

Détection par SVM - Application à la Détection de Churn en Téléphonie Mobile Prépayée

Cédric Archaux ***, Arnaud Martin **, Ali Khenchaf **

* Bouygues Telecom, 20 quai du point du jour, 92100 Boulogne Billancourt
carchaux@bouyguetelecom.fr
<http://www.bouyguetelecom.fr>

** ENSIETA, 2 rue François Verny, 29806 Brest cedex 9
[archauce, Arnaud.Martin, Ali.Khenchaf]@ensieta.fr
<http://www.ensieta.fr/e3i2>

Dans le cadre de la téléphonie mobile prépayée, les clients ne sont pas engagés contractuellement et peuvent cesser leur activité sans préavis. Les opérateurs doivent donc estimer le risque de fuite (churn) que présentent les clients afin d'estimer l'effort de fidélisation qui peut être engagé au cas par cas. Nous nous plaçons ici dans un problème de détection de deux classes de clients, ou nous distinguons les clients à fort risque de départ des autres. En complément des analyses de survie présentées dans [Rosset et al., 2002], nous comparons la classification par réseau de neurones multicouche tel que celui présenté dans [Mani et al., 1999] aux machines à support de vecteurs (SVM) présentés dans [Vapnik 1998] à travers le logiciel SvmFU développé par [Rifkin et al., 2002]. Les SVM sont particulièrement efficaces en classification, et sont particulièrement adaptés aux données bruitées grâce à leur robustesse. Nous séparons les clients par l'hyperplan optimal qui garantit que l'écart entre les deux classes soit maximal. Les nouveaux clients pour lesquels nous devons détecter le churn, pourront ainsi ne pas être trop similaires à ceux employés pour trouver l'hyperplan mais être tout de même situés franchement d'un côté ou l'autre de la frontière. Nous appliquons les détecteurs sur des bases d'apprentissage réelles et appliquons les modèles à des bases de volume significativement supérieur aux bases d'apprentissage afin de tester la robustesse des modèles. Le taux de 85% de bonne détection obtenu par les SVM sur les bases de test est significativement meilleur que le taux de 80% obtenu par les réseaux de neurones. Nous montrons donc que les SVM permet d'obtenir de meilleurs résultats qu'un réseau de neurones multicouche pour le problème de détection de la fuite des clients en téléphonie mobile prépayée.

Références

- [Mani et al., 1999] Mani D.R., Drew J., Betz A., Datta P. (1999), Statistics and data mining techniques for lifetime value modeling, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 94-103, 1999.
- [Rifkin et al., 2002] Rifkin R., Moreno P., Nicpanski H., Alvira M., Paris J., King V., Nadermann M., SvmFu Documentation, (2002), <http://five-percent-nation.mit.edu/SvmFu/>.
- [Rosset et al., 2002] Rosset S., Neumann E., Eick U., Vapnik N., Idan Y. (2002), Customer lifetime value modeling and its use for customer retention planning, Proceedings of the eighth ACM SIGKDD, pp. 332-340, 2002.
- [Vapnik, 1998] Vapnik V. (1998), Statistical Learning Theory. John Wiley & Sons, 1998.