Mass Function Derivation and Combination in Multivariate Data Spaces☆

Hui Wang, Jun Liu, Juan Carlos Augusto

School of Computing and Mathematics,University of Ulster, UK

Abstract

In this paper, we consider the problem of how to derive mass functions systematically from data samples. We also consider the ensuing problem of how to combine different mass functions derived in this way. We show that a mass function can be efficiently and systematically derived from multivariate data. We also demonstrate that combining mass functions obtained in this manner can be done easily. The way of deriving and combining mass functions is illustrated with a simple example.

Key words: Dempster-Shafer theory of evidence, mass function, Dempster rule of combination

1. Introduction

The *mathematical theory of evidence*, also known as the *Dempster-Shafer Theory of Evidence* (DSTE) [27], is a generalization of the Bayesian theory of probability. In recent years, there has been increased interest in advancing this theory, developing efficient algorithms, and applying the theory itself to a wide range of engineering and business problems, see [21, 1, 25, 12, 37, 14, 17, 35, 13, 26, 34, 15, 30, 19].

Using DSTE to solve a specific problem usually involves three steps. First, we obtain a *mass function* that represents uncertainties existing in the problem using independent items of evidence. Second, we compute various belief

☆This is an extended version of a paper [31] presented at the 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid, Spain, 21 - 24 September 2008.

Email address: {h.wang,j.liu,jc.augusto}@ulster.ac.uk (Hui Wang, Jun Liu, Juan Carlos Augusto)

functions on the basis of the mass function. Third, if necessary, we combine multiple mass functions into a single one. The key problem is how to obtain a mass function in a systematic way.

The mass function is usually assumed to be given, e.g., by experts, but in practice we need to derive it from data in one way or another. For a particular task, we can derive a single mass function from data (see e.g., [9]). Alternatively, we can also derive multiple mass functions from several sources of data, then combine them into a single mass function (see e.g. [11]).

If we want to derive mass functions from data, it is important that the information in data is utilized as much as possible and is encoded in the mass function. One way of achieving this objective is to derive mass functions in a systematic way.

In this paper, we review recent developments in the systematic derivation of mass functions from multivariate data sets (see [33]) and discuss the ensuing problem of how to combine the already derived mass functions.

2. A Brief Review of the Dempster-Shafer Theory of Evidence

Let V be a finite set. The power set of V, 2^V, is called a *frame of discernment*.

Definition 1. *A* mass function *is a mapping* $m : 2^V \rightarrow [0, 1]$ *such that* $\sum_{x \in 2^V} m(x) = 1$ *and* $m(\emptyset) = 0$.

The mass $m(x)$ expresses the amount of belief that is allocated to x.

With every mass function, there are associated belief (bel) and plausibility (pls) functions, which are mappings $2^V \rightarrow [0, 1]$ such that for any $e \in 2^V$, $bel(e) = \sum_{x \in 2^V, x \subseteq e} m(x)$ and $pls(e) = \sum_{x \in 2^V, x \cap e \neq \emptyset} m(x)$. The interval $[bel(e), pls(e)]$ contains the precise probability of $e \subseteq V$ in the classical sense. That is, $bel(e) \leq P(e) \leq pls(e)$. It is known that $pls(e) = 1 - bel(\bar{e})$, where $\bar{e}$ is the complement of e with respect to V.

Given two independent mass functions, we sometimes need to combine them into a single mass function. This can be done using a number of rules (e.g., *Dempster rule* [10], Yager's *modified Dempster rule* [36], Inagaki's *unified combination rule* [20], Zhang's *center combination rule* [38], Dubois and Prade's *disjunctive consensus rule* [16]). The best known is the Dempster rule of combination, which is a generalization of the Bayes rule. This rule strongly emphasises the agreement between multiple sources of evidence and ignores the disagreement by the use of a normalization factor.

Let m_1 and m_2 be two mass functions. The combined mass function, obtained through the Dempster rule of combination, is $m_{12} : 2^V \rightarrow [0, 1]$, such that for any $A \in 2^V$,

$$\begin{aligned} m_{12}(A) &= 0, && \text{if } A = \emptyset \\ m_{12}(A) &= \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K}, && A \neq \emptyset \end{aligned}$$

where $K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$. The normalization factor, $1 - K$, has the effect of completely ignoring conflict and attributing any mass associated with conflict to the null set [39]. Consequently, this operation yields counterintuitive results in the face of significant conflict present in certain contexts.

3. A Systematic Derivation of Mass Functions

Many applications are data driven, such as those encountered in data mining, where data samples (i.e., observations) are exploited to solve the problem. If we can systematically derive a mass function from data we can then apply the Dempster-Shafer theory to solve the problem. In what follows we propose a general method for this purpose that is aimed to maximally capture the available information in data.

3.1. Multivariate Data Space

The starting point of our investigations is a probability space $< V, \mathcal{F}, P >$, where V is a set or *data space*, $\mathcal{F}$ is a σ-field [3] on V called *event space*, and P is a probability function $P : \mathcal{F} \rightarrow [0, 1]$. If V is finite, the power set 2^V is an event space over V. For different types of data, the notions of data space and event space must be precisely defined.

Multivariate data are one of the commonly encountered forms of data. We assume that the data space is defined by a set of attributes $R = \{a_1, a_2, \cdots, a_n\}$. The domain of attribute a_i is denoted by $dom(a_i)$, but for simplicity we will also use a_i to represent the domain of attribute a_i in equations assuming that there is no risk of confusion. The multivariate data space is regarded as the Cartesian product of the domains, i.e., $V \stackrel{\text{def}}{=} \prod_{i=1}^{n} dom(a_i)$. A data set D is a sample of V, that is, $D \subseteq V$.

We consider two types of attributes, *numerical* and *categorical* [29]. A numerical attribute has an inherent order defined over its domain, and the operations allowed there are equality, inequality, 'greater than', 'less than', addition, subtraction, multiplication and division. A categorical attribute has no order defined over its domain, and the only operations allowed are equality and inequality. For simplicity, we assume the domain of any attribute is finite unless otherwise stated. In the case of numerical attributes, we assume their domains are discretized and then represented as natural numbers.

An element $t \in V$ is a simple tuple $t \stackrel{\text{def}}{=} \langle t_1, t_2, \cdots, t_n \rangle$, where $t_i \in dom(a_i)$. We consider a generalized tuple $h =< h_1, h_2, \cdots, h_n >$ where $h_i \subseteq dom(a_i)$ if a_i is categorical, and h_i is an interval over $dom(a_i)$ if it is numerical. Such a generalized tuple is called a *hypertuple over* R [32]. We consider the event space as the set of all hypertuples. It can be shown that this set is a σ-field on V. More formally, $\mathcal{F} = \prod_{i=1}^{n} \mathcal{F}_i$, where $\mathcal{F}_i$ is the power set over $dom(a_i)$ if a_i is categorical, and $\mathcal{F}_i$ is the set of all intervals over $dom(a_i)$ if a_i is numerical.

Consider two hypertuples h_1 and h_2, where $h_1 = \langle h_{11}, h_{12}, \cdots, h_{1n} \rangle$ and $h_2 = \langle h_{21}, h_{22}, \cdots, h_{2n} \rangle$. The hypertuple h_1 is *covered* by h_2 (or h_2 *covers* h_1), written $h_1 \leq h_2$, if for $i \in \{1, 2, \cdots, n\}$,

$$\begin{cases} \min(h_{2i}) \leq x \leq \max(h_{2i}) \text{ for all } x \in h_{1i} & \text{if } a_i \text{ is numerical} \\ h_{1i} \subseteq h_{2i} & \text{if } a_i \text{ is categorical} \end{cases}$$

3.2. Deriving mass functions

We consider a general multivariate data space V as specified in Section 3.1. The frame of discernment is taken to be the event space $\mathcal{F}$, the set of all hypertuples in the given multivariate data space V. This frame of discernment $\mathcal{F}$ is clearly more general than the classical power set.

Let D be a data set – being regarded as a set of simple tuples. For any $e \in \mathcal{F}$, we let $e^D \stackrel{\text{def}}{=} \{x \in D : x \leq e\}$, i.e., the set of all elements of D that are covered by e.

Definition 2. *We define a function* $\bar{m} : \mathcal{F} \to [0, 1]$ *such that for any* $e \in \mathcal{F}$,

$$\bar{m}(e) \stackrel{\text{def}}{=} \frac{|e^D|}{M}, \tag{1}$$

where $M = \sum_{x \in \mathcal{F}} |x^D|$ *is a normalization factor.*

Clearly, function $\bar{m}$ is a mass function over $\mathcal{F}$. In addition to the two properties required in the standard definition of mass function (Def.1), the mass function defined in Def.2 comes with the following additional properties:

Lemma 1. *Consider $x, y \in \mathcal{F}$.*

- *$\bar{m}(x) \leq \bar{m}(y)$ if $x \subseteq y$. Therefore $\bar{m}(V)$ is maximal. Note that $V \in \mathcal{F}$.*
- *$\bar{m}(x \cup y) = \bar{m}(x) + \bar{m}(y) - \bar{m}(x \cap y)$*

Computing e^D is straightforward. To compute M by definition, we need to go through all elements in $\mathcal{F}$ – this is clearly expensive in computational terms. In fact, the computational complexity is exponential in the size of V. Fortunately, as is shown in [33], computing M can be done by computing $c(e)$ for all $e \in \mathcal{F}$, which is the number of all elements in $\mathcal{F}$ that cover e, i.e., $c(e) = |\{x \in \mathcal{F} : e \leq x\}|$. More importantly, as is shown in [33], $c(e)$ can be computed efficiently as follows:

Lemma 2 ([33])**.** *For $e \in \mathcal{F}$, the number of hypertuples covering e is $c(e) = \prod_{i=1}^{|R|} c(e_i)$. Note that R is the set of all attributes that define the data space.*

Lemma 3 ([33])**.** *For $e \in \mathcal{F}$ and $a_i \in R$,*

$$c(e_i) = \begin{cases} (\max(a_i) - \max(e_i) + 1) \times (\min(e_i) - \min(a_i) + 1), \\ \qquad\qquad\qquad\qquad \textit{if } a_i \textit{ is numerical} \\ 2^{|a_i| - |e_i|}, \qquad\qquad\quad \textit{if } a_i \textit{ is categorical} \end{cases}$$

where $\max(a_i)$ *is the maximal value in the domain of attribute a_i, and* $\max(e_i)$ *is the maximal value in e_i, which is a subset of the attribute's domain.*

We can then compute M as follows:

Lemma 4 ([33])**.** *Let V, $\mathcal{F}$ and D be given as above. Then $M = \sum_{x \in D} c(x)$.*

This lemma states that, although $\mathcal{F}$ may be very large (i.e., the power set of V), M can be computed by looking at all elements in D and aggregating their c values, which can be computed in a constant time regardless of the size of D or V. Consequently, the mass function, as defined in Def.2, can be derived systematically. It is shown in [33] that under such mass function, the belief functions can be computed in polynomial times. It is however open how to combine the different mass functions derived in this way. This question is answered in the next section.

4. Combining mass functions

Consider two mass functions $\bar{m}_1$ and $\bar{m}_2$ that are derived from two data sets D_1 and D_2, respectively. We can state this fact by using notation $\bar{m}_i = r(D_i)$ or alternatively, $D_i = r^{-1}(\bar{m}_i)$, where r represents the procedure implied by Eq.(1) to construct a mass function from a data set.

The data sets D_1 and D_2 are assumed to be homogeneous in the sense that they are samples from the same data space (i.e., they are characterised by the same attributes). The data sets can be understood as observations from different perspectives (i.e., different sensors of the same type).

Since the two mass functions are derived from data samples, we can combine them by first merging the data samples and then deriving a new mass function from the merged sample. Therefore, conceptually, the mass functions can be combined as follows:

$$\bar{m} \stackrel{\text{def}}{=} \bar{m}_1 \oplus \bar{m}_2 = r(D_1 \cup D_2) = r(r^{-1}(\bar{m}_1) \cup r^{-1}(\bar{m}_2))$$

Technically, the mass functions can be combined as follows. Let $D = D_1 \cup D_2$. For $e \in \mathcal{F}$,

$$\bar{m}(e) = \frac{|e^D|}{M} = \frac{|e^{D_1}| + |e^{D_2}|}{M_1 + M_2} = \frac{\bar{m}_1(e)M_1 + \bar{m}_2(e)M_2}{M_1 + M_2} \tag{2}$$

It is clear that this combining procedure is efficient since M_i can be computed efficiently according to Lemma 4, and $\bar{m}_i(e)$ is available for every $e \in \mathcal{F}$.

5. An illustrative example

To illustrate the combination method, we consider a data space defined by two attributes a_1 and a_2 that have the same domain $\{1, 2, 3, 4, 5\}$. Two data samples, D_1 and D_2, are shown in Table 5.1, each of which corresponds to a mass function as defined according to Eq.(1). The attributes can be categorical or numerical, and we will look at both cases separately.

5.1. Numerical case

We assume that both attributes are numerical. Then the two data samples, along with the union of them, can be displayed in 2D grids as in Figure 5.1. Each unit square represents a simple tuple, which is the coordinate of the square.

	a_1	a_2
t_{11}	3	2
t_{12}	2	3
t_{13}	4	4
t_{14}	5	4
t_{15}	4	5

D_1

	a_1	a_2
t_{21}	3	2
t_{22}	3	3
t_{23}	3	4
t_{24}	4	2
t_{25}	4	3

D_2

Table 5.1: A toy example.

D_1 $\qquad$ D_2 $\qquad$ $D_1 + D_2$

Figure 5.1: *Data samples displayed in the same data space*

- All hypertuples: It is not difficulty to note that there are 15 intervals in $\mathcal{F}_i$ for every attribute, which are $[1, 1]$, $[2, 2]$, $[3, 3]$, $[4, 4]$, $[5, 5]$, $[1, 2]$, $[2, 3]$, $[3, 4]$, $[4, 5]$, $[1, 3]$, $[2, 4]$, $[3, 5]$, $[1, 4]$, $[2, 5]$ and $[1, 5]$. As a result, there are $15 \times 15 = 225$ hypertuples in $\mathcal{F}$. An example is $\langle \{2, 3\}, \{4, 5\} \rangle$, which is generated by two intervals, $\{2, 3\}$ and $\{4, 5\}$, from the two attributes.

- All hypertuples that cover e: Consider $e = \langle \{3, 4\}, \{2, 3\} \rangle$. According to Lemma 2, there are $(5-4+1) \times (3-1+1) = 6$ intervals in attribute a_1 that can be used to construct a hypertuple that covers e, namely $\{[3, 4], [3, 5], [2, 4], [2, 5], [1, 4], [1, 5]\}$. Similarly, there are 6 intervals in attribute a_2, which are $\{[2, 3], [2, 4], [2, 5], [1, 3], [1, 4], [1, 5]\}$.

  As a result, there are $6 \times 6 = 36$ hypertuples that cover e. One example is $\langle [3, 5], [2, 5] \rangle$, which is generated by intervals $[3, 5]$ and $[2, 5]$ in a_1 and a_2.

- Normalization factors M_1 and M_2: According to Lemma 4 we have

$$M_1 = \sum_{x \in D_1} c(x) = c(t_{11}) + c(t_{12}) + c(t_{13}) + c(t_{14}) + c(t_{15})$$
$$= 9 * 8 + 8 * 9 + 8 * 8 + 5 * 8 + 8 * 5 = 288$$

$$M_2 = \sum_{x \in D_2} c(x) = c(t_{21}) + c(t_{22}) + c(t_{23}) + c(t_{24}) + c(t_{25})$$
$$= 9 * 8 + 9 * 9 + 9 * 8 + 8 * 8 + 8 * 9 = 361$$

- Mass functions $\bar{m}_1$ and $\bar{m}_2$: Since there are 225 hypertuples in the data space and both functions are defined over the same data space, the mass functions are defined for all 225 hypertuples. Consider $e = \langle \{3, 4\}, \{2, 3\} \rangle$ again.

$$\bar{m}_1(e) = \frac{e^{D_1}}{M_1} = \frac{1}{288}$$

$$\bar{m}_2(e) = \frac{e^{D_2}}{M_2} = \frac{4}{361}$$

- Combined mass function $\bar{m}$: Let $\bar{m}$ be the combined mass function from $\bar{m}_1$ and $\bar{m}_2$. According to Eq.(2) we have

$$\bar{m}(e) = \frac{\bar{m}_1(e)M_1 + \bar{m}_2(e)M_2}{M_1 + M_2} = \frac{5}{649}$$

- We can easily verify that if we merge D_1 and D_2 and follow the same procedure, we will get the same mass function value for e.

5.2. Categorical case

Now we assume both attributes are categorical and go through the procedure on the basis of this assumption.

- All hypertuples: There are $2^5 = 32$ subsets in each attribute that can be used to generate hypertuples. As a result, there are $32 \times 32 = 1024$ hypertuples.

- All hypertuples that cover e: Consider $e = \langle\{3,4\},\{2,3\}\rangle$. There are $2^3 = 8$ subsets in attribute a_1 that can be used to generate hypertuples that cover e. There are the same number of subsets in attribute a_2. Therefore, there are altogether $8 \times 8 = 64$ hypertuples that cover e.

- Normalization factors M_1 and M_2: According to Lemma 4, we have

$$\begin{aligned} M_1 &= \sum_{x \in D_1} c(x) = c(t_{11}) + c(t_{12}) + c(t_{13}) + c(t_{14}) + c(t_{15}) \\ &= 16 * 16 + 16 * 16 + 16 * 16 + 16 * 16 + 16 * 16 = 1280 \end{aligned}$$

$$\begin{aligned} M_2 &= \sum_{x \in D_2} c(x) = c(t_{21}) + c(t_{22}) + c(t_{23}) + c(t_{24}) + c(t_{25}) \\ &= 16 * 16 + 16 * 16 + 16 * 16 + 16 * 16 + 16 * 16 = 1280 \end{aligned}$$

- Mass functions $\bar{m}_1$ and $\bar{m}_2$: The mass functions are defined for all 1024 hypertuples. Consider $e = \langle\{3,4\},\{2,3\}\rangle$ again.

$$\bar{m}_1(e) = \frac{e^{D_1}}{M_1} = \frac{1}{1280}$$

$$\bar{m}_2(e) = \frac{e^{D_2}}{M_2} = \frac{4}{1280}$$

- Combined mass function $\bar{m}$: Let $\bar{m}$ be the combined mass function from $\bar{m}_1$ and $\bar{m}_2$. According to Eq.(2), we have

$$\bar{m}(e) = \frac{\bar{m}_1(e)M_1 + \bar{m}_2(e)M_2}{M_1 + M_2} = \frac{5}{2560}$$

- We can easily verify that if we merge D_1 and D_2 and follow the same procedure, we will get the same mass function value for e.

6. Generalisation of the new combination rule

If the mass functions in the sense of Def.2 are given by users, $\bar{m}$ may be undefined for some elements of $\mathcal{F}$, thus leading to the problem of incompleteness. Consequently, the combination function (2) may be limited.

Therefore, we need to generalise this function by estimating $\bar{m}(e)$ for every $e \in \mathcal{F}$. Then we will be able to use function (2) to combine different mass functions in cases of incompleteness (i.e., when the mass function is undefined for some elements).

Definition 3. *Let $\bar{m}$ be a mass function, possibly given by a user. A generalised mass function is $\bar{m}' : \mathcal{F} \to [0, 1]$ such that, for $e \in \mathcal{F}$*

$$\bar{m}'(e) = \sum_{x \in \mathcal{F}} \bar{m}(x) \frac{f(e \cap x)}{K}$$

where $K = \sum_{e \in \mathcal{F}} \sum_{x \in \mathcal{F}} \bar{m}(x) f(e \cap x)$, and $f(e)$ is a measure function. One possible form of this function is $f(e) = |e|$.

Lemma 5. *The generalised mass function $\bar{m}'$ defined above has the following properties:*

- $\bar{m}'(\emptyset) = 0$.
- $\sum_{e \in \mathcal{F}} \bar{m}'(e) = 1$.
- $\bar{m}'(x) \leq \bar{m}'(y)$ *if* $x \subseteq y$.
- $\bar{m}'(x \cup y) = \bar{m}'(x) + \bar{m}'(y) - \bar{m}'(x \cap y)$.

The proofs of these properties are straightforward. Based on the concept of generalised mass function we can combine any mass functions, possibly incomplete, according to the following combining rule:

Rule 1. *Let $\bar{m}_1$ and $\bar{m}_2$ be two mass functions. The two mass functions are combined to generate a new mass function $\bar{m} : \mathcal{F} \to [0, 1]$ such that, for $e \in \mathcal{F}$,*

$$\begin{aligned}
\bar{m}(e) &= \frac{\bar{m}'_1(e) M_1 + \bar{m}'_2(e) M_2}{M_1 + M_2} \\
&= \frac{\sum_{x \in \mathcal{F}} \bar{m}_1(x) \frac{f(e \cap x)}{K_1} M_1 + \sum_{x \in \mathcal{F}} \bar{m}_2(x) \frac{f(e \cap x)}{K_2} M_2}{M_1 + M_2} \\
&= \sum_{x \in \mathcal{F}} \bar{m}_1(x) \times f(e \cap x) \times \alpha_1 + \sum_{x \in \mathcal{F}} \bar{m}_2(x) \times f(e \cap x) \times \alpha_2
\end{aligned}$$

where $\alpha_i = \frac{M_i}{(M_1 + M_2) \times K_i}$ for $i = 1, 2$. To some extent, α_i ($i = 1, 2$) can be understood as the weighting coefficients for the two mass functions.

7. Some reflection from a data fusion perspective

Data fusion is the process of combining different sources of information to improve the performance of a system. An obvious illustration of fusion is the use of various sensors to detect a target. The different inputs may originate from multiple sensors distributed over space, from a single sensor at different moments, or even from experts who give their opinions on data. Fusion is useful for many data analysis tasks such as detection, recognition, identification, tracking, and decision making. These tasks may be encountered in many application domains such as defense, robotics, and medicine.

The benefits of fusion include the following:

- Improved confidence in decisions due to the use of complementary information
- Improved performance of countermeasures (it is very hard to camouflage an object in all possible wave-bands)
- Improved performance in adverse environmental conditions.

Fusion processes are often categorized as low-, intermediate- or high-level fusion depending on the processing stage at which fusion takes place [18]. Low-level fusion, also called data fusion, combines several sources of raw data to produce new raw data that is expected to be more informative and synthetic than the inputs. In intermediate-level fusion or feature-level fusion, various features such as edges, corners, lines, and texture parameters are combined into a feature map that may then be used for further processing. High-level, also called decision fusion, combines the decisions of experts. In practice, the applied fusion procedure is often a combination of the previously mentioned three levels.

The Dempster-Shafer theory, in particular the rule of combination, is often used for decision fusion [28, 4, 23, 22, 6, 7, 8, 24, 5, 2]. Other methods for decision fusion include voting methods, statistical methods, and fuzzy logic-based methods. Questions are still being asked why the decisions (from experts) are combined as such.

The combination method presented in this paper is to some extent an attempt to answer such questions. We assume that the mass functions are systematically derived from data. Such a mass function can be regarded as a generalised probability distribution, which can then be used for detection, recognition, identification, tracking, and decision making. Such a mass function can also be regarded as the decisions of 'rational' experts who base their decisions on facts or data.

If we adopt the systematic derivation view of mass functions, combination should naturally be conducted as follows: merge multiple data sets into a single one, and then derive a new mass function from the merged data set. Fortunately, as shown earlier, we do not need to keep all data sets from different sources. We only need to know the mass functions, as is the case in the Dempster-Shafer theory of evidence. Therefore, we believe that our method of combination is sound in principle.

8. Conclusion

In this paper, we revisit the Dempster-Shafer theory of evidence. We first review a recent development in the systematic derivation of mass functions from multivariate data. We then proceed to consider the problem of how to combine the mass functions thus derived.

For a multivariate data space, the frame of discernment can be taken to be the set of all hypertuples. For a given sample of the data space, a mass function can be derived such that for each hypertuple its mass is proportional to the number of data items that are covered by the hypertuple. For two mass functions derived in this way from different data samples of the same data space, we combine the functions conceptually by first merging the data samples and then deriving a new mass function from the merged data sample. We presented a formula by which the combined mass function can be obtained efficiently without having to merge the data samples in the first place.

In future work, we will apply the method to sensor data fusion in the context of smart homes.

References

[1] A. Aregui and T. Denoeux. Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities. *International Journal of Approximate Reasoning*, 49(3):575–594, 2008.

[2] M. Arif, T. Brouard, and N. Vincent. A fusion methodology based on dempster-shafer evidence theory for two biometric applications. In *Proceedings of ICPR06*, pages IV: 590–593, 2006.

[3] R.B. Ash and Catherine Doléans-Dade. *Probability and Measure Theory*. Academic Press, 2000.

[4] O. Basir and X. Yuan. Engine fault diagnosis based on multi-sensor information fusion using dempster-shafer evidence theory. *Information Fusion*, 8(4):379–386, 2007.

[5] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski. Multisensor image segmentation using dempster-shafer fusion in markov fields context. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8):1789–1798, August 2001.

[6] Y. Bi, J. Guan, and D. Bell. The combination of multiple classifiers using an evidential reasoning approach. *Artificial Intelligence*, 172(15):1731–1751, 2008.

[7] I. Bloch. Defining belief functions using mathematical morphology - application to image fusion under imprecision. *International Journal of Approximate Reasoning*, 48(2):437–465, 2008.

[8] E. Bosse and J. Roy. Fusion of identity declarations from dissimilar sources using the dempster-shafer theory. *Optical Enginerring*, 36(3):648–657, March 1997.

[9] A.-O. Boudraa, A. Bentabet, F. Salzenstein, and L. Guillon. Dempster-Shafer's Basic Probability Assignment Based on Fuzzy Membership Functions. *Electronic Letters on Computer Vision and Image Analysis*, 4(1):1–9, 2004.

[10] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Statistics*, 28:325–339, 1967.

[11] T. Denoeux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics:*, 25:804–813, 1995.

[12] T. Denoeux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42(3):228–252, 2006.

[13] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2-3):234–264, 2008.

[14] T. Denoeux. Extending stochastic ordering to belief functions on the real line. *Information Sciences*, 179(9):1362–1376, 2009.

[15] T. Denoeux and P. Smets. Classification using belief functions: Relationship between case-based and model-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(6):1395–1405, 2006.

[16] D. Dubois and H. Prade. On the combination of evidence in various mathematical frameworks. *Reliability Data Collection and Analysis*, pages 213–241, 1992.

[17] M. Grabisch. Belief functions on lattices. *International Journal of Intelligent Systems*, 24(1):76–95, 2009.

[18] D.L. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23, 1997.

[19] V.-N. Huynh, Y. Nakamori, T.-B. Ho, and T. Murai. Multiple-attribute decision making under uncertainty: The evidential reasoning approach revisited. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 36(4):804–822, 2006.

[20] T. Inagaki. Interdependence between safety-control policy and multiple-sensor schemes via dempster-shafer theory. *IEEE Transactions on Reliability*, 40(2):182–188, 1991.

[21] L. Liu. Special Issue on the Dempster-Shafer Theory of Evidence: An Introduction. *International Journal of Intelligent Systems*, 18:1–4, 2003.

[22] S. Maskell. A bayesian approach to fusing uncertain, imprecise and conflicting information. *Information Fusion*, 9(2):259–277, 2008.

[23] D. Mercier, B. Quost, and T. Denoeux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258, 2008.

[24] R.R. Murphy. Dempster-shafer theory for sensor fusion in autonomous mobile robots. *IEEE Transactions on Robotics and Automation*, 14(2):197–206, April 1998.

[25] K. Premaratne, D.A. Dewasurendra, and P.H. Bauer. Evidence combination in an environment with heterogeneous sources. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 37(3):298–309, 2007.

[26] P. Sevastianov. Numerical methods for interval and fuzzy number comparison based on the probabilistic approach and dempster-shafer theory. *Information Sciences*, 177(21):4645–4661, 2007.

[27] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, New Jersey, 1976.

[28] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.

[29] S. Stevens. *Mathematics, Measurement, and Psychophysics (Handbook of Experimental Psychology)*. Wiley, 1951.

[30] E. Straszecka. Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences*, 176(20):3026–3059, 2006.

[31] H. Wang and J. Liu. Combining evidence in multivariate data spaces. In *Proceedings of the 8th International FLINS Conference on Computational Intelligence in Decision and Control*, pages 11–16, Madrid, Spain, 2008.

[32] H. Wang, I. Düntsch, G. Gediga, and A. Skowron. Hyperrelations in version space. *International Journal of Approximate Reasoning*, 36(3):223–241, 2004.

[33] H. Wang and S. McClean. Deriving evidence theoretical functions in multivariate data space. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(2):455–465, 2008.

[34] Y.-M. Wang, J.-B. Yang, D.-L. Xu, and K.-S. Chin. On the combination and normalization of interval-valued belief structures. *Information Sciences*, 177(5):1230–1247, 2007.

[35] W.-Z. Wu. Attribute reduction based on evidence theory in incomplete decision systems. *Information Sciences*, 178(5):1355–1371, 2008.

[36] R. Yager. On the dempster-shafer framework and new combination rules. *Information Sciences*, 41:93–137, 1987.

[37] R.R. Yager. Comparing approximate reasoning and probabilistic reasoning using the dempster-shafer framework. *International Journal of Approximate Reasoning*, 50(5):812–821, 2009.

[38] L. Zhang. *Representation, independence, and combination of evidence in the Dempster-Shafer theory*, pages 51–69. John Wiley & Sons, Inc., New York, 1994.

[39] Wikimedia Foundation. Wikipedia. http://www.wikipedia.org.