The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0263-5577.htm

Predicting customer churn in mobile industry using data mining technology

Eui-Bang Lee, Jinwha Kim and Sang-Gun Lee
Department of Business Administration, Sogang University, Seoul, Korea

Abstract

Purpose – The purpose of this paper is to identify the influence of the frequency of word exposure on online news based on the availability heuristic concept. So that this is different from most churn prediction studies that focus on subscriber data.

Design/methodology/approach – This study examined the churn prediction through words presented the previous studies and additionally identified words what churn generate using data mining technology in combination with logistic regression, decision tree graphing, neural network models, and a partial least square (PLS) model.

Findings – This study found prediction rates similar to those delivered by subscriber data-based analyses. In addition, because previous studies do not clearly suggest the effects of the factors, this study uses decision tree graphing and PLS modeling to identify which words deliver positive or negative influences.

Originality/value – These findings imply an expansion of churn prediction, advertising effect, and various psychological studies. It also proposes concrete ideas to advance the competitive advantage of companies, which not only helps corporate development, but also improves industry-wide efficiency.

Keywords Data mining, Logistic regression, Decision tree, Heuristic availability, Neural network, Predicting churn

Paper type Research paper

1. Introduction

The global ICT market has grown rapidly and reached market saturation, and as such, its players have been engaged in an ever-intensifying competition (Bae, 2012; Cho *et al.*, 2009; Kim *et al.*, 2010; Lee *et al.*, 2014). In particular, the mobile industry, which is the largest ICT hardware market, has exploded with the penetration rate jumping from 15.5 percent in 2001 to 93.1 percent in 2013 and expected to reach 95.5 percent in 2014 (ITU, 2014). Moreover, in 2011, the number of subscribers exceeded the country's population in 105 countries (Lee *et al.*, 2014). Such market saturation has shifted the dynamics of the competitive landscape from attracting new-to-market customers to acquiring competitors' customers (Lee *et al.*, 2011).

Accordingly, studies on customer churn have attracted many researchers' attention and have become increasingly important. Most of these studies focus on the causes of churning for the purpose of churn management (Ahn *et al.*, 2006; Jahanzeb and Jabeen, 2007; Kaya and Williams, 2005; Lee *et al.*, 2011; Lim *et al.*, 2006). In general, studies on customer churn highlight customer outflows, while those on customer acquisition pay attention to attracting customers. The latter closely examines methods to satisfy consumers in order to attract new customers (Hansotia and Wang, 1997; Thomas, 2001; Villanueva *et al.*, 2008; Wangenheim and Bayón, 2007).

Previous studies on customer churn focused on service quality (Ahn *et al.*, 2006; Coussement and Van Den Poel, 2008; Dierkes *et al.*, 2011; Ferreira *et al.*, 2004; Jahanzeb and Jabeen, 2007; Lim *et al.*, 2006; Mozer *et al.*, 2000), demographics of subscribers (Ahn

This work was supported by the Sogang University Research Grant of 201410023.01.

Industrial Management & Data Systems
Vol. 117 No. 1, 2017
pp. 90-109

0263-5577
DOI 10.1108/IMDS-12-2015-0509

et al., 2006; Burez and Van Den Poel, 2009; Ferreira *et al.*, 2004; Lee *et al.*, 2011; Lemmens and Croux, 2006), customer satisfaction/dissatisfaction (Ahn *et al.*, 2006; Lim *et al.*, 2006), and economic value (Lim *et al.*, 2006). These studies have been built on subscriber data or information gathered via surveys, which emphasized service receivers' individual aspects.

At the present time, companies are ceaselessly trying to attract customers through use of new advertisement and promotion strategies. Meanwhile, marketing studies have been examining how television advertising affects price sensitivity, what influence brand loyalty wields (Kanetkar *et al.*, 1992), and how advertising value changes advertising attitude (Ducoffe, 1996). However, these studies have limitations on practical applications because of the survey-oriented methodology they employed.

Thus, this study intends to identify the causes of churning by analyzing online media in order to predict customer churn rates on the basis of the availability heuristic (Teversy and Kahneman, 1974; Tversky and Kahneman, 1973), a psychological concept stating that individuals are more likely to choose what they can recollect in their minds. Therefore, this study suggests that online news information should be analyzed using data mining technology. To set up independent variables, the researcher of this study built a dictionary based on words identified by previous studies, and then crawled and parsed them. The actual churn data was extracted in order to set up a dependent variable. Next, an artificial neural network, a decision tree, and logistic regression models were applied to predict churn rates, as done in previous studies.

This study aims at: first, verifying the effects of advertising through data mining analysis; second, suggesting new directions for churn rate prediction based on internet information by expanding current prediction models that use individual commerce transactions; and third, proposing influence level of churn factors in addition to those suggested by previous studies.

The remainder of this paper is organized as follows. Section 2 explores the concepts of relevant studies on availability heuristic and customer churn, and defines web mining; Section 3 presents the research sequence and methods; Section 4 describes data collection and analysis results; Section 5 presents the conclusions; and Section 6 explains some of the limitations of this study.

2. Literature review

2.1 Availability heuristic and bias

Tversky and Kahneman (1973) defined heuristics as a tool to facilitate decision making that assists an individual or group in making a prompt and accurate decision. Among the three types of heuristics that Tversky and Kahneman (1973) detailed, the availability heuristic asserts that humans tend to value information which they are able to recollect easily, promptly, and vividly. This concept has a significant impact on decision making. For example, since availability is generally associated with an ecological frequency, it is affected by other factors, including the frequencies of classes of words, of combinatorial outcomes, and of repeated events, and is hence accompanied by systematic biases. As described by Tversky and Kahneman (1973), "the phenomenon of illusory correlation is explained as an availability bias" (p. 207).

Availability bias has been explored in a number of research papers. For example, people who have frequently encountered news reports in which a cat survived after falling from a tall tree would expect cats to survive falls from high places. However, in reality most cats cannot survive (Teversy and Kahneman, 1974). Many people believe that there are more deaths from shark attacks than from plane crashes, but this is also not true (Read, 1995). Similarly, people overestimate the number of kidnapping cases because such news reports are given more significant emphasis than other more common day-to-day incidents. That is, sensational items get more news coverage, thereby giving an impression that these incidents occur more frequently (Briñol *et al.*, 2006). Consequently, the public is affected by the vivid descriptions of televised news reporting. As a result, this type of reporting develops participants' social reality beliefs accordingly (Riddle, 2010). Sjöberg and Engelberg (2010) conducted experiments concerning the influences of news information and found that people became less affected by their risk awareness after they experienced their companies' risk situations and assessed the risk in a move.

The results of these experiments demonstrate that individuals who encounter televised news or other types of media reporting come to believe that the incidents or accidents in those reports occur more frequently than the actual numbers reveal. Such studies confirm that decision making is affected by the media. That is, when people are repeatedly exposed to a company or a product by the media, they begin to remember it, which in turn affects their decision making. It is expected that the mobile communications industry will also see the influence media has on consumers. When customers encounter information on other telecommunication company, the information stays in their minds, which may cause them to change to that company when they replace their mobile phones.

2.2 Customer churn

Customer churn refers to the loss of customers, as defined by a loss of rates during a certain period (Jahanzeb and Jabeen, 2007).This term is further divided into external and internal churns (Mattison, 2001). External churn is explained as customer churn due to non-voluntary variables (non-motivational/circumstantial factors, such as death or default) and voluntary variables (motivational factors, such as change of service provider or moving out of country) (Kaya and Williams, 2005). Internal churn refers to a change of service, such as moving from prepaid to postpaid. Churn also includes outbound moves (customers lost to competitors) and inbound moves (customers gained from competitors).

Various studies have identified that the causes of customer churn are related to the quality of service (Ahn *et al.*, 2006; Coussement and Van Den Poel, 2008; Dierkes *et al.*, 2011; Ferreira *et al.*, 2004; Jahanzeb and Jabeen, 2007; Lim *et al.*, 2006; Mozer *et al.*, 2000), demographic factors (Ahn *et al.*, 2006; Burez and Van Den Poel, 2009; Ferreira *et al.*, 2004; Lee *et al.*, 2011; Lemmens and Croux, 2006), customer satisfaction/dissatisfaction (Ahn *et al.*, 2006; Lim *et al.*, 2006), and economic value (Lim *et al.*, 2006). Table I summarizes the findings of previous studies.

These studies look into customer services (such as quality, pricing, pricing options, billing data, usage data, switching cost, mobile service quality, poor voice quality, billing issues, call minutes, and calls), the demographics of subscribers (customer demographics data, customer status, the number of churners, the number of predictive features in the data set, the number of observations, the number of mobile subscribers, etc.), customer satisfaction/dissatisfaction (emotional value) and economic value (Ahn *et al.*, 2006; Jahanzeb and Jabeen, 2007; Lee *et al.*, 2011; Lim *et al.*, 2006), and churn prediction (Burez and Van Den Poel, 2009; Coussement and Van Den Poel, 2008; Dierkes *et al.*, 2011; Ferreira *et al.*, 2004; Lemmens and Croux, 2006; Mozer *et al.*, 2000).

2.3 Data mining(DM)

Database mining is a technology used to extract useful information from large-scale databases in order to discover new or unknown information from the data (Agrawal *et al.*, 1993). The technology was initially developed using the association rule and clustering models. It was later named data mining, and its development was accelerated by seven key models: cluster detection, memory-based reasoning, market basket analysis, genetic

Table I. Churn research

Author	Source	Factors	Methodology	Results
Mozer *et al.* (2000)	Subscriber data	Call quality Pricing options Corporate capability Customer service	Neural network Decision tree Logistic regression	To predict churn rates, data re-combined in the naïve and sophisticated methods are analyzed by three models. Sophisticated neural network was found to deliver the best prediction results
Ferreira *et al.* (2004)	Subscriber data in Brazil	Billing data Usage data Customer demographics data Customer relationship data Market data	Neural network Decision tree Neuro Fuzzy (GA) Rule Evolver Each analysis reenacted data based on simple and enhanced representation models	To verify churn prediction, prediction rate of analysis models and cost savings were presented. Prediction rates differ by data reproduction
Lemmens and Croux (2006)		Behavioral predictors Company interaction predictor Customer demographic	Binary logit model Bagging Stochastic gradient boosting	With bagging and boosting, Gini coefficients increased by 16%, with a 26% of top-decline lift
Ahn *et al.* (2006)	Subscriber data in Korea	Customer status Customer dissatisfaction Switching cost	Multinomial logistic regression model Canonical correlation	Heavy users deliver higher churn rates, and customer dissatisfaction affects churn
Lim *et al.* (2006)		Mobile service quality Economic value Satisfaction Emotional value	*t*-Test Structural equation modeling	Loyalty model is comprised of economic and emotional values, service quality (e.g. network data) and customer satisfaction
Jahanzeb and Jabeen (2007)	Mobile service user data in Pakistan	Poor voice quality Limited cellular coverage Network problems High prices Billing issues	*t*-Test	Service provider's strategy is one of the prime causes of churn; low price and voice quality are critical factors of service strategy assessment

(*continued*)

Author	Source	Factors	Methodology	Results
Coussement and Van Den Poel (2008)	Transactional, marketing-related information and call center e-mail	Marketing information Call center e-mail	Logistic regression	When call center e-mail information is added to a marketing information-based model (e.g. client/company interaction variables), the AUC of the result went up to 77.75 from 73.8
Burez and Van Den Poel (2009)	Related studies	Number of observations Number of churners Percentage of churners Number of predictive features in the data set	Logistic regression Random forest	AUC of random forest changed by 1 to +7%, (compared to AUC of logistic regression), demonstrating better AUC and lifts
Dierkes *et al.* (2011)	Subscriber Data in China	Call minutes Calls Short Messaging Service (SMS) Multimedia Messaging Service (MMS) Game download	Logistic regression Logistic regression with propositionalization Markov logic networks (MLN)	When churn prediction and game downloading prediction were applied to MLN, accuracy increased by 8% and sensitivity by 19.7%, compared with logistic regression
Lee *et al.* (2011)	Subscribers number in Korea	Number of mobile subscribers of SKT, KT, and LGT	Bass diffusion model	When market is saturated, churn is caused by innovators' activities and imitators continue to follow the innovator

Table I.

algorithms, link analysis, decision trees, and neural networks (Berry and Linoff, 1997). Such developments were required to analyze large-scale data more efficiently in order to resolve data bottlenecks as technological developments produced tremendous amounts of data. Knowledge discovery in databases is a more comprehensive concept, and is comprised of crawling (to extract data), data cleaning (to enhance efficiency), data reduction and transformation (to find useful information), parsing (to identify specific information), analysis models (such as decision tree), data mining (for actual analysis), pattern evaluation and knowledge presentation, and deployment (to advance knowledge) (Cooley *et al.*, 1999). Data mining techniques, including logistic regression, decision trees, and neural network models, are applied in this study, similar to previous studies on churn prediction.

DM enables us to see a new aspect through intelligent text-mining system, which elicits pieces of knowledge that are associated with users' request. Besides, DM helps aggregate historical information from large-scale client/server-application to make decisions (Lee and Siau, 2001). DM could help knowledge management in two aspects; one is to organize context of business intelligent that has been extracted from data miner, and the other is to find customers' purchase patterns from sales database to enlarge human knowledge (Wang and Wang, 2008). DM is also utilized to reveal growing influence factors that affect purchase by airline customers (Liou and Tzeng, 2010). Furthermore, DM is employed in customer relation management, which puts customer information to use, to find and predict factors that attract customers and establish sufficient strategies (Lam *et al.*, 2014). The use of data mining has been expanded to the measurement of personal behavior pattern by utilizing smart environment sensor (Gebremeskel *et al.*, 2015).

Data mining is used as a tool to collect historical information from social media, and then predict performance by analyzing the type of word-of-mouth (Gaikar *et al.*, 2015). It is also applied to use big data architecture, which extracts information from the social media's large-scale user-generated data, to extract information that has been saved in the database through data mining technology, and use social network analysis method to analyze (Ch'ng, 2015). Instead of small size data analysis, big data analysis is a methodology for decision making by analyzing large-scale information. Knowledge discovery in database is used in order to analyze large-scale data simply, and the critical analysis part, which is the vital of the process, is called data mining. Data mining technology can be utilized in big data analysis, such as data classification, data clustering, prediction, and descriptive statistic (Cheng *et al.*, 2016). Since data mining technology is experiencing rapid developments, more various technologies are applied. This study used logistic regression, decision tree, and neural network, which is explained as below.

2.3.1 Logistic regression. Logistic regression is a regression that has binary values of dependent variables. Logistic regression does not follow a normal distribution as it has dependent variables of only 0 or 1. It accompanies the concept of ODDs $= p/1-p$, where p refers to the probability that an event takes place (refer to Formula (1)) This model is frequently used to compare churn prediction models (e.g. as in studies by Lemmens and Croux, 2006; Coussement and Van Den Poel, 2008; Burez and Van Den Poel, 2009; Dierkes *et al.*, 2011, and is used with decision trees and neural networks. However, its prediction level is lower than that of other models (Mozer *et al.*, 2000):

$$\text{ODDs} = \ln\left(\frac{\text{pi}}{1-\text{pi}}\right) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + e_i, e_i \sim Normal\ \left(0, \delta^2\right) \qquad (1)$$

2.3.2 Decision tree. A decision tree is a diagram of decision-making rules used to categorize subjects into a few groups or to make predictions. It is specified according to the grouping method as a CART Decision Tree (Breiman, 1993), C4.5 Decision Tree (Quinlan, 1993), or CHAID Decision Tree (Kass, 1980) in which a CART Decision Tree uses a Gini index, a C5.0

Decision Tree uses an entropy index, and a CHAID Decision Tree uses a χ^2 index. Decision trees are particularly useful when an analytical process needs to be explained. The tree structure makes understanding and interpretation easier. In addition, when two or more variables are combined, their interactive influences are better presented. Furthermore, it does not require any assumption of linear distribution, normal distribution, or homoscedasticity (Kim *et al.*, 2010).

2.3.3 Neural networks. A neural network travels from an input layer through a hidden layer to an output layer. During the travel, the correct weight equal to the input and output is identified by multiplying different weights for the connection. It is a repetitive learning model that operates through a back-propagating algorithm. A neural network delivers low classification error rates even when the number of data is not big enough or when significant data noise is present (Nagappan and Ball, 2007). That is why this model is often used as a prediction tool, such as for default and churn predictions (Bae, 2012; Bae and Kim, 2011; Cho *et al.*, 2009; Lim *et al.*, 2006; Nagappan and Ball, 2007).

3. Research model and analysis procedure

3.1 Research model

For the purpose of understanding the effect of churn rate depending on exposure of words on online news from a heuristic perspective, this study gathered up words that have been found affecting churn in various research, and then categorized these word to examine their effect.

In order to precisely identify the affect, we selected nine words in the aspects of service price, quality, and variety, three words in the aspect of technology, two words from the aspect of competitor, and one word in the aspect of social relation (see Figure 1). Before verifying the model, we investigated whether churn rate occurs due to the selected words using predictive values. The verification will measure whether the difference between predictive values in this study and in previous studies is huge. Next, we will examine the effect of the words through structural equation partial least square (PLS) analysis.

3.2 Analysis procedure

Availability bias refers to the influence that words frequently used in media reporting have on decision making. Thus, because the internet has become a channel that is able to confirm

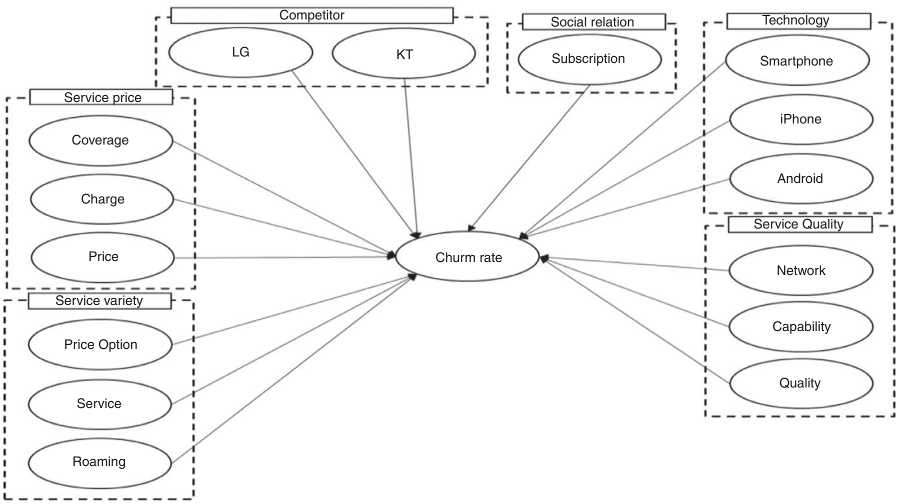


Figure 1.
Research model

the information provided by a variety of media such as television and newspaper, this study aims to explain churn through the words that are used in online new media.

The analysis procedure of this study is presented in Figure 2. Step 1 is sampling and data collection, where data sources are determined to identify online news and churn demands. Step 2 is dictionary creation for parsing by crawling the new websites selected in the previous step. Step 3 is parsing and calculation of churn ratio. Step 4 is data partition into training data and validation data. Step 5 is analysis. Step 6 is model assessment based on extracted information. Step 7 is additional analysis based on the PLS model.

4. Analytical results and discussion

4.1 Data collection

This study is limited to the South Korean market because the market is already saturated, despite its late entry into ICT services (Lee *et al.*, 2014). In fact, the country was ranked at the top of worldwide ICT development in 2011. In Korea, three major players provide mobile services: first mover SKT, fast follower KT (or KTF), and LGT (LGU+) (Lee *et al.*, 2011). SKT was chosen as the subject of this study because it has secured the top position since the very beginning of mobile services. In Step 1, "Naver" was selected as a portal service player for Korea because it provides a news clipping service that encompasses 322 media, including newspapers, broadcasting/telecommunications, internet papers, television, specialized magazines, and local papers. To identify churn demands, the number of subscribers who

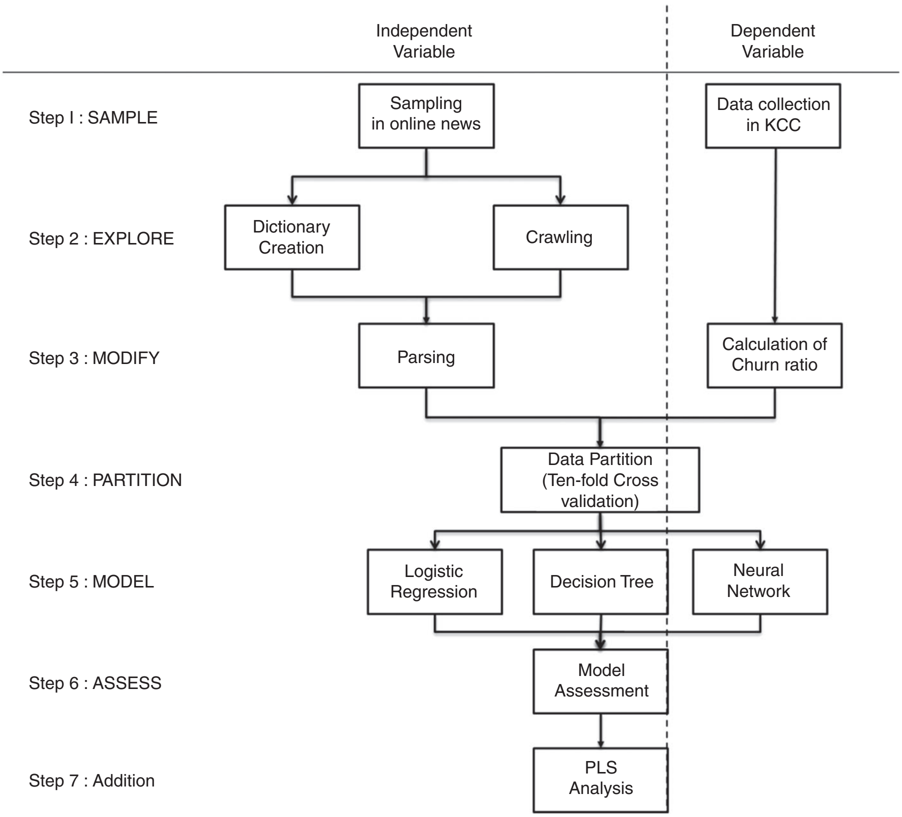


Figure 2.
Analysis procedure

changed their service providers from 2010 to 2012 was extracted from the mobile subscriber statistics published by the Korea Communications Commission (KCC).

In Step 2, previous studies on churn prediction were analyzed, and their independent variables were combined to create a dictionary (Table II) in order to define independent variables for this study. It is a Korean dictionary because the online news articles (analyzed in Step 1) were written in Korean. In Step 4, ten-fold cross-validation is used (Bae and Kim, 2010). When the processes of Steps 4 through 6 were completed, the data were randomized and the same process (Steps 4 through 6) were repeated ten times to calculate the average. Then the average was compared to the initial value. This ten-fold cross-validation method is frequently used to eliminate prediction rate bias, which may be incurred through training data and validation data bias (Bae and Kim, 2010; Kim *et al.*, 2010). For the data partition, 80 percent of the data were used as training data and 20 percent for validation.

The results of crawling SKT-related news articles (26,408 in total) from January 2010 to December 2012 on news.naver.com are presented in Table III. Table IV is the extraction result by the dictionary (Table II). Since KT shares a spelling (KT) with SKT, the search key word includes spaces, like (Space)KT(Space). Parsing is conducted as follows:

= SUMPRODUCT((LEN(,)-LEN(SUBSTITUTE(,”“))))/LEN()

* = space to be analyzed within Excel, = name of the words to be analyzed.

To identify the dependent variables, KCC data were utilized and the results are presented in Table V. To calculate prediction rates, this study designated SI for SKT inbound (positive churn) and SO for SKT outbound (negative churn). The sample cycle for this study was one month. This is because daily samples may not immediately accommodate daily churn rates. Thus, the sample size is $n = 36$ for three years.

4.2 Analysis method

SAS e-Miner 9.3, which is frequently used for logistic regression, decision trees, and neural network analysis, is utilized to run the data. In SAS e-Miner 9.3, the prediction rate is

Researcher	PO	CG	PC	QL	AR	SP	IP	NW	SV	RM	CB	CG	SS	KT	LG
Mozer *et al.* (2000)	PO								CS		CC	CV	CT		
Coussement and Van Den Poel (2008)		MM											SR		
Jahanzeb and Jabeen (2007)			HP	PQ	LC	LC	LC	NP							
Ahn *et al.* (2006)					HM	HM	HM								
Ferreira *et al.* (2004)									TS	CR				CP	CP
Lemmens and Croux (2006)									MS						

Notes: PO, pricing option; MM, mean total monthly recurring charge; HP, high price; PQ, poor voice quality; lC, limited cellular; HM, handset manufacturer; NP, network problems; TS, type of service used; CS, customer service; MS, months in service; CR, cost of roaming; CC, corporate capability; CV, coverage; CT, customer communications; SR, subscription renewed; CP, competitor rate; CG, charge; PC, price; QL, quality; AR, Android; SP, smartphone; IP, iPhone; NW, network; SV, service; RM, roaming; CB, capability; CG, coverage; SS, subscription

Table II.
Dictionary

Year/month	10/01	10/02	10/03	10/04	10/05	10/06	10/07	10/08	10/09	10/10	10/11	10/12
No. of reports	465	601	568	707	834	897	1,050	725	674	809	604	522
Year/month	11/01	11/02	11/03	11/04	11/05	11/06	11/07	11/08	11/09	11/10	11/11	11/12
No. of reports	600	628	1,039	816	1,116	900	925	961	733	642	1,213	567
Year/month	12/01	12/02	12/03	12/04	12/05	12/06	12/07	12/08	12/09	12/10	12/11	12/12
No. of reports	444	707	655	613	801	736	605	604	737	551	656	703

Table III.
Number of reports

Year/month	PO	CG	PC	QL	AR	SP	IP	NW	SV	RM	CB	CG	SS	KT	LG
10/01	11	4	1	0	55	63	13	0	30	3	1	12	27	4	3
10/02	38	31	0	17	57	46	21	0	50	9	0	1	22	9	43
10/03	23	6	0	2	22	70	0	2	40	12	0	0	8	3	25
10/04	39	19	0	0	23	67	38	7	63	2	0	0	21	8	14
10/05	30	3	2	25	20	82	19	1	70	22	0	0	51	14	30
10/06	19	0	1	0	10	50	26	7	45	11	0	0	4	12	9
10/07	59	0	1	0	24	92	25	6	72	25	0	3	31	22	8
10/08	33	0	1	3	65	82	9	7	70	7	0	1	33	15	6
10/09	48	0	1	6	3	38	1	3	65	14	0	0	30	19	11
10/10	34	0	1	4	9	73	16	8	89	12	0	2	27	8	8
10/11	20	1	1	0	3	91	7	0	39	19	0	4	21	9	19
10/12	5	0	2	0	3	58	2	1	70	0	1	0	3	5	35
11/01	15	0	0	0	1	66	11	2	39	15	0	1	14	7	17
11/02	46	0	2	3	4	62	138	1	30	27	0	2	9	23	15
11/03	58	0	5	3	0	45	289	2	94	48	0	0	66	25	14
11/04	19	0	2	22	0	33	26	10	36	5	0	0	42	19	55
11/05	58	0	1	0	0	44	22	3	28	15	0	0	22	4	8
11/06	52	0	0	4	1	43	16	2	68	22	0	0	12	6	120
11/07	19	2	2	3	0	10	5	3	46	28	2	1	15	6	48
11/08	56	0	4	2	16	35	14	2	38	33	3	0	13	24	36
11/09	131	0	1	1	1	35	3	0	28	22	0	0	20	2	25
11/10	37	0	40	0	0	21	11	0	41	19	0	0	46	2	59
11/11	25	2	28	4	1	12	177	1	21	15	0	0	35	22	34
11/12	27	0	11	10	0	15	7	2	62	44	1	0	44	18	45
12/01	54	8	8	2	3	24	7	1	23	26	0	0	36	10	11
12/02	22	0	10	9	0	37	1	6	47	20	0	1	17	1	29
12/03	37	0	9	15	2	16	0	11	52	28	0	0	33	17	22
12/04	24	1	6	16	0	6	0	0	71	31	0	0	22	17	20
12/05	57	0	4	1	0	11	0	1	65	1	0	1	21	8	17
12/06	23	2	2	29	0	23	0	8	80	31	0	1	36	7	26
12/07	22	1	2	4	0	20	0	8	48	22	1	0	63	6	29
12/08	19	0	0	17	0	14	5	1	42	26	0	0	16	6	44
12/09	25	0	0	12	0	9	89	0	37	42	1	1	39	16	22
12/10	1	0	2	29	0	15	72	2	34	0	1	0	25	6	7
12/11	9	0	10	11	1	10	46	4	18	14	0	0	21	7	20
12/12	22	4	3	0	1	16	168	0	34	6	0	0	41	25	34

Notes: PO, pricing option; CG, charge; PC, price; QL, quality; AR, Android; SP, smartphone; IP, iPhone; NW, network; SV, service; RM, roaming; CB, capability; CG, coverage; SS, subscription

Table IV. Measuring of independent variables

represented by 1 – misclassification rate. The misclassification rate refers to the error in the classification of predicted values, calculated by the following equation (SAS, 2013):

$$\text{Missclassification rate} = \frac{\sum i(1\{F\_Y_i \neq I\_Y_i\}f_i)}{\sum if_i} \tag{2}$$

$\sum if_i$ the total frequency, I_Yi the classified, Yi the input and F_Yi at normalized levels the response (output), showing the ratio of output error to input. The misclassification rate is equivalent to the "1 – accuracy" as specified in previous studies (Burez and Van Den Poel, 2009; Dierkes *et al.*, 2011) and is represented by the following equation:

$$\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{FN}+\text{TN}+\text{TP}) \tag{3}$$

TP is the number of true positives; TN the number of true negatives; FN the number of false negatives; and FP the number of false positives

Table V. Churn rate table

				S K T			
Year/month	KT→S	LGU+→S	SKT→K	SKT→L	Total Subscribers	CR (%)	Churn
10/01	117,910	75,755	115,768	75,570	24,435,145	0.010	SI
10/02	156,186	94,511	154,627	92,275	24,628,773	0.015	SI
10/03	171,409	105,742	171,766	106,322	24,824,527	−0.004	SO
10/04	116,385	73,963	93,937	80,107	24,932,320	0.065	SI
10/05	232,225	130,866	231,043	127,334	25,071,500	0.019	SI
10/06	214,427	122,752	204,133	120,079	25,146,337	0.052	SI
10/07	253,598	139,607	215,069	137,837	25,242,567	0.160	SI
10/08	217,275	124,926	215,007	125,133	25,332,382	0.008	SI
10/09	264,473	127,295	289,258	105,259	25,445,309	−0.011	SO
10/10	198,408	110,145	210,726	109,187	25,498,479	−0.045	SO
10/11	247,307	94,793	246,997	94,345	25,613,970	0.003	SI
10/12	177,262	101,727	172,717	98,030	25,705,049	0.032	SI
11/01	224,194	144,802	261,620	118,773	25,811,727	−0.044	SO
11/02	189,129	116,665	201,419	106,612	25,882,959	−0.009	SO
11/03	175,898	106,723	172,334	107,649	25,988,510	0.010	SI
11/04	180,357	101,096	177,023	100,145	26,069,453	0.016	SI
11/05	253,696	124,733	252,292	127,941	26,203,566	−0.007	SO
11/06	226,537	108,129	238,305	119,811	26,268,972	−0.089	SO
11/07	216,484	116,951	202,645	118,409	26,319,110	0.047	SI
11/08	220,438	128,961	218,212	129,601	26,399,912	0.006	SI
11/09	180,074	105,005	181,006	131,287	26,420,872	−0.103	SO
11/10	209,473	118,716	238,974	112,411	26,444,500	−0.088	SO
11/11	239,147	132,782	249,434	133,456	26,489,013	−0.041	SO
11/12	206,145	115,861	201,798	121,926	26,552,716	−0.006	SO
12/01	215,555	128,522	201,353	133,924	26,571,498	0.033	SI
12/02	192,105	115,522	177,969	128,731	26,546,876	0.003	SI
12/03	213,089	132,369	177,562	150,099	26,556,148	0.067	SI
12/04	239,808	140,585	186,654	178,175	26,586,316	0.059	SI
12/05	240,430	158,748	218,538	187,088	26,610,958	−0.024	SO
12/06	184,566	143,454	168,173	157,140	26,658,998	0.010	SI
12/07	261,102	193,816	229,145	211,486	26,678,718	0.054	SI
12/08	272,375	188,429	300,982	208,011	26,715,610	−0.180	SO
12/09	298,234	181,047	291,399	192,565	26,777,852	−0.017	SO
12/10	133,862	88,367	116,993	105,931	26,825,007	−0.003	SO
12/11	185,427	116,320	171,862	146,218	26,874,009	−0.061	SO
12/12	268,664	148,445	262,521	181,411	26,961,045	−0.099	SO

Note: CR, churn rate

Accordingly, the churn prediction rate of this study will be based on the value of 1 – misclassification rate. As a choice resulting from a decision tree, the most frequently used C5.0 Decision Tree is adopted (Bae and Kim, 2010).

4.3 Analysis results

t-Tests were conducted to examine the average differences of independent variables by dependent variables (SO and SI), and the results are presented in Table VI. Within the significance level of 0.1, "network," "service," and "Android" deliver significant differences. This confirms that certain words do deliver significant statistical differences in the average service churn rates.

The results of the decision tree analysis are presented in Figure 3. It found that churn rates are affected by the terms "price option," "smartphone," "network," and "charge," and

Factor	t	df	Sig. (two-tailed)	Mean difference	SE difference	95% confidence interval of the difference Lower	Upper
Price option	−0.776	34	0.443	−6.053	7.795	−21.894	9.788
Charge	1.640	19.860	0.117	3.084	1.880	−0.840	7.007
Price	−1.129	34	0.267	−3.012	2.669	−8.437	2.412
Quality	0.606	34	0.549	1.777	2.934	−4.185	7.739
Android	2.400	20.622	0.026	12.201	5.084	1.616	22.787
Smartphone	1.613	34	0.116	13.728	8.511	−3.568	31.024
iPhone	−0.933	34	0.357	−19.579	20.976	−62.207	23.049
Network	2.374	27.856	0.025	2.328	0.981	0.319	4.337
Service	1.783	33.256	0.084	11.359	6.371	−1.599	24.318
Roaming	0.353	34	0.726	1.474	4.174	−7.009	9.956
Capability	1.099	34	0.279	0.245	0.223	−0.208	0.697
Coverage	1.207	34	0.236	0.851	0.705	−0.582	2.285
Subscription	0.785	34	0.438	3.969	5.057	−6.308	14.246
KT	0.562	34	0.578	1.399	2.490	−3.661	6.460
LG	−0.831	34	0.412	−6.006	7.229	−20.697	8.685

Table VI.
t-Test result

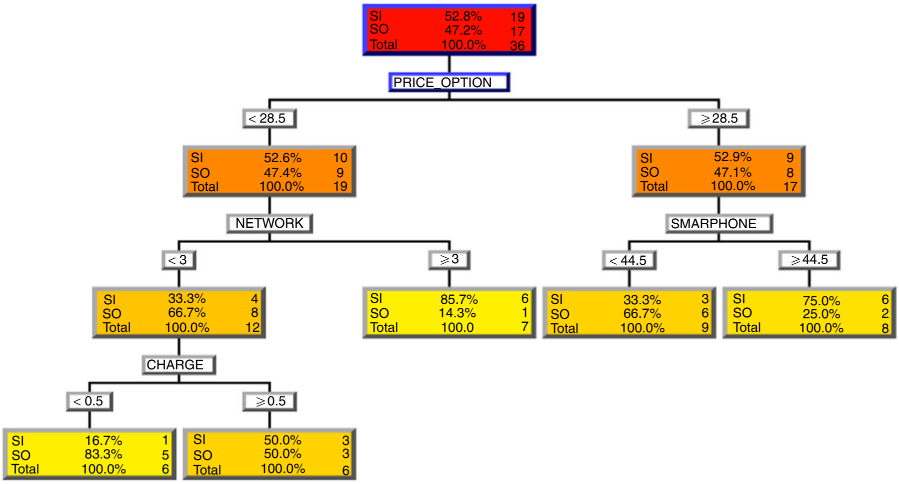


Figure 3.
Decision tree

their frequencies are presented in Appendix 1. The decision tree analysis found the following rules:

Rule 1: when charge < 0.5, network < 3, and price option < 28.5, SO was 5 and SI 1, meaning more outflows from SKT to its competitors than inflows. This may be attributed to the fact that SKT's rate plans are more expensive than its competitors'.

Rule 2: when charge $\geqslant 0.5$, network < 3, and price option < 28.5, SO was 3 and SI 3, meaning that inflows and outflows were almost the same. This may demonstrate that when the high-price policy is less known, outflows tend to be mitigated.

Rule 3: when network $\geqslant 3$ and price option < 28.5, SI was 6 and SO 1, showing more inflows into SKT than outflows from it. This demonstrates that promotion and advertising of its network quality was an effective strategy. This may be interpreted as a result of SKT's high-quality strategy.

Rule 4: when Smartphone < 44 and price option $\geqslant 28.5$, SO was 6 and SI 3, showing more outflows than inflows. This means 1) when new technology, such as smartphones, is yet to be promoted, outflows tend to increase, and 2) rate plan changes involving new technology also facilitate churn.

Rule 5: when Smartphone $\geqslant 44$ and price option $\geqslant 28.5$, SI was 6 and SO 2, showing more inflows than outflows. This means that less promotion of new technology, such as smartphones, increases inflows into SKT.

Considering all the rules comprehensively, the greatest SI is delivered by Rule 3, meaning that increased exposure of SKT's network information via online news boosts inflows into SKT. When network-related information is not well exposed, inflows and outflows become even only if charge-related information is not well exposed. When more information on charge is exposed, outflows increase with the greatest SO shown in Rule 1. In addition, the use of more words related to smartphones increases inflows, and vice versa (Rule 4 $\geqslant$ 3).

After identifying the decision tree above, training data and validation data were measured to predict churn rates. The results are presented in Table VII, with levels of significance of 45.71 percent for logistic regression, 68.57 percent for decision tree, and 72.86 percent for neural network.

4.4 Results of additional analysis

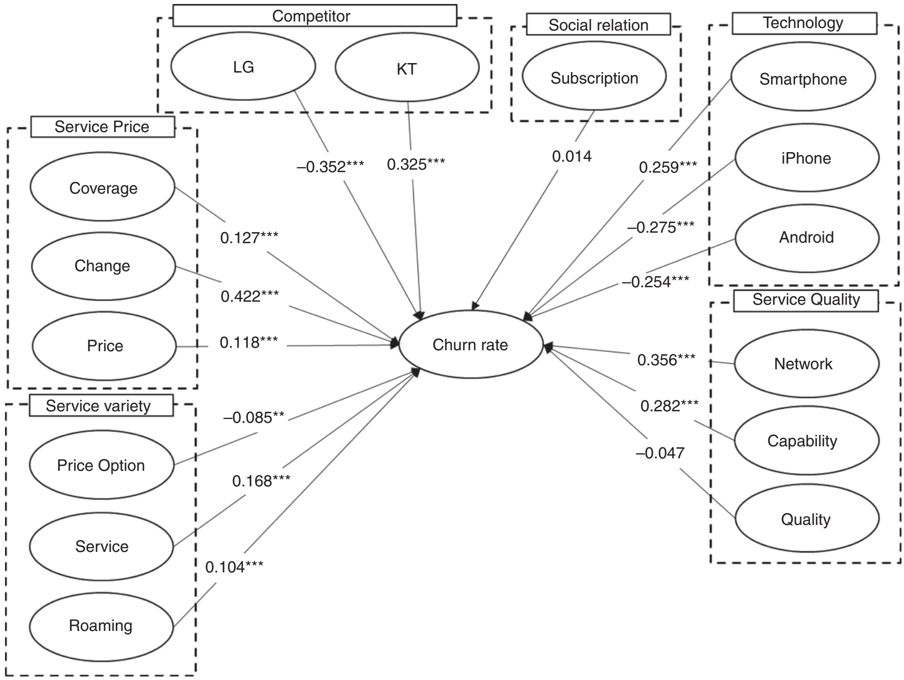
After the analysis results were verified as statistically significant, a PLS model was applied to further clarify the effect of the words. PLS is widely adopted as a structural equation model. Unlike AMOS or LISLE, it does not presume a chi-square distribution, but instead represents the same level of association relations. For the purpose of the PLS analysis, the churn rate shown in Table V is not changed into a binary predictor but is used as a dependent variable. Smart PLS is used as an analysis tool. The adequacy of the PLS model was verified by goodness-of-fit, which is calculated by multiplying the average of R^2 by the average of redundancy and extracting the square root thereof. Goodness-of-fit must be no < 0.1 (Tenenhaus *et al.*, 2005). The model was verified to be fit as presented in Appendix 2.

The square root of the average and the correlation coefficient were extracted to confirm convergent validity and discriminant validity as shown in Appendix 3. When the correlation coefficient is lower than the square root average, the corresponding variable is confirmed as valid (Hair *et al.*, 1998). The analysis confirmed the validity.

PLS analysis results are as presented in Figure 4 and Table VIII. To summarize, when words such as charge (0.422***), network (0.356***), KT (0.325***), capability (0.282***), smartphone (0.259***), service (0.168***), coverage (0.127***), price (0.118***), and roaming (0.104***) appear in online news articles, subscribers move to SKT from its competitors. "Charge," "network," and "smartphone" have already been confirmed as having influence on churn by decision tree analysis. This result reveals that the first mover and the fast follower are constantly compared, and when both SKT and KT simultaneously appear, the first mover tends to gain a greater inflow of subscribers, on the back of its comparative advantage. In addition, as SKT maintains its high-quality service strategy, "network," "service," "smartphone," "coverage," and "roaming" continue to have a positive impact. Its strong capability contributes to a stable corporate image, and "charge" and "prices" also have a positive impact. However, Subscription (0.014) and Quality (−0.047) turn out to have

Table VII.
Prediction rate

Prediction Rate	Training (goodness-of-fit)	Validation (verified prediction rates)
C5.0 decision tree	0.69656	0.6857
Neural network	0.779310345	0.728571429
Regression	0.648275862	0.457142857



Notes: *$p<0.05$; **$p<0.01$; ***$p<0.005$ (one-side test)

Figure 4.
PLS model

Table VIII.
PLS result

	Original sample (O)	Sample mean (M)	SD	SE	*t*-Statistics |O/SE|
KT→Churn rate	0.32555	0.325424	0.034914	0.034914	9.324443
LG→Churn rate	−0.350879	−0.352209	0.022536	0.022536	15.569946
Price→Churn rate	0.11663	0.118415	0.029245	0.029245	3.988062
Subscription→Churn rate	0.013903	0.013834	0.032991	0.032991	0.421411
Charge→Churn rate	0.422149	0.422166	0.039599	0.039599	10.660522
Network→Churn rate	0.35381	0.355769	0.031543	0.031543	11.216875
Roaming→Churn rate	0.104731	0.104534	0.037964	0.037964	2.758675
Coverage→Churn rate	0.130826	0.127238	0.030329	0.030329	4.313558
Service→Churn rate	0.169289	0.168353	0.042013	0.042013	4.029444
Smartphone→Churn rate	0.259163	0.259108	0.039909	0.039909	6.493881
iPhone→Churn rate	−0.274404	−0.274877	0.035738	0.035738	7.678167
Android→Churn rate	−0.256583	−0.253698	0.043725	0.043725	5.868099
Price option→Churn rate	−0.086203	−0.084939	0.034683	0.034683	2.485444
Capability→Churn rate	0.28036	0.282202	0.039184	0.039184	7.154918
Quality→Churn rate	−0.046986	−0.047664	0.030842	0.030842	1.523453

little influence. As such, the word "quality" itself does not affect churn rates and neither does "subscription."

Lastly, when LG (−0.352***), iPhone (−0.275***), Android (−0.254***), and price option (−0.085**) appear in online news articles, subscribers move from SKT to its competitors. This may be explained by the fact that when the first mover is compared with the follower, LG's quality and price competitiveness are frequently cited, giving comparative advantage to the follower. Given the fact that KT was the sole distributor of iPhone when it was first launched in the Korean market, "iPhone" delivers a negative impact on SKT's churn rates. The negative influence of "Android" is attributed to the mobile devices offered by service providers, while that of "price option" is attributed to SKT's high-price policy.

5. Conclusion

Businesses continuously work on various activities such as marketing and sales, in order to use and sell their own corporate products. Thus, businesses pay keen attention to the effect of such activities on sales. Especially, since the changes in sales made from mass media are measured by comparing sales before and after advertising, it is essential to understand to what extend the words in an advertisement are able to attract users, and on the contrary, to what extend the words in an advertisement could be avoided. However, previous studies neglected the influence of words in advertisements.

This study focused on advertising as a corporate activity used to identify factors influencing churn. It examined words exposed on online news articles, introduced psychological concepts, and applied data mining technology to reach its conclusions. To this end, this study harnessed neural network, decision trees, and logistic regression models, with a PLS model as an additional test. The analysis delivered meaningful prediction rates, with logistic regression scoring 45.71 percent, decision tree 68.57 percent, and neural network 72.86 percent. These results are similar to those reported in previous studies based on commercial transaction data. For example, many studies evaluate 68 percent prediction rate(Nath and Behara, 2003), 77 percent prediction rate(Ferreira *et al.*, 2004), 78.3 percent prediction rate(Nath and Behara, 2003), and 77.07 percent prediction rate(Dierkes *et al.*, 2011), which once again confirm the statistical significance of this analysis. In addition, a PLS model was applied to verify statistical significance of the words analyzed in this study.

The results of this study help measure churn increases or decreases associated with the mass media, a factor which has not been addressed in previous studies. Because it delivers prediction rates similar to those of previous studies based on subscriber data, the methodology of this study is confirmed as valid for churn prediction. In addition, this study found that the more the word "network" is exposed on online news media, the more customer inflows SKT has. However, when "network" is not sufficiently exposed, inflows and outflows become almost even only if the word "charge" is not frequently exposed. Moreover, outflows become greater when exposure of "charge" increases. Lastly, when the word "smartphone" is frequently exposed, inflows into SKT increase. The additional analysis confirms the positive impact of the terms "charge," "network," "KT," "capability," "smartphone," "service," "coverage," "price," and "roaming" as well as the negative impact of "LG," "iPhone," "Android," and "price option." This finding shows that specific comments about high quality cause an increase of inflows to SKT, while comments about high prices cause an increase of outflows from SKT. In other words, the more the company's strengths appear in the media, the more inflows it gets. Thus, SKT may be more successful if it could expose its high-quality, high-price policy through online media.

The contributions of this study are as follows: first, it expands the scope of churn prediction research by introducing data mining technology, as distinct from the subscriber data-based approach; second, it demonstrates how to utilize web information, in contrast to most of the previous studies on advertising effects which were built upon surveys; third, it suggests specific guidelines for companies in order to secure competitive advantages; and fourth, it proposes specific models to analyze churn of subscribers by media exposure on the basis of psychological concepts. Based on such results of this research, it became possible to examine the increase in exposures of subscribers to various media such as news by analysis of not only the effect of the whole advertisement, but also the effect of texts within advertisements, and thus to suggest strategies for holding effective economic activities.

6. Limitations

The data cycle for the data mining model of this study is on a monthly basis, and 36 samples that cover three years were analyzed for the purpose of this study. This study has the following limitations: first, it focuses on macro perspectives, therefore, failing to consider that churn factors may be affected by individual tendencies; second, its dictionary is built upon previous studies, therefore, failing to consider that other words may also have significant impact; and third, it only examines the Korean market, therefore, failing to analyze the differences between markets. Accordingly, future studies may be able to identify more in-depth the specific causes of churn by covering other national markets, using different independent variables, and employing a greater scope.

References

Agrawal, R., Imielinski, T. and Swami, A. (1993), "Database mining: a performance perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5 No. 6, pp. 914-925.

Ahn, J.-H., Han, S.-P. and Lee, Y.-S. (2006), "Customer churn analysis: churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, Vol. 30 No. 10, pp. 552-568.

Bae, J.K. (2012), "Predicting financial distress of the South Korean manufacturing industries", *Expert Systems with Applications*, Vol. 39 No. 10, pp. 9159-9165.

Bae, J.K. and Kim, J. (2010), "Integration of heterogeneous models to predict consumer behavior", *Expert Systems with Applications*, Vol. 37 No. 3, pp. 1821-1826.

Bae, J.K. and Kim, J. (2011), "Combining models from neural networks and inductive learning algorithms", *Expert Systems with Applications*, Vol. 38 No. 5, pp. 4839-4850.

Berry, M.J. and Linoff, G. (1997), *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, Inc, New York, NY.

Breiman, L. (1993), *Classification and Regression Trees*, CRC press, Belmont.

Briñol, P., Petty, R.E. and Tormala, Z.L. (2006), "The malleable meaning of subjective ease", *Psychological Science*, Vol. 17 No. 3, pp. 200-206.

Burez, J. and Van Den Poel, D. (2009), "Handling class imbalance in customer churn prediction", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 4626-4636.

Ch'ng, E. (2015), "Social information landscapes: automated mapping of large multimodal, longitudinal social networks", *Industrial Management & Data Systems*, Vol. 115 No. 9, pp. 1724-1751.

Cheng, S., Zhang, Q. and Qin, Q. (2016), "Big data analytics with swarm intelligence", *Industrial Management & Data Systems*, Vol. 116 No. 4, pp. 646-664.

Cho, S., Kim, J. and Bae, J.K. (2009), "An integrative model with subject weight based on neural network learning for bankruptcy prediction", *Expert Systems with Applications*, Vol. 36 No. 1, pp. 403-410.

Cooley, R., Mobasher, B. and Srivastava, J. (1999), "Data preparation for mining world wide web browsing patterns", *Knowledge and Information Systems*, Vol. 1 No. 1, pp. 5-32.

Coussement, K. and Van Den Poel, D. (2008), "Integrating the voice of customers through call center emails into a decision support system for churn prediction", *Information & Management*, Vol. 45 No. 3, pp. 164-174.

Dierkes, T., Bichler, M. and Krishnan, R. (2011), "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks", *Decision Support Systems*, Vol. 51 No. 3, pp. 361-371.

Ducoffe, R.H. (1996), "Advertising value and advertising on the web", *Journal of Advertising Research*, Vol. 36 No. 5, pp. 21-36.

Ferreira, J.B., Vellasco, M., Pacheco, M.A. and Barbosa, C.H. (2004), "Data mining techniques on the evaluation of wireless churn", *ESANN2004 Proceedings-European Symposium on Artificial Neural Networks Bruges*.

Gaikar, D.D., Marakarkandy, B. and Dasgupta, C. (2015), "Using twitter data to predict the performance of bollywood movies", *Industrial Management & Data Systems*, Vol. 115 No. 9, pp. 1604-1621.

Gebremeskel, G.B., Yi, C., Wang, C. and He, Z. (2015), "Critical analysis of smart environment sensor data behavior pattern based on sequential data mining techniques", *Industrial Management & Data Systems*, Vol. 115 No. 6, pp. 1151-1178.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W. (1998), *Multivariate Data Analysis*, Pearson Prentice Hall, Upper Saddle River, NJ.

Hansotia, B.J. and Wang, P. (1997), "Analytical challenges in customer acquisition", *Journal of Interactive Marketing*, Vol. 11 No. 2, pp. 7-19.

ITU (2014), "Individuals using the internet per 100 inhabitants", available at: www.itu.int/en/ITU-D/…/2014/stat_page_all_charts_2014.xls (accessed 6, October 2015).

Jahanzeb, S. and Jabeen, S. (2007), "Churn management in the telecom industry of Pakistan: a comparative study of Ufone and Telenor", *Journal of Database Marketing & Customer Strategy Management*, Vol. 14 No. 2, pp. 120-129.

Kanetkar, V., Weinberg, C.B. and Weiss, D.L. (1992), "Price sensitivity and television advertising exposures: some empirical findings", *Marketing Science*, Vol. 11 No. 4, pp. 359-371.

Kass, G.V. (1980), "An exploratory technique for investigating large quantities of categorical data", *Applied Statistics*, Vol. 29 No. 2, pp. 119-127.

Kaya, S. and Williams, B. (2005), "Effective churn management for business", *Journal of Corporate Real Estate*, Vol. 7 No. 2, pp. 154-163.

Kim, J., Won, C. and Bae, J.K. (2010), "A knowledge integration model for the prediction of corporate dividends", *Expert Systems with Applications*, Vol. 37 No. 2, pp. 1344-1350.

Lam, H.Y., Ho, G.T.S., Wu, C.H. and Choy, K.L. (2014), "Customer relationship mining system for effective strategies formulation", *Industrial Management & Data Systems*, Vol. 114 No. 5, pp. 711-733.

Lee, J.S. and Siau, K. (2001), "A review of data mining techniques", *Industrial Management & Data Systems*, Vol. 101 No. 1, pp. 41-46.

Lee, S.-G., Lee, E.-B. and Yang, C.-G. (2014), "Strategies for ICT product diffusion: the case of the Korean mobile communications market", *Service Business*, Vol. 8 No. 1, pp. 65-81.

Lee, S.-G., Yu, M., Yang, C. and Kim, C. (2011), "A model for analyzing churn effect in saturated markets", *Industrial Management & Data Systems*, Vol. 111 No. 7, pp. 1024-1038.

Lemmens, A. and Croux, C. (2006), "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, Vol. 43 No. 2, pp. 276-286.

Lim, H., Widdows, R. and Park, J. (2006), "M-loyalty: winning strategies for mobile carriers", *Journal of Consumer Marketing*, Vol. 23 No. 4, pp. 208-218.

Liou, J.J. and Tzeng, G.-H. (2010), "A dominance-based rough set approach to customer behavior in the airline market", *Information Sciences*, Vol. 180 No. 11, pp. 2230-2238.

Mattison, R. (2001), *Telecom Churn Management: The Golden Opportunity*, Apdg Publishing, Fuquay-Varina.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. (2000), "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry", *IEEE Transactions on Neural Networks*, Vol. 11 No. 3, pp. 690-696.

Nagappan, N. and Ball, T. (2007), "Using software dependencies and churn metrics to predict field failures: an empirical case study", First International Symposium on Empirical Software Engineering and Measurement, ESEM First International Symposium, September 20-21.

Nath, S.V. and Behara, R.S. (2003), "Customer churn analysis in the wireless industry: a data mining approach", *Proceedings-Annual Meeting of the Decision Sciences Institute*, pp. 505-510.

Quinlan, J.R. (1993), *C4. 5: Programs for Machine Learning*, Morgan Kaufmann, San Francisco.

Read, J.D. (1995), "The availability heuristic in person identification: the sometimes misleading consequences of enhanced contextual information", *Applied Cognitive Psychology*, Vol. 9 No. 2, pp. 91-121.

Riddle, K. (2010), "Always on my mind: exploring how frequent, recent, and vivid television portrayals are used in the formation of social reality judgments", *Media Psychology*, Vol. 13 No. 2, pp. 155-179.

SAS (2013), *Sas/stat(r) 9.2 User's Guide*, 2nd ed., SAS, available at: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_logistic_sect015.htm (accessed October 6, 2015).

Sjöberg, L. and Engelberg, E. (2010), "Risk perception and movies: a study of availability as a factor in risk perception", *Risk Analysis*, Vol. 30 No. 1, pp. 95-106.

Tenenhaus, M., Vinzi, V.E., Chatelin, Y.-M. and Lauro, C. (2005), "PLS path modeling", *Computational Statistics & Data Analysis*, Vol. 48 No. 1, pp. 159-205.

Teversy, A. and Kahneman, D. (1974), "Judgment under uncertainty: heuristics and biases", *Science*, Vol. 185 No. 4157, pp. 1124-1131.

Thomas, J.S. (2001), "A methodology for linking customer acquisition to customer retention", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 262-268.

Tversky, A. and Kahneman, D. (1973), "Availability: a heuristic for judging frequency and probability", *Cognitive Psychology*, Vol. 5 No. 2, pp. 207-232.

Villanueva, J., Yoo, S. and Hanssens, D.M. (2008), "The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth", *Journal of Marketing Research*, Vol. 45 No. 1, pp. 48-59.

Wang, H. and Wang, S. (2008), "A knowledge management approach to data mining process for business intelligence", *Industrial Management & Data Systems*, Vol. 108 No. 5, pp. 622-634.

Wangenheim, F.V. and Bayón, T. (2007), "The chain from customer satisfaction via word-of-mouth referrals to new customer acquisition", *Journal of the Academy of Marketing Science*, Vol. 35 No. 2, pp. 233-249.

Corresponding author
Sang-Gun Lee can be contacted at: slee1028@sogang.ac.kr

(The Appendix follows overleaf.)

Appendix 1

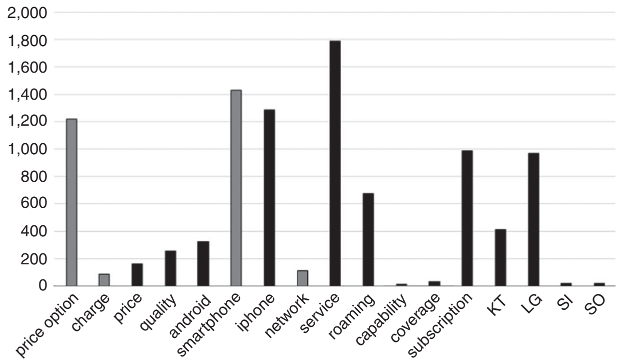


Figure A1.
Frequency

Appendix 2

	AVE	CR	R^2	Cronbach's α	Communality	Redundancy
KT	1	1		1	1	
LG	1	1		1	1	
Churn rate	1	1	0.576405	1	1	0.044434
Price	1	1		1	1	
Subscription	1	1		1	1	
Charge	1	1		1	1	
Network	1	1		1	1	
Roaming	1	1		1	1	
Coverage	1	1		1	1	
Service	1	1		1	1	
Smartphone	1	1		1	1	
iPhone	1	1		1	1	
Android	1	1		1	1	
Price option	1	1		1	1	
Capability	1	1		1	1	
Quality	1	1		1	1	
AVG			0.576405		1	0.044434
Goodness-of-fit	0.759213					

Table AI.
Model fit

Appendix 3

Table AII. Square root average and correlation coefficients

Factor	KT	LG	CR	PC	SS	CG	NW	RM	CG	SV	SP	IP	AR	PO	CB	QL
KT	1.000															
LG	−0.107	1.000														
CR	0.231	−0.374	1.000													
PC	0.019	0.213	−0.194	1.000												
SS	0.335	−0.036	0.089	0.290	1.000											
CG	−0.110	0.022	0.170	−0.130	−0.057	1.000										
NW	0.033	−0.122	0.416	−0.140	0.140	−0.108	1.000									
RM	0.354	0.143	0.048	0.110	0.348	−0.252	−0.089	1.000								
CG	−0.134	−0.276	0.143	−0.161	−0.071	0.033	−0.140	−0.179	1.000							
SV	0.148	0.028	0.290	−0.211	0.223	−0.018	0.287	0.183	−0.091	1.000						
SP	−0.023	−0.255	0.296	−0.366	−0.182	0.133	0.101	−0.249	0.369	0.283	1.000					
IP	0.579	−0.097	−0.123	0.161	0.370	−0.031	−0.213	0.224	−0.086	−0.011	−0.100	1.000				
AR	−0.012	−0.215	0.253	−0.231	−0.066	0.527	0.019	−0.333	0.486	0.109	0.531	−0.150	1.000			
PO	0.083	0.010	−0.096	−0.040	0.044	0.015	−0.124	0.247	−0.154	0.055	0.077	0.033	−0.027	1.000		
CB	0.076	0.098	0.164	−0.084	−0.089	−0.118	−0.123	0.206	0.111	−0.127	−0.218	−0.107	0.017	−0.138	1.000	
QL	0.023	0.103	−0.002	−0.086	0.228	0.070	0.193	0.117	−0.202	0.122	−0.310	−0.085	−0.083	−0.312	−0.052	1.000

Notes: PO, pricing option; CG, charge; PC, price; QL, quality; AR, android; SP, smartphone; IP, iPhone; NW, network; SV, service; RM, roaming; CB, capability; CG, coverage; SS, subscription; CR, churn rate