Generating Sets of Classifiers for the Evaluation of Multi-Expert Systems

D. Impedovo, G. Pirlo
Dipartimento di Informatica – Università degli Studi di Bari - via Orabona 4, 70126 – Bari
Dipartimento di Elettrotecnica ed Elettronica-Politecnico di Bari – Via Orabona 4, 70126 – Bari
Centro “Rete Puglia” – Università degli Studi di Bari – via G. Petroni 15/F.1 70100 Bari
pirlo@di.uniba.it

Abstract
This paper addresses the problem of multi-classifier system evaluation by artificially generated classifiers. For the purpose, a new technique is presented for the generation of sets of abstract classifiers with different characteristics at the individual-level (i.e. recognition performance) and at the collective-level (i.e. degree of similarity).

The technique has been used to generate sets of abstract classifiers simulating different working conditions in which the performance of combination methods can be estimated. The experimental tests demonstrate the effectiveness of the approach in generating simulated data useful to investigate the performance of combination methods for abstract-level classifiers.

1. Introduction
Classifier combination is a diffuse paradigm for high-performance classification for all those cases in which the complexity of the classification problem do not allow the development of classifiers as good as required for practical applications [1].

Although the potential of multi-classifier systems has been confirmed, the problem of multi-classifier system design is still open. This is mainly due to difficulty in evaluating combination methods. Theoretical analysis of combination methods is very complex and in fact results have been obtained only for very simple combination schemes. When the combination methods are too complex, theoretical analysis can be impracticable and the evaluation of performance can be estimated on experimental basis. In this case the result depends on the specific working conditions of the test and no useful information can be obtained on the performance of the combination method under different working conditions [2].

More recently, the need to define general approaches for evaluating combination methods has led many researchers to use of synthetic classifiers to artificially determine different working conditions under which the performance of the combination method can be estimated [3, 4].

This paper starts from the consideration that combined classifiers significantly outperforms the individual classifiers, if classifiers are diverse enough from each other. In this case the combination method should be able to exploit the strength of some individual classifiers to overcome the weaknesses of others. Therefore the degree of similarity of the set of individual classifiers represents an important parameter for the evaluation process of combination methods. According to this consideration, this paper presents an effective technique for the generation of sets of abstract classifiers with different degrees of similarity, to be used for multi-expert system evaluation.

The organization of this paper is as follows. Section 2 presents the Similarity Index, that is an estimator of similarity among abstract classifiers. Section 3 describes the technique for generating and using sets of abstract classifiers for multi-expert system evaluation. Section 4 presents some experimental results, showing relevant properties of some of the most important combination methods for abstract-level classifiers. The conclusion of this work is presented in Section 5.

2. Abstract-level Classifiers: a Similarity-based Analysis
In a parallel multi-classifier system the input pattern is fed in parallel to K individual classifiers and their responses are combined to obtain the final classification result, according to a suitable combination strategy [1, 2]. In other words let be

\[ X = \{x_1, x_2, \ldots, x_N\} \] a set of patterns;

\[ C = \{C_1, C_2, \ldots, C_M\} \] a set of classes;

\[ A = \{A_1, A_2, \ldots, A_K\} \] a set of abstract-level classifiers; where , for any \(i=1,2,\ldots,K:\)

\[ A_i(x_t) = C_m \quad x_t \in X ; C_m \in C; \]
a parallel multi-classifier system based on the set of classifiers A performs the classification of a test pattern \( x_i \) as follows: \( E(x_i) = E(A_1(x_i), A_2(x_i), \ldots, A_k(x_i)) = C_m \), being \( x_i \in X \) and \( C_m \in C \).

Of course, the effectiveness of a multi-classifier system depends not only on the performance of the individual classifiers but also on the collective behavior of the entire set of classifier. The combination method should be able to overcome the weakness of the individual classifiers in classifying specific patterns by using the ability of other classifiers in classifying the same patterns. Therefore, a combination method cannot be useful if the individual classifiers are similar each other [1,2].

![Table](image)

**Fig. 1** Variability of the Similarity Index

In order to evaluate the similarity among classifiers, a well-suited estimator has been proposed, named the Similarity Index [3]. Let \( A_1, A_2 \) be two classifiers and \( A_1(x_i) \) and \( A_2(x_i) \) respectively the top-candidates provided for pattern \( x_i \), with \( x_i \) belonging to the set \( X=\{x_1, x_2, \ldots, x_N\} \) and let be

\[
I(A_1(x_i), A_2(x_i)) = \begin{cases} 
1 & \text{if} \ A_1(x_i) = A_2(x_i) \\
0 & \text{otherwise}
\end{cases} \quad (1)
\]

The Similarity Index between \( A_1, A_2 \) is defined as:

\[
\rho_{A_1,A_2} = \frac{1}{\text{Card}(X)} \sum_{x_i \in X} I(A_1(x_i), A_2(x_i)) \quad (2)
\]

For example, Figure 1a shows the outputs of two classifiers \( A_1 \) and \( A_2 \), for 10 input patterns belonging to the set of the ten numeral digits. The recognition rate of both classifiers is 80%. Moreover it is easy to verify that \( A_1 \) and \( A_2 \) always provide the same response. Thus, we have \( \rho=1 \). Also the recognition rate of both classifiers in Figure 1b is equal to 80%. However, in this case \( \rho=0.6 \).

In general, for a set \( A=\{A_i\}_{i=1,2,\ldots,K} \) of \( K \) classifiers, the Similarity Index is defined as [3]:

\[
\rho_{A_i,A_j} = \frac{\sum_{i, j = 1 \ldots K} \rho_{A_i, A_j}}{K} \quad (3)
\]

Now, the degree of similarity \( \rho \), for a set of \( K \) abstract-level classifiers, can vary in a well-defined range \([\rho_{\text{min}}, \rho_{\text{max}}]\) depending on the characteristics of the individual classifiers, where [5]:

\[
\rho_{\text{min}} = \frac{K-k'}{K} \quad (4)
\]

\[
\rho_{\text{max}} = 1 - \frac{2 \sum_{i=1}^{K} i \cdot R_i - (K+1) \sum_{i=1}^{K} R_i}{K} \quad (5)
\]

For example, Fig. 2 shows the outputs of \( K=4 \) classifiers with recognition rate equal to 0.7 (i.e. \( R_1=R_2=R_3=R_4=0.7 \)). The degree of similarity (eq. (3)) is equal to 0.65. Note that from eqs. (4) and (5) it results \([\rho_{\text{min}}, \rho_{\text{max}}]=[0.6,1]\).

![Table](image)

**Fig. 2.** A set of four classifiers

### 3. Evaluating Multi-Expert Systems

The generation of artificial classifiers with diverse characteristics offers the possibility to estimate the performance of each combination method under different working conditions. In particular, the performance \( E(K, R, \rho) \) of a combination method is here considered as a function of:

- \( K \): the number of combined classifiers;
- \( R=(R_1,R_2,\ldots,R_k) \): the recognition rates;
- \( \rho \): the degree of similarity of the set of classifiers.

Of course, it is worth noting that, for a set of \( K \) classifiers \( A=\{A_i\}_{i=1,2,\ldots,K} \), with \( R=(R_1,R_2,\ldots,R_k) \) being \( R \), the recognition rate of \( A_k \) the exhaustive analysis on the performance of combination method as a function of the degree of similarity among classifiers can be performed by generating sets of synthetic
classifiers with similarity index ranging in \([\rho_{\text{min}}, \rho_{\text{Max}}]\). The performance of each combination method \(E(K,R,\rho)\) is then computed as the average performance of the method, when sets of classifiers belonging to the class \((K,R,\rho)\) are considered \([5]\).

For the purpose, in this paper an effective technique is proposed for generating sets of artificial classifiers having a degree of similarity ranging in \([\rho_{\text{min}}, \rho_{\text{Max}}]\). The input data are \(K\) (number of classifiers) and \(R=(R_1,R_2,...,R_K)\) (the recognition rates of the classifiers), the output data are the output simulating the sets of \(K\) classifiers with recognition rates \(R=(R_1,R_2,...,R_K)\) and degree of similarity ranging in \([\rho_{\text{min}}, \rho_{\text{Max}}]\).

![Figure 3. Set of Artificial Classifiers](image)

In order to develop the set of artificial classifiers simulating different working conditions, each classifier \(A_i\), \(i=1,2,...,K\), is here considered to be a discrete random variable producing a simple class label as its decision (output) corresponding to each input pattern. More precisely, if a set of \(N_j\) input patterns for each class label \(C_j\) is supposed to be input to the classifier \(A_i\), with recognition rate \(R_i\), \(A_i\) will generate a list of \(N_j\) class labels which simulate the classifier decisions. The list contains (in random order):

- \(N_j\) - \(R\) recognitions (that are indicated by \(R\));
- \(N_j\) - (1-\(R\)) misclassifications (that are indicated by \(S1,S2,S3,...\)).

Of course, misclassifications are obtained by uniformly picking in the set \(\{C_1, C_2,..., C_m\}\) - \(\{C_j\}\).

Figure 3 shows the list of outputs simulating a set of \(K=4\) classifiers with \(R=(R_1,R_2,R_3,R_4)=(0.7, 0.6, 0.7, 0.6)\) and \(\rho=0.7\). In this case we suppose \(N=10\) to be the number input patterns of different classes, therefore we indicate as “\(R\)” the recognitions and with “\(S1\)”, “\(S2\)”, “\(S3\)” etc. the substitutions (being \(Si\neq Sj\), if \(i\neq j\)).

In this paper an effective technique is proposed for generating sets of artificial classifiers with diverse characteristics. In fact, traditional approaches - based on iterative procedures for random number generation - are not effective since they produce sets of classifiers showing similar characteristics \([6]\). Conversely, the new technique derives, from an initial list of outputs, several other sets of classifiers by well-suited operators named CHANGE+, CHANGE-, SWAP+, SWAP-. These operators have the aim to generate sets of synthetic classifiers with the same individual characteristics (i.e. the same vector \(R=(R_1,R_2,...,R_K)\)) and different degree of similarity. A detailed description of the operators is reported in the following:

- The CHANGE operators act on substitutions. Two classifiers \(A_i, A_j\) and one position \(r\) in the lists of outputs are selected.
  - When the CHANGE+ operator is used, if \(A_r\) and \(A_s\) are substitutions, and \(A_r\neq A_s\), they are made equal to increase the value of the Similarity Index (for instance \(A_r(\leq A_s(\leq A_r)\) without varying the recognition rate of the two classifiers.
  - When the CHANGE- operator is used, if \(A_r\) and \(A_s\) are substitutions, and \(A_r(\neq A_s(\neq A_r)\), one of them is changed with a different wrong value to reduce the value of the Similarity Index (for instance \(A_r(\leq A_s(\neq A_r)\).

- The SWAP operators act on recognitions and substitutions. Two classifiers \(A_i, A_j\) and two positions \(r\) and \(s\) are randomly selected.
  - When the SWAP+ operator is used, if \(A_r\) and \(A_s\) are recognitions and \(A_r\neq A_s\), \(A_r\) and \(A_s\) are swapped to increase the Similarity Index.
  - When SWAP- is used, if \(A_r\) and \(A_s\) are recognitions and \(A_r\neq A_s\), \(A_r\) and \(A_s\) are swapped to decrease the Similarity Index.

On the basis of these operators, sets of synthetic classifiers can be generated. Precisely, given the integer \(K\) and the vector of recognition rates \(R=(R_1,R_2,...,R_K)\), the generation procedure works as follows (being \(N\) the number of outputs required):

1. Generate the initial list of \(K\cdot N\) outputs simulating \(K\) classifiers having recognition rate equal to \(R\) and compute the value of \(\rho\).
2. Repeat until the lists of outputs generated is equal to \(M\) (being \(M\) the number of sets of classifiers to be generated);
3. Repeat until \(\rho=\rho_{\text{Max}}\).
   - Modify the previous list of output by applying randomly CHANGE+ and SWAP+ (Increase \(\rho\))
   - Compute the new value of \(\rho\)
   - Store the new list of output
   - End Repeat;
4. Repeat until \(\rho=\rho_{\text{Min}}\).
   - Modify the previous list of output by applying randomly CHANGE- and SWAP- (Reduce \(\rho\))
Compute the new value of $\rho$
Store the new list of output
End Repeat;
End Repeat;

4. Experimental results

The procedure for generating artificial sets of classifiers has been used and compared to the traditional technique. A typical result is shown in Figure 4, that demonstrates the capability of the new technique in generating sets of classifiers with diverse degrees of similarity (in the example we have K=6, N=1000, M=10, $R_i=0.8$, i=1,2,...6, and therefore $[\rho_{\text{min}}, \rho_{\text{Max}}]=[0.613,1]$). Conversely, the generation technique based on random number generation is not able to generate sets of classifiers with degree of similarity spreading over the entire range of variability.

When the sets of generated classifiers are used to evaluate the performance of a multi-classifier system, they allows to derive relevant information on the behaviour of different combination methods, depending on the degree of similarity between classifiers. Figure 5 shows the performance of three well-known combination methods for abstract-level classifiers: Majority Vote (MV), Dempster-Shafer (DS) and Behaviour Knowledge Space (BKS). The complete description of these methods is beyond the aim of this paper and can be found in the literature [1,8,9]. In this case the performance has been estimated by the cost function $CF=E+2R$, being $E$ the error rate and $R$ the rejection rate of the combination method [5]. The result demonstrate important properties of combination methods. For instance, in this case, it results that the best method is DS, for $p_e \in [0.613,0.73]$; MV, for $p_e \in [0.73, 1]$. Conversely, the results demonstrate that BKS is never the best combination method, in these operative conditions.

5. Conclusion

This paper presents a new technique for generating artificial sets of abstract-level classifiers, for the evaluation of multi-expert systems. The experimental results demonstrate the effectiveness of the new technique.

References