International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Customer Churn Prediction using Association Rule Mining

Mie Mie Aung, Thae Thae Han, Su Mon Ko

Information Technology Support and Maintenance Department,
University of Computer Studies, Meiktila, Myanmar

How to cite this paper: Mie Mie Aung | Thae Thae Han | Su Mon Ko "Customer Churn Prediction using Association Rule Mining" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1886-1890, **https://doi.org/10.31142/ijtsrd26818**

IJTSRD26818

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 company, 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. In this paper, we apply the method to the retail sales and marketing company customer churn data set. This paper provides an extended overview of the literature on the use of data mining in customer churn prediction modeling. It will help the retail sales and marketing company to present the targeted customers with the estimated loss of clients or customers for the promotion in direct marketing.

KEYWORDS: *Data Mining, Customer Churn Prediction, Association Rule Mining, FP-Growth*

1. INTRODUCTION

Data 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, we apply the FP-Growth method to the retail sales and marketing company customer churn data set. One of the currently fastest and most popular algorithms for frequent item set mining is the FP-growth algorithm. It is based on a prefix tree representation of the given database of transactions (called an FP-tree), which can save considerable amounts of memory for storing the transactions. 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 Association Rule Mining based technique. FP-Growth is used to find the number of customers churns. Customer churn is the action of the customer who is like to leave the company and it is one of the mounting issues of today's rapidly growing and competitive the retail sales and marketing company. To minimize the customer churn, prediction activity to be an important part of the retail sales and marketing company's vital decision making and strategic planning process.

1.1 Churn Prediction

Today numerous the retail sales and marketing companies are prompt all over the world. The retail sales and marketing company is (facing a severe) loss of revenue due to increasing competition among them and loss of potential customers. Churn is the activity of the retail sales and marketing company is the customers leaving the current company and moving to another company. Many companies are finding the reasons of losing customers by measuring customer loyalty to regain the lost customers. To keep up with the competition and to acquire as many customers, most operators invest a huge amount of revenue to expand their business in the beginning. In the retail sales and marketing company each company provides the customers with huge incentives to attract them to change to their services, it is one of the reasons that customer churn is a big problem in the company nowadays. To prevent this, the company should know the reasons for which the customer decides to move on to another company. The Churns can be classified into two main categories: Involuntary and Voluntary. Involuntary are easier to identify. Involuntary churn is those customers whom the

retail sales and marketing company decides to remove as a subscriber. They are churned for fraud, non-payment and those who don't use the service. Voluntary churn is difficult to determine because it is the decision of the customer to unsubscribe from the service provider. Voluntary churn can further be classified as incidental and deliberate churn . The former occurs without any prior planning by the churn but due to change in the financial condition, location, etc. Most operators are trying to deal with these types of churns mainly.

1.2 Churn Management

Churn management is very important for reducing churns as acquiring a new customer is more expensive than retaining the existing ones. Churn rate is the measurement for the number of customers moving out and in during a specific period of time. If the reason for churning is known, the providers can then improve their services to fulfill the needs of the customers. Churns can be reduced by analyzing the past history of the potential customers systematically. A large amount of information is maintained by the retail sales and marketing company for each of their customers that keep on changing rapidly due to a competitive environment. The information includes the details about billing, calls and network data. The huge availability of information arises the scope of using Data mining techniques in the company's database. The information available can be analyzed in different perspectives to provide various ways to the operators to predict and reduce churning. Only the relevant details are used in the analysis which contributes to the study from the information given. Data mining techniques are used for discovering the interesting patterns within data and it helps to learn to predict whether a customer will churn or not based on customer's data stored in the database.

2. 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 thedecision 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.

3. 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.

3.1 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 classification; 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 attribute 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.

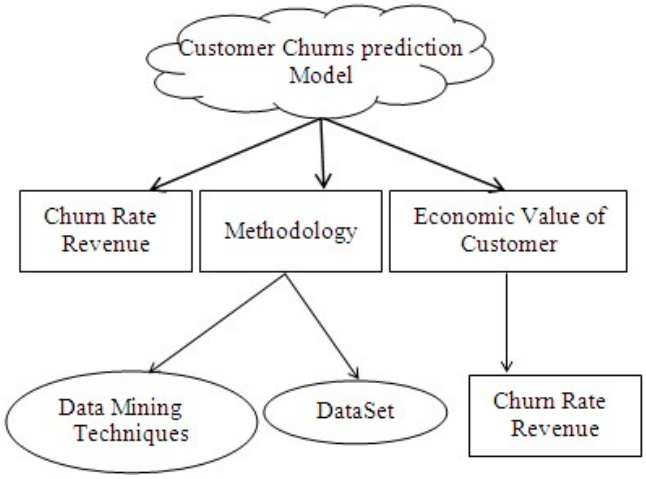

Figure.1 Customer Churn Prediction Model

3.2 Association Rule Mining

Association rule mining, one of the most important and well researched techniques of data mining, was first introduced in. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc. Various association mining techniques and algorithms will be briefly introduced and compared later. Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two subproblems. One is to find those itemsets whose occurrences exceed a predefined threshold in the database; those itemsets are called frequent or large itemsets. The second problem is to generate association rules from those large itemsets with the constraints of minimal confidence. Suppose one of the large itemsets is Lk, Lk = {I1, I2, … , Ik},

association rules with this itemsets are generated in the following way: the first rule is {I1, I2, … , Ik-1}⇒ {Ik}, by checking the confidence this rule can be determined as interesting or not. Then other rule are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty. Since the second subproblem is quite straight forward, most of the researches focus on the first subproblem. The first sub-problem can be further divided into two sub-problems: candidate large itemsets generation process and frequent itemsets generation process. We call those itemsets whose support exceed the support threshold as large or frequent itemsets, those itemsets that are expected or have the hope to be large or frequent are called candidate itemsets.

Association Rule Mining can be viewed as a two-step process:

1. **Find all frequent item sets**
   - Apriori Method
   - FP Growth Method (Frequent Pattern)

2. **Generate strong association rules from the frequent item sets:**
   - By definition, these rules must satisfy minimum support and minimum confidence

3.3 Basic Concepts & Basic Association Rules Algorithms

Let I=I1, I2, … , Im be a set of m distinct attributes, T be transaction that contains a set of items such that T ⊆ I, D be a database with different transaction records Ts. An association rule is an implication in the form of X⇒Y, where X, Y ⊂ I are sets of items called itemsets, and X ∩ Y =∅. X is called antecedent while Y is called consequent, the rule means X implies Y. There are two important basic measures for association rules, support(s) and confidence(c). Since the database is large and users concern about only those frequently purchased items, usually thresholds of support and confidence are predefined by users to drop those rules that are not so interesting or useful. The two thresholds are called minimal support and minimal confidence respectively. Support(s) of an association rule is defined as the percentage/ fraction of records that contain X ∪ Y to the total number of records in the database. Suppose the support of an item is 0.1%, it means only 0.1 percent of the transaction contain purchasing of this item. Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain X ∪ Y to the total number of records that contain X. Confidence is a measure of strength of the association rules, suppose the confidence of the association rule X⇒Y is 80%, it means that 80% of the transactions that contain X also contain Y together. In general, a set of items (such as the antecedent or the consequent of a rule) is called an itemset. The number of items in an itemset is called the length of an itemset. Itemsets of some length k are referred to as k-itemsets. Generally, an association rules mining algorithm contains the following steps:

- The set of candidate k-itemsets is generated by 1-extensions of the large (k -1)-itemsets generated in the previous iteration.
- Supports for the candidate k-itemsets are generated by a pass over the database.
- Itemsets that do not have the minimum support are discarded and the remaining itemsets are called large k-itemsets.

This process is repeated until no more large itemsets are found. The AIS algorithm was the first algorithm proposed for mining association rule. In this algorithm only one item consequent association rules are generated, which means that the consequent of those rules only contain one item, for example we only generate rules like X ∩ Y⇒Z but not those rules as X⇒Y∩ Z. The main drawback of the AIS algorithm is too many candidate itemsets that finally turned out to be small are generated, which requires more space and wastes much effort that turned out to be useless. At the same time this algorithm requires too many passes over the whole database.

Apriori is more efficient during the candidate generation process. Apriori uses pruning techniques to avoid measuring certain itemsets, while guaranteeing completeness. These are the itemsets that the algorithm can prove will not turn out to be large. However there are two bottlenecks of the Apriori algorithm. One is the complex candidate generation process that uses most of the time, space and memory. Another bottleneck is the multiple scan of the database. Based on Apriori algorithm, many new algorithms were designed with some modifications or improvements.

3.4 Frequent Pattern Growth (FP Growth)

Finding frequent item sets without candidate generation

1. First, compress the database representing frequent items into a frequent pattern tree or Data classification is a two-step process. In the first FP tree, which retains the itemset association information. FP-tree is an extended prefix-tree structure storing crucial, quantitative information about frequent patterns. Only frequent length-1 items will have nodes in the tree, and the tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing nodes than less frequently occurring ones. FP-Tree scales much better than Apriori because as the support threshold goes down, the number as well as the length of frequent itemsets increase dramatically. The candidate sets that Apriori must handle become extremely large, and the pattern matching with a lot of candidates by searching through the transactions becomes very expensive. The frequent patterns generation process includes two sub processes: constructing the FT-Tree, and generating frequent patterns from the FP-Tree. The mining result is the same with Apriori series algorithms. To sum up, the efficiency of FP-Tree algorithm account for three reasons. First the FP-Tree is a compressed representation of the original database because only those frequent items are used to construct the tree, other irrelevant information are pruned. Secondly this algorithm only scans the database twice. Thirdly, FP-Tree uses a divide and conquer method that considerably reduced the size of the subsequent conditional FP-Tree.

2. Then devide the compressed database into a set of conditional databases (a special kind of projected database), each associated with one frequent item or “pattern fragment”, mines each such database separately.

No	Variable Name	Description
1	Age, Gender, Occupation	Demographic variables considered
2	The number of purchase	Identifies the number of customer is purchased
3	Frequently used purchase	Identifies the most frequently purchase by the consumer
4	Churn	Identifies whether customer have changed company or not
5	Product innovation	Determines whether product innovation is necessary for sustaining customers
6	Product purchase amount (DpM)	Approximates the amount used to purchase product a month
7	Credit purchase amount (CpM)	Approximates the amount used to purchase call credits a month
8	Tariffs	The type of customer, whether a prepaid or post-paid customer
9	Tenure	Length of time a customer has been with a particular subscriber

Table1: The Variables Used In Dataset for This Research

3.5 FP-growth Algorithm

In this section we examine the FP-growth algorithm over a hypothetical dataset for a sailing company. This example is picked up from the textbook *Data-Mining Concepts and Techniques* (Han & Kamber., 2006). The dataset is a collection of transaction records. Each transaction has a unique ID and each item is represented by an index Ij. The dataset is represented in Table 1. The algorithm starts with the first scan of the database which derives the set of frequent items (1-itemsets) and their support counts (frequencies). Let the minimum support count is 2. The set of frequent items is sorted in the order of descending support count. This resulting set or *list* is denoted as L. Thus, we have:

L = {I2: 7, I1: 6, I3: 6, I4: 2, I5: 2}

TID	List of items Ids
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Table2: Transactional Data for a Sailing Company

An FP-tree is then constructed as follows. First, create the root of the tree, labeled with "null". Scan database *D* a second time. The items in each transaction are processed in *L* order (i.e., sorted according to descending support count), and a branch is created for each transaction.

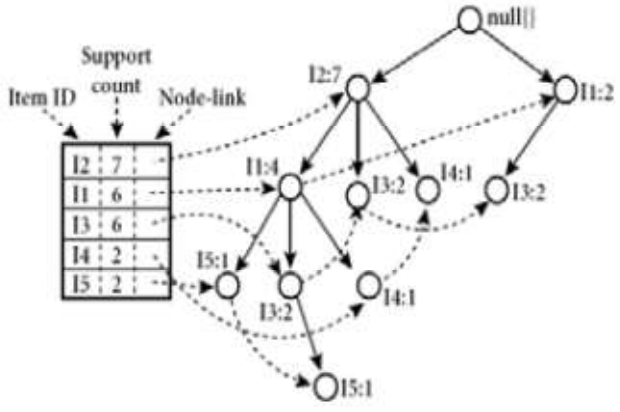

Figure2: An FP-tree registers compressed, frequent pattern information.

The tree obtained after scanning all of the transactions is shown in Figure 1 with the associated node-links. In this way, the problem of mining frequent patterns in databases is transformed to that of mining the FP-tree. The FP-tree is mined as follows: Start from each frequent length-1 pattern (as an initial suffix pattern); construct its conditional pattern base (a "sub database" which consists of the set of *prefix paths* in the FP-tree co-occurring with the suffix pattern), then construct its (*conditional*) FP-tree, and perform mining recursively on such a tree. Mining of the FP-tree is summarized in Table 3.

Item	Conditional Pattern Base	Conditional FP-tree	Frequent Pattern
15	{{I2,I1:1}, {I2,I1,I3:1}}	<I2:2,I1:2>	{I2,I5:2}, {I1,I5:2}, {I2,I1,I5:2}
14	{{I2,I1:1}, {I2:1}}	<I2:2>	{I2,I1:2}
13	{{I2,I1:2}, {I2:2}, {I1:2}}	<I2:4,I1:2>, <I1:2>	{I2,I3:4},{I1,I3:4},{I2,I1,I3:2}
12	{{I2:4}}	<I2:4>	{I2,I1:4}

Table3: Mining the FP-tree by creating conditional (sub-) pattern bases

4. CONCLUSION

This paper deals with the customer churn analysis and predicting the most profitable customer in the retail sales and marketing system. Customer churn is one of the most important metrics for a growing business to evaluate. As churn management is a major task for companies to retain valuable customers, the ability to predict customer churn is necessary. This paper mainly focused on the customer classification and prediction in Customer Relationship Management concerned with data mining based on FP Growth technique. This technique is used to finding frequent item sets without candidate generation.

References

[1] A Lemmens, & S. Gupta, "Managing Churn to Maximize Profit", Harvard Business Schol Working Paper, (14-020), (2013).

[2] A. Fazlzadeh, M. M. Tabrizi, & K. Mahboobi, "Customer Relationship Management in Small-Medium Enterprises", the case of science and technology parks of Iran, African Journal of Business Management.5(15),6160-6168,(2011).

[3] C. Rygielski, J. C. Wang, & D. C. Yen, "Data Mining Technique for Customer Relationship Management", Technology in society, 24(4), 483-502(2002).

[4] D. Pyle, "Data Preparation for Data Mining", Morgan Kaufmann Publishers, Los Altos, California, (1999).

[5] Sharma, D. Panigrahi, & P. Kumar, "A Neural Network Based Approach for Predicting Customer Churn in Cellular Network Services", arXiv preprint arXiv: 1309.3945,(2013).

[6] Jiawei Han, Jian Pei, Yiwen Yin: Mining Frequent Patterns without Candidate Generation in Proceedings of the 2000 ACM SIGMOD international Conference on Management of Data (Dallas, Texas, United States, May 15-18, 2000). SIGMOD'00. ACM Press, New York, NY, 1-12.

[7] V. Umayaparvathi and K. Iyakutti, "A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics," International Research Journal of Engineering and Technology (IRJET), vol. 03, no. 04, April 2016.