

Exploration des corrélations dans un classifieur Application au placement d'offres commerciales

Vincent Lemaire*, Carine Hue**

* Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion
vincent.lemaire@orange-ftgroup.com,
<http://perso.rd.francetelecom.fr/lemaire>

**GFI Informatique, 11 rue Louis de Broglie, 22300 Lannion
chue@gfi.fr

Résumé. Cet article présente une nouvelle méthode permettant d'explorer les probabilités délivrées par un modèle prédictif de classification. L'augmentation de la probabilité d'occurrence de l'une des classes du problème étudié est analysée en fonction des variables explicatives prises isolément. La méthode proposée est posée et illustrée dans un cadre général, puis explicitement dédiée au classifieur Bayésien naïf. Son illustration sur les données du challenge PAKDD 2007 montre que ce type d'exploration permet de créer des indicateurs performants d'aide à la vente.

1 Introduction

Etant donné une base de données, une question classique est de chercher à relier un phénomène dit "à expliquer" à un ou plusieurs phénomènes explicatifs. L'extraction de connaissances passe alors couramment par l'élaboration d'un modèle qui explicite cette relation. Pour chaque individu de la base, un modèle probabiliste permet, étant données les valeurs de l'individu pour chaque variable explicative, d'estimer les probabilités d'occurrence de chaque classe cible ainsi que la classe cible prédite. Ces probabilités ou scores sont réinjectés dans le système d'information pour par exemple personnaliser la relation clients : le choix des offres, de l'interface des services, du canal de communication, du canal de distribution... Néanmoins la connaissance extraite sur un phénomène par le score n'est pas toujours exploitable directement. Par exemple, si un modèle prédit pour un client son potentiel à adhérer ou non à une offre, autrement dit son appétence à cette offre, il ne dit rien sur l'action ou les actions à entreprendre pour rendre plus probable son adhésion. Il semble ainsi nécessaire de posséder une méthodologie qui, pour chaque client, (i) permettra d'identifier l'importance des variables explicatives (ii) permettra d'identifier le placement des valeurs de ces variables explicatives et (iii) proposera d'entreprendre une action pour augmenter son appétence à l'offre.

Nous proposons de traiter ce troisième point en explorant la relation existante, au sens du classifieur, entre les variables explicatives prises indépendamment et la variable cible. Cette exploration, à réaliser pour chacun des clients, produit une connaissance qui sera ensuite exploitée dans un processus de "Customer Relationship Management" (CRM) pour, par exemple, fournir une information personnalisée dans l'argumentaire des téléopérateurs lors d'une campagne de promotion.