Multiple Classifiers Combination Based on Interval-valued Fuzzy Permutation

Jiangtao HUANG¹,†, Minghui WANG¹,², Bo GU², Zhixiang CHEN²

¹Institute of Image and Graphics, Sichuan University, Chengdu 610064, China
²School of Computer Science, Sichuan University, Chengdu 610065, China

Abstract

Multiple classifiers combination is a technique that combines the decisions of different classifiers as to reduce the variance of estimation errors and improve the overall classification accuracy. A new multiple classifiers fusion method integrated classifier selection and classifier combination is proposed in this paper. It is base on interval-valued fuzzy permutation. Firstly, normalize all classifier posterior probabilities using the priori knowledge of corresponding classifier recognition rate. And then, convert the decision matrix of multiple classifier system into interval-valued fuzzy decision matrix. Thirdly, determine the grade of possibility of each class for input sample in multiple classifier system. Finally, select the best classifier in current pattern recognition task using interval-valued fuzzy permutation and use the best classifier to make final decision. The experiments have shown that the new multiple classifiers fusion approach using interval-valued fuzzy permutation can provide much better accuracy compared to independent classifiers and some other fusion methods.

Keywords: Classifiers Combination; Classifier Selection; Interval-valued Fuzzy Permutation; Normalized Decision Space

1. Introduction

There are many techniques or algorithms available for classification problems such as Bayes classifier, k-NN classifier, various distance classifiers, neural network based classifiers, SVM classifier and PCA classifier etc. Sometimes, a method might override the others in classification performance on a specific problem, but in general, it is not possible that one method always outperforms all the other methods for every possible situation [1]. Many researchers have realized that there exist limitations on using a single classification technique. Multiple classifiers can avoid the risk of picking the output of a single classifier, and, consequently, overcome the weaknesses of individual classifiers [2]. Multiple classifiers fusion has been widely investigated in diverse fields, such as image segmentation, data mining from noisy data streams, credit card fraud detection, sensor networks, image, speech and handwriting recognition, and fault diagnosis, to name a few [2].

The most often used classifiers fusion approaches include the majority voting [3], the weighted combination (weighted averaging) [4], the probabilistic schemes [5,6], various rank-ordered rules, such as the Borda count [1,7], the sum rule (averaging), product-rule, max-rule, min-rule, median rule [5], the Bayesian approach [3,4,8], the Dempster-Shafer (D-S) theory of evidence [3,9], the behavior-knowledge
space method (BKS) [10,11], the fuzzy integral [4,12,13], fuzzy templates [14], decision templates [4,15],
combination through order statistics [16,17], combination by a neural network [18]. Roughly, these
techniques can be decomposed into two categories: classifier selection (CS) and classifier combination
(CC). The CS techniques select a single “best” classifier from base classifiers for the final decision. And
CS is usually divided into two types: static classifier selection (SCS) and dynamic classifier selection
(DCS). SCS selects the best classifier in training phase, but DCS selects the best classifier during the
classification phase. The CC techniques combine the results from all base classifiers to work out the final
decision.

In this paper, a novel multiple classifiers combination method using interval-valued fuzzy permutation
(IVFP) will be proposed. The new approach dynamic selects the most locally accurate classifier to estimate
the class of each particular test pattern. Different to other CS techniques in common use, the DCS process
in the new method seeks a base classifiers permutation which obtains the maximum evaluation value in
IVFP. And, the evaluation value is dynamic updated by combining all classifiers. That is to say, the new
approach integrates CS and CC.

The remaining of this paper is organized as follows. Section 2 gives an overview of IVFP, and the
schema of new fusion method will be described in Section 3. In Section 4, experimental evaluation is done.
The final Section of this paper provides conclusions.

2. Overview of IVFP

2.1. Preliminary

A multi-attribute decision making problem can be concisely expressed in a decision matrix, whose element
indicates the evaluation or value of the $i$-th alternative, $A_i$, with respect to the $j$-th attribute, $x_j$ [19]. Ting-Yu
Chen and Jih-Chang Wang [19] extend the canonical matrix format to interval-valued fuzzy decision
matrix $D$, that is, decision makers are expected to assign an extent of membership grades that captures the
degree of the alternative $A_i$ satisfies the attribute $x_j$ according to their opinions. Let $X$ be the discussion
universe containing decision attributes in the multi-attribute decision problem setting. Denote the set of all
attributes $X = \{x_1, x_2, \ldots, x_n\}$. Let $\text{Int}([0,1])$ stand for the set of all closed subintervals of $[0,1]$. An
interval-valued fuzzy set (IVFS) $A_i$ of the $i$-th alternative on $X$ is given by:

$$
A_i = \{x_j, M^-_{A_i}(x_j) > | x_j \in X \} 
$$

(1)

where $M^-_{A_i}: X \rightarrow \text{Int}([0,1])$, such that $x_j \rightarrow M^-_{A_i}(x_j) = [M^-_{A_i}(x_j), M^+_A(x_j)]$. $M^-_{A_i}$ indicates
the possible degree to which the alternative $A_i$ satisfies attribute $x_j$. $M^-_{A_i}(x_j)$ and $M^+_A(x_j)$ are the
lower bound and the upper bound, respectively, of the interval $M^-_{A_i}(x_j)$.

In IVFS theory, let \( \mu_A (x_j) \) be the degree to which the alternative \( A_i \) satisfies attribute \( x_j \), where \( \mu_A (x_j) : X \rightarrow [0,1] \). Similarly, let \( \nu_A (x_j) \) be the degree to which the alternative \( A_i \) does not satisfy attribute \( x_j \) with \( \nu_A (x_j) : X \rightarrow [0,1] \). And for all \( x_j \in X, 0 \leq \mu_A (x_j) + \nu_A (x_j) \leq 1 \).

Let \( M_A^- (x_j) = \mu_A (x_j) \) and \( M_A^+ (x_j) = 1 - \nu_A (x_j) \), and thus \( [M_A^- (x_j), M_A^+ (x_j)] \) = \([\mu_A (x_j), 1-\nu_A (x_j)]\). An interval \( [M_A^- (x_j), M_A^+ (x_j)] \) can be mapped bijectively onto a couple \( (\mu_A (x_j), 1-\nu_A (x_j)) \) [20].

For each element \( x_j \in X \), the intuitionistic index of \( x_j \) in \( A_i \) is defined as follows [21][22]:

\[
\pi_A (x_j) = 1 - \mu_A (x_j) - \nu_A (x_j)
\]

where \( \pi_A (x_j) \in [0,1], \forall x_j \in X \). \( \pi_A (x_j) \) reflects the fact that the decision maker may not always be certain of membership grades. In other words, an interval \( [M_A^- (x_j), M_A^+ (x_j)] \) shows all possible degrees of membership and the decision maker is hesitated to the extent \( \pi_A (x_j) \).

### 2.2. IVFP Method

Let \( A \) and \( B \) denote two IVFSs of the universe of discourse \( X \). [23] defined the following expressions:

\[
A \leq B \text{ if and only if } M_A^- (x) \leq M_B^- (x) \text{ and } M_A^+ (x) \leq M_B^+ (x) \quad \forall x \in X
\]

(3)

\[
A \prec B \text{ if and only if } M_A^- (x) \leq M_B^- (x) \text{ and } M_A^+ (x) \geq M_B^+ (x) \quad \forall x \in X
\]

(4)

In addition, \( A \geq B \) if and only if \( B \leq A \), \( A \succ B \) if and only if \( B \prec A \). \( A = B \) if and only if \( M_A^- (x) = M_B^- (x) \) and \( M_A^+ (x) = M_B^+ (x) \) for all \( x \in X \).

Consider \( i=1,2, \ldots, m; j=1,2, \ldots, n \) and \( [M_A^- (x_j), M_A^+ (x_j)] \) representing the performance measure of the \( i \)-th alternative in terms of the \( j \)-th attribute. The interval-valued fuzzy decision matrix \( D \) is defined as (5).

Since all attributes cannot be assumed to be of equal importance, [19] constructs a set of grades of importance, denoted as \( \omega \), from the decision maker. The IVFS can also be expressed as the subjective importance of decision attributes during the decision maker's evaluation process. An IVFS \( \omega \) in \( X \) is an object having the form as (6).
Consider the interval-valued fuzzy decision matrix $D$ that refers to $m$ alternatives on $n$ attributes. Then, $m!$ permutations of the ranking of the alternatives exist. Let $P_i$ denote the $i$-th permutation:

$$P_i = (\ldots, A_k, \ldots, A_1), \quad \text{for} \quad i = 1, 2, \ldots, m!$$

where $A_k$ is ranked higher than $A_l$. [19] defines the concordance set $C_{kl}$, midrange concordance set $C_{kl}'$, weak concordance set $C_{kl}''$, discordance set $D_{kl}$, midrange discordance set $D_{kl}'$, and weak discordance set $D_{kl}''$ as (8)-(13). And the evaluation value $E(P_i)$ of the $i$-th permutation $P_i$ is defined by (14).

$$E(P_i) = \sum_{j \in C_{kl}} \omega_j + \frac{2}{3} \sum_{j \in C_{kl}'} \omega_j + \frac{1}{3} \sum_{j \in C_{kl}''} \omega_j - \sum_{j \in D_{kl}} \omega_j - \frac{2}{3} \sum_{j \in D_{kl}'} \omega_j - \frac{1}{3} \sum_{j \in D_{kl}''} \omega_j$$

For each permutation $P_i$, its optimal weight value can be computed via the linear programming (LP) as (15). And finally, the permutation with the maximum evaluation value is the optimal ranking order of the alternatives.
3. Classifiers Combination Method using IVFP

In this section, an IVFP-based multiple classifiers fusion method will be proposed. The new approach integrates the various results generated by multiple classifiers into a single result which can give better decision information.

Given a pattern space \( P \) consisting of \( M \) mutually exclusive sets \( P = C_1 \cup \ldots \cup C_M \) with each of \( C_i, \forall i \in \Lambda = \{1, \ldots, M\} \) representing a set of specified patterns called a class. Assume that there are \( K \) individual classifiers in multiple classifier system (MCS). Each classifier (denoted \( e_k, k=1, \ldots, K \)) outputs one index \( j \in \Lambda \) as a class label for input sample \( x \) from \( P \), which can be written as \( e_k=j \).

In [3], the output information that various classification algorithms supply or are able to supply can be divided into three levels: abstract level, rank level and measurement level. An abstract level classifier \( e \) only outputs a class label \( j \) or outputs a subset \( J \subset \Lambda \), and the rank level classifiers output ranks for all class. The measurement level classifiers can give more information than the abstract and the rank level classifiers. A measurement level classifier \( e \) attributes each class in \( \Lambda \) a measurement value to indicate the possibility that the input sample \( x \) belongs to the class. The new fusion method based on IVFP is a new measurement level classifiers combination method.

The decision profile of a measurement level MCS which including \( K \) classifiers for input sample \( x \) can be got as follow:

\[
D(x) = \begin{bmatrix}
P_1(C_1 \mid x) & \cdots & P_1(C_M \mid x) \\
\vdots & \ddots & \vdots \\
P_K(C_1 \mid x) & \cdots & P_K(C_M \mid x)
\end{bmatrix} = \begin{bmatrix}
o_{11} & \cdots & o_{1M} \\
\vdots & \ddots & \vdots \\
o_{KM}
\end{bmatrix}
\]  

(16)

Let \( N_{ij}(e_k) \) be the number of patterns that the output of classifier \( e_k \) is index \( j \) for input sample whose true class is \( C_i \). Then, the recognition rate (Rec.) of classifier \( e_k \) for class \( C_i \) is calculated as:

\[
\text{Rec.}(e_k) = \frac{N_{ii}(e_k)}{\sum_{j \in \Lambda} N_{ij}(e_k)}
\]  

(17)
In MCS, different classifier output has different decision reliability. Before multiple classifiers fusion for final decision, each classifier output is normalized into the same decision confidence criterion by integrating corresponding classifier recognition rate as follow:

\[ \hat{\omega}_{ki} = o_{kj} \times \text{Rec.}(e_{ki}) \]  \hfill (18)

\( \hat{\omega}_{ki} \) is seen as support degree of classifier \( e_k \) for input sample \( x \) coming from class \( C_i \). Because the support degree bases on normalized decision reliability, each classifier all class support degree can not cover the whole decision space as raw outputs. The rest space is hesitation space. In this case, each class hesitation space of each classifier is equal when input sample \( x \) is same. Let \( \overline{\omega}_{ki} \) be reject degree of classifier \( e_k \) for sample \( x \) coming from class \( C_i \), and it can be calculated as follow:

\[ \overline{\omega}_{ki} = \sum_{i' \in \Lambda} \hat{\omega}_{ki'}, \quad i' \in \Lambda \]  \hfill (19)

Reject degree \( \overline{\omega}_{ki} \) is the width of the reject space of classifier \( e_k \) rejecting input sample coming from class \( C_i \). Now, \( \hat{\omega}_{ki} + \overline{\omega}_{ki} \leq 1 \). That is to say, the decision space for each class membership is re-subdivided into three spaces: support space, hesitation space and reject space. All these three spaces in MCS are on the normalizing decision confidence criterion. The hesitation degree denoting the width of hesitation space can be calculated as follow:

\[ \pi_k(C_i) = 1 - \hat{\omega}_{ki} - \overline{\omega}_{ki} \]  \hfill (20)

where \( \pi_k(C_i) \in [0,1], \forall k \in \{1, \ldots, K\}, i \in \Lambda \). \( \pi_k(C_i) \) reflects the fact that the classifier \( e_k \) may not always be certain of \( C_i \) membership grades. In other words, an interval \([M^-_k(C_i), M^+_k(C_i)]\) shows all possible degrees of membership and classifier \( e_k \) is hesitated to the extent \( \pi_k(C_i) \). The hesitation margin plays an important role for final decision. In whole decision space, the upper bound of support space is the lower bound of hesitation space, and the lower bound of reject space is the upper bound of hesitation space. Then, \( M^-_k(C_i) \) and \( M^+_k(C_i) \) is equal to \( \hat{\omega}_{ki} \) and \( 1 - \overline{\omega}_{ki} \) respectively.

Then, the decision profile of MCS as (16) shows can be converted to interval-valued fuzzy decision matrix \( D' \) as the follow form:

\[
D'(x) = \begin{bmatrix}
[M^-_1(C_1), M^+_1(C_1)] & \cdots & [M^-_1(C_M), M^+_1(C_M)] \\
\vdots & \ddots & \vdots \\
[M^-_K(C_1), M^+_K(C_1)] & \cdots & [M^-_K(C_M), M^+_K(C_M)]
\end{bmatrix}
\]  \hfill (21)

Interval-valued fuzzy decision matrix \( D' \) is constructed basing on the classifier accuracy in recognition each class. The classifier accuracy is priori knowledge which is a statistical quantity of historic training.
data. However, the classifier performance is not necessarily consistent with the priori knowledge when a new sample input to the classifier. This uncertain factor of classifier instantaneous decision reliability is taken into account in the paper. And the grade of possibility of each class is dynamic update and regarded in MCS instead of in single classifier. The minimum and maximum posterior probabilities in each class are selected as the lower bound and upper bound of the possibility grade. Using these possibility grades \([M^-(C_i), M^+(C_i)](\forall i \in \Lambda)\) with interval-valued fuzzy decision matrix \(D'\), a optimal permutation of all classifiers \(P_i (i \in \{1, \ldots, m!\})\) which has the largest optimal weight value \(E(P_i)\) among all permutation evaluation values can be sought out. Finally, we can select the first classifier in the optimal permutation to make final decision.

In last section, we find that the concordance set and discordance set easily confuse when \(M^-_A(x_j) = M^-_A(x_j)\), \(M^+_A(x_j) = M^+_A(x_j)\) and \(\pi_A(x_j) = \pi_A(x_j)\). And from (8)-(13), it is clear that each set is constructed by judging the relations of \(M^-, M^+\) and \(\pi\). A new rate rule will be defined for multiple classifiers combination as (22). In (22), the fact that \(M^-_k(C_i) > M^-_i(C_i)\) will be rate \(\omega_j\), \(M^-_k(C_i) = M^-_i(C_i)\) being rated 0 and \(M^-_k(C_i) < M^-_i(C_i)\) being rated \(-\omega_j\). Then, each partial ranking will be composed by three facts such as \([M^-_k(C_i) > M^-_i(C_i), M^+_k(C_i) > M^+_i(C_i), \pi_k(C_i) > \pi_i(C_i)]\). And all relations assembles are given as (23) with their rates, where \(\emptyset\) denotes that this relations assemble can not be appeared in MCS. The LP in (15) is changed into (24).

\[
\begin{array}{cccc}
M^- & M^+ & \pi \\
> & 1 & 1 & -1 \\
e & 0 & 0 & 0 \\
< & -1 & -1 & 1 \\
\end{array}
\]  

(22)

The algorithm for multiple classifiers combination using IVFP (IVFP_CS) is as follows:

**Step1.** Normalize decision space of all classifier in MCS.

**Step2.** Convert the multiple classifier system decision output matrix into interval-valued fuzzy matrix.

**Step3.** Construct each class possibility grade in current classification task.

**Step4.** Array all classifiers and calculate the optimal weight value of each permutation.

**Step5.** Select the best permutation which has highest optimal weight.

**Step6.** Select the top classifier in the best permutation, and use its output for final decision.
\[
[M^{-1}M^{+}\pi] \quad \text{rate} \quad [M^{-1}M^{+}\pi] \quad \text{rate} \quad [M^{-1}M^{+}\pi] \quad \text{rate}
\]

\[
[\ggg] \quad 1 \quad [\ggg] \quad 0 \quad [\ggg] \quad -1
\]

\[
[\ggg] \quad 2 \quad [\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset
\]

\[
[\ggg] \quad 3 \quad [\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset
\]

\[
[\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset \quad [\ggg] \quad -2
\]

\[
[\ggg] \quad \emptyset \quad [\ggg] \quad 0 \quad [\ggg] \quad \emptyset
\]

\[
[\ggg] \quad 2 \quad [\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset
\]

\[
[\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset \quad [\ggg] \quad -3
\]

\[
[\ggg] \quad \emptyset \quad [\ggg] \quad \emptyset \quad [\ggg] \quad -2
\]

\[
[\ggg] \quad 1 \quad [\ggg] \quad 0 \quad [\ggg] \quad -1
\]

\[
\max \quad E(P_i) = \sum_{j=1}^{\omega} \text{rate}_j \omega_j
\]

subject to \quad \mathbf{M}^{-1}(x_j) \leq \omega_j \leq \mathbf{M}^{+}(x_j) \quad (j = 1, 2, \ldots, n)

\[
\sum_{j=1}^{n} \omega_j = 1
\]

for each \quad i = 1, 2, \ldots, m!

\section{Experiments}

In this section, we report experiments on some multi-class data sets using MCS. Firstly, the data sets and base classifiers of MCS are discussed in Section 4.1. In Section 4.2, we present and discuss the results.

\subsection{Data Sets and Base Classifiers}

Four data sets are taken from the UCI Machine Learning Repository \cite{uci} to use in the experiments, and the specifications of these four data sets are summarized in Table 1. Training data sets are extracted from each data set by the rate of 2/3, and the whole data set for testing.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of classes</th>
<th>Features</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>3</td>
<td>4</td>
<td>150</td>
</tr>
<tr>
<td>Wine</td>
<td>3</td>
<td>13</td>
<td>178</td>
</tr>
<tr>
<td>Vehicle</td>
<td>4</td>
<td>18</td>
<td>846</td>
</tr>
<tr>
<td>Glass</td>
<td>6</td>
<td>10</td>
<td>214</td>
</tr>
</tbody>
</table>

In this work, nearest centroid classifier (NCC), \(k\)-nearest-neighbor rule (\(k\)-NN) \cite{NN}, \(k\)-means clustering classifier and Parzen classifier are used as base classifier in MCS. In the experiments, the parameter \(k\) is set \(k=5\) in \(k\)-NN and \(k\)-means.
4.2. Results

Table 2 gives the test results of different base classifiers on four data sets. And the test results of the novel method IVFP_CS is presented in Table 3. Two common used multiple classifiers fusion methods majority voting and sum rule are also tested on these four data sets whose performance are showed in Table 3 for comparing with IVFP_CS.

Table 2 Classification Accuracy of Base Classifiers (%)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Iris</th>
<th>Wine</th>
<th>Vehicle</th>
<th>Glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCC</td>
<td>92.64</td>
<td>72.41</td>
<td>39.23</td>
<td>89.39</td>
</tr>
<tr>
<td>k-NN</td>
<td>97.13</td>
<td>72.98</td>
<td>66.53</td>
<td>98.23</td>
</tr>
<tr>
<td>k-Means</td>
<td>92.49</td>
<td>71.66</td>
<td>47.94</td>
<td>84.35</td>
</tr>
<tr>
<td>Parzen</td>
<td>94.41</td>
<td>87.48</td>
<td>79.09</td>
<td>99.39</td>
</tr>
</tbody>
</table>

Table 3 Classification Accuracy of IVFP_CS and Some Other Fusion Methods (%)

<table>
<thead>
<tr>
<th>Fusion Method</th>
<th>Iris</th>
<th>Wine</th>
<th>Vehicle</th>
<th>Vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority Voting</td>
<td>94.84</td>
<td>77.24</td>
<td>67.24</td>
<td>98.00</td>
</tr>
<tr>
<td>Sum Rule</td>
<td>96.51</td>
<td>88.53</td>
<td>83.79</td>
<td>99.34</td>
</tr>
<tr>
<td>IVFP_CS</td>
<td>96.97</td>
<td>90.79</td>
<td>86.07</td>
<td>99.47</td>
</tr>
</tbody>
</table>

From the two tables, we can find that the new fusion method is almost always better than single classifiers in MCS. The performance of k-NN classifier is better than IVFP_CS on iris data set, but its performance on other data sets are worse than fusion methods. Because the data set iris is easy to classification than others. And From Table 3, it is easy to see that IVFP_CS can obtain higher accuracy than max-voting and sum-rule. The advantage of IVFP_CS is prominent when handling complex classification task.

5. Conclusions

In this paper, we discussed interval-valued fuzzy permutation method for multiple classifiers fusion. A new multiple classifiers fusion method IVFP_CS is proposed in this paper. IVFP_CS is a fusion method integrated classifier selection and classifier combination. Experimental results on 4 UCI data sets show that IVFP_CS can obtain better performance compared with base classifiers in MCS and some common used fusion methods.

References


