PRIORITIZING THE CUSTOMERS USING DATA MINING TECHNIQUES FOR AN EFFICIENT CRM

VIT University, Vellore.
Email: nithya.s@vit.ac.in

Received on: 18.10.2016
Accepted on: 11.11.2016

Abstract

Companies and their way of interaction with customers has been changed a lot over past years. A customers containing business cannot be guaranteed for so long time. So, companies found a key point that need of understanding customers is playing a vital role. So, they are approaching data mining as a platform to achieve the CRM (Customer Relationship Management). Different fields are following different techniques like association rule mining, in-depth mining, frequent pattern mining. By this paper we are going to plant an enhanced version of Apriori algorithm of Association rule mining.

Keywords: CRM (Customer Relationship Management), Data Mining, Apriori, Association Rule Mining.

1. Introduction

Data Mining is a process of extraction of useful knowledge from large database. It allow the data exploration, data analysis and data visualization of large databases. It is a method which uses different methods and techniques to extract patterns and relationships in data that is used to make predictions. It gives you a path to select customers on whom you should focus and consider to pay some additional interest. In return you get an excellent revenue because of a greatly improved ability to each individual in their best way.

Fig1 Data mining[1].
II. Related Work

1. **Data Mining Task**: Data mining deals with mining of the patterns. On the basis of the type of data that is to be mined, because of this reason may be it is known that “data mining is misnomer”. The word misnomer means inaccurate designation. This is called so because, we call it as knowledge mining also due to the extraction of knowledge from database. So, we can call this with any name because it extracts different kind of knowledge. There are two categories of data mining task, they are mentioned below. [2]

![Fig2 Data mining task](image)

**A. Predictive Mining:**

The purpose it’s going to serve is to predict the form of future outcome whether it is numeric value or is categorized. These models are supervised learning functions which predict the target. Predictive methods are being used in next phase of empirical study. Some variables are used to predict the unknown values of other variables. Even comparison is placed between these supervised approaches to get details about the strength and weakness of each approach. One of the goals of the study is to check whether these methods are suitable or not for the knowledge.

1. **Classification:**

Comparing all the predictive data mining techniques we can say that the classification is much easier to understand. The features are same as the supervised learning and assigns new value to the well-defined classes. These techniques are being used in different applications like modelling businesses, credit analysis and many more.
2. Regression:
Regression is another predictive data mining model which comes under supervised learning. It is used to analyze the dependency of attributes that depend on the other attribute values which are present in same item. The attributes of new cases are predicted by developing this model. Regression deals with numerical target attributes and classification deals with categorical target attributes, this is the difference between them.

3. Time-series Analysis:
It is an application with one or more time-dependent attributes which predict the numeric outcomes like future price of individual stock. A collection of values obtained from sequential measurements are represented by this time-series. The ability of visualizing the shape of data is a natural feature of the time-series data mining. The subsequent values are within predictable ranges of one another. These time-series are very long and considered smooth. It have many applications like stock market analysis, economic forecasting and more. Many other applications like observations of natural phenomena like atmosphere, temperature, wind and earthquake. [11]

B. Descriptive mining:
The tasks like correlation generation, frequency, etc. are accomplished using these descriptive mining techniques. The main task is to discover regularities of data and patterns to be uncovered. The additional use is to find interesting subgroups out of bulk data.

a. Clustering: It is considered as very important technique in data mining techniques. The process of clustering can be considered as partitioning data into a set of classes called as groups. It has ability to find natural groups from database called clusters. There will be similar characteristics among the data items in a group. It is mainly used to group the similar items. It identify set of objects with unknown class.

b. Summarization: It can be defined as mapping data with simple descriptions into subsets. The general overview of summarized data set is given by aggregate information. Different levels of abstraction and view of different angles can be projected by summarization. The technique to find compact description of dataset is accomplished by summarization.

c. Association: the relationship between attributes and items are defined by associations. One pattern presence implies the other i.e., item is related to cause and effect of another. Statistical relationships among different interdependent variables of a model are commonly established. It is an important technique for market basket analysis to derive the
possible combinations of interesting product group. These associations uncover the relationships between unrelated data
in a transaction database or other repository. [11]

1. Association Rule Mining:
The discovery of interesting patterns from huge database is known as association rule mining. By using the
interestingness of measures strong rules are found. The correlations among items are also found in transactional database
or relational database. Analyzing data for frequent pattern is recreated by using the support and confidence. Here,
support holds the intimacy of the item in the database and confidence defines the count of appearance. There are various
algorithms being used in this.

2. Frequent item set mining:
It is being an important part of data mining for several years. There is a distinctive progress and numerous efficient
algorithms has been developed to reveal the patterns in huge database in this field. This frequent pattern mining is being
used in many applications in real world. They can be used in any kind of business for market basket analysis.

Fig3 Association rule algorithms [2]

If X and Y are two disjoint sets then the implication of X->Y can be known as association rules. Where, X is called the
antecedent and Y is called the consequent. The prediction of correlation between the data items [1, 2] is done by the
association rules. Support, Confidence, Lift are given below which are the measures of the association rule mining. [4]

- **Support:** Support is an indication of how frequently the item-set appears in the database.
  
  The support value of X is defined by sup (X).

- **Confidence:** Confidence is an indication of how often the rules has been found to be true.
Confidence can be well-defined as
\[ \text{Con}(X \Rightarrow Y) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)}. \]

- **Lift**: the lift of a rule is calculated as
\[ \text{Lift}(X \Rightarrow Y) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)} \times \text{sup}(Y). \]

As we seen there are many algorithms to generate frequent item set. Let’s take a view on Apriori algorithm.

**II. Literature Survey**

In [1] they approached the association rule mining to resolve a problem in number of edges in lattice. They proposed a solution by using non redundant rules and same algorithm in efficient way.

In [2] they approached Multi-level association rule mining to resolve the problem of quality and quantity. They made it by proposing a solution using Multi-minimum support.

In [3] they approached the apriori algorithm to resolve the problem with delay in cross selling analysis. They solved it by deleting the invalid business.

In [4] they approached apriori algorithm to solve the problem in complexity level of MS-apriori. They made an enhancement with basic rules and proposed MSB-apriori algorithm.

In [5] they approached Negative association rule mining to solve the problem of deficiency of quality. They came up with item-set called FAP (frequent absence and presence).

In [6] they approached In-depth mining to solve the problem with lack of customized marketing. They solved it by applying data mining techniques to CRM customer value.

In [7] they approached data mining to deal the problem with interaction of customers. They proposed a solution of mining techniques for anticipation.

In [8] they have approached the frequent pattern mining to solve the defects in existing approach. They proposed an approach which is different from the existing approaches of finding the patterns from uncertain data.

In [9] they have approached data mining to solve the problem with hidden facts. So, they discovered the interesting facts from circulation data.
In [10] they have approached frequent pattern mining to handle the problem with data extraction. They resolved it by identifying predictive patterns in existing data.

To accomplish the task we can approach any kind of technology in data mining but we approached the association rule mining because, we are going to generate the item set of the customers. Even going with association rule mining we have many algorithms to make use of. We have selected the apriori algorithm because of its efficiency in working. Let me frame some properties or characteristics of the apriori, the apriori algorithm is based on the bottom-up approach and a property called downward closure which follows a rule, any superset of the infrequent item set remain as infrequent.

**III. Apriori Algorithm**

Apriority is used on databases containing transactions of different kind. The data like transactional data is viewed as an item set. Assume C as a threshold, the subsets of database which are called as item sets and those item sets which are of subsets of at least C transactions are found using this algorithm. The working of this algorithm is based on the bottom-up approach. It follows anti-monotone rule, if an item-set is frequent then all the subsets of that item set should also be frequent. The process will be like there will be candidate set generated from given data and an item set is generated from that candidate set and it prolong up to k number of sets are generated. Even the drawback is the same thing that the data should be scanned every time of candidate set generation. But finally the frequent pattern set is generated by omitting the less occurred items, which is done using support and confidence.

```plaintext
Apriori(T, C)
L_1 ← {large 1 − itemsets}
k ← 2
while L_{k-1} ≠ ∅
    C_k ← \{a ∪ \{b\} \mid a ∈ L_{k-1} \land b ∉ a\} − \{c \mid \{s \mid s ⊆ c \land |s| = k-1\} ∉ L_{k-1}\}
    for transactions t ∈ T
        C_t ← \{c \mid c ∈ C_k \land c ⊆ \{t\}\}
        for candidates c ∈ C_t
            count[c] ← count[c] + 1
        L_k ← \{c \mid c ∈ C_k \land count[c] ≥ \epsilon\}
        k ← k + 1
    return ∪_k\ L_k
```

**Fig4: pseudo code of apriori algorithm.**
In this paper first we made a brief discussion about the Data Mining. The main goal of the paper is to discuss about the usage of data mining techniques in CRM (Customer Relationship Management). Previously there are many data mining implementations in business oriented organizations.

IV. Existing Approach:

Apriori usually process under a horizontal layout for representing the transaction database and the frequency of an item set computing by its occurrence in every transaction. The main drawback is the database will be scanned for many times. Even though it is easy to implement the apriori algorithm. Every time scanning result generation of candidate set. It needs a huge memory to store the data. The execution time is more. The other algorithms have their own features, they are mentioned in simple way in table.

Table1. Comparison of algorithms [2]

<table>
<thead>
<tr>
<th>CATEGORIES</th>
<th>Techniques</th>
<th>Database scan</th>
<th>Drawback</th>
<th>Advantages</th>
<th>Data format</th>
<th>Storage structure</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALGORITHMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APRIORI</td>
<td>Uses breadth first search</td>
<td></td>
<td>Candidate set is generated every time</td>
<td>Large memory needed</td>
<td>Easy to implement</td>
<td>Horizontal Array</td>
<td>More execution time</td>
</tr>
<tr>
<td>RARM</td>
<td>Depth first search</td>
<td></td>
<td>Scan for few times</td>
<td>Usage in interactive systems is difficult</td>
<td>No candidate set is generated</td>
<td>Horizontal Tree</td>
<td>Less</td>
</tr>
<tr>
<td>ECLAT</td>
<td>Depth first search &amp; interaction of t-id.</td>
<td></td>
<td>Scan for few times</td>
<td>It requires the virtual memory</td>
<td>No need to scan data each time</td>
<td>Vertical Array</td>
<td>Less</td>
</tr>
<tr>
<td>FP-GROWTH</td>
<td>Divide and conquer</td>
<td>Two times</td>
<td>Expensive to build</td>
<td>Database is scanned only two times</td>
<td>Horizontal Tree</td>
<td>Less</td>
<td></td>
</tr>
<tr>
<td>ASPMS</td>
<td>BSM(branch sort method)</td>
<td>One time</td>
<td>-----</td>
<td>Requires less memory</td>
<td>Horizontal Tree</td>
<td>Less</td>
<td></td>
</tr>
</tbody>
</table>
The technique like association rule mining, pattern mining, negative association rule mining, in depth mining are used in different fields for different goals. Let me take an example of one of the applications of data mining in business industries. In a company called SHANHUA carpet company used apriori algorithm and its deployment for their group cross selling analysis, this algorithm helps in deleting lots of invalid business, reduce the records for the following scanning, which raises the efficiency of data mining, scale of database is decreased and scanning time is saved. [3]

In the same way the online shopping websites are using these techniques to assist their customers in their purchase. The recommended items are generated through frequent patterns and displayed to make the customer purchase as simple as possible.

These are all some of the applications of data mining in business. We are going to filter certain customers and maintain the relationship between the customer and vendor. The customers who maintain intimacy with the company and a good gross will remain as preferable by company. So, by using apriori algorithm we are going to generate the item set containing the customers with this features. These particular set of customers are offered special care from the company.

V. Proposed Work

The usage of data mining techniques in online shopping is applied a far before. The existing online Websites use the data mining techniques to recommend the products for the customers, which are similar to the previous purchase or the product they are interested. By this paper we are going to propose a new ideology to prioritize the customers through data mining techniques. Here we are going to separate the customers through apriori algorithm with an enhancement of gross and intimacy. In brief we can say as selection of customers by means of frequency of purchase and total amount of purchase till date.

VI. Conclusion

Customer Relationship Management is serving a very good role in market. It became must for any kind of business to maintain good relation with their customers. When you take care of consumer needs then you will make a leading position in market.

To accomplish this task there is a need of some platform and data mining is a smart choice. The data mining techniques serve various benefits and one must know how to utilize that content. The better way you use the data mining the better results you get.
References


4. Xiangjun Dong, Tiantian Xu, “Mining frequent patterns with multiple minimum supports using basic apriori” 957-961, 2013, IEEE Conference Publications.


