

A Study on Various Approaches of Data Mining classification Techniques

Deep Kumar

Software Engineer, Igniva Solutions Private Limited, Mohali, Punjab, 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. 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 contains the information about the relations of the customer with an enterprise. The dataset has to be classified using rules for extraction of information. Mainly Churn, appetency, up selling and score are the major entities which will be considered in the proposed work. To overcome the problems of CRM database a new hybrid algorithm is introduced which will be the combination of GA and Fuzzy KNN classification.

Keywords : Data Mining, Fuzzy KNN, SVM, CRM, Classification.

I. INTRODUCTION

1.1 Data Mining

Data mining is crucial for extracting and identifying useful information from a large amount of data that is why retailing companies operate purchase databases in a long way, such that all transactions are stored in arranged order. A record-of-transaction database typically contains the transaction date and the products bought in the course of a given transaction.

1.2 Architecture for Data Mining

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining.

The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as promotional campaign management, fraud detection, new product rollout, and so on.

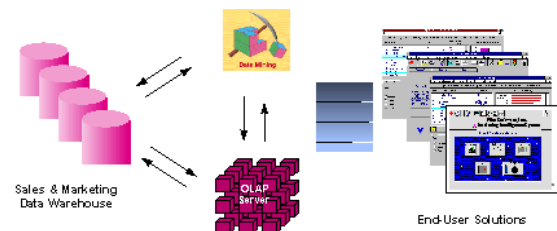


Figure 1. Integrated Data Mining Architecture

The ideal starting point is a data warehouse containing a combination of internal data tracking all customer contact coupled with external market data about competitor activity. Background information on potential customers also provides an excellent basis for prospecting. This warehouse can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and

should be optimized for flexible and fast data access. An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user business model to be applied when navigating the data warehouse. The multidimensional structures allow the user to analyze the data as they want to view their business – summarizing by product line, region, and other key perspectives of their business. The Data Mining Server must be integrated with the data warehouse and the OLAP server to embed ROI-focused business analysis directly into this infrastructure.

1.3 Customer Relationship Management (CRM): CRM is the core business strategy that integrates internal processes and functions of the organization, to create and deliver value to targeted customers at a profit. It is mainly grounded on high quality customer related data and enabled by information technology. CRM is an information industry term that helps an enterprise to manage customer relationships in an organized way and helping the company to provide better services to its customers. CRM is the process of managing all aspects of interaction of the company with its existing customers and the new customers, including prospecting, sales and service. CRM applications try to provide insight into and help in improving the company/customer relationship by combining all these views of customer interaction into one picture.

1.4 Types of CRM

1.4.1 Operational CRM: The primary goal of CRM systems is to integrate and automate sales, marketing, and customer support. Therefore, these systems typically have a dashboard that gives an overall view of the three functions on a single page for each customer that a company may have. The dashboard may provide client information, past sales, previous marketing efforts, and more, summarizing all of the relationships between the customer and the firm. Operational CRM is made up of 3 main components: sales force automation, marketing automation, and service automation.

1.4.2 Analytical CRM: The role of analytical CRM systems is to analyze customer data collected through multiple sources, and present it so that business managers can make more informed decisions. Analytical CRM systems use techniques such as data mining, correlation, and pattern recognition to analyze the customer data. These analytics help improve customer service by finding small problems which can be solved, perhaps, by marketing to different parts of a consumer audience differently. For example, through the analysis of a customer base's buying behavior, a company might see that this customer base has not been buying a lot of products recently. After scanning through this data, the company might think to market to this subset of consumers differently, in order to best communicate how this company's products might benefit this group specifically.

II. REVIEW OF LITERATURE

Jo-Ting Wei et al [1] “Customer relationship management in the hairdressing industry: An application of data mining techniques” explained With the increase of living standards and the sustainable changing patterns of people’s lives, nowadays, hairdressing services have been widely used by people. This paper adopts data mining techniques by combining self-organizing maps (SOM) and K-means methods to apply in RFM (recency, frequency, and monetary) model for a hair salon in Taiwan to segment customers and develop marketing strategies. The data mining techniques help identify four types of customers in this case, including loyal customers, potential customers, new customers and lost customers and develop unique marketing strategies for the four types of customers.

Wen-Yu Chiang et al [2] “Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan”, explained the purpose of this research is to mine high value family travelers for CRM system for airline & travelers agencies. The market of online air travel

shopper is divided into six markets. The study also applies the C5-0 decision tree algorithm on the discovered rank market.

Alexander Tuzhilin et al [3] “Customer relationship management and Web mining: the next frontier” After a decade of successful development of new Web mining technologies, it is a good time to examine novel promising areas that will advance Web mining over the next decade. This paper argues that CRM is such an area that can benefit from and contribute to further advancements of the Web mining research. This is the case because CRM is an underexplored field that has many open and interesting problems important to the industry and academia. This paper reviews some of the key aspects of CRM, describes certain problems and promising research directions in the field, and discusses how Web mining can contribute to solving these problems.

Siavash Emtiyaz et al [4] “Customers Behavior Modeling by Semi-Supervised Learning in Customer Relationship Management” Leveraging the power of increasing amounts of data to analyze customer base for attracting and retaining the most valuable customers is a major problem facing companies in this information age. Data mining technologies extract hidden information and knowledge from large data stored in databases or data warehouses, thereby supporting the corporate decision making process. CRM uses data mining (one of the elements of CRM) techniques to interact with customers. This study investigates the use of a technique, semi-supervised learning, for the management and analysis of customer-related data warehouse and information. The idea of semi-supervised learning is to learn not only from the labeled training data, but to exploit also the structural information in additionally available unlabeled data. The proposed semi-supervised method is a model by means of a feed-forward neural network trained by a back propagation algorithm (multi-layer perceptron) in order to predict the category of an unknown

customer (potential customers). In addition, this technique can be used with Rapid Miner tools for both labeled and unlabeled data.

Pareesh Tanna et al [5] “Using Apriori with WEKA for Frequent Pattern Mining” Knowledge exploration from the large set of data generated as a result of the various data processing activities due to data mining only. Frequent Pattern Mining is a very important undertaking in data mining. Apriori approach applied to generate frequent item set generally espouse candidate generation and pruning techniques for the satisfaction of the desired objective. This paper shows how the different approaches achieve the objective of frequent mining along with the complexities required to perform the job. This paper demonstrates the use of WEKA tool for association rule mining using Apriori algorithm.

Shrey Bavisi et al [6] “A Comparative Study of Different Data Mining Algorithms” Data Mining is used extensively in many sectors today, viz., business, health, security, informatics etc. The successful application of data mining algorithms can be seen in marketing, retail, and other sectors of the industry. The aim of this paper is to present the readers with the various data mining algorithms which have wide applications. This paper focuses on four data mining algorithms K-NN, Naïve Bayes Classifier, Decision tree and C4.5. An attempt has been made to do a comparative study on these four algorithms on the basis of theory, its advantages and disadvantages, and its applications. After studying all these algorithms in detail, we came to a conclusion that the accuracy of these techniques depend on various characteristics such as: type of problem, dataset and performance matrix.

III. APPROACHES USED

K-nearest neighbor classifiers: K-nearest neighbor classifiers ([347-352]) represent a totally different approach to classification. They do not build any explicit global model, but approximate it only locally

and implicitly. The main idea is to classify a new object by examining the class values of the K most similar data points. The selected class can be either the most common 21 class among the neighbors or a class distribution in the neighborhood. The only learning task in K-nearest neighbor classifiers is to select two important parameters: the number of neighbors K and distance metric d. An appropriate K value can be selected by trying different values and validating the results in a separate test set. When data sets are small, a good strategy is to use leave-one-out cross-validation. If K is fixed, then the size of the neighborhood varies. In sparse areas the nearest neighbors are more remote than in dense areas. **Support vector machines:** Support vector machines (SVMs) (Vapid, 1998) are an ideal method, when the class boundaries are non-linear but here is too little data to learn complex non-linear models.. The main advantage of SVMs is that they find always the global optimum, because there are no local optima in maximizing the margin. Another benefit is that the accuracy does not depend on the dimensionality of data and the system is very robust to over fitting. This is an important advantage, when the class boundary is non-linear. Most other classification paradigms produce too complex models for non-linear boundaries. However, SVMs have the same restriction as neural networks: the data should be continuous numerical (or quantized); the model is not easily interpreted, and selecting the appropriate parameters (especially the kernel function) can be difficult. Outliers can cause problems, because they are used to define the class borders. Usually, the problem is avoided by soft margins.

Linear regression: Linear regression is actually not a classification method, but it works well, when all attributes are numeric. For example, passing a course depends on the student's points, and the points can be predicted by linear regression. In linear regression, it is assumed that the target attribute (e.g. total points) is a linear function of other, mutually independent attributes.

IV. CONCLUSION

Data mining is the process for extraction of different data values on the basis of features. In this process different approaches have been used for classification on the basis of features. The different attributes have been used for classification on the basis distance between dataset attributes of training data and testing data. In this process these approaches measure distance from the test samples to the each training samples and on the basis of the distance these test samples have been labels with a single class label. In the purposed work the fuzzy based membership function has been used for classification this approach assigns the weight age fir dataset attributes on the basis of fuzzy membership rules. After the assignment of weight age the data set distance has been computed using Euclidian classifier and the function has been used that use both weight age and distance factor for prediction of class label to a single dataset.

V. REFERENCES

- [1]. 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.
- [2]. 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.
- [3]. Alexander Tuzhilin "Customer relationship management and Web mining: the next frontier", Springer conf. on CRM & WM, 2012, pp 584-612.
- [4]. 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.

- [5]. Shu-Hsien Liao "Data mining techniques and applications-A decade review from 2000 to 2011", *Expert Systems with Applications*, 2012, PP 11303–11311.
- [6]. Farnoosh Khodakarami "Exploring the role of customer relationship management (CRM) systems in customer knowledge creation", *Conf. on CRM*, 2014, PP 56-70.
- [7]. Paresh Tanna "Using Apriori with WEKA for Frequent Pattern Mining", *International Journal of Engineering Trends and Technology (IJETT)*, 2014, pp. 127-131.
- [8]. Shrey Bavisi "A Comparative Study of Different Data Mining Algorithms", *International Journal of Current Engineering and Technology*, 2015, pp. 3248-3252.
- [9]. Manjari Anand "Customer Relationship Management using Adaptive Resonance Theory", *International Journal of Computer Applications*, 2013, pp. 43-47.
- [10]. Ms. Saranya, "Decision Support System for CRM in Online Shopping System", *International Journal of Advances in Computer Science and Technology*, 3(2), February 2014, 148, 2014, pp. 148-151.
- [11]. Kamal R. "Adaptive Pointing Theory (APT) Artificial Neural Network", *International Journal of Computer and Communication Engineering*, 2014, pp. 212-215.
- [12]. Mohammed Al-Maolegi "An Improved Apriori Algorithm For Association Rules", *International Journal on Natural Language Computing (IJNLC)*, 2014, pp. 21-29.