Expert Systems with Applications 38 (2011) 1999–2006

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Employee churn prediction

V. Vijaya Saradhi *, Girish Keshav Palshikar

Tata Research Development and Design Center, Tata Consultancy Services, 54-B, Hadapsar Industrial Estate, Pune, Maharashtra, India

ARTICLE INFO

Keywords:
Customer churn
Employee churn
Predictive model
Support vector machines
Data mining
Machine learning

ABSTRACT

Customer churn is a notorious problem for most industries, as loss of a customer affects revenues and brand image and acquiring new customers is difficult. Reliable predictive models for customer churn could be useful in devising customer retention plans. We survey and compare some major machine learning techniques that have been used to build predictive customer churn models. *Employee churn* (or attrition) closely related but not identical to customer churn is similarly painful for an organization, leading to disruptions, customer dissatisfaction and time and efforts lost in finding and training replacement. We present a case study that we carried out for building and comparing predictive employee churn models. We also propose a simple *value* model for employees that can be used to identify how many of the churned employees were "valuable". This work has the potential for designing better employee retention plans and improving employee satisfaction.

1. Introduction

The purpose of this paper is twofold. First, we discuss the customer churn problem and survey the use of data mining and statistical techniques that have been used to build predictive customer churn models. Second, we discuss a closely related (but not identical) problem of employee churn (aka attrition) and demonstrate, via a real-life case study that we carried out, how the same data mining and statistical techniques can be used to build predictive employee churn models.

1.1. Motivation, scope and contribution

Customer churn is a key problem in high competitive service markets. This problem gained attention due to its severity and resulting losses organizations has to bear. Employee churn a similar, but not quite exact as customer churn is a problem for every organization. For IT service organizations in particular the employee churn rate is approximately 12–15%. This churn rate is quite high and assuming even a lower churn rate of 5%, the cost involved in an employee leaving a firm is approximately 1.5 times the annual salary of an employee. Assuming that organizations strength (in terms of number of employees) is 140,000 (organization under study) and employee's average salary is $12,000.00 annually, then organization has to lose $12,60,00,000 ($7000 * 1.5 * 12000$). This amount is certainly not a good news for organizations with high employee churn rate (attrition).

Organizations have little insight in taking pro-active measures. In the case of customer churn, a host of solutions have been proposed in the literature for taking pro-active measures (retention strategies) by developing predictive model for customer churn and identifying valuable customers by building customer lifetime value models (Gupta et al., 2006; Rosset, Neumann, Eick, & Vatnik, 2003).

In the present work, we focus on building a predictive model for employee churn problem. We present a case study to demonstrate the use of existing predictive models in the context of employee churn. We also focus on developing an employee value model (EVM) and draw analogies to customer lifetime value model. Ground truth validation has been carried out in order to measure the effectiveness of the proposed EVM.

1.2. The customer churn problem

The phenomenon of customer churn is commonly observed in volatile consumer service markets such as mobile phones (Archaux & Martin, 2004; Hung, Yen, & Wang, 2006; Rosset et al., 2003; Wei & Chiu, 2002) insurance (Morik & Köpcke, 2004), subscription services (Coussement & den Poel, 2008) and banking (Lariviere & den Poel, 2005). We review these attempts in Section 2.1. A customer c is said to churn the service (or product) s when he/she discontinues using s, although the customer may continue to use other products or services. Customer churn is a serious problem for many industries for several reasons.

1. Acquiring a new customer is much harder and more expensive
2. Process-related costs for terminating a customers service are high.

* Corresponding author. Tel.: +91 2066086303.
E-mail addresses: saradhi@iitg.ernet.in (V.V. Saradhi), gk.palshikar@tcs.com (G.K. Palshikar).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.134

3. Losing a customer leads to loss of revenue as well as negative impact on the bottom line.
4. Loss of customers adversely affects many functions within an organization. E.g., churned customers affect the brand-value and may influence prospective customers.

Figures even within a single industry indicate the seriousness of the customer churn problem. For example, in 2001, customer churn rates were around 40% for wireless telephone carriers (Telephony online, 2002) and churn-related costs for them were around US $10 billion.

A voluntary churn happens when a customer voluntarily terminates the services of a company and usually switches to the services of a competitor company. An involuntary churn happens when a company on its own terminates the services of a customer, typically for non-payment of dues. The time period for which a customer stays with the company (for a specific service) is called the customers lifetime. Primary reasons for voluntary churn are availability of more competitors that offer better products/services, more flexibility/convenience, better pricing, better customer support etc. In addition, availability of many competitors makes the customer switch companies for even small reasons e.g., customer support related issues. Involuntary churn happens because of inadequate screening at the timing of acquiring a customer (low entry barriers), lack of flexible payment options, lack of early warnings (e.g., inaccurate or untimely billing) so that customers can attempt to improve the usage etc. In this paper we focus on voluntary churn; similar techniques can be applied to involuntary churn as well. Credit-scoring techniques are also useful for involuntary churn prediction.

A number of strategies can be designed to reduce customer churn; these include:

1. Improve services offered; e.g., wider coverage, better voice quality, cheaper call or activation rates, innovation for competitive differentiation.
2. Offer discounts or other promotional benefits; improve billing services (accuracy, frequency), offer more payment options. Rosset et al. (2003) have shown the uses of customer lifetime value model (CLV) in retention planning.
3. Improve customer support service e.g., improve service time, minimize waiting time.

Several critical issues arise when attempting to achieve a business goal such as reduce customer churn by 50%. The following are relevant from the business angle:

1. *Root cause:* One must identify the root causes for the churning of the customers. This is important because not knowing the real root causes can lead to incorrect remedial measures. For example, if quality of customer support does not affect customer churn then there is no need to add more people in that function.
   Often the root cause of why a customer churned is not directly available (e.g., the customer may not specify it when leaving) and may have to be (correctly) inferred from the data related to the history of that customers business interactions with the company. Obviously, there is a need to validate the suggested root causes.
2. *Retention:* There is very little one can do to make a customer change his/her mind, when the customer has already decided to churn. On the other hand, if there is an early warning (a red flag) then it may be possible to take some pro-active actions (depending on the nature of the red flag) to prevent the customers churn. For example, if the red flag is (“customer duration $\leqslant$3 months AND current monthly payment is delayed”) then the customer can be offered a discount for the next payment. For this purpose, each red flag must be a strong indicator (i.e., predictor) of the impending churn event. Ideally, the red flag should also indicate (or at least have good correlation with) a root cause of why that customer is likely to churn. Thus there is a need for a predictive model which can be applied to individual customers (say, every month) to predict accurately and in time whether or not the customer will churn.
   In Section 2.1 we survey various data mining and machine learning techniques that have been used for building such predictive churn models based on the historical data of customer transactions. Machine learning techniques view churned/not-churned as two class labels for customers and hence consider prediction of churn as a pattern recognition (classification) problem. Statistical techniques view churn as a probabilistic event that occurs (or does not occur) with some probability and hence consider prediction of churn as a problem of building probabilistic models. Both approaches use the historical data for building the predictive models.
3. *Value model:* It is well-known that not all customers are equally valuable (or profitable) for a company. For example, customers who make many international calls are usually more valuable than customers who make only a few local calls. Given the fact that it may not be possible to stop customer churn altogether, one area of focus then could be to prevent the churn of high-value customers. For this purpose, there is a need for an accurate member lifetime value model, which can associate a measure of value with each customer; e.g., expected profits over the customers lifetime with the company. Thus, a predictive churn model should also take into account the value of a customer and should deliver more accurate churn predictions for more valuable customers.
4. *Expectations from models:* Finally, for effective deployment, predictive churn models should be supplemented with the following:
   (a) Models that predict (i.e., suggest) for different types of red flags appropriate actions that will help in retaining that customer.
   (b) Methods to detect significant changes in customer behavior.
   (c) Detect new/emerging needs of the customers.
   (d) Detect new/emerging segments of customers with distinctive needs.

1.3. Employee churn

It is a fact that employees churn i.e., they leave an organization for various reasons. Employee churn (aka attrition) is a serious issue for all organizations, but particularly for high-tech industries, service organizations, etc. Losing an employee is a problem for various reasons:

1. It is difficult to find suitable replacements for employees, particularly those with high experience and special skills.
2. It takes time, effort and money to recruit new employees.
3. Loss of an employee adversely affects ongoing projects and services, which leads to dissatisfaction among customers and other stake-holders.
4. It takes time and efforts for new employees to achieve the same levels of expertise and productivity.
5. Loss of an employee costs money. Employee churn rates can be as high as 12–15% annually.

In the present work, we focus on voluntary employee churn where employees leave an organization for their own reasons. There are many reasons why employees churn. The positive reasons include offer of better things (work, pay, perks, work conditions and facilities, career growth, leadership, location, etc.). Negative reasons include conflicts with supervisors or colleagues, lack of things (appreciation, interesting work, training, career growth, focus/direction, etc.), low pay, bad working conditions, etc. While exit interviews provide a good idea of the reasons why an employee is leaving, there is still a need to infer corroborate the stated reasons by other means such as analysis of the persons work history.

Just as for customer churn, predictive models for employee churn would be useful to understand root causes for employee churn, plan retention strategies, plan recruitment and improve team management. Not all employees are equally good performers; e.g., employees who have demonstrated excellence in specific tasks and employees with specialized knowledge are perhaps more valuable. Hence, predictive models for employee churn focused on accurately identifying churn of "valuable" employees would be more beneficial. Towards that end, a model of employee lifetime value would be useful.

1.4. Comparison

Customer churn and employee churn are two similar yet not identical problems which are of concern for organizations. Reasons for customer churn and employee churn are significantly different. As noted in the previous two sections, in a highly competitive service markets, customer churn for very small reasons (better products, offerings, customer support and the like). In the case of employee churn, the reasons are quite complex and also involve human element in the decisions (examples include spouse transferred to a different location, relations with the supervisor and the like).

Customer retention strategies are based on the reasons why a customers leave a particular service. Though the reasons are not explicitly available, upon careful examining of the past customer data reason can be better understood to device retention strategies (Rosset et al., 2003; Gupta et al., 2006) and effectiveness of the retention strategies can immediately be measured. However the retention strategies in the employee churn context are quite complex and often there is no direct handle for measuring effectiveness of the strategies. An example of retention strategy in the case of customer churn for a cellular telephone services company could be free battery, reduced price of handset upgrade (Rosset et al., 2003). However these strategies are not specific to any customer. In the case of employee churn retention strategies could be very specific (offering a location transfer).

In the case of customer churn, every customer is valued according to the value the customer brings for the organization. This value is computed for the lifetime of the customer (time for which the customer is associated with the organization for utilizing the services offered). Typically the customer lifetime value (CLV) is computed using historic data of customer and the value computed is a *projected value* in future time.

Employees do not bring revenue directly for organizations. Employees are valued often on varieties of ways; for example, for meeting deadlines of projects assigned to employee. Also, it is difficult to project employee value in future (unlike in the context of customer churn).

Despite the differences, models build to predict both the problems are same and the degree of success varies from predictive model to predictive model and the attributes that characterize customers and employees.

This paper is organized in the following way: we review some of the predictive models used in the context of customer churn in Section 2.1. In Section 2.2 we present a case study that is carried out on an large organization in which a small portion of employees work for a big client of the organization under study. In Section 3.1, we recollect the customer lifetime value models proposed in the literature. Employee value model is proposed in Section 3.2. Validation of the proposed employee value model is carried out in Section 3.3. We conclude the paper in Section 4.

2. Predictive models

Predictive models are now widely used in many business applications to predict occurrences of events. In this section, we review some of the predictive modeling techniques that have been used for customer churn. We will also mention some case studies in the research literature for customer churn prediction. We review some approaches for measuring customer lifetime value and how they are incorporated in churn prediction models.

The problem of supervised learning consists of learning classification rules by generalizing from given training examples. This supervised learning problem is stated informally as follows: given a set of objects each labeled with one of k distinct class labels, learn (discover) a function f which classifies every (new) object into one of the k classes. The set of labeled objects is called training data. The new objects for which the class label is to be predicted are usually not part of the training data; they are called test data. The method used to find function f is referred to as a learning algorithm. In the present context, given a set of customer records labeled with churned or not-churned, the supervised learning task is to discover a rule which can then be used to classify a new customer record into one of the two categories. Learned function f is thus a predictive model (Saradhi, 2008).

Validation of the learned function is carried out using several techniques. Popular among them are (1) dividing, at random, the available data into 80:20 ratio. Eighty percent of records are used for learning the function f and remaining 20% of records are used for testing the effectiveness of learned function, (2) using a separate validation data set and (3) cross validation: the data is split into k equal sets and function f is learned using $k-1$ sets as training data and effectiveness is tested against the set which was left out.

2.1. Predictive models for customer churn

In this subsection we review predictive models that have been employed in the literature for customer churn problem.

2.1.1. Naïve Bayes

Naïve Bayes is a popular classification technique in the machine learning literature and has attracted attention for its simplicity and performance (Mitchell, 1997). This method computes a posteriori probabilities for each class. In the customer churn context these are the probabilities of observing churn and observing non-churn given a customer record. These a posteriori probabilities for each class given a specific customer record are computed using the Bayes rule and the Naïve Bayesian assumption. Thus the learned function f is nothing but a table of probabilities. The Bayes decision rule is to assign a new customer record to that class which has maximum a posteriori probability. An important assumption in the Naïve Bayes classifier is that the attributes used for describing customers are conditionally independent. Naïve Bayes technique has demonstrated its success in the text categorization domain. This technique has been employed in many case studies and in particular wireless telephony industry. However, its success in the customer churn prediction is limited.

2.1.2. Support vector machines

Support vector machine (SVM) is a supervised learning algorithm used for binary classification. Considering that the input training data points (each with a class label 1 or +1) are "scattered" in an n-dimensional space, SVM identifies a suitable linear function (hyperplane) which optimally separate the training data points belonging to the two classes. Thus all points with class label 1 are on one side of the discovered hyperplane and all points with class label +1 are on the other side of the discovered hyperplane. SVMs use what is widely known as the kernel trick to obtain a non-linear decision function for data points that are not linearly separable. In the kernel trick, data points in the input space are mapped using a non-linear mapping $\varphi(\cdot)$ into a high dimensional space (called the feature space). The inner product of two points from the feature space is replaced with a kernel function: $K(\mathbf{X}_i, \mathbf{X}_j) = \langle \varphi(\mathbf{X}_i), \varphi(\mathbf{X}_j) \rangle$. The class of algorithms that employ this kernel trick for learning is known as kernel methods.

The optimal separating hyperplane is found by maximizing the margin i.e., the distance between the closest data points lying on either side of the hyperplane. This maximization is formulated as a quadratic programming problem with the constraints that all points belonging to one class (say + 1) should lie on one side of the hyperplane and points belonging to different class (say -1) should lie on other side of the hyperplane.

Primal formulation 2.1. Given the training data set $\{(\mathbf{X}_i, y_i)\}_{i=1}^N$, find $\mathbf{w}$, b and ξ that minimize the objective function subject to the given constraints:

$$\begin{aligned} &\text{minimize}_{\mathbf{w},b,\xi} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i, \\ &\text{subject to} \quad y_i(\langle \mathbf{w}, \mathbf{X}_i\rangle + b) \geqslant 1 - \xi_i \quad \forall i = 1, \ldots, N, \\ &\qquad\qquad\quad \xi_i \geqslant 0, \end{aligned} \tag{1}$$

where C is an user defined parameter. This free parameter plays a critical mediator role between the margin of separation and the number of training errors. For small values of C the margin of separation is larger, while for larger values of C we get fewer training errors (Pontil & Verri, 1998).

Dual formulation of the above primal formulation is as given below:

Dual formulation 2.1. Given the training data set $\{(\mathbf{X}_i, y_i)\}_{i=1}^N$, find the optimal αs that maximize the objective function subject to the given constraints:

$$\begin{aligned} &\text{maximize}_{\alpha} \quad \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \langle \mathbf{X}_i, \mathbf{X}_j\rangle, \\ &\text{subject to} \quad \sum_{i=1}^{N}\alpha_i y_i = 0, \\ &\qquad\qquad 0 \leqslant \alpha_i \leqslant C \quad \forall i = 1, \ldots, N. \end{aligned} \tag{2}$$

where α_i's are Lagrangian multipliers associated with each constraint in the primal formulation. $\langle \cdot, \cdot \rangle$ denotes inner product.

Those data points for which $\alpha_i > 0$ lie on the hyperplane and are known as *support vectors* (SVs). The weight vector is given by $\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{X}_i$. The label of a test data point $\mathbf{X}$ is determined by which side of the hyperplane the data point is located. The predicted label (+1 or -1) is given by

$$f(\mathbf{X}) = \text{sign}\left(\sum_{i=1}^{N}\alpha_i y_i \langle \mathbf{X}_i, \mathbf{X}\rangle + b\right). \tag{3}$$

Churn in subscription application: SVMs has been employed in a Belgian newspapers subscription case study (Coussement & den Poel, 2008). Subscribers for this newspaper publishing company have to pay a fixed amount to get the newspaper for a specified period of time. Subscribers are not allowed to discontinue the contract before the specified period. Data for a period of 4 years and 8 months was collected. Attributes that characterize customer include client–company interaction information (e.g., the total number of complaints for the subscription), renewal related information (last renewal, how many times customer has renewed, how many days before contract expiration period the customer has renewed the current subscription, etc.) and customers socio-demographic information (age, gender, person/company, etc.). Forty five thousand records were used for training the SVM of which 50% were churned and 50% were non-churned customers. Each record was described using 82 attributes. Test data set contained 45,000 records of which 11.14% records stand for customers who have not renewed their subscription. SVMs performance on the test data set is reported to be around 88.63% and is claimed to be competing with the best algorithm namely random forests (whose accuracy on the same test data set is around 89.14%).

Customer churn in insurance industry: Another relevant case study in the insurance domain has been carried out by using SVMs (Morik & H. Köpcke, 2004). A Swiss-Life insurance companies data comprising of 217,586 policies for a total of 163,745 customers have been used. Domain specific term frequency and inverse document frequency like features were extracted from this data. SVMs were employed using two instances of this data; one in which time stamp information was ignored and another in which time stamp information was taken into account. The data having no time information was shown to perform poorly (including SVMs). Predictive accuracy for the data which takes into account time-stamped information was reported to be around 99.71%.

2.1.3. Decision tree and random forests

Decision trees gained popularity due to the ease of interpretation of the discovered rules. Given a training data set, this learning algorithm constructs a tree in which each node is an attribute and branches of the nodes are corresponding attribute values. Nodes in the decision tree are arrived on the basis of the explanatory power of the attributes (measured in terms of a quantity called information gain) (Duda, Hart, & Stork, 2001).

One problem with decision tree learning algorithm is that it is an unstable algorithm (small changes in the training data lead to large variations in the classification performance). To overcome this problem, Breiman proposed random forests (Breiman, 2001). Idea in this approach is to construct multiple decision trees on sampled data (obtained through bootstrap re-sampling technique) using only a subset of attributes. The final class label for a new data point is obtained by combining (using a voting based scheme) the decisions of all the trees thus built.

Despite decision trees instability, in practice, they have been applied in many case studies. The Taiwan wireless telecommunication case study reported in Hung et al. (2006) uses decision trees for customer churn prediction. The data set comprises of 160,000 customers with 8.75% customer churn (14,000). Each customer record is characterized by socio-demographic (age, tuner, gender, etc.), billing and payment analysis (monthly fee, billing amount, etc.), call details (call types) and customer care service attributes. Performance of random forests is close to 98% in this case study.

Random forests have been successfully used in case studies like news paper subscription (Coussement & den Poel, 2008) and banking (Lariviere & den Poel, 2005). In the news paper subscription case study (Coussement & den Poel, 2008), Section 2.1.2, random forests are shown to perform better than SVMs at instances with recognition rates 89.14%.

In the banking domain case study carried in Lariviere and den Poel (2005), data set of a Belgian financial institution is used for building random regression forests (which is the extension of random forests to the case where class labels are real numbers not finite discrete in number). A total of 50,000 customer records were used which were collected over a period of 8 months. Attributes extracted on this data include: past customer behavior (specific banking products owned, self banking activity, number of products owned, monitory value, cross products, etc.), customer demographics (age, gender, geographical region, etc.). A classification accuracy of 75.1% using random forests has been reported.

2.1.4. Logistic regression

This technique obtains the a posteriori probabilities by assuming a model for the same and estimates the parameters involved in the assumed model. The form of the model is given below:

$$p(\text{churn}|\mathbf{w}) = \frac{1}{1 + e^{-\left[w_0 + \sum_{i=1}^{N} w_i \mathbf{x}_i\right]}}. \tag{4}$$

The parameters $\mathbf{w}$ are estimated using maximum likelihood estimation technique (King & Zeng, 2001). The logistic regression technique has been shown to compete with other learning models (SVMs and random forests). In the Belgian banking case study (Lariviere & den Poel, 2005), logistic regression performance is close to that of random forests with classification accuracy of 74.5% on the test data; In the case of Belgian news paper publishing company case study (Coussement & den Poel, 2008), this method has yielded a classification accuracy of 88.47%.

2.1.5. Comparison

Various learning techniques discussed above have been used in customer churn prediction problem. Non parametric models, such as SVMs, have registered their success in the domains when there are many events available for a given customer. A comparative study of SVM, logistic regression, random forests in subscription domain has been carried out by Coussement and den Poel (2008) and it is argued that random forests are efficient. Widely varying views were reported which argue for the superiority of one technique over others in the literature for the churn prediction problem. However, we feel the true potential of kernel based methods is yet to be explored in the customer churn problem.

2.2. Predicting employee churn

In this sub section, we present a case study that is carried out for building a predictive model for employee churn using employee work history data.

2.2.1. Input data set

We consider only a subset of employees working in a specific client unit within a large organization. The data set is collected over a period of two years. Details of each employee at the end of every month were recorded throughout this time window. Therefore, in a given month, the data set contains as many records as the number of employees in that given month. Each record stands for a specific employee and associated attributes are recorded. If an employee has resigned (or has been transferred out of the unit) in a particular month, that employee's record will not appear in the following months.

The data includes 25 attributes for each employee: (1) parent organization employee identification number, (2) client organization identification number, (3) parent organization joining date, (4) client organization reporting date, (5) billing start date, (6) past experience, (7) experience in parent organization, (8) experience in parent organization (years), (9) experience in parent organization categorization, (10) level, (11) client organization experience, (12) client organization experience in months, (13) client organization experience categorization, (14) parent organization designation, (15) client organization designation, (16) employee location, (17) on-site/off-shore, (18) delivery location, (19) billed/not billed, (20) project name, (21) group name, (22) sub group name, and (23) month recorded and two technical attributes related to on-site/off-shore.

Apart from the above attributes, we use parent organization specific attributes such as age, gender, marital status, department, designation and branch. The class labels are resigned, released, and retained. This data differentiates between employees who have resigned from employees who have been released. The released employees have either joined a different project or moved out for a different branch. A qualitative attribute gives a reason for the resignation/release.

Total number of employees in the client unit is 1575. In the two years window, 186 employees resigned (approximately 11.81%) and 212 employees were released (approximately 13.46%). The percentage of working employees is much higher compared to that of released or resigned employees. The learning algorithm should account for this class imbalance.

This case study differentiates resigned, released and retained employees. Accordingly three models were built using SVMs, random forests and Naïve Bayes classifiers. Details of the data sets used for building different models are presented in Table 1. The data set is divided at random into 80:20 and 80% of the records were used as training input for the classifier and 20% of the records were used as testing input.

2.2.2. Importance of derived attributes

The first step in building any predictive model is data cleaning/preprocessing and arriving at a set of features that account for the class label to be predicted (resign and/or release in the present case study). In the data cleaning step, if some attributes have not been recorded, they are filled up using appropriate domain knowledge (e.g., if the past experience in number of years is not available, it is assumed that the employee is a fresher with 0 years of past experience). Categorical attributes were appropriately assigned values. For example, marital status can take three values: married, single and other which are mapped to 1, 2 and 3, respectively. Other categorical attributes are mapped similarly (e.g., department, designation, etc.).

After preprocessing, the raw attributes were used to build a predictive model. Discouraging results were observed when only the given (original) attributes were used. The following experiment justifies our observation. A total of 12 attributes are used in this experiment: not all attributes among the original 25 attributes are useful in predictive modeling; e.g., employee number, project name, etc. The relevant 12 attributes include employee: (1) age, (2) designation, (3) gender, (4) department, (5) qualification, (6) past experience in years, (7) employee location, (8) experience in parent organization, (9) experience in client organization, (10) billed or not billed, (11) on-site/off-site, and (12) designation in client organization.

SVM was employed for learning from the data set at hand. We choose a Gaussian kernel function with varying spread widths and user parameter C. Best values for these free parameters (γ, C) on training data set were chosen as the model parameters and

Table 1
Data set summary.

Model		Training data	Test data
I	Resigned + released	329	69
	Retained	931	246
II	Resigned	156	30
	Retained	934	242
III	Released	176	36
	Retained	935	241

tested against the test input data set. A classification accuracy of 71.31% is obtained on the test data set (Model 1 in Table 1). Though the overall classification accuracy is not discouraging, the predictive model fails to accurately classify those employees who have resigned (or released) known as true positives (TP). The TP accuracy for thus built model is around 42.10%.

To improve the classification accuracy and the TP recognition rate, derived attributes are introduced, which are summaries of the employees work history within the 2 years time window. Note that we have access to the employee data with the original 25 attributes. Using which we have derived the following attributes. Having access to full data set is quite difficult and hence we have computed summaries of the various attributes. These include: (1) number of promotions in parent organization, (2) number of promotions in client organization, (3) number of project an employee involved in the client organization, (4) number of billed months, (5) number of non-billed months, (6) number of months employee spent on-site, (7) number of months employee spent off-site, (8) relative stay of employee in the client organization (computed as ratio of experience of employee in client organization to experience in parent organization), (9) relative stay of employee in parent organization (ratio of experience of employee in parent organization to past experience), (10) how long an employee spent in the same designation in the parent organization, (11) how long an employee spent in the same designation in the client organization, (12) number of transfers employee taken, (13) number of levels employee achieved, (14) number of groups in which employee involved in client organization, (15) number of sub-groups in which employee involved, (16) number of months delayed in billing to the client, (17) percentage billed months, (18) percentage non-billed months, (19) percentage months on-site, and (20) percentage months off-site.

The above 20 derived attributes were included as features for building a predictive model. Three models (Model 1, Model 2 and Model 3, Table 2) were built using SVM, random forests and Naïve Bayes classification methods on the training data set. Class imbalance in SVM is handled through adjusting class penalties 2. Other classification techniques are limited by class imbalance problem. The results of three learning methods are presented in Table 2. Each cell in Table 2 stands for classification accuracy (percentage) on test data set for corresponding classification technique and Model.

Following observations can be made from Table 2:

1. Introducing derived attributes helped in discriminating resign/release and retained employees; this is evident through increase in the TP recognition rate.
2. SVMs success over random forests and Naïve Bayes classifier is attributed to introducing individual class penalties in SVMs. True positives accuracy is very high compared to other methods.
3. Random forests and Naïve Bayes compete with SVMs on the total accuracy front; however, they fail to impress along the TP recognition rate.

Table 2
Experimental results using three learning techniques for all models.

Method	I	II	III	Remarks
SVM	80.00 (252/315)	81.99 (223/272)	84.12 (233/277)	Total accuracy
	81.16 (56/69)	**80.00 (24/30)**	**77.78 (28/36)**	TP
	79.67 (196/246)	82.23 (199/242)	85.06 (205/241)	True negatives (TN)
Random forests	83.49 (263/315)	92.28 (251/272)	88.81 (246/277)	TA
	53.62 (37/69)	50.00 (15/30)	47.20 (17/36)	TP
	91.90 (226/246)	97.50 (236/242)	95.00 (229/241)	TN
Naïve Bayes	73.97 (233/315)	79.04 (215/272)	77.62 (215/277)	TA
	55.10 (38/69)	63.33 (19/30)	44.40 (16/36)	TP
	79.30 (195/246)	81.00 (196/242)	82.60 (199/241)	TN

3. Value models

3.1. Customer lifetime value model

Customer lifetime value (CLV) is *"the present value of all future profits obtained from a customer over his/her life of relationship with a firm"* (Gupta et al., 2006). Gupta et al. (2006) have extensively surveyed various modeling techniques for customer lifetime value and identified six approaches for modeling CLV. These include: (a) recency frequency and monitory value model (RFM) (Fader, Hardie, & Lee, 2005), (b) probability models focus on predicting the observed behavior (David, Morrison, & Colombo, 1987; Reinartz & Kumar, 2000), (c) econometric models whose philosophy is similar to that of probability models, specifically they use hazard models used for modeling customer acquisition (Thomas, 2001), retention (Lewis, 2003; Venkatesan & Kumar, 2004) and expansion (Thomas, Blattberg, & Fox, 2004), (d) persistence models captures the modeling behavior of acquisition, retention and cross-selling (Julian, Yoo, & Hanssens, 2006; Shijin & Hanssens, 2005), (e) computer science models random forests, support vector machines (SVMs) (Breiman, 2001; Vapnik, 1998) have been employed for customer churn, and (f) diffusion and growth models (Kumar, 2006).

Rosset et al. (2003) captured dependency of the following three quantities for modeling CLV:

1. Customer value: how valuable a customer is to the company. This is estimated using customer's value $v(t)$ (for time $t \geqslant 0$).
2. Length of service (LOS). This describes customer's churn probability over time described using survival function $S(t)$. $S(t)$ describes the probability that customer will still be active at time t. For LOS, the hazard function $h(t)$ is the most commonly modeled function.
3. Discounting factor $D(t)$ which describes how much the profit gained in future is worth at present; usually an exponential decay function is used for $D(t)$.

CLV is expressed using the above three quantities as $\int S(t)v(t)D(t)\,dt$, where the integral is taken over the entire time domain.

Customer's value $v(t)$ is estimated from the customer's data in combination with business knowledge. The LOS is computed through one of the following three methods: (a) parametric approach; typically a distribution is assumed whose parameters depend on the data variables, (b) semi-parametric approach; in this the model comprises of time independent linear function involving the data and time dependent risk, (c) non-parametric approach which makes use of the data for estimating $S(t)$.

In the customer retention model, CLV is modified such that retention strategies are reflected. A detailed case study on mobile telephony has been carried out in Rosset et al. (2003) using customer CLV. This case study, which has been carried out for an Amdocs cellular telephony customer, uses non-parametric approach to model CLV. The goal in this case study is to extract useful information from customers CLV in the perspective of business and marketing activities. In particular, this case study not only estimates CLV for a given segment of customers, it also provides valuable information in the presence of various retention strategies. The case study examines the effect of providing free caller-id service or upgrading of handset at discounted price on CLV and in turn churn rate.

3.2. Employee value model

Though total accuracy along with TP and TN accuracy is encouraging, it is not good enough for large organizations. If the total number of employees is 140,000, the above learning algorithms would predict around 28,000 employees as potential churners (assuming 20% error in churn). For retention such large numbers are certainly a challenge, as it is difficult talking individually to each employee to assess their reasons for churning.

Alternatively, we start with the observation that all churning employees are not equally valuable to the organization. We develop a model for an employees value to an organization and use the model to identify only the valuable ones among those predicted to churn. In the case of CLV, it is well understood that CLV depends on the monitory value that a customer brings to a firm. However, employees in any organization brings value indirectly. To model an employee value, it is quite hard to attribute uniquely a feature that totally accounts for employee value. The following table note a few differences between CLV and EVM.

In the present case study, an employee value model is developed that associates a value with each employee based on the following attributes: (a) importance of the project an employee is working, (b) number of months a client was billed with the employee, (c) number of months a client was *not billed* with the employee, (e) number of months an employee was deputed on-site assignment, and (f) number of months an employee was off-site assignment. In order to associate importance, a non-linear model is employed with the identified attributes. Fig. 1 depicts the value model. This model in Fig. 1 assigns large value to those employees who are involved in **important projects** and with **high** number of billed months. This idea is extended over the identified attributes. Value of i^{th} employee is computed as

CLV	EVM
Derives revenues from creation and sustenance of long term relationship with customer	Employees help in achieving creation and sustenance of long term relationship with customer
Marketing serves the purpose of maximizing the CLV	Who is responsible for maximizing EVM? Human resource personnel?
A dis-aggregated *metric* to identify profitable customers'	May help in identifying *valuable employees*
	Value: depends on various characteristics of an employee
Good proxy for a firms value	Can EVM help in assessing a firms value?

$$\mathrm{EVM}_i = \mathrm{project\_weight} \times \sum_{p_j \in \mathrm{projects}} \Big(\mathrm{months\_onsite}_{p_j} + \mathrm{months\_offsite}_{p_j} + \mathrm{billed\_months}_{p_j} - \mathrm{non\text{-}billed\_months}_{p_j}\Big).$$

3.3. Validation and comparison

Using the non-linear model, values are assigned for each of the 1575 employees. The values thus obtained are arranged in sorted order and a threshold is chosen (e.g., mean). Valuable employees are those whose value is greater than a specified threshold. Main challenge in the EVM is validating the model. In the present work,

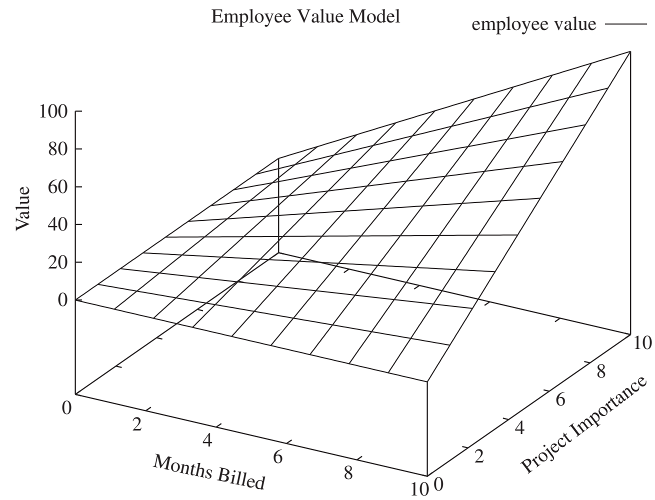

Fig. 1. A non-linear employee value model.

the above employee value model is validated using ground truth information. When an employee resigns or is released, a qualitative attribute (indicating whether the resign/release is a **regret** or **not regret**) is added to that employees data. This qualitative attribute has been used for validation.

Intersection of the set of valuable employees and set of employees who have resigned/released can now be easily computed. Fig. 2 shows important/unimportant employees among those who have resigned/released. Using the employee value model, 41 resigns out of 186 (i.e., 22.04%) were identified as valuable and rest of the resigns as not-valuable. Similarly, using the employee value model, 51 released out of 212 (i.e., 24.06%) were identified as valuable and rest of the released as not-valuable.

The challenge in the present context is to validate the model. How many of the 41 resigned employees identified as valuable (by the employee value model) were really important? The results using employee value model are validated using the ground truth data provided. The ground truth provides a qualitative attribute (whether a particular resign/release is regret or not). We use this information to evaluate the employee value model. Out of the 41 resigns identified as valuable, 26 (i.e., 63.41%) were labeled as regret (i.e., important) by the organization. Out of the 51 released identified as valuable, 10 (i.e., 19.61%) were labeled as regret (i.e., important) by the organization. Fig. 3 shows what percentage of

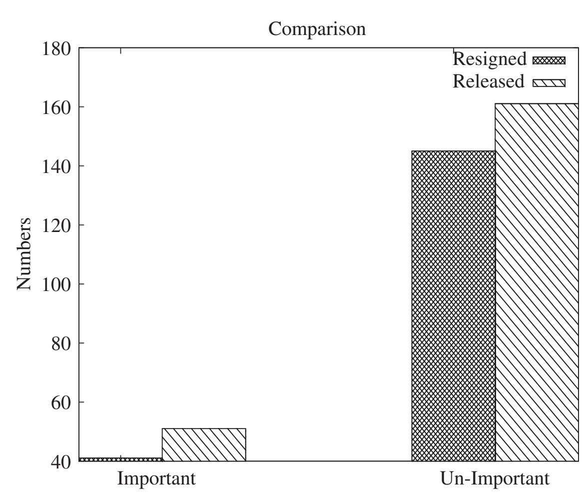

Fig. 2. Valuable employees obtained from employee value model.

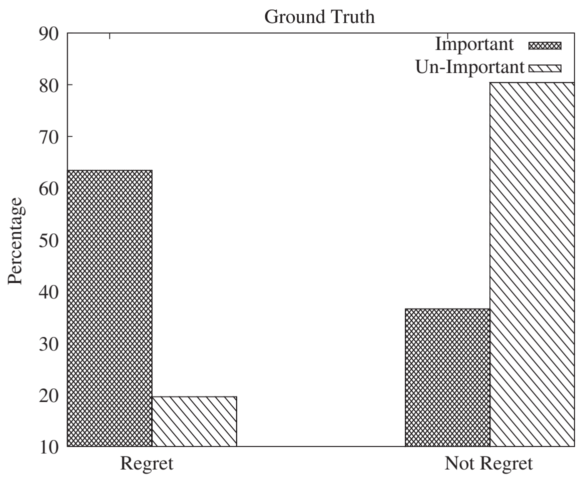

Fig. 3. Percentage valuable employee predicted as churners by SVM.

employees obtained using employee value model followed by thresholding is actually an important resign/release (i.e., the resign/release is regret).

As releases from a project are *pre-planned*, the number of valuable employees getting released is less and hence our results are consistent with respect to both resigned and released employees.

4. Conclusions and future work

Customer churn is a notorious problem for most industries, as loss of a customer affects revenues and brand image and acquiring new customers is difficult. It is important to build reliable predictive models for customer churn which could be useful in devising customer retention strategies. Employee churn (or attrition) closely related but not identical to customer churn is similarly painful for an organization, leading to disruptions, customer dissatisfaction and time and efforts lost in finding and training replacement. The purpose of this paper was twofold. First, we discussed the customer churn problem and surveyed and compared the use of data mining and statistical techniques that have been used to build predictive customer churn models. Second, we presented a case study that we had carried out for building and comparing predictive employee churn models. We also proposed a simple value model for employees that can be used to identify how many of the churned employees were valuable. This work has the potential for designing better employee retention plans and improving employee satisfaction.

The major contribution of this paper is the demonstration that machine learning techniques (e.g., SVM) can be used to build reliable and accurate predictive models for employee churn. We also proposed a quantitative model of the value of an employee for an organization. Such a model can be used to design employee retention plans targeted towards retaining valuable employees. The value model is very simple and needs considerable experimentation before it can be in widespread use. Further, unlike customer lifetime value models, our employee value model does not make use of too much data about employee work history nor does it project the discount value of the employee in future. In that sense, we only consider the present value of an employee. We are working on improving the accuracy of the employee churn prediction models, including enriching the input data representation. We will be working with a more event-oriented temporal representation of the employee work history data and will use survival analysis techniques to build employee churn prediction models.

Acknowledgments

The authors thank Prof. Harrick Vin for continuous support and encouragement. We owe a lot to TCS HR, Dinanath Kholkar, Nanndkishore Bhavsar and other associates for their enthusiasm and commitment. Thanks to Shailesh Deshpande and the QUEST team for help. We appreciate the contributions of Dr. Maitreya Natu. Girish Palshikar would like to thank Dr. Manasee Palshikar for her encouragement.

References

Archaux, C., Martin, A., & Khenchaf, A. (2004). An SVM based churn detector in prepaid mobile telephony. In *International conference on information and communication technologies: From theory to applications*.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Coussement, K., & den Poel, D. V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications, 34*, 313–327.

David, S. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who they are and what will they do next? *Management Science, 33*(January), 1–24.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (second ed.). John Wiley and Sons Inc..

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., et al. (2006). Modeling customer lifetime value. *Journal of Services Research, 9*(2), 139–155.

Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications, 31*, 515–524.

Julian, V., Yoo, S., & Hanssens, D. M. (2006). *The impact of marketing-induced vs. word-of-mouth customer acquisition on customer equity*. Working paper. Los Angeles: University of California, Anderson School of Management.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9*(2), 137–163.

Kumar, V. (2006). *Customer lifetime value*. Thousand Oaks, CA: Sage.

Lariviere, B., & den Poel, D. V. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications, 29*, 472–484.

Lewis, M. (2003). *Customer acquisition promotions and customer asset value*. Working paper. Gainseville: University of Florida.

Mitchell, T. (1997). *Machine learning*. USA: McGraw Hill.

Morik, K., & Köpcke, H. (2004). Analysing customer churn in insurance domain – A case study. In J.-F. Boulicaut, et al. (Eds.), *Lecture notes in artificial intelligence* (Vol. 3202, pp. 325–336).

Pontil, M., & Verri, A. (1998). Properties of support vector machines. *Neural Computation, 10*(4), 955–974.

Reinartz, W., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*(64), 17–35.

Rosset, S., Neumann, E., Eick, U., & Vatnik, N. (2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery, 7*, 331–339.

Saradhi, V. V. (2008). *Kernel methods: Some improvements and some analysis*. Ph.D. Thesis. Kanpur, UP, India: Indian Institute of Technology Kanpur.

Shijin, Y., & Hanssens, D. M. (2005). *Modeling the sales and customer equity effects of the marketing mix*. Working paper. Los Angeles: University of California, Anderson School of Management.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). RFM and CLV: Using iso-CLV curves for customer base analysis. *Journal of Marketing Research, 42*(4), 415–430.

Telephony online (2002). *Standing by your carrier*. <http://currentissue.telephonyonline.com/> March 2002.

Thomas, J. (2001). A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research, 38*(2), 262–268.

Thomas, J., Blattberg, R., & Fox, E. (2004). Recapturing lost customers. *Journal of Marketing Research*(16), 31–45.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley and Sons.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*(4), 106–125.

Wei, C.-P., & Chiu, I.-T. (2002). Turning telecommunication call details to churn prediction: A data mining approach. *Expert Systems and Applications*(23), 103–112.