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 cannot 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
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 \( x_q \in \mathbb{R}^d \). 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 \( x_q \). Classifier \( e_k \) assigns a class label \( j_k \) to sample \( x_q \), i.e., \( e_k(x_q) = j_k, \ k = 1, 2, \ldots, n, j_k \in \{c_1, \ldots, c_C\} \).

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

3) Measurement Level: the classifier assigns each label a measurement value such as \( a \) posteriori probability or membership function value. For \( x_q \), each member classifier \( e_k \) brings out an output vector \([\omega_k(c_1), \omega_k(c_2), \ldots, \omega_k(c_M)]\), 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**

<table>
<thead>
<tr>
<th>( e_i ) correct (1)</th>
<th>( e_j ) incorrect (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_{ij}^{10} )</td>
<td>( N_{ij}^{10} )</td>
</tr>
<tr>
<td>( N_{ij}^{10} )</td>
<td>( N_{ij}^{01} )</td>
</tr>
</tbody>
</table>

\[
Q_{i,j} = \frac{N_{ij}^{10}N_{ij}^{00} - N_{ij}^{01}N_{ij}^{10}}{N_{ij}^{11}N_{ij}^{00} + N_{ij}^{01}N_{ij}^{10}} \tag{1}
\]

\[
R_{i,j} = \frac{N_{ij}^{11}N_{ij}^{00} - N_{ij}^{01}N_{ij}^{10}}{\sqrt{(N_{ij}^{11} + N_{ij}^{10})(N_{ij}^{01} + N_{ij}^{00})(N_{ij}^{11} + N_{ij}^{01})(N_{ij}^{10} + N_{ij}^{00})}} \tag{2}
\]

\[
D_{i,j} = \frac{N_{ij}^{01} + N_{ij}^{10}}{N_{ij}^{11} + N_{ij}^{00} + N_{ij}^{01} + N_{ij}^{10}} \tag{3}
\]

\[
DF_{i,j} = \frac{N_{ij}^{00}}{N_{ij}^{11} + N_{ij}^{00} + N_{ij}^{01} + N_{ij}^{10}} \tag{4}
\]

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 Windett 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 propose the MCS diversity measure from a different viewpoint, and expect to obtain good properties.
II. 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 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 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 classifier, where \( i = 1, \ldots, L \). 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 \( \Lambda_i \). For the \( L \) member classifiers, there exist corresponding \( L \) rankings. Given two member classifiers \( e_i, e_j \) and their corresponding rankings \( \Lambda_i \) and \( \Lambda_j \), their ranking distance can be calculated.

![Fig. 2. Classification Accuracy Ranking.](image)

E.g., one can choose the commonly used Spearman distance\(^1\) as [23]:

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

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 \( \Lambda_i \) and \( \Lambda_j \) 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

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

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

\[
\text{Div}(MCS_A) = \frac{\rho(\Lambda_1^A, \Lambda_2^A) + \rho(\Lambda_1^A, \Lambda_3^A) + \rho(\Lambda_2^A, \Lambda_3^A)}{C_3^2} = \frac{1}{3} (1.5 + 1.5 + 1.5) = 1.5
\]

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

\[
\text{Div}(MCS_B) = \frac{\rho(\Lambda_1^B, \Lambda_2^B) + \rho(\Lambda_1^B, \Lambda_3^B) + \rho(\Lambda_2^B, \Lambda_3^B)}{C_3^2} = \frac{1}{3} (0.5 + 0.0 + 0.5) = 1/3
\]

Therefore, we have \( \text{Div}(MCS_A) > \text{Div}(MCS_B) \). 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 higher 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.

\[ e_{C_i}(\cdot)(i = 1, \ldots, V) \text{ in Fig. 3 represent the overproduced individual classifiers based on training samples, and } e_{j}(\cdot)(j = 1, \ldots, L) \text{ represent the member classifiers selected out of the overproduced ensemble. Here the “overproduce” means that the number of the produced classifiers is no less than that of classifiers chosen for constructing the MCS. The selection of member classifiers is converted into an optimization problem whose objective function is based on the joint use of} \]
the proposed diversity measure and the average classification accuracy of the selected individual classifiers:

\[ \text{Obj}(MCS) = w_D \cdot \text{Div}(MCS)/2 + w_A \cdot \text{Acc}_{\text{ave}}(MCS) \]  

(9)

where \( \text{Acc}_{\text{ave}}(MCS) \) 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 \( \text{Div}(MCS)/2 \) 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 \( \text{Obj}(MCS) \) in (9):

\[ \text{MCS}_{\text{Best}} = \arg \max_{MCS} \{ \text{Obj}(MCS) \} \]  

(10)

\( \text{MCS}_{\text{Best}} \) 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]

\[ J = \text{tr}(S_w) / \text{tr}(S_b) \]  

(11)

where \( \text{tr} \) 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

\[ M = \frac{1}{C} \sum_{i=1}^{C} \left( \frac{1}{N_i} \sum_{x \in c_i} x \right) \]  

(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.

Assume that three member classifiers will be selected to construct an MCS. Then, there will be \( C^3 \) = 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.

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
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 shown in Table IV.

![TABLE IV](https://example.com/tableIV.png)

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:

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

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

where $P(f(c_i))$ 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](https://example.com/tableV.png)

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.

![Fig. 6](https://example.com/fig6.png)

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.

### 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