International Journal of Scientific Research & Engineering Trends
Volume 3, Issue 4, July-2017, ISSN (Online): 2395-566X

Efficient and Scalable Multiple Class Classification: A Review

Tarun Yadav
Email: yadavec@gmail.com

***Abstract* – Data mining a nontrivial extraction of the novel, implicitly and actionable knowledge from large data sets is an evolving technology, which is a direct result of the increasing use of computer databases for the purpose of storing and retrieve effective way of information may also known as knowledge discovery in databases (EDC) and enables data mining, data analysis and visualization of data from large databases to a level 'high abstraction without specific hypotheses in mind. The operation of the extraction is to use a method called modeling with him to make predictions. Technical data mining are the result of a long process of research and product development, including neural networks, decision trees and genetic algorithms.**

***Keywords* – Data Mining, Classification, Multi Class Classification, Bee Colony, Naïve Bayes, Probabilistic Classification.**

I. INTRODUCTION

Data mining [1] a nontrivial extraction of the novel, implicitly and actionable knowledge from large data sets is an evolving technology, which is a direct result of the increasing use of computer databases for the purpose of storing and retrieve effective way of information may also known as knowledge discovery in databases (EDC) and enables data mining, data analysis and visualization of data from large databases to a level 'high abstraction without specific hypotheses in mind. The operation of the extraction is to use a method called modeling with him to make predictions. Technical data mining are the result of a long process of research and product development, including neural networks, decision trees and genetic algorithms. This data recovery that the technology helps mining. Data mining can be considered the result of the natural evolution of information technology. This technology provides high availability of large amounts of data and the imminent convert this data into useful information and knowledge needs. Data mining is the extraction of patterns or knowledge of many interesting facts. It may be known by different names such as knowledge discovery (mining) in databases (KDD), knowledge extraction, data analysis / design, archeology, data dredging data collection information, the business intelligence and other. Data mining can be viewed as a result of the natural evolution of information technology. The database system industry has witnessed an evolutionary path in the development. The term "data mining" [2] is nothing but analysis of data in a database using tools which look for trends or anomalies without the knowledge of meaning of the data and is primarily used by statisticians, database researchers and business communities. A data mining software does not just change the presentation, but discovers previously unknown relationships among the data. The information on which the data mining process operates is contained in a historical database of previous interactions. In principle, data mining is not specific to one type of media or data. Data mining should be applicable to any kind of information repository. Some kinds of information that is collected are as follows:

Business Transactions

Every transaction in the business industry is (often) "memorized" for perpetuity. Such transactions are usuallytime related and can be inter-business or intra-businessoperations effective use of the data in a reasonable timeframe for competitive decision making is definitely the most important problem to solve for businesses that struggle tosurvive in a highly competitive world.

Scientific Data

Our society is amassing colossal amounts of scientific data that need to be analyzed. Unfortunately, wecan capture and store more new data faster than we cananalyze the old data already accumulated.

Medical And Personal

Data from government census to personnel andcustomer files, very large collections of information arecontinuously gathered about individuals and groups. Thistype of data often reveals if the information is collected,used and even shared. When correlated with other data thisinformation can shed light on customer behavior.

II. RELATED WORK

Zhen- Hui Song &Yi Li [6] Describe in the field of data classification as the associative classification (AC) has shown promising results for many other classification techniques on the set of static data. However, the increasing importance of data streams from a wide range of advanced applications has posed a new challenge for him. The author describes and CA-DS, a new classification algorithm for associative data stream based on the account of Lossy estimation mechanism (LC) and evaluates landmark model window. AC-DS applied for mining of various data sets from the UCI Machine Learning Repository and the result shows that the algorithm is effective and efficient. A partnership approach to the classification based on association rules for mining data streams. Empirical studies have shown its

effectiveness in the use of a large number of examples. ACD application of a high speed flow is running.

S.P. Syed Ibrahim1 K. R. Chandran [7] Describe in the field of mining weighted data classification association rules reflect semantic meaning of the subject, given its weight. extracts classification rules and build a classifier for predicting new instance of data. The author proposed compact associative method of weighted classification, integrating mining weighted association rules and classification for the construction of an effective weighted associative classifier. compact associative classification algorithm weighted random no defined class of attribute data and any association The weighted class rules is generated based on this attribute. The weight of the item is considered as one of the parameters of the generation of weighted class association rules. weight of the element is calculated by taking into account the quality of the transaction using the basic linkage model. Experimental results show that the proposed system generates less high quality standards. The objective of the integration of classification and extraction rules weighted association is to meet some of the needs created by modern data mining process. The development of weighted associative classification algorithm using the weighted compact associative classification (CWAC), reduce the number of rules in the classifier greatly. It shows how compact class generating weighted association rules, which can greatly improve the classification accuracy. This plays an essential role in the analysis of the consumption basket, medical diagnosis and many other applications. Experimental results show that the proposed method weighted compact associative Classification (CWAC) exceeded the ABC method. This work may be applied to a larger number of reference data.

Pei-Yi Hao[8]Describe in the field of data classification and the classification of the Association not only widely adopted, but also achieved good results in the data extraction. Literature argued that small disjunction and use multiple class association rules have a significant effect on the classification accuracy. The author proposes a CMAR (Classification based on several rules-Class Association) and Adriano Veloso proposed Lazy associative classifier algorithm for small mining disjunction. In addition, we have placed a new method of calculating weight in our algorithm to solve the problem through weight CMAR. The author uses ICU data 26 for the experience in our proposed algorithm. Finally, it is convincingly demonstrated that our proposed algorithm is very accurate. The author proposes a CMAR algorithm, and successfully integrates management methodology LAC algorithm. Experience shows that the use of the method of calculating the ALC algorithm support for small disjunction rule. After combining with multiple associative method rule, the accuracy can effectively increase. Furthermore, the author proposes, but also higher than the 1.01% L3 algorithm a new method of calculating weight not only greater than CMAR and the LAC algorithm. Furthermore, the author proposes an algorithm usually has a high precision if the number of rules and the number of candidates who are high or low limit restriction. Therefore, experience shows that our algorithm has stable characteristics and high precision. In some parts followed, this document does not perform any L3 used data series. Therefore, the tracker monitoring may be considered established so that it can use our proposed algorithm if you still have a high and stable accuracy additional data. Secondly, the use of the extraction of FP-growth rule is not suitable for the method of CLA support calculation algorithm so far. Due to the method of calculating support the ALC algorithm differs from the other classification algorithm. However, we can always try another data structure or algorithm to match the mining of the FP-growth rule. It could shorten the time achieved in the mining sector of the calculation rule and support for all the time production of the rule could be reduced

You Wan [9]Describe in the field of data classification and spatial models are similar to the rules of collocation association, but to explore spatial autocorrelation more dependent. They represent subsets of spatial characteristics Boolean whose cases are often found nearby. Existing models of co-location of mining research concern only the spatial attributes, and few of them can handle the huge amount of non-spatial attributes of spatial data sets. In addition, distance threshold used to define spatial neighborhood. However, it is difficult to decide the distance threshold for each set without specific prior knowledge of spatial data. In addition, spatial data sets are usually not even distributed, so a value of only the distance, can not fit all of the irregular distribution or a spatial data. Here, they have proposed a qualitative spatial model of co-location, which contains both spatial and non-spatial information. And the k closest features (k-NC) neighborhood relations are defined to establish the spatial relationship between the different types of spatial characteristics. all k-NF instantiation of an entity to measure closely with other functions are used. To find qualitative co-location patterns in large spatial datasets received formal definitions, and proposed a QuCOM (qualitative spatial models of mining co-location) algorithm. Experimental results on the map data of US QuCOM thesis show that the algorithm is accurate and effective, and based model contains the most interesting information. When considering both spatial and non-spatial attributes to the analysis of spatial data, you can get more detailed and meaningful information. The author proposes a new model of spatial association: qualitative spatial distribution of co-location. Mines this model of spatial data sets, we gave several definitions and algorithm proposed QuCOM. analysis comparison experiments gave clear evidence that the qualitative spatial colocalization mining can be more interesting than traditional results. The ability to manage spatial characteristics of the line remains a challenge in the field of mining spatial association rules, because the definition of their close relationship is more difficult than the points and polygons. Moreover, the algorithm can not handle the item QuCOM

polygons and spatial characteristics as well. The extending handle all kinds of spatial characteristics could be our ongoing work.

Achilleas Tziatzios [10]Describe in the field of data classification as the ability to learn the data classification rules is important and useful in a variety of applications. Although many methods have been proposed for this purpose, can derive some classification rules with numerical ranges (intervals). We consider how field-based classification rules can be obtained from the digital data and propose a new method inspired by the Mining Association classification rules. This research method similar ranges partners in how sets of related items are sought categorical attributes mining association rules, but uses class values to guide the search path, so that only those beaches that are found relevant for the calculation of classification rules. Our preliminary experiments demonstrate the effectiveness of our method. The author proposes a new method for finding classification rules based range from digital data. Our method is based on the approach to classification rules mining association. In other words, looking for associations between the numeric ranges, but our research is guided by the class values. This enables precise classification rules based on the resulting sphere effectively.

Rupalihaldulakar[11]Describe in the field of data classification as Strong rule generation is an important area of data mining. The author proposed a design a novel method for generation of strong rule. In which a general Apriori algorithm is used to generate the rules after that they use the optimization techniques. Genetic algorithm is one of the best ways to optimize the rules .In this direction for the optimization of the rule set we design a new fitness function that uses the concept of supervised learning then the GA will be able to generate the stronger rule set. In this direction we optimize association rule mining using new fitness function. In which fitness function divide into two classes' c1 and c2 one class for discrete rule and another class for continuous rule. Through this direction we get a better result. To make genetic algorithm more effective and efficient it can be incorporated with other techniques so it can provide a best result.

XING Xue [12]Describe in the field of data classification as association rules which applied in data mining that aims to analyze large source data and reveal knowledge hidden in the database. The author proposed a presents the principle of association rules in data mining. It has been viewed as an important evolution in information processing. The author proposed a association rules mining to the software of examination paper evaluation system, obtaining the useful information which is hidden in the database. It's concluded that the algorithm provides a valuable analysis of information to the examination paper evaluation system. Keywords-association rules. Association rules is the one most important theory in data mining, which have a wide range of applications in the various fields, but, applied to the Evaluation of reliable, it can be said that has just begun, with the mining association rules theoretical the constant further research and using, the rational, efficient and objective analysis of The examination, all from the Association Rules theoretical support. And, judging from the current access to a large number of information, Association rules, applied to the analysis of the test research, has aroused the expert's wide attention.

III. CONCLUSION

Recent research having lower predictive accuracy which lead to tends, combine existing log-linear model with probabilistic techniques. While a search for informative aggregate features is computationally expensive, when it succeeds, the new aggregate features can increase the predictive accuracy. There are several possibilities for a combined hybrid approach. Once good aggregate features are found, they can be treated like other features and used in a decision tree. A simple decision forest is fast to learn and can establish a strong baseline for evaluating the information gain due to a candidate aggregate feature. The regression weights can be used to quickly prune uninformative join tables with or small weights, which allows the search for aggregate features to focus on the most relevant link paths. Whereas in a hybrid mining algorithms to improve the classification accuracy rates of decision tree (DT) and naïve Bayes (NB) classifiers for the classification of multi-class problems but it's don't have any genetic algorithms, rough set approaches and fuzzy logic, be used to deal with real-time multi-class classification tasks under dynamic feature sets.

REFERENCES

[1] B.N. Lakshmi. #1, G.H. Raghunandhan. #2 "A conceptual Overview of Data Mining" Proceedings of the National Conference on Innovations in Emerging Technology-2011 Kongu Engineering College, Perundurai, Erode, Tamilnadu, pp.27-32. India.17 & 18 February, 2011.

[2] Han J. and M. Kamber (2000), Data Mining: Concepts and Techniques, Academic Press, San Diego, CA.

[3] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth "From Data Mining to KDD in Databases" pp. 0738-4602 1996.

[4] Xun Zhu1, Hongtao Deng2, Zheng Chen3 "A Brief Review On Frequent Pattern Mining" PP-4-11 2011 IEEE.

[5] Thair Nu Phyu "Survey of Classification Techniques in Data Mining" Vol I Imecs2009, March 18 - 20, 2009, Hong Kong.

[6] Zhen- Hui Song &Yi Li, "Associative classification over Data Streams", IEEE, PP.2-10, 2010.

[7] S.P.Syed Ibrahim1 K. R. Chandran2 M. S. Abinaya3 "Compact Weighted Associative Classification" IEEE pp.8-11, 2011.

[8] Pei-yihao, yu-de Chen "a novel associative classification algorithm: a combination of LAC and CMAR with new measure of Weighted effect of each rule group" IEEE pp.9-11, 2011.

[9] You Wan#1, Chenghu Zhou*2" QuCOM: k nearest features neighborhood based qualitative spatial co-location patterns mining algorithm" IEEE pp.8-11, 2011.
[10] Achilleas Tziatzios and Jianhua Shao, GrigoriosLoukides" A Heuristic Method for Deriving Range-Based Classification Rules" IEEE pp.6-11, 2011.
[11] Rupalihaldulakar, prof. Jitendra agrawal" Optimization of Association Rule Mining through Genetic Algorithm" (IJCSE) Vol. 3 No. 3 Mar 2011.
[12] XING Xue, CHEN Yao. WANG Yan-en" Study on Mining Theories of Association Rules and Its Application"IEEE PP.2-10, 2010.
[13] Yingqin Gu1,2, Hongyan Liu3, Jun He1,2, Bo Hu1,2 and Xiaoyong Du1,2 "A Multi-relational Classification Algorithm based on Association Rules" pp.4-9 2009 IEEE.
[14] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient Classification Based on Multiple Class-Association Rules", Proceedings of the ICDM, IEEE Computer Society, San Jose California, 2001, pp. 369-376.
[15] X. Yin, and J. Han, "CPAR: Classification based on Predictive Association Rules", Proceedings of the SDM, SIAM, Francisco California, 2003.
[16] Zhen Peng, Lifeng and Wu Xiaoju Wang "Research on Multi-Relational Classification Approaches" pp.3-9 2009 IEEE.
[17] S. Dzeroski, N. Lavrac eds. Relational data mining. Berlin: Springer, 2001.
[18] He J, Liu HY, Du XY. Mining of multi-relational association rules. Journal of Software, 2007, 18(11): 2752-2765.
[19] M.J. Zaki, and C.j. Hsiao, "CHARM: An Efficient Algorithm for Closed Item set Mining", Proceedings of SIAMOD International Conference on Data Mining, 2002, pp. 457-473.
[20] L. Xu, and K. Xie, "A Novel Algorithm for Frequent Item set Mining in Data Warehouses", Journal of Zhejiang University, Journal of Zhejiang University, Zhejiang China, pp.216-224, 2006.
[21] R. agrawal, t. imielinski and a. swami. "Mining association rules between sets of items in large databases". In proc. of the ACM sigmoid conference on management of data, Washington, D.C. May 1993.
[22] B. Liu, W. Hsu, and Y. Ma. "Integrating classification and association rule mining". In KDD 98, New York, NY, Aug.1998.
[23] B. Liu, Y. Ma, and C.-K. Wong, "Improving an association rule based classifier," in Proc.4th Eur. Conf. Principles Practice Knowledge Discovery Databases (PKDD-2000), 2000.
[24] Rajdev Tiwari, Manu Pratap Singh "Correlation-based Attribute Selection using Genetic Algorithm" International Journal of Computer Applications (0975 – 8887) Volume 4– No.8, August 2010.
[25] Kalyanmoy Deb, "Introduction to Genetic Algorithms", Kanpur Genetic Laboratory (Kangal), Depart of Mechanical Engineering, IIIT Kanpur 2005.