# Customer Churn Prediction Using Neural Network Approach

Su Mon Ko<sup>1</sup>, Mie Mie Aung<sup>2</sup>, Yin Yin Cho<sup>3</sup>

<sup>1,2,3</sup>Information Technology Support and Maintenance Department, University of Computer Studies (Meiktila), Meiktila, Myanmar

Abstract— Customer churn is one of the most important metrics for a growing business to evaluate. It is a business term used to describe the loss of clients or customers. In the retail sales and marketing industry, customers have multiple choices of services and they frequently switch from one service to another. In these competitive markets, customers demand best products and services at low prices, while service providers constantly focus on getting hold of as their business goals. An increase in customer retention of just 5% can create at least a 25% increase in profit. Therefore, customer churn rate is important because it costs more to acquire new customers than it does to retain existing customers. This present of this paper is to classify customers in Customer Relationship Management (CRM) based upon a list of selected customer information by using Backpropagation Feed Forward approach. It will help the retail sales and marketing company to present the targeted customers with the estimated loss across all non-churners for the promotion in direct marketing.

*Keywords*— Data Mining, Customer Churn Prediction, Neural Network, Customer Relationship Management, Backpropagation Approach

#### I. INTRODUCTION

ata mining (DM) methodology has a tremendous contribution for researchers to extract the hidden knowledge and information which have been inherited in the data used by researchers and it is to extract the knowledge and information which have been hidden in a large volume of data. The rapid growth of the market in every sector is leading to a bigger subscriber base for service providers. Service providers have realized the importance of the retention of existing customers. Satisfying customer's needs is the key for business success. Customer Relationship Management (CRM) is a business strategy that aims to understand, anticipate and manage the needs of an organization's current and potential customers. Customer retention has become a significant stage in CRM, which is also the most important growth point of profit. Retail Sales and Marketing across the world are approaching saturation levels. Therefore, the current focus is to move from customer acquisition towards customer retention. In this paper, churn prediction can be viewed as a supervised classification problem where the behavior of previously known churners and non-churners are used to train a binary classifier. Data mining techniques are used to implement customer classification in CRM because mass volume of data is needed to analyze by implementing an efficient and effective Neural Network based technique. Backpropagation is used to classify the selected customers into churners and non-churners to maximize the financial profits of the retention program. Profit-based loss function is used to calculate the estimate loss for the non-churners.

## II. RELATED WORKS

Berry and Linoff (2000) defines data mining as the process of exploring and analyzing huge datasets, in order to find patterns and rules which can be important to solve a problem. Berson et al. (1999); Lejeune extract or detect hidden patterns or information from large databases. Data mining is motivated by the need for techniques to support the decision maker in analyzing, understanding and visualizing the huge amounts of data that have been gathered from business and are stored in data warehouses or other information repositories. Data mining is an interdisciplinary domain that gets together artificial intelligence, database management, machine learning, data visualization, mathematic algorithms, and statistics data mining is considered by some authors as the core stage of the Knowledge Discovery in Database (KDD) process and consequently it has received by far the most attention in the literature (Fayyad et al., 1996a). Data mining applications have emerged from a variety of fields including marketing, banking, finance, manufacturing and health care (Brachman et al., 1996). Moreover, data mining has also been applied to other fields, such as spatial, telecommunications, web and multimedia.

#### III. THEORETICAL BACKGROUND

Data Mining is very famous technique for churn prediction and it is used in many fields. It refers to the process of analyzing data in order to determine patterns and their relationships. It is an advanced technique which goes deep into data and uses machine learning algorithms to automatically shift through each record and variable to uncover the patterns and information that may have been hidden. Data mining is used to solve the customer churn problem by identifying the customer behavior from large number of customer data. Its techniques have been used widely in churn prediction context such as Support Vector Machines (SVM), Decision Tree (DT), Artificial Neural Network (ANN) and Logistic regression.

#### A. Customer Churn Prediction Model

Customer Relationship Management (CRM) system have been developed and it is applied in order to improve customer acquisition and retention. Increase of profitability and to support important analytical tasks such as predictive modeling and classificaiton; CRM applications hold a huge set of information regarding each individual customer. This information is gained from customers' activity at the company, data entered by the customer in the process of registration.

The size of gathered data is usually very large, which results in high dimensionality, making to analyze a complex and challenging task. Therefore, before beginning to use a churn prediction method a data reduction technique is used, deciding with application domain knowledge which attributes can be of use and which can be ignored. Missing values should also be regarded – on attribure level these can be ignored if they are with low significance, whereas on record level they have to be replaced with a reasonable estimate. Providing a good estimate for these missing values is an important issue for proper churn prediction.



Backpropagation Neural Network

Fig.1 Customer Churn Prediction Model

#### B. Regression Analysis

Regression is considered to be a good technique for identifying and predicting customer satisfaction. For each of the variables in a regression model the standard error rate is calculated using SPSS. Then the variables with the most significance in respect to linear regressions for churn prediction are obtained and a regression model is constructed. The logistic regression model is simply a non-linear transformation of a linear regression model. The standard representation of logistic regression is referred as logistic function. The estimated probability of churn is estimated with the function,

$$Pr[churn] = 1/1 + e - T(1)$$

#### Where T = a + BX.

Here a is a constant term, X represents the predictor attributes vector and B is the coefficient vector for the predictor attributes. If T equals 0, the probability is 0.5. This mesns that it is equi-probable that a customer is a churner and non-churner. With T growing large the probability comes closer to 1, so the customer becomes a more probable churner, when T is becoming small the probability of churn is tending to be 0.

#### C. Naïve Byes

Naïve Bayes is a type of supervised-learning module that contains examples of the input-target mapping the model tries to learn. Such models make predictions about new data based on the examination of previous data. The Naïve Byes algorithm uses the mathematics of Bayes' Theorem to make its predictions.

#### Pr[A|B] = Pr[B|A] Pr[A]/Pr[B]

#### D. Decision Tree

Decision trees are the most commonly used tool for predictions and classification of future events. The development of such trees is done in two major steps: building and pruning. During the first phase, the data set is partitioned recursively until most of the records in each partition contain identical value. The second phase then removes some branches which contain noisy data. When evaluating a customer data set the classification is done by traversing throuch the tree until a leaf node is reached. Figure shows a simplified churn prediction decision tree for the Retail Sales and Marketing sector.

#### E. Back-propagation Feed-forward Neural Network

In a feed-forward Neural Network, each neuron in a given layer is connected to all other neurons in the next layer. No connections are allowed in the backward direction. This type of ANN is made up of a series of layes including input layer, optional hidden layer(s) and an output layer. Backpropagation is one of the methods that can be used to train this type of ANN. Training involves providing the NN with initial data to enable it to establish a pattern as it executes this data. Establishing a pattern during execution of the training set is referred to as learning by the Neural Network. Learning involves adjusting the weights and biases of the neurons, these weights are randomly generated and in most cases lies between -1 and 1.

#### F. Min-Max Normalization

Min-max normalization performs a linear transformation on the original data. Suppose that minA and maxA are the minimum and maximum values of an attribute, A. Min-max normalization maps a value,

$$v_{i}^{'} = \frac{v_{i} - min_{A}}{max_{A} - min_{A}} (new_{max_{A}} - new_{min_{A}}) + new_{min_{A}}$$

Min-max normalization preserves the relationship among the original data values. It will encounter an "out-of-bounds" error if a future input case for normalization falls outside of the original data range for A.

## G. Profit-based loss function

In a classification context, the most commonly-used loss functions penalize classification errors in predictions. The targeting profit depends not only on whether the customer is churner or not, but also on a number of other customer-specific parameters that influence her targeting profitability. There could indeed be cases where a customer is a churner (y1=1) but is not associated with a positive targeting profit. The loss associated with targeting a non-profitable customer equals

$$|\pi_i = |-\varphi_i \delta| = \varphi_i \delta$$

When she is a non-churner, and

$$|\pi_i| = \gamma_i (\delta - V_i)$$

When she is a churner.

## F. Creating the Network Customer Classification

The NN structure consists of input layer, hidden layer and an output layer. The input layer has 4 neurons, representing the four factors listed in follows:

1) *Churn-rate*: is a measure of the number of individuals or items movingout of a collective group over a specific period of time. The following equation is used to calculate the churn-rate.

# Churn\_Rate= BasicQuantity/TotalQuantity

2) *Revenu-rate:* is the annualized revenue of a company and it can be calculated as follows:

Revenue\_Rate= BasicRevenue/TotalRevenue

3) *Gross Profit*: is the difference between the purchase price and the costs of bringing to market.

4) *Quantity*: is the amount of products that are ordered by each customer.

There are guidelines to selecting the number of hidden layers and neurons per layer. Selecting a given number of layers depends on the function or decisions the NN is expected to execute/make, but in most situations one hidden layer will suffice. This study wants to achieve, one hidden layer with 7 neurons was used. This was arrived at upon trial and erroe, until the best result was obtained. One output layer with one neuron was used corresponding with either the customer is churner and non-churner. A 1 output implies that the customer is churner and a 0 output imply the customer is non-churner.

## IV. IMPLEMENTATION

In this paper, it is focused on a Retail Sales and Marketing system to predict the churners and non-churners for profit maximization. This paper can suggest the compnay which customers can maximize the profit among the selected customers and how to do this customers for retention action. To predit whether the customer is churn or non-churn, the customer data are collecting from the Retail Sales and Marketing System and are predicting by using Backpropagation algorithm such as revenue, remaining payment, quantity, profits and retention action costs. The process first starts with identifying a customer and maintaining all the corresponding details into the database. In implementation of classifying customer churn for profit maximization system, there are three conceptually separate pards to predict customer churn 1) selecting the target customer 2) preprocessing steps for customer data and 3) predicting the customer churn by using Backpropagation Neural Network Algorithms.

# A. Selecting Target Customer

This paper is used for a retail sales and marketing company to predict the customers for targeted retention programs. The company might set up a retention report with two attributes: how much quantity ordered by the customer and how much revenue to express the profitability of a company per customer as basis.

# B. Preprocessing and Normalizing Data

In this paper, the four neurons are input into the neural network to train the selected customer data. The input neurons are revenue-rate, remaining payment, churn-rate and gross profit. Therefore, it is needed to calculate the churn rate and revenue-rate as the preprocessing step. After preprocessing step, the amount of remaining payment and gross profit are too large and it is rarely handled for training data. To solve the problem, the data are normalized by using min-max normalization for a linear transformation on the original data.

# C. Classification Churner or Non-churner

In the training, the network has only three layers input, hidden and output. The first step is to initialize the network with small ramdom weights and four input pattern are presented as the input layer of the network. Then the forward pass used to determine the activation levels of the neurons. The activation levels of the hidden layer will be calculated. This is accomplished by input values by the weight matrix between the input layer and the hidden layer and then running the results through a sigmoid function.

After a neural network is trained, the value of total error and average error is calculated. And then, the customers are divided into churner and non-churner according to the assumption- it is churner if the error of each customer which is grater than average error, otherwise, it is non-churner. This paper is focused on the customer retention program and estimating the profitable customers amoung the selected customers. Therefore, the estimated loss for each customer need to calculate by using the profit-based loass function. It is reported to the user which customer is the most profitable customer among the non-churner and which customers are needed to give the retention action.

#### V. CONCLUSION

This paper deals with the customer churn analysis and predicting the most profitable customer in retail sales and marketing system. Artificial Neural Network is used in solving the problem domains. It is very useful classification techniques for prediction and it is used for prediction of customer churn in direct marketing system. This paper mainly focused on the customer classification and prediction in Customer Relationship Management concerned with data mining based on Backpropagation technique. This technique is used to average the total error of the network to predict two different types of customers that is churn customers and nonchurn customers. This work can be extended by using Regression analysis and Rule based learning.

#### REFERENCES

- Sharma, D. Panigrahi, & P. Kumar, "A Neural Network Based Approach for Predicting Customer Churn in Cellular Network Services", arXiv preprint arXiv:1309.3945,(2013).
- [2]. A Lemmens, & S.Gupta, "Managing Churn to Maximize Profit", Harvard Business Schol Working Paper, (14-020), (2013).
- [3]. A. Fazlzadeh, M. M. Tabrizi, & K. Mahboobi, "Customer Relationship Management in Small-Medium Enterprises", the case of science and technology parks of Iran, African Journalof Business Management.5(15),6160-6168,(2011).
- [4]. C.Rygielski, J. C. Wang, & D. C. Yen, "Data Mining Technique for Customer Relationship Management", Technology in society, 24(4), 483-502(2002).
- [5]. D. Pyle, "Data Preparation for Data Mining", Morgan Kaufmann Publishers, Los Altos, California, (1999).