Re-investigating Dempster's idea on evidence combination.

Liu, W., & Hong, J. (2000). Re-investigating Dempster's idea on evidence combination. Knowledge and Information Systems, 2(2), 223-241. DOI: 10.1007/s101150050012

Published in:
Knowledge and Information Systems

Document Version:
Peer reviewed version

Re-investigating Dempster's Idea on Evidence Combination

Weiru Liu and Jun Hong

School of Information and Software Engineering
University of Ulster at Jordanstown
Newtownabbey, Co. Antrim BT37 0QB, UK
{w.liu, j.hong}@ulst.ac.uk

Abstract. In this paper, we investigate the problem encountered by Dempster's combination rule in view of Dempster's original combination framework. We first show that the root of Dempster's combination rule (defined and named by Shafer) is Dempster's original idea on evidence combination. We then argue that Dempster's original idea on evidence combination is, in fact, richer than what has been formulated in the rule. We conclude that, by strictly following what Dempster has suggested, there should be no counterintuitive results when combining evidence.

Keywords: Dempster's combination rule, evidential reasoning, Dempster's combination framework.

1 Introduction

Methods for dealing with uncertainty in many areas of Artificial Intelligence have received considerable attention for more than a decade. Several numerical and symbolic methods have been proposed for handling uncertain information (see [3], [10], [11] for details). The traditional numerical method is *Bayesian probability*. The well known expert system *Prospector* ([5]) is a typical example of using this method. Bayesian updating rule provides this method with the ability to revise a result in the light of new evidence. Generally speaking, Bayesian probability (subjective and conditional) is powerful for a certain class of problems, but not suitable for all situations. Alternatives have been investigated by many researchers. The *certainty factor model* in expert system MYCIN ([20]) is one of the alternatives.

The Dempster-Shafer theory of evidence (DS theory) (sometimes called evidential reasoning or belief function theory) is a mechanism formalised by Shafer ([14]) for representing and reasoning with uncertain, imprecise and incomplete information. It is based on Dempster's original work ([4]) on the modelling of uncertainty in terms of upper and lower probabilities that are induced by a multivalued mapping rather than as a single probability value. DS theory reduces to standard Bayesian reasoning when an agent's knowledge is accurate but it is more flexible in representing and dealing with ignorance and uncertainty [1]. DS theory has been popular since early 1980s when AI researchers were searching for different mechanisms to cope with those situations where Bayesian probability is powerless. Its relationships with related theories have been intensively discussed [23].

There are two main reasons why DS theory has attracted so much attention. It has the ability to model information flexibly, without requiring a probability (or a prior) to be assigned to each element in a set, and it provides a convenient and simple mechanism (Dempster's combination rule) for combining two or more pieces of evidence under certain conditions. The former allows an agent to describe ignorance because of lacking of information, and the latter allows an agent to narrow down the possible solution space as more evidence is accumulated.

Even though DS theory has been widely used, it has however been found that Dempster's combination rule gives counterintuitive results in some cases. The condition under which the rule is used is crucial to the successful application of the theory but the condition was not fully defined when Shafer gave the rule in the first instance [14]. Various discussions and criticisms of the rule have appeared in the literature. A mathematical description of the condition of applying Dempster's combination rule is formalised in [22].

In this paper, we first introduce the background of DS theory. We then study the probabilistic basis of basic functions defined by Shafer to see how we can derive a mass function from a probability distribution through a multivalued mapping [4]. We closely examine Dempster's original idea on evidence combination. We show that, in fact, Dempster suggested two alternative approaches to combine evidence. One approach requires the construction of a combined source covering several original sources, and propagates the combined probability distribution on the combined source to the target space. Another approach encourages the individual propagation from each original source to the target space, and then combines the propagation results on the target space. Dempster's combination rule in DS theory is a simplified form of the second approach.

We argue that the simplified form of combination (*i.e.* Dempster's combination rule) does not carry enough information to allow an agent to judge whether two pieces of evidence can be combined. In fact under Dempster's original idea, some counterintuitive examples of using Dempster's combination rule are not counterintuitive at all, because these examples cannot be and should have not been dealt with in DS theory in the way they were.

2 Basic Concepts in Dempster-Shafer Theory

In DS theory, a piece of information is usually described as a mass function on a frame of discernment. We first give some definitions of the theory [14].

Definition 1 (Frame of Discernment). *A set is called a frame of discernment (or simply a frame) if it contains mutually exclusive and exhaustive possible answers to a question. It is usually denoted as Θ. The set is required that at any time, one and only one element in the set is true.*

For instance, if we assume that Emma lives in one of the cities, $city_1, ..., city_6$, then $\Theta = \{city_1, city_2, city_3, city_4, city_5, city_6\}$ is a frame of discernment for the question 'In which city does Emma live?'.

Definition 2 (Mass Function). *A function $m : 2^\Theta \to [0,1]$ is called a mass function on frame Θ if it satisfies the following two conditions:*

1. *$m(\emptyset) = 0$, and*
2. *$\Sigma_A m(A) = 1$,*

where $\emptyset$ is an empty set and A is a subset of Θ.

A mass function is also called *a basic probability assignment*, denoted as *bpa*.

For instance, if we know that Emma lives in the area covering the six cities, but we have no knowledge about in which city she lives, then we can only give a mass function $m(\Theta) = 1$. Alternatively, if we know that Emma lived in $city_3$ two years ago and she intended to move to other cities and tried to find a job somewhere within these six cities, but we have no definite information about where she lives now, then a mass function could be defined as $m(\{city_3\}) = p, m(\Theta) = 1 - p$, where p stands for the degree of our belief that she still lives in $city_3$.

Definition 3 (Belief Function). *A function $bel : 2^\Theta \to [0,1]$ is called a belief function if bel satisfies:*

1. *$bel(\Theta) = 1$;*
2. *$bel(\cup_1^n A_i) \geq \Sigma_i bel(A_i) - \Sigma_{i>j} bel(A_i \cap A_j) + ... + (-1)^{-n} bel(\cap_i A_i)$.*

It is easy to see that $bel(\emptyset) = 0$ for any belief function. A belief function is also called *a support function*. The difference between $m(A)$ and $bel(A)$ is that $m(A)$ is our belief committed to the subset A excluding any of its subsets while $bel(A)$ is our degree of belief in A as well as all of its subsets.

In general, if m is a mass function on frame Θ then bel defined in (1) is a belief function on Θ

$$bel(B) = \Sigma_{A \subseteq B} m(A). \tag{1}$$

Recovering a mass function from a belief function is as follows [18]:

$$m(A) = \Sigma_{B \subseteq A} (-1)^{|B|} bel(B).$$

For any finite frame, it is always possible to get the corresponding mass function from a belief function and the mass function is unique.

A subset A with $m(A) > 0$ is called a *focal element of this belief function*. If all focal elements of a belief function are the singletons of Θ then the corresponding mass function is exactly a probability distribution on Θ. So mass functions are generalised probability distributions in this sense.

If there is only one focal element for a belief function and the focal element is the whole frame Θ, this belief function is called a *vacuous belief function*. It represents total ignorance (because of lack of knowledge).

Definition 4 (Plausibility Function). *A function pls defined below is called a plausibility function*

$$pls(A) = 1 - bel(\neg A).$$

$pls(A)$ represents the degree to which the evidence fails to refute A. From a mass function, we can get its plausibility function as [18]:

$$pls(B) = \Sigma_{A \cap B \neq \emptyset} m(A). \tag{2}$$

When more than one mass function is given on the same frame of discernment, the combined impact of these pieces of evidence is obtained using a mathematical formula called *Dempster's combination rule*. If m_1 and m_2 are two mass functions on frame Θ, then $m = m_1 \oplus m_2$ is the mass function after combining m_1 and m_2.

$$m(C) = \frac{\Sigma_{A \cap B = C} m_1(A) m_2(B)}{1 - \Sigma_{A \cap B = \emptyset} m_1(A) m_2(B)}.$$

$\oplus$ means that Dempster's combination rule is applied on two (or more) mass functions. The condition of using the rule is stated as "two or more pieces of evidence are based on distinct bodies of evidence" [14]. This description is a bit confusing and causes a lot of misapplications and counterintuitive results [22].

3 Probability Background of Mass Functions

Even though Shafer has not agreed on the idea that belief function theory is generalised probability theory and has regarded belief function theory as a new way of representing evidence and knowledge, some people have argued that the theory has strong links with probability theory [6], [7]. We argue that in Dempster's paper [4], Dempster implicitly gave the prototype of mass functions. Shafer's contribution has been to explicitly define the mass function and to use it to represent evidence directly.

3.1 Dempster's probability prototype of mass functions

Definition 5 (Dempster's Probability Space). *A structure (X, τ, μ) is called a Dempster's probability space where*

1. *X is a sample space containing all the possible worlds;*
2. *τ is a class of subsets of X;*
3. *μ is a probability measure which gives $\mu : \tau \to [0, 1]$, and $\mu(X) = 1$.*

Definition 6 (Multivalued Mapping). *Function $\Gamma : X \to 2^S$ is a multivalued mapping from space X to space S if Γ assigns a subset $\Gamma(x) \subseteq S$ to every $x \in X$.*

From a multivalued mapping Γ, a probability measure μ on spsce X can be propagated to space S in such a way that for any subset T of S, the lower and upper bounds of probabilities of T are defined as

$$P_*(T) = \mu(T_*)/\mu(S^*), \tag{3}$$

$$P^*(T) = \mu(T^*)/\mu(S^*), \tag{4}$$

where

$$T_* = \{x \in X, \Gamma(x) \neq \emptyset, \Gamma(x) \subseteq T\},$$

$$T^* = \{x \in X, \Gamma(x) \cap T \neq \emptyset\}.$$

Eqs (3) and (4) are defined only when $\mu(S^*) \neq 0$. The denominator $\mu(S^*)$ is a renormalizing factor necessitated by the fact that the subset $\{x \mid \Gamma(x) = \emptyset\}$, which does not map into a meaningful subset of S, should be removed from X and the measure of the remaining set S^* renormalized to unity.

A multivalued mapping Γ from space X to space S says that if the possible answer to a question described in the first space X is x, then the possible answer to a question described in the second space S is in $\Gamma(x)$.

For the case that $S = \{s_1, s_2, ..., s_n\}$ is finite, the propagation procedure can be done as follows. Suppose that $S_{\gamma_1\gamma_2...\gamma_n}$ denotes the subset of S which contains s_i if $\gamma_i = 1$ and excludes s_i if $\gamma_i = 0$ for $i = 1, 2, ..., n$. If for each $S_{\gamma_1\gamma_2...\gamma_n}$, we define $X_{\gamma_1\gamma_2...\gamma_n}$ as

$$X_{\gamma_1\gamma_2...\gamma_n} = \{x \in X, \Gamma(x) = S_{\gamma_1\gamma_2...\gamma_n}\}, \tag{5}$$

then all the non-empty subsets of X defined in Eq (5) form a partition[1] of X and

$$X = \cup_{\gamma_1\gamma_2...\gamma_n} X_{\gamma_1\gamma_2...\gamma_n}. \tag{6}$$

The idea of constructing $X_{\gamma_1\gamma_2...\gamma_n}$ is that each $X_{\gamma_1\gamma_2...\gamma_n}$ contains those elements in X which have the same mapping environment in S.

In order to calculate $P_*(T)$ and $P^*(T)$, Dempster assumed that each non-empty $X_{\gamma_1\gamma_2...\gamma_n}$ is in τ, then for any $T \subset S$, $P_*(T)$ and $P^*(T)$ are uniquely determined by the 2^n quantities $p_{\gamma_1\gamma_2...\gamma_n}$ as,

$$p_{\gamma_1\gamma_2...\gamma_n} = \mu(X_{\gamma_1\gamma_2...\gamma_n}). \tag{7}$$

We use an example to demonstrate the idea.

Example 1 (From [4])

Assume that $S = \{s_1, s_2, s_3\}$. Using $p_{\gamma_1\gamma_2...\gamma_n}$, the lower and upper bounds of probabilities of all subsets of S are given in Table 1.

[1] A list of subsets $X_1, X_2, ..., X_n$ of X is called a partition of X if $X_i \cap X_j = \emptyset$ and $\cup_{i=1}^{i=n} X_i = X$.

Table 1. Upper and lower probabilities of all subsets of S.

T	$P^*(T)$	$P_*(T)$
$\emptyset$	0	0
$\{s_1\}$	$(p_{100}+p_{110}+p_{101}+p_{111})/k$	p_{100}/k
$\{s_2\}$	$(p_{010}+p_{110}+p_{011}+p_{111})/k$	p_{010}/k
$\{s_3\}$	$(p_{001}+p_{101}+p_{011}+p_{111})/k$	p_{001}/k
$\{s_1,s_2\}$	$(p_{100}+p_{010}+p_{110}+p_{101}+p_{011}+p_{111})/k$	$(p_{100}+p_{010}+p_{110})/k$
$\{s_1,s_3\}$	$(p_{100}+p_{001}+p_{110}+p_{101}+p_{011}+p_{111})/k$	$(p_{100}+p_{001}+p_{101})/k$
$\{s_2,s_3\}$	$(p_{010}+p_{001}+p_{110}+p_{101}+p_{011}+p_{111})/k$	$(p_{010}+p_{001}+p_{011})/k$
S	1	1

Here k is defined as $\mu(S^*)=1-p_{000}$. Given a subset T of S, the corresponding lower and upper subsets in X are known. For instance, if $T=S_{110}=\{s_1,s_2\}$, then $T_*=X_{100}\cup X_{010}\cup X_{110}$ and $T^*=X_{100}\cup X_{010}\cup X_{110}\cup X_{101}\cup X_{011}\cup X_{111}$.

If we define a function m on S as $m(S_{\gamma_1\gamma_2\ldots\gamma_n})=m_{\gamma_1\gamma_2\ldots\gamma_n}=p_{\gamma_1\gamma_2\ldots\gamma_n}/(1-p_{00\ldots0})$, as shown in Table 2, then function m is exactly a mass function when S is a frame. Certainly some of $m_{\gamma_1\gamma_2\ldots\gamma_n}$ may be 0. P_* and P^* define a belief function and a plausibility function on S respectively. We assume that this is the model for defining mass functions in Shafer's style. S being a frame of discernment is a special case of S being a space in Dempster's paper.

Table 2. Upper and lower probabilities of all subsets of S using function m.

T	$P^*(T)=pls$	$P_*(T)=bel$	$m(T)$
$\emptyset$	0	0	0
$\{s_1\}$	$m_{100}+m_{110}+m_{101}+m_{111}$	m_{100}	m_{100}
$\{s_2\}$	$m_{010}+m_{110}+m_{011}+m_{111}$	m_{010}	m_{010}
$\{s_3\}$	$m_{001}+m_{101}+m_{011}+m_{111}$	m_{001}	m_{001}
$\{s_1,s_2\}$	$m_{100}+m_{010}+m_{110}+m_{101}+m_{011}+m_{111}$	$m_{100}+m_{010}+m_{110}$	m_{110}
$\{s_1,s_3\}$	$m_{100}+m_{001}+m_{110}+m_{101}+m_{011}+m_{111}$	$m_{100}+m_{001}+m_{101}$	m_{101}
$\{s_2,s_3\}$	$m_{010}+m_{001}+m_{110}+m_{101}+m_{011}+m_{111}$	$m_{010}+m_{001}+m_{011}$	m_{011}
S	1	1	m_{111}

The vital requirement of calculating probability bounds in Dempster's prototype is that every non-empty subset $X_{\gamma_1\gamma_2\ldots\gamma_n}$ should be in τ. If τ, a collection of subsets of X, does not suit this requirement, then the rest of the calculation in Dempster's paper could not be carried out.

3.2 Deriving mass functions from probability spaces

Definition 7 (Probability Space). *A probability space (X,χ,μ) has:*

X*: a sample space containing all the possible worlds;*
χ*: a σ-algebra containing some subsets of X, which is defined as containing X and closed under complementation and countable union.*

μ: *a probability measure* $\mu : \chi \to [0, 1]$ *with the following features:*

1. $\mu(X_i) \geq 0$ *for all* $X_i \in \chi$;
2. $\mu(X) = 1$;
3. $\mu(\cup_{j=1}^{\infty} X_j) = \Sigma_{j=1}^{\infty} \mu(X_j)$, *if the* X_j*'s are pairwise disjoint members of* χ.

A subset χ' of χ is called a *basis* of χ if it contains non-empty and disjoint elements, and if χ consists precisely of countable unions of members of χ'. For any finite χ there is a unique basis χ' of χ and it follows that

$$\Sigma_{X_i \in \chi'} \mu(X_i) = 1.$$

For any subset X_i of X, if X_i is not in χ, it is only possible to get two probability bounds of X_i, usually called the *inner measure*, denoted as μ_*, and the *outer measure*, denoted as μ^*, with

$$\mu_*(X_i) = sup\{\mu(X_j) \mid X_j \subseteq X_i, X_j \in \chi\}, \tag{8}$$

$$\mu^*(X_i) = inf\{\mu(X_j) \mid X_j \supseteq X_i, X_j \in \chi\}. \tag{9}$$

It has been proved in [6], [7] that μ_* is a belief function on X when X is a frame.

Given a probability space (X, χ, μ), assume there is a multivalued mapping function Γ from X to frame S. For a subset T of S, we define

$$bel(T) = \begin{cases} \mu(T_*)/k & \text{when } T_* = \{x \mid \Gamma(x) \neq \emptyset, \Gamma(x) \subseteq T\} \in \chi, \\ \mu_*(T_*)/k & \text{Otherwise.} \end{cases} \tag{10}$$

bel is also a belief function on S. Here $k = \mu_*(\{x \in X, \Gamma(x) \cap S \neq \emptyset\})$ which has the same meaning as it has in Table 1.

In a Dempster's probability space, those non-empty subsets $X_{\gamma_1\gamma_2\ldots\gamma_n}$ of X form a partition of X, and equation $\Sigma_{\gamma_1\gamma_2\ldots\gamma_n} \mu(X_{\gamma_1\gamma_2\ldots\gamma_n}) = 1$ holds. This suggests that these non-empty subsets possess the properties of a basis of a $\sigma-$algebra. If we use χ' to denote this partition, we can derive a $\sigma-$algebra of X using this basis. So a Dempster's probability space can always be translated into a normal probability space. Therefore a probability space is more general than a Dempster's probability space. Given a probability space (X, χ, μ) and a multivalued mapping Γ from X to S, Eq (10) is suitable for deriving a belief function on S, while given a Dempster's probability space (X, τ, μ) and a multivalued mapping Γ from X to S, Eq (3) is adequate to define a belief function on S. In either case, a belief function can always be derived, so belief function theory is closely related to probability theory and comes out of probability theory.

In the following, we use probability spaces to stand for both normal and Dempster's probability spaces.

4 Problems with Dempster's Combination Rule

A multivalued mapping from space X_1 to space S generates a mass function on S given a probability distribution on X_1 or a collection of subsets of X_1 (either τ or χ) as we have seen in the previous section.

If there is another multivalued mapping from space X_2 to S which also defines a mass function on S, then there are two mass functions available on S. In DS theory, when there are two or more mass functions on the same frame of discernment, it is desirable to combine them using Dempster's combination rule, if it is believed that they are derived from 'independent sources'. The debate about the precise meaning of 'independent sources' has appeared in many research papers, including some by Shafer himself, since the interpretation of this phrase determines whether an agent can apply Dempster's rule correctly. However, Dempster's combination rule only involves mass functions, which cannot tell anything about the relationships between their sources. Therefore, it is very difficult to describe the condition of using the rule if an agent's attention is focused on the rule itself.

As Dempster's combination rule is motivated by the idea in Dempster's original paper, we believe that re-examining the idea of combination suggested by Dempster would be very beneficial to the clarification of the condition of using the rule.

4.1 Dempster's combination framework

In [4], Dempster first discussed the procedure of generating the lower and upper bounds of probabilities on space S from a probability space (X, χ, μ) through a multivalued mapping Γ, as discussed in Section 3. Dempster then considered the situation where n probability spaces are available and each of which has a multivalued mapping relation with the same space S, given that these n probability spaces are independent.

In summary, Dempster's idea on the combination of independent sources of information can be stated as follows. Suppose there are n pieces of evidence which are given in the form of n probability spaces (X_i, χ_i, μ_i), each of which has a mapping relation with the same space S through a multivalued mapping Γ_i. These n sources are said to be independent and explained by Dempster as *"opinions of different people based on overlapping experiences could not be regarded as independent sources. Different measurements by different observations on different equipment would often be regarded as independent ... the sources are statistically independent"* [4].

Under his assumption of independence, Dempster [4] suggested that these sources can be combined using Eq (11) to derive a combined source (X, χ, μ) and a unified Γ as

$$X = X_1 \otimes X_2 \otimes ... \otimes X_n,$$

$$\chi = \chi_1 \otimes \chi_2 \otimes ... \otimes \chi_n, \tag{11}$$

$$\mu = \mu_1 \otimes \mu_2 \otimes ... \otimes \mu_n,$$

$$\Gamma(x) = \Gamma_1(x) \cap \Gamma_2(x) \cap ... \cap \Gamma_n(x).$$

The fourth formula can also be stated and explained as

$$\Gamma(x) = \Gamma_1'(x) \cap \Gamma_2'(x) \cap ... \cap \Gamma_n'(x)$$

where $\Gamma_i'(x) = \Gamma_i(x_i)$ when $x \in X_1 \otimes ... \otimes X_{i-1} \otimes \{x_i\} \otimes ... \otimes X_n$.

Here $\otimes$ is the set product operator as usual. $\mu = \mu_1 \otimes \mu_2$ is defined as $\mu(\{(x_{1i}, x_{2j})\}) = \mu_1(\{x_{1i}\}) \times \mu_2(\{x_{2j}\})$ for every $(x_{1i}, x_{2j}) \in X$ where $x_{1i} \in X_1$ and $x_{2j} \in X_2$. The combined source (set product measure space) summarises the message carried by all separate sources. The combined multivalued mapping Γ from the combined source to space S suggests $x = (x_1, x_2, ..., x_n) \in X$ is consistent with $s_i \in S$, if and only if s_i belongs to all $\Gamma_i(x_i)$ simultaneously.

Through the combined source (X, χ, μ) and the unified mapping Γ, for any subset T of S, a pair of lower and upper bounds of probabilities can be calculated using Eqs (3) and (4).

Apart from using the set of formulas in Eq (11), Dempster also recommended an alternative calculation method for calculating the final bounds of probabilities on S using function q

$$q(T) = q_1(T) \times q_2(T) \times, ..., \times q_n(T), \tag{12}$$

where

$$q_j(T) = \mu_j(\tilde{T}_j),$$

and

$$\tilde{T}_j = \{x_j \in X_j, \Gamma_j(x_j) \supseteq T\}.$$

When S is finite, each q_j can be calculated through $p^j_{\gamma_1\gamma_2...\gamma_n}$. For instance $q^j_{100} = p^j_{100} + p^j_{110} + p^j_{101} + p^j_{111}$ when S has three elements.

If we examine Eqs (11) and (12) carefully, it is easy to see that these two different calculation methods suggest two different procedures of combination. The former calculates the combined source first while the latter propagates the effect of each source to S first.

Dempster implicitly assumed that the results obtained in these two procedures are the same under the condition that the n sources are *statistically independent*.

If we refer to the levels containing spaces (X_i, χ_i, μ_i) as the *original information level*, and space S as the *target information level*, then Dempster's condition of independence is assumed at the original information level. This requirement is called *DS-independent* in [22]. Now the two approaches in Dempster's combination framework can be summarised as follows.

Approach 1: Combining information at the original information level by producing a joint space and a single probability distribution on the space. This procedure should consider the different mappings from the joint space to the target information space, unify the original mappings into one mapping and propagate the joint probability distribution to the target information level. The intuitive meaning of this method is shown in Figure 1.

Approach 2: Propagating different pieces of evidence at the original information level to the target information level separately and then combining them at the target information level. Figure 2 illustrates this procedure explicitly.

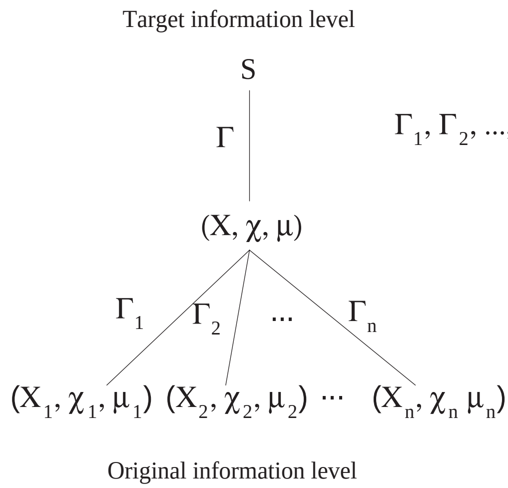

Fig. 1. Find a common probability space first and then propagate the combined probabilities distribution to the target space.

4.2 The condition of using Dempster's combination rule

When giving Dempster's combination rule in his book [14], Shafer followed the spirit of *Approach 2* suggested by Dempster, where function q is replaced by function m in Figure 2. On S, Shafer proposed a mathematical formula named Dempster's combination rule to combine the mass functions. If we follow the idea that the rule proposed by Shafer was derived from Dempster's combination framework *i.e., Approach 2*, then Dempster's combination rule should obey the condition defined by Dempster. Later, in some of his papers, Shafer indeed addressed the importance of independence among original sources. For example in [16], [17], Shafer stated that the condition of using Dempster's combination rule is that 'two or more belief functions on the same frame but based on independent arguments or items of evidence' and in [15] he used randomly encoded messages to describe the condition of using Dempster's combination rule. These explanations are much closer to the definition given by Dempster in his combination framework. However, as Dempster's combination rule does not require

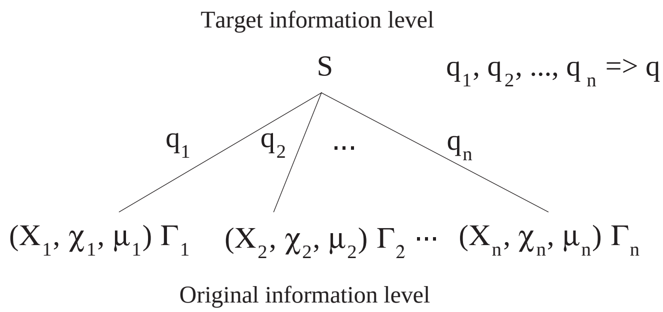

Fig. 2. Propagate evidence individually to the target space first and then carry out combination at the target information level.

or reflect any information about the sources which support the corresponding belief functions and it only needs belief functions (or mass functions) in order to carry out the combination, it is, therefore, difficult to describe independent condition precisely using only belief functions.

In contrast to the two views on belief functions in [8], we argue that the main cause of giving counterintuitive results[2] in using Dempster's combination rule is *overlooking (or ignorance of)* the condition of combination given in Dempster's original paper. In Shafer's simplified combination mechanism (*i.e.* Dempster's combination rule) the original sources are hidden. The invisibility of the original sources in the simplified combination mechanism makes it difficult to state the condition of the mechanism. In other words, Dempster's combination rule is too simple (compared to Dempster's combination framework) to show (or carry) enough information for a precise mathematical description of the dependent (or independent) relations between multiple pieces of evidence.

Proposition 1. *Two belief functions on a frame can be combined using Dempster's combination rule if the two sources,* (X_1, χ_1, μ_1) *and* (X_2, χ_2, μ_2)*, from which the two belief functions are derived, are statistically independent (or called DS-independent).*

[2] These do not include the situations such as Example 6.12 in [22] and the murderer case in Section 5.2 in [21]. Example 6.12 in [22] demonstrates that Dempster's combination rule agrees that a mass value assigned to a subset should not be evenly split among its elements but be kept for all the elements in this subset. The murderer case in Section 5.2 in [21] suggests that the normalisation function of Dempster's combination rule may not produce sensible results when two pieces of evidence are almost contradictory with each other. These examples are not within the province of this paper.

The idea of describing and judging dependent relations among the original probability spaces has also been mentioned implicitly by Shafer [15], Shafer and Tversky [19] and Voorbraak [22] but not explicitly defined at the original information level.

5 Examples

In this section, we use three examples to illustrate the delicate difference between the two approaches in Dempster's combination framework. The first two examples come from [16] where the former shows that when two pieces of evidence are statistically independent, both approaches in Dempster's combination framework are applicable, and the latter shows that when they are not statistically independent, only the first approach works. The third example ([22]) not only demonstrates that the careless application of Dempster's combination rule will yield a wrong result but also reveals the possibility of using *Approach 1* to combine two dependent pieces of evidence when their common probability space is known (as also investigated in [16], [17], [12]), although Dempster's original intention was aimed at dealing with independent probability spaces.

Example 2:

Suppose that Shafer wants to know whether the street outside is slippery, instead of observing this himself, he asks another person Fred. Fred tells him that 'it is slippery'. However Shafer knows that Fred is sometimes careless in answering questions. Based on his knowledge about Fred, Shafer estimates that 80% of the time Fred reports what he knows and he is careless 20% of the time. So Shafer believes that there is only a 80% chance that the street is slippery. In fact, Shafer forms two frames X_1 and S in his mind in order to get his conclusion in this problem, where X_1 is related to Fred's truthfulness and S is related to the possible answers of slippery outside.

$$X_1 = \{truthful, careless\},$$
$$S = \{yes, no\}.$$

Here *yes* and *no* stand for 'it is slippery' and 'it is not slippery' respectively. A probability measure p_1 on X_1 is defined as

$$\mu_1(\{truthful\}) = 0.8,$$
$$\mu_1(\{careless\}) = 0.2.$$

Shafer derives his conclusion when this probability measure is propagated from frame X_1 to frame S through a multivalued mapping function between X_1 and S as

$$\Gamma_1(truthful) = \{yes\},$$
$$\Gamma_1(careless) = \{yes, no\}.$$

Shafer obtains a belief function on S based on Eq (10).

$$bel_1(\{yes\}) = 0.8, \qquad bel_1(\{no\}) = 0.0. \tag{13}$$

Furthermore, suppose Shafer has some other evidence about whether the street is slippery: his trusty indoor-outdoor thermometer says that the temperature is 31 degrees Fahrenheit, and he knows that because of the traffic, ice could not form on the street at this temperature. However he knows that the thermometer could be wrong even though it has been very accurate in the past. Suppose that there is a 99% chance that the thermometer is working properly, so he could form another frame X_2 with its probability distribution as

$$X_2 = \{working, not\_working\},$$

$$\mu_2(\{working\}) = 0.99, \qquad \mu_2(\{not\_working\}) = 0.01,$$

and a mapping function Γ_2 as

$$\Gamma_2(working) = \{no\},$$

$$\Gamma_2(not\_working) = \{yes, no\}.$$

Therefore another belief function on S is calculated using Eq (10)

$$bel_2(\{yes\}) = 0.0, \qquad bel_2(\{no\}) = 0.99. \tag{14}$$

Now there are two pieces of evidence available regarding the same question 'slippery or not?' and Shafer wants to know what the joint impact of the evidence on S is. In the following, we try to solve this problem for Shafer in Dempster's combination framework.

Using Approach 1 in Dempster's combination framework:

First of all, there are two probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) carrying the original information (with $\chi_1 = 2^{X_1}$ and $\chi_2 = 2^{X_2}$), which are believed to be independent as Fred's answer is independent of the output of the thermometer. According to Dempster's explanation about independence of sources, Eq (11) is applied to combine these two sources to obtain a combined source and a joint multivalued mapping. As a result, the combined source is (X, χ, μ) where $\chi = 2^X$ with

$$X = X_1 \otimes X_2$$

and

$$\mu(\{(x_1, x_2)\}) = \mu_1(\{x_1\}) \times \mu_2(\{x_2\}),$$

such as

$$\begin{aligned}\mu(\{(truthful, working)\}) &= \mu_1(\{truthful\}) \times \mu_2(\{working\}) \\ &= 0.8 \times 0.99 = 0.792,\end{aligned}$$

as shown in the second column in Table 3. In the meanwhile, the joint multivalued mapping Γ from X to S is defined as

$$\Gamma((truthful, working)) = \emptyset,$$

$$\Gamma((truthful, not)) = \{yes\},$$

$$\Gamma((careless, working)) = \{no\},$$

$$\Gamma((careless, not)) = \{yes, no\},$$

after taking into account both what Fred and the thermometer have said. Here '*not*' means '*not_working*'. Element $(truthful, working)$ is the only element which matches the empty set in S, so that $\mu(S^*) = 1 - 0.792 = 0.208$. As a result, using Eq (3),

$$P_*(\{yes\}) = \mu(\{yes\}_*)/\mu(S^*) = \mu(\{(truthful, not)\})/\mu(S^*) \\ = 0.008/0.208 = 0.04.$$

This is the degree of our belief that the road is indeed slippery.

Alternatively, it is also possible to use Eq (10) to calculate the degree of our belief

$$bel(\{yes\}) = 0.04,$$

which is the same as P_*.

Table 3. The combined source is simply the set product of the original sources

$X = X_1 \otimes X_2$ combined source	$\mu = \mu_1 \otimes \mu_2$ joint probability	$\Gamma(x) \subseteq S$ joint mapping
$(truthful, working)$	0.792	$\emptyset$
$(truthful, not)$	0.008	$\{yes\}$
$(careless, working)$	0.198	$\{no\}$
$(careless, not)$	0.002	$\{yes, no\}$

Using Approach 2 in Dempster's combination framework:

In this approach, function q is used to calculate the joint impact on S without constructing the joint probability space, according to Eq (12).

Given two original probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) as defined in the first part of this example, it is possible to calculate q on S as

$$q(\emptyset) = q_1(\emptyset) \times q_2(\emptyset) = 1 \times 1 = 1,$$
$$q(\{yes\}) = q_1(\{yes\}) \times q_2(\{yes\}) = 1 \times 0.01 = 0.01,$$
$$q(\{no\}) = q_1(\{no\}) \times q_2(\{no\}) = 0.2 \times 1 = 0.2,$$
$$q(S) = q_1(S) \times q_2(S) = 0.2 \times 0.01 = 0.002.$$

Dempster also suggested that when S is finite, q can be expressed in the form of $p_{\gamma_1,\ldots\gamma_n}$. In this example, S contains only two elements, it is therefore possible to re-write function q as follows:

$$\begin{aligned}
&q(\emptyset) = q_{00} = p_{00} + p_{10} + p_{01} + p_{11} = 1,\\
&q(\{yes\}) = q_{10} = p_{10} + p_{11} = 0.01,\\
&q(\{no\}) = q_{01} = p_{01} + p_{11} = 0.2,\\
&q(S) = q_{11} = p_{11} = 0.002.
\end{aligned}$$

This set of equations determines $p_{\gamma_1\gamma_2}$ with following values:

$$p_{00} = 0.792, \qquad p_{10} = 0.008,$$

$$p_{01} = 0.198, \qquad p_{11} = 0.002.$$

Therefore, using the method shown in Table 1 or Table 2, the degree of our belief in the statement 'the outside is slippery' is

$$P_*(\{yes\}) = p_{10}/(1 - P_{00}) = 0.008/0.208 = 0.04 = bel(\{yes\}).$$

This result is consistent with what has been obtained in *Approach 1*.

Alternatively, as we believe that Dempster's combination rule, proposed by Shafer, is inspired by and an variation of Eq (12), it is also possible to apply the rule to two belief functions in (13) and (14) on S directly. The combined result is shown in Table 4.

Table 4. The direct application of Dempster's combination rule.

m		$\{yes\}$	0.8	$\{yes, no\}$	0.2
$\{no\}$	0.99	$\emptyset$	0.792	$\{no\}$	0.198
$\{yes, no\}$	0.01	$\{yes\}$	0.008	$\{yes, no\}$	0.002

The first row and the first column stand for the two mass functions derived from the two belief functions defined in (13) and (14) respectively. The combined mass function on *yes* is $m(\{yes\}) = 0.008/0.208 = 0.04$, so that $bel(\{yes\}) = 0.04$ which is identical to the result in both approaches.

$\diamondsuit$

Let us not forget Shafer's initial independence assumption between Fred's opinion and the output of the thermometer when we attempted to solve this problem. Next we will see when this assumption no longer holds, what results these approaches will offer.

Example 3:

Continuing with Example 2, assume Shafer believes that the Fred's answer relates to the thermometer as Fred accesses to the thermometer regularly to see whether it is working. If it is not working properly then Fred would be careless in answering questions. Assume that Fred has a 90% chance of being careless if the thermometer is not working, then Fred's answer is somehow affected when the thermometer is not working.

Using Approach 1 in Dempster's combination framework:

Because the two original probability spaces are not statistically independent, it is impossible to get the joint probability distribution μ on the joint space X by simply applying $\otimes$ to μ_1 and μ_2. Under this circumstance, Shafer somehow worked out an alternative joint probability distribution μ' on the same joint space X ($X = X_1 \otimes X_2$). The probability that μ' assigns on each element in X is shown in the second column in Table 5. Based on this newly created joint probability space (X, χ, μ') and the original multivalued mapping Γ between X and S (shown in the third column in Table 5), Eq (3) (or Eq (10)) once again is used to calculate the degree of our belief

$$\begin{aligned} P_*(\{yes\}) = \mu'(\{yes\}_*)/\mu'(S^*) &= \mu'(\{(truthful, not)\})/\mu'(S^*) \\ &= 0.001/0.201 = 0.005 = bel(\{yes\}), \end{aligned}$$

where $\mu'(S_*) = 1 - 0.799$.

Table 5. The combined space and its new probability distribution from the two original spaces, when the original two spaces are not independent

$X = X_1 \otimes X_2$ combined source	μ' joint probability	$\Gamma(x) \subseteq S$ joint mapping
$(truthful, working)$	0.799	$\emptyset$
$(truthful, not)$	0.001	$\{yes\}$
$(careless, working)$	0.191	$\{no\}$
$(careless, not)$	0.009	$\{yes, no\}$

This result is obviously different from the result in Example 2 because of the relationship between the two pieces of evidence.

Using Approach 2 in Dempster's combination rule:

If we took a chance and believed that *Approach 2* in Dempster's combination framework could be applied, then we would end up with exactly the same calculation procedure, and the same result of course, as in Example 2, because the two original sources are still the same. When function q is being calculated, there are no requirements on the joint probability of the originial sources, so the change Shafer made in μ' has no way to be reflected. Therefore, the second approach should not be used in the non-independent situation at all. As a straightforward consequence, Dempster's combination rule should not be allowed to apply in this case as well[3], as the rule is practically a different way of calculating q.

The summary of this analysis is shown in Table 6.

Table 6. Comparison of Dempster's original combination framework and Dempster's combination rule in different situations

	Dempster's combination framework		Dempster's combination rule
	Approach 1	Approach 2	bel_1 bel_2
Example 2	Applicable and correct	Applicable and correct	Applicable and correct
Example 3	Applicable and correct	Inapplicable	Inapplicable

We can see from this table that whenever *Approach 2* is applicable, Dempster's rule is guaranteed applicable and whenever *Approach 2* is **not** applicable, Dempster's combination rule **is not** applicable either.

$\diamondsuit$

If an agent uses Dempster's combination rule under the condition that the rule comes from Dempster's combination framework, then the agent can normally use the rule correctly. However, if an agent does not have Dempster's combination framework in mind when he intends to apply Dempster's combination rule, then it is sometimes very difficult to judge whether the rule is applicable, given two belief functions. We use next example to further emphasise the importance of bearing in mind the fact that Dempster's combination rule comes from *Approach 2* in Dempster's combination framework whenever Dempster's combination rule is being used.

Example 4 (from [22])

There are 100 labelled balls in an urn. Each ball must have either label a or b, in addition to some extra labels, x or y or xy.

Agents A and B give separate observations of drawing a ball from the urn as follows.

Agent A: the drawn ball has label x. The space of labels describing those balls is $X_1 = \{axy, ax, bxy, bx\}$ with probability distribution

$$\mu_1(\{axy\}) = \mu_1(\{ax\}) = 4/28,$$

$$\mu_1(\{bxy\}) = \mu_1(\{bx\}) = 10/28.$$

The probability $\mu_1(\{axy\}) = 4/28$ means that a drawing ball has label axy with probability 4/28 when it is known that the label definitely contains x.

[3] If Dempster's combination rule were applied, the result would still be $bel(\{yes\}) = 0.4$, which would have not shown the inter-relationship between the two mass functions.

Agent B: the drawn ball has label y. The space of labels describing those balls is $X_2 = \{axy, ay, bxy, by\}$ with probability distribution

$$\mu_2(\{axy\}) = 4/50, \quad \mu_2(\{ay\}) = 16/50,$$

$$\mu_2(\{bxy\}) = 10/50 \quad \mu_2(\{by\}) = 20/50.$$

The probability $\mu_2(\{axy\}) = 4/50$ means that a drawing ball has label axy with probability $4/50$ when it is known that the label definitely contains y.

Based on these two pieces of evidence, we are interested in knowing the degree of our belief that the drawn ball also has label b.

First of all, we try to work out the solution using Dempster's combination rule.

Using Dempster's combination rule:

Let $\{a, b\}$ be a frame of discernment, where a stands for 'the drawn ball has label a' and b stands for 'the drawn ball has label b'. Two mass functions are defined on S based on the information carried by two agents A and B as:

$$m_X(a) = 2/7, \ m_X(b) = 5/7,$$

$$m_Y(a) = 2/5, \ m_Y(b) = 3/5,$$

where $m_X(a)$ is the mass value on a given by agent A's observation which represents the possibility of a ball having label a when the ball is observed having label x and $m_Y(a)$ is the mass value on a given by agent B's observation which represents the possibility of a ball having label a when the ball is observed having label y.

The result of applying Dempster's combination rule to m_X and m_Y is $m(b) = m_X \oplus m_Y(b) = 15/19$. So $bel(b) = 15/19$.

Before judging whether this result is right or wrong, let us see what results the two approaches in Dempster's combination framework and Bayesian probability can offer.

Using Bayesian probability:

In probability theory, the full information of probability distribution on every possible label must be known. This information is provided in Table 7.

The probability that a ball has both labels x and y is

$$p(x)p(y) = 0.28 \times 0.5 = 0.14 = p(x \wedge y).$$

Therefore, the conditional probability that a drawn ball also has label b is $p(b \mid x \wedge y) = 5/7$ when both labels x and y are observed.

Using Approach 1 in Dempster's combination framework:

Table 7. All the possible labels, the number of balls with each possible label, and the prior probability of drawing a ball having a particular label.

Set of labels X	Number of balls having that label	μ
axy	4	0.04
ax	4	0.04
ay	16	0.16
a	16	0.16
bxy	10	0.1
bx	10	0.1
by	20	0.2
b	20	0.2

In fact, the information provided by agents A and B are carried by two probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) with $\chi_1 = 2^{X_1}$ and $\chi_2 = 2^{X_2}$. Let us construct another space containing labels a and b only, $S = \{a, b\}$, then there can be two multivalued mapping functions between the two spaces (provided by A and B) and S as

$$\Gamma_1(axy) = \Gamma_1(ax) = \{a\}, \Gamma_1(bxy) = \Gamma_1(bx) = \{b\},$$

$$\Gamma_2(axy) = \Gamma_2(ay) = \{a\}, \Gamma_2(bxy) = \Gamma_2(by) = \{b\}.$$

The idea of *Approach 1* is to get a combined space first before the combined probability is propagated to space S. In this case, these two probability spaces are not independent as they are from the same original probability space. Agents A and B have only provided partial information. So the combined probability distribution on the combined space $X = \{axy, bxy\}$ is not $\mu_1 \otimes \mu_2$ but the posterior probability distribution $\mu'(\bullet) = \mu(\bullet/x \wedge y)$ as given in column 3 in Table 8, after agents A and B's opinions are considered. The combined multivalued mapping function is detailed in the fourth column in Table 8.

Table 8. Possible labels after A and B's opinions are considered, the number of balls having each possible label, the posterior probability distribution, and a multivalued mapping between X and S

Set of labels X	Number of balls having that label	μ'	$\Gamma(\bullet) \subseteq S$
axy	4	4/14	a
bxy	10	10/14	b

Therefore, according to Eq (3), the degree of our belief that the drawn ball also has label b is $P_*(\{b\}) = \mu'(\{b\}_*)/\mu'(S^*) = (10/14)/1 = 5/7$, when agents A and B confirmed that the label of a drawn ball contains both x and y.

Using Approach 2 in Dempster's combination framework:

This is a typical situation where the two original probability spaces are not independent, so *Approach 2* cannot be used.

Discussion: Obviously the result obtained in DS theory is different from that obtained in probability theory and that in Dempster's combination framework, and the result given in DS theory is wrong. The very reason of this wrong result is that since the two pieces of evidence are not statistically independent, *Approach 2* cannot be applied. Therefore Dempster's combination rule should not be applied either.

$\diamondsuit$

The importance of considering relations among the original information sources has once again been discussed above. The result tells us that it is more natural to consider the combination at both the original information level and the target information level than only at the target information level.

6 Summary

In this paper, we have examined DS theory from the perspective of probability theory and tried to clarify the independence requirement in DS theory by defining the original information level and the target information level. We argue that any independent judgement in DS theory should be made explicitly at the original level. Merely considering Dempster's combination rule alone without examining the original information will cause problems. However Dempster's combination rule does not give us (or requires from us) any information about what the original sources are. So the conclusion we get from the above analysis is that some counterintuitive examples given in some articles [2], [9], [12], and [13][4] are caused by ignoring the independence requirement defined by Dempster in his combination framework. In the sense of statistically independence required by Dempster's combination framework, those examples do not satisfy this requirement, so Dempster's combination framework is not applicable. However if we merely consider Dempster's combination rule and believe that those examples satisfy the independence requirement needed by Dempster's combination rule then Dempster's combination rule is applicable, but the combination results are counterintuitive. From the former point of view, they are caused by the misapplication of the framework; and from the latter point of view, they are caused by the weakness of the combination rule. Neither of them is able to deal with those cases. Based on such a discussion, those belief functions, which can only be viewed as generalised probabilities, are precisely the cases which fail to sat-

[4] Some of the examples in [22] show the sensitivity of choosing frames. The author argued that the accuracy of reasoning results depends on at which level the frame is constructed such as Example 4.1. Some other examples explain that even though Dempster's combination rule can be used in some situations, the results are still counterintuitive (violate with common sense) like Example 6.12, due to the normalisation of the combination rule. This problem is also discussed in [21], [24] and [25]. Readers can refer to those papers if interested in more details.

isfy the requirement of DS-independence. So Dempster's combination rule is not suitable for coping with them.

References

1. J.A. Barnett. Computational methods for a mathematical theory of evidence. In P. J. Hayes (ed.), *Proc. of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81)*, William Kaufmann, 1981, pp.868-875.
2. P. Black. Is Shafer general Bayes? *Proc. of Third AAAI Uncertainty in Artificial Intelligence workshop*, 1987, pp.2-9.
3. D.A. Clark. Numerical and symbolic approaches to uncertainty management in AI. *Artificial Intelligence Review* **4**,109-146, 1990.
4. A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.* **38**,325-339, 1967.
5. R.O. Duda, P.E. Hart, N.J. Nisson. Subjective Bayesian methods for rule-based inference systems. *Proc. of the 1976 National Computer Conference (AFIPS)* 45, 1976, pp.1075-1082.
6. R. Fagin, J. Halpern. Uncertainty, belief and probability. N.S. Sridharan (ed.), *Proc. of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, Detroit, MI, USA, August. Morgan Kaufmann, pp.1161-1167.
7. R. Fagin, J. Halpern. Uncertainty, belief and probability, Research Report of IBM, RJ 6191. 1989.
8. J.Y. Halpern, R.Fagin. Two views of belief: belief as generalized probability and belief as evidence, *Artificial Intelligence* **54**, 275-317, 1992.
9. D. Hunter. Dempster-Shafer vs. probabilistic logic, *Proc. Third AAAI uncertainty in artificial intelligence workshop*, 1987, pp.22-29.
10. R. Kruse, E.Schwecke, J.Heinsohn. *Uncertainty and Vagueness in Knowledge-Based Systems*. Springer Verlag. 1992.
11. P. Krause, D. Clark. *Representing uncertain knowledge: An artificial intelligence approach*. Kluwer Academic Publishers: the Netherlands, 1993.
12. P. Lingras, S.K.M.Wong. Two perspectives of the Dempster-Shafer theory of belief functions, *Int. J. Man-Machine Studies* **33**, 467-487, 1990.
13. Pearl,J., *Probabilistic Reasoning in Intelligence Systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc., USA, 1988.
14. G. Shafer. *A Mathematical theory of evidence*, Princeton university press, 1976.
15. G. Shafer. Belief functions and parametric models, *J. R. Stat. Soc. Ser.*(B) **44**, 322-352, 1982.
16. G. Shafer. Probability judgement in artificial intelligence. In Kanal and Lemmer (eds.), *Uncertainty in Artificial Intelligence 2*, 1986, pp. 127-135.
17. Shafer,G., Probability judgement in artificial intelligence and expert systems, *Stat. Sci.* **2**, 3-16, 1987.
18. G. Shafer. Perspectives on the theory and practice of belief functions, Working paper No.218. Univ. of Kansas, School of Business. 1990.
19. G. Shafer. A.Tversky. Languages and Designs for Probability Judgement, *Cognitive Science* **9**, 309-339. 1985.
20. E.H. Shortliffe, B. Buchanan. A Model of Inexact Reasoning in Medicine, *Mathematical Biosciences*, **23**, 351-379, 1975.
21. Ph. Smets. Belief functions. In Smets, Mamdani, Dubois and Prade (eds.), *Non-Standard Logics for Automated Reasoning*, 1988, pp. 253-286.

22. F. Voorbraak. On the justification of Dempster's rule of combination, *Artificial Intelligence* **48**, 171-197. 1991.
23. Yager, R.R., Fedrizzi, M. and Kacprzyk, J. (eds). *Advances in the Dempster-Shafer Theory of Evidence* John Wiley & Sons, Inc., 1994.
24. L.A. Zadeh. A mathematical theory of evidence (book review), *AI Mag.* **5**(3), 81-83, 1984.
25. L.A. Zadeh. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination, *AI Mag.* **7**,85-90, 1986.

Dr. Weiru Liu is a lecturer of Computer Science at the University of Ulster at Jordanstown. Dr. Liu received her BSc and MSc degrees in Computer Science from Jilin University, and her PhD degree in Artificial Intelligence from the University of Edinburgh. Dr. Liu's main research interests include knowledge representation and reasoning, probabilistic logic, approximate reasoning, and knowledge discovery in databases.

Dr. Jun Hong is a lecturer of Computer Science at the University of Ulster at Jordanstown. Dr. Hong received his BSc and MSc degrees in Computer Science from Jilin University, and his PhD degree in Artificial Intelligence from the University of Ulster at Jordanstown. Dr. Hong has worked in several areas of Artificial Intelligence, including plan recognition, reasoning under uncertainty, intelligent user interfaces, intelligent tutoring systems and software agents.