Online ISSN 1848-3380, Print ISSN 0005-1144
ATKAFF 56(4), 522–534(2015)

Uroš Droftina, Mitja Štular, Andrej Košir

Predicting Influential Mobile-Subscriber Churners using Low-level User Features

DOI 10.7305/automatika.2016.01.880
UDK 004.65.048:654.1.01.03

Original scientific paper

In the last years, customer churn prediction has been very high on the agenda of telecommunications service providers. Among customers predicted as churners, highly influential customers deserve special attention, since their churns can also trigger churns of their peers. The aim of this study is to find good predictors of churn influence in a mobile service network. To this end, a procedure for determining the weak ground truth on churn influence is presented and used to determine the churn influence of prepaid customers. The determined scores are used to identify good churn-influence predictors among 74 candidate features. The identified predictors are finally used to build a churn-influence-prediction model. The results show that considerably better churn prediction results can be achieved using the proposed model together with the classical churn-prediction-model than by using the classical churn-prediction model alone. Moreover, the successfully predicted churners by the combined approach also have a greater number of churn followers. A successful retention of the predicted churners could greatly affect churn reduction since it could also prevent the churns of these followers.

Key words: Churn prediction, User influence, Social network, Weak ground truth, Churn-influence model

Predviđanje odljeva utjecajnih mobilnih pretplatnika korištenjem značajki niske razine. Posljednjih godina, predviđanje odljeva korisnika jedna je on važnijih tema među pružateljima telekomunikacijskih usluga. Među odlazećim korisnicima, oni najutjecajniji zaslužuju posebnu pažnju, jer njihov odljev može okinuti i odljev sljedbenika. Cilj ovog članka je pronalazak dobrih prediktora utjecaja odljeva na mobilne uslužne mreže. U tu svrhu, razvijena je metoda za njihovu identifikaciju među 74 potencijalna kandidata. Identificirani prediktori su potom korišteni za konačnu izgradnju modela predviđanja odljeva korisnika. Znatno bolji rezultati ostvaruju se kada se koristi predloženi model u kombinaciji s klasičnim modelom, nego kada se klasični model koristi zasebno. Štoviše, kombiniranim predviđanjem izdvojeni utjecajni korisnici imaju veći broj sljedbenika. Uspješno zadržavanje predviđenog odljeva moglo bi uvelike utjecati na njegovo smanjenje, pošto bi samim time spriječilo i odljev sljedbenika.

Ključne riječi: predviđanje odljeva, utjecaj korisnika, društvena mreža, slabi referentni podaci, model utjecaja odljeva

1 INTRODUCTION

The telecommunication sector is one of the most profitable and the most competitive in the world. It has been estimated that around 5 trillion dollars were spent in 2013 in this sector [1]. Competition between telecommunication service providers has become even stronger where the public regulators enabled customers to easily churn, i.e. move from one service provider to another. Annual churn rates with certain service providers extend up to 40 % [2], but also considerably smaller churn rates (under 5 % annually) can inflict a great damage to the company. Due to many telecommunication markets around the world getting saturated, service providers are faced with two options to increase or at least retain their market share: (i) to retain existing customers, and (ii) to acquire new customers from competing providers. Since the cost of retention is about one-sixth of that of acquisition [3], providers mainly focus on the first action, known also as customer churn prevention. This can be achieved by improving customer satisfaction, e.g., by offering competitive bundles and by improving their services [4]. This approach, however, is not optimal, since it acts on all customers of the provider, both future churners and non-churners. A more effective way to churn prevention is by a more personalised approach, i.e., by predicting possible churners before they actually churn, and only acting on them.

The most widely used approaches to churn prediction use machine learning and data mining techniques to build classification models [5]. An extensive benchmarking of various known classification techniques from the literature was performed by Verbeke et al. [6]. Using such models (henceforth, we refer to them as the classical models), the customers are assigned churn probability scores, which are necessary for labelling them as churners or non-churners. Then, the top k-percent (typically $\{k : 0.1\% \leq k \leq 10\%\}$ [7]) of customers determined most likely to churn are offered incentives to prevent them from actually churning. However, not all customers are equally valuable to the company. Hence, not all predicted churners should be treated equally. One intuitive way of ranking customers by their value to the company is considering the average revenue that a customer creates to the company (known as Average Revenue Per Unit – ARPU). However, user influence is arguably even more important when determining the value of a customer to the company. An influential customer could potentially inflict greater damage to the company if she churned, because her churn could initiate a chain reaction within her social network (SN). Therefore, successfully predicting influential churners, and preventing them from churning, can result in greater savings than only by taking actions on potential churners with higher ARPU.

A number of different techniques for determining the most influential users in a SN has been examined [8–13], but most of them assume that a social graph with edges weighted with influence scores is already given. In fact, there is a gap between the work done on this assumption, and real SN dataset that do not contain influence scores but some proxy scores. Example of such proxy scores is the amount of interactions between users (calls, messages) [14, 15] or the number of direct connections of each user in her social network [16]. Such scores do not necessarily reflect the influence of the observed user.

Smailovic et al. [16, 17] already proposed a definition of general social influence of mobile network users, but their work was limited to users, who also use popular online social networks such as Facebook or Twitter. Furthermore, general user influence does not necessarily reflect the more specific churn influence. In fact, little attention has been given to identify the proxy scores that are good indicators of churn influence.

In this work we present a methodology for the discovery and the evaluation of these indicators. To this end, we evaluate numerous low-level user features as possible churn-influence indicators. Then, we build a model for predicting influential churners using features determined as relevant regarding churn influence. Finally, we evaluate the performance of the proposed churn-influence-prediction model by comparing the performance of the combined models (i.e., the proposed and the classical) against the classical model alone.

A user can churn for many different reasons, e.g., price of services, poor coverage or poor services, but by our beliefs, the reason that can best indicate an influential churner is peer pressure. Therefore, to build a churn-influence-prediction model, a ground truth on peer pressure as the reason for churn is needed. Unfortunately, such data are not typically available. Thus, we design and propose a procedure for determining the weak-ground-truth (WGT) on churn influence. This is done using the real data on consecutive churns among peers in a limited period of time, where we assume that the user who churned first, influenced her peer and caused her to churn as well. The term weak-ground-truth is used because such ground truth data represent the best approximation available to the firm ground truth, which is not available [18]. We base our WGT definition on known models of peer influence (see Subsection 3.2) that exploit available historic churn data to determine WGT scores on churn influence for each user in the dataset.

The contributions of this work include (i) a novel approach for determining WGT on churn influence, (ii) identification of good churn-influence predictors, (iii) and a hybrid approach for predicting the most influential churners in a service provider network.

The remainder of the paper is structured as follows. Section 2 provides a short review of the literature concerning the identification of influential users. Section 3 presents the methodology section, where the procedure for churn-influence WGT determination, extraction of features as churn-influence predictors, and building of a churn-influence prediction model, are presented. Section 4 discusses the experimental results and evaluates the churn-influence prediction model with comparable methods. The paper closes in Section 5 with a conclusion and discussion of open issues and plans for future work.

2 IDENTIFICATION OF INFLUENTIAL USERS

Many studies in recent years have confirmed the spread of influence through social ties for numerous cases, such as the adoption of products and services [19, 20], influence on voting decisions [21], general influence of mobile network users [16, 17], and also influence on churn decisions [14, 15]. Kusuma et al. [22] used real data from a mobile service provider and showed that churn rate of the users with half of their peers already churned is twice the churn rate of all the users. Therefore, to reduce churn most effectively, it would be best to retain the most influential users. To this end, a subset of the most influential users must first be identified. In this section, we provide a short review of the literature concerning the identification of influential users, determine some of the open issues in the literature, and describe the proposed solution.

2.1 Approaches for identifying influential users

Richardson and Domingos [8, 9] were the first to address the problem of identifying the most influential users algorithmically using a probabilistic approach. Kempe et al. [10], on the other hand, considered the influence maximisation problem as an optimisation problem. They determined the problem of finding the most influential nodes to be NP-hard, and presented a solution using a greedy approach that provided an approximation within 63% of the optimal solution to the problem. The authors determined that influence functions have to meet two conditions, monotonicity and submodularity. The proposed method can be explained as follows. At every moment, each node in a SN is either active or inactive. Inactive users become active by adopting a product or performing an action, such as churn. The probability of each node becoming active monotonically increases with every additional active neighbour (monotonicity). Additionally, the more active neighbours a node already has in her SN, the less influence a new active neighbour has on the node (submodularity).

The problem of the proposed approach is its computational complexity, which makes it inappropriate for the usage on large-scale SNs. Therefore, several improvements to the simple greedy algorithm have been proposed in the literature. Leskovec et al. [11] presented an algorithm called Cost-Effective Lazy Forward (CELF) and compared it with the simple greedy algorithm on the task of selecting the most influential blogs from the dataset of 45,000 blogs with 1 million links between them. The usage of CELF exhibited a significant improvement in run time by being up to 700 times faster over the simple greedy algorithm. However, this speed is still not adequate for large-scale networks. Additional improvements were made by Chen et al. [12], who proposed a combination of the simple greedy algorithm by Kempe et al. [10] and CELF [11], which resulted in additional 34% speed up over CELF. Additionally, they proposed new heuristics that improved the spread of influence by six orders of magnitude compared to all greedy algorithms designed up to that point. Chen et al. performed their experiments on an arXiv.org database with the task of finding the 50 most influential nodes in two different networks with (i) 15,000 nodes and 59,000 edges and (ii) 37,000 nodes and 232,000 edges. A different approach was proposed by Wang et al. [13]; instead of running any of the aforementioned algorithms on an entire SN, they proposed a two-step algorithm that (i) clusters nodes into different communities in a SN and (ii) selects appropriate communities to find influential users inside of them using a selected influence maximisation algorithm. Using an effective clustering or community finding algorithm, this approach is able to find the most influential users within a reasonable time even for the largest-scale SNs.

All of the aforementioned influence maximisation algorithms assume that a social graph with influence-weighted edges is given. This is usually not the case with real SNs, the edges of which are usually weighted with the amount of interaction between users, e.g., calls or re-Tweets. Goyal et al. [23] addressed this problem by proposing several models of influence using a social graph and a log of actions.

2.2 Problem Statement

The shortcoming of the approach proposed by Goyal et al. [23] is that it is only useful for predicting user influence regarding the targeted action where users perform the action repeatedly. This is not the case with determining churn influence, because users usually only churn once. Also, when a user churns, it is already too late for the service provider to persuade her against churning.

In this work, we propose determining the WGT on churn influence using the data on connected users that churned successively. We use a modified approach to that proposed by Goyal et al. [23]. The introduction of churn-influence WGT is necessary because firm ground truth on churn influence is usually not known. Then, we extract numerous low-level user features and evaluate them in connection with the WGT to identify good predictors of churn influence. Finally, we use the features determined as relevant regarding churn influence to build a churn-influence-prediction model.

3 MATERIALS AND METHODS

The aim of this study was twofold. Firstly, using the real telco data, we calculated churn-influence scores that were used as the WGT. We classified the churn-influence WGT scores into a binary variable that distinguished “very influential” from “weakly influential or non-influential” future churners. Secondly, using this variable as the target variable, and the features determined as good influence predictors we built a predictive churn-influence model. We evaluated the models using a lift curve [24].

3.1 Data

For our analysis, a European mobile service provider with approximately 1 million customers provided us with their Call Detail Records (CDR) of July 2012 for all their residential prepaid customers, which represented approximately 370 thousand of their users. The data were limited to only prepaid users because they do not have any contractual obligations to service providers and can therefore easily churn at any time. The events included in the CDR were limited only to calls. Each call in the CDR contained the anonymous ID of a caller and a callee, the time stamp of the call, and the duration and cost of each call.

CDR data were used for extracting different user variables and for building a social graph of users as nodes and calls as edges. Among all users in CDR, only active users were considered, i.e., users who made at least one incoming on-net call, one outgoing on-net call, one incoming off-net call, and one outgoing off-net call. This summed up to 150 thousand users with approximately 225 thousand connections between them. Since not all phone connections reflect influence, e.g., calling a mechanic, making an inquiry on a product, or making a hotel reservation, only connections between users with at least five calls and at least one call in each direction were considered as candidates for influence propagation.

Additionally, a mobile number portability (MNP) log was provided for the time period from August 2012 to June 2013. It included all mobile number transfers from a considered service provider to any of its competitors. We used this log as an indicator for churns. The MNP log included two parameters, the anonymous ID of a user that churned or transferred her number, and the date of her churn.

3.2 Churn-influence WGT model

Here, we describe the construction of the churn-influence WGT model, that is based on the continuous time model introduced by Goyal [23]. WGT model alone cannot be used as a method for predicting churn influence since it requires historical data on churns. By the time WGT model detects a churner, she would have already changed the provider.

3.2.1 Rules for constructing the churn-influence WGT model

A SN of all users of a service provider can be represented as a social graph $G = (U, E)$. Users are represented as nodes U, and social ties between users are represented as edges $(u_i, u_j) \in E$. Additionally, a log of churns $C = U \times T$ is established in the form of (u_i, t_c) that describes a user $u_i \in U$ churning at time t_i. The construction of the churn-influence WGT model is based on the following set of simple rules:

RULE 1 Churn-influence of a user u_i to her directly connected user u_j is determined using past pairs of churns $P = C_i \times C_j$ where each pair $p(c_i, c_j)$ consists of a churn c_i of user u_i at time t_i and a churn c_j of user u_j at time $t_j; t_i < t_j$.

RULE 2 The value of single-event influence between directly connected users u_i and u_j for each considered pair $p(c_i, c_j)$ is determined using a single-event influence function $\zeta_e(u_i, u_j, \Delta t)$, that monotonically decays by increasing time $\Delta t = t_j - t_i$ between churns c_i and c_j. In the limit of an infinitely large Δt, the value of ζ_e reaches 0.

RULE 3 The influence of each user u_i is calculated using a churn-influence function $\zeta(u_i)$ that considers all single-event-influence values of user u_i. Function $\zeta(u_i)$ must meet two criteria, monotonicity and submodularity [10].

3.2.2 Churn-influence function

The set of rules defined in the previous subsection, and the churn frequency distribution dependent on the elapsed time between past neighbouring churns, are used to determine the single-event influence function $\zeta_e(u_i, u_j, \Delta t)$. This function is used to determine churn-influence scores for each pair of churns among connected users, and these scores are then used to determine the churn-influence score of each individual user.

The churn frequency distribution shows how many different connected pairs of customers churned 1, 2, 3, or more days or weeks apart. A real example of such distribution extracted from the provided data is shown in Fig. 1, at daily and weekly resolution. The y-axes of the plots are normalised due to the confidential nature of the data. The histogram exhibiting daily resolution (the left plot) shows a weekly fluctuation of churns, which is mainly a consequence of the increase in churns (or mobile number transfers) on Mondays. The reason for this is that service providers usually perform mobile number transfers on the next working day after they receive the number transfer request. Therefore, Monday churn labels actually consist of three aggregated days of churns, i.e., from Friday, Saturday, and Sunday labels.

To eliminate weekly fluctuations, the use of a histogram with the weekly resolution is more appropriate (see the right plot in Fig. 1). Nevertheless, both plots in Fig. 1 clearly show exponential decay behaviour, which indicates that the influence of a user u_i on her neighbours is strongest immediately after the user performs an event. However, we need to consider that both, (i) churns due to influence, and (ii) churns due to other reasons are captured in Fig. 1. We assume that the reasons under (ii) are not dependent on the time difference between neighbouring churns, and we consider these reasons to be equally distributed with respect to the time between neighbouring churns. As a result, we propose a churn function $\omega(t)$ as the sum of an exponential function $\psi(t)$ (influence from neighbours that churned) and a constant value C that includes other reasons for churn independent of neighbour churn, as shown in Eq. (1) and (2), and in Fig. 2.

$$\omega(t) = \psi(t) + C \tag{1}$$

$$\psi(t) = k\mathrm{e}^{-\lambda \Delta t} = \zeta(u_i, u_j, \Delta t) \tag{2}$$

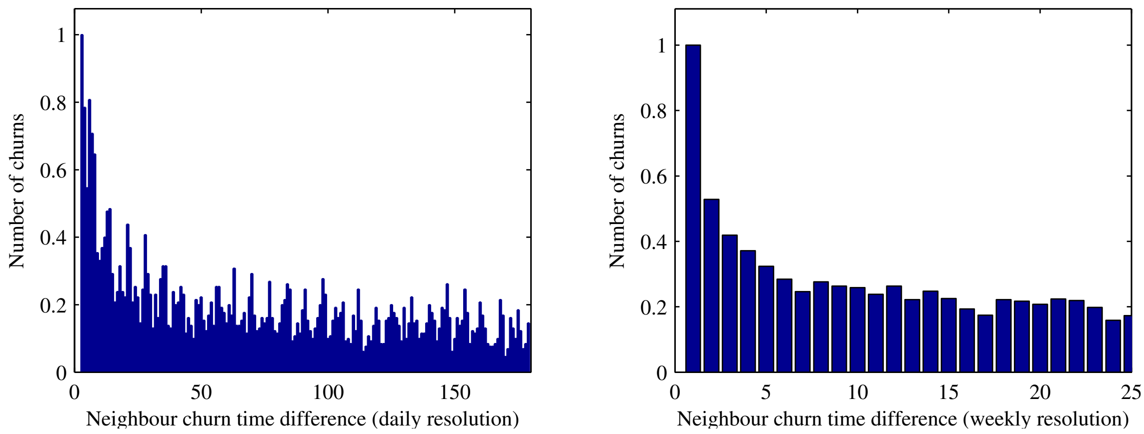

Fig. 1. A histogram of neighbour churn time differences at daily and weekly resolutions. The daily histogram shows a weekly fluctuation because most churns occur on Mondays (after weekends). The y-axes of the plots are normalised due to the confidential nature of the data.

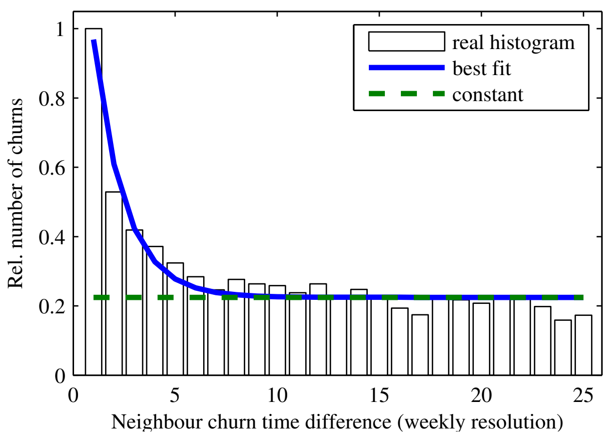

Fig. 2. The best fit to histogram of neighbour churn time differences at weekly resolution. The histogram is fitted with a three-parameter equation of the form $y = ke^{\lambda t} + C$, a sum of an exponential function (neighbour churn influence) and a constant C (other churn reasons).

We fit the parameters k, λ, and C in Eq. (1) and (2) to the already discussed churn frequency distribution using an appropriate fitting method (e.g., least squares). As we focus only on explaining churn due to influence, we only use the exponential part of the equation, $\psi(t)$, as the single-event influence function $\psi(t) = \zeta(u_i, v_i, \Delta t)$. As we are only interested in the ratios between churn-influence scores, we can simplify the function $\psi(t)$ by setting $k = 1$. Therefore, the churn-influence scores of neighbouring churners are only affected by the parameter λ. Since churn influence is naturally decreasing with increasing time Δt, the value of λ must be a real negative number. The closer the value of λ is to zero, the flatter the exponential curve becomes. The value of λ that we determined to be the best fit to the churn frequency distribution in Fig. 2, was $\lambda = -0.66$. The fitting was determined using the non-linear least squares method.

Churn-influence scores are calculated by Eq. (2) for all pairs of u_i and u_j, where both u_i and u_j churned at times t_i and t_j, respectively, and $t_i < t_j$, whereas influence is left undefined for the remaining pairs. Δt is equal to the difference between the events of users $\Delta t = t_j - t_i$.

The churn influence of each user u can be calculated by a number of different approaches that meet the conditions of monotonicity and submodularity. In this work, we use Eq. (3), similar to the solution proposed in [23], where monotonicity and submodularity have already been proven. We denote all neighbours of a user u_i as N_{u_i} and a subset of neighbours of user u_i for which we can determine $\zeta_e(u_i, u_j, \Delta t)$ as $S_{u_i} \subseteq N_{u_i}$.

$$\zeta(u_i) = 1 - \prod_{u_j \in S_{u_i}} (1 - \zeta_e(u_i, u_j, \Delta t_{u_i,u_j})) \tag{3}$$

3.2.3 Calculating the churn-influence WGT scores

Using the Eq. (2) and (3), we were able to calculate churn-influence scores for all the users who already churned, with at least one of their connections churning after them. For simplicity, we set $\zeta(u_i, u_j, \Delta t) = 0$, where $\Delta t > \Delta T_{max}$. Consequently, we eliminated very small and noisy contributions to influence. In practice, if more than ΔT_{max} elapsed between churns of users u_i and u_j, then we can assume that u_j was no longer influenced by u_i, and vice versa. In this work we intuitively set ΔT_{max}

to 180 days (or approximately 26 weeks). Here, the value of $\zeta_e(u_i, u_j, \Delta t)$ is already approximately 10 million times smaller that the value of $\zeta_e(u_i, u_j, \Delta t)$ after 1 week, using Eq. (2).

Additionally, a minimal shift between neighbouring churns ΔT_{min} needed to be determined to calculate churn-influence scores, due to the following reasons:

- Some time is needed for service providers to analyse user and churn data and then perform appropriate retention actions;
- Users are not influenced by their peers instantly but require some time (a few days) to process the new information and take appropriate action;
- Neighbouring users who churn on the same day cannot be considered to influence one another but can be considered to be influenced by another third user or by another external factor;
- By delaying the week counter for at least 3 days, weekend gaps are also considered, when service providers normally do not transfer numbers.

Because the influence on user u_j is still the strongest immediately after the action of her neighbour u_i, ΔT_{min} must be set to the lowest possible value. In consideration of the foregoing arguments, we propose setting $\Delta T_{min} = 3$ days. This setting was also used in this work. Also, a weekly resolution of time difference between churns was used, which was determined to be the most appropriate, as discussed in Subsection 3.2.2. Since ΔT_{min} was set to 3 days, the values of Δt between 3 and 9 days were translated into week 1, the values of Δt between 10 and 16 days into week 2, and so on.

When calculating churn-influence scores, certain factors must be considered. The first factor is the number of users for whom we can calculate influence using Eq. 1 and Eq. 3. Because churn is a relatively rare event (approximately 5% annually for the considered service provider) and churn in pairs of neighbours within a time window $\Delta T_{min} < \Delta t < \Delta T_{max}$ is even rarer, influence can only be determined for a very limited number of users. Another factor to consider is the use of different price plans by service providers. Postpaid users are generally bound to their service provider with a contract and usually do not leave the service provider until their contract expires, although they could be strongly influenced by neighbours who have recently churned. Therefore, considering only prepaid users is more logical because these users are not bound by a contract and can churn at any time without financial consequences.

Considering these constraints, we were able to extract and determine the churn-influence WGT scores for 146 users from the entire available dataset. Although the number of users for whom we could calculate the churn-influence WGT scores was considerably smaller than the number of all users, it was still sufficient for our case, which was using these scores to evaluate numerous extracted user features as possible churn-influence predictors.

3.3 Candidates for churn-influence predictors

The features that we use for churn-influence prediction model must be able to be extracted for most of the service provider users so that we can determine the churn-influence for the widest range of users possible. To this end, 74 different low-level user features were extracted from the data provided as candidates for influence predictors. The extracted candidate features can be grouped into different categories:

Price features This group contains the absolute monthly cost of calls a user made to users in the same network (on-net calls), users in different networks (off-net calls), and all users together (all calls). It also includes the relative costs of on-net and off-net calls vs. all calls. Users with higher monthly costs can save significantly more money by churning to a competitive provider, compared to those with smaller costs. Therefore, it is expected that these users are more informed about the offers of competitive service providers, to choose the provider that suits them best. Being more informed, these users could also have greater influence among their friends and family.

Degree features This group only contains two features, the number of first-degree and the number of second-degree neighbours. A higher degree and therefore a greater number of friends and other contacts could indicate a more influential user; however, more friends does not necessarily mean that a user also has more true friends among whom her influence is the strongest.

Call stats - all calls This group of call stats features contains numbers and durations of on-net calls, off-net calls, and all calls together. These features include the sum of incoming and outgoing calls. More overall calls could indicate a more popular user and, consequently, a more influential user.

Call stats - outgoing calls This group contains similar features as the group *Call stats - all calls* but only for outgoing calls instead of all calls together. The outgoing calls feature is partially related to the features in the group *Price* *features* because outgoing calls are charged by duration. However, price and duration are not completely correlated because the call-charging system of the considered service provider charges for the first full minute even if the call duration is only seconds. Therefore, similar results to those obtained for the group *Price features* are expected.

Call stats - incoming calls This group contains similar features as the group *Call stats - all calls* but only for incoming calls instead of all calls together. Many incoming calls may indicate that other users are asking for advice. Therefore, many incoming calls may also indicate a high level of influence of the user.

Ratios of call stats features This group contains many different variations of ratios between *call stats features*, such as ratios between (i) incoming and outgoing calls, (ii) on-net and off-net calls, (iii) on-net and all calls, and (iv) off-net and all calls. Altogether, this group includes 36 different ratio features.

SPA diffusion model The last group contains the energy parameters extracted from the first four iterations of the spreading activation-based (SPA) technique proposed by Dasgupta et al. [14] used to model the spread of information (or influence) through SNs. The extracted features include initial user energy, user energy after each of the first four iterations of SPA, and incoming and outgoing energy of users in each of the first four iterations. Only the first four iterations are used because users are mainly (if not only) influenced by users connected closely to them, while every additional iteration adds the influence from users that are additional degree of separation away.

3.4 Churn-influence prediction model fitting

The described features can be used to construct a predictive model for churn influence. Because the influence of users determined by the WGT model is a numerical score, it must be converted to a nominal class feature by setting a meaningful threshold. Service providers are mostly interested in the most influential users who have the fastest impact on their friends; therefore, we determine "the most influential" users to be those who have at least one friend who churned after them within a short period of time, and we define others as "weakly influential or non-influential". To the best of our knowledge, there are no data in the literature regarding the real ratios of influential people. Based on preliminary analysis, where different churn-influence WGT score thresholds were evaluated, we determined the value 30 % of churners with the highest values of $\zeta(u_i)$ to be the optimal ratio of "very influential" users. The remaining 70 % of user were labeled as "weakly influential or non-influential". However, a different ratio could also be used.

First, we extracted the most relevant subset of features among all the described candidate features using an appropriate feature selection method, called the Correlation Feature Selection (CFS) [25]. This method evaluates all of the possible subsets of input features by taking into account the level of correlation between each input feature and the class feature (influential vs. non-influential users), and the level of correlation between the input features themselves. The idea is to find such subset of features that highly correlate with the class feature, but are uncorrelated with each other. To determine the level of correlation in CFS, Pearson's correlation is used.

Second, we fitted the model to the influence labels using the most relevant subset of candidate features according to CFS results. The model was built using logistic regression classifier, validated using a 10-fold cross-validation scheme and evaluated using a lift curve [24], that shows the ratio of true positives successfully identified by capturing a specific amount of all instances (users) using a specific model. As a baseline, the models are usually compared with the random approach, which is represented as a straight line connecting points (0,0) to (1,1). Basically, a random model assumes that by selecting n-percent of all users, also n-percent of all true positives is captured. We can also use area under the curve (AUC), e.g., under the lift, to evaluate classification models without limiting to a specific amount of captured instances.

Feature selection, model building, its validation and validation were all performed in Weka, an open-source machine learning software [26], and in Matlab.

4 RESULTS AND DISCUSSION

In this section, we present the results of our study. First, we present the evaluation of churn-influence-prediction model built using the predictors, determined appropriate using CFS. Second, we evaluate the performance of the model in combination with the classic churn-prediction model on a dataset of 150 thousands real users, by comparing it with two other methods, (i) the combination of classical churn-prediction model and the ARPU based scoring model, and (ii) the classical churn-prediction model alone.

4.1 Churn-influence-prediction model

Here, we present the churn-influence-prediction model built with logistic regression and the evaluation of this model using lift.

Among 74 candidates, the best subset of features for building the model was determined using CFS, the feature selection method. Interestingly, only four variables were determined as appropriate by CFS. We present these features together with their respective values of standardized regression coefficients in Table 1. This table reveals that the probability for outcome "very influential" increases with increasing number of on-net calls and decreases with the duration of outgoing on-net calls, the ratio between outgoing and incoming on-net calls, and the ratio between outgoing off-net calls and all outgoing calls. Other candidate features were found either correlated with these four features, or non-correlated with the the target feature, i.e., the churn-influence feature.

Table 1. Standardized regression coefficient values β of the churn-influence prediction model for the outcome "very influential"

Feature	β
number of on-net calls	0.4313
duration of outgoing on-net calls	-0.0313
outgoing vs. incoming on-net calls	-1.4033
off-net vs all outgoing calls	-0.2243
constant (β_0)	-1.1454

Surprisingly, none of the features from the predictor group *SPA diffusion model* were selected as the appropriate features for the model. The SPA diffusion model [14] is based on the fact that in every iteration of the diffusion process, energy is transferred from each user to her peers. At the end of the iterative process, users' energies are used as the scores for their propensities to churn. The more energy a user has, the more she is influenced by others. Therefore, the transfer of energy from user u_i to u_j is considered to be the case of a user u_i influencing a user u_i. Based on this reasoning, energy transfer was expected to be a strong predictor for churn influence; however, it was found to be non-correlated to churn-influence label.

Another feature group that, despite our expectations, showed no connection with churn influence, was the group *Degree features*, in which the variable *degree - order 1* represented the degree centrality of the user. This feature, however, was shown to be a good predictor of influence in the work by Kiss et al. [27]. Therefore, our findings indicate that the use of a ground truth for influence is necessary to determine strong influence predictors.

The evaluation of the built logistic regression model is presented with a lift curve in Fig. 3, in comparison to a random approach.

The lift curve shows that despite using only four variables for the model, the improvement over the random model is clearly presented. By successfully retaining influential users that would churn otherwise, a service provider could significantly reduce the number of churns in their network, because also the number of influence-induced churns would be reduced. In the next section we evaluate the proposed churn-influence model in comparison to a classical churn-prediction model.

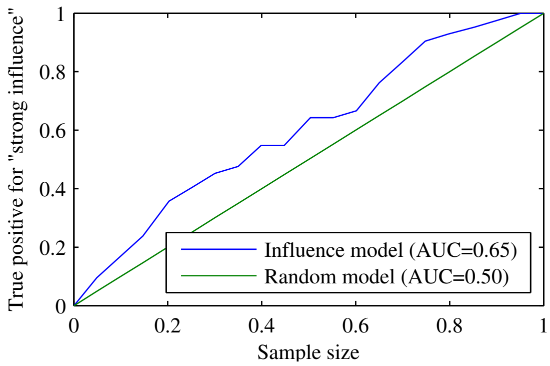

Fig. 3. Lift curve of the churn-influence model in comparison with the random model.

4.2 Evaluation of the churn-influence-prediction model

To determine the value of our approach we combined the obtained 4-variable churn-influence model with a churn-prediction model. Then, we evaluated the combined approach by comparing it to the churn-prediction model alone and the ARPU based scoring model.

We built the churn-prediction model using logistic regression and the same set of user variables as for the churn-influence model building.

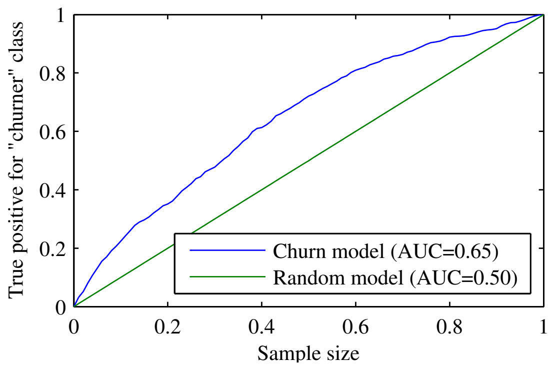

Fig. 4. Lift curve of the churn-prediction model in comparison with the random model.

The dataset included the features, described in Section 3.3, of roughly 150 thousands of prepaid users that were active in July 2012. Additionally, the absolute and relative number of past neighbouring churners, and the number of outgoing SMS messages, were included in the dataset. The users were labelled as churners if they

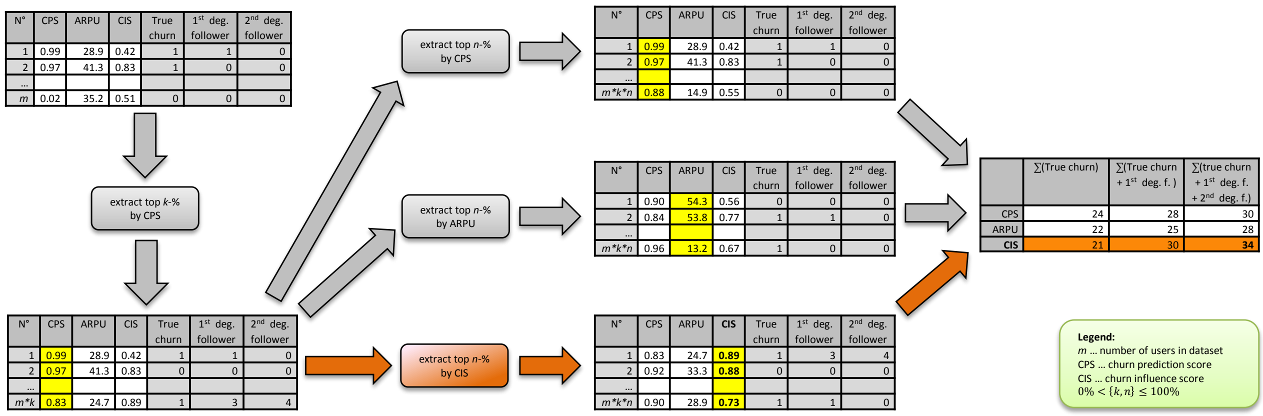

Fig. 5. Graphical presentation of the evaluation of three ranking methods, (i) by churn-prediction score (baseline model), (ii) by ARPU, and (iii) by churn-influence-prediction score (our proposed approach marked with the orange flow).

churned in the period from August 2012 to October 2012. The number of all churners in the dataset was almost 1900 (approx. 1.3%), which means that the annual churn rate of the the considered service provider was around 5% in the prepaid segment in the observed time period. The whole dataset was split to equally sized training and test sets. The churn rate was preserved in each of the sets. Using the CFS, 11 features were determined as the appropriate predictor variables for churn prediction and used for building the churn-prediction model. These features included: (i) price of outgoing off-net calls, (ii) number of all calls, (iii) duration of all off-net calls, (iv) number of all incoming calls, (v) number of incoming off-net calls, (vi) duration of incoming off-net calls, (vii) outgoing vs. incoming duration of all calls, (viii) outgoing vs. incoming number of off-net calls, (ix) relative number of outgoing off-net calls, (x) number of all outgoing SMS messages, and (xi) absolute number of past neighbouring churners. We evaluated the model on the test set with the lift curve (see Fig. 4).

After the churn-prediction model was built, the top k-percent of users with the highest churn-prediction scores according to the model, were extracted. Then, these users were ranked by three different methods and the top n-percent of users was further extracted from the sorted list, for evaluation. The first ranking method extracted the top n-percent of predicted churners with the highest churn-prediction scores, i.e., using the model described above. This method was used as the baseline model. The second method extracted the top n-percent of predicted churners with the highest ARPU, which was obtained by summing the cost of all calls for each user in the CDR dataset. Finally, the third method extracted the top n-percent of predicted churners with the highest churn-influence scores, obtained by the model proposed in this work. The graphical presentation of this evaluation is presented in Fig. 5.

In the first part of evaluation we set n to 40% and changed the value of k from 1% to 20% in 1%-steps. Although, only 20% of the whole possible range of values of k is used, this covers the whole applicable range typically used by service providers $\{k : 0.1\% \leq k \leq 10\%\}$ [7]. For each configuration of k and n and each of the three ranking methods we counted the true positives (predicted churners that actually churned) in the test dataset. Additionally, we counted the neighbours of the true positives that churned within 90 days after them, which we refer to as the 1st degree followers. Also, the number of 2nd degree followers was extracted, i.e. the number of neighbours of the 1st degree followers, that churned within 90 days after them. To properly compare the three ranking methods, we divided the numbers of true positives with the numbers of true positives obtained by the baseline model, and presented the results in Fig. 6. Plot (a) in Fig. 6 compares the three ranking methods by only considering the actual predicted churner, plot (b) also considers the 1st degree followers, and finally, the plot (c) displays the results of considering the true predicted churners together with the 1st and 2nd degree followers.

In the second part, we performed a similar evaluation of the three ranking methods, but by setting a fixed value of $k = 5\%$ (a commonly used value in churn prevention campaigns) and changing the value of n from 5% to 100% in 5%-steps. The results are presented in Fig. 7.

Both, Fig. 6 and 7 show that the best results were achieved using the churn-influence-prediction model as the ranking method. Fig. 6 shows that by considering only the predicted churners (without the followers) the best lift over the baseline model is already achieved using the churn-influence-prediction model, although, only at k between 4% and 6%. The real power of our approach is shown in the Figs. 6.b, 6.c, 7.b, and 7.c, that include the 1st and 2nd

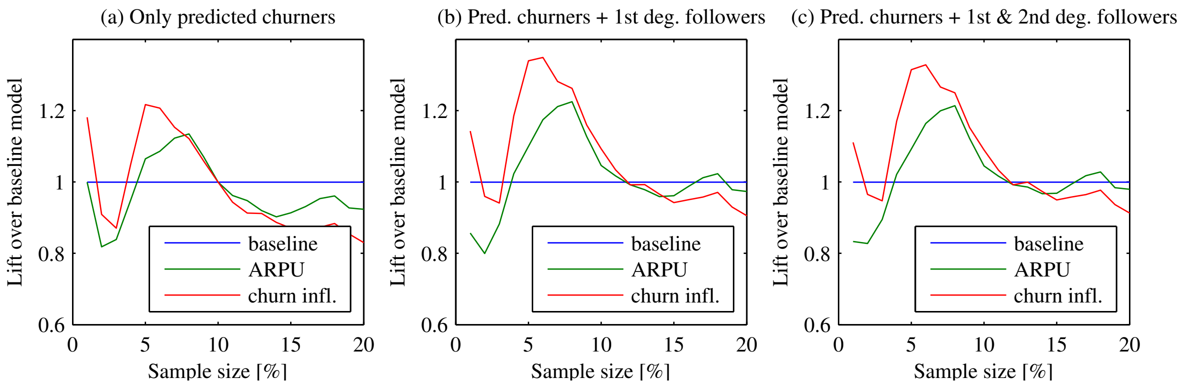

Fig. 6. Lift of the churn-influence scoring- and the ARPU based scoring-model over the baseline churn-prediction model at different sample sizes (top k-percent of users by churn-prediction scores) and fixed subsample size by scoring method - $n = 40\%$.

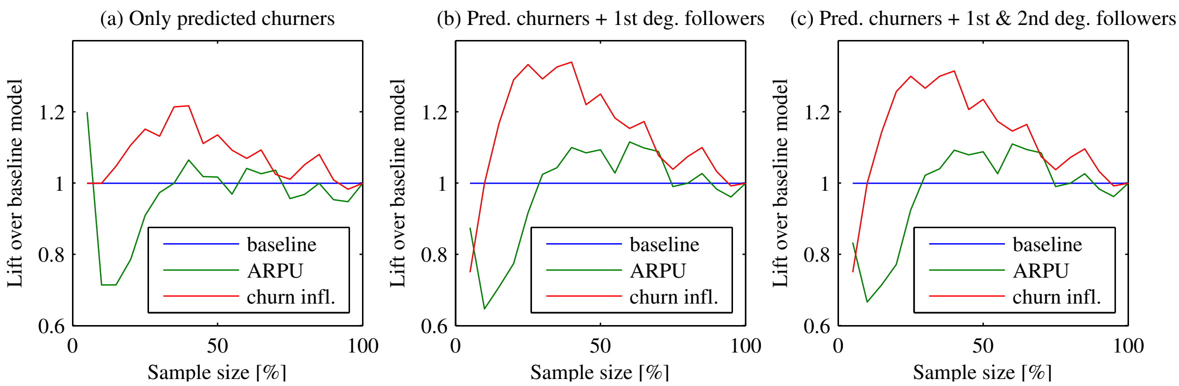

Fig. 7. Lift of the churn-influence scoring- and the ARPU based scoring-model over the baseline churn-prediction model at fixed sample sizes (top 5% of users by churn-prediction scores) and different subsample sizes by scoring method.

degree followers, where our approach is superior at the values of k between 3% and 10%. Consider, that the values of k usually used in practice are between 0.1% and 10% [7]. Therefore, at 70% of this interval, our method outperforms the other two. It is interesting to see that the middle and the right plot in Fig. 6 are almost identical. This observation can indicate that the users generally only influence their directly connected users, while the influence rarely spreads beyond the first degree of neighbours. Fig. 7 shows that when setting k to 5%, our approach outperforms the baseline model and the ARPU based ranking model on almost the entire interval of value n.

As we can see from Fig 6 and 7, the results are highly dependent on the settings of parameters k and n. However, finding the most optimal setting of k and n is not our goal since the most optimal setting is dependent on the dataset used and the specifics of the retention task in mind. However, we showed that our approach outperformed the classical churn prediction model and the ARPU model (two approaches, commonly used in practice) at a wide range of the applicable values of parameters k and n. Note that when predicting influential churners, we target at the very small portion of future churners, who will also initiate churns of some of the users connected with them.

4.3 Revenue-savings evaluation of the churn-influence-prediction model

To further evaluate the proposed approach, we compared the churn-influence-prediction model with the ARPU-based model and the classical churn prediction model, considering the revenue savings in the case of successfully retaining all the identified churners, including their 1st and 2nd degree followers.

Results of the evaluation are presented in Table 2, which shows lift of the ARPU-based model and the churn-influence-prediction model over the classical churn prediction model. For this evaluation, the parameter k was set to a fixed value, $k = 5\%$. This is the value often used in churn retention campaigns by service providers. We varied the values of parameter n from 10% to 40% by 5%-steps.

Table 2. Comparison of ARPU-based and churn-influence prediction (CIP) model with lift regarding financial savings, compared to classical churn prediction model

n [%]	lift_{ARPU}	lift_{CIP}
10	1.08	**1.17**
15	1.03	**1.12**
20	1.11	**1.19**
25	**1.25**	1.22
30	**1.29**	1.21
35	**1.33**	1.25
40	**1.38**	1.27

Table 2 reveals that both, the ARPU-based approach and the churn-influence-prediction model outperform (lift > 1) the classical churn-prediction model in the whole evaluated range of parameter n. Up to $n = 20\%$, the greatest savings are achieved by our proposed approach. With greater values of n, the ARPU-based method achieves better revenue savings. Since higher values of n in case of ARPU-based method also mean greater number of identified churners that spend more money, the amount of revenue savings shifts in favour of the ARPU-based method, even though more churners and their followers are identified with the churn-influence-prediction model (see Fig. 7). There results suggest, that the values of influential users used in campaigns of retaining influential churners should be rather small, e.g., up to 25%.

5 CONCLUSION

Identifying influential users is an important task in every market in which user decisions that impact company income can be influenced by other users in their SN. A typical example of such a market is the telco market, in which user decisions regarding churn have been shown to be significantly influenced by other users in their SN [14]. Several studies in the literature describe the procedure for determining user influence, but they are based on known influence data between pairs of users. In the case of churn, service providers are interested in identifying influential churners before they churn and consequently initiate a chain reaction.

In this paper, a procedure for predicting churn influence prior to a user churning, was presented. To identify good churn-influence predictors, a ground truth on churn influence is needed. Since ground truth on churn influence is usually not available, a procedure for determining WGT on churn influence was designed and proposed. The proposed WGT model was used to determine churn-influence scores using the dataset provided from a European mobile service provider. Then, 74 different user features were evaluated in connection with the churn-influence labels, obtained from WGT scores, using CFS feature selection method. As a result, four variables were determined as good predictors of churn-influence: (i) the number of on-net calls, (ii) the duration of outgoing on-net calls, (iii) the ratio between outgoing and incoming on-net calls, and (iv) the ratio between outgoing off-net calls and all outgoing calls. These features were used as the predictor variables in building a churn-influence-prediction model using logistic regression. The lift curve of the model clearly presented the improvement over the random approach despite using only four variables for the model.

The practical utility of the model lies in the determination of the most influential group of customers among those predicted as churners by one of the known churn-prediction methods. Strong retention campaigns directed toward an influential subset of predicted churners could not only prevent the churn of customers (successfully predicted as future churners) but also prevent the churn of users influenced by them. We show through our results, that our model, when used together with a classical churn-prediction model performs better in determining the true churners than the churn-prediction model alone. Also, we show that using this combination, considerably more true future churners are detected, which, in turn, have lots of followers who are potential churners as well. An additional evaluation of the models by expected revenue savings shows that in a limited range of sample size, our proposed approach also outperforms the compared methods.

As noted earlier, ground truth on churn influence is usually not available. However, to more accurately investigate customer churn for influential reasons, a strong firm truth would have to be obtained, e.g., by surveying churners about their reasons for churn and, if the reason is social influence, obtaining the necessary data to build a real SN of churners with their corresponding influence. We leave this issue open for future work.

Although the methodology presented in this work was focused only on churn, it could easily be used in solving other related issues such as determining user influence regarding new product adaptation.

ACKNOWLEDGEMENT

Operation part financed by the European Union, European Social Fund.

REFERENCES

[1] J. W. Plunkett, *Plunkett's Telecommunications Industry Almanac 2014: Telecommunications Industry Market Research, Statistics, Trends & Leading Companies*. Houston, Texas: Plunkett Research, Ltd., 2013.

[2] T. Verbraken, W. Verbeke, and B. Baesens, "Profit optimizing customer churn prediction with Bayesian network classifiers," *Intelligent Data Analysis*, vol. 18, no. 1, pp. 3–24, 2014.

[3] L. J. Rosenberg and J. A. Czepiel, "A Marketing Approach for Customer Retention," *Journal of Consumer Marketing*, vol. 1, no. 2, pp. 45–51, 1984.

[4] Y.-L. Lai and J.-R. Jiang, "Pricing Resources in LTE Networks through Multiobjective Optimization," *The Scientific World Journal*, vol. 2014, pp. 1–9, Jan. 2014.

[5] G. Kraljevic and S. Gotovac, "Modeling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services," *AUTOMATIKA*, vol. 51, no. 3, pp. 275–283, 2010.

[6] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European Journal of Operational Research*, vol. 218, pp. 211–229, Apr. 2012.

[7] D. Baras, A. Ronen, and E. Yom-Tov, "The Effect of Social Affinity and Predictive Horizon on Churn Prediction Using Diffusion Modeling," tech. rep., IBM, 2012.

[8] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, (New York, New York, USA), p. 61, ACM Press, 2002.

[9] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, (New York, New York, USA), pp. 57–66, ACM Press, 2001.

[10] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146, ACM, 2003.

[11] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 420–429, ACM, 2007.

[12] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 199–208, ACM, 2009.

[13] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1039–1048, ACM, 2010.

[14] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in *Proceedings of the 11th international conference on Extending database technology Advances in database technology - EDBT '08*, (New York, New York, USA), p. 668, ACM Press, 2008.

[15] U. Droftina, M. Štular, and A. Košir, "A diffusion model for churn prediction based on sociometric theory," *Advances in Data Analysis and Classification*, 2014.

[16] V. Smailovic, D. Striga, and V. Podobnik, "Advanced User Profiles for the SmartSocial Platform: Reasoning upon Multi-Source User Data," in *Web Proceedings of the 6th ICT Innovations Conference 2014* (A. Madevska-Bogdanova and D. Gjorgjevikj, eds.), (Ohrid, Macedonia), pp. 258–268, 2014.

[17] V. Smailovic, D. Striga, D.-P. Mamic, and V. Podobnik, "Calculating User's Social Influence through the SmartSocial Platform," in *Proceedings of the 22th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2014)*, (Split, Croatia), pp. 1–5, IEEE, 2014.

[18] M. Tkalčič, A. Odić, and A. Košir, "The impact of weak ground truth and facial expressiveness on affect detection accuracy from time-continuous videos of facial expressions," *Information Sciences*, vol. 249, pp. 13–23, Nov. 2013.

[19] J. Bjelland, G. Canright, K. Engo-Monsen, P. R. Sundsoy, and R. S. Ling, "A Social Network Study of the Apple vs. Android Smartphone Battle," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 983–987, IEEE, Aug. 2012.

[20] P. l. R. Sundsø y, J. Bjelland, G. Canright, K. Engø-Monsen, and R. Ling, "Product Adoption Networks and Their Growth in a Large Mobile Phone Network," in *2010 International Conference on Advances in Social Networks Analysis and Mining*, pp. 208–216, IEEE, Aug. 2010.

[21] J. Bello and M. Rolfe, "Is influence mightier than selection? Forging agreement in political discussion networks during a campaign," *Social Networks*, pp. 1–13, July 2013.

[22] P. D. Kusuma, D. Radosavljevik, F. W. Takes, and P. van den Putten, "Combining Customer Attribute and Social Network Mining for Prepaid Mobile Churn Prediction," in *BENELEARN 2013: Proceedings of the 22nd Belgian-Dutch Conference on Machine Learning*, (Nijmegen, The Netherlands), pp. 50–58, 2013.

[23] A. Goyal, F. Bonchi, and L. Lakshmanan, "Learning influence probabilities in social networks," in *Proceedings of the third ACM international conference on Web search and data mining*, pp. 241–250, ACM, 2010.

[24] S. Tufféry, *Data Mining and Statistics for Decision Making*. Wiley, 2nd ed., 2011.

[25] M. Hall and L. Smith, "Feature subset selection: a correlation based filter approach," in *International Conference on Neural Information Processing and Intelligent Information Systems*, (Berlin), pp. 855–858, Springer, 1997.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software," *ACM SIGKDD Explorations Newsletter*, vol. 11, p. 10, Nov. 2009.

[27] C. Kiss and M. Bichler, "Identification of influencers — Measuring influence in customer networks," *Decision Support Systems*, vol. 46, pp. 233–253, Dec. 2008.

Uroš Droftina received his PhD degree at the Faculty of Electrical Engineering, University of Ljubljana. Since 2010 he has been employed as a researcher at Telekom Slovenije, d.d. where he has been working in various research areas including customer churn prediction, analysis of social interactions between users, TV recommender systems, and smart grids research.

Mitja Štular graduated from the Faculty of Electrical Engineering and Computer Science of the University of Ljubljana, Slovenia in 1994. He successfully defended his master thesis in 1997 and Ph.D. thesis in 2000. In 2001 he began a career at Mobitel as a consultant to the CEO and also as the UMTS project director and been involved in a variety of research and development projects. In 2006 he took on the position of CTO responsible for networks, services and IT, and in 2009 a position of a Senior Strategy Adviser to the CEO, CSO.

Andrej Košir received the Ph.D. degree in electrical engineering from the University of Ljubljana in 1999. Since 2014 he is a full professor at the Faculty of Electrical Engineering, University of Ljubljana, and a head of User-adapted communications and ambient intelligence lab. He is active in a broad research fields including: user modeling and personalization, signal processing and statistical analysis of datasets and experiment design. Highlights from international cooperation include: guest researcher at University of Westminster, London, UK, and guest researcher at University of Waterloo, Canada and guest researcher at North Carolina State University, USA. International projects include: IST VICTORY, LIVE, IST Content4All, IST Tiramisu, P2People, Share It, IST Muffins, IST MyTV, ACTS STORit, ACTS SMASH. Currently he is co-editing an international monograph Emotions and Personality in Personalized Systems by Springer.

AUTHORS' ADDRESSES

Uroš Droftina, Ph.D.
Mitja Štular, Ph.D.
Telekom Slovenije, d.d.,
Cigaletova 15, 1000 Ljubljana, Slovenia
email: uros.droftina@telekom.si, mitja.stular@telekom.si
Prof. Andrej Košir, Ph.D.
Faculty of Electrical Engineering,
University of Ljubljana,
Tržaška 25, 1000 Ljubljana, Slovenia
email: andrej.kosir@fe.uni-lj.si

Received: 2014-05-26
Accepted: 2015-07-07