

Projet OKA : Optimisation Knowledge Accurated

Les statisticiens traitent la masse d'information sur laquelle est appliquée un ensemble de processus de reconnaissance de connaissances : mesures (probabilistes ou autres), segmentation et classification (supervisées ou non), scoring (normé ou pas). Or, les aspects technique et scientifique rigoureux limitent la capacité des équipes études à créer et générer de la connaissance à la même vitesse qu'augmente l'accès aux données.

Le projet répond aux enjeux suivants :

- de s'affranchir des données manquantes par la modélisation non paramétrique des événements
- d'automatiser des classifieurs et des segmentations robustes
- détecter les événements rares et actualiser dynamiquement les connaissances
- d'intégrer les techniques de fusion d'information symboliques et numériques ([Chauvin, 1999], [Desarachy, 1994]), de faire de la multi segmentation un outil performant de mesures décisionnelles sur de gros volumes de données.

Les résultats scientifiques de détection de connaissances dans la masse d'information sont suffisamment mûrs pour être exploités industriellement, de manière récurrente (mensuellement, a fortiori hebdomadairement et quotidiennement) et fiable (robuste aux artefacts des informations nouvelles, à l'échantillonnage et au bruit d'information) et reproductibles (capacité à s'adapter aux nouveaux contextes). Ce projet mettra en exergue les techniques pertinentes et fiables pour la prise de décision rapide sous ces trois contraintes.

Les thèmes traités sont les suivants :

Sujets scientifiques :

- Mesure : les données (qualitatives, quantitatives, symbolique et numériques, algèbre des indicateurs et des indices) mesurer par les Ensembles Flous, les Probabilités et Dempster et Shafer.
- Qualité statistique de la décision : mesures, robustesse (ROC et Lift).
- Echantillonnage : technique et industrialisation (optimisation des poids)
- Classification dynamique des critères cibles homogènes et / ou hétérogènes (supervisée et non supervisée) : arbre de décision, réseau de neurones, règles 'association, scoring, algorithme de Galois, cluster
- Benchmark des solutions de scoring.
- Mesure temporelle des classifications : mesure de prédiction de classes (optimisation dynamique de seuils statistiques)
- Combinaison de classification (fusion)

Sujets techniques :

- Blog OKA : point de convergence de référencement des pratiques et des démarches projet de type fusion de données
- Benchmarking de logiciel :
 - partenariat industriel avec des constructeurs
 - état de l'art technique sur l'optimisation d'un OLAP
 - référencement des outils pouvant offrir des réponses en temps réel.
 - architecture technique
- Conception logiciel OKA (plate forme d'intégration de multiple modèles) : modèle prévisionnel sur matrice décisionnelle (prévision sur les classifications).

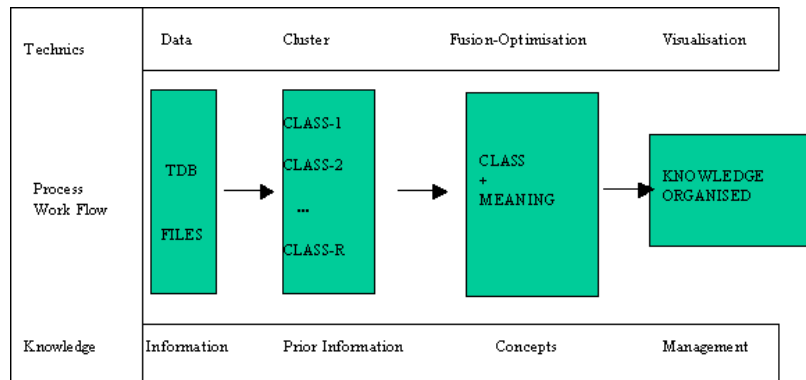
Sujets fonctionnels :

- Benchmarking des méthodes : expliciter la démarche Six Sigma DMAIC pour l'approche Fusion de données.
- Dynamique temporelle : actualiser les matrices décisionnelles et détecter les évolutions atypiques
- Blog OKA :
 - transfert de compétence, cours, liens utiles
 - mise à disposition d'indicateurs métier et études de marché optimisées par les méthodes dites de Knowledge Discovery et knowledge management
 - applications Marketing, Sondage, Météorologique, Industrielles.

A. Livrables :

- Générer de multiple calculs et de multiple raisonnements pour extraire des connaissances
- Réduire la complexité de l'interprétation des données en injectant de nouvelles connaissances
- Constituer une méthodologie efficace à l'analyse des données adaptée à de multiples sujets d'étude différents sur très gros volumes.

B. Méthodologie Six Sigma pour le datamining :



Fusion Specified For Data Mining: integration Software analysing Knowledge, Research Report for European Community, TRACK ESPRIT PROJECT n° 25.517, 1999.

C. Mots clés :

- Scientifiques : qualité statistique de la décision & benchmark classification dynamique (supervisée et non supervisée), décisions distribuées / fusion de données.
- Techniques : benchmark outil OLAP & Visualisation, architecture des traitements et méta données pour le datamining, Knowledge Management.

D. Références :

- [Chauvin, 1999] System Optimised by Fusion of Information for Data Mining Problems, S. Chauvin, C. L. Dunis, L. Jáñez, and J. Laws, Rapport ITC & UE, 1999, Madrid.
- [Desarachy, 1994] B. Desarachy, Decision Fusion, IEEE Computer Society Press, 1994.