



Calling communities analysis and identification using machine learning techniques

Keivan Kianmehr^a, Reda Alhajj^{a,b,*}

^a Department of Computer Science, University of Calgary, Calgary, Alberta, Canada

^b Department of Computer Science, Global University, Beirut, Lebanon

ARTICLE INFO

Keywords:

Social communities
Classification
Clustering
Customer behavior
Machine learning

ABSTRACT

The analysis of social communities related logs has recently received considerable attention for its importance in shedding light on social concerns by identifying different groups, and hence helps in resolving issues like predicting terrorist groups. In the customer analysis domain, identifying calling communities can be used for determining a particular customer's value according to the general pattern behavior of the community that the customer belongs to; this helps the effective targeted marketing design, which is significantly important for increasing profitability. In telecommunication industry, machine learning techniques have been applied to the Call Detail Record (CDR) for predicting customer behavior such as churn prediction. In this paper, we pursue identifying the calling communities and demonstrate how cluster analysis can be used to effectively identify communities using information derived from the CDR data. We use the information extracted from the cluster analysis to identify customer calling patterns. Customers calling patterns are then given to a classification algorithm to generate a classifier model for predicting the calling communities of a customer. We apply different machine learning techniques to build classifier models and compare them in terms of classification accuracy and computational performance. The reported test results demonstrate the applicability and effectiveness of the proposed approach.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Though started as subfield of sociology to identify and study the relations between people and the things they do, with the advance in the computing based prediction and analysis techniques, social network analysis in general and community identification in particular has received considerable and increasing attention over the past decade in different domains, including web mining (Flake, Lawrence, & Giles, 2000; Kleinberg, 1999), biological networks (Girvan & Newman, 2002), among others. So, identifying social communities is an emerging research area that has already attracted the attention of many research groups. The main theme is to analyze logs that reflect social communication between different parties. The analysis leads to valuable discoveries that may have essential social and economical impact. From social perspective, the discoveries may highlight terrorist groups, family relationships, friendship, etc. From economical perspective, the analysis may lead to certain target customer groups. We concentrate on the latter perspective in this paper. In particular, we investigate customer relationships by analyzing call detail records obtained from a telecommunication company. Other researchers concen-

trated on identifying terrorist groups (e.g., Nasrullah & Larsen, 2006).

In telecommunication, a Call Detail Record (CDR) is a record containing information about recent system usage, such as the identities of sources (points of origin), the identities of destinations (endpoints), the duration of each call, the amount billed for each call, the total usage time in the billing period, the total free time remaining in the billing period, and the running total charged during the billing period. The format of the CDR varies among telecom providers and call-logging softwares.

In recent years, we witness dramatic increase in the competition among telecommunication companies in order to detain their current customers and acquire new ones. For this reason, the ability to dynamically classify and predict customers' behaviors according to their calling patterns obtained from CDR data has attracted considerable attention in the research community; it is beneficial for (Yan, Fassino, & Baldasare, 2005):

- *Churn Prediction:* The goal is to understand when and why company's customers are likely to leave so that appropriate action can be planned. Customers become "churners" when they discontinue their subscription and move their business to a competitor company. This has been developed in the telecommunication industry using data mining techniques. Data mining is applied in this area to perform two major tasks:

* Corresponding author. Address: Department of Computer Science, University of Calgary, Calgary, Alberta, Canada

E-mail address: alhajj@cpsc.ucalgary.ca (R. Alhajj)

1. predict whether a particular customer will churn and when this will happen;
2. understand why particular customers churn.

By predicting which customers are likely to churn, the telecommunication company can reduce the rate of churn by offering the customers new incentives to stay.

- *Identifying Calling Communities*: This can be used for determining a particular customer's value according to the general pattern behavior of the community that the particular customer belongs to. This helps the effective targeted marketing design, which is significantly important for increasing profitability in the telecommunication industry.

The main focus of this study is to use an unsupervised machine learning technique, namely clustering to classify customers of a mobile service provider into appropriate calling communities according to the statistics extracted from the CDR data.

Clustering is one of the most prominent approaches for identifying unknown classes amongst a group of objects, and has been successfully used as a tool in many fields such as biology, image analysis, finance, etc. The classification algorithms evaluated in this paper use an unsupervised learning mechanism, wherein unlabeled training data is grouped based on similarity. Once an acceptable clustering has been found using the similarities and dissimilarities in the training data set, the clustering is transformed into a classifier by employing a classification technique.

In our approach, the clusters are labeled, and a new object is classified with the label of the cluster to which it is most similar. However, the type of the classifier that is to be used to identify calling communities and customers' value is very crucial. For instance, in a marketing campaign, selecting inappropriate customers is very costly for verifying the impact of the marketing campaign when it is intended to modify the strategy toward the target group. In this study, the agglomerative hierarchical clustering approach has been applied for the clustering task. It can produce an ordering of the objects (cluster tree), which may be more informative for the nature of the data being analyzed and investigated.

For building the classifier model, we use supervised Machine Learning (ML) techniques to automatically classify and identify customers' communities based on customers' characteristics extracted from clusters. A ML algorithm constructs a classifier by exploring through data objects (training set) to find a set of rules which determine the class of each object according to its attributes. These rules are later used to predict the class or missing attribute value of unseen objects whose class might not be known. Supervised machine learning techniques have been widely used in real world classification applications such as increasing revenue, preventing theft, medical diagnosis, market analysis signaling potentially fraudulent transactions, etc. In this paper, we evaluate and compare the performance of our classifier model, in terms of accuracy and time complexity, by applying several machine learning techniques. The reported results are promising.

The rest of this paper is organized as follows. Section 2 covers CDR data and calling neighbors. Section 3 presents details of the proposed models. Section 4 reports experimental results. Section 5 is conclusions.

2. CDR Data and calling neighbors

Because of the confidential nature of the customers' information and to preserve the privacy, the information provided by telecommunication companies is very limited for modeling except the CDR data. From the CDR data, it is possible to extract customer's calling destination numbers, duration and frequency for each des-

ination number for a particular period of time. Customer's calling neighbors can be also identified by using the links of who calls whom (Yan et al., 2005). There are two types of calling neighbors:

1. *Direct calling neighbor*: a person who calls the customer or whom the customer calls. The majority of these neighbors of a customer may be outside of the service provider's own network, and no information about them is available. An example of direct calling neighbor can be described as follows: members of a research community call each other heavily.
2. *Indirect calling neighbor*: a person who calls the same number(s) as another customer does. For example, employees of a large organization such as a bank who work in different local branches may call their headquarter frequently, and each employee is possibly a calling neighbor of other employees. Even though employees of different branches may not know each other, they can be classified into the community of colleagues.

To measure the closeness of the first type of calling neighbors, the calling frequency or duration has been used. For the second type of neighbors, a distance measure, which will be presented in more detail in Section 3.2, has been applied to quantify the closeness of a neighbor. Discovering calling neighbors will result in forming calling communities. However, there are two major challenges in using the CDR data for finding the calling communities:

- The majority of the destination phone numbers are outside of the service provider's network. Therefore, information about these customers such as their calling records and links between them is not available.
- The other type of phone numbers are those in the service provider's network, and each of them corresponds to a customer of the service provider. The connectivity between these customers is very sparse because customers more likely call numbers outside their home network.

After identifying the customers communities, the information derived from the calling communities can be used for building a classifier model. This classifier model is able to assign a particular customer to a community according to his/her general calling pattern extracted from the CDR data as well as the closeness of his/her direct and indirect calling neighbors. In terms of market campaign design, the characteristics of communities can be used for increasing the profitability of the service provider company by targeting appropriate communities and customers. Intuitively, people in the same community have more likely similar behaviors because of their influences on the community. For instance, if a customer more frequently calls member of a community in which members generally accept service promotion offers, that particular customer can be considered for targeted service promotions.

3. Model details

In this section, the classifier model that has been developed for the course of this study, will be described in more detail. The system consists of three major phases: (1) data preprocessing, (2) clustering, and (3) classification.

3.1. Data preprocessing

The CDR data used in this work was given by a telecommunication company providing wireless services. Because of the confidential nature of the data and business rules, the CDR data for a short period of time is given. However, the results show that the classifier model built based on the information extracted from calling

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات