Incorporating Neighborhood Effects in Customer Relationship Management Models

Philippe Baecke and Dirk Van den Poel

Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, B-9000 Ghent, Belgium

Abstract. Traditional customer relationship management (CRM) models often ignore the correlation that could exist in the purchasing behavior of neighboring customers. Instead of treating this correlation as nuisance in the error term, a generalized linear autologistic regression can be used to take these neighborhood effects into account and improve the predictive performance of a customer identification model for a Japanese automobile brand. In addition, this study shows that the level on which neighborhoods are composed has an important influence on the extra value that results from the incorporation of spatial autocorrelation.

Keywords: Customer Intelligence; Predictive Analytics; Marketing; Data Augmentation; Autoregressive Model; Automobile Industry.

1 Introduction

Besides the data mining technique, the success of a CRM model also depends on the quality of the information used as input for the model [1]. Traditional CRM models often ignore neighborhood information and rely on the assumption of independent observations. This means that customers’ purchasing behavior is totally unrelated to the behavior of others. However, in reality, customer preferences do not only depend on their own characteristics but are often related to the behaviors of other customers in their neighborhood. Using neighborhood information to incorporate spatial autocorrelation in the model can solve this violation and significantly improve the predictive performance of the model.

Several studies have already proven that spatial statistics can produce interesting insights in marketing [2-8]. However, only a limited number of studies use spatial information to improve the accuracy of a predictive CRM model. In reference [9], customer interdependence is estimated based on geographic and demographic proximity. The study indicates that geographic reference groups are more important than demographic reference groups in determining individual automobile preferences. Reference [10] shows that taking zip-code information into account can significantly improve a model used for the attraction of new students by a private university. The focus of this study will also be on incorporating physical geographic interdependence to improve CRM models, but, compared to this previous literature, this study includes a large number of independent socio-demographic and lifestyle variables that are typically available at an external data vendor. This should avoid that the predictive improvement could...
be caused by the absence of other important variables that can easily be obtained for customer acquisition models.

In addition, this article introduces an extra complexity that is mostly ignored in previous literature. Customers can often be clustered in neighborhoods at multiple levels (e.g. country, district, ward, etc.). In order to incorporate these neighborhood effects efficiently, the level of granularity should be carefully chosen. If the neighborhood is chosen too large, interdependences will fade away because the preferences of too many surrounding customers are taken into account that do not have any influence in reality. On the other hand, choosing neighborhoods that are too small can affect the reliability of the measured influence and ignore the correlation with some customers that still have an influence. This study will compare the relevance of taking spatial neighborhood effects into account at different levels of granularity.

In this paper, neighborhood information is used to incorporate spatial autocorrelation in a customer acquisition model for a Japanese car brand. Within CRM models, customer acquisition models suffer often the most from a lack of data quality. A company’s customer database is typically single source in nature. The data collection is limited to the information a company retrieves from its own customers. As a result, for customer acquisition campaigns the company has to attract data from external data vendors. Nevertheless, this data still only contains socio-demographic and lifestyle variables [11]. Especially in such situation, incorporating extra neighborhood information can improve the identification of potential customers.

The remainder of this article is organized as follows. Section 2 describes the methodology, consisting of the data description and the generalized linear autologistic regression model used in this study. Next, the results are reported in Section 3 and this paper ends with conclusions in Section 4.

2 Methodology

2.1 Data Description

Data is collected from one of the largest external data vendors in Belgium. This external data vendor possesses data about socio-demographics and lifestyle variables from more than 3 million respondents in Belgium. Furthermore, it provides information about automobile ownership in December 2007 of a Japanese automobile brand.

The model in this study has a binary dependent variable, indicating whether the respondent owns the Japanese automobile brand. Based on this model, potential customers with a similar profile as the current owners can be identified. These prospects can then be used in a marketing acquisition campaign. Because a customer acquisition model typically cannot rely on transactional information, 52 socio-demographic and lifestyle variables are included as predictors.

Further, also information about the geographical location of the respondents is available. Table 1 illustrates that respondents can be divided into several mutually exclusive neighborhoods at different levels of granularity. This table presents seven granularity levels together with information about the number of neighborhoods at each level, the average number of respondents and the average number of owners in each neighborhood.