An Algebraic Formalization of the Relationship between Evidential Structures and Data Tables *

Churn-Jung Liau
Institute of Information Science, Academia Sinica, Taipei, Taiwan
Fax: (+)886-2-7824814, E_mail: liaucj@iis.sinica.edu.tw

Abstract

In this paper, we would like to investigate the relationship between evidential structures (ES)—the basic qualitative structures of Dempster-Shafer theory, and the data table based knowledge representation systems(KRS) subject to rough set analysis. It is shown that an ES has a natural representation as a data table and from a given data table and two of its attributes, an ES can be extracted. We also show that some important operations on ES's can be realized in relational algebra. The results are then generalized to the fuzzy case. Consequently, we further clarify the connection between evidence theory and rough set theory.

1. Introduction

Theory of knowledge has been a commonly important topic of many academic branches such as philosophy, psychology and artificial intelligence, whereas the storage and retrieval of data is the main concern of information science. In the modern experimental science, knowledge is usually acquired from observed data. The data can provide the causal-effect or associational relationship between attributes of the observed objects. However, when the amount of data is large, it becomes a difficult task to analyze the data and extract knowledge from them. The rough set theory proposed by Pawlak provides an effective tool in extracting knowledge from data tables[7]. With the aid of computers, the large amount of data stored in relational data tables can be transformed into symbolic knowledge automatically.

On the other hand, the evidence theory proposed by the statistician Dempster and later developed further by Shafer[12](also known as Dempster-Shafer theory) provides a framework to represent numeric knowledge on a set of propositions derived from a probability distribution on the frame of evidence. In general, if there are two sets of exclusive and exhaustive propositions (called antecedent frame and consequent frame respectively) and given a compatibility relation between them, then the numeric measures on the antecedent frame can be transferred to the consequent one by the compatibility relation. Though many works on evidence theory have been focused on the theoretical properties and applications of the numeric measures, it should by no means be ignored that the qualitative compatibility relation between frames also plays an important role.

As pointed out by Skowron[13, 14], the common feature of rough set theory and evidence theory is the classification of objects. A frame of discernment in evidence theory

*The preliminary version of the paper has appeared as [5].

can be viewed as a partition of a set of objects according to some of their attributes, whereas the notions of partition and equivalence relation are just the central concept of rough set theory. Based on the key observation, Skowron suggests a method to compute the basic functions of evidence theory such as belief and plausibility measures from a given decision table that is a special kind of data tables. This shows that there exist some types of evidential information in data tables. Inspired by Skowron's result, the current paper will explore the relationship between the basic qualitative structures of evidence theory and the data table style knowledge representation systems.

In what follows, we will first review the basic notions of evidence theory and rough set theory. In section 3, it is shown that an evidential structure has a natural representation as a (two column) data table and from a given data table and any two attributes, an evidential structure can be derived. All frames derived from a given data table are then shown to form a lattice. In section 4, three important operations on evidential structures are considered. They can all be realized in relational algebra[16]. Next, we show how fuzzy decision tables are related to the fuzzification of evidential structures. Finally, some related works are discussed and a brief summary concludes the paper.

2. Preliminaries

2.1. Evidence Theory

Let Θ be a finite set, called a frame of discernment, then a function $m : 2^\Theta \to [0,1]$ is called a basic probability assignment (bpa) if (1) $m(\emptyset) = 0$ and (2) $\sum_{A \subseteq \Theta} m(A) = 1$. Intuitively, a frame of discernment denotes a set of exclusive answers to a given question and $m(A)$ is considered to be the measure of the belief that is committed *exactly* to A. Condition (1) means that no belief is committed to $\emptyset$ and is called closed world assumption[1]. Condition (2) denotes the convention of the measure of total belief being one. Note the belief that is committed *exactly* to A is different from the *total belief* committed to A. The latter is defined as

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

i.e., the summation of all belief committed exactly to the subsets of A. The function $Bel : 2^\Theta \to [0,1]$ so defined is called a *belief function* over Θ. A dual function, called *plausibility function,* is defined as $Pl : 2^\Theta \to [0,1]$ such that

$$Pl(A) = 1 - Bel(A^c)$$

where A^c is the complement set of A with respect to Θ.

According to the above definition, both belief and plausibility measures can be derived from the bpa. However, where does the bpa come from? We can imagine two frames of discernment Ω and Θ and a mapping $C : \Omega \to 2^\Theta$ satisfying the following two requirements:

1. $C(\omega) \neq \emptyset$, for all $\omega \in \Omega$,
2. $\bigcup_{\omega \in \Omega} C(\omega) = \Theta$.

[1] For the distinction of closed and open world assumptions, see [15].

Generally speaking, the frame Ω is called the antecedent frame and denotes a set of observable evidence and Θ is called the consequent frame and denote a set of hypotheses, whereas $\theta \in C(\omega)$ if the evidence ω is compatible with the hypothesis θ. Thus, C is called a *compatibility mapping* from Ω to Θ. It is easy to extend the domain of C to 2^{Ω} by letting $C(X) = \bigcup_{\omega \in X} C(\omega)$ for all $X \subseteq \Omega$, so we will not distinguish between the original C and its extension. Suppose an *a priori* probability distribution Pr exists on the frame Ω, then the function $m : 2^{\theta} \rightarrow [0.1]$ defined by

$$m(A) = Pr(\{\omega \mid C(\omega) = A\}$$

is a bpa on Θ.

Even when the *a priori* probability distribution is not available in the antecedent frame, the compatibility mapping still provides very valuable information about the support relation between the two frames. Thus we define an *evidential structure* (ES) as a triplet (Ω, Θ, C), where Ω and Θ are both finite sets and C is a compatibility mapping between them. When $\{C(\omega) \mid \omega \in \Omega\}$ forms a partition of Θ, Θ is called a refinement of Ω, Ω is called a coarsening of Θ, and C is a refining between these two frames. When C is a refining, we also call the whole ES a refining.

The support relation between evidence and hypotheses can then be summarized in the following basic set assignment(bsa) $m : 2^{\Theta} \rightarrow 2^{\Omega}$,

$$m(A) = C^{-1}(\{A\}) = \{\omega \mid C(\omega) = A\}.$$

The intuitive meaning of $m(A)$ is the set of evidence that supports exactly the proposition A. Naturally, we have the total support of A and a dual function, denoted by $\underline{C}$ and $\overline{C}$ respectively,

$$\underline{C}(A) = \bigcup_{B \subseteq A} m(B)$$

$$\overline{C}(A) = \Omega - \underline{C}(A^c).$$

Obviously, we have

$$\underline{C}(A) = \{\omega \mid C(\omega) \subseteq A\}$$

and

$$\overline{C}(A) = \{\omega \mid C(\omega) \cap A \neq \emptyset\}.$$

When C is a refining, $\underline{C}$ and $\overline{C}$ are called inner reduction and outer reduction respectively in [12]. We will see later that $\underline{C}$ and $\overline{C}$ have very close relationship with lower and upper approximations in rough set theory. However, before discussing the relationship, we will review rough set theory first.

2.2. Rough Set Theory

Let U be a finite set of objects (the universe) and R be an equivalence relation on U, then for any $X \subseteq U$, we can associate two subsets with X,

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$$

$$\overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\},$$

where $[x]_R$ denotes the equivalence class containing x. $\underline{R}X$ and $\overline{R}X$ are called the R-lower and R-upper approximation of X respectively. From a practical viewpoint, R can be considered as an indiscernibility relation, so for a given concept X, we can only know that X contains at least all elements in $\underline{R}X$ and does not contain any element outside $\overline{R}X$. The pair $(\underline{R}X, \overline{R}X)$ is called the rough approximation of X and any such pair is called a rough set.

Rough set theory is very useful in the analysis of data table based knowledge representation system (KRS). A KRS or data table is a pair $S = (U, A)$, where U is a nonempty, finite set (the universe) and A is a nonempty, finite set of primitive attributes. Every $a \in A$ is a total function $a : U \to V_a$, where V_a denotes possible values of a. If $B = \{a_1, \ldots, a_n\} \subseteq A$, then $V_B = V_{a_1} \times \cdots \times V_{a_n}$. Thus B is also considered as a function from U to V_B. An equivalence relation $IND(B)$ is associated with every subset of attributes $B \subseteq A$, and defined by

$$xIND(B)y \Leftrightarrow a(x) = a(y), \quad \forall a \in B.$$

$IND(B)$ is called an indiscernibility relation. We will write $IND(a)$ instead of $IND(\{a\})$ for all $a \in A$. Obviously, $IND(B) = \bigcap_{a \in B} IND(a)$. Since $IND(B)$ is an equivalence relation, we can define $IND(B)$-lower and $IND(B)$-upper approximation of X for any $X \subseteq U$. We will write $\underline{B}X$ and $\overline{B}X$ instead of $\underline{IND(B)}X$ or $\overline{IND(B)}X$.

The definitions are used in the analysis of dependency between attributes in a data table. Let us say that attribute B_2 depends on B_1, denoted by $B_1 \Rightarrow B_2$, iff $IND(B_1) \subseteq IND(B_2)$, i.e., any two objects in U with same values in their attributes B_1 will have also same ones in B_2. It is easily to show that $B_1 \Rightarrow B_2$ iff $\underline{B_1}X = X$ for all X that is an equivalence class of $IND(B_2)$.

From now on, we will use the following symbols (possibly with indices) with fixed meaning.

- $E = (\Omega, \Theta, C)$: an ES,
- ω, θ : an element of Ω and Θ respectively,
- $S = (U, A)$: a KRS,
- $X, Y, Z, \ldots$: subsets of Ω, Θ, or U, determined by the context,
- $x, y, z, \ldots$: elements of Ω, Θ, or U, determined by the context,
- B, D, F : subsets of A,
- U/B : the set of all equivalence classes under $IND(B)$,
- a, b, c : elements of A, sometimes denotes $\{a\}, \{b\}, \{c\}$
- V_a, V_b, V_c : domains of a, b, c respectively.

3. Correspondence Results

3.1. The table representation of evidential structures

It is quite simple to provide a data table style representation for a given $E = (\Omega, \Theta, C)$. Let us define $\delta(E) = (U, \{a, b\})$, where

$$U = \{(\omega, \theta) \mid \omega \in \Omega, \theta \in C(\omega)\}$$

and

$$a((\omega, \theta)) = \omega, b((\omega, \theta)) = \theta,$$

for all $(\omega, \theta) \in U$. For any $X \subseteq \Theta$ and $Y \subseteq \Omega$, let $\widehat{X}, \widehat{Y} \subseteq U$ denotes the following sets respectively,

$$\widehat{X} = \{(\omega, \theta) \in U \mid \theta \in X\},$$
$$\widehat{Y} = \{(\omega, \theta) \in U \mid \omega \in Y\}.$$

Then we have the following result.

Theorem 1 *Let $E = (\Omega, \Theta, C)$ be an ES and $\delta(E) = (U, \{a, b\})$, then for all $X \subseteq \Theta$ and $Y \subseteq \Omega$,*

$$\underline{a}\widehat{X} = \widehat{\underline{C}X}$$
$$\overline{a}\widehat{X} = \widehat{\overline{C}X}$$

Proof. We prove the first equation, and the second is analogous. On one hand, by definition,

$$\underline{C}X = Y = \{\omega \mid C(\omega) \subseteq X\},$$

so

$$\widehat{\underline{C}X} = \{(\omega, \theta) \in U \mid \omega \in Y\} = \{(\omega, \theta) \in U \mid C(\omega) \subseteq X\}.$$

On the other hand,

$$\underline{a}\widehat{X} = \{(\omega, \theta) \in U \mid [(\omega, \theta)]_a \subseteq \widehat{X}\} = \{(\omega, \theta) \in U \mid C(\omega) \subseteq X\},$$

so the result follows immediately. ■

3.2. Evidential Structures in a Data Table

Let $S = (U, A)$ be a KRS, and B_1 and B_2 be subsets of A, then we can define $\varepsilon(S, B_1, B_2) = (U/B_1, U/B_2, C)$, where $C(\omega) = \{\theta \in U/B_2 \mid \omega \cap \theta \neq \emptyset\}$ for all $\omega \in U/B_1$. It can be easily verified that C satisfies the two requirements of compatibility mapping.

Lemma 1
1. *If $S = (U, A)$ is a KRS, and B_1 and B_2 are subsets of A, then $\varepsilon(S, B_1, B_2)$ is an ES.*
2. *$\varepsilon(S, B_1, B_2)$ is a refining iff $IND(B_2) \subseteq IND(B_1)$ iff $B_2 \Rightarrow B_1$. In particular, $\varepsilon(S, B_1, A)$ is always a refining.*

Proof.

1. First, if $\omega \in U/B_1$, then there exists $u \in U$ such that $\omega = [u]_{B_1}$, so $[u]_{B_2} \in C(\omega)$ and $C(\omega) \neq \emptyset$. Second, by a symmetric argument, if $\theta \in U/B_2$, we can find an $u \in U$ such that $\theta = [u]_{B_2}$ and hence $\theta \in C([u]_{B_1})$, so $\bigcup_{\omega \in U/B_1} C(\omega) = U/B_2$.

2. By definition, $\varepsilon(S, B_1, B_2)$ is a refining iff

$$\forall u, v \in U, [u]_{B_1} \neq [v]_{B_1} \Longrightarrow C([u]_{B_1}) \cap C([v]_{B_1}) = \emptyset$$

iff

$$\forall u, v \in U, [u]_{B_1} \neq [v]_{B_1} \Longrightarrow [u]_{B_2} \neq [v]_{B_2}$$

iff $B_2 \Rightarrow B_1$.

■

Theorem 2 *Let $S = (U, A)$ be a KRS and $\varepsilon(S, B_1, B_2) = (U/B_1, U/B_2, C)$ be an ES corresponding to attributes B_1 and B_2, then for each $X \subseteq U/B_2$, we have*

$$\bigcup(\underline{C}X) = \underline{B_1}(\bigcup X)$$

and

$$\bigcup(\overline{C}X) = \overline{B_1}(\bigcup X).$$

Proof. We prove the first equation as follows.

$$\begin{aligned}
\bigcup(\underline{C}X) &= \bigcup\{[u]_{B_1} \mid C([u]_{B_1}) \subseteq X\} \\
&= \{u \in U \mid \forall v \in [u]_{B_1} \Rightarrow [v]_{B_2} \subseteq X\} \\
&= \{u \in U \mid [u]_{B_1} \subseteq \bigcup X\} \\
&= \underline{B_1}(\bigcup X)
\end{aligned}$$

The second equation follows dually. ■

Let us now use an example to illustrate the last theorem.

Example 1 The example is a simplification of that presented in [7](p.150). Let us first define the following abbreviation.

1. n = temperature is normal
2. s = temperature is subfeb.
3. h = temperature is high
4. c = dry-cough
5. a = headache
6. m = muscle pain,

and let $\overline{c}$, $\overline{a}$, and $\overline{m}$ denote the complement of c, a, and m respectively. Then the following is a symptom table of nine patients.

U	s_1	s_2	s_3	s_4
1	n	$\overline{c}$	$\overline{a}$	$\overline{m}$
2	n	$\overline{c}$	a	m
3	s	$\overline{c}$	a	m
4	s	c	$\overline{a}$	$\overline{m}$
5	s	c	$\overline{a}$	$\overline{m}$
6	h	$\overline{c}$	$\overline{a}$	$\overline{m}$
7	h	c	$\overline{a}$	$\overline{m}$
8	h	c	$\overline{a}$	$\overline{m}$
9	h	c	a	m

Now, if $B_1 = \{s_1, s_2\}$ and $B_2 = \{s_3, s_4\}$, then $U/B_1 = \{n\overline{c}, s\overline{c}, sc, h\overline{c}, hc\}$ and $U/B_2 = \{\overline{am}, am\}$, where

$$\begin{array}{ll}
n\overline{c} = \{1, 2\} & s\overline{c} = \{3\} \\
sc = \{4, 5\} & h\overline{c} = \{6\} \\
hc = \{7, 8, 9\} & \overline{am} = \{1, 4, 5, 6, 7, 8\} \\
am = \{2, 3, 9\}, &
\end{array}$$

and C is such that

$$C(\omega) = \begin{cases} \{\overline{am}\} & \text{if } \omega = sc \text{ or } h\overline{c}, \\ \{am\} & \text{if } \omega = s\overline{c}, \\ \{\overline{am}, am\} & \text{otherwise.} \end{cases}$$

Thus, if $X = \{\overline{am}\}$, then $\bigcup X = \overline{am}$, thus $\underline{B_1} \bigcup X = \{4, 5, 6\}$. On the other hand, $\underline{C}X = \{sc, h\overline{c}\}$, so $\bigcup(\underline{C}X) = \{4, 5, 6\}$, too. ■

If a data table $S = (U, A)$ satisfies $U/A = U$, then we call it *simple* data table. For a simple data table $S = (U, A)$ and its associated ES $\varepsilon(S, B, A) = (U/B, U, C)$, the theorem above is simplified to

$$\bigcup(\underline{C}X) = \underline{B_1}X$$

and

$$\bigcup(\overline{C}X) = \overline{B_1}X,$$

for any $X \subseteq U$. Thus, the B-lower and B-upper approximations of X correspond precisely to the inner and outer reductions of X under the refining C.

Let us now consider the collections

$$\mathcal{F} = \{U/B \mid B \subseteq A\}$$

and

$$\mathcal{C} = \{C \mid \varepsilon(S, B_1, B_2) = (U/B_1, U/B_2, C), B_2 \Rightarrow B_1\}$$

for a given KRS $S = (U, A)$.

Lemma 2 *The collections $\mathcal{F}$ and $\mathcal{C}$ for a given KRS satisfy the following four properties:*

(i) If $C_1 : 2^{\Theta_1} \to 2^{\Theta_2}$ and $C_2 : 2^{\Theta_2} \to 2^{\Theta_3}$ are in $\mathcal{C}$, then $C_2 \circ C_1$ is also in $\mathcal{C}$.

(ii) If $C_1 : 2^{\Theta} \to 2^{\Omega}$ and $C_2 : 2^{\Theta} \to 2^{\Omega}$ are in $\mathcal{C}$, then $C_1 = C_2$.

(iii) If $C_1 : 2^{\Theta_1} \to 2^{\Omega}$ and $C_2 : 2^{\Theta_2} \to 2^{\Omega}$ are in $\mathcal{C}$, and for each $\theta_1 \in \Theta_1$ there exists $\theta_2 \in \Theta_2$ such that $C_1(\theta_1) = C_2(\theta_2)$, then $\Theta_1 = \Theta_2$ and $C_1 = C_2$.

(iv) If $\Theta_1, \Theta_2 \in \mathcal{F}$, then there exists $\Omega \in \mathcal{F}$ such that $C_1 : 2^{\Theta_1} \to 2^{\Omega}$ and $C_2 : 2^{\Theta_2} \to 2^{\Omega}$ are in $\mathcal{C}$.

Proof.

(i) This follows from the transitivity of the dependency relation.

(ii) This holds because in the definition of $\varepsilon(S, B_1, B_2) = (U/B_1, U/B_2, C)$, C is uniquely defined by U/B_1(i.e. Θ) and U/B_2(i.e. Ω).

(iii) First, let $\Omega = U/B$, $\Theta_1 = U/B_1$, and $\Theta_2 = U/B_2$, then $B \Rightarrow B_1$ and $B \Rightarrow B_2$, so $\theta_i = \bigcup C_i(\theta_i)$ for all $\theta_i \in \Theta_i (i = 1, 2)$ by definition. This means $\Theta_1 \subseteq \Theta_2$ by the assumption. However, since each equivalence class is nonempty and $\bigcup \Theta_1 = U = \bigcup \Theta_2$, we have $\Theta_1 = \Theta_2$ and thus $C_1 = C_2$ by (ii).

(iv) If $\Theta_1 = U/B_1$ and $\Theta_2 = U/B_2$, then $\Omega = U/(B_1 \cup B_2)$ is the required frame.

■

In [12], $\mathcal{F}$ is called a family of compatible frames with refining $\mathcal{C}$ if it satisfies the above four properties and two additional requirements—existence of coarsenings and refinings. However, because our KRS is with finite attributes, it is impossible to have arbitrary coarsenings and refinings for frames in $\mathcal{F}$. In other words, the family of frames induced by a KRS is limited by its attributes. However, the four properties listed above are sufficient to prove the existence of the minimal common refinement between two frames in $\mathcal{F}$.

Theorem 3 *Suppose Θ_1, Θ_2 are elements of $\mathcal{F}$. Then there exists a unique element $\Omega \in \mathcal{F}$ such that*

(i) for $i = 1, 2$, $C_i : 2^{\Theta}_i \to 2^{\Omega} \in \mathcal{C}$, and

(ii) for each $\omega \in \Omega$, there exist elements $\theta_i \in \Theta_i (i = 1, 2)$ such that $\{\omega\} = C_1(\theta_1) \cap C_2(\theta_2)$.

Such Ω is denoted by $\Theta_1 \otimes \Theta_2$ and called minimal common refinement of Θ_1 and Θ_2 in $\mathcal{F}$.

Proof. This is a corollary of theorem 6.4 of [12] and the preceding lemma. ■

An alternative (and more natural) characterization of minimal common refinement is as follows.

Theorem 4 *If Ω is a common refinement of Θ_1 and Θ_2, then $\Theta_1 \otimes \Theta_2$ is a coarsening of Ω. Furthermore, $\Theta_1 \otimes \Theta_2$ is the unique element in $\mathcal{F}$ with such property.*

Proof. This is a corollary of theorem 6.5 of [12] and the preceding lemma. ■

However, unlike the definition of families of compatible frames in [12], we have also common coarsening for each pair of frames in $\mathcal{F}$.

Theorem 5 *Suppose Θ_1, Θ_2 are elements of $\mathcal{F}$. Then there exists a unique element $\Theta_1 \oplus \Theta_2 \in \mathcal{F}$ such that*

(i) $\Theta_1 \oplus \Theta_2$ is a common coarsening of Θ_1 and Θ_2, and

(ii) if Ω is a common coarsening of Θ_1 and Θ_2, then $\Theta_1 \oplus \Theta_2$ is a refinement of Ω

Proof. : If $\Theta_1 = U/B_1$ and $\Theta_2 = U/B_2$, then find greatest subsets E_1 and E_2 of A such that $B_i \Rightarrow E_i (i = 1, 2)$. Note that the greatest subsets exist because $B_i \Rightarrow E$ and $B_i \Rightarrow F$ imply $B_i \Rightarrow E \cup F$. Let $\Theta_1 \oplus \Theta_2 = U/(E_1 \cap E_2)$, then it satisfies requirement (i) because $B_i \Rightarrow (E_1 \cap E_2)$. To prove (ii), assume $\Omega = U/B$, then $B_i \Rightarrow B (i = 1, 2)$, so $B \subseteq (E_1 \cap E_2)$ by the definition of E_i, and the result follows. ■

As a corollary of the above-mentioned theorems and lattice theory[1], we have,

Corollary 1 *The structure $(\mathcal{F}, \oplus, \otimes)$ is a lattice with the finest element U/A and the coarsest element $U/\emptyset$.*

3.3. Translation Between ES and KRS

In the preceding two subsections, we have shown how a KRS can be constructed from a given ES and how an ES can be extracted from a KRS by given attributes. Now, we will show such constructions preserve the original information so that we can recover the original structures from the translated ones. First, let us define some notions.

Definition 1 *Two ES $E_1 = (\Omega_1, \Theta_1, C_1)$ and $E_2 = (\Omega_2, \Theta_2, C_2)$ are said to be isomorphic, denoted by $E_1 \cong E_2$, iff there exist bijective functions $f : \Omega_1 \to \Omega_2$ and $g : \Theta_1 \to \Theta_2$ such that for all $\omega \in \Omega_1$ and $\theta \in \Theta_1$, $\theta \in C_1(\omega)$ iff $g(\theta) \in C_2(f(\omega))$.*

Definition 2 *Two KRS $S_1 = (U_1, A_1)$ and $S_2 = (U_2, A_2)$ are said to be isomorphic, denoted by $S_1 \cong S_2$, iff there exist bijective functions $f : U_1 \to U_2$, $g : A_1 \to A_2$, and $h_a : V_a \to V_{g(a)}$ for all $a \in A_1$ such that for all $u \in U_1$ and $a \in A_1$, $h_a(a(u)) = g(a)(f(u))$*

Definition 3 *Let $S = (U, A)$ be a KRS and $B \subseteq A$, then the projection of S on B, denoted by $\pi_B(S)$, is the KRS $(U/B, B)$ such that for each $b \in B$ and $[u]_B \in U/B$, set $b([u]_B)$ in $\pi_B(S)$ as $b(u)$ in S.*

If $B = A$, then $\pi_B(S)$ just simplifies S into a simple data table by removing the duplicated rows.

Definition 4 *Let $S = (U, A)$ be a KRS and $B_i \subseteq A$ for $1 \leq i \leq n$, then the coding of S according to $B_1, \ldots, B_n$, denoted by $\gamma_{B_1,\ldots,B_n}(S)$, is the KRS $(U, \{b_i \mid 1 \leq i \leq n\})$ such that for each $u \in U$ and $1 \leq i \leq n$,*

$$b_i(u) = (a_{i1}(u), a_{i2}(u), \ldots, a_{ik_i}(u)),$$

where $B_i = \{a_{i1}, a_{i2}, \ldots, a_{ik_i}\}$.

If $A = \bigcup_{i=1}^{n} B_i$, then $\gamma_{B_1,\ldots,B_n}(S)$ is essentially equivalent to S.

According to these definitions, we have the following main correspondence results.

Theorem 6 *1. Let E be an ES, and $\delta(E) = (U, \{a, b\})$, then*

$$E \cong \varepsilon(\delta(E), \{a\}, \{b\}).$$

2. If S is a KRS and B_1, B_2 are its attributes, then

$$\delta(\varepsilon(S, B_1, B_2)) \cong \gamma_{B_1,B_2}(\pi_{B_1 \cup B_2}(S)).$$

Proof.

1. Let $E = (\Omega, \Theta, C_1)$ and the compatibility mapping of $\varepsilon(\delta(E), \{a\}, \{b\})$ be C_2, then define
$$f : \Omega \to U/a, \quad f(\omega) = \{(\omega, \theta) \mid \theta \in C_1(\omega)\}$$
and
$$g : \Theta \to U/b, \quad g(\theta) = \{(\omega, \theta) \mid \theta \in C_1(\omega)\}.$$
Obviously, f and g are bijective. Furthermore, $\theta \in C_1(\omega)$ iff $(\omega, \theta) \in f(\omega) \cap g(\theta)$ iff $f(\omega) \cap g(\theta) \neq \emptyset$ iff $g(\theta) \in C_2(f(\omega))$, so f and g are the isomorphisms.

2. If $S = (U, A)$, then

$$\delta(\varepsilon(S, B_1, B_2)) = (U_1, \{a, b\}),$$

where

$$U_1 = \{(\omega, \theta) \mid \omega \in U/B_1, \theta \in U/B_2, \omega \cap \theta \neq \emptyset\},$$

and

$$\gamma_{B_1,B_2}(\pi_{B_1 \cup B_2}(S)) = (U_2, \{b_1, b_2\}),$$

where $U_2 = U/(B_1 \cup B_2)$ and $b_i([u]_{B_1 \cup B_2}) = B_i(u)$ for all $u \in U$ and $i = 1, 2$. Now, define

$$f : U_1 \to U_2, \quad f((\omega, \theta)) = \omega \cap \theta,$$

$$g : a \mapsto b_1, b \mapsto b_2,$$

$$h_a : U/B_1 \to V_{B_1}, \quad h_a(\omega) = B_1(u), \quad \forall u \in \omega,$$

$$h_b : U/B_2 \to V_{B_2}, \quad h_b(\theta) = B_2(u), \quad \forall u \in \theta,$$

then these functions are all well-defined since $\omega \cap \theta \in U/(B_1 \cup B_2)$ for all $(\omega, \theta) \in U_1$ and for all $u \in \omega$(resp. θ), $B_1(u)$(resp. $B_2(u)$) is the same. Obviously, they are all bijective. Moreover,

$$h_a(a((\omega, \theta))) = h_a(\omega) = B_1(u), \forall u \in \omega,$$

$$g(a)(f((\omega, \theta))) = b_1(\omega \cap \theta) = B_1(u), \forall u \in \omega,$$

so $\delta(\varepsilon(S, B_1, B_2)) \cong \gamma_{B_1,B_2}(\pi_{B_1 \cup B_2}(S))$.

■

Example 2 Let us continue in using the table in Example 1 to illustrate the theorem. If S is the table in Example 1, then $\delta(\varepsilon(S, B_1, B_2))$ will be the following one,

U_1	a	b
$(n\overline{c}, \overline{am})$	$n\overline{c}$	$\overline{am}$
$(n\overline{c}, am)$	$n\overline{c}$	am
$(s\overline{c}, am)$	$s\overline{c}$	am
$(sc, \overline{am})$	sc	$\overline{am}$
$(h\overline{c}, \overline{am})$	$h\overline{c}$	$\overline{am}$
$(hc, \overline{am})$	hc	$\overline{am}$
(hc, am)	hc	am

,and $\gamma_{B_1,B_2}(\pi_{B_1 \cup B_2}(S))$ will be the following one.

U_2	b_1	b_2
$\{1\}$	$(n, \overline{c})$	$(\overline{a}, \overline{m})$
$\{2\}$	$(n, \overline{c})$	(a, m)
$\{3\}$	$(s, \overline{c})$	(a, m)
$\{4, 5\}$	(s, c)	$(\overline{a}, \overline{m})$
$\{6\}$	$(h, \overline{c})$	$(\overline{a}, \overline{m})$
$\{7, 8\}$	(h, c)	$(\overline{a}, \overline{m})$
$\{9\}$	(h, c)	(a, m)

Obviously, these two tables are isomorphic. ■

3.4. Remarks on Numeric Uncertainty

So far, we have discussed the relationship between qualitative evidential structures and data table style knowledge representation systems. However, there may be more things in a given data tables, that is, the numeric belief on any subsets of the universe U that one can induce from the data tables. In general, a subset of the universe can be considered as a proposition [2], so the numeric belief reflect the degree to which we believe the proposition is true according to the data in the tables.

The problem is when we can induce such numeric knowledge from a data tables. The answer of the problem will depend on how the data are collected into our KRS. More specifically, let us consider a KRS $S = (U, A)$ and two subsets of A, B_1 and B_2. We have seen from the preceding subsections that $\varepsilon(S, B_1, B_2)$ forms an ES with the antecedent frame U/B_1 and the consequent frame U/B_2, so if we have a probability distribution on U/B_1, then we can construct the numeric belief function on the subsets of the consequent frame according to Dempster-Shafer's theory.

However, how can we get the probability distribution on U/B_1? In general, if the objects in U are drawn evenly from a population independent of the values of attribute B_1, then we can assume the probability distribution $Pr : U/B_1 \to [0,1]$ such that $Pr(x) = \frac{|x|}{|U|}$ for each equivalence class $x \in U/B_1$. The assumption is implicitly made in [13, 14], so the numeric belief measure can be obtained there. However, the assumption of uniformity and independence does not always hold. For an extreme example, assume we would like to consider the relations between two attributes a and b of a fixed population, and the possible values of a are in $\{1, 2, \ldots, n\}$. Then the population may be partitioned into n classes (possibly not equal size) according to their a attributes. Now, if we draw m experimental cases from each class, observe their b attributes, and record all results in a KRS, then we get a data table with $m \cdot n$ objects. If we made the assumption of uniformity and independence, then for a fixed i, the probability that the a attribute of a randomly chosen object has value i is equal to $\frac{1}{n}$. However, it is obviously wrong, because we do not know the exact size of each a-equivalence class. Therefore, the proportion $\frac{|x|}{|U|}$ only reflects how the objects are selected into the data tables, not the proper distribution of the whole population according to some attributes.

In the case that the assumption of uniformity and independence does not hold or we do not know that it holds, what can still be obtained is the qualitative evidential structures (of course, with the closed world assumption). Thus, in this sense, an ES is the more conservative knowledge one can get from a data table when the numeric measures are not available. This is also the reason why we consider it is meaningful to divorce a qualitative ES from the belief measures imposed on it.

Another reason why we can not made the uniformity and independence assumption arbitrarily is that if we do, then we can get the probability distribution on the consequent frame, so the Dempster-Shafer's belief function is not needed in this case. If all objects of our data table is drawn evenly from a population, then we would have a Bayesian causal link model[8] between any two frames of the data table. Given a KRS $S = (U, A)$ and $B_1, B_2 \subseteq A$, then define $\beta(S, B_1, B_2) = (U/B_1, U/B_2, P)$, where

[2] We can even represent the propositions in a language, called decision language(DL), see Section 4.3 and [7]

$P : U/B_1 \times U/B_2 \to [0,1]$ is such that for $x \in U/B_1$ and $y \in U/B_2$,

$$P(y|x) = \frac{|x \cap y|}{|x|}.$$

So if we have a probability distribution Pr_{B_1} on U/B_1, then it is easy to compute

$$Pr_{B_2}(y) = \sum_{x \in U/B_1} Pr_{B_1}(x) \cdot P(y|x)$$

for all $y \in U/B_2$. Since Pr_{B_2} is more specific than the belief measure derived by Dempster-Shafer's theory, it is unnecessary to use the latter in this case.

4. Multiple Data Tables

In the preceding section, we have shown that an ES can be represented in a single (in fact, two-attribute) data table and from a data table, we can extract several evidential structures. However, sometimes, the data are collected from different sources and are represented by more than one data tables. In such case, we would like to see what evidential structures will be extractable from the joining table. On the other hand, in Dempster-Shafer theory, there are some operations on evidential structures, we are also interested in exploring the corresponding ones in data tables. In this section, we will discuss three main operations on evidential structures and their relationship with the join operation of data tables. We first introduce a variant of the translation ε.

Definition 5 *Let $S = (U, A)$ be a KRS, and B_1 and B_2 be subsets of A, then we can define $\tilde{\varepsilon}(S, B_1, B_2) = (B_1(U), B_2(U), C)$, where $C(\omega) = \{\theta \in B_2(U) \mid \exists u \in U, B_1(u) = \theta, B_2(u) = \omega\}$ for all $\omega \in B_1(U)$.*

In other words, $\tilde{\varepsilon}$ is same as ε except that $[u]_B$ is replaced by the value $B(u)$ for each u in the original universe and each subset of attributes B. It is obvious that $\tilde{\varepsilon}(S, B_1, B_2)$ is isomorphic to $\varepsilon(S, B_1, B_2)$

4.1. Composition

Let $E_1 = (\Theta_1, \Omega_1, C_1)$ and $E_2 = (\Omega_2, \Theta_2, C_2)$ be two evidential structures, then the composition of E_1 and E_2, denoted by $E_2 \circ E_1$, is defined as the ES, $(\Theta_1', \Theta_2', C_3)$, where

$$\Theta_1' = \{\theta_1 \mid C_2 \circ C_1(\theta_1) \neq \emptyset\},$$
$$\Theta_2' = \{\theta_2 \mid \exists \theta_1 \in \Theta_1', \theta_2 \in C_2 \circ C_1(\theta_1)\},$$

and C_3 is the restriction of $C_2 \circ C_1$ to $\Theta_1' \times \Theta_2'$.

As for the join of two data tables, it is defined analogously to that in relational algebra[16].

Definition 6 *Assume $S_1 = (U_1, A_1 \cup A)$ and $S_2 = (U_2, A_2 \cup A)$ are two data tables where A_1, A_2 and A are pairwisely disjoint and $A \neq \emptyset$, then the natural join of S_1 and S_2, denoted by $S_1 \bowtie S_2$, is the KRS $(U, A_1 \cup A_2 \cup A)$, where*

$$U = \{(x, y) \in U_1 \times U_2 \mid a_1(x) = a_2(y), \forall a \in A\}$$

and for all $a \in A_1 \cup A_2 \cup A$,

$$a((x,y)) = \begin{cases} a_1(x), & \text{if} a \in A_1, \\ a_2(y), & \text{otherwise.} \end{cases}$$

In the definition, we adopt the convention a_i to denote the attribute function of a in data table S_i for $i = 1, 2$.

According to the definitions, we have the following results.

Theorem 7 *1. Let $E_1 = (\Theta_1, \Omega, C_1)$ and $E_2 = (\Omega, \Theta_2, C_2)$ be two evidential structures and assume $\delta(E_1) = (U_1, \{a, b\})$ and $\delta(E_2) = (U_2, \{b, c\})$ such that $a \neq c$, then*

$$\delta(E_2 \circ E_1) \cong \pi_{a,c}(\delta(E_1) \bowtie \delta(E_2)).$$

2. Let S_1 and S_2 be two KRS's as defined in definition 6, then

$$\varepsilon(S_1 \bowtie S_2, A_1, A_2) \cong \tilde{\varepsilon}(S_1 \bowtie S_2, A_1, A_2) = \tilde{\varepsilon}(S_2, A, A_2) \circ \tilde{\varepsilon}(S_1, A_1, A),$$

Proof.

1. By definitions, $\pi_{a,c}(\delta(E_1) \bowtie \delta(E_2)) = (U, \{a, c\})$, where

$$U = \{[\theta_1, \omega, \theta_2]_{\{a,c\}} \mid \omega \in C_1(\theta_1), \theta_2 \in C_2(\omega)\},$$

and $\delta(E_2 \circ E_1) = (U', \{a, c\})$, where

$$U' = \{(\theta_1, \theta_2) \mid \exists\omega, \omega \in C_1(\theta_1), \theta_2 \in C_2(\omega)\}.$$

Let us define

$$f : U \to U', \quad f([\theta_1, \omega, \theta_2]_{\{a,c\}}) = (\theta_1, \theta_2),$$

and g, h_a, and h_c as identity mappings in respective domains. Then these functions construct the isomorphism immediately.

2. Define $C' : V_{A_1} \to 2^{V_{A_2}}$ as follows:

$$v_2 \in C'(v_1) \Leftrightarrow \exists x_1 \in U_1, x_2 \in U_2, A_1(x_1) = v_1 \wedge A_2(x_2) = v_2 \wedge A(x_1) = A(x_2)$$

Then by definition, both sides of the equality can be expanded into (V_1, V_2, C), where

$$V_1 = \{v_1 \in V_{A_1} \mid C'(v_1) \neq \emptyset\},$$
$$V_2 = \bigcup_{v_1 \in V_1} C'(v_1),$$

and C is the restriction of C' to V_1 and V_2.

■

4.2. Combination

One of the main components of Dempster-Shafer's theory is the so-called Dempster's combination rule. Here, because we are mainly interested in the evidential structures, we will first present a qualitative version of the Dempster's rule.

Definition 7 *If $E_1 = (\Theta_1, \Omega_1, C_1)$ and $E_2 = (\Theta_2, \Omega_2, C_2)$ are two evidential structures, and let $R \subseteq \Theta_1 \times \Theta_2$ be a binary relation (called dependency relation) on the two antecedent frames, then the combination of E_1 and E_2 according to R, denoted by $E_1 \oplus_R E_2$, is the ES (Θ, Ω, C), where $\Theta = \{(x, y) \in R \mid C_1(x) \cap C_2(y) \neq \emptyset\}$, $C((x, y)) = C_1(x) \cap C_2(y)$ for all $(x, y) \in \Theta$, and $\Omega = C(\Theta)$.*

When $R = \Theta_1 \times \Theta_2$, we say that E_1 and E_2 are independently combined. As a matter of fact, the original Dempster's rule is defined only for the independent case, so the definition given here is slightly more general than the original one.

Let us remark further on the dependency relation before proceeding. We first note that R is just an ordinary binary relation, not necessarily a compatibility relation. Second, we can abuse the notation R to denote a KRS $(R, \{a, b\})$ such that $a((x, y)) = x$ and $b((x, y)) = y$ for all $(x, y) \in R$. By such abuse, we can define the dependency degree of attribute a on b and vice verse. According to rough set theory, $\kappa_a(b) = \frac{|POS_a(b)|}{|R|}$ is the dependency degree of attribute b on a, where $POS_a(b) = \bigcup_{X \in R/b} \underline{a}X$, and $\kappa_b(a)$ is defined analogously. Then the dependency degree of R is $\max(\kappa_a(b), \kappa_b(a))$. It is easy to see that if R is independent, then its dependency degree is equal to zero.

Now, we have the following results.

Theorem 8 *1. Let $E_1 = (\Theta_1, \Omega, C_1)$ and $E_2 = (\Theta_2, \Omega, C_2)$ be two evidential structures, and let $R \subseteq \Theta_1 \times \Theta_2$ be a dependency relation. Then*

$$E_1 \oplus_R E_2 = \tilde{\varepsilon}(\delta(E_1) \bowtie R \bowtie \delta(E_2), \{a, b\}, c),$$

where (a, c), (a, b), and (b, c) are the attributes of $\delta(E_1)$, R, and $\delta(E_2)$ respectively.

2. Let $S_1 = (U_1, A_1 \cup B)$, $S_2 = (U_2, A_2 \cup B)$ and $R = (U_3, A_1 \cup A_2)$ be three KRS's such that A_1, A_2 and B are pairwisely disjoint, then

$$\tilde{\varepsilon}(S_1 \bowtie S_2 \bowtie R, A_1 \cup A_2, B) = \tilde{\varepsilon}(S_1, A_1, B) \oplus_{R'} \tilde{\varepsilon}(S_2, A_2, B),$$

where $R' = \{(A_1(u), A_2(u)) \mid u \in U_3\}$ is a subset of $V_{A_1} \times V_{A_2}$.

Proof.

1. By definition, $\delta(E_1) \bowtie R \bowtie \delta(E_2) = (U, \{a, b, c\})$, where

$$U = \{((\theta_1, \omega), (\theta_1, \theta_2), (\theta_2, \omega)) \mid (\theta_1, \theta_2) \in R, \omega \in C_1(\theta_1) \cap C_2(\theta_2)\}$$

$$a : U \to \Theta_1, \quad a(((\theta_1, \omega), (\theta_1, \theta_2), (\theta_2, \omega))) = \theta_1$$
$$b : U \to \Theta_2, \quad b(((\theta_1, \omega), (\theta_1, \theta_2), (\theta_2, \omega))) = \theta_2$$
$$c : U \to \Omega, \quad c(((\theta_1, \omega), (\theta_1, \theta_2), (\theta_2, \omega))) = \omega.$$

   Thus $\tilde{\varepsilon}(\delta(E_1) \bowtie R \bowtie \delta(E_2), \{a, b\}, c)$ can be expanded into $E_1 \oplus_R E_2$ exactly.

2. By definition, $S_1 \bowtie S_2 \bowtie R = (U, A_1 \cup A_2 \cup B)$, where

$$\begin{aligned} U &= \{((u_1, u_2), u_3) \mid u_i \in U_i (1 \le i \le 3), \\ &\quad A_1(u_1) = A_1(u_3), A_2(u_2) = A_2(u_3), B(u_1) = B(u_2)\} \end{aligned}$$

$$A_1(((u_1, u_2), u_3)) = A_1(u_3),$$
$$A_2(((u_1, u_2), u_3)) = A_2(u_3),$$
$$B(((u_1, u_2), u_3)) = B(u_1) = B(u_2).$$

   Thus

$$\tilde{\varepsilon}(S_1 \bowtie S_2 \bowtie R, A_1 \cup A_2, B) = (\Theta, \Omega, C),$$

where $\Theta = (A_1 \cup A_2)(U)$, $\Omega = B(U)$, and $\omega \in C(\theta)$ iff there exist $u_i \in U_i (1 \leq i \leq 3)$ such that $A_1(u_1) = A_1(u_3)$, $A_2(u_2) = A_2(u_3)$, $B(u_1) = B(u_2)$, $\theta = (A_1(u_3), A_2(u_3))$, and $\omega = B(u_1)$.

On the other hand, for $i = 1, 2$, $\tilde{\varepsilon}(S_i, A_i, B) = (\Theta_i, \Omega_i, C_i)$, where $\Theta_i = A_i(U_i)$, $\Omega_i = B(U_i)$, and $\omega_i \in C_i(\theta_i)$ iff there exist $u_i \in U_i$ such that $A_i(u_i) = \theta_i$ and $B(u_i) = \omega_i$. Then it can be shown that $\tilde{\varepsilon}(S_1, A_1, B) \oplus_{R'} \tilde{\varepsilon}(S_2, A_2, B) = (\Theta, \Omega, C)$ by the definition of combination operation.

■

When E_1 and E_2 are combined independently, the above theorem is just the qualitative counterpart of Theorem 4. in [14].

4.3. Conditioning

Let $E = (\Theta, \Omega, C)$ be an ES and X be a (proper) subset of Ω, then the conditioning of E on X, denoted by $E|X$, is the structure (Θ', X, C'), where $\Theta' = \{\theta \in \Theta \mid C(\theta) \cap X \neq \emptyset\}$ and $C'(\theta) = C(\theta) \cap X$ for all $\theta \in \Theta'$. It is obvious that $E|X$ is the combination of E and the special structure $(\{x, y\}, \Omega, \{x \mapsto X, y \mapsto \Omega \backslash X\})$ under the dependency relation $R = \Theta \times \{x\}$, so the join mechanism used in the last subsection can be used to simulate the conditioning operation.

However, we have an alternative characterization of conditioning by using the basic relational algebra operation—selection. To express selection operation, we need a Boolean language for expressing constrains. For the present purpose, we will use Pawlak's decision logic (DL) language[7]. In DL, the atomic expression is of the form (a, v), where a is an attribute constant and v is a value constant, and any formulas of DL is the Boolean combination of atomic expressions via $\neg$, $\vee$, and $\wedge$. We will use $a \in V$ to abbreviate $\bigvee_{v \in V}(a, v)$. If S is a KRS, and φ is a formula of DL, then $|\varphi|_S$ denotes all objects in the universe of S that satisfy the constraints expressed by φ. Using these definitions, we can define the selection of a KRS S under φ, denoted by $\sigma_\varphi(S)$, as $(|\varphi|_S, A)$ with the attribute A unchanging. Then, the following results can be shown.

Theorem 9 *1. Let $E = (\Theta, \Omega, C)$ be an ES and X be a (proper) subset of Ω. If $\delta(E) = (U, \{a, b\})$, then*

$$\delta(E|X) = \sigma_{b \in X}(\delta(E)).$$

2. Let $S = (U, A)$ be a KRS and φ be a formula of DL involving with only attribute constants in a proper subset B of A, then

$$\varepsilon(\sigma_\varphi(S), \overline{B}, B) \cong \varepsilon(S, \overline{B}, B)|X,$$

where $X = (|\varphi|_S / B)$ and $\overline{B} = A \backslash B$.

Proof.

1. First, by the definition of $E|X$, $\delta(E|X) = (U', \{a, b\})$, where $U' = \{(\theta, x) \mid x \in C(\theta) \cap X\}$. Second, $\sigma_{b \in X}(\delta(E)) = (|b \in X|_{\delta(E)}, \{a, b\})$. However, $|b \in X|_{\delta(E)} = \{u \in U \mid b(u) \in X\} = U'$ since $U = \{(\theta, x) \mid x \in C(\theta)\}$.

2. First, note that $X \subseteq U/B$ since φ only involves with the attributes in B. Thus, we have

$$\varepsilon(S, \overline{B}, B)|X = (\Theta_1, X, C_1),$$

where

$$\begin{aligned} \Theta_1 &= \{\theta_1 \in U/\overline{B} \mid \exists x \in X, x \cap \theta_1 \neq \emptyset\} \\ &= \{\theta_1 \in U/\overline{B} \mid \theta_1 \cap |\varphi|_S \neq \emptyset\}, \end{aligned}$$

and

$$\forall x \in X, \theta_1 \in \Theta_1, x \in C_1(\theta_1) \Leftrightarrow x \cap \theta_1 \neq \emptyset.$$

On the other hand,

$$\varepsilon(\sigma_\varphi(S), \overline{B}, B) = (\Theta_2, X, C_2),$$

where

$$\Theta_2 = \{\theta_1 \cap |\varphi|_S \mid \theta_1 \in U/\overline{B}, \theta_1 \cap |\varphi|_S \neq \emptyset\},$$

and

$$\forall x \in X, \theta_1 \in \Theta_1, x \in C_2(\theta_1 \cap |\varphi|_S) \Leftrightarrow x \cap \theta_1 \cap |\varphi|_S \neq \emptyset.$$

Therefore the mapping $f : \Theta \rightarrow \Theta_2$, $f(\theta_1) = \theta_1 \cap |\varphi|_S$ constructs the desired isomorphism.

■

4.4. Remarks

We have emphasized the analogy between evidential structures and data tables in the discussion above, however, we must also note some difference between these two representation formalisms. A careful reader may have found some asymmetry between the translation mappings δ and ε. The former is an one-argument mapping while the latter is a three-argument one. The main reason is that in an evidential structure, the compatibility mapping is directed from the antecedent frame to the consequent frame, whereas data tables are essentially not directional. Thus, to extract an ES from a data table, we must specify the direction of compatibility mapping. This also explain why composition and combination correspond to the same join mechanism in data tables, although they are so different operations in evidence theory. Moreover, due to the same reason, we can extract many different evidential structures from a data table, whereas an ES can only correspond to a KRS.

5. Fuzzification

In this section, we will consider a special kind of data tables, called *decision table.* A decision table is a triplet $S = (U, A, B)$, where U is a set of rule names and A and B are two sets of primitive attributes defined as in the KRS's, called condition attributes and action attributes respectively. A decision table is a special case of KRS's because $(U, A \cup B)$ is exactly a KRS. For each $u \in U$, we can write the corresponding rule as

$$\bigwedge_{a \in A} (a, a(u)) \supset \bigwedge_{b \in B} (b, b(u)),$$

in the language of DL. Intuitively, each rule says that if the condition attributes fulfill some values, then some particular actions specified in the action attributes will be adopted. Note that a decision table may be nondeterministic, that is, we may have two rules with the same values in their condition attributes but with different values in the action ones.

The analysis and optimization of decision tables has been treated from a rough set theory viewpoint in [7]. Here, we will consider the extension of decision tables to the fuzzy logic case[21]. Define a fuzzy decision table (FDT) as a quadruple (U, A, B, d) such that (U, A, B) is a decision table and $d : U \rightarrow (0, 1]$ is the degree function of the rules. As usual, $d(u)$ denotes the strength or the certainty of the rule u. Here, we assume that there are no duplicate rules in (U, A, B), i.e., each element of $U/(A \cup B)$ is a singleton. Because if there are more than two elements in any equivalence class of $U/(A \cup B)$, we can replace all of them by a new element with the degree being the maximum degrees of all the original elements in that class, the assumption in fact does not result in any loss of generality.

On the other hand, we can generalize the definition of evidential structures to the fuzzy case. Define a fuzzy evidential structures(FES) as a triple (Ω, Θ, C) just like an ES except that $C : \Omega \rightarrow \mathcal{F}(\Theta)$ is now a mapping form Ω to the collection of all fuzzy subsets of Θ and the second requirement must be modified as follows:
2. for all $\theta \in \Theta$, there exists $\omega \in \Omega$ such that $\mu_{C(\omega)}(\theta) > 0$.

Given an FES described as above, we can define the crisp support for any fuzzy subset of Θ. Let X be a fuzzy subset of Θ, then

$$\underline{C_c}X = \{\omega \in \Omega \mid C(\omega) \subseteq X\},$$

and

$$\overline{C_c}X = \Omega \backslash \underline{C_c}X^c,$$

where "$\subseteq$" denotes the fuzzy set inclusion relation (i.e. $X \subseteq Y$ iff $\mu_X(\theta) \leq \mu_Y(\theta)$ for all $\theta \in \Theta$) and X^c is the fuzzy complement of X w.r.t. Θ.

The fuzzy support can be defined by using the degrees of inclusion and consistency between fuzzy subsets[19]. Let $X, Y \in \mathcal{F}(\Theta)$, then define

$$Con(X, Y) = \sup_{\theta \in \Theta} \min(\mu_X(\theta), \mu_Y(\theta)),$$

and

$$Inc(X, Y) = 1 - Con(X, Y^c).$$

Then we can define $\underline{C_f}X$ and $\overline{C_f}X$ as fuzzy subsets of Ω with the following membership functions,

$$\mu_{\underline{C_f}X}(\omega) = Inc(C(\omega), X)$$

and

$$\mu_{\overline{C_f}X}(\omega) = Con(C(\omega), X).$$

Obviously, from these definitions, we have $\underline{C_f}X = (\overline{C_f}X^c)^c$. Given the fuzzy support, we can compute the support degrees of any fuzzy subsets of the antecedent frame on those of the consequent frame. Let $X \in \mathcal{F}(\Omega)$ and $Y \in \mathcal{F}(\Theta)$, then

$$Sp(X, Y) = Inc(X, \underline{C_f}Y)$$

is the support degree of X on Y.

By exploiting the techniques presented in this paper, we can extract an FES from an FDT, so we can discover the support relation (either crisp or fuzzy) between condition and action attributes. In fact, we can transform an FES into an FDT and vice verse. For a given FES $E = (\Omega, \Theta, C)$, define $U = \{(x, y) \in \Omega \times \Theta \mid \mu_{C(x)}(y) > 0\}$, $A = \{a\}$, $B = \{b\}$, and for all $(x, y) \in U$, $a((x, y)) = x$, $b((x, y)) = y$, and $d((x, y)) = \mu_{C(x)}(y)$. Then $\delta(E) = (U, A, B, d)$ is an FDT. Conversely, given an FDT $S = (U, A, B, d)$, we can define $\Omega = U/A$, $\Theta = U/B$, and for all $x \in \Omega$, define $C(x)$ is the fuzzy subset of Θ with the membership function

$$\mu_{C(x)}(y) = \begin{cases} d(x \cap y), & \text{if} x \cap y \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

Then, $\varepsilon(S) = (\Omega, \Theta, C)$ is an FES.

It is a straightforward work to extend the definitions of isomorphism and coding to FES's and FDTs. Then, we have the correspondence result for the fuzzy case.

Theorem 10 *1. Given an FDT S, then $\gamma_{A,B} S \cong \delta(\varepsilon(S))$.*

2. Given an FES E, then $E \cong \varepsilon(\delta(E))$.

Proof. The proof is analogous to that for theorem 6. ■

6. Related Works and Conclusion

As indicated in Section 1, the current work is directly inspired by those in [13, 14]. We have also remarked in Section about the difference between their works and ours. However, it is fair to say that the structural similarity between qualitative ES's and KRS's revealed here has been implied by their works. Here, we just give an algebraic formalization for it.

The second influential work along the same direction we know is by Wong et al.[18]. They term $\underline{C}X$ and $\overline{C}X$ for an ES as the lower preimage and upper preimage of X respectively and show that the two preimages satisfy the basic properties of rough sets(p. 142, [18]). Their result can then be seen as a natural corollary of our Theorem 1 and Pawlak's rough set theory.

As for the fuzzy case, the fuzzy Dempster-Shafer theory and fuzzy rough sets have been proposed in [20] and [3] respectively. According to our results here, we can say that fuzzy rough sets are a special case of FES's. More specifically, let (Ω, Ω, C) be an FES and if $\{C(\omega) \mid \omega \in \Omega\}$ forms a fuzzy partition, then $(\underline{C_f}X, \overline{C_f}X)$ is a fuzzy rough set in the sense of [3].

Finally, there are some connection among modal logic[2], evidence theory and rough set theory. On the one hand, we can represent a Kripke model in the modal system **T** as an ES. A Kripke model for **T** is a triple (W, R, V) where W is a set of possible worlds, $R \subseteq W \times W$ is a reflexive binary relation on W, and V assigns a subset of W to each formula in the modal language. Let $|\varphi|$ denote the set of possible worlds that satisfy the formula $|\varphi|$, then according to Kripke semantics,

$$|\Box\varphi| = \{w \in W \mid \forall u \in W(R(w, u) \supset u \in |\varphi|)\}$$

and

$$|\Diamond\varphi| = \{w \in W \mid \exists u \in W(R(w, u) \wedge u \in |\varphi|)\}.$$

If we define the ES (W, W, C) with C being such that $C(w) = \{u \in W \mid R(w, u)\}$, then it is obvious that

$$|\Box\varphi| = \underline{C}|\varphi|$$

and

$$|\Diamond\varphi| = \overline{C}|\varphi|$$

hold for each formula φ. This is essentially the viewpoint taken in [11].

On the other hand, given a KRS (U, A), U can be seen as a set of possible worlds and for each $B \subseteq A$, $(U, IND(B))$ define an epistemic structure(i.e. an **S5** Kripke structure). Based on this viewpoint, each frame $U/a (a \in A)$ induced from the KRS may be understand as the epistemic structure of some agent. The common refinement of two frames is then interpreted as the strong distributed knowledge structure of two groups of agents in the sense of [4]. This viewpoint is taken in [9, 10] and Corollary 1 are also derived there in the context of epistemic structures.

In summary, the relationship between evidence theory and modal logics are suggested in [11] and the one between rough set theory and modal logics are described in [9, 10, 6, 17], while the works reported here (and of course the works on which our results depend) provide a strong connection between evidence theory and rough set theory. Although there are some overlap between our presentation and the previous works, we hope that the framework introduced in this paper clarify further the connection between these theories.

References

[1] G. Birkhoff and S. MacLane. *A Survey of Modern Algebra.* Macmillan Publishing Co., 1977.

[2] B.F. Chellas. *Modal Logic : An Introduction.* Cambridge University Press, 1980.

[3] D. Dubois and H. Prade. "Rough fuzzy sets and fuzzy rough sets". *International Journal of General Systems*, 17:191–200, 1990.

[4] J.Y. Halpern and Y. Moses. "Knowledge and common knowledge in a distributed environment". *JACM*, 37(3):549–587, 1990.

[5] C.J. Liau. "On the relationship between evidential structures and data tables". In M. De Glas and Z. Pawlak, editors, *Proceedings of the 2nd World Conference on the Fundamentals of Artificial Intelligence*, pages 205–216. Angkor, France, 1995.

[6] E. Orlowska. "Logic for reasoning about knowledge". *Zeitschrift f. Math. Logik und Grundlagen der Math*, 35:559–572, 1989.

[7] Z. Pawlak. *Rough Sets–Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, 1991.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference.* Morgan Kaufmann Publisher, 1988.

[9] C.M. Rauszer. "Approximation methods for knowledge representation systems". In J.Komorowski and Z. W. Raś, editors, *Proceedings of the 7th ISMIS*, LNAI 689, pages 326–337. Springer-Verlag, 1993.

[10] C.M. Rauszer. "Rough logic for multi-agent systems". In M. Masuch and L. Pólos, editors, *Knowledge Representation and Reasoning under Uncertainty*, LNAI 808, pages 161–181. Springer-Verlag, 1994.

[11] G. Resconi, G.J. Klir, and U.ST. Clair. "Hierarchical uncertainty metatheory based upon modal logic". *International Journal of General Systems*, 21:23–50, 1992.

[12] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[13] A. Skowron. "Boolean reasoning for decision rules generation". In J.Komorowski and Z. W. Raś, editors, *Proceedings of the 7th ISMIS*, LNAI 689, pages 295–305. Springer-Verlag, 1993.

[14] A. Skowron and J. W. Grzymala-Busse. "From rough set theory to evidence theory". In R. R. Yager, M. Federizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 193–236. John Wiley and Sons, 1994.

[15] P. Smets. "Belief functions". In P. Smets, A. Mamdani, D. Dubois, and H. Prade, editors, *Non-Standard Logics for Automated Reasoning*, pages 253–286. Academic Press, 1988.

[16] J. D. Ullman. *Principles of Database Systems.* Pitman Publishing Ltd., 1982.

[17] D. Vakarelov. "Modal logics for knowledge representation". In A.R. Meyer and M.A. Taitslin, editors, *Proceedings of Symposium on Logical Foundations of Computer Science*, LNCS 363, pages 257–277. Springer-Verlag, 1989.

[18] S.K.M. Wong, Y.Y. Yao, and L.S. Wang. "An application of rough sets in knowledge synthesis". In K. S. Harber, editor, *Proceedings of the 7th ISMIS(Poster Session)*, ORNL/TM-12375, pages 140–150. Center for Engineering Systems Advanced Research, Oak Ridge National Laboratory, 1993.

[19] R. R. Yager. "Using approximate reasoning to represent default knowledge ". *Artificial Intelligence*, 31:99–112, 1987.

[20] J. Yen. "Generalizing the Dempster-Shafer theory to fuzzy sets". *IEEE Transactions on Systems, Man, and Cybernetics*, 20(3):559–570, 1990.

[21] L. A. Zadeh. "Fuzzy set ". *Information and Control*, 8:338–353, 1965.