chaque étape du processus est susceptible d'engendrer des composants défectueux
- Pour les détecter au plus tôt, de très nombreux tests sont effectués
- Que ce soit pour prédire les défaillances ou en analyser les origines, des algorithmes de recherche multidimensionnelle d'outliers (valeurs atypiques, données erronées, signaux faibles) ont facilité ce travail

Pharmacovigilance (1/2)

- En pharmacovigilance, on cherche à évaluer les effets indésirables résultant de l'usage des médicaments.
- Actuellement, la détection repose principalement sur les médecins (obligation réglementaire).
- L'établissement de la causalité est possible uniquement si elle est évidente et immédiate.
- Les affaires du Mediator ou de l'OxyContin sont des exemples patents des limites du système actuel de la pharmacovigilance.

Pharmacovigilance (2/2)

- Les laboratoires d'analyse médicale font transiter de nombreux résultats d'analyse biologique pour les acheminer jusqu'aux médecins
- Des analyses sont faites au fil de l'eau pour trouver les grandes tendances
- Quand une tendance devient identifiable avec les outils classiques d'informatique décisionnelle, il est trop tard
- Avec des méthodes d'apprentissage et des algorithmes d'extraction des renversements de tendances, on isole les phénomènes au plus tôt

Jeux vidéo

- Dans les jeux vidéo en ligne ou dans l' E-Sport, la triche peut gacher l'expérience collective et générer des pertes financières
- La triche peut être définie comme une déviation volontaire de l'espérance de gain d'un joueur
- La triche est parfois liée à un contournement des règles mais aussi à l'exploitation répétée des déséquilibres intrinsèques à la mécanique de jeu
- Détecter les changements de tendance et les valeurs atypiques permet d'identifier les triches potentielles

Réseaux sans fil

- Pour contourner la limite de la portée dans les réseaux sans fil, de plus en plus de réseaux de capteurs adoptent des technologies dites maillées
- Dans un réseau maillé, tous les noeuds peuvent être amenés à relayer les informations des autres noeuds à portée
- Agir comme un relais consomme de l'énergie et les capteurs ont une autonomie limitée
- Les méthodes d'analyse de séquences temporelles, ont permis de déterminer l'ensemble des relais maximisant l'autonomie globale du réseau

Conclusion

- L'émergence du Big Data puis de l'IA a apporté un nouvel éclairage aux travaux de fouille de données.
- La profusion de données non encore exploitables est une mine d'or pour le chercheur en data mining.
- Il y a encore de nombreuses questions de fiabilité, d'éthique ou d'impact environnemental qui ne sont que peu traitées.