Expert Systems With Applications 106 (2018) 55–65

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Time series for early churn detection: Using similarity based classification for dynamic networks

María Óskarsdóttir[a,*], Tine Van Calster[a], Bart Baesens[a,b], Wilfried Lemahieu[a], Jan Vanthienen[a]

[a] KU Leuven, Dept. of Decision Sciences and Information Management, Naamsestraat 69, Leuven B-3000, Belgium
[b] University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom

A R T I C L E I N F O

Article history:
Received 29 January 2018
Revised 13 March 2018
Accepted 3 April 2018
Available online 6 April 2018

Keywords:
Multivariate time series
Churn prediction
Call detail records
Time series classification
Social networks
Dynamic networks

A B S T R A C T

The success of retention campaigns in fast-moving and saturated markets, such as the telecommunication industry, often depends on accurately predicting potential churners. Being able to identify certain behavioral patterns that lead to churn is important, because it allows the organization to make arrangements for retention in a timely manner. Moreover, previous research has shown that the decision to leave one operator for another, is often influenced by the customer's social circle. Therefore, features that represent the churn status of their connections are usually good predictors of churn when it is treated as a binary classification problem, which is the traditional approach.

We propose a novel method to extract time series data from call networks to represent dynamic customer behavior. More precisely, we use call detail records of the customers of a telecommunication provider to build call networks on a weekly basis over the period of six months. From each network, we extract features based on each customer's connections within the network, resulting in individual time series of link-based measures. The time series are then classified using the recently proposed similarity forests method, which we further extend in various ways to accommodate multivariate time series. We show that predicting churn with customer behavior represented by time series is a suitable option. According to our results, the similarity forests method together with some of our proposed extensions, perform better than the one-nearest neighbor benchmark for time series classification. Using a time series of a single feature only, the similarity forests method performs as good as traditional churn prediction methods using more features. In fact, compared to traditional methods, similarity forests based approaches perform better when predicting further in the future, and are therefore better at detecting churn early, allowing organizations to make arrangements for retention in a timely manner.

1. Introduction

The mobile telecommunication industry (telco) is a saturated and fast-moving market where customers can easily change providers to obtain a subscription which they are satisfied with. At the same time, it is less profitable for providers to attract new customers than to prevent current customers from quitting (De Caigny, Coussement, & De Bock, 2018). Therefore, providers engage in building predictive models in order to identify which customers are most likely to leave –or churn– so they can be offered promotions to persuade them to stay (Verbeke, Martens, Mues, & Baesens, 2011). In business-oriented contexts, where costs and benefits are a concern, it is furthermore valuable to detect the potential churners early enough for the campaigns to achieve optimal success and maximize their return (Verbraken, Verbeke, & Baesens, 2013). Research has shown that features that incorporate the influence of prior churners in a customer's social circle –their ego-net– are usually good predictors of churn when it is treated as a binary classification problem (Óskarsdóttir et al., 2017; Phadke, Uzunalioglu, Mendiratta, Kushnir, & Doran, 2013; Verbeke, Martens, & Baesens, 2014). These features are extracted from call networks which are constructed based on call detail records (CDR) by linking together customers who have called or texted each other. The result is a representation of the customers' calling behavior and social circle that can be utilized to predict churn with enhanced accuracy (Óskarsdóttir et al., 2017).

* Corresponding author.
E-mail addresses: maria.oskarsdottir@kuleuven.be (M. Óskarsdóttir), tine.vancalster@kuleuven.be (T. Van Calster), bart.baesens@kuleuven.be (B. Baesens), wilfried.lemahieu@kuleuven.be (W. Lemahieu), jan.vanthienen@kuleuven.be (J. Vanthienen).

https://doi.org/10.1016/j.eswa.2018.04.003
0957-4174/

Most studies consider static networks by aggregating the call records for a fixed period of time, e.g. one month (Backiel, Baesens, & Claeskens, 2016; Kim, Jun, & Lee, 2014; Zhang, Zhu, Xu, & Wan, 2012). This approach has an obvious drawback since it fails to take into account the temporal nature of the networks, which in turn might result in a partial and biased representation (Navarro, Miritello, Canales, & Moro, 2017). As human behavior is susceptible to changes over time, it is reasonable to assume that during the time leading up to a customer's decision to churn, they start behaving differently. Furthermore, the behavior among churners might be similar and distinguishable from the behavior of non-churners (Chen, Fan, & Sun, 2012).

In this study, we propose a way of incorporating the time dimension of customer behavior and present a dynamic end-to-end approach to the churn prediction task in telco, which even leads to higher model performance. More precisely, we construct call networks on a weekly basis for a period of six months and extract network features from each customer's ego-net to capture the dynamics of customer churn behavior, using three distinct datasets. This is a novel procedure of processing CDR data as far as we know. To predict customer churn, we then perform classification with the resulting multivariate time series.

Although time series classification is an active research field, an extensive benchmarking study of the numerous methods that have been proposed over the years showed that only complex methods outperformed the one-nearest neighbor (1-NN) with dynamic time warping (DTW) benchmark (Bagnall, Lines, Bostrom, Large, & Keogh, 2017). However, this 1-NN classifier with DTW as a similarity measure, computes all pairwise similarities between the time series which makes it computationally expensive, especially for a large dataset. Although extensive, the benchmarking study (Bagnall et al., 2017) did not include classification of multivariate time series and in fact, the literature on the topic is scarce. Therefore, we turn to a recently proposed classification method called similarity forests, which has not been applied to time series before (Sathe & Aggarwal, 2017). This extension of random forests, which is applicable to any type of data for which similarity between observations is defined, computes only a fraction of the pairwise similarities and is thus orders of magnitude faster and computationally less expensive than the 1-NN DTW benchmark. In addition, it is robust to noise and missing values and performs well in terms of accuracy. We propose novel extensions to the similarity forests method to make it applicable to multivariate time series by, on the one hand, incorporating a multivariate distance measure for time series and, on the other hand, exploiting the functionality of random forests in two distinct ways.

We apply our extended methods together with the original similarity forests to the multivariate time series to classify the customers and compare their performance with the 1-NN DTW benchmark. Thereby, we perform the analyses with variation in the length of the time series, to better understand the dynamics of the churn process. Our results show that the similarity forests perform substantially better than the benchmark method. Finally, we compare our dynamic approach to the more traditional static one, by extracting features from a static network and predicting churn with two commonly used classifiers, i.e. logistic regression and random forests. We demonstrate that time series classification is better suited for detecting churn early. Concretely, we show that by using our proposed method for representing dynamic behavior of customers together with a powerful binary classifier, i.e. similarity forests, telcos have an optimal chance of detecting churners well in advance to take appropriate action for retention.

In the next section, we discuss related literature on dynamic networks and time series classification. In Section 3 we introduce our proposed method of extracting multivariate time series from CDR data to represent dynamic customer behavior. Thereafter, in Section 4, we describe the similarity forests method and our adaptations of the method for multivariate time series. In Section 5, we explain the setup of the experiments conducted in the paper and present the results in Section 6. Finally, we summarize our findings and contributions and discuss future work in Section 7.

2. Theoretical background

2.1. Dynamic networks

Although data used to build networks typically arrives as streams of time stamped transactions, the analysis is most often conducted on static networks, with the information from the data streams aggregated. However, a more natural way, and one that is more representative of the reality of social networks and human behavior and interaction, is to base the analysis on dynamic or temporal networks. This was the topic of a recent special issue of Machine Learning, that focussed on innovative techniques that handle time-evolving networks, while also acknowledging the challenges that arise with such data (Rouveirol, Pensa, & Kanawati, 2017).

In the current literature, there are mainly two approaches for dealing with dynamic networks. One way is to capture the time-dependent network structure in a single network by using temporal edges between the same vertices at different moments. The other approach is to look at time series of networks, i.e. a sequence of static networks, where each network is built by aggregating interactions over a fixed period of time and thus gives snapshots of the network at given times (Santoro, Quattrociocchi, Flocchini, Casteigts, & Amblard, 2011; Simmhan et al., 2014).

Much of the research on dynamics of social networks focuses on the evolution and discovery of communities, and to discover what drives network formation (Kossinets & Watts, 2006). For example, Aiello and Barbieri (2017) looked at the evolution of ego-nets in social media. Using a sequence of network snapshots, they studied how link recommendations affect the expansion of people's social circles. In the case of community detection, the recently proposed Tiles method extracts communities and tracks their evolutions over time (Rossetti, Pappalardo, Pedreschi, & Giannotti, 2017). Furthermore, Lin, Chi, Zhu, Sundaram, and Tseng (2008) proposed FacetNet, a method to analyse, in a single process, communities and how they evolve. Finally, to distinguish peer-to-peer influence from homophily, Aral, Muchnik, and Sundararajan (2009) studied the diffusion of a mobile service product over a social network. Evidently, these studies are mostly aimed towards descriptive analytics, while prediction and classification in dynamic networks, which is the main application in this paper, is scarce in the literature. A few recent publications include an approach for outlier detection in dynamic networks (Ranshous, Chaudhary, & Samatova, 2017), link prediction (Ranshous et al., 2017) and the modeling of epidemic spreading (Joneydi, Khansari, & Kaveh, 2017) and peer influence (Wölbitsch, Walk, & Helic, 2017).

2.2. Time series classification

A lot of research has been conducted on time series classification in the last years and numerous algorithms have been proposed. Many of these were included in an extensive benchmarking study, which showed that the few techniques that outperformed the simple 1-NN classifier with DTW were complex ensembles of methods (Bagnall et al., 2017). The benchmarking study did however not include methods designed for classification of multivariate time series.

Many methods developed for multivariate time series classification depend on either some kind of featurization of the time series data to construct a tabular dataset or on dimensionality reduc-