A New Efficient Fuzzy Diversity Measure in Classifier Fusion

Abbas Golestani *Abbas_golestani@iust.ac.ir*
Javad Azimi *Ja_azimi@comp.iust.ac.ir*
Morteza Analoui *analoui@iust.ac.ir*

Computer faculty of Iran University of Science and Technology

Abstract. The combination of classifiers is one of the important and efficient algorithms in pattern recognition. In classifiers combination, the diversity rate among classifier outputs is one of the most important discussions. There are different methods calculating the diversity among classifiers. In this paper, we intend to find a new method for calculating the diversity among classifiers by using the fuzzy logic. We have checked this method over the total of different classifiers set and compare them with different methods. We will observe that according to the powerful credit of fuzzy discussions, this method has more advantages than the previous methods.

Keywords: Diversity, classifier ensembles, Fuzzy, Diversity Measures.

1 Introduction

The optimal classifier in every case is highly dependent upon to the problem domain. In practice, one might come across a case where no single classifier can achieve an acceptable level of accuracy. In such cases, it would be better to pool the results of different classifiers to achieve the optimal accuracy. Every classifier operates well on different aspects of the training or test feature vector. As a result, assuming appropriate conditions and combining multiple classifiers may improve classification performance when compared with any single classifier.

Instead of looking for the best set of features and the best classifier, now we look for the best set of classifiers and then the best combination method [1].

If we have a perfect classifier that makes no errors, then we do not need an ensemble. If, however, the classifier does make errors, then we seek to complement it with another classifier, which makes errors on different objects. The diversity of the classifier outputs is therefore a vital requirement for the success of the ensemble. Intuitively, we want the ensemble members to be as correct as possible, and in case they make errors, these errors should be on different objects.

Diversity is studied in two ways: Producing and measuring. In order to optimally and robustly integrate classifiers, one needs a diversity of component partitions for combination. Generally, this diversity can be obtained from several sources:

1) Using different classification algorithms to produce partitions for combination [2].

2) Using different features or feature extraction [3, 4, 5, 6].

3) Partitioning different subsets of the original data [7].

The diversity among classifiers combinations is an important part in combination of classifiers [8]. Having more diversity means that ensemble members are more different to each other. There are several methods for evaluating diversity in classifier ensembles [9].

In this paper we propose a new diversity measure based on fuzzy logic concepts. First we explain the concept of measure of fuzziness in fuzzy logic. Then we discuss about the efficiency of this parameter in fuzziness of a fuzzy set. At last, the fuzziness concept is used to introduce a new fuzzy diversity measure.

Simplicity, Symmetry and Accuracy are some characteristics of the proposed method. Experimental results demonstrate the effectiveness of proposed method in measuring diversity. The rest of the paper is organized as follows: Section 2 presents the diversity measures, Section 3 describes measure of fuzziness, in section 4 the proposed method has been introduced, in Section 5 the mathematical confirmation of proposed method has been illustrated and in section 6 an example has been introduced to exhibit the ability of proposed method.

2 Diversity measures

The diversity among the combination of classifiers is defined as: if one classifier has some errors, then for combination, we look for classifiers which have errors on different objects [10]. Among a set of classifiers, basically, there are two kinds of diversity measure methods, pairwise and nonpairwise.

2.1 Pairwise Measures

These measures, and the ones discussed hitherto, consider a pair of classifiers at a time. An ensemble of L classifiers will produce $\frac{L(L-1)}{2}$ pairwise diversity Values. To get a single value we average across all pairs.

2.1.1 The Disagreement Measures:

The disagreement measure is probably the most intuitive measure of diversity between a pair of classifiers. Table1 shows the classification probability of two classifiers *i* and *k* for a given sample. The probability that the two classifiers will disagree on their decisions is considered as the diversity between two classifiers [11, 12]. Therefore, $D_{i,k} = (b+c)$.

Table1. the confusion matrix of two classifier D_i , D_k

	D_k correct(1)	D_k wrong(0)
D_i correct(1)	A	B
D_i wrong(0)	c	d

2.2 Nonpairwise Measures

The measures of diversity introduced below consider all the classifiers together and Calculate directly one diversity value for the ensemble.

2.2.1 Entropy evaluation method:

For a particular $z_j \in Z$ when $\lfloor L/2 \rfloor$ of the votes are 0s (1s) and the other $L - \lfloor L/2 \rfloor$ votes are 1s (0s). If they all were 0s or all were 1s, there is no disagreement, and the classifiers cannot be deemed Diverse. One possible measure of diversity based on this concept is

$$E = \frac{1}{N} \frac{2}{L-1} \sum_{j=1}^{N} \min\{ (\sum_{i=1}^{L} y_{i,j}), (L-\sum_{i=1}^{L} y_{i,j}) \}$$

E varies between 0 and 1, where 0 indicates no difference and 1 indicates the highest possible diversity. Let all classifiers have the same individual accuracy p. Then while value 0 is achievable for any number of classifiers L and any *p*, the value 1 can only be attained for $p \in \left[\frac{L-1}{2L}, \frac{L+1}{2L} \right]$ [13].

This method has some limitations, for example the precision of all classifiers must be *p*.

Generally, there are 10 famous methods for measuring diversity which are shown in Table 2. The symbol (↑) means, increasing the output of respective methods shows more diversity and the symbol (↓) means, decreasing the output of respective methods shows more diversity.

Table 2. List of some popular diversity measures

Name	Notation	Type	**Pairwise or nonpairwise**
Q-statistic[14]	Q	↓	*Pairwise*
Correlation coefficient[15]	ρ	↓	*Pairwise*
Disagreement[16,17]	D	↑	*Pairwise*
Double-fault[18]	DF	↓	*Pairwise*
Kohavi-Wolpert variance[19]	Kw	↑	*Nonpairwise*
Measurement of interrater agreement [20]	κ	↓	*Nonpairwise*
Entropy[21]	Ent	↑	*Nonpairwise*
Measure of difficulty[22]	θ	↓	*Nonpairwise*
Generalized diversity[23]	GD	↑	*Nonpairwise*
Coincident failure diversity[24]	CFD	↑	*Nonpairwise*

3 Measure of fuzziness

Classical logic is based on binary logic with two values of truth. These two values are 1 and 0. Fuzzy logic is a multi-valued logic with truth represented by a value on the closed interval [0, 1]. Values in (0, 1) indicate varying degrees of truth which is named as membership function. The membership 0.5 means maximum uncertainty.

There are some relations for calculation of measure of fuzziness. One of them is known as the SHANON relation [25]:

$$(x_i , \mu_{(x_i)}) \rightarrow \sum_i \{- \mu_{(x_i)} \mathrm{Ln}(\mu_{(x_i)}) - (1 - \mu_{(x_i)}) \mathrm{Ln}(1 - \mu_{(x_i)})\} \qquad (1)$$

Where, x_i is a member of set and $\mu_{(x_i)}$ is its membership.

In above equation, the rate of fuzziness increases when the results of equation increase.

Another method for calculating the fuzziness is based on the uncertainty difference between a set and its complement. Mr. Yager has performed a method for calculation of fuzziness [25] which is shown as below:

$$D_P(\tilde{A} , \not\subset A) = [\sum_{i=1}^{n} \left| \mu_{A(x_i)} - \mu_{\not\subset A(x_i)} \right|^P]^{\frac{1}{P}} \quad , \quad P = 1,2,3,... \qquad (2)$$

Where A is the proposed set and n is the number of its members.

4 The Proposed Diversity Measure

We intend to perform a new diversity measure for two or more classifier combinations. This method has been defined on the concept of measure of fuzziness in fuzzy logic.

Suppose we have *n* classifiers and *m* classifiers can correctly classify a given sample. Therefore the membership $(\frac{m}{n})$ is set for the given sample $(0 \leq \frac{m}{n} \leq 1)$. The obtained membership is applied to SHANON function to measure its fuzziness which is studied as diversity:

$$(x_i , \mu_{(x_i)}) \rightarrow \sum_i \{- \mu_{(x_i)} \mathrm{Ln}(\mu_{(x_i)}) - (1 - \mu_{(x_i)}) \mathrm{Ln}(1 - \mu_{(x_i)})\}$$

Where $\mu_{(x_i)}$ is the membership function of sample x_i.

5 Mathematical confirmation

Since the purpose of classifiers combination is reaching to a higher precision, it is very vital to understand the diversity concepts in classifier combinations. The best combination of classifier is a combination whose classifiers fail in different samples which can obtain by diversity.

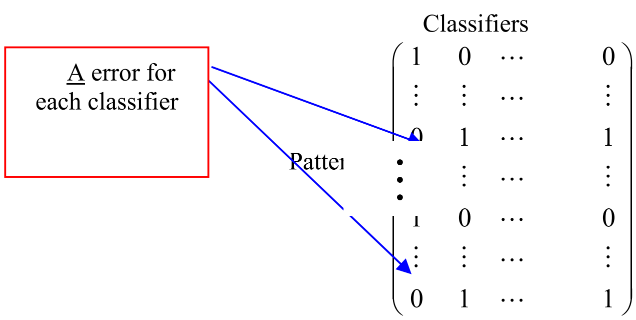
The function calculating the diversity parameter should be accurate.

In the following, the correct response for the new method is shown in two situations. At first we propose a method when the number of samples and the number of classifiers are equal. Then we propose a method when the number of samples is more than the number of classifiers.

5.1 The number of samples and the number of classifiers are equal

Suppose that:
The number of classifiers: m
Number of samples: m
Precision in each classifier: $\frac{m - A}{m}$

According to above hypothesis, the number of samples is equal with the number of classifiers. Since each classifier has *A* faults, the best case in diversity form is to have exactly *A* faults in each sample to create a minimum overlap among the classifier faults.



Therefore, we can calculate the diversity fuzziness method, for the above matrix.

$$\text{diversity1} = m \left[-\frac{m - A}{m} . \operatorname{Ln}\left(\frac{m - A}{m}\right) - \left(1 - \frac{m - A}{m}\right) . \operatorname{Ln}\left(1 - \frac{m - A}{m}\right) \right] \quad (3)$$

Now if the fault density increases in one of the samples in the simplest case, one of the samples has A+1 fault and another A-1 fault, the diversity is calculated by using the following fuzziness method:

$$\text{diversity2} = (m - 2) \left[-\frac{m - A}{m} . \operatorname{Ln}\left(\frac{m - A}{m}\right) - \left(1 - \frac{m - A}{m}\right) . \operatorname{Ln}\left(1 - \frac{m - A}{m}\right) \right] +$$
$$\left[-\frac{m - A + 1}{m} . \operatorname{Ln}\left(\frac{m - A + 1}{m}\right) - \left(1 - \frac{m - A + 1}{m}\right) . \operatorname{Ln}\left(1 - \frac{m - A + 1}{m}\right)\right] +$$
$$\left[-\frac{m - A - 1}{m} . \operatorname{Ln}\left(\frac{m - A - 1}{m}\right) - \left(1 - \frac{m - A - 1}{m}\right) . \operatorname{Ln}\left(1 - \frac{m - A - 1}{m}\right)\right] \quad (4)$$

If we prove that the diversity1 is bigger than diversity2, means that the new method calculates the diversity correctly. The numbers of n-2 from samples have similar diversity, and only two samples differentiate with each other that inequality is as follow:

$$2 . f\left(\frac{m - A}{m}\right) > f\left(\frac{m - A + 1}{m}\right) + f\left(\frac{m - A - 1}{m}\right)$$
$$\left(f(x) = -x . \operatorname{Ln}(x) - (1 - x) . \operatorname{Ln}(1 - x) \right) \quad (5)$$

According to the following figure, since the f(x) function is convex, the above inequality is resulted.

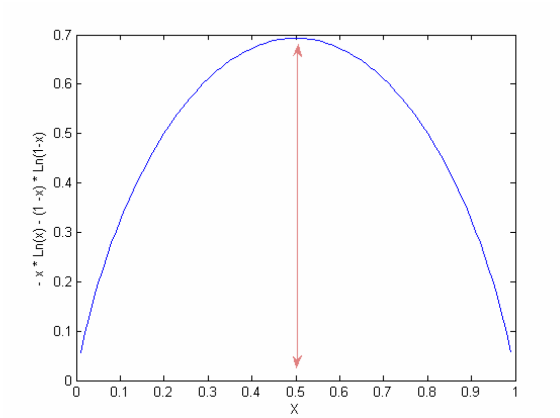


Figure1: The formula, which was used in previous method, has some features which we can use this function based on x axis values between zero and one. Thus we see this function has the maximum amount in amount of $\mu_{(x_i)} = 0.5$

5.2 the number of samples is more than classifiers:

Suppose that:
The number of classifiers = n
The number of samples = m = k (nA)
Precision in each classifier = $\frac{m - A}{m}$

According to above hypothesis, the number of samples is equal with the factor of all classifier faults. Since each classifier has A fault according to the diversity, the base case is having exactly k faults in each sample to have the minimum overlap among the classifiers faults. Therefore, this is like part 1 which was proven.

6 Examples

The following table shows all the possible combinations, correct or incorrect classifications, of ten samples by three classifiers. The accuracy of each classifier is set as 0.6; therefore each classifier can only correct 6 samples.

In the first line, the case of each column is consisted of a binary number by the form of “010" which any of these three numbers is a classifier output. The value 1 shows the correct classification and vise versa. For example the number 3 in column b, which is labeled as 101, means that there are 3 samples which both classifier 1 and 3 classify them correctly and classifier 2 classifies them incorrectly. Therefore by applying simple majority voting to the results, these 3 samples classifies correctly. The column which is labeled as P_{maj} shows the accuracy of majority voting [26]. The majority voting algorithm classifies the samples of column *a, b, c and e* correctly. The last column of table 2 shows the value of $P_{maj} - P$, that *P* is the accuracy of each classifier (0.6). In rows 1 to 12 the value of $P_{maj} - P$ is greater than 0 which means that classifier ensemble improves final accuracy. In rows 13 to 28 the value of $P_{maj} - P$ is less than 0 which means that classifier ensemble can not improve final accuracy.

Table2: possible combinations of correct or incorrect classifications of 10 samples by three classifiers.

No.	***111 a***	***101 b***	***011 c***	***001 D***	***110 e***	***100 F***	***010 g***	***000 H***	P_{maj}	P_{maj} - P
1	0	3	3	0	3	0	0	1	0.9	0.3
2	2	2	2	0	2	0	0	2	0.8	0.2
3	1	2	2	1	3	0	0	1	0.8	0.2
4	0	2	3	1	3	1	0	0	0.8	0.2
5	0	2	2	2	4	0	0	0	0.8	0.2
6	4	1	1	0	1	0	0	3	0.7	0.1
7	3	1	1	1	2	0	0	2	0.7	0.1
8	2	1	2	1	2	1	0	1	0.7	0.1
9	2	1	1	2	3	0	0	1	0.7	0.1
10	1	2	2	1	2	1	1	0	0.7	0.1
11	1	1	2	2	3	1	0	0	0.7	0.1
12	1	1	1	3	4	0	0	0	0.7	0.1
13	6	0	0	0	0	0	0	4	0.6	0.0
14	5	0	0	1	1	0	0	3	0.6	0.0
15	4	0	1	1	1	1	0	2	0.6	0.0
16	4	0	0	2	2	0	0	2	0.6	0.0
17	3	1	1	1	1	1	1	1	0.6	0.0
18	3	0	1	2	2	1	0	1	0.6	0.0
19	3	0	0	3	3	0	0	1	0.6	0.0
20	2	1	1	2	2	1	1	0	0.6	0.0
21	2	0	2	2	2	2	0	0	0.6	0.0
22	2	0	1	3	3	1	0	0	0.6	0.0
23	2	0	0	4	4	0	0	0	0.6	0.0
24	5	0	0	1	0	1	1	2	0.5	-0.1
25	4	0	0	2	1	1	1	1	0.5	-0.1
26	3	0	1	2	1	2	1	0	0.5	-0.1
27	3	0	0	3	2	1	1	0	0.5	-0.1
28	4	0	0	2	0	2	2	0	0.4	-0.2

Now, we use the different method of the diversity calculation for the mentioned states in above table and compare them with the new method. It must be noticed that, this method is from ($\uparrow$) kind and has a good conformity with diversity among the classifiers. If we calculate the average of diversity values between the diversity values in two parts, the precision has increased and the precision has decreased, by the new fuzziness method and GD which is the best reported method, we can see the improvement of the new method in measuring the diversity.

Table3: comparison between diversity calculation methods

NO.	*D*	*DF*	*KW*	κ	*E*	θ	*GD*	*CFD*	*Fuzziness*	
1	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.90	5.72862	
2	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.75	3.81908	
3	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.83	5.09211	
4	0.67	0.07	0.22	-0.39	1.00	0.02	0.83	0.90	6.36514	
5	0.67	0.07	0.22	-0.39	1.00	0.02	0.83	0.90	6.36514	
6	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.50	1.90954	
7	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.64	3.18257	1
8	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	4.45559	
9	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	4.45559	
10	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	5.72862	
11	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	5.72862	
12	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	5.72862	
13	0.00	0.40	0.00	1.00	0.00	0.24	0.00	0.00	0.00000	
14	0.13	0.33	0.04	0.72	0.20	0.20	0.17	0.30	1.27302	
15	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	2.54605	
16	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	2.54605	
17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	3.81908	
18	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	3.81908	
19	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	3.81908	
20	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	5.09211	
21	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	5.09211	2
22	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	5.09211	
23	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	5.09211	
24	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.30	1.90954	
25	0.33	0.27	0.11	0.31	0.50	0.13	0.42	0.50	3.18257	
26	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	4.45559	
27	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	4.45559	
28	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.50	3.81908	

Fuzziness method:

The average of diversity number in part 1 = 58.56

The average of diversity number in part 2 = 56.01

$$\rightarrow \quad \frac{58.56}{56.01} = 1.04552$$

GD method:

The average of diversity number in part 1 = 7.66

The average of diversity number in part 2 = 7.34

$$\rightarrow \quad \frac{7.66}{7.34} = 1.43596$$

Here we see that in this ratio, the amount of high diversity to high precision in fuzziness method is more than GD.

There are two parameters about the diversity measures which make one method more favor than others, simplicity and symmetry. The existing simplicity in these methods means that the rate and concept of diversity calculation are very easy and understandable. The symmetry means that if all classifiers change their decisions, for example in binary classification, the diversity value does not change. It can be seen that the proposed method has mentioned characteristics, simplicity and symmetry.

7. Conclusion

Since the diversity among the classifiers is a vital parameter, in this paper we introduced a new fuzzy method measuring the diversity among classifiers outputs based on Shanon formula. Accuracy, simplicity and symmetry are the characteristics of the proposed method. Experimental results demonstrated that the proposed method calculates the diversity values more accurate than previous popular methods, especially GD. The mathematical confirmation and a simple example introduced in this paper show the ability of proposed method in contrast to other methods.

References

1. T. K. Ho. Multiple classifier combination: Lessons and the next steps. In A. Kandel andH. Bunke, editors, Hybrid Methods in Pattern Recognition. World Scientific Publishing, 2002, pp. 171–198.
2. L.I. Kuncheva, (2005) combining pattern classifier wiley
3. A. Strehl & J. Ghosh,. Cluster ensembles—a knowledge reuse framework for combining partitionings, in: Proc. of 11-th National Conf. on Artificial Intelligence, Edmonton, Alberta, Canada (2002) (pp. 93–98).

4. Cunningham, P., Carney, J., (2000) Diversity versus Quality in Classification Ensembles Based on Feature Selection, Trinity College Dublin, Computer Science Technical Report: TCD-CS-2000-02, submitted to ECML-2000.
5. Guerra-Salcedo, C., Whitley, D., (1999b). Feature Selection Mechanisms for Ensemble Creation: A Genetic Search Perspective, in Data Mining with Evolutionary Algorithms.
6. Opitz, D. (1999), Feature selection for ensembles, *in* 'Proc. of the 16th National Conf. on Artificial Intelligence', AAAI Press, pp. 379–384.
7. Freund, Y. & Schapire,R. E. (1996), Experiments with a new boosting algorithm, in L. Saitta, ed.,`Machine Learning: Proceedings of the Thirteenth National Conference',Morgan Kaufmann, pp. 148{156.
8. L. Lam. Classifier combinations: implementations and theoretical issues. In J. Kittler and F. Roli, editors, Multiple Classifier Systems, volume 1857 of lecture Notes in Computer Science, pages 78-86, Cagliari, Italy, 2000. Springer.
9. Ludmila I.Kuncheva and Christopher J.Whitaker. Measures of diversity in classifier ensembles.
10. L.I. Kuncheva, (2005) combining pattern classifier *wiley*
11. T. K. Ho. The random space method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8):832–844, 1998.
12. P. Cunningham and J. Carney. Diversity versus quality in classification ensembles based on feature selection. Technical Report TCD-CS-2000-02, Department of Computer Science, Trinity College, Dublin, 2000.
13. P. Cunningham and J. Carney. Diversity versus quality in classification ensembles based on feature selection. Technical Report TCD-CS-2000-02, Department of Computer Science, Trinity College, Dublin, 2000.

14. S. W. Looney. A statistical technique for comparing the accuracies of several classifiers. Pattern Recognition Letters, 8:5–9, 1988.
15. J. L. Fleiss. Statistical Methods for Rates and Proportions. John Wiley & Sons, 1981.
16. G. S. Sebestyen. Decision-Making Process in Pattern Recognition. The Macmillan Company, N.Y., 1962.
17. R. P. W. Duin. PRTOOLS(Version 2). A Matlab toolbox for pattern recognition. Pattern Recognition Group, Delft University of Technology, June 1997.
18. N. J. Nilsson. Learning Machines. McGraw-Hill, New York, 1962.
19. J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3):226–239, 1998.
20. M. Minsky and S. Papert. Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge, Mass, 1969.
21. G. T. Toussaint. Bibliography on estimation of misclassification. IEEE Transactions on Information Theory, 20:472–479, 1974.
22. E. A. Patrick. Fundamentals of Pattern Recognition. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1972.
23. A. K. Jain and J. Mao. Guest editorial: Special issue on artificial neural networks and statistical pattern recognition. IEEE Transactions on Neural Networks, 8(1):1–3, 1997.
24. R. P. W. Duin. A note on comparing classifiers. Pattern Recognition Letters, 17:529– 536, 1996.

25. L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their application to handwriting recognition. IEEE Transactions on Systems, Man, and Cybernetics, 22:418–435, 1992.