# Belief function combination: comparative study within the classifier fusion framework

Asma Trabelsi (✉)[1], Zied Elouedi[1], and Eric Lefèvre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC , Tunisia
trabelsyasma@gmail.com, zied.elouedi@gmx.fr
[2] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France
eric.lefevre@univ-artois.fr

**Abstract.** Data fusion under the belief function framework has attracted the interest of many researchers over the past few years. Until now, many combination rules have been proposed in order to aggregate beliefs induced form dependent or independent information sources. Although the choice of the most appropriate rule among several alternatives is crucial, it still requires non-trivial effort. In this investigation, we suggest to evaluate and compare some combination rules when dealing with independent information sources in the context of the classifier fusion framework.

Keywords - data fusion, belief function theory, combination rules, independent information sources, classifier fusion.

## 1 Introduction

Pattern recognition has been widely studied to solve classification problems owing to its capacity to achieve the greatest possible classification accuracy [15]. One of the proposed solutions is based on an advanced method named ensemble classifiers. Hence, various combination approaches have been proposed to combine multiple classifiers such as voting-based systems, plurality, Bayesian theory, belief function theory [8]. This latter, also known as Dempster-Shafer theory, is regarded as a convenient method for representing and managing different kinds of imperfect data [14] and has proved to be an efficient approach for combining a set of classifiers. Thus, it provides several combination rules which mainly differ according to the way of managing the mass assigned to the empty set also called conflict [1, 2, 5, 7, 12]. Basically, in this paper, we are interested in the Dempster rule [1], the conjunctive rule [12], the combination with adapted conflict rule (CWAC rule) [5] and the improved CWAC rule [2]. It is noteworthy that the combination rule of Dempster does not support the value of the conflict generated when combining pieces of evidence and consequently this latter should be proportionally distributed over all focal elements. However, in the conjunctive combination rule the mass allocated to the empty set should be kept in the purpose of reflecting the degree of conflict between the combined sources. Nevertheless, this conflict has an absorption effect: when we apply a large number of conjunctive combinations, the mass assigned to the conflict tends towards 1 and hence the conflict loses its initial role. The CWAC and the improved CWAC rules are defined by an adaptive weighting between the Dempster and the conjunctive rules in order to give the conflict its paramount role as an alarm signal. With this diversity of combination rules, the choice of the most efficient one becomes really a challenging task. So, in this work, we propose to compare the CWAC and the improved CWAC rules with the conjunctive and the Dempster rules within the classifier fusion framework in order to pick out the most appropriate combination rule. The rest of this paper is structured as follows. Section 2 provides a brief overview of the basic concepts of the belief function theory. We outline some combination rules dealing with distinct pieces of evidence in Section 3. Section 4 is devoted to discussing our comparative approach. The experiments and the results are presented in Section 5. The conclusion is reported in Section 6.

## 2 Belief function theory: basic concepts

Let $\Theta$ be a finite non-empty set of $N$ elementary events related to a given problem, these events are assumed to be exhaustive and mutually exclusive. Such $\Theta$ is called the frame of discernment. The power set of $\Theta$, denoted by $2^{\Theta}$, is composed of all the subsets of $\Theta$.

The impact of evidence assigned to each subsets of the frame of discernment $\Theta$ is named basic belief assignment (bba). It is defined as:

$$\begin{aligned} & m : 2^{\Theta} \rightarrow [0,1] \\ & \sum_{A \subseteq \Theta} m(A) = 1 \end{aligned} \tag{1}$$

The amount $m(A)$, known as basic belief mass (bbm), expresses the degree of belief committed exactly to the event $A$.

To make decision within the belief function framework, we must transform the bba into a probability measure called pignistic probability denoted $BetP$ and defined as follows [13]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)} \;\forall\; A \in \Theta \tag{2}$$

where $|B|$ denotes the cardinality of B.

The reliability of each information source $S$ can be quantified. In fact, if $S$ is not fully reliable then the bba provided by $S$ should be discounted using a reliability factor denoted 1 - $\alpha$ [11]. The discounted bba is obtained as follows:

$$m^{\alpha}(A) = (1-\alpha)m(A) \;\forall\; A \subset \Theta \qquad (3)$$
$$m^{\alpha}(\Theta) = \alpha + (1-\alpha)m(\Theta)$$

Given two bbas $m_1$ and $m_2$, according to [4], the distance measure between them is computed as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D(m_1 - m_2)} \qquad (4)$$

with $D$ is the Jaccard index matrix, the elements of which are calculated as follows:

$$D(A,B) = \begin{cases} 1 & \text{if } A{=}B{=}\,\emptyset \\ \dfrac{|A \cap B|}{|A \cup B|} & \forall\; A,B \in 2^{\Theta} \end{cases} \qquad (5)$$

## 3 Combination rules dealing with independent information sources

As mentioned earlier, there exist several combination rules assuming items of evidence combined to be independent. In this section, we present only the conjunctive rule [12], the Dempster rule [1], the CWAC rule [5] and the improved CWAC rule [2].

1. The conjunctive rule, proposed by Smets, is used to combine two bbas provided by reliable and distinct information sources [12]. The resulting bba, denoted $m_1 \bigcirc\!\!\!\!\cap\, m_2$, is defined by:

$$(m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B).m_2(C) \qquad (6)$$

   The mass assigned to the empty set $(m_1 \bigcirc\!\!\!\!\cap\, m_2(\emptyset))$ quantifies the degree of disagreement between the two combined sources.

2. The Dempster rule, based on the orthogonal sum, is a normalized version of the conjunctive rule where the mass of the empty set must be reallocated over all focal elements in the case where $m_1 \bigcirc\!\!\!\!\cap\, m_2(\emptyset) \neq 0$ thanks to a normalization factor, denoted $K$ [11]. This rule, assuming pieces of evidence combined to be reliable and distinct, is defined as follows:

$$(m_1 \oplus m_2)(A) = K(m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) \qquad (7)$$

   and

$$(m_1 \oplus m_2)(\emptyset) = 0 \qquad (8)$$

   where

$$K^{-1} = 1 - (m_1 \bigcirc\!\!\!\!\cap\, m_2)(\emptyset) \qquad (9)$$

3. In [5] the authors have proposed the CWAC combination rule which is defined by an adaptive weighting $D$ between the conjunctive and the Dempster rules. This adaptive weighting offers an effective way to obtain the same behavior as the conjunctive rule when the bbas are contradictory and the same behavior as the Dempster rule when the bbas are similar. The CWAC rule uses the Jousselme distance to measure the dissimilarity between sources. Assume we have $M$ bbas, denoted as $m_1$,...., $m_M$, the result of their combination using the CWAC operator is noted as $m_{\ominus}$ and is defined as follows:

$$m_{\ominus}(A) = Dm_{\bigcirc\!\!\!\!\cap}(A) + (1 - D)m_{\oplus}(A) \tag{10}$$

and

$$m_{\ominus}(\emptyset) = 1 \text{ when } m_{\bigcirc\!\!\!\!\cap}(\emptyset) = 1 \tag{11}$$

with

$$D = \max_{i,j}[d(m_i, m_j)] \ \forall \ i,j \in [1,M] \tag{12}$$

$$m_{\bigcirc\!\!\!\!\cap}(A) = (\underset{i}{\bigcirc\!\!\!\!\cap} m_i)(A) \text{ and } m_{\oplus}(A) = (\underset{i}{\oplus} m_i)(A) \ \forall \ i \in [1,M] \tag{13}$$

4. The improved CWAC rule [2], inspired from the spirit of the CWAC rule, is employed to combine reliable and distinct pieces of evidence. Authors in [2] have proved that the improved CWAC rule enhances the ability of the CWAC rule to preserve the conflict as an alarm signal and also it truly reflects the opposition between bbas in the combination. Assume we have $M$ bbas, denoted as $m_1$,.....,$m_M$, the result of their combination using the improved CWAC operator, denoted $\ominus^I$, is defined as follows:

$$m_{\ominus^I}(A) = \bar{D}m_{\bigcirc\!\!\!\!\cap}(A) + (1 - \bar{D})m_{\oplus}(A) \tag{14}$$

and

$$m_{\ominus^I}(\emptyset) = 1 \text{ when } m_{\bigcirc\!\!\!\!\cap}(\emptyset) = 1 \tag{15}$$

with

$$\bar{D} = \frac{\sum_{i=1,j>i}^{M} d(m_i, m_j)}{\frac{M(M-1)}{2}} \forall \ i,j \in [1,M] \tag{16}$$

$$m_{\bigcirc\!\!\!\!\cap}(A) = (\underset{i}{\bigcirc\!\!\!\!\cap} m_i)(A) \text{ and } m_{\oplus}(A) = (\underset{i}{\oplus} m_i)(A) \ \forall \ i \in [1,M] \tag{17}$$

As the CWAC and the improved CWAC rules are defined by an adaptive weighting between the conjunctive and the Dempster rules, we suggest to make a comparative approach that allows to select the most efficient rule within the classifier fusion framework.

# 4 Comparative approach

Ensemble classifier systems, also known as multiple classifiers, is considered as an efficient way to solve pattern recognition issues. As well, the fusion of a set of classifiers in the context of the belief function framework has been extensively explored in several studies [6, 10]. In this paper, ensemble classifier systems will be used as a way for evaluating and comparing some fusion rules dealing with independent information sources. The process of combining classifiers within the belief function framework composed of two distinct parts. The first one consists of the construction of mass functions from classifiers' outputs and the second one focuses on the combination of these mass functions across some combination rules.

## 4.1 Mass functions construction from classifiers' outputs

Consider a pattern recognition problem where $B = \{x_1, ..., x_n\}$ is a database with $n$ instances, $C = \{C_1, \ldots, C_M\}$ is a set of $M$ classifiers and $\Theta = \{w_1, \ldots, w_N\}$ is a set of $N$ class labels. $B$ should be randomly split into learn and test sets. We first construct classifiers from the learning set and then we apply them to predict the label class of all test patterns. In order to combine classifiers within the belief function framework, classifiers outputs must be transformed into bbas. Since classifiers outputs may differ from one classifier to another, each pattern test should have $M$ bbas obtained as follows:

$$\begin{aligned} &m_i(\{w_j\}) = 1 \\ &m_i(A) = 0 \ \forall \ A \subseteq \Theta \text{ and } A \neq \{w_j\} \end{aligned} \tag{18}$$

where $m_i(\{w_j\})$ expresses the part of belief assigned exactly to the predicted class $w_j$ through the classifier $C_i$.

Results supplied by classifiers can be unreliable, therefore the bbas generated must be discounted by taking into consideration the reliability rate of each classifier. The reliability $r_i$ of a classifier $C_i$ is computed as follows:

$$r_i = \frac{\text{Number of well classified instances}}{\text{Total number of classified instances}} \tag{19}$$

If $r_i$ equals 1 then the classifier $C_i$ is absolutely reliable, by against the classifier $C_i$ is totally unreliable in the case where $r_i$ is equal to 0. The discounted mass functions, using Equation 3, are obtained as follows:

$$\begin{aligned} &m_i^{\alpha_i}(\{w_j\}) = r_i \\ &m_i^{\alpha_i}(\Theta) = 1 - r_i \end{aligned} \tag{20}$$

with $\alpha_i$ = 1 - $r_i$.

## 5 Classifier fusion

Let us remind that the process of combining classifiers within the belief function framework consists on two main steps: classifiers' outputs modeling and classifiers' combination. So, if the outputs of all classifiers are converted into bbas, we move on to classifier fusion using combination rules mentioned in Section 3. Combination results allows us to assess and compare these alternative rules in order to select the most efficient one. Thus, we rely on two assessment criteria: the PCC and the distance.

- The PCC criterion that represents the percent of correctly classified instances will be used to compare the CWAC and the improved CWAC rules of combination with the Dempster one. Such case demands the use of three variables $n_1$, $n_2$, $n_3$ which respectively correspond to the number of well classified, misclassified and rejected instances. For each combination rule, we proceed as follows:
  1. Firstly, we set a tolerance thresholds $S$=[0.1,1]. For any threshold $s \in S$, we examine the mass of the emptyset ($m(\emptyset)$) induced by each model test as follows:
     - if $m(\emptyset) > s$, the classifier chooses to reject instance instead of misclassifying it. As a result, we increment the number of rejected instances $n_3$.
     - if $m(\emptyset) \leq s$, we calculate the pignistic probability ($BetP$) in order to choose the most probable class. Accordingly, the chosen class will be compared to the real one: in the case where the chosen class is similar to the real one, we increment the number of well classified instance $n_1$, inversely we increment the number of the misclassified instances $n_2$.
  2. Secondly, having the well classified, misclassified and rejected instances, we calculate the PCC for each threshold $s \in S$ using the following formula:

$$PCC = \frac{n_1}{n_1 + n_2} * 100 \tag{21}$$

  The most appropriate combination rule is the one that has the highest value of PCC $\forall \ s \in S$.

- The distance criterion, which corresponds to the Jousselme distance between two bbas, will be employed to compare the CWAC and the improved CWAC rules with the conjunctive one. Thus, for each combination rule, we track the following steps to compute the distance criterion:
  1. The real class $w_j$ of each pattern test must be transformed into a mass function. It is obtained as follows:

$$\begin{aligned} &m_r(\{w_j\}) = 1 \\ &m_r(A) = 0 \ \forall \ A \subseteq \Theta \text{ and } A \neq \{w_j\} \end{aligned} \tag{22}$$

2. Then, we compute for each test pattern the Jousselme distance (See Equation 4) between the mass function relative to its real class ($m_r$) and the mass function generated when combining bbas induced from $M$ classifiers.
3. At last, we sum the Jousselme distances obtained by all test patterns in order to get the total distance.

The best combination rule is the one that has the minimum total distance.

# 6 Simulation and experimentations

## 6.1 Experimental setup

To evaluate our alternative combination rules, our experiments are performed using several real world databases obtained from the U.C.I repository [9] described in Table 1. These databases have different number of instances and different number of attributes. However, their classe numbers are equal to 2 or 3. It is noteworthy that our alternative combination rules can support databases with a number of classes greater than 3.

**Table 1.** Description of databases

| **Databases** | **#Instances** | **#Attributes** | **#Classes** |
|---|---|---|---|
| Diabetes | 768 | 2 | 2 |
| Fertility | 100 | 10 | 2 |
| Heart | 270 | 13 | 2 |
| Hepatitis | 155 | 19 | 2 |
| Iris | 150 | 4 | 3 |
| Parkinsons | 195 | 23 | 2 |

We have carried out experiments using four machine learning algorithms implemented in Weka [3]: the Naive Bayes (NB), the Decision tree(DT), the $k$-Nearest Neighbor where $k$ equals 1 (1-NN), and the Neural Network (NN) algorithms. These latter were run based on the leave one out cross validation approach. The accuracy values of the single classifiers are given in Table 2.

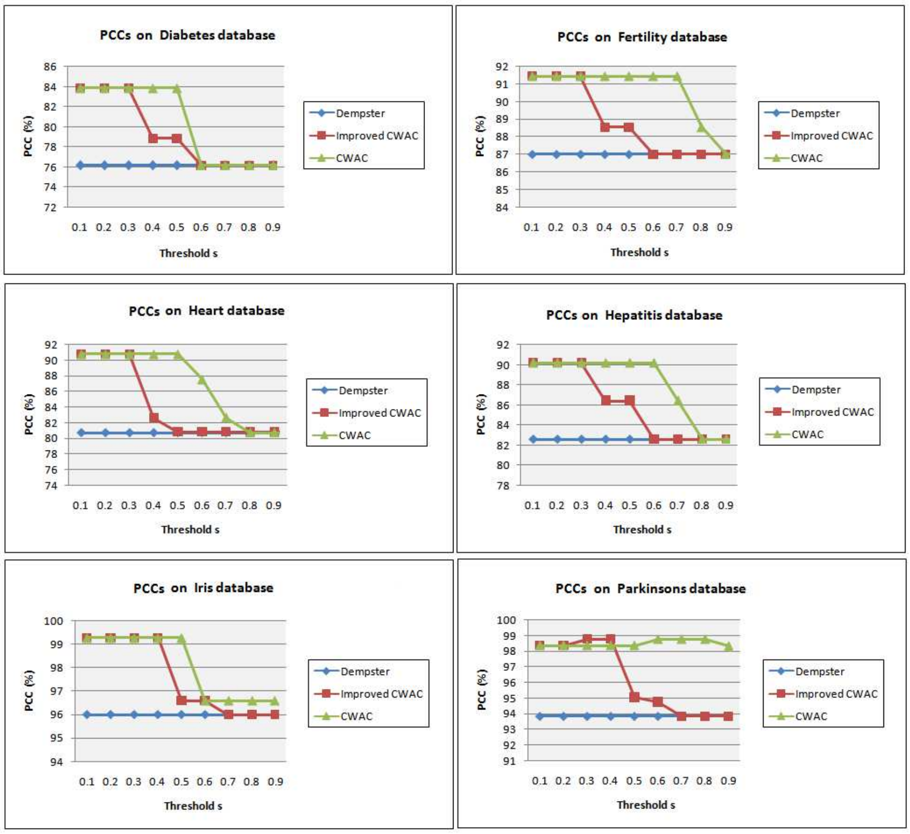
**Table 2.** Single classifier accuracies (%)

| **Databases** | **NB** | **DT** | **1-NN** | **NN** |
|---|---|---|---|---|
| Diabetes | 75.65 | 70.57 | 73.82 | 74.21 |
| Fertility | 87.00 | 83.00 | 85.00 | 87.00 |
| Heart | 82.96 | 75.55 | 75.18 | 78.88 |
| Hepatitis | 83.22 | 81.29 | 80.00 | 79.35 |
| Iris | 95.33 | 95.33 | 95.33 | 96.66 |
| Parkinsons | 69.74 | 96.41 | 87.17 | 91.79 |

### 6.2 Results and discussion

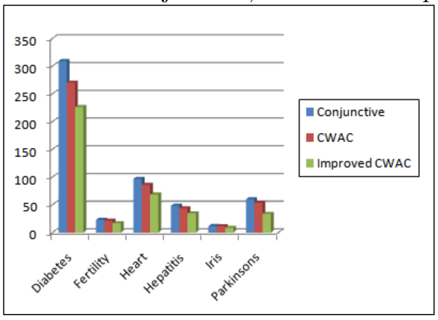
Let us start by comparing the CWAC and the improved CWAC rules with the Dempster one in term of the PCC criterion. Figure 1 illustrates the PCCs for the Dempster, the CWAC and the improved CWAC rules relative to all the mentioned databases. The results, as seen in Figure 1, indicate that for any

**Fig. 1.** PCC values for all databases




threshold $s$ the value of PCC relative to the Dempster rule is constant due to the fact that $n_3$ is always equal to 0 (no rejected instances). It should be noted that the PCC values of the CWAC and the improved CWAC rules are greater or equal to those corresponding to the Dempster rule for the different databases. For instance, the values of $PCC$ relative to the Diabetes database for both CWAC and improved CWAC rules varies from 84% to 76% with the variation of $s$, whereas they are equal to 76% for the Dempster rule $\forall\ s \in S$. This result might be due to the fact that the average number of rejected instance correspond to the CWAC and the improved CWAC rules are greater than 0, whereas that correspond to the Dempster rule is equal to 0. It should be emphasized that this interpretation is available for the remaining databases. Then, we can conclude that the CWAC and the improved CWAC rules are more efficient than the Dempster rule according to the *PCC* criterion. To this end, we move to the comparison of the CWAC and the improved CWAC rules with the conjunctive one. Figure 2 illustrates the conjunctive, the CWAC and the improved CWAC distances correspond to the different databases.

**Fig. 2.** Distance results of the conjunctive, CWAC and improved CWAC rules.




From Figure 2, we can notice that the CWAC and the improved CWAC rules achieve best results compared with the conjunctive rule. In fact, the distances correspond to the CWAC and the improved CWAC rules are lower than those correspond to the conjunctive rule. For example, for Pima Indian Database, the CWAC and the improved CWAC distances are respectively 270.39 and 226.60, whereas the conjunctive distance is equal to 270.42. For Hepatitis database, we have 48.62 as conjunctive distance, 43.63 as CWAC distance and 34.75 as improved CWAC distance. As far, we can assume that this result can be applied to all the other databases. Thus, we can conclude that the CWAC and the improved CWAC rules are more adequate than the conjunctive one according to the distance criterion. Also, from Figure 1, we can remark that the PCCs obtained by the CWAC rule are higher than those obtained by the improved CWAC rule (for all $s \in S$). Moreover, the total distances of the improved CWAC rule are less than those relative to the CWAC rule. Therefore, we can note that the improved CWAC rule is better than the CWAC rule according to the distance criterion but worse than the CWAC rule according to the PCC criterion.

## 7 Conclusion

In this paper, we have outlined some combination rules assuming the pieces of evidence combined to be distinct. Then, we relied on the ensemble classifier system to carry out some experimental tests in the purpose of comparing these alternative rules. The obtained results show the efficiency of the CWAC rule and the improved CWAC rule compared with the conjunctive and Dempster ones. Results of experimentations show also that the improved CWAC rule is the best combination rule according to the distance criterion, whereas the CWAC rule is considered as the best rule of combination in term of the PCC criterion.

## References


1. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
2. X. Deng, Y. Deng, and F. T. S. Chan. An improved operator of combination with adapted conflict. *Annals OR*, 223(1):451–459, 2014.
3. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
4. A. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information fusion*, 2(2):91–101, 2001.
5. E. Lefèvre and Z. Elouedi. How to preserve the conflict as an alarm in the combination of belief functions? *Decision Support Systems*, 56:326–333, 2013.
6. E. Mandler. Combining the classification results of independent classifiers based on the Dempster-Shafer theory of evidence. *Pattern recognition and artificial intelligence*, pages 381–393, 1988.
7. A. Martin. About conflict in the theory of belief functions. In *Belief Functions: Theory and Applications*, pages 161–168. Springer, 2012.
8. D. Mercier, G. Cron, T. Denœux, and M. Masson. Fusion of multi-level decision systems using the transferable belief model. In *7th International Conference on Information Fusion, FUSION'2005*, volume 2, pages 655–658. IEEE, 2005.
9. P. Murphy and D. Aha. UCI repository databases. *http://www.ics.uci.edu/mlearn*, 1996.
10. M. Ponti. Combining classifiers: from the creation of ensembles to the decision fusion. In *24th Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 1–10. IEEE, 2011.
11. G. Shafer. A mathematical theory of evidence, princeton university press. *Princeton university press Princeton*, 1976.
12. P. Smets. The application of the transferable belief model to diagnostic problems. *International Journal of Intelligent Systems*, 13:127–157, 1998.
13. P. Smets. The transferable belief model for quantified belief representation. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1:267–301, 1998.
14. M. Vatsa, R. Singh, A. Noore, and S. K. Singh. Belief function theory based biometric match score fusion: Case studies in multi-instance and multi-unit iris verification. In *7th International Conference on Advances in Pattern Recognition (ICAPR)*, pages 433–436, 2009.
15. S. M. Weiss and I. Kapouleas. An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. In *11th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 781–787. Morgan Kaufmann, 1989.