



Airline customer lifetime value estimation using data analytics supported by social network information

Ahmet Birol Çavdar^{a,*}, Nilgün Ferhatosmanoğlu^b

^a HAVELSAN A.Ş., Training and Simulation Systems, Turkey

^b University of Turkish Aeronautical Association (UTAA), Industrial Engineering, Turkey

ARTICLE INFO

Keywords:

Social network analysis
Regression and classification models
Time series
Forecasting methods
Customer profiling
Data analytics
Customer relationship management

ABSTRACT

Companies can improve their customer relationships and business performance via analytical applications such as estimation of customer lifetime value (CLV) and profitability, customer profiling and classification, customer retention and churn analyses. Customer Relationship Management (CRM) tools can now have access to relationship and interaction data of the customers, besides the traditional data sets such as billing information. While there has been a sharp increase in mining social and interaction data, integration of this information with the current data analytical models is limited. In this paper, we develop a new model for estimating the customer lifetime value in airline industry that integrates customers' social network and flight information. We first adopt a regression model for airline customers that can be used to estimate their CLVs. We then present a methodology to enhance this base model with customers' social network information to incorporate indirect contributions the customers make. We compare the performances of both models to show that our proposed method may improve the accuracy and reliability of models that make use of only flight related factors. We provide examples to potential customer analyses using our models for use by airline CRM applications.

1. Introduction

Estimating the customer lifetime value (CLV) is an essential problem in many industries, including air transportation (Tirenni et al., 2007). The traditional approach to assess the customer value is to use customer centric data such as flight records. There are several problems with this approach. First, the company might have incomplete or inaccurate information about some of their customers. In fact, airlines have difficulties in collecting even basic flight data about their customers. Frequent flyer programs help with this regard, yet they are limited to a small number of customers and they may still be incomplete. Second, even when the data about each customer are complete, CLV is not limited to what the customer herself pays. Some customers might have indirect yet significant ways of contribution for the company that is not measured by the traditional methods that consider each customer independently. In this paper, we hypothesize that relationship among customers and their social network information can help to resolve these problems of incompleteness and indirect value estimation. We propose to incorporate social network information into estimation of CLV and develop a methodology that can be used for this purpose.

We aim to infer useful information about customers by using the data about their neighbors in the social network, or data about people

with similar characteristics in the network. For this purpose, we propose to combine the customers' flight data and their social network information as independent variables in a multiple linear regression model.

One can potentially determine the relationships between users using public information such as Facebook friendships. Besides public social networks, companies can also create their own social networks from their customer activity logs and customer databases. The same applies for the air transportation domain. With the help of this social network information, business intelligence products can be developed. For example, if a person's many friends have begun to prefer another company, the possibility of choosing the same preference will increase for that person (Dasgupta et al., 2008). Such analyses for social networking have an important role in business intelligence and customer relationship management applications.

There is a wide range of social network information that we investigate in our context, such as the social scores including degree centrality, closeness centrality, betweenness centrality, page rank and hub score. The structure of a social network is defined as a graph. There exist studies in the literature that try to find patterns on a graph using theoretical graph algorithms. A fundamental example of these studies are network influence measures that are generally used by the search

* Corresponding author.

E-mail addresses: abcavdar@havelsan.com.tr (A.B. Çavdar), nilgunf@thk.edu.tr (N. Ferhatosmanoğlu).

engines such as PageRank (Brin and Page, 1998) and Hyperlink-Induced Topic Search (HITS; also known as hubs and authorities) (Kleinberg, 2007). In their original forms, these algorithms have been developed to model the relationships between the web pages instead of the social network nodes. However, different versions of them for social network analysis studies have been also developed. There are also many studies showing that people are affected by each other or similar people do similar behaviors (Trusov et al., 2009). We are not aware of any approach in the literature that incorporates social network information into estimation of CLVs for air transportation industry.

The studies in social network literature can be grouped into two categories: i) Structural analysis of social networks, and ii) Data mining using social networks. The research in the first group, which has also been the subject of sociological studies, has started with the experiment of Milgram (1967) that follows chains of mail letters manually. Similarly, Granovetter (1973) presented one of the first studies of examining the indirect structure of the networks. Watts and Strogatz (1998) have observed the small world, grouping and short path features in many social network applications. It is also observed that people affect each other's behaviors and adjacent people act similarly (Trusov et al., 2009). As a result of this, in addition to the traditional media, social networks have emerged as a new media for advertising. Such methods and models for social networks have been discussed in detail in the books: Wasserman and Faust (1994), Watts (2004), and Carrington et al. (2005).

With the advancement of communication technology, the social network data generated by electronic chat environments like electronic mail, Skype, Google Talk or MSN Messenger have increased significantly (Leskovec and Horvitz, 2008). Finally, Web 2.0 technologies such as Facebook, Twitter, Myspace and Orkut have further diversified the social network data structures. As a result, mining social media has also gained momentum (Kleinberg, 2007).

The need for interpreting various contents such as text and multimedia over social networks has led to the use of content analysis methods used in communication, journalism, sociology, psychology and many other disciplines. One of the content analysis application areas related to our study is the computer-aided examination of the social media posts. Neuendorf (2016) covers all aspects of content analysis including theoretical background, how it can be applied to different types of content, and the tools that can be used in computer-aided applications.

The real physical and social environments of the people affect their flight patterns and travel habits. Considering these factors in defining airline customer value makes it possible to create more customized and more comprehensive models. Information related to the real physical environment is typically gathered from the location data. Advances in localization techniques such as geo-tagging enable a deeper understanding of users' preference and behavior by utilizing the location data that bridge the gap between the physical and digital worlds. Bao et al. (2015) survey the novel recommendation systems that make use of the geo-spatial data in social networks by analyzing the data source used, the methodology employed, and the objective of the recommendation. Information related to real social environment of the people, which is generally called as ground-truth communities, can be extracted from the interest based social groups in the social networks. Yang and Leskovec (2015) propose a methodology which allows comparison and quantitative evaluation of how different structural definitions of network communities correspond to ground-truth communities. These findings about real physical and social environment of the customers can be used to selection of the social network related candidate independent variables of the CLV model.

There are studies that formally define the *influence* concept (Tang et al., 2009). Many work on network mining lies on the basis of these studies. Another area of social networking has found a wide application area is security applications. Especially significant research is conducted on monitoring the terrorist activities (Krebs, 2001) and

epidemic diseases (Eubank et al., 2004).

Data mining for CRM has been successfully used in the telecommunications and financial sectors. For example, support vector machine (SVM) based data classification is used in customer credit scoring system (Chen et al., 2009). Regression is used in customer loyalty (Lariviere and Vandenpoel, 2005). Decision tree, naive Bayes, and k-NN are used in marketing (Jiang and Tuzhilin, 2006). These approaches are not integrated with the social network information either. However, applications in the field of credit scoring, churn analyses and marketing can be extended with the customers' social network information. While our study is mainly focused on customer value determination and scoring, our methods can be applied in similar other CRM applications, as well.

In the rest of the paper, we present our method to model the customer value in air transportation industry which enriches the traditional data analysis methods with the customer's social position and relationships. We present the performance evaluation results which show significant increases in several metrics, such as the *adjusted R²* and *predicted R²* values and confidence intervals. The results confirm that replacing the base model factors with social score factors lead to significant increases in the determination coefficient of the model. Although the increases are significant, the results need to be used cautiously because they are indicative numbers based on mostly synthetic data. Synthetic data usage mainly resulted from the fact that our work requires access of commercially secret and/or private information such as customers' flight transactions, ticket fares, customers' friendship relationships etc. In fact, the airline companies' customer relationship departments (or other related units) have access to this commercially secret and private information. Within this context, we aimed to show how related departments of the airline companies can combine the flight and the social network information in order to determine the customer value of their customers.

We observe that when the social factors added to the model, besides the traditional factors, majority of the list of most valuable customers change, which show the promise of including social network information for estimation of CLV. We categorize customers based on their estimated CLVs and visualize the results. We also exemplify a churn analysis scenario using our model.

2. Materials and methods

Our method is based on a multiple linear regression model enhanced with social scores that we gather from an underlying social network. We first present our methodology to generate a realistic data set by combining an anonymized customer flight data set and a publicly available social network data set. We present our proposed method by illustrating how we model CLV using this data set. We first fit a model using customer independent attributes, we then investigate whether enhancing this model with social scores improves the estimation accuracies. Throughout our work, we have used Microsoft Excel for editing data and converting data files to other formats and RStudio software, which is a development tool for R statistical computing environment, for all kinds of our programming needs (web ref [1] and [2]).

2.1. Flight data

We have used an anonymized dataset that includes flight information about the members of an airline frequent flyer program. This dataset consists of Date, Flight, Class, Activity, Description, Bonus Miles and Status Miles attributes as seen in Table 1. We have added a Distance attribute, measured in miles, to this dataset by using source and destination airport information as inputs for the flight distance calculator service provided by the travelmath web site (web ref [3]).

The data for the mile program members have been converted to the format given in Table 2. All but one of the attributes, MonthlyRecency,

Download English Version:

<https://daneshyari.com/en/article/7435237>

Download Persian Version:

<https://daneshyari.com/article/7435237>

[Daneshyari.com](https://daneshyari.com)