

Outils pour le Data Mining



Comment Choisir un outil

- **Systemes commerciaux de data mining possèdent peu de propriétés communes :**
 - Différentes méthodologies et fonctionnalités de data mining
 - Différents types d'ensembles de données
- **Pour la sélection d'un outil, on a besoin d'une analyse multi- critère des systèmes existants**

Comment Choisir un outil

- **Types de données** : relationnel, transactionnel, texte, séquences temporelles, spatiales ?
- **Issues systèmes**
 - Support systèmes d'exploitation ?
 - Architecture client/serveur ?
 - Fournit une interface Web et permet des données XML en entrée et/ou en sortie ?
- **Sources des données** :
 - Fichiers texte ASCII, sources de données relationnels multiples, ...
 - Support ODBC (OLE DB, JDBC) ?

Comment Choisir un outil

- **Fonctionnalités et méthodologies**
 - une vs. plusieurs fonctions de data mining
 - une vs. plusieurs méthodes par fonction
- **Couplage avec les systèmes de gestion de base de données et les entropots de données**
- **Outils de visualization** : visualisation des données, visualisation des résultats obtenus, visualisation du processus, visualisation interactive (split attribut, ...), etc.

Comment Choisir un outil

- **Extensibilité (Scalability)**
 - instances (Taille de la base de données)
 - attributs (dimension de la base)
 - Extensibilité en terme d'attributs est plus difficile à assurer que l'extensibilité en terme d'instances
- **Langage de requête et interface graphique (IHM)**
 - easy-to-use et qualité de l'interface
 - data mining interactif

Exemple d'outils (1)

- **Intelligent Miner d'IBM**
 - Intelligent Miner for Data (IMA)
 - Intelligent Miner for Text (IMT)
 - Tâches : groupage de données, classification, recherche d'associations, etc.
- **Entreprise Miner de SAS**
 - SAS : longue expérience en statistiques
 - Outil «complet» pour le DM
- **Darwin de Thinking Machines**
 - Trois techniques : réseaux de neurones, arbres de décision et régression.
 - Client-Serveur

Exemples d'outils (2)

- **MineSet de Silicon Graphics**
 - Fonctionnalités interactives et graphiques
 - Techniques sous-jacentes : classification, segmentation, recherche de règles d'association.
- **Outils/librairies libres**
 - SIPINA
 - WEKA
- **Data-Miner Software Kit (DMSK)**
 - Kit de programmes : méthodes statistiques, segmentation, groupage, réseaux de neurones, etc.
 - Il existe une version en java
- **etc.**