

P51 Visualisation et clustering de données en grandes dimensions

Encadrant 1 : Vincent Gripon

Département : Élec

Encadrant 2 : Grégoire Mercier

Département : ITI

Encadrant 3 : Dominique Pastor

Département : SC

Partenaire extérieur : *Mathias Herberts*, Mathias.Herberts@cityzendata.com, *Cityzen Data*

Mots clés : grande dimension, apprentissage, visualisation, données réelles

● CONTEXTE :

Les technologies actuelles permettent la manipulation de données en grande dimension. Ces données riches sont cependant difficiles à appréhender, en particulier à cause de la malédiction de la dimension. Visualiser et classer de telles données nécessite la mise en place d'algorithmes spécifiques faisant appel à des notions mathématiques complexes. La littérature scientifique sur le sujet est assez récente et la mise en pratique de ces modèles pour des données réelles peut mener à des écarts conséquents avec la prédiction théorique. Ce projet a pour but de mettre en œuvre certaines de ces techniques pour des données réelles mises à disposition par le partenaire extérieur et de tester leur efficacité.

● DESCRIPTIF SUCCINCT DU PROJET :

Lorsque l'on fait face à des données en grandes dimensions se pose le problème de la représentation des dépendances internes. Un outil générique permettant de représenter ces dépendances est le graphe. Récemment une littérature scientifique fournie s'est développée autour des idées qu'un graphe portant un signal peut être exploité pour spécifier des transformées adaptées, lesquelles généralisent les transformées de signal classiques type Fourier ou ondelettes. L'objectif du projet est la mise en œuvre de techniques de transformation de signal s'appuyant sur des graphes pour la visualisation et la classification de données en grande dimension.

● LIVRABLES :

- Étude bibliographique sur les outils du traitement du signal sur graphes 20%
- Spécification de méthodes de visualisation à base de graphes 20%
- Développement d'applications basées sur la plateforme Cityzen Data et son langage Einstein dédié à la manipulation de séries spatio-temporelles ainsi que sur l'emploi du langage Java et/ou d'un autre langage s'exécutant sur la JVM (Java Virtual Machine). Les langages HTML5 et Javascript seront utilisés pour la partie graphique. Les

applications développées devront permettre la visualisation et la classification de données en grande dimension 40%

- Tests et comparaison des méthodes proposées avec celles disponibles dans la littérature/ sur le marché 20%

● OBJECTIFS PEDAGOGIQUES

- Mettre en œuvre de façon raisonnable d'un point de vue computationnel les méthodes de transformations s'appuyant sur des graphes,
- Maîtriser les outils pour l'indexation multimedia de vecteurs en grandes dimensions,
- Afficher de façon synthétique et intuitive les résultats des algorithmes mis en œuvre.

● PRE-REQUIS:

- Connaissances en programmation linéaire,
- Maîtrise d'outils de calcul parallèle,
- Savoir-faire avec les bibliothèques d'affichage.