

Customer Relationship Management Classification by Hybridizing Genetic Algorithm and Fuzzy K-Nearest Neighbor - A Review

Jashandeep Kaur
Punjabi University Regional Centre for IT and
Management Mohali,
160062(India)

Rekha Bhatia
Punjabi University Regional Centre for IT and
Management Mohali,
160062(India)

ABSTRACT

Data mining is the process of extraction of information from various datasets on the basis of different attributes. Mining has to be done to extract hidden relationship between various database entities. In the customer relationship management, a choice of relational attributes are available in the dataset. In this paper various techniques are used for the customer relationship.

Keywords

Data Mining, Customer Relationship Management, SVM, Naïve bayes, GA, Fuzzy KNN, J48.

1. INTRODUCTION

Customer relationship management one of centre point for many industries such as banking, retail, telecommunication and insurance. CRM takes customer as the focal point and optimizes the business process. There are major challenges for achieving high performance CRM classification models for the real-world application. Since data quality is a significant obstacle for CRM classifications in that different types of data anomaly which complicate the data preparation and classification methods [1]. It is difficult to find one methodology that rectifies all data mining problems in the CRM data set such as Heterogeneous, High dimensional, Severe data anomaly and Imbalanced.



Fig 1.1: CRM

Generally, the data set will not contains all the data because of missing data by reluctant clients who do not furnish all information misconception and human errors. High dimensional data may contain large number of redundant and irrelevant information that might influence the performance of learning algorithms. Therefore, feature selection becomes very important for machine learning tasks[1].Heterogeneous data is gathered from several number of sources mainly unknown and unlimited, and in many different formats either numeric or nominal. Data Anomaly is the logical or structural in database. For example, if we have a several instances of

same information in a table it may appear that the data is not consistent after operations like update. This may result in the loss of data integrity. Imbalanced classification is known to result in biased models that ignore instances belonging to the minority class. For designing the algorithms of data mining there are various challenges are as follows:

1. Huge datasets and high dimensionality.
2. Determining the statistical significance.
3. Understanding of patterns.
4. Non-standard incomplete data and data integration
5. Mixed changing and redundant data.

The rest of this paper is organized as follows. First, Section 2 reviews related work. Section 3 describes research gaps. Section 4 describes approaches used. Section 5 describes challenges & issues in data mining. Finally, Section 6 concludes the paper.

2. RELATED WORK

In this paper different data mining techniques has been proposed for improving the data quality and to remove the redundancy of CRM data set and feature relevancy after pre-processing. In this a new feature selection method is proposed [1]. This paper enhances the performance of the associative algorithms or their quality in terms of rules [2]. Medical data mining has been utilized in the health care industry for the prediction of diseases. From the distinct data sets the hidden patterns has been analyzed and these patterns has been used for taking the appropriate decision so that effective medicine is giving to the patients. In this paper two major Data Mining techniques FP-Growth and A-priori have been proposed. These techniques have been used in the diabetes dataset and association rules are being developed by both of these algorithms [3]. For the rough set as fuzzy sets data mining is achieved at granual level. During a knowledge transformation the knowledge properties should be kept unchanged. Thus, a data mining process is not only mining knowledge from data but also from human. This is the key concept of Domain-oriented Data-driven Data Mining (3DM) [4]. The proposed framework maintains multiple populations of membership functions in which one population for one item's membership functions. For making the best set of membership functions collected from all the populations is used to effectively mine fuzzy association rules [5]. Author thus proposes a GA-based approach to speed up the partition process. A new encoding representation and a transformation scheme are constructed to help the search process. Experimental results also show that the algorithm can get a proper partition with good efficiency [6]. Now a days, hairdressing services have been generally

used by people due to the living standards and the sustainable changing patterns of people's lives. In this paper data mining adopts techniques through combination of K-means methods and self-organizing maps (SOM) to apply in RFM recency, frequency, and monetary) model for a hair salon in Taiwan to partition customers and establishing marketing strategies [7].

In this customer relationship management (CRM) systems support customer knowledge creation processes including internalization, socialization, combination and externalization. The types of CRM systems are collaborative, operational and analytical. Analytical systems strongly support the combination process. Collaborative systems indulge the big support to externalization. The customers facilitate socialization through operational systems. Collaborative and analytical systems support the internalization process through providing learning opportunities [12].

3. RESEARCH GAP

In this paper different authors proposed different methods and gives an idea to overcome the problems. Data mining is the process of extraction of information from various datasets on the basis of different attributes. Mining has to be done to extract hidden relationship between various database entities. On the basis of these entities different types of decisions are taken for the extraction of different relationships. In the customer relationship management, different relational attributes are available in the dataset. This dataset have information about the relations of the customer with an enterprise. Classification of data set has to be done with rules for extraction of information. Mainly Churn, appetency, up selling and score are the main entities which will be supposed in the proposed work. On the basis of these values features have been selected from database. Naïve bayes, SVM, J48 has been used for the classification of database. These algorithms do not provide better classification of data due to incompatibility with dataset. To overcome the problems of CRM database a new hybrid algorithm is introduced which will be the combination of GA and Fuzzy KNN classification.

4. APPROCHES USED

4.1 Genetic Algorithm

The genetic algorithms are motivated by Darwin's theory about development. A Solution to a problem solved by genetic algorithms is evolved. This algorithm is started with a set of solutions(represented by chromosomes)called population. These solutions from one population are taken and used to form a new population. This is enthused by a expect that the new population will be enhanced than the old one. According to their fitness the solutions which are selected to form new solutions (offspring), the more suitable they are the more chances they have to reproduce [8].

4.2 Fuzzy Knn (K Nearest Neighbor)

The FUZZY KNN algorithm utilizes power of check sample into any class called fuzzy class membership. The fuzzy classification rule has been produced by the membership of the fuzzy class [10].

4.3 K-Nearest Neighbor

It is a non-parametric method used for classification and deterioration.

- KNN is the category of instance based learning. In this classification and regression is used to appoint weight to the contribution of the neighbors which is close to neighbours that will contribute more to the average as compare to the more distant ones.

4.4 Fuzzy k-Nearest Neighbor Algorithm (FKNN)

This is the K-Nearest Neighbor algorithm is one of the oldest & simplest non parametric pattern classification methods. In the KNN algorithm a class is appointed according to the most common class amongst its k nearest neighbors. According to this approach rather than individual classes as in KNN the fuzzy memberships of samples are assigned to different categories according to the following formulation.

5. CHALLENGES & ISSUES

For better customer relationship management there are various challenges they are:

- *Non-trivial results almost always need a combination of DM techniques:*

Data Mining techniques played a crucial role in the CRM in order to exploring the CRM data from the different angles and analyzing them from various aspects[13]. Every application needs different types of DM techniques and their application have different "slices" of data in an interactive and continual fashion. After this it needs to use several DM operators and then combine (chain) them into a single "exploration plan"[13].

- *Need of data integration before data mining:*

Data comes from the different resources therefore data is unified before the analyzing of data begins.

- *Discrete data types are often encountered, which needed for the integrated mining of diverse and heterogeneous data:*

Customer data comes in the form of structured report of various data types (e.g. demographic data and chat-room data), temporal data (e.g. weblogs),sometimes audio (e.g. recorded phone conversations with customers),text (e.g email), consumer reviews, blogs.

- *CRM deal with noisy data:*

In CRM, due to the webcrawlers the CRM dataset has large number of noise.For giving the better decisions to customer the data has to be cleaned.

- *Real-world validation of results is necessary to acceptance:*

In CRM, as in many DM applications the explored patterns are usually behaved like as hypotheses that has to be tested for new data using rigorous statistical tests for the actual acceptance of the results. This is even more so for taking or recommending actions, especially in such high-risk applications such as in the financial and medical field. Example: recommending investments to customers (it is actually illegal in the US to let software give investment advice).

6. CONCLUSION

Data mining is the process of extraction of information from various datasets on the basis of different attributes. Mining has to be done to extract hidden relationship between various database entities. On the basis of these entities various types of decisions are taken for the extraction of different relationships. Relational attributes are available in the dataset in the CRM. The relations of the customer with an enterprise are available in the data set. For the extraction of information there will be various rules have been used for the classification of data in the data set. The features have been

selected from database on the basis of values of entities. SVM, Naïve bayes, J48 has been used for the classification of database. These algorithms do not provide better classification of data due to incompatibility with dataset.

To overcome the problems of CRM database a new hybrid algorithm is introduced which will be the combination of GA and Fuzzy KNN classification.

7. REFERENCES

- [1] S.Ummugulthum Natchiar “Customer Relationship Management Classification Using Data Mining Techniques”, International Conference on Science, Engineering and Management Research, 2014, pp 223-234.
- [2] Nedaabdelhamid, Aladdin Ayesh and FadiThabtah “Emerging trends in associative classification data mining” International journal of electronics and electrical engineering Feb 2015.
- [3] Sankaranarayanan, S. “Diabetic Prognosis through Data Mining Methods and Techniques”, International Conf. on Intelligent Computing Applications (ICICA), 2014, pp. 162 – 166.
- [4] Wang, Guoyin “Granular computing based data mining in the views of rough set and fuzzy set” IEEE Conf. on Granular Computing, 2008, pp. 67.
- [5] Tzung-Pei Hong “Using divide-and-conquer GA strategy in fuzzy data mining” IEEE Conf. on Computers and Communications, 2004, pp. 116 - 121 Vol.1.
- [6] Tzung-Pei Hong “GA-based item partition for data mining” IEEE Conf. on Systems, Man, and Cybernetics (SMC), 2011, pp. 2238 – 2242.
- [7] Jo-Ting Wei “Customer relationship management in the hairdressing industry: An application of data mining techniques”, IEEE Conf. on Expert Systems with Applications, 2013, pp Pages 7513–7518.
- [8] Wen-Yu Chiang “Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan”, Conf. on Data Mining, 2014, pp 89-97.
- [9] Alexander Tuzhilin “Customer relationship management and Web mining: the next frontier”, Springer conf. on CRM & WM, 2012, pp 584-612.
- [10] Siavash Emtiyaz “Customers Behavior Modeling by Semi-Supervised Learning in Customer Relationship Management”, Advances in information Sciences and Service Sciences (AISS), 2011, PP 56-67.
- [11] Shu-Hsien Liao “Data mining techniques and applications – A decade review from 2000 to 2011”, Expert Systems with Applications, 2012, PP 11303–11311.
- [12] Farnoosh Khodakarami “Exploring the role of customer relationship management (CRM) systems in customer knowledge creation”, Conf. on CRM, 2014, PP 56-70.
- [13] Md. Rashid Farooqi “A Comprehensive Study of CRM through Data Mining Techniques” Proceedings of the National Conference; NCCIST-2011, pp.6-11.