A Ranking Distance Based Diversity Measure for Multiple Classifier Systems

Yi Yang
SKLSVMS
School of Aerospace
Xi'an Jiaotong University
Xi'an, Shaanxi, China 710049
jiafeiyy@mail.xjtu.edu.cn

Deqiang Han
Institute of Integrated Automation
School of Electronic and Information Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China 710049
deqhan@gmail.com

Jean Dezert
ONERA
The French Aerospace Lab
Chemin de la Hunière
F-91761 Palaiseau, France
jean.dezert@onera.fr

***Abstract*—Multiple classifier fusion belongs to the decision-level information fusion, which has been widely used in many pattern classification applications, especially when the single classifier is not competent. However, multiple classifier fusion can not assure the improvement of the classification accuracy. The diversity among those classifiers in the multiple classifier system (MCS) is crucial for improving the fused classification accuracy. Various diversity measures for MCS have been proposed, which are mainly based on the average sample-wise classification consistency between different member classifiers. In this paper, we propose to define the diversity between member classifiers from a different standpoint. If different member classifiers in an MCS are good at classifying different classes, i.e., there exist expert-classifiers for each concerned class, the improvement of the accuracy of classifier fusion can be expected. Each classifier has a ranking of classes in term of the classification accuracies, based on which, a new diversity measure is implemented using the ranking distance. A larger average ranking distance represents a higher diversity. The new proposed diversity measure is used together with each single classifier's performance on training samples to design and optimize the MCS. Experiments, simulations, and related analyses are provided to illustrate and validate our new proposed diversity measure.**

***Index Terms*—Multiple classifier system (MCS); multiple classifier fusion; diversity; ranking distance; pattern classification**

I. Introduction

To handle pattern classification problems in a complicated environment, a single classifier is usually incompetent. The multiple classifier system (MCS) [1] theory and method have been proposed to build multiple classifiers and then aggregate their outputs for the final decision-making. In machine learning community, MCS belongs to the ensemble learning. MCS can also be considered as a decision-level information fusion. Over the past decade, MCSs have been actively exploited for improving classification accuracy and reliability over individual classifiers. MCSs have been widely used in areas such as the handwriting character recognition [2], [3], biometric identification [4], remote sensing [5], fault diagnosis [6], network security [7] and automatic object recognition [8].

To implement an MCS, one should generate multiple individual member classifiers first. Note that using multiple classifiers cannot assure the improvement of classification accuracy in general. It would be meaningless to combine multiple redundant classifiers. The complementarity among member classifiers is crucial for the improvement of classification accuracy. Such a complementarity is called diversity [9] in the field of MCS. The diversity can be implemented qualitatively, e.g., using different samples, different feature spaces (or subspaces), different types of classifiers, and different parameter settings for classifiers to generate different member classifiers, and thus expect to obtain "larger" diversity. For the convenience of practical use, the diversity is expected to be implemented quantitatively.

Diversity measures have already become a research focus in the field of MCSs, and various diversity measures have been proposed so far. In 2005, the journal "Information Fusion" published a special issue on "Diversity Measure in Multiple Classifier Systems", paying a special attention to definitions of diversity measures (e.g., Q-statistics, Double Fault, Difficulty, Correlation Coefficient, Disagreement, etc) in terms of their prediction ability of the combining performance [9]. Fan et al. [10] proposed a new diversity measure for the classifiers with soft label output in 2008. In 2009, Trawinski et al. [11] jointly use the diversity measure and the classification accuracy for MCSs. In 2011, Nascimento et al. [12] proposed an approach to jointly use the available diversity measures. Haghighi et al. [13] used the support vector data descriptor (SVDD) to implement a diversity measure for MCSs. In 2013, we proposed a dynamic diversity measure and modelled the MCS with the theory of belief functions [14]. In 2014, Krawczyk et al. [15] proposed the diversity measure for the single-class member classifiers. Diez-Pastor et al. [16] studied diversity measures for the MCS given the unbalanced data set. In 2016, Kadkhodaei et al. [17] proposed an entropy-based diversity measure. In 2016, Cavalcanti et al. [18] also combined diversity measures for the MCSs. Till now, almost all the available diversity measures are based on the average sample-wise classification consistency between different member classifiers, and there is no prominent relation between the diversity and MCS accuracy.

In this paper, we attempt to design the MCS diversity measure from a different standpoint. We think that if different member classifiers are good at classifying different classes, then the corresponding MCS is diverse. In another word, there

exists the corresponding expert-classifier(s) for the different class(es). Note that each classifier has its own ranking of classes in terms of the classification accuracies, therefore, we implement the diversity measure for MCSs using the ranking distance. A larger average ranking distance represents a higher diversity in an MCS, since it means that different member classifiers are likely to be "experts" of different classes. Our new proposed diversity measure is illustrated and validated by experiments, simulations, and related analyses provided.

II. Basics of MCSs and Diversity

The construction of an MCS includes the generation of member classifiers and the combination of the member classifiers' outputs.

Various approaches to generating member classifiers have been proposed, e.g., using different training samples, different feature spaces (or subspaces), and different types of classifiers, etc. The specific combination method for the MCS depends on the output types of individual classifiers. Suppose that a query sample is $\boldsymbol{x}_q \in R^d$. We consider that the class space is $\{c_i,\, i = 1, \ldots, C\}$. The output of a member classifier can be categorized into three types [2]:

1) Abstract Level: the classifier produces a unique class label for $\boldsymbol{x}_q$. Classifier e_k assigns a class label j_k to sample $\boldsymbol{x}_q$, i.e., $e_k(\boldsymbol{x}_q) = j_k$, $k = 1, 2, ..., n$, $j_k \in \{c_1, ..., c_C\}$.

2) Rank Level: the classifier e_k ranks all possible labels and outputs a rank Λ_k with the label at top being the first choice.

3) Measurement Level: the classifier assigns each label a measurement value such as *a posterior* probability or membership function value. For $\boldsymbol{x}_q$, each member classifier e_k brings out an output vector $[\omega_k(c_1), \omega_k(c_2), ..., \omega_k(c_M)]$, where $\omega_k(c_i) \in [0, 1]$ can be considered as the membership function for the given query sample belonging to class c_i.

If the outputs are the abstract level, one can use the voting rules to combine member classifiers; if the outputs are the rank level, one can use the voting rules and ranking aggregation rules to combine them; if the outputs are the measurement level, one can use various rules including voting rules, Behavior Knowledge Space (BKS) [19], fuzzy logic and the theory of belief functions to combine according to the outputs' specific representation (e.g., the probability, membership function or belief function) at the measurement level.

MCS cannot assure the improvement of the classification accuracy in general. Large diversity is a necessary condition for improving the classification performance. Using different ways to generate member classifiers can be considered as qualitative ways to implement the diversity for MCSs. To design diversity measures is the quantitative way.

Diversity measures quantify the diversity or complementarity among member classifiers. Available diversity measures can be categorized into two major types [15]:

1) Pairwise measures: Pairwise measures are calculated between two member classifiers. Table I shows the joint counts N_{ij}^{ab} of two classifiers e_i and e_j. For example N_{ij}^{01} denotes that e_i obtains an incorrect result and e_j obtains a correct result. Here, subscript ij for N has been omitted for the simplicity. Some representative pairwise diversity measures, e.g., the Q-statistic (Q), correlation coefficient (R), disagreement measure (D) and double-fault measure (*DF*), are recalled in (1)–(4).

TABLE I
THE JOINT COUNTS FOR OUTPUTS OF TWO CLASSIFIERS

	e_j *correct* (1)	e_j *incorrect* (0)
e_i *correct* (1)	N^{11}	N^{10}
e_i *incorrect* (0)	N^{01}	N^{00}

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \tag{1}$$

$$R_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11}+N^{10})(N^{01}+N^{00})(N^{11}+N^{01})(N^{10}+N^{00})}} \tag{2}$$

$$D_{i,j} = \frac{N^{01} + N^{10}}{N^{11} + N^{00} + N^{01} + N^{10}} \tag{3}$$

$$DF_{i,j} = \frac{N^{00}}{N^{11} + N^{00} + N^{01} + N^{10}} \tag{4}$$

For an ensemble of L classifiers, the averaged diversity measure over all classifiers is given by

$$Diversity_{ave} = \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} Diversity_{i,j} \tag{5}$$

where $Diversity_{i,j}$ can be either $Q_{i,j}$, $R_{i,j}$, $D_{i,j}$ or $DF_{i,j}$.

2) Non-pairwise measures: Non-pairwise measures are calculated directly over all member classifiers. They can be calculated using the proportion of classifiers that misclassify randomly selected samples. A non-pairwise measure (Entropy measure) is [20]

$$E = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{(L - \lceil L/2 \rceil)} \min\{l(x^j), L - l(x^j)\} \tag{6}$$

where L is the number of classifiers, N is the number of training samples, $\lceil \cdot \rceil$ is the ceiling function and $l(x^j)$ represents the number of classifiers that correctly classify the sample x^j. If for all samples, all classifiers agree, then E reaches its minimum value of 0. If for each sample x^j, $l(x^j)$ is close to $L - l(x^j)$, i.e., about half classifiers are not accordant to their counterparts, then E is close to its maximum value 1.

As we see, traditional diversity measures are usually designed using the classification results on training samples, i.e., the consistency or inconsistency of the classification results are used to establish the diversity measures. They are defined in a statistical sense. A good diversity measure should have the relation with the MCS accuracy as shown in Fig. 1.

As argued by both Windeatt et al. [20] and Didaci et al. [21], there is no prominent relation between existing diversity measures and the MCS accuracy. Traditional ways to define diversity measures might not lead to a fertile, but to a dead-end. Therefore, we proose the MCS diversity measure from a different viewpoint, and expect to obtain good properties.

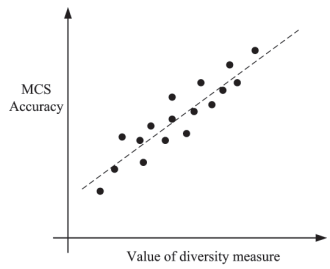


Fig. 1. Expected relation between MCS accuracy and the diversity measure.

III. New Diversity Measure Based on Ranking Distance

If different member classifiers in an MCS are good at classifying different classes, then the MCS is diverse. That is, the different class has its corresponding expert-classifiers as the MCS_A illustrated in Fig. 2, then the MCS is more diverse, and the improvement of the classification accuracy can be expected; if all the member classifiers have the same expert-class as the MCS_B illustrated in Fig. 2, then the MCS is less diverse, and has less potential to improve the classification accuracy. According to such an idea, we propose a new diversity measure for MCSs. Suppose that e_i is a

Member classifiers	MCS_A			MCS_B		
	Class-wise Ranking			Class-wise Ranking		
	1st	2nd	3rd	1st	2nd	3rd
Classifier1	●	●	●	●	●	●
Classifier2	●	●	●	●	●	●
Classifier3	●	●	●	●	●	●

● Class1 ● Class2 ● Class3

Fig. 2. Classification Accuracy Ranking.

member classifier, where $i = 1, ..., L$. There are M classes in the classification task. Given a training set S, the class-wise classification accuracy ranking of e_i is denoted by Λ_i. For the L member classifiers, there exist corresponding L rankings. Given two member classifiers e_i, e_j and their corresponding rankings Λ_i and Λ_j, their ranking distance can be calculated. E.g., one can choose the commonly used Spearman distance[1] as [23]:

$$\rho(\Lambda_i, \Lambda_j) = 2 - \frac{6 \cdot \sum_{k=1}^{M} \left(\Lambda_i(k) - \Lambda_j(k)\right)^2}{M(M^2-1)} \tag{7}$$

where M denotes the number of items to rank (the number of classes). Clearly $\rho \in [0, 2]$. $\rho = 2$ means a total positive correlation between the ranks, while $\rho = 0$ means a total negative one.

If the distance between Λ_i and Λ_j is larger, then e_i and e_j's expert-class are more different. The average distance between all the member classifiers in an MCS is large, then the expert-classes are more diverse. Therefore, the ranking distance based diversity measure is defined as

[1]There are also other ranking distance definitions such as the footrule distance, Kendall distance, etc (see [22] for details). In this paper, we use Spearman distance, since it is commonly used and easy to implement.

$$Div = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L} \rho(\Lambda_i, \Lambda_j)}{C_L^2} = \frac{1}{L(L-1)} \sum_{i=1}^{L} \sum_{j=1, j\neq i}^{L} \rho(\Lambda_i, \Lambda_j) \tag{8}$$

For example, the class-wise classification accuracy ranking of Classifier 1 in MCS_A is $\Lambda_1^A = [1, 3, 2]$, which means that the accuracy on class 1 is the highest (1st place); the accuracy on class 2 is the lowest (3rd place); the accuracy on class 3 takes the 2nd place. The class-wise classification accuracy ranking of Classifier 2 in MCS_A is $\Lambda_2^A = [2, 1, 3]$, which means that the accuracy on class 1 takes the 2nd place; the accuracy on class 2 takes the 1st place; the accuracy on class 3 takes the 3rd place; The class-wise classification accuracy ranking of Classifier 3 in MCS_A is $\Lambda_3^A = [3, 2, 1]$, which means that the accuracy on class 1 takes the 3rd place; the accuracy on class 2 takes the 2nd place; the accuracy on class 3 takes the 1st place. Therefore, the average ranking distance in MCS_A is

$$\begin{aligned} Div(MCS_A) &= \frac{\rho\left(\Lambda_1^A, \Lambda_2^A\right) + \rho\left(\Lambda_1^A, \Lambda_3^A\right) + \rho\left(\Lambda_2^A, \Lambda_3^A\right)}{C_3^2} \\ &= \frac{1}{3}\left[1.5 + 1.5 + 1.5\right] = 1.5 \end{aligned}$$

The class-wise classification accuracy ranking of Classifier 1 in MCS_B is $\Lambda_1^B = [2, 1, 3]$; the class-wise classification accuracy ranking of Classifier 2 in MCS_B is $\Lambda_2^B = [3, 1, 2]$; the class-wise classification accuracy ranking of Classifier 3 in MCS_B is $\Lambda_3^B = [2, 1, 3]$. Therefore, the average ranking distance in MCS_B is

$$\begin{aligned} Div(MCS_B) &= \frac{\rho\left(\Lambda_1^B, \Lambda_2^B\right) + \rho\left(\Lambda_1^B, \Lambda_3^B\right) + \rho\left(\Lambda_2^B, \Lambda_3^B\right)}{C_3^2} \\ &= \frac{1}{3}\left[0.5 + 0.0 + 0.5\right] = 1/3 \end{aligned}$$

Therefore, we have $Div(MCS_A) > Div(MCS_B)$. This is intuitive, since each member classifier in MCS_A has the different "expert-class", while in MCS_B, the different member classifier has the same "expert-class" (class 2 here).

Although the diversity is crucial, it is only a necessary condition but not a sufficient condition for MCSs' improvement of classification performance. That is, only a larger diversity can not assure a better performance. If the MCS has higer diversity and at the same time the member classifiers have high classification accuracies, then the higher fusion-based classification accuracy can be expected. Therefore, one can construct an MCS by jointly using the diversity and individual classification accuracy as illustrated in Fig. 3.

$e_{O_i}(\cdot)(i = 1, ..., V)$ in Fig. 3 represent the overproduced individual classifiers based on training samples, and $e_j(\cdot)$ $(j = 1, ..., L)$ represent the member classifiers selected out of the overproduced ensemble. Here the "overproduce" means that the number of the produced classifiers is no less than that of classifiers chosen for constructing the MCS. The selection of member classifiers is converted into an optimization problem whose objective function is based on the joint use of

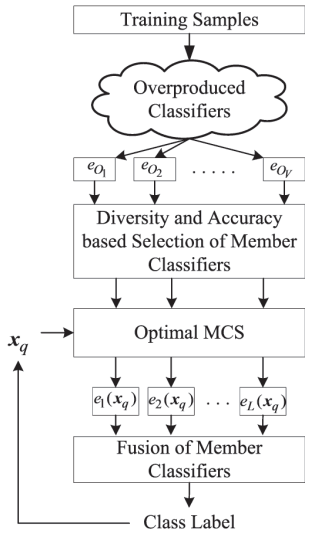


Fig. 3. Implementation of MCS based on diversity.

the proposed diversity measure and the average classification accuracy of the selected individual classifiers:

$$Obj(MCS) = w_D \cdot Div(MCS)/2 + w_A \cdot Acc_{ave}(MCS) \quad (9)$$

where $Acc_{ave}(MCS)$ is the average classification accuracy of the member classifiers in an MCS. w_D, w_A are the weights of the diversity and average accuracy, respectively. Note that the ranges of $Div(MCS)/2$ and that of the accuracy are both $[0, 1]$. The weighting parameters selection depends on the users' preference. $w_D = w_A = 1$ is suggested indicating an equal-treat attitude. One can use some optimization algorithm to find the best MCS by maximizing $Obj(MCS)$ in (9):

$$MCS_{Best} = \arg\max_{MCS}\{Obj(MCS)\} \quad (10)$$

MCS_{Best} obtained is with high diversity and simultaneously with high average accuracy of the member classifiers included.

Note that when the member classifiers are generated based on different feature subspaces, the classification accuracy can also be replaced by the discriminability defined as [24]

$$J = tr(S_w)/tr(S_b) \quad (11)$$

where tr denotes the trace of a matrix. Suppose that there are C classes and each class c_i has N_i samples.

Here x is a feature (vector) of a sample and

$$M = \frac{1}{C}\sum_{i=1}^{C}\left(\frac{1}{N_i}\sum_{x \in c_i} x\right) \quad (12)$$

is the mean of all the classes' centroids. If J of some feature (or set of features) in Eq. (11) is smaller, then such a feature (or set of features) is crisper and more discriminable.

Here, an illustrative example is provided to show the selection of member classifiers. Suppose that there are four candidate classifiers as shown in Fig. 4.

Individual classifiers	Class-wise Ranking			Classification Accuracy
	1st	2nd	3rd	
Classifier1	Class1	Class3	Class2	70%
Classifier2	Class2	Class1	Class3	80%
Classifier3	Class3	Class2	Class1	70%
Classifier4	Class2	Class1	Class3	70%

Class1 Class2 Class3

Fig. 4. Individual classifiers for member classifier selection.

TABLE II
MEMBER CLASSIFIERS SELECTION

MCS	Ranking Dist. $Div(MCS)/2$	Average Accu. $Acc_{ave}(MCS)$	$Obj(MCS)$
Classifiers {1, 2, 3}	0.75	0.73	**1.48**
Classifiers {1, 2, 4}	0.50	0.73	1.23
Classifiers {1, 3, 4}	0.75	0.70	1.45
Classifiers {2, 3, 4}	0.50	0.73	1.23

Assume that three member classifiers will be selected to construct an MCS. Then, there will be $C_4^3 = 4$ possible MCSs. Their objective function values are shown in Table II.

According to Eq. (9) and Eq. (10), classifiers 1, 2, and 3 are selected as member classifiers to construct an MCS, which has larger diversity and at the same time has higher average classification accuracy, therefore, better fusion-based classification accuracy can be expected.

IV. EXPERIMENTS

Experiments based on artificial datasets are provided to verify our proposed diversity measure and MCS construction.

A. On 1-D artificial dataset with Gaussian distribution

A three-class artificial dataset is generated. Each class has 100 samples. We generate five feature subspaces for the samples, where each feature subspace has one dimensions with Gaussian distribution. The five subspaces are shown in Fig. 5.

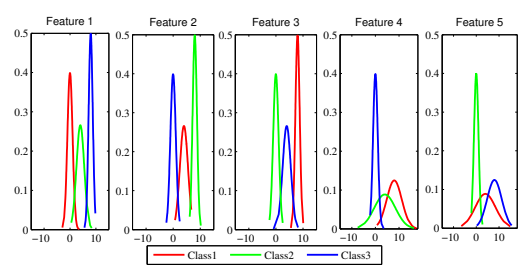


Fig. 5. 1-D artificial dataset.

The Gaussian distribution parameter is listed in Table III.

As we can see in Fig. 5, Feature 1 can well discriminate Class 1 and Class 3, while it cannot well discriminate Class 2 from both Class 1 and Class 3; Feature 2 can well discriminate Class 2 and Class 3, while it cannot well discriminate Class 1 from both Class 2 and Class 3; Feature 3 can well discriminate

TABLE III
DISTRIBUTION PARAMETERS OF DIFFERENT FEATURES

Feature		Mean	Std
1	Class 1	0.0	1.0
	Class 2	4.0	1.5
	Class 3	8.0	0.8
2	Class 1	4.0	1.5
	Class 2	8.0	0.8
	Class 3	0.0	1.0
3	Class 1	8.0	0.8
	Class 2	0.0	1.0
	Class 3	4.0	1.5
4	Class 1	8.0	3.2
	Class 2	4.0	4.5
	Class 3	0.0	1.0
5	Class 1	4.0	4.5
	Class 2	0.0	1.0
	Class 3	8.0	3.2

Class 1 and Class 3, while it cannot well discriminate Class 2 from both Class 1 and Class 3. The feature discriminability calculated based on Eqs. (11) is show in Table IV.

TABLE IV
FEATURE DISCRIMINABILITY FOR 1D DATA

Feature	J
1	0.1268
2	0.1332
3	0.1100
4	0.8926
5	1.0782

We use the k-nearest neighbors (k-NN) as the individual classifier. Suppose that there are M classes. For a test sample, find its k nearest neighbors. In k nearest neighbors, calculate the ratio of the each class's samples, respectively, as:

$$P(c_i) = kk(i)/\sum_{j=1}^{M} kk(j) \tag{13}$$

where $P(c_i)$ represents the ratio of class i and $kk(i)$ represents the number of samples belonging to class i in the k nearest neighbors, $i = 1, 2, ..., M$.. Obviously, $k = \sum_{j=1}^{M} kk(j)$. There are five individual classifiers according to the five different features. Three member classifiers are selected to form an MCS. The fusion rule of the MCS is

$$P^f(c_i) = \sum_{m=1}^{3} P^m(c_i)/3 \tag{14}$$

where P^f is the classification probability obtained based on fusion, and P^m is the member classifier m's classification probability. The class who takes the maximum probability is assigned to the query sample.

5-fold cross-validation is used here for evaluation. The average classification accuracy and the corresponding objective function value in Eq. (8) are calculated. Suppose that the number of member classifiers in an MCS is 3, the results (including the best MCS, the worst MCS and the middle one's corresponding accuracy and value of the objective function) are listed in Table V. Here member classifier i uses feature i.

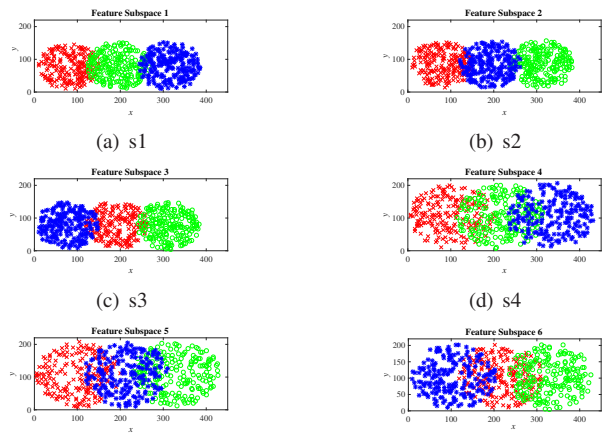
TABLE V
MCS' OBJECTIVE FUNCTION VALUE AND FUSION-BASED CLASSIFICATION ACCURACY ON 1D DATA

Members in MCS	Fusion-based Accuacry	Obj(MCS)
{1, 2, 3}	99.98%	10.8132
{2, 3, 4}	98.00%	3.0418
{2, 4, 5}	95.98%	1.9012

As we see, the MCS with the maximum (minimum) Obj value has the highest (lowest) fusion-based classification accuracy.

B. On 2-D artificial dataset with uniform distribution

A three-class artificial dataset with samples is generated. Each class has 200 samples. We generate six feature subspaces for the samples, where each feature subspace has two dimensions with uniform distribution, as shown in Fig. 6.



(a) s1 (b) s2 (c) s3 (d) s4 (e) s5 (f) s6

Fig. 6. 2-D artificial dataset.

As we see in Fig. 6, Feature subspace 1 can well discriminate Class 1 (Red) and Class 3 (Blue), while it cannot well discriminate Class 2 (Green) from both Class 1 and Class 3; Feature subspace 2 can well discriminate Class 1 and Class 2, while it cannot well discriminate Class 3 from both Class 1 and Class 2; Feature subspace 3 can well discriminate Class 2 and Class 3, while it cannot well discriminate Class 1 from both Class 2 and Class 3. Feature subspace 4 (5 and 6) is similar to Feature subspace 1 (2 and 3), however, Feature subspace 4 (5 and 6) is more ambiguous (i.e., with a deeper overlap). The feature discriminability is show in Table VI.

TABLE VI
FEATURE DISCRIMINABILITY FOR 2D DATA

Feature subspace	J
1	0.3842
2	0.3822
3	0.3907
4	0.5716
5	0.5983
6	0.5697

We still use the k-NN classifier [24] and generate the probability according to Eq. (13) and apply the fusion rule in

Eq. (14). 5-fold cross-validation is used here for evaluation, and the average classification accuracy and the corresponding objective function value in Eq. (8) are calculated. We can use our new diversity based approach to generate the optimal MCS. Suppose that the number of member classifiers in an MCS is 3, the results (including the best MCS, the worst MCS and the middle one's corresponding fusion-based accuracy and objective functions) are listed in Table VII. As we see in this experiment, the MCS with the maximum (minimum) Obj value has the highest (lowest) classification accuracy.

TABLE VII
MCS' OBJECTIVE FUNCTION VALUE AND FUSION-BASED CLASSIFICATION ACCURACY ON 2-D DATA

Members in MCS	Fusion-based Accuacry	Obj(MCS)
{1, 2, 3}	99.83%	3.4567
{1, 3, 4}	98.48%	2.9129
{3, 4, 5}	96.32%	2.2994

V. CONCLUSION

A new ranking distance based diversity measure for the MCS is proposed, which is positively correlated to the MCS's classification accuracy in experiments provided. In future work, our proposed diversity measure will be compared with prevailing ones and we will also try to use more types of ranking distance in defining the ranking distance based diversity measure and make related comparisons and analyses. Note that the construction of MCS based on the diversity measure is actually an optimization, which might cause the local optimal problem and high computational cost especially when the number of classifiers to select is large. Therefore, we will analyze the characteristic of the object function, and try to propose more efficient construction method for the MCS.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation (No. 61671370, No. 61573275), Postdoctoral Science Foundation of China (No. 2016M592790), Postdoctoral Science Research Foundation of Shaanxi Province (No. 2016BSHEDZZ46), and Fundamental Research Funds for the Central Universities (No. xjj2016066).

REFERENCES

[1] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, 1994.

[2] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[3] I. Mayire, M. Kamil, and H. Askar, "Multi-classifier combination scheme for recognition of online handwritten uyghur characters," *Computer Engineering and Applications*, vol. 48, no. 31, pp. 140–145, 2012.

[4] F. Wang, J. Han, and X. Yao, "Mutimodal biometric fusion approach based on iris and face," *Journal of Xian Jiaotong University*, vol. 42, no. 2, pp. 133–137, 2008.

[5] B. Zhong, A. Yang, A. Nie, Y. Yao, H. Zhang, S. Wu, and Q. Liu, "Finer resolution land-cover mapping using multiple classifiers and multisource remotely sensed data in the heihe river basin," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4973–4992, Oct 2015.

[6] F. de Assis Boldt, T. W. Rauber, T. Oliveira-Santos, A. Rodrigues, F. M. Varejo, and M. P. Ribeiro, "Binary feature selection classifier ensemble for fault diagnosis of submersible motor pump," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 19-21 June 2017, pp. 1807–1812.

[7] B. Biggio, I. Corona, Z.-M. He, P. P. K. Chan, G. Giacinto, D. S. Yeung, and F. Roli, "One-and-a-half-class multiple classifier systems for secure learning against evasion attacks at test time," in *Multiple Classifier Systems*. Cham: Springer International Publishing, 2015, pp. 168–180.

[8] Y. Wu, J. Li, and J. Bai, "Multiple classifiers-based feature fusion for rgb-d object recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 05, p. 1750014, 2017. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218001417500148

[9] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181–207, May 2003.

[10] T.-G. Fan, Y. Zhu, and J.-M. Chen, "A new measure of classifier diversity in multiple classifier system," in *2008 International Conference on Machine Learning and Cybernetics*, vol. 1, Kunming, China, 12-15 July 2008, pp. 18–21.

[11] K. Trawinski, A. Quirin, and O. Cordón, "On the combination of accuracy and diversity measures for genetic selection of bagging fuzzy rule-based multiclassification systems," in *Proceedings of the International Conference on Intelligent Systems Design & Applications*, Pisa, Italy, Nov. 30–Dec 2 2009, pp. 121–127.

[12] D. S. C. Nascimento, A. M. P. Canuto, L. M. M. Silva, and A. L. V. Coelho, "Combining different ways to generate diversity in bagging models: An evolutionary approach," in *The 2011 International Joint Conference on Neural Networks*, San Jose, USA, July 31 – Aug. 5 2011, pp. 2235–2242.

[13] M. S. Haghighi, A. Vahedian, and H. S. Yazdi, "Creating and measuring diversity in multiple classifier systems using support vector data description," *Applied Soft Computing*, vol. 11, no. 8, pp. 4931–4942, 2011.

[14] D. Han, X. R. Li, and S. Liang, "Design of dynamic multiple classifier systems based on belief functions," in *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, 9-12 July 2013, pp. 418–425.

[15] B. Krawczyk and M. Wozniak, "Diversity measures for one-class classifier ensembles," *Neurocomputing*, vol. 126, pp. 36–44, 2014.

[16] J. F. Diez, Pastor, J. J. Rodriguez, C. I. Garcia-Osorio, and L. I. Kuncheva, "Diversity techniques improve the performance of the best imbalance learning ensembles," *Information Sciences*, vol. 325, pp. 98–117, 2015.

[17] H. Kadkhodaei and A. M. E. Moghadam, "An entropy based approach to find the best combination of the base classifiers in ensemble classifiers based on stack generalization," in *2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA)*, Qazvin, Iran, 27-28 Jan 2016, pp. 425–429.

[18] G. D. Cavalcanti, L. S. Oliveira, T. J. Moura, and G. V. Carvalho, "Combining diversity measures for ensemble pruning," *Pattern Recognition Letters*, vol. 74, pp. 38–45, 2016.

[19] Y. S. Huang and C. Y. Suen, "The behavior-knowledge space method for combination of multiple classifiers," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Jun 1993, pp. 347–352.

[20] T. Windeatt, "Diversity measures for multiple classifier system analysis and design," *Information Fusion*, vol. 6, no. 1, pp. 21 – 36, 2005, diversity in Multiple Classifier Systems.

[21] L. Didaci, G. Fumera, and F. Roli, "Diversity in classifier ensembles: Fertile concept or dead end?" in *Multiple Classifier Systems*, Z.-H. Zhou, F. Roli, and J. Kittler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–48.

[22] A. Argentini, "Aggregation based on belief function theory," PhD Dissertation, University of Trento, Trento, Canada, 2012.

[23] J. L. Myers, A. Well, and R. F. Lorch, *Research Design and Statistical Analysis*, 3rd ed. Hove, UK: Routledge, 2010.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: John Wiley & Sons, 2012.