# Classifier fusion within the belief function framework using dependent combination rules

Asma Trabelsi, Zied Elouedi, Eric Lefevre

# Classifier fusion within the belief function framework using dependent combination rules

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.** The fusion of imperfect data within the framework of belief functions has been studied by many researchers over the past few years. Up to now, there are some proposed combination rules dealing with dependent information sources. Moreover, the choice of one rule among several alternatives is crucial but the criteria to be based on are still non clear. Thus, in this paper, we evaluate and compare some dependent combination rules for selecting the most efficient one under the framework of classifier fusion.

Keywords - belief function theory, combination rules, dependent information sources, multiple classifier fusion.

## 1 Introduction

Pattern recognition has been extensively explored in last decades almost always using ensemble classifiers. Thus, several combination approaches have been proposed to combine multiple classifiers such as plurality, Bayesian theory, belief function theory, etc [6]. This latter has many interpretations such as the Transferable Belief Model (TBM) [6] which offers numerous combination rules. Some of these rules assume the independence of information sources [5], [8] while others deal only with dependent information sources [1, 2]. The choice of the convenient rule is a crucial task but it has not been yet deeply explored. In this paper, we are interested in the combination of multiple classifiers within the framework of belief functions to evaluate and compare some dependent combination rules in order to pick out the most efficient one among them. Basically, we compare the most known dependent combination rules: the cautious conjunctive rule [2], the normalized cautious rule [2] and the cautious Combination With Adapted Conflict rule [1]. The remaining of this paper is organized as follows: we provide in Section 2 a brief overview of the fundamental concepts of the belief function theory. We present three combination rules dealing with dependent sources of information in Section 3. Section 4 is devoted to describing our comparative approach. Experiments and results are outlined in Section 5. Section 6 draws conclusion.

## 2 Fundamental concepts of belief function theory

Let $\Theta$ be a finite non-empty set of $N$ elementary events related to a given problem, called the frame of discernment. The beliefs held by an agent on the different subsets of the frame of discernment $\Theta$ are represented by the so-called basic belief assignment (bba). The bba is defined as follows:

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

The quantity $m(A)$ states the degree of belief committed exactly to the event $A$. From the basic belief assignment, we can compute the commonality function ($q$). It is defined as follows: $q(A) = \sum_{B \supseteq A} m(B)$.

Decision making aims to select the most reasonable hypothesis for a given problem. In fact, it consists of transforming beliefs into probability measure called the pignistic probability denoted by $BetP$ and defined as follows [10]:

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

The dissimilarity between two bbas can be computed. One of the well-known measures is the one proposed by Jousselme [4]:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D (m_1 - m_2)} \tag{3}$$

where $D$ is the Jaccard similarity measure defined by:

$$D(A,B) = \begin{cases} 1 & \text{if } A{=}B{=}\emptyset \\ \dfrac{|A \cap B|}{|A \cup B|} & \forall \ A{,}B \in 2^{\Theta} \end{cases} \tag{4}$$

## 3 Combination of pieces of evidence

The TBM framework offers several tools to aggregate a set of bbas induced form dependent information sources:

1. The cautious conjunctive rule, denoted $\textcircled{\wedge}$, has been proposed by [2] in order to aggregate pieces of evidence induced from reliable dependent information sources using the conjunctive canonical decomposition proposed by Smets [9]. Let $m_1$ and $m_2$ be two non-dogmatic bbas ($m(\Theta) > 0$) and let $m_1 \textcircled{\wedge} m_2$ be the result of their combination. We get:

$$m_1 \textcircled{\wedge} m_2(A) = \textcircled{\cap}_{A \subset \Theta} A^{w_1(A) \wedge w_2(A)} \tag{5}$$

where $w_1(A) \wedge w_2(A)$ represents the weight function of a bba $m_1 \textcircled{\wedge} m_2$ and $\wedge$ denotes the minimum operator. The weights $w(A)$ for every $A \subset \Theta$ can be obtained from the commonalities as follows: $w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|-1}}$.

2. The normalized version of the cautious conjunctive rule, denoted $\textcircled{\wedge}^*$, is obtained by replacing the conjunctive operator $\textcircled{\cap}$ by the Dempster operator $\oplus$ [1] in order to overcome the effect of the value of the conflict generated by the unnormalized version. It is defined by the following equation:

$$m_1 \textcircled{\wedge}^* m_2(A) = \underset{\emptyset \neq A \subset \Theta}{\oplus} A^{w_1(A) \wedge w_2(A)} \tag{6}$$

3. The cautious CWAC rule, based on the cautious rule and inspired from the behavior of the CWAC rule, is defined by an adaptive weighting between the unormalized cautious and the normalized ones [1]. The cautious CWAC rule is then defined as follows $\forall\ A \subseteq \Theta$ and $m_{\textcircled{\wedge}}(\emptyset) \neq 1$:

$$m_{\bigcirc}(A) = D m_{\textcircled{\wedge}}(A) + (1 - D) m_{\textcircled{\wedge}^*}(A) \tag{7}$$

with $D = \underset{i,j}{max}[d(m_i, m_j)]$ is the the maximum Jousselme distance between $m_i$ and $m_j$.

## 4 Comparative study

In our investigation, ensemble classifiers, based on the combination of the outputs of individual classifiers, have been proposed as tools for evaluating and comparing dependent combination rules. Let us consider a pattern recognition issue where $B = \{x_1, ..., x_n\}$ be a data set with $n$ examples, $C = \{C_1, \ldots, C_M\}$ be a set of $M$ classifiers and $\Theta = \{w_1, \ldots, w_N\}$ be a set of $N$ class labels. $B$ will be partitioned into train and test sets. The classifiers must be built from the training set and then we apply them to predict the label class $w_j \in \Theta$ of any pattern test $x$. The outputs from $M$ classifiers should be converted into bbas by taking into account the reliability rate $r$ of each classifier. In fact, for each pattern test we have $M$ bbas obtained as follows:

$$m_i(\{w_j\}) = 1 \text{ and } m_i(A) = 0\ \forall\ A \subseteq \Theta \text{ and } A \neq \{w_j\} \tag{8}$$

with $r_i = \dfrac{\text{Number of well classified instances}}{\text{Total number of classified instances}}$.

Note that $m_i(\{w_j\})$ denotes the part of belief given exactly to the predicted class $w_j$ by the classifier $C_i$.

Once the outputs of all classifiers are transformed into bbas, we move to the combination of classifier through dependent fusion rules. The combination results will allow us to evaluate and compare these rules in the purpose of selecting the most appropriate one based on two popular evaluation criteria: the distance and the Percent of Correct Classification (PCC). This is justified by the fact that the cautious conjunctive rule does not keep the initial alarm role of the conflict due to its absorbing effect of the conflictual mass, the normalized cautious rule ignores the value of the conflict obtained by combining pieces of evidence whereas the cautious CWAC rule gives the conflict its initial role as an alarm signal.

– The PCC criterion, representing the percent of the correctly classified instances, was employed to compare the cautious CWAC rule of combination with the normalized cautious rule. Such case requires the use of three variables $n_1$, $n_2$ and $n_3$ which respectively represent the number of well classified, misclassified and rejected instances. Hence, for each combination rule, we propose the following steps:
  1. We define a tolerance thresholds $S = \{0.1, 0.2, \ldots, 1\}$. For each threshold $s \in S$, we check the mass of the empty set $m(\emptyset)$ induced by any test instance. If $m(\emptyset)$ is greater than $s$, our classifier chooses to reject instance instead of misclassifying it. Consequently, we increment $n_3$. Inversely, we compute the $BetP$ in order to make a decision about the chosen class. Accordingly, we increment $n_1$ if the current class is similar to the real one else we increment $n_2$.
  2. Once we have calculated our well classified, misclassified and rejected instances, we compute then the $PCC$ for each threshold $s \in S$ as follows:

$$PCC = \frac{n_1}{n_1 + n_2} * 100 \tag{9}$$

  The best rule is the one that has the highest values of $PCC \; \forall \; s \in S$.
– The distance criterion, corresponding to the Jousselme distance between two mass functions [4], was used to compare the cautious CWAC rule of combination with the cautious conjunctive rule. Thus, for each combination rule we proceed as follows:
  1. The real class $w_j$ of each pattern test should be converted into a mass function: $m_r(\{w_j\}) = 1$.
  2. Then, we calculate for the instance $x$ the Jousselme distance between the mass function corresponding to its real class ($m_r$) and the mass function produced by combining bbas coming from $M$ classifiers.
  3. Finally, we aggregate the Jousselme distances obtained by all test patterns in order to obtain the total distance.

  The most appropriate rule is the one that has the minimum total distance.

## 5 Empirical evaluation

### 5.1 Experimental settings

In order to evaluate our combination rules, we have performed a set of experiments on several real world databases with different number of instances, different number of attributes and different number of classes obtained from the U.C.I repository [7]. We have conducted experiments with four machine learning algorithms implemented in Weka [3]. These learning algorithms including Naive Bayes, $k$-Nearest Neighbors, Decision tree and Neural Network were run based on a validation approach named leave one out cross validation. This method divides a data set with $N$ instances into $N$-1 parts for training and the remaining instance for testing. This process should be repeated $N$ times where each instance is used once as a test set. Thus, from each classifier we get $N$ test patterns with their predicted class labels.

### 5.2 Experimental results

Let's lead off by comparing the cautious CWAC rule of combination with the normalized cautious one according to the PCC criterion. Figure 1 presents the PCCs for both the normalized cautious and the cautious CWAC rules relative to all the mentioned databases. From Figure 1, we can deduce that for the different

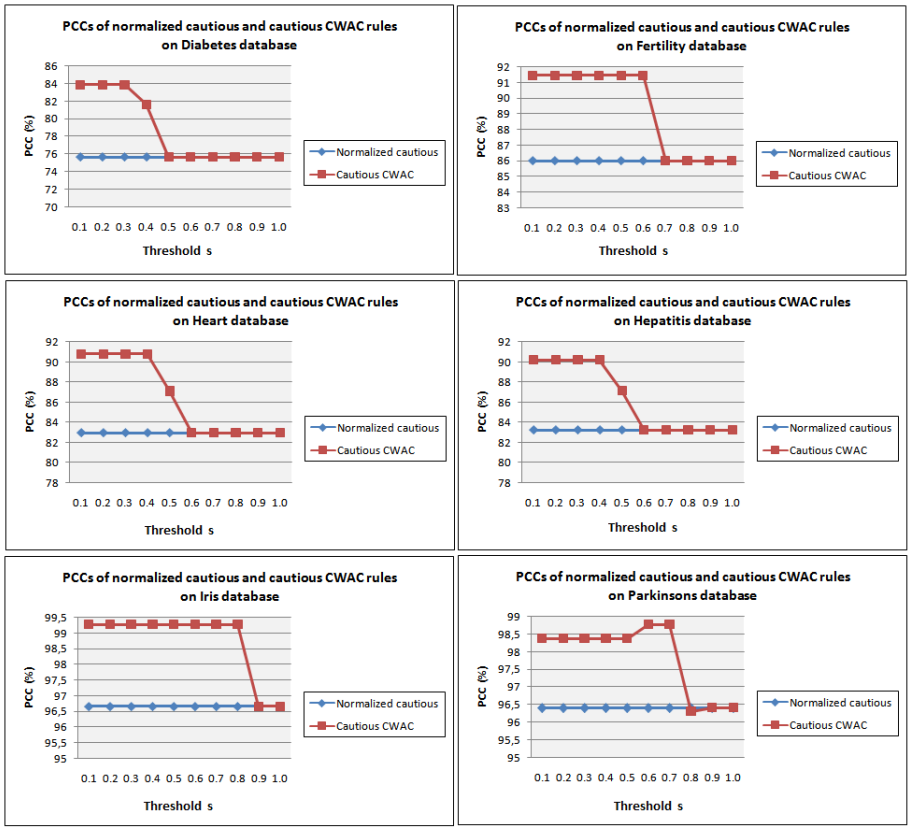


**Fig. 1.** PCCs for all databases

values of $s$, the $PCC$ values of the cautious CWAC rule are greater or equal to those relative to the normalized cautious one for all the mentioned databases. So, we can conclude that the cautious CWAC rule is more efficient than the normalized cautious one in term of $PCC$ criterion.

As shown in Table 1, the cautious CWAC rule achieves best results compared with the cautious conjunctive one. In fact, the distance relative to the cautious CWAC rule is lower than that relative to the cautious conjunctive one. Accordingly, we can conclude that the cautious CWAC rule is more efficient than the cautious conjunctive one in term of distance criterion.

**Table 1.** Distance results of the cautious conjunctive and cautious CWAC rules.

| **Datasets** | **Cautious conjunctive** | **Cautious CWAC** |
|---|---|---|
| Pima Indians Diabetes | 334.87 | 312.45 |
| Fertility | 28.06 | 26.81 |
| Statlog (Heart) | 104.60 | 96.79 |
| Hepatitis | 54.90 | 51.56 |
| Iris | 15.09 | 14.78 |
| Parkinsons | 60.92 | 50.31 |

## 6 Conclusion

In this paper, we have studied some fusion rules when dealing with dependent information sources. Then, we have conducted experimental tests based on multiple classifier systems to judge the efficiency of the cautious CWAC rule compared with the cautious conjunctive and the normalized cautious ones.

## References


1. J. Boubaker, Z. Elouedi, and E. Lefèvre. Conflict management with dependent information sources in the belief function framework. In *14th International Symposium of Computational Intelligence and Informatics (CINTI)*, volume 52, pages 393–398. IEEE, 2013.
2. T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2):234–264, 2008.
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. 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.
7. P. Murphy and D. Aha. UCI repository databases. *http://www.ics.uci.edu/mlearn*, 1996.
8. F. Pichon and T. Denoeux. The unnormalized Dempster's rule of combination: A new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87, 2010.
9. P. Smets. The canonical decomposition of a weighted belief. In *14th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 95, pages 1896–1901, 1995.
10. P. Smets. The transferable belief model for quantified belief representation. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1:267–301, 1998.