Title	Discounting and combination scheme in evidence theory for dealing with conflict in information fusion
Author(s)	Huynh, Van-Nam
Citation	Lecture Notes in Computer Science, 5861/2009: 217-230
Issue Date	2009
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/9070
Rights	This is the author-created version of Springer, Van-Nam Huynh, Lecture Notes in Computer Science, 5861/2009, 2009, 217-230. The original publication is available at www.springerlink.com, http://dx.doi.org/10.1007/978-3-642-04820-3_20
Description	Proceedings of the 6th International Conference, MDAI 2009, Awaji Island, Japan, November 30-December 2, 2009.

Discounting and Combination Scheme in Evidence Theory for Dealing with Conflict in Information Fusion*

Van-Nam Huynh

School of Knowledge Science
Japan Advanced Institute of Science and Technology
Nomi, Ishikawa, 923-1292, JAPAN
Email: huynh@jaist.ac.jp

Abstract. Recently combination rules as well as the issue of conflict management in Dempster-Shafer theory have received considerable attention in information fusion research. Mostly these studies considered the combined mass assigned to the empty set as the conflict and have tried to provide alternatives to Dempster's rule of combination, which mainly differ in the way of how to manage the conflict. In this paper, we introduce a hybrid measure to judge the difference between two bodies of evidence as a basis for conflict analysis, and argue that using the combined mass assigned to the empty set *as a whole* to quantify conflict seems inappropriate. We then propose to use the *discounting* operator in association with the *combination* operator to resolve conflict when combining evidence, in which the discount rate of a basic probability assignment is defined using the entropy of its corresponding pignistic probability function. Finally, an application of this *discounting and combination scheme* to fusion of decisions in classifier combination is demonstrated.

1 Introduction

The Dempster-Shafer theory of evidence (D-S theory, for short), originated from the work by Dempster [6] and then developed by Shafer [32], has appeared as one of the most popular theories for modeling and reasoning with uncertainty and imprecision in intelligent systems. In the D-S theory, Dempster's rule of combination plays a pivotal role serving as a powerful tool for combining evidence from distinct sources of information. According to Dempster's rule [32], the combined mass assigned to the empty set considered as the conflict is distributed proportionally to the other masses. Critically, Zadeh [41] presented an example showing that applying Dempster's rule to conflicting evidence yields counterintuitive results. After Zadeh's example, many alternatives have been proposed in the literature, most notably Smets' unnormalized combination rule [33], Yager's combination rule [39], Dubois and Prade's disjunctive combination rule [10].

* This work was supported by JSPS Grant-in-Aid for Scientific Research (C) #20500202

Lefevre *et al.* [21] have proposed a generic framework for evidence combination which provides a flexible way of distributing the conflict, i.e. the combined mass assigned to the empty set, among subsets of the frame of discernment and allows Dempster's rule as well as the three just mentioned rules of combination to be retrieved within the framework. Recently, motivated by the practical difficulty of verifying the distinctness assumption imposed on combined sources of evidence, Denoeux [9] has proposed two new rules of combination, namely the cautious conjunctive rule and its dual bold disjunctive rule, which are suggested to be suitable for combining belief functions from possibly overlapping bodies of evidence. Although there have been a numerous number of combination rules developed so far, Dempster's rule of combination [32] together with its unnormalized version [33] have been well justified theoretically and have greatly dominated the other rules in information fusion applications, e.g., [2, 3, 5, 8, 7, 19, 31, 38].

In most previous studies on conflict management, it is mainly assumed that the conflict is identified by using the combined mass assigned to the empty set before normalization, denoted by $m_{\oplus}(\emptyset)$, and the thinking of how to manage this mass has basically raised interesting ideas for developing alternatives such as in [13, 21, 39]. Recently, Liu [26] has argued that the use of $m_{\oplus}(\emptyset)$ alone to quantify the conflict might lead to a wrong claim when considering what combination rule would be appropriate for combining conflicting evidence. Instead, Liu proposed to use a pair of quantitative measures, the combined mass allocated to the empty set before normalization, i.e. $m_{\oplus}(\emptyset)$, and the so-called distance between betting commitments, to justify when two pieces of evidence are in conflict. This formal definition of conflict can be served as a prerequisite for selecting appropriate combination rules [26]. Smets [36] has eventually provided a throughout examination of perhaps all existing combination rules and proposed an expert system approach for resolving conflict in evidence combination.

Note that the difference between two distinct bodies of evidence may be not only due to the conflict between two sources of evidence but also due to the complement of each other. For example, different sensors observe an object from different angles may provide different but complementary evidence about it. Although disjunctive consensus rules proposed in the literature such as Dubois and Prade's disjunctive combination rule [10] may be properly applied for combining complementary sources of evidence, the issue of detection of complement between combined bodies of evidence has been completely ignored so far. In the following of this paper, we first introduce a hybrid measure consisting of two components, the *quantitative distance between two mass assignments* and the *qualitative distance between two families of focal sets*, to judge the difference between two bodies of evidence. This hybrid measure can be used as a basis for conflict and complement analysis later on. We then argue that only a part of $m_{\oplus}(\emptyset)$ reflecting the conflict whilst the remainder representing the mass of uncommitted belief as a result of combination.

On the other hand, observing from the previous studies on the conflict analysis which mostly cited Zadeh's famous counterexample [41] to criticize Dempster's rule, we can see that ones assumed combined sources of evidence are still fully reliable to be combined even a large conflict has been identified between them. Naturally, once realized that there is a conflict between sources of evidence, one should behave as if at least one of the sources would be not fully reliable. This issue has been critically discussed by Haenni in [23, 24]. One of reasonable solutions to tackle such situations is to use discounting operator in association with combination [24, 30]. A problem naturally arises here is how to determine which source of evidence is not fully reliable and to what discount rate it should be applied. Haenni [24] and Smets [36] suggested to use a meta-belief structure on combined sources of evidence for modeling this problem. However, it seems practically difficult to obtain such a meta-belief especially in information fusion for pattern recognition applications. In this paper, motivated from Smets' two-level model of belief [34], we propose to define the discount rate of a basic probability assignment based on how sure its commitment is if we use it alone for decision making. More particularly, the discount rate applied to a body of evidence is defined using the entropy of its corresponding pignistic probability function and intuitively, the more committed a basic probability assignment is, the lower discount rate it is applied.

The rest of this paper is organized as follows. In Section 2, we recall necessary concepts in the D-S theory. Section 3 devotes to the analysis of conflict and difference between two bodies of evidence. We particularly ague that the conventional view of $m_\oplus(\emptyset)$ as a whole to reflect conflict may be inappropriate. In Section 4, we propose to use the *discounting and combination scheme* for resolving conflict when combining evidence. Section 5 then illustrates an application of this scheme to ensemble learning for the problem of word sense disambiguation. Finally, some conclusions are presented in Section 6.

2 Basic of Dempster-Shafer Theory of Evidence

In the D-S theory [32], a problem domain is represented by a finite set Θ of mutually exclusive and exhaustive hypotheses, called *frame of discernment*. An important concept of the theory is the so-called *basic probability assignment* (BPA, for short), also called *mass function* or *basic belief assignment* (Smets [34]), $m : 2^\Theta \to [0, 1]$ satisfying

$$m(\emptyset) = 0, \text{ and } \sum_{A \in 2^\Theta} m(A) = 1$$

The quantity $m(A)$ can be interpreted as a measure of the belief that is committed exactly to A, given the available evidence. Note that the condition of $m(\emptyset) = 0$ corresponding to the "closed-world assumption" is not required in the Transferable Belief Model (TBM) introduced by Smets [33]. A subset $A \in 2^\Theta$ with $m(A) > 0$ is called a *focal element* of m. A BPA m is called to be *vacuous* if $m(\Theta) = 1$ and $m(A) = 0$ for all $A \neq \Theta$.

Let us denote $\mathcal{F}_m$ the set of focal elements of m, i.e.

$$\mathcal{F}_m = \{A \in 2^\Theta | m(A) > 0\}$$

Union of all elements in $\mathcal{F}_m$ defines the core of m and the pair $\mathcal{B} = (\mathcal{F}_m, m)$ is called a *body of evidence* (BOE).

Two useful operations that especially play an important role in the evidential reasoning are *discounting* and *Dempster's rule of combination* [32]. The discounting operation is used when a source of information provides a BPA m, but knowing that this source has probability α of reliability. Then one may adopt $(1 - \alpha)$ as one's *discount rate*, resulting in a new BPA m^α defined by

$$m^\alpha(A) = \alpha \times m(A), \quad \text{for any } A \subset \Theta \tag{1}$$

$$m^\alpha(\Theta) = (1 - \alpha) + \alpha \times m(\Theta) \tag{2}$$

Consider now two pieces of evidence on the same frame Θ represented by two BPAs m_1 and m_2. Dempster's rule of combination is then used to generate a new BPA, denoted by $m_\oplus = (m_1 \oplus m_2)$ (also called the orthogonal sum of m_1 and m_2), which is defined, for any $A \in 2^\Theta \setminus \emptyset$, as follows

$$m_\oplus(A) = \frac{\sum\limits_{B \cap C = A} m_1(B)m_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} m_1(B)m_2(C)} \tag{3}$$

where

$$\sum_{B \cap C = \emptyset} m_1(B)m_2(C) \triangleq m_\oplus(\emptyset) \tag{4}$$

is the combined mass assigned to the empty set before normalization. Note that the orthogonal sum combination is only applicable to such two BPAs that verify the condition $m_\oplus(\emptyset) < 1$.

According to Smets' two-level view in TBM [34], when a decision needs to be made, a BPA m encoded the available evidence must be transformed into a so-called *pignistic probability function* $BetP_m : \Theta \to [0, 1]$ defined by

$$BetP_m(\theta) = \sum_{A \subseteq \Theta, \theta \in A} \frac{m(A)}{|A|} \tag{5}$$

where $|A|$ is the cardinality of A. A justification for the necessity of the pignistic transformation in TBM framework is provided in [35]. Here we assume, however, to work under the closed-world assumption, i.e. $m(\emptyset) = 0$.

3 Conflict and Difference Between Two BOEs

3.1 Conflict Revisited

In the research community of Dempster-Shafer theory, the mass associated with $m_\oplus(\emptyset)$ when combining two bodies of evidence with Dempster's rule has long been commonly taken as the only quantity indicating the conflict between two sources of information. The extreme case of fully conflict appears when $m_\oplus(\emptyset) = 1$. Recently, Liu [26] argued that value $m_\oplus(\emptyset)$ cannot be used as a quantitative measure of conflict between two bodies of evidence but only represents the mass of uncommitted belief as a result of combination.

Example 1. Let us consider Liu's example of two identical BPAs $m_1 = m_2$ on frame $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ and $m_1(\theta_i) = 0.2$ for $i = 1, ..., 5$. Then we get $m_\oplus(\emptyset) = 0.8$, which is quite high whilst it appears the total absence of conflict as two BPAs are identical.

More generally, we always get $m_\oplus(\emptyset) > 0$ with two identical BPAs whenever their focal elements define a partition of the frame. Simultaneously, Liu also proposes to use an addition criterion based on the difference between the pignistic probabilities together with value $m_\oplus(\emptyset)$ for judging whether two bodies of evidence are in conflict. Formally, two BPAs m_1 and m_2 are said to be *in conflict* if and only if

$$m_\oplus(\emptyset) > \epsilon \text{ and } \mathrm{difBetP}(m_1, m_2) > \epsilon \tag{6}$$

where $\epsilon \in [0, 1]$ is a threshold of conflict tolerance and $\mathrm{difBetP}(m_1, m_2)$ is defined by

$$\mathrm{difBetP}(m_1, m_2) = \max_{A \subseteq \Theta}(|BetP_{m_1}(A) - BetP_{m_2}(A)|)$$

and called the *distance between betting commitments* of the two BPAs [26].

Basically, by the conclusion that "value $m_\oplus(\emptyset)$ cannot be used as a quantitative measure of conflict between two beliefs, contrary to what has long been taken as a *fact* in the Dempster-Shafer theory community." ([26], page 913) Liu tries to look into an addition criterion, namely $\mathrm{difBetP}(m_1, m_2)$, in order to use in association with value $m_\oplus(\emptyset)$ for revealing the relationship between two BPAs.

Let us consider the following example.

Example 2. Suppose that we have the following pair of BPAs on the same frame $\Theta = \{\theta_i | i = 1, \ldots, 7\}$

$$m_1(\{\theta_1, \theta_2, \theta_3, \theta_4\}) = 1; \text{ and } m_2(\{\theta_4, \theta_5, \theta_6, \theta_7\}) = 1$$

Then, combining these two BPAs produces $m_\oplus(\emptyset) = 0$. That is, in the qualitative view of conflict defined by Liu [26], they do not contradict with each other, or in other words these two BPAs are not in conflict at all. However, using the second criterion we easily get $\mathrm{difBetP}(m_1, m_2) = 0.75$.

In this example, note that m_1 and m_2 have assigned, by definition, the total mass exactly to $\{\theta_1, \theta_2, \theta_3, \theta_4\}$ and $\{\theta_4, \theta_5, \theta_6, \theta_7\}$, respectively, and *to none of the proper subsets of them.* So intuitively these two BPAs are partly in conflict. Clearly, such a partial conflict does not be judged by means of $m_\oplus(\emptyset)$ but $\mathrm{difBetP}(m_1, m_2)$ as shown above. However, they are not in conflict in the sense of (6).

On the other hand, in some information fusion situations, evidence come from different sources may offer complementary information each other but not only being in conflict.

Example 3. Consider the following two BPAs on the frame $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$

$$m_1(\{\theta_1, \theta_2\}) = 0.4, \, m_1(\Theta) = 0.6$$
$$m_2(\{\theta_3, \theta_4\}) = 0.6, \, m_2(\Theta) = 0.4$$

That is, while the event $\{\theta_1, \theta_2\}$ is observable from the first source and becomes unseen from the second one, its complementary event $\{\theta_3, \theta_4\}$ is vice versa. The masses assigned to these events are based on available evidence of corresponding sources, and the unassigned masses are attributed to the whole frame due to ignorance. Intuitively, these two sources of evidence provide complementary information each other rather than they are in conflict. However, we obtain

$$m_{\oplus}(\emptyset) = 0.24, \text{ and difBetP}(m_1, m_2) = 0.4$$

which allows us, in light of Liu's definition above, to conclude that two BPAs are in conflict to some extent.

The above observations suggest that taking $m_{\oplus}(\emptyset)$ as a whole for identifying the conflict seems inappropriate, except the extreme case of fully conflict, i.e. when $m_{\oplus}(\emptyset) = 1$. In the following subsection, we propose a more direct approach to judging the difference between two bodies of evidence, which then together with value $m_{\oplus}(\emptyset)$ can serve for conflict analysis. In the other words, we need to look at the difference between two bodies of evidence before using value $m_{\oplus}(\emptyset)$ for analyzing conflict.

3.2 Difference Between Two BOEs

Let $\mathcal{B}_1 = (\mathcal{F}_{m_1}, m_1)$ and $\mathcal{B}_2 = (\mathcal{F}_{m_2}, m_2)$ be two bodies of evidence on the same frame Θ derived from two distinct sources of information. We first directly define the distance between two BPAs m_1 and m_2, denoted by $d(m_1, m_2)$, as follows

$$d(m_1, m_2) = \max_{A \subseteq \Theta}(|m_1(A) - m_2(A)|) \tag{7}$$

Obviously, $d(m_1, m_2) = 0$ if and only if $m_1 = m_2$. This distance is considered as a *quantitative measure* for judging the difference between two bodies of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$. Now let us denote $\mathrm{dif}_{\mathcal{F}}(m_1, m_2)$ the symmetric difference between two families of focal elements $\mathcal{F}_{m_1}$ and $\mathcal{F}_{m_2}$, i.e.

$$\mathrm{dif}_{\mathcal{F}}(m_1, m_2) = (\mathcal{F}_{m_1} \setminus \mathcal{F}_{m_2}) \cup (\mathcal{F}_{m_2} \setminus \mathcal{F}_{m_1}) \tag{8}$$

It is easily seen that if $\mathrm{dif}_{\mathcal{F}}(m_1, m_2) = \mathcal{F}_{m_1} \cup \mathcal{F}_{m_2}$, and $A \cap B = \emptyset$ for any $A \in \mathcal{F}_{m_1}$ and $B \in \mathcal{F}_{m_2}$, then $m_{\oplus}(\emptyset) = 1$, which corresponds to the extreme case of fully conflict mentioned above. If $\mathrm{dif}_{\mathcal{F}}(m_1, m_2) = \emptyset$ and $d(m_1, m_2) > 0$, then qualitatively two sources are not in conflict but having different preferences in distributing their masses to focal elements. This qualitative measure $\mathrm{dif}_{\mathcal{F}}(m_1, m_2)$ allows us to see how different between two sources in realization of the question of where the true hypothesis lies.

Let us denote

$$\mathrm{dif}(\mathcal{B}_1, \mathcal{B}_2) = \langle d(m_1, m_2), \mathrm{dif}_{\mathcal{F}}(m_1, m_2) \rangle \tag{9}$$

and call it the *difference measure of two bodies of evidence*. It is clearly that the conflict between two bodies of evidence originates from either or both of

$d(m_1, m_2)$ (quantitative) and $\mathrm{dif}_{\mathcal{F}}(m_1, m_2)$ (qualitative). Actually, Liu's criterion of using $\mathrm{difBetP}(m_1, m_2)$ is somewhat weaker than using the direct distance of $d(m_1, m_2)$. For example, consider the pair of BPAs given in Example 2 we have $d(m_1, m_2) = 1$ whilst $\mathrm{difBetP}(m_1, m_2) = 0.75$. Note further that if $m_1 = m_2$ we have $\mathrm{difBetP}(m_1, m_2) = 0$ but the reverse does not hold in general.

We now argue that only a part of value $m_\oplus(\emptyset)$ should be used to quantify a conflict qualitatively stemming from $\mathrm{dif}_{\mathcal{F}}(m_1, m_2)$. Let

$$m_\oplus^{\mathrm{comb}}(\emptyset) = \sum_{A,B \in \mathcal{F}_1 \cap \mathcal{F}_2, A \cap B = \emptyset} m_1(A) m_2(B) \tag{10}$$

Clearly, $m_\oplus^{\mathrm{comb}}(\emptyset)$ is a part of $m_\oplus(\emptyset)$ and intuitively representing the mass of uncommitted belief as a result of combination rather than a conflict, which, however, may be properly represented by the remainder of $m_\oplus(\emptyset)$, i.e.

$$m_\oplus(\emptyset) - m_\oplus^{\mathrm{comb}}(\emptyset) \stackrel{\triangle}{=} m_\oplus^{\mathrm{conf}}(\emptyset) \tag{11}$$

Interestingly enough, with this formulation of conflict, the fact used to question the validity of Dempster's rule that two identical probability measures are always conflicting becomes inappropriate.

Example 4. Consider again two BPAs considered in Example 1, which are identical. Then we get $m_\oplus^{\mathrm{comb}}(\emptyset) = 0.8$ and $m_\oplus^{\mathrm{conf}}(\emptyset) = 0$, and hence no conflict appears between the two sources at all. Generally, we always get $m_\oplus^{\mathrm{conf}}(\emptyset) = 0$ whenever two BPAs being combined are identical. Now, looking at Zadeh's famous counterexample with two BPAs m_1 and m_2 defined on $\Theta = \{a, b, c\}$ as: $m_1(a) = 0.99$, $m_1(b) = 0.01$ and $m_2(c) = 0.99$, $m_2(b) = 0.01$, we have $m_\oplus^{\mathrm{conf}}(\emptyset) = 0.98$, which accurately reflects a very high conflict between two BPAs. With such a high conflict but still assuming both sources are fully reliable to proceed with directly applying Demspter's rule on them (to get unsatisfactory results) seems irrational.

Intuitively, the information from $\mathrm{dif}(\mathcal{B}_1, \mathcal{B}_2)$ and $m_\oplus^{\mathrm{conf}}(\emptyset)$ may properly provide helpful suggestions for conflict management on selecting appropriate combination rules in some typical situations.

- If $\mathrm{dif}_{\mathcal{F}}(m_1, m_2) = \mathcal{F}_1 \cup \mathcal{F}_2$ and $A \cap B = \emptyset$ for any $A \in \mathcal{F}_{m_1}$ and $B \in \mathcal{F}_{m_2}$, we have $m_\oplus^{\mathrm{conf}}(\emptyset) = 1$ and two sources are fully conflict. In this case a discounting and then combination strategy should be applied, where different attitudes may suggest different combination rules for use.
- If $\mathrm{dif}_{\mathcal{F}}(m_1, m_2) = \emptyset$ and $d(m_1, m_2) > 0$, we have $m_\oplus^{\mathrm{conf}}(\emptyset) = 0$ and two sources qualitatively are not in conflict but having different beliefs attributed to focal elements. In this situation, a compromise attitude may suggest to use the *trade-off* rule [10], or its special case of averaging operator.
- If $\mathrm{dif}_{\mathcal{F}}(m_1, m_2) \neq \emptyset$, then we have $d(m_1, m_2) > 0$. In this situation, if $m_\oplus(\Theta) = m_1(\Theta) m_2(\Theta) > 0$ two sources may provide complementary information each other as in the case of Example 3 above, and then Dempster's

rule can be applied. If $m_{\oplus}(\Theta) = 0$, two sources may be in a partial conflict and then depending on value $m_{\oplus}^{\mathrm{conf}}(\emptyset)$ whether it is tolerated and information on meta-belief is available or not, one may apply discounting and then combination strategy or a disjunctive combination rule.

The issue of justifying whether two bodies of evidence are in conflict plays an especially important role in selecting alternative combination rules [36]. Thus, identification of conflict should be analyzed as carefully as possible. An accurate determination of the origin of conflicts can also help to manage them properly. In addition, it is our opinion that justifying whether two bodies of evidence are in complementary each other, which has been ignored so far, also have some impact in the mentioned selection problem and should be incorporated into the conflict analysis. Intuitively, the role of value $m_{\oplus}(\Theta)$ may play for this purpose in a somehow similar fashion to that of $m_{\oplus}^{\mathrm{conf}}(\emptyset)$ for conflict analysis as roughly mentioned above; however, this is not a main topic of this paper.

4 Discounting and Combination Scheme

Previously, a common explanation for counterintuitive results yielded by applying Dempster's rule of combination is that possible conflicts between different sources of evidence are mismanaged by Dempster's rule, and this explanation has motivated for developing alternatives combination rules, which are mainly different in the way of managing possible conflicts [24]. Unfortunately, these alternatives are generally not associative, e.g. [10, 21, 39], and thus making them difficult to be applied in practice. In [23], Haenni also presented a critical note on the increasing number of possible combination rules.

Once possible conflicts have been identified, we may naturally wonder about the reliability of different sources of evidence being combined. If a meta-belief of the sources is available, we can first use the discounting operator for BPAs envolved and then apply Dempster's rule to discounted BPAs for combining them. The idea of using the discounting operator to resolve conflict has already been suggested in, i.e., [23, 30, 36]. However, in practice such a beta-belief is not always available, particularly in situations of applying the Dempster-Shafer theory to, for instance, information fusion in pattern recognition (see, e.g., [2, 3, 19, 29, 38]).

According to Smets' two-level view of evidence [34], to make decisions based on evidence, beliefs encoding evidence must be transformed into probabilities using the so-called pignistic transformation. Guided by this view, we propose to discount a BPA involving in combination based upon how sure in its decision when it is used alone for decision making. More particularly, we provide a method for defining discount rates of BPAs being combined using the entropy of their corresponding pignistic probability functions.

Let m_1 and m_2 be two BPAs on the frame Θ and $BetP_{m_1}$ and $BetP_{m_2}$ be pignistic probability functions of m_1 and m_2, respectively. For $i = 1, 2$, we denote

$$H(m_i) = -\sum_{\theta \in \Theta} BetP_{m_i}(\theta) \log_2(BetP_{m_i}(\theta))$$

the Shannon entropy expression of pignistic probability distribution $BetP_{m_i}$. This measure has been used in Jousselme *et al.* [14] as an *ambiguity measure* of belief functions.

Clearly, $H(m_i) \in [0, \log_2(|\Theta|)]$. We now define the *discount rate* of BPA m_i $(i = 1, 2)$, denoted by $\delta(m_i)$, as follows

$$\delta(m_i) = \frac{H(m_i)}{\log_2(|\Theta|)} \tag{12}$$

That is, the higher uncertainty (in its decision) a source of evidence is, the higher discount rate it is applied. Once discount rates have been defined, the discounting and combination strategy applied to two BPAs m_1 and m_2 can be generally formulated in the following form

$$m_{\oplus} = m_1^{(1-\delta(m_1))} \oplus m_2^{(1-\delta(m_2))} \tag{13}$$

where $\oplus$ is a combination operator in general and $m_i^{(1-\delta(m_i))}$ is the discounted BPA obtaining from m_i after discounting at a rate of $\delta(m_i)$ [refer to (1)-(2)].

It is of interest to note that if, for example, $\delta(m_1) = 1$, i.e. $BetP_{m_1}$ is the uniform distribution on Θ or m_1 is at the most uncertain in its decision, $m_1^{(1-\delta(m_1))}$ becomes a vacuous BPA and then plays no role in combination if Dempster's rule is applied. In other words, a decision made using the combined evidence represented by m then depends on the second source of evidence represented by m_2 only.

As for illustration, this discounting and combination strategy will be applied for combining multiple classifiers in the following section. Here Dempster's rule and averaging operator are used for combination. Thanks to its associativity, we can develop an efficient algorithm for combining multiple classifiers with Dempster's rule, where soft decisions by individual classifiers typically are represented in forms of probability distributions over the set of possible classes. Also, although simple in computation, averaging is suggested as providing a good solution to balance multiple evidence [27].

5 An Illustrative Application

Applying the D-S theory to classifier combination has received attention since early 1990s, e.g., [2, 3, 29, 38]. In these methods, it is usually assumed that the involved individual classifiers provide fully reliable sources of information for identifying the label of a particular input pattern, i.e. discounting operator plays no role there. In this section, we present an illustration for applying the discounting and combination scheme discussed above to ensemble learning for the problem of word sense disambiguation (WSD) [12], which has received much interest and concern since the 1950s and is still one of the most challenging tasks in NLP.

Actually, Le *et al.* [19] recently have attempted to apply the D-S theory for weighted combination of classifiers for WSD, in which the weighting is also modeled by the discounting operator. However, their method of defining discounting factors for individual classifiers is based on the strength of individual classifiers, which is determined by testing them on a designed sample data set and therefore does not be influenced by an input pattern under classification. Here, in the context of classification problem, the discounting method discussed above in this paper provides a new way of adaptively weighting individual classifiers based on ambiguity measures associated with their outputs corresponding to a particular pattern under consideration.

5.1 WSD

Roughly speaking, WSD is the task of associating a given word in a text or discourse with an appropriate sense among numerous possible senses of that word. This is an "intermediate task" which necessarily accomplishes most natural language processing tasks such as grammatical analysis and lexicography in linguistic studies, or machine translation, man-machine communication, message understanding in language understanding applications [12].

During the last two decades, many machine learning techniques and algorithms have been applied for WSD, including Naive Bayesian (NB) model, decision trees, exemplar-based model, support vector machines (SVM), maximum entropy models (MEM), etc. [1, 20]. On the other hand, as observed in studies of classification systems, the set of patterns misclassified by different learning algorithms would not necessarily overlap [25]. This means that different classifiers may potentially offer complementary information about patterns to be classified. This observation highly motivated the interest in combining classifiers to build an ensemble classifier which would improve the performance of the individual classifiers. Particularly, classifier combination for WSD has been received considerable attention recently from the community as well.

5.2 Individual Classifiers In Combination

To build individual classifiers for combination, we use three well-known statistical learning methods including the Naive Bayes (NB), Maximum Entropy Model (MEM), and Support Vector Machines (SVM). The selection of these learning methods is basically guided by the direct use of output results for defining BPAs in the present work. Clearly, the first two classifiers produce classified outputs which are probabilistic in nature. Although a standard SVM classifier does not provide such probabilistic outputs, the issue of mapping SVM outputs into probabilities has been studied [28] and recently become popular for applications requiring posterior class probabilities [3, 22]. We have used the library implemented for maximum entropy classification available at [37] for building the MEM classifier, whilst the SVM classifier is built based upon LIBSVM implemented by Chang and Lin [4], which has the ability to deal with the multiclass classification problem and output classified results as posterior class probabilities.

Due to the limitation of page number, the technical detail of these methods as well as the discounting and combination strategy applied to them is omitted here (see [18] for the detail). Informally, the output of individual classifiers is used to define corresponding BPAs. Then we apply the discounting and combination strategy discussed in Section 4 to these BPAs and the final decision is made based on the resulted BPA. Two combination rules are applied in this application, namely Dempster's rule of combination and averaging. Accordingly, we develop two algorithms corresponding to these combination rules, namely *discounting-and-orthogonal sum combination algorithm* and *discounting-and-averaging combination algorithm*, respectively.

5.3 Experimental Results

Test Data. As for evaluation of exercises in automatic WSD, three corpora so-called Senseval-1, Senseval-2 and Senseval-3 were built during three corresponding workshops held in 1998, 2001, and 2004 respectively. Here, the developed combination algorithms will be tested on English lexical samples of Senseval-2 and Senseval-3. Currently, these two datasets are widely used in current WSD studies. The detail of these data sets can be referred to Kilgarriff [15] for Senseval-2 and to Mihalcea *et al.* [17] for Senseval-3.

Like Le *et al.* [19], we use the evaluation method proposed by Melamed and Resnik in [16], which provides a scoring method for exact matches to fine-grained senses as well as one for partial matches at a more coarse-grained level. Also, like most related studies, the fine-grained score is computed in the following experiments.

Results. Table 1 below provides the experimental results obtained by three individual classifiers and two combination algorithms developed, where DCA_1 and DCA_2 stand for the discounting-and-orthogonal sum combination algorithm and the discounting-and-averaging combination algorithm, respectively. The obtained results show that combined classifiers always outperform individual classifiers participating in the corresponding combination. It is of interest to see that the results yielded by the discounting-and-averaging combination algorithm (i.e., DCA_2) are comparable or even better than that given by the discounting-and-orthogonal sum combination algorithm (i.e., DCA_1), while the former is computational more simple than the latter. Although the averaging operation was actually mentioned briefly by Shafer [32] for combining belief functions, it has been almost completely ignored in the studies of information fusion and particularly classifier combination with D-S theory. Interestingly, Shafer [32] did show that discounting in fact turns combination into averaging when all the information sources being combined are highly conflicting and have been sufficiently discounted. This might, intuitively, provide an interpretation for a good performance of DCA_2.

To have a comparative view of obtained results, Table 2 provides comparative results of the developed algorithms with previous studies, namely the best systems in the contests for the English lexical sample tasks of Senseval-2 [15], Senseval-3 [17], and the method developed by Le *et al.* [19]. The best system of

Table 1. Experimental results

%	Individual			Combination	
	NB	MEM	SVM	DCA_1	DCA_2
Senseval-2	65.6	65.5	63.5	66.3	66.5
Senseval-3	72.9	72.0	72.5	73.3	73.3

Senseval-2 contest also used a combination technique: the output of subsystems (classifiers) which were built based on different machine learning algorithms were merged by using weighted and threshold-based voting and score combination [40]. The best system of Senseval-3 contest used the Regularized Least Square Classification (RLSC) algorithm with a correction of the a priori frequencies (for more details, see [11]). This comparative result shows that both developed combination algorithms deriving from the discounting and combination scheme yield an improvement in overall accuracy compared to previous work for WSD in the tests with Senseval-2 and Senseval-3.

Table 2. A comparative result

%	Best systems	Le [19]	DCA_1	DCA_2
Senseval-2	64.2	64.7	66.3	66.5
Senseval-3	72.9	72.4	73.3	73.3

6 Conclusions

In this paper, we have introduced a difference measure of two bodies of evidence serving as a basis for conflict analysis in Dempster-Shafer theory. We argued that the combined mass allocated to the empty set should be divided into two parts, one part represents the mass of uncommitted belief as a result of combination whilst the other reflects the conflict. Interestingly, this analysis might help to solve the question of the validity of Dempster's rule by the fact that two identical probability measures are always conflicting. We have also proposed the use of the discounting operator together with the combination operator for resolving conflict when combining evidence, in which an entropy-based method for defining discounting factors was introduced. As for illustrating the applicability of the proposed discounting and combination scheme, we have also provided an experimental study in combining multiple classifiers for WSD which produces better results in comparison to previous related studies.

References

1. E. Agirre, P. Edmonds (Eds.), *Word Sense Disambiguation: Algorithms and Applications* (Springer, Dordrecht, the Netherlands 2006).
2. A. Al-Ani, M. Deriche, A new technique for combining multiple classifiers using the Dempster–Shafer theory of evidence, *J. Artif. Intell. Res.* **17** (2002) 333–361.
3. D. Bell, J. W. Guan, Y. Bi, On combining classifiers mass functions for text categorization, *IEEE Trans. Know. Data Eng.* **17** (2005) 1307–1319.
4. C. C. Chang, C. J. Lin, LIBSVM: A library for support vector machines, http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.
5. F. Delmotte, P. Smets, Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model, *IEEE Trans. Syst., Man, Cybern. A* **34** (2004) 457–471.
6. A. P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Stat.* **38** (1967) 325–339.
7. T. Denoeux, A neural network classifier based on Dempster-Shafer theory, *IEEE Trans. Syst., Man, Cybern. A* **30** (2000) 131–150.
8. T. Denoeux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, *IEEE Trans. Syst., Man, Cybern.* **25** (1995) 804–813.
9. T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, *Artif. Intell.* **172** (2008) 234–264.
10. D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, *Comput. Intell.* **4** (1988) 244–264.
11. C. Grozea. Finding optimal parameter settings for high performance word sense disambiguation, *Proc. of ACL/SIGLEX Senseval-3*, Barcelona, Spain, July 2004, pp. 125–128.
12. N. Ide, J. Véronis, Introduction to the special issue on word sense disambiguation: The state of the art, *Comput. Ling.* **24** (1998) 1–40.
13. A. Josang, The consensus operator for combining beliefs, *Artif. Intell.* **141** (2002) 157–170.
14. A. L. Jousselme, C. Liu, D. Grenier, E. Bosse, Measuring ambiguity in the evidence theory, *IEEE Trans. Syst., Man, Cybern. A* **36** (2006) 890–903.
15. A. Kilgarriff, English lexical sample task description, *Proc. of Senseval-2: Second Inter. Workshop on Evaluating Word Sense Disambiguation Syst.*, 2001, Toulouse, France, pp. 17–20.
16. I. D. Melamed, P. Resnik, Tagger evaluation given hierarchical tag sets, *Comp. and The Human.* **34** (1-2) (2000) 79–84.
17. R. Mihalcea, T. Chklovski, A. Killgariff, The Senseval-3 English lexical sample task, *Proc. of ACL/SIGLEX Senseval-3*, Barcelona, Spain, July 2004, pp. 25–28.
18. V.-N. Huynh, T. T. Nguyen, C. A. Le, Adaptively entropy-based weighting classifiers in combination using Dempster-Shafer theory for word sense disambiguation, *Comp. Speech Lang.* (to appear).
19. C. A. Le, V.-N. Huynh, A. Shimazu, Y. Nakamori, Combining classifiers for word sense disambiguation based on Dempster-Shafer theory and OWA operators, *Data Know. Eng.* **63** (2007) 381–396.
20. Y. K. Lee and H. T. Ng, 2002. An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. *Proc. of EMNLP*, pages 41–48.
21. E. Lefevre, O. Colot, P. Vannoorenberghe, Belief function combination and conflict management, *Infor. Fusion* **3** (2002) 149–162.

22. H.-T. Lin, C.-J. Lin, R. C. Weng, A note on Platts probabilistic outputs for support vector machines, *Mach. Learn.* **68** (2007) 267–276.
23. R. Haenni, Are alternatives to Dempsters rule of combination alternatives?, *Infor. Fusion* **3** (2002) 237–241.
24. R. Haenni, Shedding new light on Zadeh's criticism of Dempster's rule of combination, in: FUSION'05, *8th Inter. Conf. on Infor. Fusion*, pp. 879–884, 2005.
25. J. Kittler, M. Hatef, R. P. W. Duin, J. Matas, On combining classifiers, *IEEE Trans. Patt. Anal. Mach. Intell.* **20** (1998) 226–239.
26. W. Liu, Analysing the degree of conflict among belief functions, *Artif. Intell.* **170** (2006) 909–924.
27. C. Murphy, Combining belief functions when evidence conflicts, *Dec. Sup. Syst.* **29** (2000) 1–9.
28. J. Platt, Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A. Smola, P. Bartlett, B. Schölkopf, D. Schuurmans (Eds.), *Advances in Large Margin Classifiers* (Cambridge: MIT Press, 2000).
29. G. Rogova, Combining the results of several neural network classifiers, *Neural Networks* **7** (1994) 777–781.
30. E. H. Ruspini, J. D. Lowrance, T. M. Strat, Understanding evidential reasoning, *Inter. J. Approx. Reason.* **6** (1992) 401–424.
31. R. J. Safranek, S. Gottschlich, A. C. Kak, Evidence accumulation using binary frames of discerment for verification vision, *IEEE Trans. Robot. Autom.* **6** (1990) 405–417.
32. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).
33. P. Smets, The combination of evidence in the transferable belief model, *IEEE Trans. Patt. Anal. Mach. Intell.* **12** (1990) 447–458.
34. P. Smets, R. Kennes, The transferable belief model, *Artif. Intell.* **66** (1994) 191–234.
35. P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, *Inter. J. Approx. Reason.* **38** (2004) 133–147.
36. P. Smets, Analyzing the combination of conflicting belief functions, *Infor. Fusion* **8**(2007) 387–412.
37. Y. Tsuruoka, A simple C++ library for maximum entropy classification, http://www-tsujii.is.s.u-tokyo.ac.jp/∼tsuruoka/maxent/, 2006.
38. L. Xu, A. Krzyzak, C. Y. Suen, Several methods for combining multiple classifiers and their applications in handwritten character recognition, *IEEE Trans. Syst., Man, Cybern.* **22** (1992) 418–435.
39. R. R. Yager, On the Dempster-Shafer framework and new combination rules, *Infor. Sci.* **41** (1987) 93–138.
40. D. Yarowsky, S. Cucerzan, R. Florian, C. Schafer, R. Wicentowski, The Johns Hopkins Senseval2 system descriptions, *Proc. of SENSEVAL2*, 2001, pp. 163–166.
41. L. A. Zadeh, Reviews of Books: A Mathematical Theory of Evidence, *AI Magazine* **5** (1984) 81–83.