REPRESENTING HEURISTIC KNOWLEDGE AND PROPAGATING BELIEFS IN THE DEMPSTER-SHAFER THEORY OF EVIDENCE

Weiru Liu
Dept of AI
Univ. of Edinburgh
Edinburgh EH1 1HN

John G. Hughes, and Michael F. McTear
Dept. of Information Systems
Univ. of Ulster at Jordanstown
Co. Antrim BT37 0QB,U.K.

July, 1991

Abstract

The Dempster-Shafer theory of evidence has been used intensively to deal with uncertainty in knowledge-based systems. However the representation of uncertain relationships between evidence and hypothesis groups (heuristic knowledge) is still a major research problem. This paper presents an approach to representing such heuristic knowledge by **evidential mappings** which are defined on the basis of mass functions. The relationships between evidential mappings and multivalued mappings, as well as between evidential mappings and Bayesian multi- valued causal link models in Bayesian theory are discussed. Following this the detailed procedures for constructing evidential mappings for any set of heuristic rules are introduced. Several situations of belief propagation are discussed. Shafer's partition technique is used to get the mass function in a complex evidence space when the antecedent frame of a rule is made of several variables. Matrices and vectors are used to represent uncertain relationships in heuristic knowledge and pieces of evidence respectively.

REPRESENTING HEURISTIC KNOWLEDGE AND PROPAGATING BELIEFS IN THE DEMPSTER-SHAFER THEORY OF EVIDENCE

Weiru Liu[1] John G. Hughes Michael F. McTear
Dept. of Information Systems
Univ. of Ulster at Jordanstown
Co. Antrim BT37 0QB,U.K.

Abstract

The Dempster-Shafer theory of evidence has been used intensively to deal with uncertainty in knowledge-based systems. However the representation of uncertain relationships between evidence and hypothesis groups (heuristic knowledge) is still a major research problem. This paper presents an approach to representing such heuristic knowledge by **evidential mappings** which are defined on the basis of mass functions. The relationships between evidential mappings and multivalued mappings, as well as between evidential mappings and Bayesian multi- valued causal link models in Bayesian theory are discussed. Following this the detailed procedures for constructing evidential mappings for any set of heuristic rules are introduced. Several situations of belief propagation are discussed. Shafer's partition technique is used to get the mass function in a complex evidence space when the antecedent frame of a rule is made of several variables. Matrices and vectors are used to represent uncertain relationships in heuristic knowledge and pieces of evidence respectively.

Keywords: evidential reasoning, representing heuristic knowledge, evidential mappings.

1 Introduction

In the design and implementation of expert systems and decision making systems, the problem of uncertain knowledge and evidence has to be solved. Several approaches can be used to deal with this problem, such as Mycin's certainty factors, Prospector's inference nets, fuzzy sets, Bayesian nets and Dempster-Shafer's belief functions.

Generally speaking, there are two kinds of problem involving uncertainty: one is caused by uncertain evidence; another is caused by uncertain knowledge, i.e. heuristic knowledge. The former is a result of ill-defined concepts in the observation, or due to inaccuracy and poor reliability of the instruments used to make the observations. The latter is a result of weak implication which occurs when the expert or model builder is unable to establish a strong correlation between premise (or evidence) and conclusion (or hypotheses) (Bonissone and Tong 1985).

The Dempster-Shafer theory of evidence provides a flexible approach to representing uncertain evidence. This theory, which is claimed as a generalization of Bayesian inference (Shafer 1976, 1981), has the advantages of representing ignorance of evidence and narrowing the hypothesis space as a result of evidence accumulation. Several AI implementations have been undertaken (Laskey et al 1989, Lowrance et al 1986, Strat 1987, Wesley 1988, Yen 1989, Zarley et al 1988) based on the theory or extended versions of the theory (Laskey and Lehner 1989; Yen 1989). In this paper we argue that it is difficult to represent uncertain heuristic knowledge in this theory; however in most complex domains, heuristic knowledge plays an important role in solving problems.

Consider the following piece of heuristic knowledge: if X *is* X_1, then Y *is* Y_1 with a degree of belief r_1. If we get a piece of evidence which says that X *is* X_1 with a degree of a_1, by invoking this rule we should be able to obtain the corresponding degree y_1 for Y *is* Y_1. Certainly the value of y_1 must be a function F of a_1 and r_1 (i.e. $y_1 = F(a_1, r_1)$).

More generally, we suppose that a set of heuristic rules R includes:

R_1: if E_1 then H_{11} with a degree of belief r_{11};
H_{12} with a degree of belief r_{12};

[1]The current address of the author is Dept. of AI, Univ. of Edinburgh, 80 South Bridge, Edinburgh EH1 1HN.

$$
\begin{array}{lll}
 & & \ldots \\
R_2\text{: if } E_2 \text{ then} & H_{21} & \text{with a degree of belief } r_{21}; \\
 & H_{22} & \text{with a degree of belief } r_{22}; \\
 & \ldots & \\
\ldots & & \\
R_n\text{: if } E_n \text{ then} & H_{n1} & \text{with a degree of belief } r_{n1}; \\
 & H_{n2} & \text{with a degree of belief } r_{n2}; \\
 & \ldots &
\end{array}
\tag{1}
$$

where $E_1, E_2, ..., E_n$ are values (or propositions) of the variable E, and E_i is called an antecedent of rule R_i. H_{ij} in rule R_i is a subset of the values (or propositions) of the variable H and it is called one of the conclusions of rule R_i. r_{ij} is called a rule strength.

Assume we have a piece of evidence which says that E_1 is confirmed with a_1, E_2 is confirmed with a_2, ..., E_n is confirmed with a_n, how can we solve the following problems:

1. what conditions should $\sum_i a_i$ satisfy?
2. what conditions should $\sum_j r_{ij}$ satisfy?
3. what is the function F to determine h_{ij}(the degree of belief on H_{ij}) from those a_i and r_{ij}?
4. if more than one set of rules is invoked and the same conclusion H_{ij} is obtained, what will be the final degree of belief on H_{ij} from those $h_{ij}, ..., h_{kl}$?

Generally, if the variable E is a Cartesian product of variables $A, B, ..., C$, that is each E_i is in a form of $(A_i \; and \; B_j and ... \; and C_k)$, assuming we know the evidence for $A, B, ..., C$, then

5. what is the function F' to determine the degree of belief on the premise $(A_i \; and B_j \; and ... \; and C_k)$?

These problems have been modelled in fuzzy theory using a fuzzy extension of modal logic, based on Zadeh's concepts of necessity and possibility (Prade 1981). They were also solved in Mycin's certainty factor model (Shortliffe and Buchanan 1976). Can these problems be solved in Dempster-Shafer theory?

In this paper we analyze these problems and propose our approaches for solving them by extending the theory of evidence. The paper is organized as follows. In section 2 the basics of Dempster- Shafer theory are introduced and the approach for representing heuristic knowledge by evidential mappings is described in which a matrix is used to represent the uncertain relationships between evidence and conclusions. In section 3 the relations between Bayesian inference and evidential mappings are examined in which it is proved that the multi-valued causal links between hypotheses space H and evidence space E (Pearl 1988) in Bayesian theory is consistent with the special case of evidential mappings. In section 4 the method of constructing a complete evidential mapping matrix for an evidential mapping of a heuristic rule is discussed. In section 5 belief propagation approaches are discussed for different situations. Finally a conclusion is given along with some consideration of related work.

2 Representing heuristic knowledge in the Dempster-Shafer Theory

The Dempster-Shafer theory of evidence (which is also called the theory of belief functions (Smets 1988, Shafer 1990)) provides an alternative approach to drawing plausible conclusions from uncertain and incomplete evidence. It is a generalization of the Bayesian theory of subjective probability, it is more flexible, and it allows us to derive degrees of belief for a question from probabilities of a related question (Shafer 1990).

2.1 The Basics of Dempster-Shafer Theory of Evidence

Suppose Θ is a finite set, which consists of mutually exclusive and exhaustive propositions of a problem or all values of a variable, 2^Θ is the set of all subsets of Θ. A function *Bel*: $2^\Theta \longrightarrow [0,1]$ is called a belief function in Shafer (1976), if it satisfies the following conditions:

1. $Bel(\emptyset) = 0$;
2. $Bel(\Theta) = 1$;
3. for every positive integer n and every collection $A_1, ..., A_n$ of subsets of Θ,

$$Bel(A_1 \cup ... \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + -... + (-1)^{n+1} Bel(A_1 \cap ... \cap A_n)$$

Such a set Θ is called a **frame of discernment**.

By knowing a belief function on a frame of discernment, another function m can be calculated as:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \quad \forall A \subseteq \Theta$$

where $|A - B|$ denotes the number of elements in the set of $A - B$.

The function m is called a **basic probability assignment** *(bpa)* or a **mass function**. Obviously a mass function has the features that $m(\emptyset) = 0$ and $\sum m(A) = 1$ for all subsets A of Θ. A subset A is called a focal element of the belief function *Bel* if $m(A) > 0$. Recovering the belief function *Bel* from a mass function m is carried out by

$$Bel(B) = \sum_{A \subseteq B} m(A)$$

If all the focal elements of a belief function are singletons of Θ, then the corresponding mass function m is a Bayesian subjective probability distribution.

A belief function (or a mass function) on a frame Θ can either be directly obtained from a piece of evidence or calculated from a probability measure P on the related frame T by a *multivalued mapping* Γ between T and Θ. The term multivalued mapping was originally introduced by Dempster in his early article (Dempster 1967) in which he defined that a multivalued mapping Γ assigns each element t of T to a subset A of Θ. Suppose we get a probability measure P on frame T, two other functions on frame Θ will be obtained by a multivalued mapping Γ as:

$$m(A) = \sum_{\Gamma(t)=A} P(t) \quad t \in T, A \subseteq \Theta$$

$$Bel(A) = \sum_{\Gamma(t)\subseteq A} P(t) \quad t \in T, A \subseteq \Theta$$

where $P(t)$ is the probability distributed on t by P and $\sum_i P(t_i) = 1$. Here we suppose that no element t of T maps to the empty set in Θ. It is easy to prove that m is a mass function and *Bel* is a belief function on Θ. Another name for multivalued mappings is compatibility relations which is used in Shafer (1987), Lowrance et al (1986), and Shafer and Srivastava (1990).

The impact of several belief functions (or mass functions) on the same frame of discernment is obtained by using Dempster's rule of combination which treats Bayesian conditioning probabilities as a special case (Shafer 1976). Dempster's rule of combining two belief functions Bel_1 and Bel_2 can be defined by a relatively simple rule in terms of the corresponding mass functions m_1 and m_2.

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

This rule requires that the combined belief functions (or their mass functions) are independent. This condition has been further enhanced as *DS-independent* in Voorbraak (1991).

2.2 Representing Heuristic Knowledge in The Dempster-Shafer Theory

It is obvious that a heuristic rule like {if X *is* X_1 then Y *is* Y_1 with a degree of belief r_1} cannot be directly represented in D-S theory. Some work concerning this topic was carried out previously (Ginsberg 1984, Yen 1988, Liu 1986, Hau and Kashyap 1990). We propose that evidential mappings which are defined on the basis of mass functions can be used to represent the uncertain relationships between evidence and conclusions.

Definition 1 *An evidential mapping is the mapping from one frame of discernment to another, which represents causal links among elements of two frames of discernment in the form of mass functions. Formally an evidential mapping from frame Θ_E to frame Θ_H is a function Γ^*: $\Theta_E \longrightarrow 2^{2^{\Theta_H \times [0,1]}}$. The image of each element in Θ_E, denoted by $\Gamma^*(e_i)$, is a collection of subset-mass pairs:*

$$\Gamma^*(e_i) = \{(H_{i1}, f(e_i \to H_{i1})), ..., (H_{im}, f(e_i \to H_{im}))\}$$

and let $\Theta_i = \cup_{j=1}^{m} H_{ij}\ \ H_{ij} \subseteq \Theta_H$
that satisfies the following conditions:

a. $H_{ij} \neq \emptyset$ $\qquad j = 1, ..., m$
b. $f(e_i \to H_{ij}) > 0$ $\qquad j = 1, ..., m$
c. $\sum_j (e_i \to H_{ij}) = 1$

There is a set of heuristic rules, denoted as **R**, related to an evidential mapping, each of which is in the form of

$$\begin{aligned} R_i : e_i &\longrightarrow H_{i1\ (f(e_i \to H_{i1}))}; \\ &...\ ; \\ e_i &\longrightarrow H_{im\ (f(e_i \to H_{im}))}. \end{aligned}$$

$f(e_i \to H_{ij})$ represents our belief exactly on H_{ij} given condition e_i, and it is in the range of [0,1].

A rule states that if e_i is true then the truth of the problem carried by Θ_H is in H_{i1} with the degree of belief $f(e_i \to H_{i1})$ exactly committed to H_{i1}, ..., in H_{im} with the degree of belief $f(e_i \to H_{im})$ exactly committed to H_{im}. The e_i is called the antecedent of rule R_i and it is an element of Θ_E. H_{ij} is called one of the conclusions of rule R_i, and it is a subset of Θ_H. We name Θ_E and Θ_H as antecedent frame and conclusion frame of R respectively.

The corresponding matrix of this evidential mapping is:

$$\begin{array}{ccccc} row/col & H_1 & H_2 & ... & \Theta \\ \\ \{e_1\} & m_{11} & m_{12} & ... & m_{1l} \\ \{e_2\} & m_{21} & m_{22} & ... & m_{2l} \\ & & ... & & \\ \{e_n\} & m_{n1} & m_{n2} & ... & m_{nl} \end{array} = M$$

The size of matrix M is $n \times l$ where n is the number of elements in Θ_E and l equals $|2^{\Theta_H}| - 1$ (except $\emptyset$). H_i is a subset of the elements of Θ_H. For any H_{ij} appearing in $(H_{ij}, f(e_i \to H_{ij}))$ there is H_k where $H_k = H_{ij}$. The (i,k)-th entry of M is defined as m_{ik} which equals $f(e_i \to H_{ij})$ if the pair $(H_{ij}, f(e_i \to H_{ij}))$ is an element of $\Gamma^*(e_i)$ and $H_k = H_{ij}$; otherwise m_{ik} equals 0. Thus those $m_{i1}, m_{i2}, ..., m_{il}$ of line i must satisfy the condition $\sum_j m_{ij} = 1$. More precisely based on

$m_{i1}, m_{i2}, ..., m_{il}$, we define a function m_i. In fact m_i is a mass function on $\Theta_E \times \Theta_H$, with its focal elements as $A_{i1} = \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H_{i1}\}, ..., A_{im} = \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H_{im}\}$ and $m_i(A_{ij}) = m_{ij}$ for $j = 1, ..., l$. So there are in total n mass functions on frame $\Theta_E \times \Theta_H$. But we define that the combination of any two of the above mass functions is meaningless.

In order to identify each row and column in M we call H_k the *title of column* k, $\{e_i\}$ the *title of row* i. We also call $[\{e_1\}, \{e_2\}, ..., \{e_n\}]$ and $[H_1, H_2, ..., \Theta]$ the *row title vector* and *column title vector* of M respectively. When we mention a matrix M of an evidential mapping, we assume the row title vector and the column title vector are known. Thus for any given evidential mapping the related heuristic rule set and the matrix are unique.

An evidential mapping from Θ_E to Θ_H states that for two related questions represented by Θ_E and Θ_H, if the truth for the question represented by Θ_E is e_i then the truth for the question represented by Θ_H is in a set Θ_i, but e_i has different inter-relationships with different subsets of Θ_i. The $f(e_i \rightarrow H_{ij})$ is used to reflect the sensitivity or strength of interrelation between e_i and H_{ij}. Certainly the total strength should be 1.

Example 1: If an evidential mapping Γ^* specifies a mapping from an evidence space Θ_E to a hypothesis space Θ_H as:

$$\begin{aligned}
\Gamma^*(e_1) &= \{(\{a_1, a_2\}, 0.7), (\{a_3, a_4\}, 0.3)\} \\
\Gamma^*(e_2) &= \{(\{a_2, a_3\}, 0.8), (\Theta_H, 0.2)\} \\
\Gamma^*(e_3) &= \{(\{a_4, a_5\}, 0.9), (\Theta_H, 0.1)\}
\end{aligned}$$

and a related set of heuristic rules is

$$\text{R:} \quad \begin{array}{ll}
e_1 \longrightarrow \{a_1, a_2\}_{\ (0.7)}; & e_1 \longrightarrow \{a_3, a_4\}_{\ (0.3)}. \\
e_2 \longrightarrow \{a_2, a_3\}_{\ (0.8)}; & e_2 \longrightarrow \Theta_{H\ (0.2)}. \\
e_3 \longrightarrow \{a_4, a_5\}_{\ (0.9)}; & e_3 \longrightarrow \Theta_{H\ (0.1)}.
\end{array}$$

where $\Theta_E = \{e_1, e_2, e_3\}$ and $\Theta_H = \{a_1, a_2, a_3, a_4, a_5\}$, then the matrix M has $2^5 - 1$ columns, most of which have only zero m_{ij} such as columns $\{a_1\}, \{a_1, a_2, a_3\}$. Usually a matrix becomes too big when Θ_H contains several elements. So we delete all those columns which have only zero m_{ij} and form another matrix. We call such a simplified matrix the **Basic Matrix** and denote it as **BM**. Thus the title vector of a basic matrix of an evidential mapping only contains those H_{ij} which appear in $\Gamma^*(e_i)$. The BM of this evidential mapping in the above example is

$$\begin{array}{cccccc}
row/col & \{a_1, a_2\} & \{a_2, a_3\} & \{a_3, a_4\} & \{a_4, a_5\} & \Theta_H \\
\{e_1\} & 0.7 & 0.0 & 0.3 & 0.0 & 0.0 \\
\{e_2\} & 0.0 & 0.8 & 0.0 & 0.0 & 0.2 \\
\{e_3\} & 0.0 & 0.0 & 0.0 & 0.9 & 0.1
\end{array} = BM$$

with row title vector $[\{e_1\}, \{e_2\}, \{e_3\}]$ and column title vector $[\{a_1, a_2\}, \{a_2, a_3\}, \{a_3, a_4\}, \{a_4, a_5\}, \Theta_H]$.

Obviously, multivalued mappings in section 2.1 and Bayesian multi-valued causal link models (Pearl 1988) can all be represented using such evidential mappings.

Corollary 1 *If all the m_{ij} in a basic matrix BM of an evidential mapping from Θ_E to Θ_H are either 1 or 0 then the evidential mapping is a multivalued mapping. For any e_i, the mass function m_i on $\Theta_E \times \Theta_H$ is a simple support function with a focal element $A_{ij}(A_{ij} = \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H_{ij}\})$, and $m_i(A_{ij}) = 1$.*

Corollary 2 *If a basic matrix BM has $|\Theta_H|$ columns, and the titles of all columns are singletons of Θ_H then the evidential mapping from Θ_E to Θ_H of this matrix is exactly a Bayesian multi-valued causal link model. For any e_i, the mass function m_i on $\Theta_E \times \Theta_H$ is a Bayesian probability distribution. We refer to this kind of evidential mappings as* Bayesian evidential mappings.

If a piece of evidence gives a probability distribution P on Θ_E, then a new function m on Θ_H can be calculated by the evidential mapping from Θ_E to Θ_H:

$$m(H_k) = \sum_i P(e_i) \times m_{ik} = \sum_i P(e_i) \times f(e_i \to H_k) \quad \text{When } H_k \text{ is the title of a column}$$
$$m(H_k) = 0 \quad \text{Otherwise} \tag{2}$$

The function m is a basic probability assignment in the hypothesis space and has the following features:

1) $m(\emptyset) = 0$
2) $\sum_k m(H_k) = 1$ *where* $H_k \subseteq \Theta$

This can be proved by the following according to definition 1, probability distribution P and features of a mass function.

$$\begin{aligned}\sum_k m(H_k) &= \sum_k \sum_i P(e_i) \times f(e_i \to H_k)\\ &= \sum_i \sum_k P(e_i) \times f(e_i \to H_k)\\ &= (\sum_i P(e_i))(\sum_k f(e_i \to H_k))\\ &= (\sum_i P(e_i))(\sum_j f(e_i \to H_{ij})) \text{ (because there exists an } H_{ij} \text{ such that } H_k = H_{ij})\\ &= 1 \times 1 = 1\end{aligned}$$

Corollary 3 *A function m is a mass function on frame Θ_H if it is given by formula 2 under the condition that P is a probability distribution on space Θ_E and Γ^* is an evidential mapping from Θ_E to Θ_H.*

The theoretical support of the formula (2) is Bayes' formula

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

where B_i is an element of an exhaustive and mutually exclusive event set (Pearl 1988).

We suppose that any evidence e giving $P(B_i/e)$ has no effect on $P(A|B_i)$. This rule is also called Jeffrey's rule of conditioning (Jeffrey 1965, Shafer 1981).

2.3 Creating Evidential Mappings for Incomplete Heuristic Rule Sets

We have seen in the above section that an evidential mapping can be associated with a set of heuristic rules. The other way around, given a set of heuristic rules in the form of (1), if all the antecedents of rules can form a frame of discernment Θ_E, all the conclusions of rules can form another frame of discernment Θ_H, and for any heuristic rule R_i the sum of r_{ij} (for $j = 1, .., m$) is 1, then an evidential mapping can be established between Θ_E and Θ_H. Unfortunately, the antecedents (or conclusions) of a set of rules normally cannot form a frame of discernment which is mutually exclusive and exhaustive and usually the sum of r_{ij} for rule R_i is less than 1. For example if there is only one rule in a set of heuristic rules: if X *is* X_1 then Y *is* Y_1 with a degree of belief r_1, then the antecedent (X *is* X_1) itself does not form a frame of discernment at all nor does the conclusion (Y *is* Y_1).

Definition 2 *If at least one of the antecedent frame and conclusion frame of a heuristic rule set R is not a frame of discernment or there is a rule R_i in rule set R where $\sum_j r_{ij} < 1$, then we define such a heuristic rule set as an* **incomplete heuristic rule set**. *Otherwise we call it a complete heuristic rule set.*

Corollary 4 *Given an incomplete heuristic rule set R, let $E = \{e_1, ..., e_n\}$ represent the antecedent set, and $H = \{h_1, ..., h_m\}$ represent the conclusion set of R,* *1) if E is not a frame of discernment then define $e_{n+1} = \neg(e_1 \vee ... \vee e_n)$ and $\Theta_E = \{e_{n+1}\} \cup E$; otherwise define $\Theta_E = E$.*

2) if H is not a frame of discernment then define $h_{m+1} = \neg(h_1 \vee ... \vee h_m)$ and $\Theta_H = \{h_{m+1}\} \cup H$; otherwise define $\Theta_H = H$.

3) if e_{n+1} exists then add the rule $R_{n+1} : (e_{n+1} \longrightarrow \Theta_{H\ (1)})$ to rule set R.

Then Θ_E and Θ_H are two frames of discernment representing the antecedent frame and the conclusion frame of R respectively.

Corollary 5 *For each rule R_i in R, if $\sum_j r_{ij} < 1$ then we add an extra conclusion Θ_H with belief $r_{ih} = (1 - \sum_j r_{ij})$ to R_i. That is, if the original R_i is*

$$R_i : \quad e_i \longrightarrow H_{i1\ (r_{i1})}; ...; e_i \longrightarrow H_{im\ (r_{im})}. \text{ where } \sum_j r_{ij} < 1$$

then a new R_i' is

$$R_i' : \quad e_i \longrightarrow H_{i1\ (r_{i1})}; ...; e_i \longrightarrow H_{im\ (r_{im})}; e_i \longrightarrow \Theta_{H\ (r_{ih})} \text{ where } r_{ih} = 1 - \sum_j r_{ij}.$$

Now the heuristic rule set R is complete and an evidential mapping from Θ_E to Θ_H can be created.

In fact the added part of a rule represents our ignorance. In other words, based on the current knowledge of a specific domain, we have no knowledge to identify any more ad-hoc relationships among elements of reasons and results.

Example 2: Suppose we have a rule set R which consists of a rule as follows:

Smoke alarm is ringing $\longrightarrow$ There is a fire $_{(0.9)}$.

We construct Θ_E={(smoke alarm is ringing), not(smoke alarm is ringing)}, Θ_H={(there is a fire), not(there is a fire)} based on corollary 4, and a new rule set R' based on corollary 5,

R' has: R_1: Smoke alarm is ringing $\longrightarrow$ There is a fire $_{(0.9)}$;
Smoke alarm is ringing $\longrightarrow \Theta_{H\ (0.1)}$.
R_2: Not (Smoke alarm is ringing) $\longrightarrow \Theta_{H\ (1)}$.

This rule set can be associated with an evidential mapping from Θ_E to Θ_H. In particular, if Θ_H is the same as Θ_E then the corresponding evidential mapping represents self-relations of Θ_E (it is also called delta-Θ_E compatibility relation by Lowrance et al (1986)).

Now we can represent any heuristic rule set (either complete or incomplete) in the Dempster-Shafer theory of evidence by the means of evidential mappings. In the following we simply use a triple (R, Θ_E, Θ_H) to represent an evidential mapping where R is a heuristic rule set, Θ_E is the antecedent frame of discernment of R and Θ_H is the conclusion frame of discernment of R.

3 The Relation Between Evidential Mappings and Bayesian Conditional Probabilities

The Dempster-Shafer theory of evidence as an generalization of Bayesian inference includes two meanings: mass functions are the general form of Bayesian subjective probabilities in representing evidence; Bayesian conditional probabilities are a special case of Dempster's rule of combination (Shafer 1976). Pearl (1988) gave a general formula to calculate posterior-probabilities (on hypotheses) or predict future events in multi-valued causal link models of Bayesian theory when a set of evidence (for evidence variable) is given. In fact, Pearl's work is the extension of traditional Bayesian inference theory to the situation when the relationships among elements of an evidence space and a hypothesis space are multi-valued causal mappings. In this section we prove that Bayesian inference performed on multi-valued causal link models can be carried out in D-S theory by using evidential mappings.

3.1 Predicting Future Events in D-S Theory

Example 3: Let S be a variable for "alarm sound" and D for "a person's call". If we use the same capital letter to represent both a variable name and the name of the frame which includes all the values of the variable, we have $S=$ {*alarm on, alarm off*} and $D=${*a person will call, a person will not call*} each of which represents an exhaustive and mutually exclusive set of propositions. Suppose the causal link between S and D is

S/D	*will call*	*will notcall*
alarm on	0.7	0.3
alarm off	0.0	1.0

Bayesian inference produces

$$P(d_i) = \sum_j P(d_i|s_j)P(s_j) \tag{3}$$

which is a shorthand notation for the statement

$$P(d_i \mid e) = \sum_j P(d_i|s_j, e)P(s_j \mid e)$$

where d_i is an element of D and s_j is an element of S and we assume that a piece of evidence has no effect on the causal link between S and D. Given a probability distribution of a piece of evidence on S, the probabilities on D can be calculated from formula (3).

Suppose $P(s_1{=}on){=}0.2686$, $P(s_2{=}off){=}0.7314$,

$$\textit{then}\ P(d_1 = \textit{will call}) = \sum_j P(d_1 \mid s_j)P(s_j) = [0.2686, 0.7314] \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix} = 0.188 \tag{4}$$

This is called *predicting future events* by Pearl in Bayesian inference.

Obviously the causal link above forms an evidential mapping from S to D in Dempster-Shafer theory. In the condition of prior probabilities $P(s_1 = on) = 0.2682$, $P(s_2 = off) = 0.7314$, applying formula (2) we get a mass function on D which is the same as that showed in formula (4).

$$m(d_1) = \sum_i P(s_i) \times f(s_i \to \{d_1\}) = 0.2686 \times 0.7 + 0.7314 \times 0.0 = 0.188$$

$$m(d_2) = \sum_i P(s_i) \times f(s_i \to \{d_2\}) = 0.2686 \times 0.3 + 0.7314 \times 1.0 = 0.812$$

In Bayesian multi-valued causal link models, the causal link between the hypothesis space H and the evidence space E is identified by a $n \times m$ matrix M, where n and m are the numbers of values of H and E respectively, and the (i, j)-th entry of M is $M_{ij} = P(e_j \mid h_i)$ (Pearl 1988).

It is easy to see (corollary 3) that the causal link model above is consistent with the special case of evidential mappings. The mass function on D obtained from formula (2) is exactly the same as the probability distribution on D obtained in Bayesian inference.

3.2 Calculating Posterior Probabilities in D-S Theory

Furthermore, in Bayesian multi-valued causal link models, given a prior probability distribution on hypothesis space H, causal link matrix M with $M_{ij} = P(e_j \mid h_i)$

$$\begin{array}{c|cccc} & \{e_1\} & \{e_2\} & ... & \{e_m\} \\ \hline \{h_1\} & p(e_1 \mid h_1) & p(e_2 \mid h_1) & ... & p(e_m \mid h_1) \\ \{h_2\} & p(e_1 \mid h_2) & p(e_2 \mid h_2) & ... & p(e_m \mid h_2) \\ & & ... & & \\ \{h_n\} & p(e_1 \mid h_n) & p(e_2 \mid h_n) & ... & p(e_m \mid h_n) \end{array} = M$$

and a set of evidence $e^1, e^2, ..., e^N$ on evidence space E, then posterior-probability $P(h_i \mid e^1, e^2, ..., e^N)$ on h_i of H is:

$$P(h_i \mid e^1, e^2, ..., e^N) = \alpha P(e^1, e^2, ..., e^N \mid h_i) P(h_i) \quad (5)$$

where $\alpha = [P(e^1, e^2, ..., e^N)]^{-1}$ is a normalizing constant to be computed by requiring that Eq.(5) sum to unity. Assuming $e^1, e^2, ..., e^N$ are independent with each other and conditional independence of respect to each h_i, Pearl (1988) indicated that,

$$P(h_i \mid e^1, e^2, ..., e^N) = \alpha P(h_i)[\Pi_{k=1}^{N} P(e^k \mid h_i)] \quad (6)$$

Here we should make it clear that Pearl assumes that for each piece of evidence e^k there exists an element e_j in E where $p(e_j) = 1$ given by e^k so that $P(e^k \mid h_i) = P(e_j \mid h_i)$.

Can these posterior probabilities be calculated in D-S theory using evidential mappings based on the above causal link matrix under these assumptions? The following theorem indicates that they can.

Theorem 1 *Let E and H be two frames of discernment, Γ^* be a Bayesian evidential mapping from H to E, BM be the basic matrix of the mapping Γ^* with (i,j)-th entry as $p(e_j \mid h_i)$. Assume the prior probability on h_i of H is $p(h_i)$, a set of evidence on E is $e^1, e^2, ..., e^N$ for each of which there exists an e_l where $p(e_l) = 1$. Then the final belief function* Bel *on H using D-S theory is*

$$Bel(h_i) = \alpha p(h_i)[\Pi_{k=1}^{N} p(e^k \mid h_i)]$$

$$where \ \ \alpha = (\sum_{1}^{n}(p(h_i)[\Pi_{k=1}^{N} p(e^k \mid h_i)]))^{-1}$$

and $p(e^k \mid h_i) = p(e_l \mid h_i)$ for each k when the evidence e^k makes $p(e_l) = 1$.

The theorem can be proved by the following steps. The mathematical proof is given in Appendix A.

step 1: form an evidential mapping from a frame of discernment H to a frame of discernment E, the corresponding basic matrix BM is M in the Bayesian multi-valued causal link model and m_{ij} in BM is $P(e_j \mid h_i)$. The titles of rows from 1 to n are $\{h_1\}, ..., \{h_n\}$, and the titles of columns from 1 to m are $\{e_1\}, ..., \{e_m\}$.

step 2: a prior probability $p(h_i)$ is transformed into a basic probability assignment $m_0(h_i) = p(h_i)$ on H.

step 3: construct an evidential mapping from E to H through the evidential mapping from H to E, the corresponding basic matrix is BM' with $(j,i)'$-th entry as

$$p'(h_i \mid e_j) = \frac{p(e_j \mid h_i)}{\sum_{i=1}^{n} p(e_j \mid h_i)} \quad (7)$$

and the titles of rows from 1 to m are $\{e_1\},...,\{e_m\}$, and the titles of columns from 1 to n are $\{h_1\},...,\{h_n\}$.

step 4: for each probability distribution p_k on E provided by a piece of evidence e^k, calculate k-th mass function m_k on H using (2).

step 5: obtain the final belief function by using Dempster's combination rule to combine all those basic probability assignments because of independence of evidence.

Example 4 (from Pearl 1988, p.39): Let a hypothesis space have four propositions H=$\{h_1, h_2, h_3, h_4\}$ and an evidence space have three propositions E=$\{e_1, e_2, e_3\}$. The causal link matrix between H and E is

$$\begin{array}{cccc} & \{e_1\} & \{e_2\} & \{e_3\} \\ \{h_1\} & 0.5 & 0.4 & 0.1 \\ \{h_2\} & 0.06 & 0.5 & 0.44 \\ \{h_3\} & 0.5 & 0.1 & 0.4 \\ \{h_4\} & 1.0 & 0.0 & 0.0 \end{array} = BM$$

Assume prior probabilities for the hypotheses in H are a vector $p(h_i)$=[0.099 0.009 0.001 0.891], two pieces of evidence are e^1 providing $P(e_3) = 1$, and e^2 providing $P(e_1) = 1$.

* In Bayesian inference, applying formula (6), the posterior probabilities are

$$P(h_1 \mid e^1, e^2) = 0.919, \qquad P(h_2 \mid e^1, e^2) = 0.0439,$$
$$P(h_3 \mid e^1, e^2) = 0.0375, \qquad P(h_4 \mid e^1, e^2) = 0.0$$

* Now we use evidential mappings in D-S theory to analyze the example again. Given the above causal link matrix we can form a set of heuristic rules with its evidential mapping as:

R:

$h_1 \to e_1\ (0.5); h_1 \to e_2\ (0.4); h_1 \to e_3\ (0.1).$

$h_1 \to e_1\ (0.06); h_1 \to e_2\ (0.5); h_1 \to e_3\ (0.44).$

$h_1 \to e_1\ (0.5); h_1 \to e_2\ (0.1); h_1 \to e_3\ (0.4).$

$h_1 \to e_1\ (1.0); h_1 \to e_2\ (0.0); h_1 \to e_3\ (0.0).$

Γ^*:

$\Gamma^*(h_1) = \{(\{e_1\}, 0.5), (\{e_2\}, 0.4), (\{e_3\}, 0.1)\}$

$\Gamma^*(h_2) = \{(\{e_1\}, 0.06), (\{e_2\}, 0.5), (\{e_3\}, 0.44)\}$

$\Gamma^*(h_3) = \{(\{e_1\}, 0.5), (\{e_2\}, 0.1), (\{e_3\}, 0.4)\}$

$\Gamma^*(h_4) = \{(\{e_1\}, 1.0), (\{e_2\}, 0.0), (\{e_3\}, 0.0)\}$

Based on this evidential mapping to construct another evidential mapping from E to H using formula (7), we get

$$\Gamma'^*(e_1) = \{(\{h_1\}, 50/206), (\{h_2\}, 6/206), (\{h_3\}, 50/206), (\{h_4\}, 100/206)\}$$
$$\Gamma'^*(e_2) = \{(\{h_1\}, 0.4), (\{h_2\}, 0.5), (\{h_3\}, 0.1), (\{h_4\}, 0.0)\}$$
$$\Gamma'^*(e_3) = \{(\{h_1\}, 10/94), (\{h_2\}, 44/94), (\{h_3\}, 40/94), (\{h_4\}, 0.0)\}$$

and

$$\begin{array}{ccccc} & \{h_1\} & \{h_2\} & \{h_3\} & \{h_4\} \\ \{e_1\} & 50/206 & 6/206 & 50/206 & 100/206 \\ \{e_2\} & 0.4 & 0.5 & 0.1 & 0.0 \\ \{e_3\} & 10/94 & 44/94 & 40/94 & 0.0 \end{array} = BM'$$

Based on the prior probability distributions on H, the first mass function on H is obtained as:

$$m_0(h_1) = 0.099;\ \ m_0(h_2) = 0.009;\ \ m_0(h_3) = 0.001;\ \ m_0(h_4) = 0.891.$$

According to two pieces of evidence on E, the evidential mapping from E to H, and the Corollary 1, we get another two mass functions:

$$
\begin{aligned}
m_1(h_1) &= P(e_3) \times f(e_3 \to \{h_1\}) = 1 \times 10/94 = 10/94 \\
m_1(h_2) &= P(e_3) \times f(e_3 \to \{h_2\}) = 1 \times 44/94 = 44/94 \\
m_1(h_3) &= P(e_3) \times f(e_3 \to \{h_3\}) = 1 \times 40/94 = 40/94 \\
m_1(h_4) &= P(e_3) \times f(e_3 \to \{h_4\}) = 1 \times 0.0 = 0.0
\end{aligned}
$$

and

$$
\begin{aligned}
m_2(h_1) &= P(e_1) \times f(e_1 \to \{h_1\}) = 1 \times 50/206 = 50/206 \\
m_2(h_2) &= P(e_1) \times f(e_1 \to \{h_2\}) = 1 \times 6/206 = 6/206 \\
m_2(h_3) &= P(e_1) \times f(e_1 \to \{h_3\}) = 1 \times 50/206 = 50/206 \\
m_2(h_4) &= P(e_1) \times f(e_1 \to \{h_4\}) = 1 \times 100/206 = 100/206
\end{aligned}
$$

Combining these three mass functions using Dempster's rule of combination we eventually get $m(h_i) = m_0 \oplus m_1 \oplus m_2(h_i)$
that is

$$
\begin{aligned}
m(h_1) &= 0.099 \times 0.1 \times 0.5\alpha = \alpha P(h_1)[\Pi_{k=1,2} P(e^k \mid h_1)] \\
m(h_2) &= 0.009 \times 0.44 \times 0.06\alpha = \alpha P(h_2)[\Pi_{k=1,2} P(e^k \mid h_2)] \\
m(h_3) &= 0.001 \times 0.4 \times 0.5\alpha = \alpha P(h_3)[\Pi_{k=1,2} P(e^k \mid h_3)] \\
m(h_4) &= 0.891 \times 0.0 \times 1.0\alpha = \alpha P(h_4)[\Pi_{k=1,2} P(e^k \mid h_4)]
\end{aligned}
$$

where α is $(0.099 \times 0.1 \times 0.5 + 0.009 \times 0.44 \times 0.06 + 0.001 \times 0.4 \times 0.5 + 0)^{-1}$. The result of D-S theory is exactly the same as what we get in Bayesian inference.

4 Constructing Complete Evidential Mapping Matrices to Propagate Mass Functions From an Evidence Space Θ_E to a Hypothesis Space Θ_H

In Dempster-Shafer theory a multivalued mapping is used to calculate a mass function on a frame based on either a probability distribution or a mass function on another frame (Lowrance et al 1986, Zarley 1988, Laskey et al 1989). What we have assumed in the previous two sections is that a piece of evidence on an evidence space (a frame of discernment Θ_E) is represented in the form of Bayesian subjective probabilities. A mass function on Θ_H will be obtained based on the probability distribution on Θ_E through an evidential mapping from Θ_E to Θ_H.

In section 2, we gave the definition of evidential mappings. Let Θ_E and Θ_H be two frames of discernment, Γ^* be an evidential mapping from Θ_E to Θ_H. Assuming a piece of evidence indicates that $m(E) = p, m(\Theta_E) = 1 - p$, E is a subset of Θ_E, what is the impact of the evidence on Θ_H? Obviously the impact of the evidence on Θ_H is easyly obtained if Γ^* is a multivalued mapping. But it is not so easy when Γ^* is an evidential mapping.

In this section, we introduce the approach to constructing complete evidential mapping matrices between two frames. A complete evidential mapping matrix between two frames allows the propagation of a mass function from one frame to another.

Definition 3 *If Γ^* is an evidential mapping from Θ_E to Θ_H, BM is the basic matrix of Γ^* with m_{ij} as its (i,j)-th entry, the titles of rows of BM are $\{e_1\}$, ..., $\{e_n\} \subseteq \Theta_E$, the titles of columns of BM are A_1, ..., $A_m \subseteq \Theta_H$, then a matrix is called a* complete evidential mapping matrix *of BM, denoted as CEM, if it is defined as:*

1. all the subsets of Θ_E except $\emptyset$ are titles of rows of CEM and $\{e_1\}$, ..., $\{e_n\}$ are the first n row titles; A_1, ..., A_m are the titles of the first m columns of CEM.

2. the m_{ij} of BM is the value of (i,j)-th entry of CEM and denoted as m'_{ij}.

3. for a row l with the title E, and $l > n$, suppose $E = \{e_{l_1}, e_{l_2} \ldots, e_{l_k}\}$, then (l,j)-th entry of CEM is

$$m_{lj} = (m_{l_1 j} + m_{l_2 j} + \ldots + m_{l_k j})/k \qquad \text{if all } m_{l_i j} \neq 0 \quad \text{for } i = 1, \ldots k$$
$$m_{lj} = 0 \qquad \text{otherwise}$$

4. for $m_{lj} = 0$, create a column r with the title A_r

let $A_r = \cup_i \Theta_i$ for $i = l_1, \ldots, l_k$. (for Θ_i see definition 1 in 2.2)

if A_r is not an element of column title vector, then add A_r as an column title and the value of (l,r)-th entry is $m_{lr} = (m_{l_1 j} + m_{l_2 j} + \ldots + m_{l_k j})/k$. Otherwise there is a column r' satisfying $A_{r'} = A_r$ and we update $m_{lr'}$ as $m_{lr'} + (m_{l_1 j} + m_{l_2 j} + \ldots + m_{l_k j})/k$

5. for any other entry (x,y), define $m_{xy} = 0$.

Obviously we have the unequal formula

$$max(m_{l_1 j}, m_{l_2 j} \ldots, m_{l_k j}) \geq (m_{l_1 j} + m_{l_2 j} + \ldots + m_{l_k j})/k \geq min(m_{l_1 j}, m_{l_2 j}, \ldots, m_{l_k j}) \tag{8}$$

The basic idea of constructing a complete evidential mapping matrix is that if the causal links from $e_{l1}, e_{l2} \ldots, e_{lk}$ to A' are $m_{l_1 j}, m_{l_2 j} \ldots, m_{l_k j}$ respectively, then the causal link from $\{e_{l1}, e_{l2}, \ldots, e_{lk}\}$ to A' is something between $max(m_{l_1 j}, m_{l_2 j}, \ldots, m_{l_k j})$ and $min(m_{l_1 j}, m_{l_2 j}, \ldots, m_{l_k j})$. Here we use the average value of $m_{l_1 j}, m_{l_2 j}, \ldots, m_{l_k j}$ to represent approximate causal link from $\{e_{l1}, e_{l2} \ldots, e_{lk}\}$ to A'.

Example 5: Assume a BM of an evidential mapping from $\Theta_E = \{e_1, e_2, e_3\}$ to $\Theta_H = \{h_1, h_2, h_3, h_4\}$ is

	$\{h_1, h_2\}$	$\{h_3\}$	$\{h_4\}$	
$\{e_1\}$	0.5	0.5	0.0	$= BM$
$\{e_2\}$	0.7	0.0	0.3	
$\{e_3\}$	0.0	0.0	1.0	

with row title vector $[\{e_1\}, \{e_2\}, \{e_3\}]$, and column title vector $[\{h_1, h_2\}, \{h_3\}, \{h_4\}]$.

Then the corresponding CEM is

	$\{h_1, h_2\}$	$\{h_3\}$	$\{h_4\}$	Θ_E	$\{h_1, h_2, h_4\}$
$\{e_1\}$	0.5	0.5	0	0	0
$\{e_2\}$	0.7	0	0.3	0	0
$\{e_3\}$	0	0	1	0	0
$\{e_1, e_2\}$	$(0.5 + 0.7)/2$	0	0	$0.5/2 + 0.3/2$	0
$\{e_1, e_3\}$	0	0	0	$0.5/2 + 0.5/2 + 1/2$	0
$\{e_2, e_3\}$	0	0	$(0.3 + 1)/2$	0	$0.7/2$
Θ_E	0	0	0	$(0.5 + 0.7)/3 + 0.5/3 + (0.3 + 1)/3$	0

with row title vector $[\{e_1\}, \{e_2\}, \{e_3\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_2, e_3\}, \Theta_E]$ and column title vector $[\{h_1, h_2\}, \{h_3\}, \{h_4\}, \Theta_E, \{h_1, h_2, h_4\}]$.

Example 6: Assume a BM of an evidential mapping from Θ_E=$\{e_1, e_2, e_3\}$ to Θ_H=$\{h_1, h_2, h_3, h_4\}$ is a multivalued mapping as:

	$\{h_1, h_2\}$	$\{h_2, h_3, h_4\}$
$\{e_1\}$	1	0
$\{e_2\}$	1	0
$\{e_3\}$	0	1

Then the corresponding CEM is

	$\{h_1, h_2\}$	$\{h_2, h_3, h_4\}$	Θ_H
$\{e_1\}$	1	0	0
$\{e_2\}$	1	0	0
$\{e_3\}$	0	1	0
$\{e_1, e_2\}$	1	0	0
$\{e_1, e_3\}$	0	0	1
$\{e_2, e_3\}$	0	0	1
Θ_E	0	0	1

It is easy to prove that a CEM is a basic matrix of an evidential mapping from 2^{Θ_E} to Θ_H. So any piece of evidence, which is in the form of *bpa* on Θ_E can be propagated to Θ_H through the CEM. If a BM is the matrix of an multivalued mapping then its related CEM is also associated with the same multivalued mapping.

Certainly if a rule in a rule set specifies the causal link between a subset E of Θ_E and A_1, ..., A_n of Θ_H, then the values of row i, with the row title as E, are $(f(E \rightarrow A_1)$, ..., $f(E \rightarrow A_n))$ in CEM. But these $f(E \rightarrow A_i)$ must satisfy the condition of (8).

5 Propagating Beliefs Using Heuristic Knowledge

Belief propagation in a rule based system such as that described above indicates that, given belief functions on an antecedent frame and a set of rules with rule strengths in the form of mass functions, the belief functions on the conclusion frame can be deduced. If (R, Θ_E, Θ_H), (R', Θ_H, Θ'_H), (R", Θ'_E, Θ_H), (R_1, Θ_E, Θ_H) and (R_2, Θ_E, Θ_H) are five triples associated with five evidential mappings, generally we need to solve the following belief propagation problems:

(i). given a piece of evidence on Θ_E, (R, Θ_E, Θ_H) is known, to deduce belief on Θ_H.

(ii). given a piece of evidence on Θ_E, (R, Θ_E, Θ_H), and (R', Θ_H, Θ'_H) are known, to deduce belief on Θ'_H.

(iii). given two pieces of evidence on Θ_E and Θ'_E respectively, (R, Θ_E, Θ_H) and (R", Θ'_E, Θ_H) are known, to deduce belief on Θ_H.

(iv). given a piece of evidence on Θ_E, (R_1, Θ_E, Θ_H) and (R_2, Θ_E, Θ_H) are given independently, to deduce belief on Θ_H.

(v). given several pieces of evidence each of which is on A, B, ..., C respectively, (R, Θ_E, Θ_H) is known where $\Theta_E = A \times B \times ... \times C$, to deduce belief on Θ_H.

These problems can be solved by the following three theorems.

Theorem 2 *Let (R, Θ_E, Θ_H) be a triple associated with an evidential mapping Γ^*, BM and CEM are the basic matrix and the complete evidential mapping matrix of Γ^*, if a mass function m on Θ_E is known, then a mass function m_1 on Θ_H is calculated by the formula*

$$[m_1(H_1)...m_1(H_m)] = [m(E_1)...m(E_n)] \times \begin{bmatrix} m_{11} & ... & m_{1m} \\ & ... & \\ m_{n1} & ... & m_{nm} \end{bmatrix} (= CEM) \qquad (9)$$

where $[E_1, ..., E_n]$ is the row title vector of CEM, and $[H_1, ..., H_m]$ is the column title vector of CEM.

Specifically, if m is a Bayesian subjective probability assignment then m_1 on Θ_H is calculated by

$$[m_1(H_1)...m_1(H_m)] = [m(e_1)...m(e_n)] \times \begin{bmatrix} m_{11} & ... & m_{1m} \\ & ... & \\ m_{n1} & ... & m_{nm} \end{bmatrix} (= BM)$$

Example 7: Assume there are two rules in a rule set R:

$R_1 \;\; p \rightarrow \{q\}_{(c)};\ p \rightarrow \neg\{q\}_{(1-d)};\ p \rightarrow \Theta_{H(d-c)}.$
$R_2 \neg p \rightarrow \Theta_{H(1)}.$

and a piece of evidence says that $[m(p) \;\; m(\neg p) \;\; m(\Theta_E)]= [a \;\; 1-b \;\; b-a]$, where a, b, c, and d are all real numbers between $[0,1]$ with conditions $b > a$, and $d > c$, what is the belief distribution on q?

Obviously the CEM of this evidential mapping is

$$\begin{array}{cccc} & \{q\} & \{\neg q\} & \Theta_H \\ \{p\} & c & 1-d & d-c \\ \{\neg p\} & 0 & 0 & 1 \\ \Theta_E & 0 & 0 & 1 \end{array} = CEM$$

with row title vector $[\{p\}, \{\neg p\}, \Theta_E]$ and column title vector $[\{q\}, \{\neg q\}, \Theta_H]$.

Applying formula (9) to (R, Θ_E, Θ_H), we get

$$\begin{aligned} & [m_1(q) \;\; m_1(\neg q) \;\; m_1(\Theta_H)] \\ & = [m(p) \;\; m(\neg p) \;\; m(\Theta_E)] \times \begin{bmatrix} c & 1-d & d-c \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\ & = [ac \;\;\; a(1-d) \;\;\; 1-ac-a+ad] \end{aligned}$$

This result is the same as formula (12) by Hau and Kashyap (1990).

Theorem 3 *Let (R, Θ_E, Θ_H) and $(R', \Theta_H, \Theta'_H)$ be two triples associated with two evidential mappings Γ^* and Γ'^*, CEM_1 and CEM_2 are two complete evidential mapping matrices of Γ^* and Γ'^*, if a mass function m on Θ_E is known, then a mass function m_1 on Θ'_H is calculated by the formula*

$$[m_1(H'_1)...m_1(H'_m)] = [m(E_1)...m(E_n)] \times \begin{bmatrix} m_{11} & ... & m_{1m} \\ & ... & \\ m_{n1} & ... & m_{nm} \end{bmatrix}$$

where $CEM = CEM_1 \times CEM_2$.

Theorem 3 indicates that if we know a series of evidential mappings from Θ_{E_1} to Θ_{E_2}, ..., from $\Theta_{E_{n-1}}$ to Θ_{E_n} and those CEM_i of evidential mappings from Θ_{E_i} to $\Theta_{E_{i+1}}$ for $i = 1,...,n-1$, then we will get an evidential mapping from Θ_{E_1} to Θ_{E_n} with its CEM as $\text{CEM}_1 \times ... \times \text{CEM}_{n-1}$.

Example 8: Assume two rule sets R and R' are as following

$$\begin{aligned} R: \;\; & A \rightarrow B_{\;(a)};\;\; A \rightarrow \neg B_{\;(1-b)};\;\; A \rightarrow \Theta_{B\;(b-a)}. \\ & \neg A \rightarrow \Theta_{B\;(1)}. \\ R': \;\; & B \rightarrow C_{\;(c)};\;\; B \rightarrow \neg C_{\;(1-d)};\;\; B \rightarrow \Theta_{C\;(d-c)}. \\ & \neg B \rightarrow \Theta_{C\;(1)}. \end{aligned}$$

CEM_1 and CEM_2 are:

$$\begin{array}{cccccl} & B & \neg B & \Theta_B & \\ \\ A & a & 1-b & b-a & = CEM_1 \\ \neg A & 0 & 0 & 1 \\ \Theta_A & 0 & 0 & 1 \end{array}$$

and

$$\begin{array}{cccccl} & C & \neg C & \Theta_C & \\ \\ B & c & 1-d & d-c & = CEM_2 \\ \neg B & 0 & 0 & 1 \\ \Theta_B & 0 & 0 & 1 \end{array}$$

Given a mass function m on Θ_A then the mass function on Θ_C is

$$\begin{array}{l} [m_1(C) \;\; m_1(\neg C) \;\; m_1(\Theta_C)] \\ = [m(A) \;\; m(\neg A) \;\; m(\Theta_A)] \times \left[\begin{array}{lll} ac & a(1-d) & a(d-c)+1-a \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array} \right] (= CEM) \end{array}$$

This formula is identical to the formula (14) in Hau and Kashyap (1990) and the formula (4) in Ginsberg (1984).

Theorems 2 and 3 can be used to solve problems (i) and (ii). Dempster's combination rule is used to deal with the problem in (iii) where we suppose that any two pieces of evidence bearing on the same frame are DS-independent (Voorbraak1991). DS-independence will guarantee that if we use Dempster's rule to combine two probability distributions we should obtain the same result as what we get from Bayesian theory.

Theorem 4 *Let $(R_1, \Theta_E, \Theta_H)$ and $(R_2, \Theta_E, \Theta_H)$ be two triples associated with two evidential mappings Γ^* and Γ'^*, m_i and m'_i are two mass functions indicating causal links from e_i to Θ_H in Γ^* and Γ'^* respectively (for $i = 1, ..., n$), that is*

$\Gamma^*(e_i) = \{(H_{i1}, f(e_i \to H_{i1})), ..., (H_{in'}, f(e_i \to H_{in'}))\}$
$m_i(A_{il}) = f(e_i \to H_{il})$ *where* $A_{il} = \{(x,y) | x \in \neg\{e_i\}$ *or* $y \in H_{il}\}$ *for* $l = 1,\ ...,\ n'$
and
$\Gamma'^*(e_i) = \{(H'_{i1}, f(e_i \to H'_{i1})), ..., (H'_{in''}, f(e_i \to H'_{in''}))\}$
$m'_i(A'_{ir}) = f(e_i \to H'_{ir})$ *where* $A'_{ir} = \{(x,y) | x \in \neg\{e_i\}$ *or* $y \in H'_{ir}\}$ *for* $r = 1,\ ...,\ n''$

then the joint impact of two evidential mappings is the third evidential mapping Γ''^* from Θ_E to Θ_H in which:

$$\begin{array}{c} \Gamma''^*(e_i) = \{(H''_{i1}, f(e_i \to H''_{i1})), ..., (H''_{ik}, f(e_i \to H''_{ik}))\} \\ m_i(A''_{ij}) = f(e_i \to H''_{ij}) = m_i \oplus m'_i(A''_{ij}) \text{ for } j = 1, ..., k \\ \text{where} \quad A''_{ij} = A_{il} \cap A'_{ir}, \quad \text{and } H''_{ij} = H_{il} \cap H'_{ir}. \end{array} \qquad (10)$$

Here $\oplus$ indicates that Dempster's rule is used to combine m_i and m'_i.

Proof: Because $m_i \oplus m'_i$ is still a mass function in the Dempster-Shafer theory, we only need to prove that any focal element $A"_{ij}$ in $m_i \oplus m'_i$ is in the form of $\{(x,y)|x \in \neg\{e_i\}$ or $y \in H"_{ij}\}$.

Given that $A"_{ij} = A_{il} \cap A'_{ir}$, $A_{il} = \{(x,y)|x \in \neg\{e_i\}$ or $y \in H_{il}\}$ and $A'_{ir} = \{(x,y)|x \in \neg\{e_i\}$ or $y \in H'_{ir}\}$, we have

$$\begin{aligned}
A"_{ij} &= A_{il} \cap A'_{ir} \\
&= \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H_{il}\} \cap \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H'_{ir}\} \\
&= (\neg\{e_i\} \times \Theta_H \cup \Theta_E \times H_{il}) \cap (\neg\{e_i\} \times \Theta_H \cup \Theta_E \times H'_{ir}) \\
&= (\neg\{e_i\} \times \Theta_H \cup \{e_i\} \times H_{il}) \cap (\neg\{e_i\} \times \Theta_H \cup \{e_i\} \times H'_{ir}) \\
&= (\neg\{e_i\} \times \Theta_H) \cup (\{e_i\} \times H_{il} \cap \{e_i\} \times H'_{ir}) \\
&= (\neg\{e_i\} \times \Theta_H) \cup (\{e_i\} \times (H_{il} \cap H'_{ir})) \\
&= \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H_{il} \cap H'_{ir}\} \\
&= \{(x,y)|x \in \neg\{e_i\} \text{ or } y \in H"_{ij} = H_{il} \cap H'_{ir}\}
\end{aligned}$$

End.

The meaning of this theorem is that if there are two independent heuristic rule sets (as in figure 1) given by different domain experts respectively, each of those specifies one kind of causal link from frame Θ_E to frame Θ_H, then the joint impact of the two causal links can be substituted by the third heuristic rule set which is produced from them.

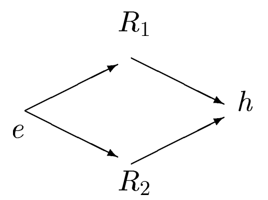

Figure 1: Two evidential mappings from e to h.

Here we need to address the issue that the meaning of this theorem is different from using theorem 2 twice through two evidential mappings. Using theorem 2 in that way gives a wrong result because of the dependency of the two mass functions on Θ_H.

The formula (6) given by Ginsberg (1984) is achieved from formula (10) in a special case when (R_1, Θ_E, Θ_H) and (R_2, Θ_E, Θ_H) are as follows:

R_1 $\quad e \to H_{(a)}$; $e \to \neg H_{(b)}$; $e \to \Theta_{H\,(1-b-a)}$.
$\quad \neg e \to \Theta_{H\,(1)}$.

and

R_2 $\quad e \to H_{(c)}$; $e \to \neg H_{(d)}$; $e \to \Theta_{H\,(1-c-d)}$.
$\quad \neg e \to \Theta_{H\,(1)}$.

Then the joint evidential mapping produces a new rule set as:

R:	$e \to H$	$(1 - \frac{(1-a)(1-c)}{1-(ad+bc)}) = x$;
	$e \to \neg H$	$(1 - \frac{(1-b)(1-d)}{1-(ad+bc)}) = y$;
	$e \to \Theta_H$	$(1 - x - y)$.
	$\neg e \to \Theta_H$	(1).

Here if $b = 0$ and $d = 0$ then it is identical with the result of parallel reduction given by Pearl (1990).

For problem (v), given a triple of an evidential mapping (R, Θ_E, Θ_H), $\Theta_E = A \times B \times ... \times C$, and a series of evidence in the form of mass functions on $A, B, ..., C$ respectively, we must first get the joint mass function on Θ_E in order to obtain the impact of those pieces of evidence on Θ_H. Shafer's partition theory and technique (Shafer 1976, Shafer, Shenoy and Mellouli 1987; Shafer and Logan 1987) provide a sound background for propagating mass functions (or belief functions) from $A, B, ...,$ C to their Cartesian product frame of discernment Θ_E. Thus applying theorem 2 the mass function on Θ_H will be calculated. Certainly computational complexity is a major problem which has been widely researched (Barnett 1981, Shafer and Logan 1987).

Generally speaking, when $\Theta_E = A \times B \times ... \times C$, we can solve problem (v) by following steps:

1) establishing evidential mappings $\Gamma_A, \Gamma_B, ..., \Gamma_C$ (in fact they are multivalued mappings) from $A, B, ..., C$ to Θ_E respectively.

$$\begin{aligned} &\Gamma_A(a_i) = \{(\{a_i\} \times B \times ... \times C, 1)\} \text{ for each } a_i \text{ in } A \\ &\Gamma_B(b_i) = \{(A \times \{b_i\} \times ... \times C, 1)\} \text{ for each } b_i \text{ in } B \\ &... \\ &\Gamma_C(c_i) = \{(A \times B \times ... \times \{c_i\}, 1)\} \text{ for each } c_i \text{ in } C \end{aligned} \tag{11}$$

2) given a series of pieces of evidence on $A, B, ..., C$, based on those evidential mappings, geting a number of mass functions on the joint frame from each simple frame.

3) suppose $A, B, ..., C$ are different from each other and the pieces of evidence are independent, using Dempster's combination rule to get the final mass function on Θ_E.

4) so based on this final mass function on Θ_E and an evidential mapping associated with (R, Θ_E, Θ_H), applying theorem 2 eventually to deduce a mass function on Θ_H.

Example 9: Assume we have four heuristic rules as follows. Given that p, q, v are certain, what is the degree of belief on c?

$$\begin{aligned} &r_1 : p \wedge q \rightarrow s \;\; (m_1); \\ &r_2 : s \wedge t \rightarrow c \;\; (m_2); \\ &r_3 : v \rightarrow u \;\; (m_3); \\ &r_4 : u \rightarrow c \;\; (m_4). \end{aligned}$$

Intuitively, the relations of the rules can be described as in figure 2. The degree of belief on c will be obtained through the following steps.

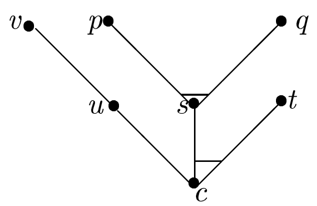

Figure 2: An AND/OR graph representing four heuristic rules

Step 1. construct a rule set for each rule above:

$$\begin{aligned} R_1 : \quad & p \wedge q \rightarrow s_{\ (m_1)}; \quad p \wedge q \rightarrow \Theta_{S\ (1-m_1)}. \\ & \neg p \wedge q \rightarrow \Theta_{S\ (1)}. \\ & p \wedge \neg q \rightarrow \Theta_{S\ (1)}. \\ & \neg p \wedge \neg q \rightarrow \Theta_{S\ (1)}. \\ R_2 : \quad & s \wedge t \rightarrow c_{\ (m_2)}; \quad s \wedge t \rightarrow \Theta_{C\ (1-m_2)}. \\ & \neg s \wedge t \rightarrow \Theta_{C\ (1)}. \\ & s \wedge \neg t \rightarrow \Theta_{C\ (1)}. \end{aligned}$$

$$\neg s \wedge \neg t \to \Theta_{C\,(1)}.$$
$$R_3: \quad v \to u_{\,(m_3)}; \quad v \to \Theta_{U\,(1-m_3)}.$$
$$\neg v \to \Theta_{U\,(1)}.$$
$$R_4: \quad u \to c_{\,(m_4)}; \quad u \to \Theta_{C\,(1-m_4)}.$$
$$\neg u \to \Theta_{C\,(1)}.$$

Step2: construct the corresponding basic matrices and complete evidential mapping matrices of those rule sets.

Step3: according to formula (11) for rule set R_1, because of p and q are certain, the joint mass function m_{pq} on Θ_{pq} is calculated:

$$m_{pq}(\{(p,q)\}) = 1.$$

(because of $m'_{pq}(\{p\} \times \Theta_q) = 1$ and $m"_{pq}(\Theta_p \times \{q\}) = 1$)

then based on theorem 2, calculate a mass function on Θ_s :

$$m_s(s) = m_1; \qquad m_s(\Theta_s) = 1 - m_1.$$

Step4: according to formula (11), for rule set R_2, because t is certain and the mass function on Θ_s is known in step3, then the joint mass function on Θ_{st} is:

$$m_{st}(\{(s,t)\}) = m_1; \qquad m_{st}(\Theta_s \times \{t\}) = 1 - m_1.$$

then based on theorem 2, a mass function on Θ_c is

$$m_c(c) = m_1 m_2; \qquad m_c(\Theta_c) = m_1(1-m_2) + 1 - m_1 = 1 - m_1 m_2.$$

Step5: using theorem 3 to rule sets R_3 and R_4, we obtain another mass function on Θ_C as

$$m'_c(c) = m_3 m_4; \qquad m'_c(\Theta_C) = 1 - m_3 m_4.$$

Step6: because m_c and m'_c are independent, we get a mass function on Θ_C by combining them

$$m_c"(c) = 1 - (1 - m_1 m_2)(1 - m_3 m_4); \qquad m_c"(\Theta_C) = (1 - m_1 m_2)(1 - m_3 m_4).$$

Eventually we have $Bel(c) = 1 - (1 - m_1 m_2)(1 - m_3 m_4)$.

In fact this example is a variation of Pearl's example (Pearl 1990, p.561). In r_3, we use the logical formula $v \to u$ with m_1 instead of $p \to u$ with m_1 in order to avoid the dependency problem. We achieve the same result as Pearl did using his own explanation (random-switch metaphor).

6 Conclusions

6.1 Related Work

Several approaches to dealing with heuristic knowledge in the Dempster-Shafer theory of evidence have been proposed (Ginsberg 1984, Liu 1986, Yen 1988, Hau and Kashyap 1990). Pearl also mentions this problem in Pearl (1990). The approach proposed in this paper is different from those approaches among which Liu's approach and Yen's approach are two proper subsets of our evidential mappings. In Ginsberg's as well as in Hau and Kashyap's representation formalisms of heuristic rules, a rule is associated with a pair of real numbers between [0,1] in the form of

if E then H with uncertainty $[c,d]$.

The meaning of c and d defined by Ginsberg is: c is the extent to which we believe a given proposition to be confirmed by the available evidence, and d is the extent to which it is disconfirmed.

That is: $E \to H$ with c and $E \to \neg H$ with d.

While Hau and Kashyap gave two explanations:

1. c is the credibility to which E supports H, d is the plausibility to which E supports H, so $1-d$ is the degree to which E supports $\neg H$.

That is $E \to H$ with c and $E \to H$ with $1-d$.

2. Let $\Theta = \Theta_E \oplus \Theta_H$ where Θ_E and Θ_H are frames of discernment of E and H, then c and d form a mass function on Θ

$$m(A) = c, m(\neg A) = 1-d, \ \ and \ m(\Theta) = d-c.$$
$$where \ \ A = \{(x,y) | x \in \neg E \ \ or \ \ y \in H\}$$

Obviously Ginsberg's representation can be incorporated into an evidential mapping from Θ_E to Θ_H by the rule set R below.

$$R: \quad E \to H_{\ (c)};\ E \to \neg H_{\ (d)};\ E \to \Theta_{H\ (1-d-c)}.$$
$$\neg E \to \Theta_{H\ (1)}.$$

Hau and Kashyap's first explanation can also be incorporated into an evidential mapping from Θ_E to Θ_H by the rule set R'

$$R': \quad E \to H_{\ (c)};\ E \to \neg H_{\ (1-d)};\ E \to \Theta_{H\ (d-c)}.$$
$$\neg E \to \Theta_{H\ (1)}.$$

In fact the second explanation given by Hau and Kashyap is to construct a mass function (furthermore a belief function) on a joint frame of discernment. Similar explanations of a rule are adopted by Laskey and Lehner (1989), by Lowrance et al (1986), and by Zarley et al (1988). This is also consistent with our explanation in section 2.2.

In section 5 we only discuss one situation involving the dependency problem. Hau and Kashyap (1990) discussed several situations based on their representation of heuristic rules.

6.2 Summary

Evidential mappings are the main concept proposed in this paper. The extended Dempster-Shafer theory is more powerful for propagating beliefs while at the same time keeping all the features of the original theory. The following are main features in our approach: 1) representing uncertain relations between evidence spaces and hypothesis spaces by evidential mappings; 2) by creating evidential mappings for incomplete heuristic rule sets, more heuristic knowledge can be represented in D-S theory; 3) by constructing complete evidential mapping matrices any piece of evidence bearing on an evidence space can be propagated to the corresponding hypothesis space; 4) when a set of heuristic rules is detailed enough to form a Bayesian multi-valued causal link model, any result produced by Bayesian inference can also be carried out by D-S theory under evidential mappings; 5) evidential mappings are consistent with other previous research work in this respect; 6) a series of belief propagation procedures are easily deduced based on evidential mappings.

Heuristic knowledge is important in knowledge based systems. Representing this kind of knowledge and propagating beliefs are the main and the most difficult tasks for designers of knowledge based systems. This paper makes some progress in this area. Future work concerning evidential mappings in the Dempster-Shafer theory should focus on exploring more features of evidential mappings and more approaches to dealing with dependency relations.

Appendix A:

Let E and H be two frames of discernment, Γ^* be a Bayesian evidential mapping from H to E, BM be the basic matrix of the mapping Γ^* with (i,j)-th entry as $p(e_j|h_i)$. Assume the prior probability on h_i of H is $p(h_i)$, a set of evidence on E is e^1, e^2, ..., e^N, for each of which there is an e_l where $p(e_l) = 1$. Then the final belief function Bel on H using D-S theory is

$$Bel(h_i) = \alpha p(h_i)\Pi_{k=1}^{N} p(e^k|h_i)$$

$$where \quad \alpha = (\sum_{1}^{n}(p(h_i)\Pi_{k=1}^{N} p(e^k|h_i)))^{-1}$$

and $p(e^k|h_i) = p(e_l|h_i)$ for each k when the evidence e^k makes $p(e_l) = 1$.

PROOF:

Step1: According to the conditions in the theorem, the given BM of the evidential mapping from H to E is:

	$\{e_1\}$	$\{e_2\}$	...	$\{e_m\}$	$\sum_j p(e_j\|h_i)$
$\{h_1\}$	$p(e_1\|h_1)$	$p(e_2\|h_1)$	...	$p(e_m\|h_1)$	$(=1)$
$\{h_2\}$	$p(e_1\|h_2)$	$p(e_2\|h_2)$	...	$p(e_m\|h_2)$	$(=1)$
...			...		
$\{h_n\}$	$p(e_1\|h_n)$	$p(e_2\|h_n)$	...	$p(e_m\|h_n)$	$(=1)$
$\sum_i p(e_j\|h_i)$	(β_1^{-1})	(β_2^{-1})		(β_m^{-1})	

where we suppose H contains n elements and E contains m elements. Then another matrix from E to H will be obtained as follows and it is also an Bayesian evidential mapping from E to H.

	$\{h_1\}$	$\{h_2\}$	...	$\{h_n\}$
$\{e_1\}$	$p(e_1\|h_1)\beta_1$	$p(e_1\|h_2)\beta_1$	...	$p(e_1\|h_n)\beta_1$
$\{e_2\}$	$p(e_2\|h_1)\beta_2$	$p(e_2\|h_2)\beta_2$	...	$p(e_2\|h_n)\beta_2$
			...	
$\{e_m\}$	$p(e_m\|h_1)\beta_m$	$p(e_m\|h_2)\beta_m$	...	$p(e_m\|h_n)\beta_m$

where $\quad \beta_j\ is\ (\sum_{i=1}^{n} p(e_j|h_i))^{-1}$ for $j = 1, ..., m$.

Step2: Assume we only get one piece of evidence e^1, and there exists an e_j that e^1 makes $p(e_j) = 1$. Applying Theorem 2 to $p(e_j) = 1$ and the Bayesian evidential mapping from E to H, a mass function m_1 on H is calculated as

$$m_1(h_i) = p(e_j|h_i)\beta_j \quad for\ i = 1, ..., n.$$

The prior probability gives another mass function on H

$$m_0(h_i) = p(h_i)$$

Applying Dempster's combination rule to m_0 and m_1, there will be

$$m_0 \oplus m_1(h_i) = \frac{(m_0(h_i)p(e_j|h_i)\beta_j)}{(\sum_{i=1}^{n} m_0(h_i)p(e_j|h_i)\beta_j)}$$

$$Let \quad \alpha_1 = (\sum_{i=1}^{n} m_0(h_i)p(e_j|h_i))^{-1}$$

Then $m'_1(h_i) = m_0 \oplus m_1(h_i)$
$= \alpha_1 m_0(h_i)p(e_j|h_i)$
$= \alpha_1 p(h_i)p(e_j|h_i)$ because of $p(h_i) = m_0(h_i)$
$= \alpha_1 p(h_i)p(e^1|h_i)$ we use e^1 instead of e_j because of e^1 making $p(e_j) = 1$
$= \alpha_1 p(h_i)\Pi_{k=1}^{1} p(e^k|h_i)$

So $Bel'_1(h_i) = \alpha_1 p(h_i)\Pi_{k=1}^{1} p(e^k|h_i)$, the theorem is true when $N = 1$.

Step3: Assume for $N-1$ pieces of evidence we have proved the theorem, that is

$$\begin{aligned} m'_{N-1}(h_j) &= m_0 \oplus m_1 \oplus ... \oplus m_{N-1}(h_i) \\ &= Bel'_{N-1}(h_i) \\ &= \alpha_{N-1} p(h_i)\Pi_{k=1}^{N-1} p(e^k|h_i) \end{aligned}$$

Spet4: Given n-th evidence e^N, e^N makes $p(e_r) = 1$ for a particular e_r, then m_N is a corresponding mass function produced from $p(e_r) = 1$ and the Bayesian evidential mapping. That is

$$m_N(h_i) = p(e_r|h_i)\beta_r \quad for \ \ i = 1, ..., n.$$

So

$$m'_N(h_i) = m_N \oplus m'_{N-1}(h_i)$$

$$= \frac{m_N(h_i)m'_{N-1}(h_i)}{(\sum_{i=1}^{n} m_N(h_i)m'_{N-1}(h_i))}$$

$$= \frac{\alpha_{N-1}p(h_i)\Pi_{k=1}^{N-1} p(e^k|h_i)p(e_r|h_i)\beta_r}{(\sum_{i=1}^{n}(\alpha_{N-1}p(h_i)\Pi_{k=1}^{N-1} p(e^k|h_i)p(e_r|h_i)\beta_r))}$$

$$= \alpha_N p(h_i)\Pi_{k=1}^{N} p(e^k|h_i) \qquad \text{by using } e^N \text{ instead of } e_r$$

where $\alpha_N = (\sum_1^n (p(h_i)\Pi_{k=1}^{N} p(e^k|h_i)))^{-1}$

Because $Bel(h_i) = m'_N(h_i)$. The theorem is true for N pieces of evidence.
END.

Acknowledgement

The research work was supported by "Information Systems Committee (of UFC)" when the first author was working in the Dept. of Information Systems, Univ. of Ulster. The authors are grateful for useful comments from Prof. Smets and Dr. Hong.

References

[1] Barnett, J.A. (1981) Computational Methods for a Mathematical Theory of Evidence. IJCAI-81, pp.868-875.

[2] Bonissone,P.P. and R.M.Tong (1985) Editorial: Reasoning with uncertainty in expert systems. Int. J. Man-Machine Studies pp. 241-250.

[3] Dempster,A.P.(1967) Upper and Lower Probabilities Induced by a Multivalued Mapping. Annals Mathematical Statistics, Vol.38, pp.325-339.

[4] Ginsberg, M. L.(1984) Non-monotonic Reasoning Using Dempster's Rule. AAAI-84, pp.126-129.

[5] Hau,H.Y. and R.L.Kashyap (1990) Belief Combination and propagation in a Lattice-Structured Inference Network. IEEE Transactions on Systems, Man and Cybernetics, Vol. 20, No. 1, Jan./Feb.,pp.45- 57.

[6] Jeffrey,R. (1965) The Logic of Decisions. New York: McGraw-Hill.

[7] Laskey,K.B., M.S.Cohen, and A.W Martin,(1989) Representing and Eliciting Knowledge about Uncertain Evidence and its Implications. IEEE Transactions on Systems, Man. and Cybernetics, Vol.19, No.3, pp.536-545

[8] Laskey, M.B. and Lehner,P.E. (1989) Assumptions, Beliefs and Probabilities. Artificial Intelligence 41 pp.65-77.

[9] Liu,G.S.H. (1986) Causal and Plausible Reasoning in Expert Systems. AAAI-86, pp.220-225.

[10] Lowrance,J.D., T. D. Garvey and T.M.Strat (1986) A Framework for Evidential-Reasoning Systems. AAAI-86, pp.896-903.

[11] Prade,H.(1981) Modal semantics and fuzzy set theory. in Yager R.R. (Ed.): Fuzzy Set and Possibility Theory: Recent Development, Pergamon press. pp.232-246.

[12] Pearl,J.(1988) Probabilistic Reasoning in Intelligence Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers,Inc..

[13] Pearl,J.(1990) Bayesian and Belief-Functions Formalisms for Evidential Reasoning: A Conceptual Analysis. in G.Shafer and J.Pearl (EDs.): Readings in Uncertain Reasoning, pp.540-574.

[14] Shafer,G.(1976) A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey.

[15] Shafer,G.(1981) Jeffrey's Rule of Conditioning. Philosophy of Science, 48 pp.337-362.

[16] Shafer,G.(1987) Probability Judgment in Artificial Intelligence and Expert Systems. Statistical Science, Vol.2 No.1, pp.3-44.

[17] Shafer,G.(1990) Perspectives on the Theory and Practice of Belief Functions. Working Paper No. 218, Univ. of Kansas.

[18] Shafer,G., P.P.Shenoy and K.Mellouli (1987) Propagating Belief Functions in Qualitative Markov Trees. International Journal of Approximate Reasoning; 1, pp.349-400.

[19] Shafer,G. and R. Logan (1987) Implementing Dempster's Rule for Hierarchical Evidence. Artificial Intelligence 33, pp. 271-298.

[20] Shafer,G. and R.Srivastava (1990) The Bayesian and Belief-Function Formalisms A General Perspective for Auditing. in G. Shafer and J. Pearl (Eds.): Readings in Uncertain Reasoning, pp.482-521.

[21] Shortliffe,E.H. and B.G.Buchanan (1976) A Model of Inexact Reasoning in Medicine. Chapter 12 of Rule-Based Expert Systems by Buchanan and Shortliffe. Addison-Wesley Publishing Company.

[22] Smets,P.(1988) Belief Functions. in P.Smets, E.H. Mamdani, D.Dubois and H.Prade (Eds.): Non-Standard Logics for Automated Reasoning, Academic Press, pp.213-252.

[23] Strat,L. (1987) Evidential-Based Control in Knowledge-Based Systems. PhD thesis, University of Massachusetts.

[24] Yen,J. (1989) Gertis: A Dempster-Shafer Approach to Diagnosing Hierarchical Hypotheses. Communications of the ACM May, Vol.32, pp.573-585.

[25] Wesley,L.P. (1988) Evidential-Based Control in Knowledge -Based Systems. PhD Thesis, University of Massachusetts, April.

[26] Zarley,D., Y.T.Hisa and G.Shafer (1988) Evidential Reasoning Using DELIEF. AAAI-88, pp.205-209.

[27] Voorbraak,F. (1991) On the justification of Dempster's rule of combination. Artificial Intelligence 48, 171-197.