

# K-moyennes contraintes par un classifieur

## Application à la personnalisation de scores de campagnes

Vincent Lemaire\*, Nicolas Creff\*\*, Fabrice Clérot\*

\* Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion

\*\* Epita 14-16 rue Voltaire 94276 Kremlin Bicêtre Cedex

**Résumé.** Lorsqu'on désire contacter un client pour lui proposer un produit on calcule au préalable la probabilité qu'il achètera ce produit. Cette probabilité est calculée à l'aide d'un modèle prédictif pour un ensemble de clients. Le service marketing contacte ensuite ceux ayant la plus forte probabilité d'acheter le produit. En parallèle, et avant le contact commercial, il peut être intéressant de réaliser une typologie des clients qui seront contactés. L'idée étant de proposer des campagnes différenciées par groupe de clients. Cet article montre comment il est possible de contraindre la typologie, réalisée à l'aide des k-moyennes, à respecter la proximité des clients vis-à-vis de leur score d'appétence.

## 1 Introduction

### 1.1 Problématique industrielle

Le data mining consiste en l'ensemble des méthodes et techniques qui permettent d'extraire des informations à partir d'une grande masse de données. Son utilisation permet d'établir des relations entre ces données et par exemple de définir des comportements type de clients dans le cadre de la gestion de la relation client.

Lorsqu'on désire contacter un client pour lui proposer un produit on calcule au préalable son appétence à ce produit. Il s'agit là de calculer la probabilité qu'il achètera ce produit. Cette probabilité (encore appelée score) est calculée à l'aide d'un modèle prédictif pour un ensemble de clients (le périmètre de la campagne). Le calcul du score exploite un grand nombre de variables explicatives issues du système d'information.

Les clients sont ensuite triés dans l'ordre décroissant de leur probabilité d'appétence. Le service marketing ne contacte ensuite que les plus appétents (nommés "tops scores"), i.e. ceux ayant la plus forte probabilité d'acheter le produit. En parallèle et avant le contact commercial il peut être intéressant de réaliser une typologie des clients qui seront contactés. L'idée étant de proposer des campagnes différenciées par segment. Un argumentaire commercial est construit pour chaque groupe de clients après analyse des caractéristiques du groupe : âge, CSP, offres actuellement détenues. Il est fréquent (pour des raisons pratiques de temps d'analyse) que l'analyse du groupe se résume en l'analyse du centre (ou représentant) du groupe.

Aujourd'hui, cette typologie est très souvent réalisée de manière non supervisée à l'aide d'une technique de partitionnement de type k-moyennes. La valeur de  $k$  est prédéfinie et la