International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Implementing Agents in Classification with Partition Method

E. Padmalatha
Asst. Professor, CBIT

Dr. C. R. K. Reddy
Professor, CBIT

Dr. Padmaja Rani
Professor, JNTUH

***Abstract* – Knowledge management needs very high end accuracy and efficiency in making decisions, this can be achieve by implement agent system with data mining. Heterogeneous multi agent coordination improves accuracy of large and complex data mining task. A working agent is a data mining agent designed to ferret out relevant information to classify the data stream .On one hand, data mining approach can drive the knowledge extraction from huge amount of data ;on the other hand,, by using a single classifier it may not provide reliable results. Therefore, to improve the classification accuracy rates, ensemble classifiers joint with agents and relied on multiple learning algorithms, can be aid to improve the performance. The main objective of the proposed system is to improve the accuracy of the classification model by using ensemble technique with the data mining agents. Data can be divided into partitions according to the partitions the number of agents will created and the partitions are processed by all the agents in the systems in equivalent, the results from the partitions are fused to consider the accuracy. Finally accuracies are compared with respective to those achieved by using classifier.**

***Keywords* – Agent, Data Stream, Ensemble Classification.**

I. INTRODUCTION

Data categorization plays key task in data mining. Building predictive models on the large and complex data will take long time and data study will not be state-of-the-art with the constant arriving data this may result in invalid predictions, so single Classifier performance and efficiency may be reduced with increase of data, unvoted data also decreases the prediction of classifier. By using the idea of Ensemble methodology a better prediction model can be build by integrating multiple models. CoLe[18] a cooperative data mining approach is addressing some of the solutions of the classification performance by discovery of mixture knowledge. To get accurate mined knowledge single classifier decision is not perfect and it may lead to wrong prediction also so the combining the decision of different classifiers with different decision may enhance the prediction accuracy of the system. [6]. Multi-agent and data mining have been generally used approach in edifice large and difficult system, (MAS) [4,17]. Data mining with multi agents can address some of issues like efficiency, scalability. When we use these approaches combining data can be divided in to partitions and agents will work parallel with all partitions, the results from partitions are then consider in decision making.

II. ENSEMBLE CLASSIFIERS

Classification is a data mining performance used to guess group association for data instances [11,12]. Eg.: Deciding if weather is sunny ,cloudy or rainy. Some of the classification techniques decision trees, Neural networks. An ensemble consider all of the accurate classifiers the learning algorithm then average their votes or weights and reduce the risk of choosing the wrong classifier Figure1 top left depict this situation The outer curve denotes the hypothesis space H The inner curve denotes the set of hypotheses that all give good accuracy on the training data The point labeled f is the true hypothesis and we can see that by averaging the accurate hypotheses we can and a good estimate to f [3,19].

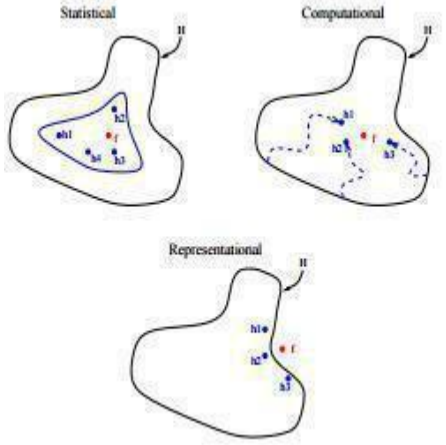

Fig.1. Explanation of ensemble technique advantage [21].

II. RELATED WORK

Past many years Data mining and Multi agents combination is playing vital role in decision making .The characteristics of Agent systems will consider as follows, Multi agents [4,12,17] can be used to solve problems that are huge to solve by a federal agent because of resource limitations [1]. Agents are capable of being autonomous n robotic in nature. Agents can interact with other external systems to manage both distributed and local knowledge by learning from their prior experience, this character of agent is important in the field of data mining as the data is persistently transform and revise.

In statistics and machine learning, ensemble[10] methods use multiple learning algorithms to obtain better prognostic performance than comparing to any of the constituent learning algorithms. A machine learning ensemble refers only to a tangible predetermined set of alternative models, but typically allows for much more stretchy structure to exist between those alternatives[3]

Copyright © 2015 IJECCE, All right reserved

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

where as statistical ensemble is usually uncontrolled. The size of the Ensemble method may degrade the performance of prediction, Pruning is added to the system to maintain the accuracy of the system. In Ensemble technique either in supervised learning or unsupervised learning a set of learning machines will try to produce more reliable and accurate predictions either by considering decisions, learning algorithms, various views of data or by considering unique characteristics .The basic model of Ensemble model is as shown in figure 2.[9]

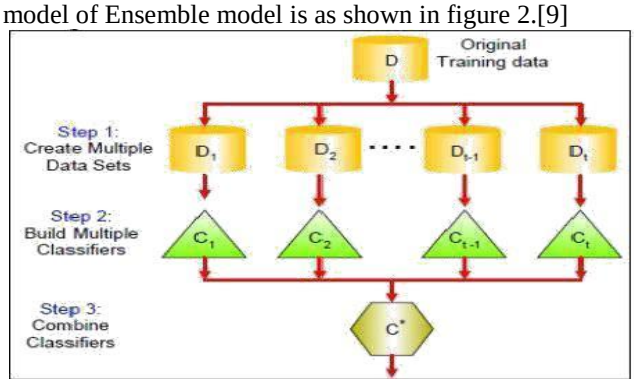

Fig.2. Basic Ensemble model[19].

Methods of Ensemble classifiers

Combination of weak learners may lead to good prediction in the knowledge management this will lead to a system with strong learning techniques, this approach replace the poor learning algorithms by executing the various extra computations which may not be computed by the weak learners. Fast algorithms such as decision trees are commonly used with ensembles(for example Random Forests),although slower algorithms can profit from ensemble learning techniques as well .An ensemble is itself a supervise algorithm [11], because it can be trained and then used All the Ensemble methods are constructed by with concept of a single learning method with different subset of training data can be used[7].

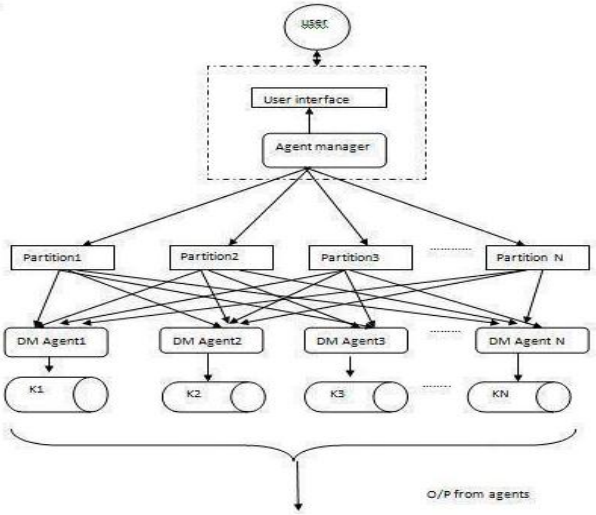

Fig.3. Classification model with ensemble agents and classifiers.

The mixture of dissimilar opinions concept is taken by human natural activity like human takes noteworthy decision by considering more than a few thoughts before taking any decision. Ensemble based classifiers generally refer to the mixture of classifiers that are negligible variants of the same base classifier, which can be considered in the broader concept of multiple classifier systems [19]. The disadvantages and weakness of ensemble classifiers [16] are increased storage ,increased computation , decreased comprehensibility.

III. IMPLEMENTATION DETAILS

Ensemble agent is implemented with JAVA code, single classifiers for comparison are taken from WEKA tool[15].

Partitioning

The dataset chosen by the user is taken in the ARFF format. Depending upon the size of the dataset, it is partitioned into some number of folds. In this system, the datasets with less than 1000 instances have been divided into 10 folds and those with instances more than 1000 and less than 5000, have been divided into 20 folds[7].Now, these folds have been divided into various partitions depending upon the size of the dataset as shown in Fig 3. Each such partition is sent to a fixed number of agents. The number of agents to be employed for the task is also dependent on the size of the dataset. Here ,the dataset with 1000 instances is divided in to 10 folds and two partitions each containing 5 folds. Two agents are used to achieve the classification. The dataset with 1000-5000 instances is divided in to 20 folds and four partitions each containing 5 folds. Four agents are used for classifying these four partitions .Each partition obtained ,as described in the above module is sent to all agents considered for performing the classification. Each agent contain four different classifiers such as J48, Decision Table, etc. The agent classifies the partition sent to it by all the algorithms present in it. Then, by comparing all the thus -obtained accuracies, the best accurate **classifier's result and confusion matrix are considered for further processing. The other classifiers'** accuracies are ignored.

Calculation of the final accuracy for the dataset

Each partition is sent to all the agents and from each agent the best accuracy is considered. For example, consider there are two partitions and two agents. The first partition is sent to both the agents. Then, the accuracies returned from the both agents are compared and the highest one is considered. Similar operation is performed for all the partitions.

		Predict Label	
		Positive	**Negative**
Known Label	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Fig.4. The possible prediction are tabulated in Confusion matrix [2].

Copyright © 2015 IJECCE, All right reserved

Now, we have the best accuracies of each partition. The final accuracy is calculated by computing the average of all the above obtained accuracies of all the partitions as shown in fig 3. Accuracy of Ensemble Agent is calculated by considering the confusion matrix which is generated show in fig.4.

Generation of accuracy table

The user is allowed to choose five different datasets. The above stated operations are Performed for each dataset and the final accuracies are obtained by the usage of Ensemble classification using agents table as shown in table 2. is generated which contain accuracies of the classified dataset by using J48 classifier ,decision table classifiers SVM,JRIP and Ensemble classifier 48 and Decision table [10],SVM ,JRIP are pre-defined classifiers in WEKA[15] .It is observed from the table that the accuracy of ensemble classifier is more than the other single classifiers.

Table 2: Showing an accuracies of different classifiers.

Classifiers/Dataset Name	Breast -Cancer	Contact-Lens	Credit-g	Diabetes	Glass
Decision table	73.076	70.83	72.2	73.95	62.14
J48	71.67	83.33	72.2	74.86	66.82
SVM	69.5	70.83	75	77.03	56.07
JRIP	70.09	75	71.7	76	68.6
Ensemble Agent	78.5	88.09	76.2	77.05	82.64

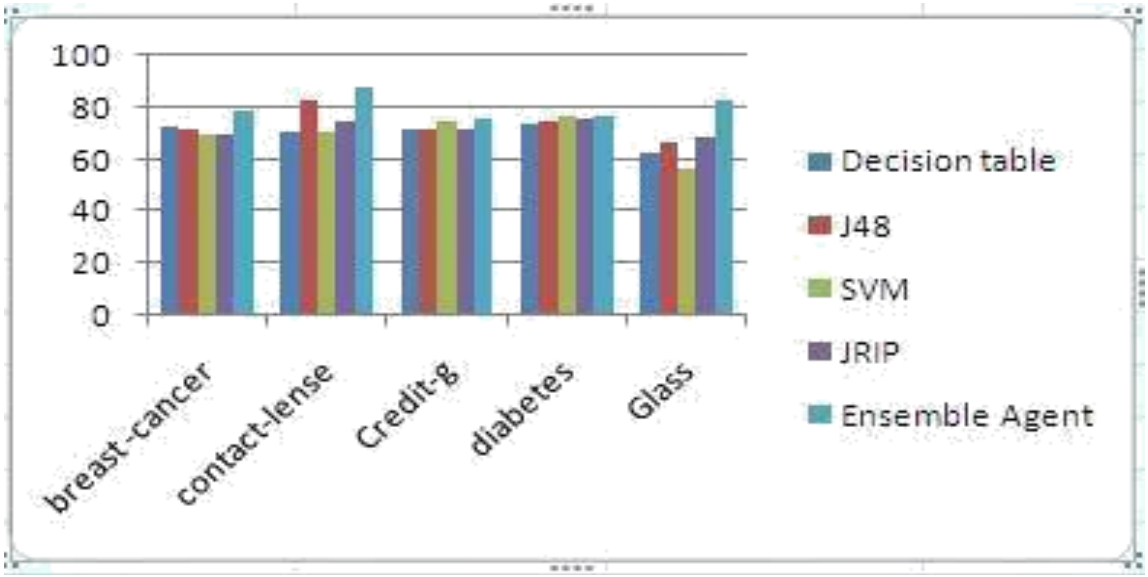

Fig.3. Comparision graph of other classifiers and proposed system.

IV. Conclusion

The main goal of the system is to explore how data partitioning and multi agent approach can help to improve the efficiency in accuracy, prediction tasks in real time system with help of data mining. Comparison of single classifier and ensemble agent proved that proposed technique increases the accuracy. Multi agent system classifier Efficiency depends on the number of partitions , too many partitions and too less partitions will decreases the accuracy of the decision system. There is no hard and tough rule to decide the number of partitions .This proposed system is not depend dent on application and not limited to specific datasets , it can be implemented in various domains where mining of knowledge plays major role and accuracy is crucial.

References

[1] M. Wooldridge, *An Introduction to Multi Agent Systems. John Wiley & Sons Ltd, 2002*. John Wiley Sons, 2002.

[2] Data Mining: Concepts and Techniques. J. Han and M. Kamber. Morgan Kaufmann, 2000.

[3] Ensemble Methods in Machine Learning Thomas G Dietterich Oregon State University Corvallis Oregon USA tgdcsorstedu WWW home page http/wwwcsorstedutgd .

[4] "Cuong Tong, Dharmendra Sharma and Fariba Shaadabi,A Multiagent approach to knowledge discovery/2010/IEEE/ WIC/ACM

[5] Cheeseman, P. 1990. On Finding the Most Probable Model. In Computational Models of Scientific Discovery and Theory Formation, eds. J. Shrager and P. Langley, 73–95. San Francisco, Calif.: Morgan Kaufmann.

[6] Cheeseman, P., and Stutz, J. 1996. Bayesian Classification (AUTOCLASS): Theory and Results. In Advances in Knowledge Discovery and Data Mining, eds.

[7] Advances in Knowledge Discovery and Data Mining. U.S. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, AAAI/MIT Press, 1996.

[8] Knowledge Discovery in Databases. G. Piatetsky- Shapiro and W. J. Frawley. AAAI/MIT Press, 1991.

[9] Tan, Pang-Ning,Michael Steinbach,and Vipin kumar. Introduction to data mining. Vol.1. Pearson Addision Wesley 2006.

[10] Ensemble Classification for Drifting Concept. E. Padmalatha. C.R.K. Reddy., B.Padmaja rani. International Journal of Computer Applications (0975-8887) Vol. 80-No 11,Oct. 2013.

Copyright © 2015 IJECCE, All right reserved

[11] Dougherty, J., R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. In Proceedings of International Conference on Machine Learning (ICML-1995), 1995.

[12] Supervised Machine Learning: A Review of Classification Techniques S. B. Kotsiantis Department of Computer Science and Technology University of Peloponnese.

[13] S. R. Ira. (2004, Intelligent Agents. *Communications of the Association for Information SystemsVolume14, 2004*pp. 275-290.

[14] Tan, Pang-Ning, Michael Steinbach, and Vipin kumar. Introduction to data mining. Vol.1.Pearson Addision Wesley 2006.

[15] https://en.wikipedia.org/wiki/Weka_(machine_learning) Weka (Waikato Environment for Knowledge Analysis).

[16] Classifier Ensemble Recommendation Pyry Matikainen1 Rahul Sukthankar2,1 Martial Hebert1 pmatikai@cs.cmu.edu rahuls@cs.cmu.edu hebert@ri.cmu.edu 1The Robotics Institute, Carnegie Mellon University 2Google Research.

[17] A Multi-Agents Approach to Knowledge Discovery Cuong Tong, Dharmendra Sharma and Fariba Shadabi Faculty of Information Sciences and Engineering, University of Canberra, ACT, 2601Australia .

[18] CoLe: A Cooperative Data Mining Approach and Its Application to Early Diabetes Detection by Jie Gao , Jörg Denzinger , Robert C. James.

[19] A survey on one class classification using ensemble method, Jay Bhatt, Nikita patel, Department of Computer science IJIRST ISSN 2349-6010.

[20] Polikar, R. (2006). "Ensemble based systems in decision making". *IEEE Circuits and Systems Magazine* 6(3): 21–45. doi:10.1109/MCAS.2006.1688199.

[21] Artificial Neural Networks Ensemble Used For Pipeline Leak Detection Systems, Inhaúma N. Ferraz,Ana C. B. Garcia, Proceedings of the 7th International Pipeline Conference IPC 2008 September 29, October 3, 2008, Calgary, Alberta, Canada Flávia C. Bernardini,

Copyright © 2015 IJECCE, All right reserved