Proposition de thèse CIFRE entre la société Groupe UP et le Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT)

Clustering prédictif pour l'analyse de trajectoires clients

Cette thèse se place dans le cadre de l'optimisation de campagnes marketing dans le contexte de la gestion de la relation client. La recommandation de programmes de stimulation et de fidélisation clients adéquats repose sur la construction de **segments clients**, ie. des groupes de clients au comportement similaire, et de leur analyse au cours du temps.

En marketing, il est crucial de comprendre **pourquoi certains clients basculent d'un segment à l'autre**, pourquoi un nouveau segment a émergé, ou un autre a disparu. La réponse à ce type de questions a été étudiée jusqu'à présent par le biais de méthodes de clustering temporel ou autorisant la caractérisation des **transitions** entre clusters [1]. Nous nous intéressons ici plus particulièrement à la prise en compte d'événements externes, et de leur impact dans les transitions de clients entre segments au cours du temps.

Nous nous proposons dans ce travail de thèse de formaliser l'analyse des transitions et des segments clients sous la forme d'un problème de clustering. Celui-ci combinera un modèle prédictif complexe avec un algorithme de clustering adapté (on parle alors de clustering prédictif) afin de produire des clusters homogènes avec les prédictions de transitions entre clusters ou d'évolutions des caractéristiques d'individus en fonction des états passés et des événements externes.

Plusieurs verrous scientifiques majeurs seront abordés par ce travail de thèse :

1. **Trouver et caractériser les transitions** sur des données exprimées sur des attributs numériques ou catégoriels. Pour ce faire, il faudra **déterminer les attributs pertinents** pour décrire les catégories d'utilisateurs et selon lesquels étudier les transitions entre catégories.
2. **Modéliser des événements externes** qui sont soit prévisibles et voulus (politique commerciale) soit imprévisibles (conflits, problèmes macro-économiques) et modéliser le lien entre les transitions observées et ces événements externes (ou leur absence).
3. **Construire une nouvelle catégorisation des clients sur la base de la prédiction de leur évolution**. Il est important de noter que ce problème ne pourrait être résolu en combinant un algorithme de clustering et un algorithme de classification, car dans ce cas les clusters seraient construits en étant agnostiques des modèles de prédiction, comme rappelé dans [2]. L'approche devra pouvoir construire un nombre

variables de catégories clients afin de s'adapter localement dans le temps à des phénomènes d'éclatement de la clientèle ou au contraire de contraction du marché.

D'une certaine manière, on peut considérer le clustering prédictif comme une méthode contrainte par l'homogénéité sur des classes prédites. Nous pensons intégrer cette nouvelle solution au démonstrateur déjà existant [3, 4] et qui intègre déjà un modèle de clustering avec préférences sur les attributs. Une perspective à plus long terme consisterait à définir un modèle qui intègre ces deux types de contraintes.

Candidatures ouvertes jusqu'au 10 juillet 2018 en envoyant CV, lettre de motivation et derniers relevés de notes.

Contact : nicolas.labroche@univ-tours.fr, patrick.marcel@univ-tours.fr

[1] M. Spiliopoulou, I. Ntoutsis, Y. Theodoridis, R. Schult (2006). MONIC: modeling and monitoring cluster transitions. In Proc. KDD, pp. 706-711

[2] V. Lemaire, N. Creff, F. Clérot (2012) K-moyennes contraintes par un classifieur. Application à la personnalisation de scores de campagnes. EGC 2012, pp. 155-166

[3] A. El Moussawi, A. Cheriati, A. Giacometti, N. Labroche, A. Soulet (2016) Clustering with Quantitative User Preferences on Attributes. ICTAI 2016, pp. 383-387

[4] A. El Moussawi, P. De Guis, A. Giacometti, N. Labroche, A. Soulet (2017) Prototype de clustering exploratoire pour l'aide à la segmentation des clients. EGC 2017, pp. 457-460