International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Implementing Hybrid GA & Fuzzy KNN to Optimize the Classification of Fuzzy KNN

Simarpreet Kaur[1], Jyoti Arora[2]

[1]Student, Desh Bhagat University

[2]Assistant Professor, Desh Bhagat University

Abstract: ***In this paper data mining is the process of fetching hidden knowledge from a wide store of raw data. The knowledge must be new, and one must be able to use it. 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. In this paper 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, Genetic Algorithm, Precision, Recall, f-measures, g-mean, sensitivity & specificity

1. Introduction

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

This white paper provides an introduction to the basic technologies of data mining. Examples of profitable applications illustrate its relevance to today's business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users.

1.1 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.

2. Review of Literature

Paresh Tanna et al "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 BavisiȦ et al "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.

Manjari Anand et al "Customer Relationship Management using Adaptive Resonance Theory" CRM is a kind of implemented model for managing a company's interactions

with their customers. CRM involves the customer classification to understand the behavior of the customer. There is a vital role of the data mining techniques for the classification. This paper presents the concept of one of the data mining technique ART for the customer classification for CRM.

Ms. Saranya, et al "Decision Support System for CRM in Online Shopping System**",** Customer Relationship management (CRM) is a seriously considered issue in today's competitive corporate world. For a firm to maintain an intact relationship with its customers, the vast amount of data within a business enterprise can be refined, mined and analyzed. This paper proposes a Decision Support System for analyzing and retaining the customers in an Online shopping System using various data mining techniques. To achieve successful Customer relationship management various mining algorithms for pattern extraction from data have been utilized. Managerial decision makers can make use of these patterns for raising the profit graph of the firm.

Kamal R. et al "Adaptive Pointing Theory (APT) Artificial Neural Network" The choice value and the testing process against the vigilance parameter, characteristic of ART Neural Network, is merged. Only, a single unique test is required to determine if a committed category node can represent the current input or not. Advantages of APT over ART are: 1-Avoid testing every committed category node before deciding to train a committed category node or a new node must be committed, 2-The vigilance parameter is fixed during training, and 3-The choice value parameter is eliminated.

3. Methodology

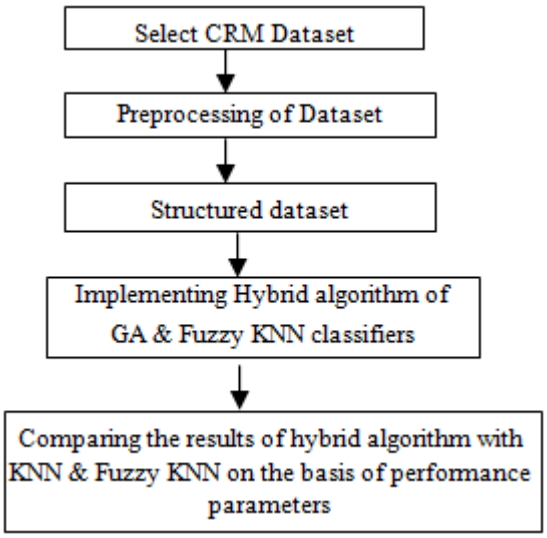

Figure 4.1: Flow of Work

In the classification of customer relationship management dataset various attributes that has to be mined on the basis of decision. Various stages that has to be carried out for development of proposed model.

Phase 1

In this phase, data is acquired using various acquisition tools. This data has to be stored in structured format. Later on mining is performed on the basis of defined rules.

Phase 2

In this phase, structured data has to be normalized and pre – processed. This database will be classified into different classes on rules of GA and Fuzzy KNN classification. After classification various attributes will be extracted from classified data and various parameters are analysed for performance evaluation.

4. Results

In the purposed work different parameters have been used for evaluation of the performance of the purposed work. In this purposed work CRM dataset have been used that has 50000 instances and 231 attributes. This dataset has been used for classification using GAF approach this approach use membership function and genetic operation to classify the data into different classes. To validate the purposed work different algorithm have been used for comparison of the purposed approach. The purposed work has been compared with Fuzzy-KNN classification and KNN classification approach on the basis of different parameters. The dataset have been classified into different classes on the basis of the value of K. the parameters have been computed by using these different values of K.

Table 5.1: Parameters for classification in two classes

Classifiers/metrics	GAF	KNN	FKNN
ACCURACY	0.98	0.87	0.93
SENSTIVITY	0.92	0.8	0.9
SPECIFICITY	0.16	0.2	0.1
PRECISION	0.92	0.73	0.83
RECALL	0.95	0.65	0.85
F-MEASURE	0.97	0.79	0.89
G-MEAN	0.85	0.73	0.89

This table represents the classification of the CRM dataset values when classification has been done on the basis of two classes. These approaches classified the testing data into two different classes on the basis of distance of different samples from the training dataset.

Table 5.2: Parameters for classification in three classes

Classifiers/metrics	GAF	KNN	FKNN
ACCURACY	0.76	0.86	0.76
SENSTIVITY	0.65	0.45	0.55
SPECIFICITY	0.25	0.55	0.35
PRECISION	0.65	0.7	0.73
RECALL	0.79	0.62	0.72
F-MEASURE	0.87	0.75	0.75
G-MEAN	0.65	0.67	0.71

This table represents the classification of the CRM dataset values when classification has been done on the basis of three classes. These approaches classified the testing data into two different classes on the basis of distance of different samples from the training dataset. As the number of classes increases in the prediction of the dataset accuracy gets decrease due to availability of actual dataset into two classes.

Table 5.3: Parameters for classification in four classes

Classifiers/metrics	GAF	KNN	FKNN
ACCURACY	0.64	0.52	0.6
SENSTIVITY	0.56	0.52	0.51
SPECIFICITY	0.56	0.5	0.4
PRECISION	0.65	0.54	0.62
RECALL	0.59	0.63	0.69
F-MEASURE	0.69	0.61	0.64
G-MEAN	0.61	0.55	0.59

This table represents the classification of the CRM dataset values when classification has been done on the basis of four classes. These approaches classified the testing data into two different classes on the basis of distance of different samples from the training dataset. As the number of classes increases in the prediction of the dataset accuracy gets decrease due to availability of actual dataset into two classes.

4.1 Comparison of Evaluation Parameters

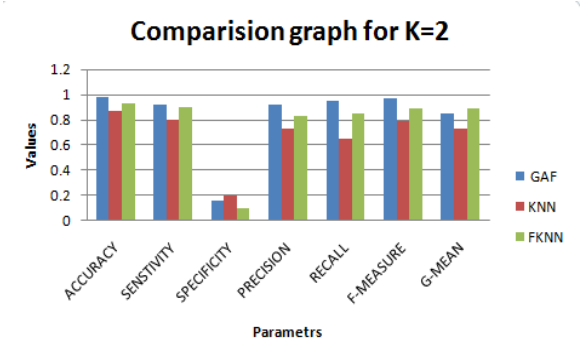

Figure 4.1: Comparison of different approaches for K=2

This figure represents the comparison between different classification approaches on the basis of different parameters. These approaches have been used for classification of data into two different classes and the parameters have been analyzed for all the approaches. On the basis of these parameters one can validate and optimized best approach for classification.

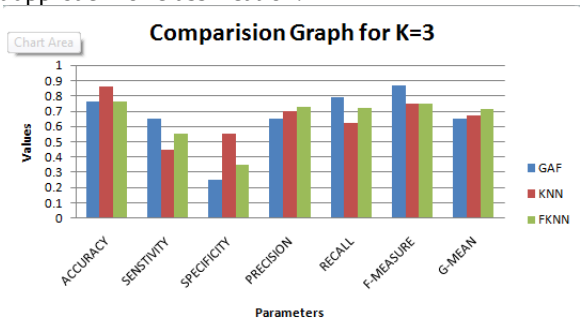

Figure 4.2: Comparison of different approaches for K=3

This figure represents the comparison between different classification approaches on the basis of different parameters. These approaches have been used for classification of data into three different classes and the parameters have been analyzed for all the approaches. On the basis of these parameters one can validate and optimized best approach for classification.

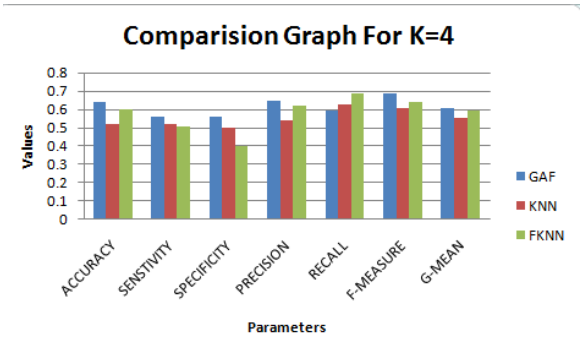

Figure 4.3: Comparison of different approaches for K=4

This figure represents the comparison between different classification approaches on the basis of different parameters. These approaches have been used for classification of data into four different classes and the parameters have been analyzed for all the approaches. On the basis of these parameters one can validate and optimized best approach for classification.

5. Conclusion & Future Scope

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. After this classification genetic algorithm has been used for optimization of predicted label. Genetic algorithm use different operators like crossover, mutation, and selection for prediction of best classification labels to a data samples. After this the actual and predicted class labels have been used for prediction of various parameters that have been used for validation of purposed work. The purposed work has been compared with various approaches like KNN and fuzzy KNN.

By analyzing various parameters like accuracy, precision, recall, f-measure, sensitivity, specificity and G-mean one can conclude that the fuzzy and genetic algorithm based approach provide better classification rather than that of the simple fuzzy classification and KNN classification.

6. Future Work

In the future reference the classification can be done by using different parameters in the real world data. In the future the data can be used for classification by using other artificial intelligence approaches for optimization of predicted labels.

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 Volume 3, Issue 1, 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] Paresh Tanna "Using Apriori with WEKA for Frequent Pattern Mining", International Journal of Engineering Trends and Technology (IJETT), 2014, pp. 127-131.
[14] Shrey BavisiÀ "A Comparative Study of Different Data Mining Algorithms", International Journal of Current Engineering and Technology, 2015, pp. 3248-3252.
[15] Manjari Anand "Customer Relationship Management using Adaptive Resonance Theory", International Journal of Computer Applications, 2013, pp. 43-47.
[16] 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.
[17] Kamal R. "Adaptive Pointing Theory (APT) Artificial Neural Network", International Journal of Computer and Communication Engineering, 2014, pp. 212-215.
[18] Mohammed Al-Maolegi "An Improved Apriori Algorithm For Association Rules", International Journal on Natural Language Computing (IJNLC), 2014, pp. 21-29.
[19] Asghar, S. "Automated Data Mining Techniques: A Critical Literature Review" 978-0-7695-3595-1, pp. 75–79, IEEE, 2009.
[20] D. Clot., "Using functional PCA for cardiac motion exploration", IEEE International Conference on Data Mining, 2002, pp. 91-98.
[21] Yin-Fu Huang., "Mining generalized association rules using pruning techniques", IEEE International Conference on Data Mining, 2002, pp. 227–234.
[22] Yuh-Jyh Hu., "Mining a set of co regulated RNA sequences", IEEE International Conference on Data Mining, 2002, pp. 625-628.