ISSN (Online) : 2277-5374
ISSN (Print) : 2277-5366

An effective and accurate fusion result from multi class ensemble classification

[*1] C. Gayathri, [2] R.Umarani

[*1] Research Scholar, Department of Computer Science, Karpagam University, Coimbatore – 641021, Tamilnadu, India.
[2] Associate professor of Computer Science, Sri Sarada College for Women, Salem – 636016, Tamilnadu, India.
[1] gayathriphd2015@gmail.com

Abstract

Background/Objectives: Financial fraud detection is the most challenging task in an online transaction oriented applications which concern more to provide the secured environment for the users. Various researches has been conducted previously that focus on providing the most secured environment to the users by finding and preventing the malicious patterns.

Methods/Statistical analysis: Classification is one of the most proved techniques for detecting the most malicious patterns that resides in the financial database by using which the malicious patterns can be identified. In our previous research work Optimal Ensemble Architecture Selection using Firefly and the dempster shafer theory based Ensemble is done for finding the fraudulent behaviour in the accurate manner. The ensemble classifier fusion approach used in the previous methodology called dempster shafer theory retrieves the fusion result as classifier output with more confidence value. This approach is computationally inefficient and doesn't concentrate on interrelation between different classifier results due to its additive measure property.

Findings: This problem is resolved in this work by introducing the fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFSSI) and the fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFSCI). These approaches can find the better and accurate Ensemble result by considering the relation between the different classifier results.

Improvements/Applications: The experimental tests conducted were proves that the proposed approach provides better result than the existing approach in terms of improved classification accuracy in the matlab simulation environment.

Keywords: Ensembling fusion, Sugeno, Choquet fuzzy integral, fuzzy values.

1. Introduction

Financial fraud detection is the most concerned and improved task in the real world environment which is considered by the most of the researchers present in the real world. The financial fraud detection ensures the secured environment for the users to precede their online transaction in the improved manner. Pattern recognition is the most effective process that can find the malicious patterns that differs from other patterns in the accurate manner. Classification is one of the most important tasks that can recognize the pattern divergence in the accurate manner.

In this research work, ensemble classification is introduced that can classify and identify the identity patterns in the accurate manner for detecting the fraudulent behaviours. This is done by training the data's by using different classifiers and combining the results of them. In our previous research Optimal Ensemble Architecture Selection using Fire Fly and dempster shafer based classification (OEAS_FFA-DSEC) is introduced which attempts to detect the fraudulent behaviour in the accurate manner by using the optimization approach that can select the most optimal ensemble from the different set of ensembles that can generate more accurate results. The Ensembling output is gained by using the methodology called the dempster shafer based fusion approach that will fuse the classifier results in terms of their confidence values that is calculated by considering its misclassification cost. This method is computationally inefficient which does not look up the relationship between different classifiers based on their varying importance level.

The main contribution of this research work is to avoid the limitation that present in our previous research work by using the fuzzy integral methods for classifier fusion. Two types of fuzzy measures are used in this research work for getting the more accurate ensemble classification result for detecting the fraudulent behaviour accurately. Two

fuzzy measures that are used in this research work are sugeno and choquet measures. By using this measure two approaches are proposed namely: fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFS) and the fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFS).

These methodologies are executed and evaluated with our previous research methodology namely OEAS_FFA-DSEC in the matlab simulation environment to prove their efficiency in terms of improved classification result. The proposed methodologies attempts to find the most optimal classification result and working scenario of the proposed research scenario are discussed detailed in the following sections.

The organization of this work is given as follows: In this section, the general introduction of the proposed research is given. In section 2, detailed description about the different research papers which was published earlier is discussed. In section 3, brief description and introduction about the proposed research work is given. In section 4, results and discussion were discussed in the detailed manner with the graphical representation. Finally in section 5, overall conclusion of this work is discussed based on the performance measure values which are retrieved in the previous section.

2. Related works

In [1] suggested an automatic classifier fusion for produce recognition. The main aim of this research scenario is focused on the supporting the automatic classification of items in a transactional settings. This research is used to produce an efficient approach for uniting the low cost classifiers. The novelty of this research scenario is depends on the utilization of different measures to discover which base classifiers are appropriate to be fusion. In [2] Discussed, on the dynamic of biometric fusion algorithms. This algorithm combines the output of several biometric classifiers to make a result about the uniqueness of an entity. This research scenario is used to deal with the issue of limitation in number of training samples and the use of a single fusion rule may not be useful due to the variety of situations. In [3] recommended the linear dependency modelling for classifier fusion. This research scenario is addressed the independent consideration issue in fusion procedure. In this scenario, the author introduced an approach called as linear classifier dependency method which is sued for improving the linear programming problems effectively.

In [4] suggested Knowledge fusion for probabilistic generative classifiers with data mining applications. In this research scenario, the author introduced a novel way to combine the classifiers in the parameters level of classification rules. The classifiers based on the fusion model and Bayesian algorithm improves the classification performance in this research. This scenario demonstrates that combining two (or more) classifiers can be completed through increasing the hyper distributions of the parameters and obtains straightforward formulas for the task. The major benefit of this fusion scheme is that the hyper distributions are preserved throughout the combination method. In [5] suggested that sparse classifier fusion for speaker verification. In speaker verification, fusion is normally executed as a weighted linear grouping of the bottom classifier gains, where the fusion weights are computed through a logistic regression approach. A different method for fusion is to use classifier ensemble assortment, which can be distinguish as sparse regularization used to logistic regression. Although score fusion is broadly considered in speaker verification, classifier ensemble range is much less calculated. Sparse regularization refers by means of adding to optimize the mixture weights for the complete classifier ensemble then execute concurrently classifier ensemble range through forcing unnecessary classifiers to contain nil weight.

In [6] presented a novel approach to classifier fusion based on upper integral. Combining the number of classifiers can typically progress the performance of individual classifiers and fuzzy integral is the communication between individual classifiers which is referred an efficient tool of fusion. In order to build the optimal usage of the individual classifiers and their fusions, in this research a novel approach of classifier fusion is introduced. The upper integral is utilized to practically organize the restricted resources, and therefore to exploit the categorization effectiveness. Through resolving an optimization issue of upper integrals, the scenario obtains a scheme for transfer proportions of instances to diverse and individual classifiers as well as their fusions. It is used to optimize the system performance rather than previous approaches. In [7] discussed about Polarimetric-spatial classification of sar images based on the fusion of multiple classifiers. Conventional image prediction techniques have issue with semantic information and their prediction results are not satisfied. To deal with the above mentioned issue, a novel object oriented method which fuses pixel based prediction and segmentation scheme for the prediction of Polarimetric synthetic space radar images. The enhanced classification approach increases the prediction accuracies after gathering the numerous classifiers, and produces the categorization maps along with more consistent regions via incorporating the spatial information, while evaluated with pixel-based classification.

In [8] discussed that the decision fusion in kernel induced spaces for hyper spectral image classification. In this research scenario, to overcome the issue of individual global transformation for the multiclass task, the author suggested one against one strategy. It is fused with support vector machine with kernel based discriminate analysis approach. Each classifier in the current method is a binary kernel discriminant investigation scheme, the discriminant complexity is decreased from the case in which all classes are treat jointly. The expensive and computation time of the training process is radically reduced. Smaller number training examples are necessary for optimizing and training the method and this research technique is efficient even in the incidence of noise. In [9] presented that genetic algorithm-based classifiers fusion for multi sensor activity recognition. Activity acknowledgment of an aged person is utilized to produce information and intellectual services to health care proficient, careers, elderly group, and their relations so that the aged people can stay in residences separately. This research inspects the use and involvement of wrist-worn multi sensors for activity detection. In this research, the techniques establish that accelerometers are the most significant sensors and heart rate information is used to boost prediction of activities along with varied heart rates. This scenario proposes a genetic algorithm-based fusion weight collection algorithm which uses genetic algorithm to discover mixture weights.

In [10] presented dempster-shafer theory based classifier fusion for improved fingerprint verification performance. This research scenario provides a Dempster Shafer theory based classifier fusion approach to progress the performance of fingerprint confirmation. The enhanced fusion approach unites decision induced contest scores of minutiae, ridge, finger code and pore based fingerprint verification approaches. It produces an enhancement in the certification accuracy compared to the individual algorithms. And also the enhanced fusion algorithm outperforms when compared with existing fusion algorithms. This research scenario is also shown that the use of Dempster's rule decreases the execution time. In [11] presented that classifier fusion as combination methods for semantic indexing in video content. Classifier fusion is explored as a novel research area to develop identification reliability through taking into account the complementarity among classifiers, in specific for repeated semantic-based video content indexing and recovery. Several fusion approaches are improved according to the kind of information produced through every classifier as well as their training and version capability. This scenario provides the current method in classifier mixture. A new training method named as weighted ten folding based on ten folding principle is improved for combining classifier. This research scenario has shown the higher efficiency of different combination methods. In [12] suggested classifier ensembles optimization guided by population oracle. This scenario is introduced a hybrid approach named as multi objective optimization algorithm to increase the classification performance. The dynamic optimization algorithm [13], [14], [15] is focused on the reduction of computational complexity and error rate significantly.

All of the research works discussed above consists of both merits and demerits together in terms of the better classification result. This approach leads to efficient finding of the fraudulent behaviour in different way in terms of improved analysis.

3. Fuzzy integral based ensemble classifier fusion

Financial fraud detection is the most important task in the real world environment which is focussed by the many of researchers in terms of reduced misclassification cost. Ensemble classification is the most concerned task which is followed by the many of the research works to perform accurate classification. In our previous research work, Optimal Ensemble Architecture Selection using Fire Fly and dempster shafer based classification (OEAS_FFA-DSEC) is introduced which is proved that the accurate classification of the financial transaction patterns, so that malicious transaction can be found easily. This approach adapts the firefly mechanism to find the better classification result by selecting the most optimal ensemble classifier and its architecture by randomly changing the combination of classifier with its architecture and subset feature in every iteration. This work ensures the appropriate selection of the ensemble classifier that can generate most accurate classification result. After selection of ensemble classifier, ensemble result is obtained by applying the dempster shafer theory approach which combines the classifier result based on the confidence score calculated for each classifier.

Dempster shafer theory is computationally in efficient and cannot consider the relation between the individual base classifier. This problem is resolved in this work proposed two approaches which are evaluated and compared with other to find the better fusion result. The proposed approaches are mainly based in fuzzy integral methods and based on the fuzzy measures. Fuzzy integrals aim at searching for the maximal agreement between the real possibilities relating to objective evidence and the expectation g which defines the level of importance of a subset of sources. The types of measures are given as follows:

- fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFSSI)
- fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFSCI)

The above approaches can generate accurate ensemble result in terms of better identification of the classification result.

3.1. Fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFSSI)

Sugeno fuzzy integral combines objective evidence of a hypothesis with the prior expectation of the importance of the evidence to the hypothesis. If we introduce a measurable space (X, Ω) and a function h : X → [0,1] then the fuzzy integral over set A ⊆ Ω of the function h with respect to a fuzzy measure g is defined by:

$$\int h(x)\circ g(.) = \sup_{E \subseteq X} [\min(\min_{X \epsilon E} h(x), g(A \cap E))]$$

Calculation of the Sugeno fuzzy integral is unexpectedly easy if we reorder elements of a set X = {x1,..., xn }so that the condition: h(x1) >= h(x2)>= ... >=h(xn) is met. Then a fuzzy integral e with respect to the fuzzy measure g over X can be computed by:

$$e = \max_{i=1}^{n} [\min (h(x_i), g(A_i))]$$

where Ai = {x1 ,..., xi}. Note that when g is the λ- fuzzy measure, the values of g(Ai) can be computed recursively as:

$$g(A_1) = g(\{x_1\}) = g^1$$
$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1})$$

For pattern recognition applications, the function hk (xi) can be treated as a partial evaluation of the degree of belonging of the object A to class k , given by the classifier associated with a group of features i x . Sugeno integral can be successfully applied to any multi-sensor systems by fusing the classifiers outputs in order to provide more accurate classification rates.

3.2. Fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFSCI)

Choquet fuzzy integral, developed by Sugeno and Murofushi, provides an extension to the Lebesgue integral which Sugeno integral do not cover. The definition of the Choquet integral refers to the Choquet functional presented in a different context. The assumptions are the same as for the Sugeno integral. The Choquet integral over set A ⊆ Í of the function h : X → [0,1] with respect to a fuzzy measure *g* is defined by:

$$\int h(x)\circ g(.) = \int_0^{+\infty} g(A_\alpha) d\alpha$$

where Aα = {x | h(x) > α} . Calculation of the Choquet fuzzy integral is similar to the numerical methods of integral calculation:

$$e = \sum_{i=1}^{n} [h(X_i) - h(X_{i-1})]_{g_i}^{n}$$

where $h(x_0) = 0$ and $g_i^j = g(\{xi , xi+1 ,..., xj \})$. The Choquet integral reduces to the Lebesque integral for a probability measure when a probability density function is used for calculations. Comparative results from the fusion of handwritten word classifiers showed similar level of classification performance for both integrals.

The integral measures that are introduced in this work can fuse the multiple classifiers results in the accurate manner. The performance evaluation were conducted in the matlab simulation environment from which it is proved the choquet integral based approach provides better results than the sugeno method in terms of improved accuracy, and precision measure which is explained detailed in the following section.

ISSN (Online) : 2277-5374
ISSN (Print) : 2277-5366

4. Experimental results

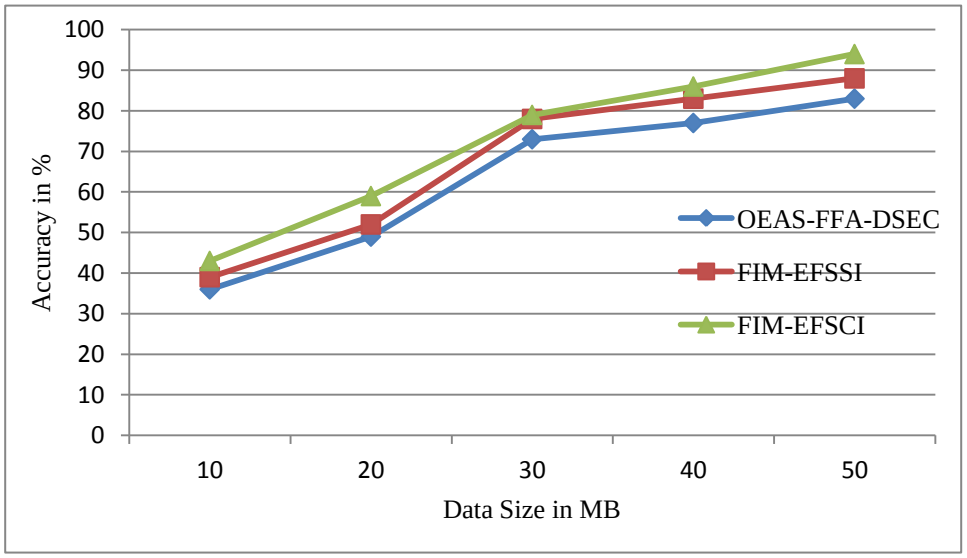
The experimental tests have been conducted in the MATLAB simulation environment to find whether the financial fraud detection occurs or not. The insurance data set would consist of different transaction behaviour of the users in the timely manner which is utilized for the accurate detection of the classification result. The performance evaluation is done between the existing methodology named Optimal Ensemble Architecture Selection using Fire Fly and dempster shafer based classification (OEAS_FFA-DSEC) and the proposed methodology named as fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFSSI) and fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFSCI). The comparison is made against the performance measures called the accuracy and precision. The performance comparison is given in the detailed manner as like follows:

4.1. Accuracy comparison

Accuracy is defined as the degrees of reduced misclassification error rate in terms of classifying different number of features present in the environment. Accuracy of the proposed research work should be more in the proposed research work in terms of reduced misclassification error rate than the existing approach. The accuracy comparisons of the proposed and existing methodologies are given as follows:

$$\text{Accuracy} = \frac{\text{Number of correctly classified data}}{\text{Total number of data}} \times 100$$

Figure 1. Accuracy comparison



The figure 1 graph depicts the performance comparison of the accuracy parameter value of both existing and proposed research scenario in terms of different data sizes. In the x axis, different data sizes are taken and in y axis accuracy value obtained while predicting the fraudulent behaviour is taken. From this graph, it is proved that the proposed research work have more accuracy than the existing approach in terms of efficient detection of fraudulent behaviour.

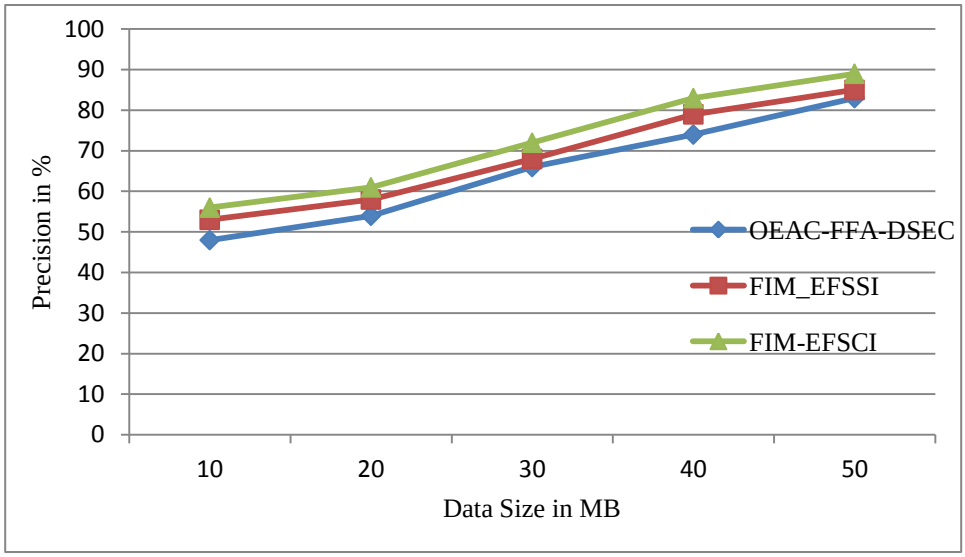
4.2. Precision comparison

Precision value is defined as the amount of correctly predicted result over a total number of predicted results. This parameter is used to indicate the overall performance improvement of the proposed methodology in terms of predicting accurate results. The precision value is calculated as like follows:

$$\text{Precision} = \frac{\text{Number of correctly classified data}}{\text{Total number of classified data's}} \times 100$$

ISSN (Online) : 2277-5374
ISSN (Print) : 2277-5366

Figure 2. Precision comparison



The figure 2 graph depicts the performance comparison of the precision parameter value of both existing and proposed research scenario in terms of different data sizes. In the x axis, different data sizes are taken and in y axis precision value obtained while predicting the fraudulent behaviour is taken. From this graph, it is proved that the proposed research work have more precision rate than the existing approach in terms of efficient detection of fraudulent behaviour.

5. Conclusion

Financial fraud detection is the most concerned task in the real world environment which needs to be done with more concern for avoiding the fraudulent behaviour. Ensemble classification is one of the better techniques for detecting the fraudulent patterns in the database. Better fusion can lead to efficient result in the ensemble classification approach. In this work, the fuzzy integral measure based ensemble fusion using sugeno integral (FIM-EFSSI) and the fuzzy integral measure based ensemble fusion using Choquet integral (FIM-EFSCI) is introduced which can fuse the multiple base classifier results in the efficient manner. The experimental tests were conducted in the matlab simulation environment from which it is proved that the FIM-EFSCI provides better result than the FIM-EFSSI in terms of improved accuracy and precision.

6. References

1. Fabio Augusto Faria, Jefersson Alex dos Santos, Anderson Rocha and Ricardo da S. Torres. Automatic Classifier Fusion for Produce Recognition. Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on 22-25 Aug. 2012.
2. Mayank Vatsa, Richa Singh, Afzel Noore, Arun Ross. On the dynamic selection of biometric fusion algorithms. *IEEE Transactions on Information Forensics and Security*. 2010; 5(3), 470-479.
3. Andy Jinhua Ma, Pong C. Yuen, Jian-Huang Lai. Linear dependency modeling for classifier fusion and feature combination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013; 35(5), 1135-1148.
4. Dominik Fisch, Edgar Kalkowski, Bernhard Sick. Knowledge fusion for probabilistic generative classifiers with data mining applications. *IEEE Transactions on Knowledge and Data Engineering*. 2014; 26(3), 652-666.
5. Ville Hautamäki, Tomi Kinnunen, Filip Sedlák, Kong Aik Lee, Bin Ma, Haizhou Li. Sparse classifier fusion for speaker verification. IEEE Transactions on Audio, Speech, And Language Processing. 2013; 21(8), 1622-1631.
6. Xi-Zhao Wang, Ran Wang, Hui-Min Feng, Hua-Chao Wang. A new approach to classifier fusion based on upper integral. IEEE Transactions on Cybernetics. 2014; 44(5), 620-635.
7. Xiaoshuang Ma, Huanfeng Shen, Jie Yang, Liangpei Zhang, Pingxiang Li. Polarimetric-Spatial Classification of SAR Images Based on the Fusion of Multiple Classifiers. IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing. 2014; 7(3), 961-971.
8. Wei Li, Saurabh Prasad, James E. Fowler. Decision Fusion in Kernel-Induced Spaces for Hyperspectral Image Classification. IEEE Transactions on Geoscience And Remote Sensing. 2014; 52(6), 3399-3411.

ISSN (Online) : 2277-5374
ISSN (Print) : 2277-5366

9. Saisakul Chernbumroong, Shuang Cang, Hongnian Yu. Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People. IEEE Journal of Biomedical And Health Informatics. 2015; 19(1), 282-289.
10. Richa Singh, Mayank Vatsa, Afzel Noore, Sanjay K. Singh. Dempster-Shafer Theory based Classifier Fusion for Improved Fingerprint Verification Performance. Computer Vision, Graphics and Image Processing. *The series Lecture Notes in Computer Science*. 2006; 4338, 941-949.
11. Rachid Benmokhtar, Benoit Huet. Classifier fusion: combination methods for semantic indexing in video content. Artificial Neural Networks – ICANN. The series Lecture Notes in Computer Science. 2006; 4132, 65-74.
12. Eulanda M. dos Santos, Robert Sabourin. Classifier ensembles optimization guided by population oracle. Evolutionary Computation (CEC), 2011 IEEE Congress on 5-8 June 2011.
13. A. Prakash, C. Chandrasekar. An optimized Multiple Semi-Hidden Markov Model. *Indian Journal of Science and Technology*. 2015; 8(2), 165-171.
14. Zahra Asheghi Dizaji, Farhad Soleimanian Gharehchopogh. A hybrid of ant colony optimization and chaos optimization algorithms approach for software cost estimation. *Indian Journal of Science and Technology*. 2015; 8(2), 128-133.
15. Reza Effatnejad, Fazlollah Rouhi. Unit commitment in power system t by combination of dynamic programming (DP), genetic algorithm (GA) and particle swarm optimization (PSO), *Indian Journal of Science and Technology*, 2015; 8(2), 134-141.

The Publication fee is defrayed by Indian Society for Education and Environment (iSee). www.iseeadyar.org

***Cite this article as**:*

C. Gayathri, R.Umarani. An effective and accurate fusion result from multi class ensemble classification.
Indian Journal of Education and Information Management. Vol 5 (1), January 2016.