Prototype de clustering exploratoire pour l'aide à la segmentation des clients

Adnan El Moussawi^{*,**}, Philippe De Guis^{**}, Arnaud Giacometti^{*},
Nicolas Labroche^{*}, Arnaud Soulet^{*}

^{*}Université François Rabelais de Tours - {prénom}.{nom}@univ-tours.fr

^{**}Group KALIDEA - {aelmoussawi, pdeguis}@kalidea.com

Résumé. Le clustering est une technique largement répandue pour la définition de profils dans le cadre de l'aide à la gestion de la relation client (CRM). Cependant, les outils classiques sont généralement limités, car ils ne prennent pas en compte la connaissance métier de l'analyste et ne permettent pas l'exploration interactive des données. Nous décrivons ici un prototype qui permet à un expert marketing d'explorer interactivement les données pour la recherche de profils des clients, mais aussi d'analyser les profils construits à l'aide de différentes visualisations synthétiques et d'étudier leurs évolutions au cours du temps.

1 Introduction

Nous considérons le problème de l'exploration de données par un expert, notamment dans le domaine du CRM où il cherche à explorer les données des clients pour construire des profils avec une sémantique compréhensible, pour leurs proposer des programmes spécifiques de fidélisation. Traditionnellement, deux méthodes peuvent être utilisées pour l'exploration de ces données : l'OLAP et les algorithmes de clustering.

Les outils OLAP permettent une analyse interactive des données avec des opérateurs de base pour sélectionner un sous-ensemble des données et spécifier les dimensions d'analyse pertinentes. Mais, ces outils ne permettent pas de révéler des modèles intéressants cachés dans les ensembles de données tels que des groupes de clients de comportements similaires.

Les algorithmes de clustering révèlent la structure naturelle des ensembles de données et peuvent résumer l'information en cas de données volumineuses. En revanche, ils manquent de l'interactivité qui permet de spécifier dynamiquement les sous-ensembles de données ou les attributs d'analyse qui intéressent l'expert. Dans les approches de clustering interactif telles que Balcan et Blum (2008) et Awasthi et al. (2013), l'interaction de l'utilisateur avec la méthode de clustering est par exemple limitée par des requêtes de scission et/ou fusion des clusters. Les approches du clustering, dites « semi-supervisée », permettent également l'interaction de l'expert via des contraintes au niveau des objets de données ou sur les caractéristiques de clusters (Grossi et al., 2016). Des travaux récents en collaboration avec l'entreprise KALIDEA ont contribué à proposer une nouvelle méthode de clustering semi-supervisée qui permet à l'expert de définir des préférences sur les attributs d'analyse (El Moussawi et al., 2016).

Le présent prototype a pour objectif de proposer un outil de segmentation semi-automatique avec une IHM adaptée à un expert non spécialiste de la fouille de données, simple à manipuler.