Abstract ampliado

RESUMEN AMPLIADO

Título: Spatial models in customer churn prediction: An application on an online grocery retailer

Autores y e-mail de todos ellos: De la Llave, M. (miguel.delalla@gmail.com); López F.A. (Fernando.lopez@upct.es)

Departamento: †Grupo de investigación en Econometría Espacial; *Departamento de Métodos Cuantitativos e Informáticos

Universidad: Politécnica de Cartagena

Área Temática: *S10 – Procesos espaciales a nivel de punto y redes espaciales*

Resumen: *(mínimo 1500 palabras)*

The impact of geography on marketing science is an important topic of research for business and management. A model becomes '*spatial*' if the behaviour of one economic agent is codetermined by nearby economic agents (Burridge et al., 2016). Spatial analysis is a new and emerging research topic in marketing – one which has not yet revealed its potential – that is attracting more interest each day due to the increasing availability of georeferenced information. In the field of Customer Relationship Management (CRM), taking advantage of the spatial correlation between customers can improve the predictive performance of models. The main contributions to CRM (including spatial effects) are in the subfield of customer acquisition (Baecke and Van den Poel, 2012, 2013; Jeonghye et al., 2010); in relation to customer churn behavior, however, no research that takes into account geography and 'space' as explicative factors has yet been undertaken. How would any business model change if we could predict which clients are more prone to leave the company? Can we actually predict customer churning accurately in first moments of the client with the company? Although there has been much literature around prediction of customer churning in the last years, none of it takes into account spatial models to do so. So, this paper is based on the belief that spatial econometrics can improve the performance of existing models.

The recent explosion of ecommerce companies makes having loyal clients an extremely complicated task. Companies spend large amounts of resources trying to gain new clients each year, but many of them consistently fail to retain such clients. Engaged customers create higher benefits than new ones (Reichheld and Sasser, 1990). The cost of gaining a new client overpasses the cost of retaining the same client. Some research has shown that this costs between 6 (Verbeke et al., 2011) and 12 times that of retaining the existing customer (Torkzadeh et al., 2006). So it seems only logical to avoid churning as much as possible in order to create a sustainable business model and also to elude adverse effects such as negative company's reputation or negative feedback which may influence potential customers (Saradhi and Palshikar, 2011). Therefore, predicting churning before it happens is a critical point for grocery e-companies. If such group of clients or policies can be detected, wherever the risk of churning is high, specific marketing actions can be developed in order to retain the clients.

The phenomenon of client churn can be frequently observed in different consumer services markets such as telecommunications (Archaux et al., 2004; Hung et al., 2006; Rosset et al., 2003), insurance (Günther et al., 2014; Risselada et al., 2010; Morik and Köpcke, 2004), subscription services (Coussement and den Poel, 2008), financial services (Lariviere and den Poel, 2005) and banking (Xie et al., 2009). A great deal of methodological approaches has been discussed in examinations of market independence. The most popular of these approaches use classification trees (Lemmens and Croux, 2006) and logistic regression (Günther et al., 2014); multiple statistical techniques[1] have also been developed in order to identify customers who are likely to churn based on their characteristics: for instance, survival analysis (e.g. Brockett et al., 2008); neural networks (Hung et al., 2006); random forest (Lariviere and Van del Poel 2005); support vector machines (Xie et al., 2009); and more recently, machine learning (bagging; boosting; staking; voting) has been applied (Risselada et al., 2010). Most of those techniques have resulted in limited gains in accuracy and substantial increases in complexity (Risselada et al., 2010). This statement is also supported by Neslin et al. (2006), who found that logistic regression models and classification trees accounted for 68% of entries in churn modelling.

[1] A full description of methodologies used in the churn prediction model, besides the most important contributions, is to be found in Table 1 in Verbeke et al., (2011); Table 1 in Soeini and Rodpysh (2012); Table 1 in Abbasimehr et al., (2014); Table 1 in Allahyari and Vahidy, (2012); and Table 1 in Tsai and Lu (2009). A comparative study is presented in Vafeiadis et al. (2015).

Small e-commerce companies collect large amount of information from their clients, which can be analyzed to find certain valuable patterns. They store data from the very first moment of the client using the service. The specific date when the client signed up and all the events of that specific moment could reveal something interesting about coming client's behavior. Channel of entrance is proved throughout this paper to be decisive to discriminate the loyal clients to the ones who become inactive in a short period of time. Also, first user's orders impact meaningfully on the prediction of future behavior of a customer. This already appears in Moriuchi and Takahashi (2015) where shopping experience appear vital in gaining consumers' trust.

There is one piece of information, which is included in all data warehouses that has never previously been used in models of churn prediction: the address of the customer. The address of a customer is an important element that enriches any churn model. First, knowing the neighborhood of the customer, then indirectly the economic position is revealed, allowing dividing customers into limited neighborhoods. Some research that has been undertaken on zip codes in churn prediction (Löchl et al. 2009; Verbeke et al., 2012; Huigevoort and Dijkman, 2015) showed ambiguous conclusions. Verbeke et al. (2012) writes, "*the number of times a customer called the helpdesk will most probably be a better predictor of churn behavior than the zip code*". Secondly and directly related to this research, knowing the exact location of a customer (map coordinates) makes it possible to identify the proximity of other customers. Nearby customer churn behavior is probably codetermined, and some mimetic conduct between them can be detected. Pinheiro and Helfert (2010) wrote, "*Some events within the network can be influenced by activities of other customers. In the example of churn, word of mouth, rumors, commentaries and mostly activities of churn of other customers may create a chain process"*. Along the same lines, Haenlein (2013) presents evidences on the importance of social interaction in customer churn decisions. Lastly, if the e-grocery company knows the exact location of a customer, it is easy to identify geographical factors (strategic geographical points) that could be related to churning. Proximity to certain supermarket is probably an essential factor that influences it.

Taking into account the state of research, the main objective of this paper is to demonstrate the impact of geographical factors on churn prediction. Using the experience of an e-commerce startup business in Madrid (Spain), it is proved in this paper that the power of using geographical data to improve the classical probit regression model using Spatial Regression Probit Models (LeSage, 2009). Spatial Autoregressive Probit Model is a popular methodology in spatial econometrics that in this paper show to perform well in churning analysis. Moreover, the parameters estimated are easily interpretable. Overall, this paper proposes a new take on the classic churning prediction analysis, which may lead to an improvement in the business model of companies and resource allocation thereof.

The paper is structured as follows: the second section describes the data and methodology; the third section presents the most important results and some potential e-companies' strategies and the last section concludes this work.

References

Abbasimehr, H., M. Setak, and M. Tarokh, 2014, A Comparative Assessment of the Performance of Ensemble Learning in Customer Churn Prediction, *The International Arab Journal of Information Technology*, 11 (6): 599-606.

Allahyari, R. and K. Vahidy, 2012, Applying Data Mining to Insurance Customer Churn Management, *IPCSIT*, 30: 82-92.

Amaral, P.V., L. Anselin, and D. Arribas-Bel, 2013, Testing for spatial error dependence in probit models, *Letters in Spatial and Resource Sciences*, 6(2): 91-101.

Archaux, C., A. Martin, and A. Khenchaf, 2004, An SVM based churn detector in prepaid mobile telephony, in: *International conference on information and communication technologies: From theory to applications*. Proceedings. 2004 International Conference on Information & Communication Technologies.

Baecke, P., and D. Van den Poel, 2012, Including spatial interdependence in customer acquisition models: a cross-category comparison, *Expert Systems with Applications*, 39 (15): 12105-12113.

Baecke, P., and D. Van den Poel, 2013, Improving customer acquisition models by incorporating spatial autocorrelation at different levels of granularity, *Journal of Intelligent Information Systems*, 41 (1): 73-90.

Brockett, P.L., L.L. Golden, M. Guillen, J.P. Nielsen, J. Parner, and A.M. Perez-Marin, 2008, Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?, *The Journal of Risk and Insurance*, 75 (3): 713-737.

Buckinx, B. , Van den Poel, D., (2005) Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting, European Journal of operational Research. 164(1). p.252-268

Burridge, P., J.P. Elhorst, and K. Zigova, 2016, Group Interaction in Research and the Use of General Nesting Spatial Models (p. 223-258), in: J. P. LeSage, K. Pace, and B. Baltagi, eds, *Advances in Econometrics: Qualitative and Limited Dependent Variables*, 37 (Amsterdam: Elsevier).

Coussement, K., and D. Van den Poel, 2008, Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, *Expert systems with applications*, 34 (1): 313-327.

Günther, C.C., I.F. Tvete, K. Aas, G. I. Sandnes, and Ø. Borgan, 2014, Modelling and predicting customer churn from an insurance company, *Scandinavian Actuarial Journal*, 2014 (1): 58-71.

Fotheringham, A.S. 2009, “The problem of spatial autocorrelation” and local spatial statistics. *Geographical Analysis*, *41*(4), 398-403.

Haenlein, M., 2013, Social interactions in customer churn decisions: The impact of relationship directionality, *International Journal of Research in Marketing*, 30 (3): 236-248.

Halleck, S., and Elhorst, J.P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339-363.

Hastie, T., Tibshirani, R., 1986, Generalized Additive Models, *Statistical Science*, Vol 1, Nº 3: 297-318

Huigevoort, C., and R. Dijkman, 2015, *Customer churn prediction for an insurance company*. Ph.D. Thesis

Hung, S.Y., D.C. Yen, and H.Y. Wang, 2006, Applying data mining to telecom churn management, *Expert Systems with Applications*, 31(3): 515-524.

İlhan, B. Y., & İşçioğlu, T. E. (2015). Effect of women’s labor market status on online grocery shopping, the case of Turkey. *Eurasian Business Review*, 5(2), 371-396.

Jeonghye Choi, Sam K. Hui, David R. Bell (2010) Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer. Journal of Marketing Research: February 2010, Vol. 47, No. 1, pp. 75-89.

Kee, H. T., & Wan, D. (2004). Intended Usage of Online Supermarkets: The Singapore Case. In *ICEB* (pp. 1308-1312).

Kelejian, H.H., and I. R. Prucha, 2001, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics*, 104(2): 219-257.

Lacombe, D., and J.P. LeSage, 2015, Use and interpretation of spatial autoregressive probit models, *Annals Regional Science,* doi:10.1007/s00168-015-0705-x.

Lai, Y., & Zeng, J. (2014), Analysis of customer churn behavior in digital libraries. *Program*, 48(4), 370-382.

Larivière, B., and D. Van den Poel, 2005, Predicting customer retention and profitability by using random forests and regression forests techniques, *Expert Systems with Applications*, 29(2): 472-484.

Lemmens, A., and Croux, C., 2006, Bagging and boosting classification trees to predict churn, *Journal of Marketing Research*, 43 (2): 276-286.

LeSage, J. P., 2000, Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models, *Geographical Analysis*, 32 (1): 19-35.

LeSage, J. P., and Pace, R. K., 2009, *Introduction to Spatial Econometrics* (CRC Press)

Löchl, M., H. R. Hauri, and K. Axhaussen, 2009, *Agents, space and market shares: a Spatial analysis of the Swiss Insurance Market*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.

Martinetti, D., and G. Geniaux, 2017, Approximate likelihood estimation of spatial probit models, *Regional Science and Urban Economics*, 64: 30-45.

McMillen, D. P., 1992, Probit with spatial autocorrelation, *Journal of Regional Science*, 32 (3): 335-348.

Morik, K., and H. Köpcke, 2004, Analysing customer churn in insurance data–a case study, in: *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 325-336) (Heidelberg, Berlin: Springer).

Moriuchi, E., & Takahashi, I. (2015). Satisfaction trust and loyalty of repeat online consumer within the Japanese online supermarket trade. *Australasian Marketing Journal*. DOI: 10.1016/j.ausmj.2016.02.006

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H., 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks 11* (3), 690–696.

Neslin, S. A., S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, 2006, Defection detection: Measuring and understanding the predictive accuracy of customer churn models, *Journal of Marketing Research*, 43(2): 204-211.

Pinheiro, C., and M. Helfert, 2010, Neural Network and Social Network to enhance the customer loyalty process, in: *Innovations and Advances in Computer Sciences and Engineering* (pp. 91-96) (Netherlands: Springer).

Pinkse, J. and M.E. Slade, 1998, Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models, *Journal of Econometrics*, 85: 125-154.

Reichheld, F.F., Sasser, W.E., 1990. Zero defections: Quality comes to services. *Harvard Business Review 68 (5)*, 105–111.

Richards, T. J., Hamilton, S. F., & Allender, W. J. (2014). Social networks and new product choice. *American Journal of Agricultural Economics*, 96(2), 489-516.

Risselada, H., P. C. Verhoef, and T.H. Bijmolt, 2010, Staying power of churn prediction models, *Journal of Interactive Marketing*, 24(3): 198-208.

Rosset, S., E. Neumann, U. Eick, and N. Vatnik, 2003, Customer lifetime value models for decision support, *Data Mining and Knowledge Discovery*, 7: 331-339.

Saradhi, V. V., and G.K. Palshikar, 2011, Employee churn prediction, *Expert Systems with Applications*, 38(3): 1999-2006.

Soeini, R., and K. Rodpysh, 2012, Applying Data Mining to Insurance Customer Churn Management, *International Proceedings of Computer Science*, 30: 82-92.

Torkzadeh, G., J. C. J. Chang, and G. W. Hansen, 2006, Identifying issues in customer relationship management at Merck-Medco, *Decision Support Systems*, 42(2): 1116-1130.

Tsai, C.F., and Y.H. Lu, 2009, Customer churn prediction by hybrid neural networks, *Expert Systems with Applications*, 36: 12547-12553.

Vafeiadis, T., K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, 2015, A comparison of machine learning techniques for customer churn prediction, *Simulation Modelling Practice and Theory*, 55: 1-9.

Verbeke, W., D. Martens, C. Mues, and B. Baesens, 2011, Building comprehensible customer churn prediction models with advanced rule induction techniques, *Expert Systems with Applications*, 38(3): 2354-2364.

Verbeke, W., K. Dejaeger, D. Martens, J. Hur, and B. Baesens, 2012, New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, *European Journal of Operational Research*, 218 (1): 211-229.

Xie, Y., X. Li, E. W. T. Ngai, and W. Ying, 2009, Customer churn prediction using improved balanced random forests, *Expert Systems with Applications*, 36(3): 5445-5449.

Palabras Clave: churn prediction, insurance, spatial autocorrelation, spatial logit model, Madrid

Clasificación JEL: