Aggregating Uncertain Access Risk Estimations from Different Sources
Invited Paper

Qun Ni* and Elisa Bertino*
*Department of Computer Science, Purdue University, West Lafayette, IN 47907
Email: [ni,bertino]@cs.purdue.edu

Abstract

Risk-based access control raises some novel problems that have not yet been investigated. In particular, the ability to aggregate uncertain risk estimations from different experts is crucial to the success of risk-based access control systems. A K-Algebra is proposed in this paper for this purpose. The algebra is not only able to support the specification of existing aggregation rules but also makes it possible to generalize these rules based on their logical explanations.

1. Introduction

The inflexibility of binary decisions from current access control systems is inadequate for novel dynamic application environments, which is the reason why risk-based access control has been proposed [1]. Unlike traditional access control systems in which all risky accesses are prohibited, risk-based access control systems permit low risky accesses if some mitigating actions based on their risk estimations, e.g. obligations, have been, are or will be executed. Therefore, risk-based access control systems may improve the overall information flow efficiency and are particularly useful in emergency or a crisis situations [1]. The recently proposed Fuzzy Multi-Level Security (Fuzzy MLS) system [2] is an example of risk-based access control systems.

In risk-based access control systems, the risk of an access is evaluated as the possible losses resulting from the access, e.g. the value of the content, multiplied by the probability or confidence degree that the subject may leak the content [2]. Such a risk estimation is in accordance with standard risk estimation methods adopted in the engineering field.

However, the distinct properties of access control systems result in some novel problems in risk estimation. The goal of access control is to prevent possible losses in the future, which is hard to predict. The sensitive nature of access control systems and the involvement of human activities make this situation even worse. As a result, we lack sufficient information or knowledge to devise *objective* equations for computing the value of the protected data and the probability that the subject will disclose the data. By *objective*, we mean that these equations are built based on well established theories or sufficient statistical results.

By contrast, we have to rely on subjective equations devised by human experts to estimate the risk of an access, including the value of data (loss) and the likelihood that the data are disclosed (probability). Such subjective estimations, unsurprisingly, contain uncertainties or vagueness. More specifically, equations from different experts may generate different or conflicting estimations concerning the loss and likelihood. Therefore, the ability to aggregate uncertain risk estimations from different sources becomes crucial to the success of risk-based access control systems, which is the focus of this paper. To address such issues, in this paper,

- we propose a K-Algebra that focuses on aggregating risk estimations from different sources;
- we suggest criteria for choosing appropriate operations for the aggregation of risk estimations;
- we show that by using the K-Algebra we can specify existing combination rules for risk estimations and present novel logic-based explanations for these rules;
- we propose a methodology to generalize existing combination rules based on the K-Algebra.

The rest of this paper is organized as follows. Section 2 introduces some running examples that we will use to illustrate the discussion. Section 3 formalizes the problem of estimating risks. Section 4 introduces the K-Algebra for the aggregation of risk estimations and discusses in detail its operations. Section 5 shows that the K-Algebra can be used to specify existing combination rules. Section 6 presents a method to generalize existing rules. Section 7 concludes this

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

paper and suggests a few future research directions.

2. Running Examples

In this section, we present two examples that motivate the operations of the K-Algebra and demonstrate the aggregation rules. These two examples are based on some examples from [3]. However, we extend them to cover more interesting cases. In the rest of the paper, we use the terms 'combination rules' and 'aggregation rules' as synonyms.

In the first example, the possible losses resulting from an access could be high (H), medium (M), or low (L). Suppose that we have two experts who estimate access risks. Given an access, the first expert believes that the loss is high with confidence degree 0.8 or medium with confidence degree 0.1. The likelihood that the expert does not know the exact loss due to insufficient information is very low, say 0.1. By contrast, the second expert believes that the loss is medium with confidence degree 0.1 or low with confidence degree 0.7. The likelihood that the second expert does not know the exact loss is 0.2. The question is "what are the final risk estimations based on the aggregation of these two experts' estimations?"

In the first example, we use ambiguous terms to describe the losses. In practice, however, interval-based data is commonly used to quantify uncertainty in the estimation of losses. The second example represents losses using intervals. Like in the first example, suppose that we have two experts. Given an access, the first expert believes that the possible loss is in the range [1000, 4000] with confidence degree 0.5, in the range [3000, 5000] with confidence degree 0.4, or in the range [1000, 6000] with confidence degree 0.1. The second expert believes that the possible loss is in the rage [1000, 2000] with confidence degree 0.1, in the range [2000, 5000] with confidence degree 0.3, in the range [3000, 6000] with confidence degree 0.4, or in the range [1000, 6000] with confidence degree 0.2. A similar question is "what are the final risk estimation based on these two experts?"

3. Problem Formalization

In the engineering field, the definition of the risk of an event, Risk(e), is often simply calculated as

$$\text{Risk}(e) = \text{Probability}(e) \times \text{Losses}(e)$$

Therefore, without loss of generality, we introduce the following definition for risk estimations in access control.

Definition 1: Let A be the universal set of losses in access control, a risk estimation m is a function $\mathcal{P}(A) \rightarrow [0, 1]$ where $\mathcal{P}(A)$ is the powerset of A. Let $S = \{x | m(x) > 0, x \subseteq A\}$, S is called the focal set of the risk estimation m.

Examples of A might be a set of predefined states or data intervals. For instance, A = {high, medium, low} in example one or $A = \{x | x \in [1000, 6000]\}$ in example two. S contains some subsets of A that represent the interesting losses in a particular estimation. The interpretation of m depends on the applications. $m(x)$ may refer to the likelihood, the subjective probability, or the confidence degree of the corresponding loss x. We use risk estimation or confidence degree to indicate the same concept thereafter. m has a unit interval [0, 1] as its codomain where 0 represents a loss that is believed to be impossible and 1 represents a loss that is believed to a must.

In the first example, the risk estimation function from expert one is the following:

$$m(x) = \begin{cases} 0.1 & \text{if } x = \{medium\} \\ 0.8 & \text{if } x = \{high\} \\ 0.1 & \text{if } x = \{low, medium, high\} \end{cases}$$

The case in which x is equal to the universal set A represents the case in which expert one cannot distinguish all losses due to the lack of information or knowledge, referred to as *ignorance*. In other words, each loss is possible in this case. Based on the description of example one, the confidence degree of ignorance is 0.1. In addition, the focal set of the risk estimation is $\{\{medium\}, \{high\}, \{low, medium, high\}\}$.

The risk estimation function from expert two is similar and thus omitted.

In the second example, the risk estimation function from expert one is the following:

$$m(x) = \begin{cases} 0.5 & \text{if } x = \{a | a \in [1000, 4000]\} \\ 0.4 & \text{if } x = \{a | a \in [3000, 5000]\} \\ 0.1 & \text{if } x = \{a | a \in [1000, 6000]\} \end{cases}$$

Notice that A is equal to $[1000, 6000]$. In this case, the confidence degree of ignorance is 0.1 based on the description of example two. The focal set S is $\{[1000, 4000], [3000, 5000], [1000, 6000]\}$.

The goal of the aggregation of risk estimation is to generate a new estimation function based on estimation functions from the two experts in both examples. Before discussing aggregation rules, we should introduce operations that will be applied to combining estimations.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

4. Operations and the K-Algebra

It is straightforward to see that the crucial operations in the aggregation of risk estimations from different sources are the combination of different confidence degrees regarding the same or similar loss (agreement) or different loss (conflict). For this purposes, we introduce some relevant operations defined on the unit interval.

4.1. Conjunction

Intuitively, we expect that the conjunction ($*$) of two confidence degrees x and y about a loss satisfies the following properties:

- $x * y$ is non-decreasing with respect to both x and y. If the confidence degree from one source, e.g. x, is increasing, the conjunct confidence can either increase or have no change, but cannot decrease.
- The evaluation order of $*$ does not matter when aggregating two or more confidence degrees.
- The value 1 (must happen) is the identity element ($1 * x = x$), and 0 (must not happen) is the zero element ($x * 0 = 0$). The intuition is as follows. When combining 1 and another confidence degree x, because 1 means that the case must happen, the conjunct confidence degree should simply be x. When combining 0 and x, because 0 means that the case must not happen, the conjunct confidence degree should be 0.

These requirements are met exactly by the following definition of t-norms.

Definition 2 (Triangular Norm(t-norm) [4]): A binary operation $*$ in the real unit interval [0,1] is a t-norm iff

1) it is associative and commutative, i.e. $\forall x, y, z \in [0, 1]$, $(x * y) * z = x * (y * z)$ and $x * y = y * z$;
2) it is monotonic in the first argument, i.e. $\forall x, y, z \in [0, 1]$, $x \leq y$ implies $x * z \leq y * z$;
3) it satisfies the boundary condition, i.e. $\forall x \in [0, 1], 1 * x = x$.

Lemma 1: A t-norm $*$ is monotonic in the second argument, i.e. $\forall x, y, z \in [0, 1], x \leq y$ implies $z * x \leq z * y$. Furthermore, a t-norm $*$ satisfies $0 * x = 0$ and $x * y \leq \min(x, y)$.

There are uncountably many t-norms [5]. Different t-norms are desirable in different settings of aggregation. The following three basic t-norms are of particular interest to us, because they represent the best operations for combining some typical risk estimations in a conjunctivemanner.

- $x *_g y = \min(x, y)$ (Gödel t-norm)
- $x *_p y = x \cdot y$ (Product t-norm)
- $x *_l y = \max(0, x + y - 1)$ (Łukasiewicz t-norm)

Obviously, we have to determine how to choose the most appropriate t-norm in a particular scenario. Roughly speaking, there are two different methods: a semantic method and an algebraic method.

The semantic method is based on the semantics of the risk estimation and/or the relevant loss. In example two, if expert one estimates the loss to be in the range [1000, 4000] with confidence degree 0.5 using methodology A and with confidence degree 0.4 using methodology B, the conjunction (the pessimistic combination) of these two confidence degrees is clearly 0.4. Therefore, $*_g$ should be applied here.

In example two, expert one estimates the loss to be in the range [1000, 4000] with confidence degree 0.5 and expert two estimates the loss to be in the range [2000, 5000] with confidence degree 0.3. Based on these two evidences, there is an agreement on the intersection loss [2000, 4000], referred to as agreement. Because in a conjunctive combination of evidences, the case in which [2000, 4000] is true relies on the case in which both estimations are true, the confidence degree of [2000, 4000], accordingly, should be $0.5 *_p 0.3 = 0.15$.

In example two, expert one estimates the loss to be in the range [3000-5000] with confidence degree 0.4, referred to as evidence one, while expert two estimates the loss to be in the range [1000-2000] with confidence degree 0.1, referred to as evidence two. These two estimations contradict each other because the two losses have no intersection, referred to as conflict. The confidence degree of the conflict relies on both evidences and may be calculated as follows if we choose a conservative approach to evaluate the combined confidence degree of the conflict.

Based on evidence one, expert one estimates all losses except for [3000-5000] with confidence degree 1 - 0.4 = 0.6. Such a collection of losses obviously contains the loss range [1000-2000]. The confidence degree of the collection (0.6) is compatible with that (0.1) of loss range [1000-2000] in evidence two. Therefore, if we choose a conservative approach, based on both evidences, the combined confidence degree of conflict is 0. $*_l$ can be applied for this situation, that is, $0.4 *_l 0.1 = \max(0, 0.4 + 0.1 - 1) = 0$.

We can clearly see that only if the sum of confidence degrees of the conflicting evidences is larger than 1, the combined confidence degree of conflict is larger than 0. Assume that expert two's confidence on loss range [1000-2000] is 0.7 rather than 0.1. Since the confidence degree on loss range [1000-2000] is at most 0.6 based on evidence one, a conflict arises. The confidence degree might be computed using the difference between

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

two confidence degrees; it would thus be equal to 0.1 (0.7-0.6). The combined confidence degree of conflict is then 0.1, which is exactly the result of $*_l$ conjunction $0.4 * 0.7 = \max(0, 0.4 + 0.7 - 1) = 0.1$.

The algebraic method is based on the algebraic properties of these t-norms. First, we have

$$(x *_g y) \geq (x *_p y) \geq (x *_l y)$$

Hence, if we want to adopt a conservative conjunction, we might choose $*_l$. By contrast, we might choose $*_g$ if we are optimistic about the conjunction.

Second, the Gödel t-norm is also the only t-norm where each $x \in [0,1]$ is an idempotent element, that is, $x *_g x = x$. This reflects the intuition that when two estimates are the same, one takes the consensus.

Third, the product t-norm belongs to an important subclasses of t-norms called strict t-norms. A *strict t-norm* is strictly increasing in both of its arguments, that is, $x_1 *_p y_1 > x_2 *_p y_2$ if either $x_1 > x_2 \wedge y_1 \geq y_2$ or $x_1 \geq x_2 \wedge y_1 > y_2$. Therefore, $*_p$ is useful if we want to ensure that larger confidence degrees always generate a larger combined confidence degree.

Fourth, the Łukasiewicz t-norm is an example of another important subclasses of t-norms, the nilpotent t-norms. A *nilpotent t-norm* ensures that $\forall x \in [0,1), \exists n \in \mathbb{N}$ such that $\overbrace{x * x * \ldots * x}^{n} = 0$. That is, it has the following property: No matter how large a confidence degree is (except for 1), when aggregated enough number of times (finite) using $*_l$, the combined confidence degree is 0.

4.2. Negation

In a situation where we need to calculate the confidence degree of losses except for a particular loss, we may need a negation operation. We define the negation of a confidence degree x, referred to as $\neg x$, as follows:

$$\neg x = 1 - x$$

For instance, if an expert estimates a high loss regarding an access with confidence degree 0.3, then we can use the negation of the confidence degree ($\neg 0.3 = 0.7$) to compute the expert's confidence degree on all of other loss cases (except for the high loss).

4.3. Disjunction

Intuitively, we expect that the disjunction ($\diamond$) of two risk values x and y satisfies the following properties (only the last one differs from the conjunction case):

- The aggregation $x * y$ is non-decreasing with respect to both x and y.
- The evaluation order does not matter.
- The value 1 (absolutely permit) is the subsuming element ($1 * x = 1$), and 0 (absolutely deny) is the identity element ($x * 0 = x$).

These properties are satisfied by any t-conorm. The definition of t-conorm is different from that of t-norm only in the boundary condition: $\forall x \in [0,1], 0 \diamond x = x$. The standard way to define t-conorms [5], [6] is to use t-norms:

$$x \diamond y = \neg(\neg x * \neg y) = 1 - (1 - x) * (1 - y)$$

Thus, the three dual t-conorms are:

- $x \diamond_g y = \max(x, y)$ (Gödel t-conorm)
- $x \diamond_p y = x + y - x \cdot y$ (Product t-conorm)
- $x \diamond_l y = \min(1, x + y)$ (Łukasiewicz t-conorm)

Because t-conorms are constructed by negation and t-norms, they have exactly the same properties of their dual t-norms. Therefore, we omit the discussion of the choice of t-conorms.

4.4. Implication

Sometimes, we may need an implication operation to evaluate "conditional" confidence degrees. For instance, suppose that we have an "unconditional" confidence degree x about a high loss based on all evidences, including both agreement and conflict, from experts, and that the confidence degree of the conflict is y. One may further ask for the confidence degree of a high loss if we ignore the conflict, i.e. adjusting the confidence degree purely based on the agreement. To achieve this goal, we may calculate the "conditional" confidence degree by implication($\rightarrow$), i.e. $\neg y \rightarrow x$. The semantics of $\neg y \rightarrow x$ is the confidence degree of a high loss if there is no conflict, i.e. adjusting the confidence degree x in the case where no conflict exists ($\neg y$). In the standard semantics of t-norm based-fuzzy logics, where conjunction is interpreted by a t-norm, an operation residuum plays the role of implication.

Definition 3: [4] Let $*$ be a left-continuous t-norm. A residuum $\rightarrow$ of t-norm $*$ satisfies $\forall x, y, z \in [0,1], x \rightarrow y = \sup\{z | x * z \leq y\}$.

Lemma 2: [4] For each left-continuous t-norm $*$ and its residuum $\rightarrow$, we have

- $x \leq y$ iff $(x \rightarrow y) = 1$.
- $(1 \rightarrow x) = x$.
- $(0 \rightarrow x) = 1$.
- Gödel residuum: $(x \rightarrow_g y) = y$ if $x > y$.
- Product residuum: $(x \rightarrow_p y) = y/x$ if $x > y$.
- Łukasiewicz residuum: $(x \rightarrow_l y) = 1 - x + y$ if $x > y$.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

The implication operation is widely used to remove the effect of specific evidences like conflict and ignorance that are detailed in Section 5.

4.5. K-Algebra

As mentioned at the beginning of this section, the crucial operations for aggregating risk estimations from different sources are the meaningful operations on the unit interval. Therefore, we introduce the following K-algebra specially designed for the operations on risk estimations.

Definition 4 (K-Algebra): An algebraic structure $\langle [0,1], *, \diamond, \neg, \rightarrow \rangle$ is a K-Algebra iff $*$, $\diamond$, $\neg$, and $\rightarrow$ are t-norms, t-conorms, negation, and implications, respectively.

The K-Algebra actually provides a class of operations that can be used for logical operations, e.g. conjunction, negation, disjunction, or implication, on risk estimations. In the following sections, we show how these operations can be used to express various risk combination rules proposed in the literature and give relevant logic explanations for these rules. In addition, we will show how to generalize these rules based on their logical explanations.

5. Demonstration of Aggregation Rules

From a set theoretic standpoint, estimations from different sources are either compatible or conflicting. If the intersection of two focal losses from different sources is not empty, the relevant estimations are compatible because they at least have an agreement on the intersection. By contrast, the relevant estimations are conflicting if the intersection is empty. Aggregation rules guide the combination of compatible and conflicting estimations from different sources.

5.1. Dempster's Rule

There are multiple possible strategies to combine compatible estimations and conflicting estimations. One strategy may strongly emphasize the agreement between compatible estimations and ignore all of the conflicting evidences. Formally the combined risk estimation (m_{12}) of a loss z is calculated from the aggregation of two risk estimations m_1 and m_2 of relevant losses x and y according to the following steps:

1) We first calculate a raw combined risk estimation $rm_{12}(z)$ of z based on the conjunction ($*_p$) of two *relevant* risk estimations x and y, i.e. $z = x \cap y$, from two sources, expert one and two, respectively.

$$rm_{12}(z = x \cap y) = m_1(x) *_p m_2(y)$$

This step represents one part in the combined risk estimation of loss z based on one evidence from two independent sources because both sources somewhat support loss z through x and y.

2) There might be more evidences for z from two independent sources, e.g. $u \cap v = z$ where u and v are different from x and y, respectively. Therefore, we need to accumulate the risk estimations of all evidences. Such an accumulated risk estimation $arm_{12}(z)$ is calculated through the disjunction ($\diamond_l$) of estimations based on all evidences.

$$arm_{12}(z) = \sum_{x \cap y = z} m_1(x) *_p m_2(y)$$

3) There might be a special class of losses from different sources such that $x \cap y = \emptyset$. Intuitively, the combined risk estimation of $\emptyset$ represents the risk estimation based on conflicting evidences, referred to as conflicting risk estimation. Likewise, we calculate the accumulated raw conflicting risk estimation $arm_{12}(\emptyset)$ by accumulating the conflicting risk estimation between multiple sources through the disjunction ($\diamond_l$) of the conjunction ($*_p$) of the estimations that have no intersection on losses.

$$arm_{12}(\emptyset) = \sum_{x \cap y = \emptyset} m_1(x) *_p m_2(y)$$

4) Finally we calculate the final risk estimation $m_{12}^d(z)$ on z by removing the effect of conflicting risk estimations from the raw risk estimation, i.e. the final risk estimation of z is the confidence degree of the statement that the raw combined risk estimation of all non-conflicting evidences (the negation of the estimation of conflicting evidences) *implies* ($\rightarrow_p$) the raw combined risk estimation of z based on evidences for z.

$$m_{12}^d(z) = (\neg arm_{12}(\emptyset) \rightarrow_p arm_{12}(z))$$

Such an implication recomputes the confidence degree of z in the case where no conflicting evidence exists. Therefore, its semantics is exactly what we need.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

If we use a mathematical equation to present the aforementioned calculation, the combination rule is

$$m_{12}^d(z) = \frac{\sum\limits_{x \cap y = z} m_1(x)m_2(y)}{1 - \sum\limits_{u \cap v = \emptyset} m_1(u)m_2(v)}$$

As we can see, this combination rule equation is exactly the Dempster's Rule of Combination [7] widely used in many applications, e.g. the aggregation of data from different sensors. To interpret the Dempster's Rule, we present a novel t-norm based explanation for it. This explanation is the base of a generalized Dempster's rule discussed in Section 6. To better understand Dempster's rule, let us look at the aggregation results in example one. Figure 1 shows the raw combined estimations of allof risk estimations by expert one and expert two.

			Expert One (m_1)			
			{L}	{M}	{H}	{L,M,H}
			0	0.1	0.8	0.1
Expert Two (m_2)	{L}	0.7	{L} 0	Ø 0.07	Ø 0.56	{L} 0.07
	{M}	0.1	Ø 0	{M} 0.01	Ø 0.08	{M} 0.01
	{H}	0	Ø 0	Ø 0	{H} 0	{H} 0
	{L,M,H}	0.2	{L} 0	{M} 0.02	{H} 0.16	{L,M,H} 0.02

Figure 1. Combined Raw Estimations for Example One

Based on Figure 1, we first calculate the raw conflicting risk estimations from expert one and expert two, i.e. the disjunction of the conjunction of estimations that have an empty intersection.

$$\begin{aligned} arm_{12}(\emptyset) &= m_1(\{M\})m_2(\{L\}) \\ &\quad +m_1(\{H\})m_2(\{L\}) \\ &\quad +m_1(\{H\})m_2(\{M\}) \\ &= 0.07 + 0.56 + 0.08 \\ &= 0.71 \end{aligned}$$

Based on Equation (5.3), we obtain the combined risk estimation as follows:

- $m_{12}^d(\{L\}) = 0.07/(1 - 0.71) = 0.24$
- $m_{12}^d(\{M\}) = 0.04/(1 - 0.71) = 0.138$
- $m_{12}^d(\{H\}) = 0.16/(1 - 0.71) = 0.55$
- $m_{12}^d(\{L, M, H\}) = 0.02/(1 - 0.71) = 0.069$

As we can see from this example, the Dempster's Rule clearly emphasizes the agreement between multiple sources and remove all of the conflicting evidences by normalization (implication operation). Now let us look at how the Dempster's Rule is used to combine estimations of data intervals. The combined raw estimations of different data intervals are shown in Figure 2.

			Expert One(m_1)		
			1000-4000	3000-5000	1000-6000
			0.5	0.4	0.1
Expert Two (m_2)	1000-2000	0.1	1000-2000 0.05	Ø 0.04	1000-2000 0.01
	2000-5000	0.3	2000-4000 0.15	3000-5000 0.12	2000-5000 0.03
	3000-6000	0.4	3000-4000 0.2	3000-5000 0.16	3000-6000 0.04
	1000-6000	0.2	1000-4000 0.1	3000-5000 0.08	1000-6000 0.02

Figure 2. Combined Raw Estimations for Example Two

We calculate the conflicting estimation first. There is only one case.

$$arm_{12}(\emptyset) = 0.04$$

Accordingly, we obtain the combined risk estimation on different intervals as follows:

- $m_{12}^d(1000 - 2000) = (0.05 + 0.01)/(1 - 0.04) = 0.0625$
- $m_{12}^d(1000 - 4000) = 0.1/(1 - 0.04) = 0.10417$
- $m_{12}^d(2000 - 4000) = 0.15/(1 - 0.04) = 0.15625$
- $m_{12}^d(2000 - 5000) = 0.03/(1 - 0.04) = 0.03125$
- $m_{12}^d(3000 - 4000) = 0.2/(1 - 0.04) = 0.20833$
- $m_{12}^d(3000 - 5000) = (0.12 + 0.16 + 0.08)/(1 - 0.04) = 0.375$
- $m_{12}^d(3000 - 6000) = 0.04/(1 - 0.04) = 0.04167$
- $m_{12}^d(1000 - 6000) = 0.02/(1 - 0.04) = 0.02083$

Since the conflict is trivial in example two, the combined risk estimations are only slightly different from the estimations obtained before removing the conflict, as expected.

5.2. Yager's Rule

Completely removing the effect of conflicting risk estimation in the final estimation may not be desired because in highly conflicting scenarios such a rule may generate counter-intuitive results [8]. For instance, the combined risk estimation of a high loss in example one is 0.55, which is counter-intuitive because expert two's confidence degree of a high loss is at most 0.2.

An obvious approach to the problem is not removing the risk estimation of conflicting evidence in the combined risk estimation. However, the question is where we place the combined risk estimation of conflict. Before answering this question, let us consider a special case, that is, the combined risk estimation of

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

the universal set A (the set of all interesting losses). Unsurprisingly, we obtain the following equation to calculate the accumulated raw combined confidence degree of the universal set A.

$$arm_{12}(A) = m_1(A)m_2(A)$$

There is no accumulation because only the intersection of two As is equal to A.

In our setting, A represents the set of *all* interesting losses, thus the risk estimation of the universal set A, i.e. $m(A)$, semantically represents the degree of *ignorance*, that is, the confidence degree of the case in which we cannot identify any difference between all possible losses due to insufficient knowledge or information.

Meanwhile, the combined confidence degree of $\emptyset$ represents the confidence degree of the case in which evidences from different sources conflict with each other. Such conflicting evidences represent a new case in which there is no sufficient information to make a judgment on which one is correct. Therefore, one reasonable solution is integrating, i.e. by using the disjunction, the combined risk estimation of $\emptyset$ and the combined risk estimation of A as the final combined risk estimation of *ignorance*, represented by $m^y_{12}(A)$. Formally,

$$m^y_{12}(A) = arm_{12}(A) + arm_{12}(\emptyset)$$

Such a combination rule for different evidences is referred to as Yager'sRule [9].

In example one, the combined risk estimations based on the Yager's rule are the following:

- $m^y_{12}(\{L\}) = 0.07$
- $m^y_{12}(\{M\}) = 0.04$
- $m^y_{12}(\{H\}) = 0.16$
- $m^y_{12}(\{L, M, H\}) = 0.02 + 0.71 = 0.73$

Compared to the combined risk estimations based on the Dempster's rule, the results based on the Yager's rule clearly show that the ignorance case dominates the combined risk estimation which is in accordance with the fact that two experts present highly conflicting estimations.

The combined risk estimations of scenario two based on the Yager's rule is the following:

- $m^y_{12}(1000 - 2000) = (0.05 + 0.01) = 0.06$
- $m^y_{12}(1000 - 4000) = 0.1$
- $m^y_{12}(2000 - 4000) = 0.15$
- $m^y_{12}(2000 - 5000) = 0.03$
- $m^y_{12}(3000 - 4000) = 0.2$
- $m^y_{12}(3000-5000) = (0.12+0.16+0.08) = 0.36$
- $m^y_{12}(3000 - 6000) = 0.04$
- $m^y_{12}(1000 - 6000) = 0.02 + 0.04 = 0.06$

Because there are very low conflicting evidences in scenario two, the combined risk estimations based on the Yager's rule are close to that based on the Dampster's rule.

5.3. Inagaki's Rule

The difficulty of combination rules for evidences from different sources is how these rules handle conflicting evidence and ignorance. Dempster's rule removes the effect of conflicting evidences while Yager's rule integrates the effect of conflicting evidences and that of ignorance. It is thus natural to suggest a different rule that completely removes the effect of conflicting evidences and ignorance, represented by the disjunction ($\diamond_l$) of $arm_{12}(\emptyset)$ and $m_{12}(A)$. This rule is referred to Inagaki's rule [10]. The crucial difference between the Inagaki's rule and the Dempster's rule is the last step: distributing the conflicting evidences and ignorance to each combined estimation. Formally, the combined estimation $m^i_{12}(z)$ of loss z is calculated as follows.

$$m^i_{12}(z) = (\neg(arm_{12}(\emptyset) \diamond_l m^d_{12}(A)) \rightarrow_p m^d_{12}(z))$$

The objective of the product implication operation in this step is to proportionally distribute the effect of conflicting evidences and ignorance into each combined risk estimation.

The corresponding mathematical equation is the following:

$$m^i_{12}(z) = \frac{\sum\limits_{x \cap y = z} m_1(x)m_2(y)}{1 - m_1(A)m_2(A) - \sum\limits_{u \cap v = \emptyset} m_1(u)m_2(v)}$$

where A represents the universal set as usual.

In example one, the combined risk estimations based on Inagaki's rule are the following:

- $m^i_{12}(\{L\}) = 0.07/(1 - 0.02 - 0.71) = 0.259$
- $m^i_{12}(\{M\}) = 0.04/(1 - 0.02 - 0.71) = 0.148$
- $m^i_{12}(\{H\}) = 0.16/(1 - 0.02 - 0.71) = 0.593$
- $m^i_{12}(\{L, M, H\}) = 0$

In the highly conflicting example one, the Inagaki's rule generates counter-intuitive combined estimation like the Dempster's rule.

The combined risk estimations in example two based on the Inagaki's rule is the following.

- $m^i_{12}(1000 - 2000) = 0.063829787$
- $m^i_{12}(1000 - 4000) = 0.106382979$
- $m^i_{12}(2000 - 4000) = 0.159574468$
- $m^i_{12}(2000 - 5000) = 0.031914894$
- $m^i_{12}(3000 - 4000) = 0.212765957$

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

- $m_{12}^{i}(3000-5000) = 0.382978723$
- $m_{12}^{i}(3000-6000) = 0.042553191$
- $m_{12}^{i}(1000-6000) = 0$

Because there are very low conflicting evidences in example two, the combined risk estimations based on the Inagaki's rule are close to that based on the aforementioned rules. In a low conflicting scenario, the Inagaki's rule is a reasonable choice if we must get rid of ignorance.

5.4. Ferson and Kreinovich's Rule

All aforementioned rules, when dealing with data intervals, calculate the resulting data interval by the intersection of two data intervals. Another choice for calculating the combined data interval is perhaps the arithmetic average of two data intervals. The calculation of risk estimations is the same as that in aforementioned rules: the disjunction of the conjunction of relevant risk estimations. This rule is referred to as Ferson and Kreinovich's rule [11].

$$m_{12}^{f}(z) = \sum_{z=\frac{x+y}{2}} m_1(x) *_p m_2(y)$$

Notice that based on the Ferson and Kreinovich's rule conflicting evidences are averaged; therefore, there is no need to particularly deal with conflicting evidences. The Ferson and Kreinovich's rule emphasizes neither the agreement nor the conflict between evidences. Instead, the rule averages all evidences from sources. Ferson and Kreinovich's rule is not for aggregating risk estimations in example one, but for aggregating risk estimations in example two. The raw combined risk estimations in example two are shown inFigure 3. The combined risk estimations are as follows.

- $m_{12}^{f}(1000-3000) = 0.05$
- $m_{12}^{f}(1000-4000) = 0.01$
- $m_{12}^{f}(1000-5000) = 0.1$
- $m_{12}^{f}(1500-4500) = 0.15$
- $m_{12}^{f}(1500-5500) = 0.03$
- $m_{12}^{f}(2000-3500) = 0.04$
- $m_{12}^{f}(2000-5000) = 0.2$
- $m_{12}^{f}(2000-5500) = 0.08$
- $m_{12}^{f}(2000-6000) = 0.04$
- $m_{12}^{f}(2500-5000) = 0.12$
- $m_{12}^{f}(3000-5500) = 0.16$
- $m_{12}^{f}(1000-6000) = 0.02$

5.5. Dubois and Prade's Disjunctive Rule

All aforementioned rules determine the raw combined risk based on a conjunctive rule on agreement,

			Expert One(m_1)		
			1000-4000	3000-5000	1000-6000
			0.5	0.4	0.1
Expert Two (m_2)	1000-2000	0.1	1000-3000 0.05	2000-3500 0.04	1000-4000 0.01
	2000-5000	0.3	1500-4500 0.15	2500-5000 0.12	1500-5500 0.03
	3000-6000	0.4	2000-5000 0.2	3000-5500 0.16	2000-6000 0.04
	1000-6000	0.2	1000-5000 0.1	2000-5500 0.08	1000-6000 0.02

Figure 3. Combined Estimations for Example Two based on the Ferson and Kreinovich's Rule

i.e. $arm_{12}(z) = \sum_{z=x\cap y} = m_1(x) *_p m_2(y)$. The resulting risk estimation must obtain supports from both sources. Accordingly, we may define a disjunctive consensus rule, i.e. the resulting risk estimation only need to obtain support from one source. Intuitively, the combined resource estimation is obtained by the disjunction of the conjunction of disjunctive risk estimations.

$$m_{12}^{\cup}(z) = \sum_{z=x\cup y} m_1(x) *_p m_2(y)$$

This rule is referred to as Dubois and Prade's rule [12]. Note that the combined losses are the union of relevant losses. Because the union does not generate any conflict and does not reject any of the information provided by the sources, no process is needed to deal with conflicting evidence. However, because this rule emphasizes the disjunctive relation between sources (experts), it may yield a more imprecise result than aforementioned methods. The combined raw estimations for example one are shown in Figure 4. The combined risk estimations are as follows:

- $m_{12}^{\cup}(\{L,M\}) = 0.07$
- $m_{12}^{\cup}(\{L,H\}) = 0.56$
- $m_{12}^{\cup}(\{M\}) = 0.01$
- $m_{12}^{\cup}(\{M,H\}) = 0.08$
- $m_{12}^{\cup}(\{L,M,H\}) = 0.28$

Most combined results do not use singletons to represent losses any more; therefore the combined results, compared to the conjunctive style rules, are fuzzier. One distinct feature of this rule is that the combined risk estimation of ignorance will significantly increase, which can be shown in both examples.

The new combined raw estimations for example two are shown in Figure 5. The combined risk estimations are as follows:

- $m_{12}^{\cup}(1000-4000) = 0.05$
- $m_{12}^{\cup}(1000-5000) = 0.04$
- $m_{12}^{\cup}(2000-5000) = 0.27$
- $m_{12}^{\cup}(3000-6000) = 0.16$

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

			Expert One (m_1)			
			{L}	{M}	{H}	{L,M,H}
			0	0.1	0.8	0.1
Expert Two (m_2)	{L}	0.7	{L} 0	{L,M} 0.07	{L,H} 0.56	{L,M,H} 0.07
	{M}	0.1	{L,M} 0	{M} 0.01	{M,H} 0.08	{L,M,H} 0.01
	{H}	0	{L,H} 0	{M,H} 0	{H} 0	{L,M,H} 0
	{L,M,H}	0.2	{L,M,H} 0	{L,M,H} 0.02	{L,M,H} 0.16	{L,M,H} 0.02

Figure 4. Combined Estimations for Example One based on the Disjunctive Rule

- $m^{\cup}_{12}(1000-6000) = 0.48$

			Expert One(m_1)		
			1000-4000	3000-5000	1000-6000
			0.5	0.4	0.1
Expert Two (m_2)	1000-2000	0.1	1000-4000 0.05	1000-5000 0.04	1000-6000 0.01
	2000-5000	0.3	2000-5000 0.15	2000-5000 0.12	1000-6000 0.03
	3000-6000	0.4	1000-6000 0.2	3000-6000 0.16	1000-6000 0.04
	1000-6000	0.2	1000-6000 0.1	1000-6000 0.08	1000-6000 0.02

Figure 5. Combined Estimations for Example Two based on the Disjunctive Rule

6. Generalization of Combination Rules

The objectives of the algebra are not only to present a new logic-based interpretation of combination rules for uncertain data, but also to introduce a new, systematic approach to generalize these combination rules. In this section, we show how to generalize the Dempster's rule. The generalization of other rules is similar and thus omitted.

In essence, the Dempster's rule emphasizes the agreement while suppressing the conflict. Agreement can be identified by conjunctive operations and collected by disjunctive operations while conflicts can be removed by implication operations. A generalized Dempster's rule is described as follows.

- Calculate a raw combined risk estimation $rm_{12}(z)$ on agreement $z = x \cap y$ by conjunction ($*$) of two relevant risk estimations of x and y.

  $$rm_{12}(z = x \cap y) = m_1(x) * m_2(y)$$

  This step represents a confidence degree of a loss based on one piece of evidence for the loss.
- Calculate an accumulated raw combined risk estimation $arm_{12}(z)$ on agreement $z = x \cap y$ by disjunction ($\diamond$) of all raw combined risk estimations.

  $$arm_{12}(z) = rm^1_{12}(z) \diamond rm^2_{12}(z) \ldots \diamond rm^n_{12}(z)$$

  This step represents the collection of confidence degrees of a loss based on many pieces of evidence for the loss.
- Calculate a final combined risk estimation $m^g_{12}(z)$ by removing the effect of conflicting evidence $arm_{12}(\emptyset)$ through implication ($\rightarrow$).

  $$m^g_{12}(z) = \neg arm_{12}(\emptyset) \rightarrow arm_{12}(z)$$

  Again, this step adjusts a confidence degree of a loss in a situation (or a precondition) in which no conflicting evidence exists.

Semantically, the generalized Dempster's rule exactly realizes its goal: emphasizing the agreement while suppressing the conflict from different evidences. The original Dempster's rule requires the sum of all confidence degrees, before and after aggregation, to be one. Such a condition can be easily realized by the following normalization process after the last step.

$$nm^g_{12}(z) = \frac{m^g_{12}(z)}{\sum\limits_{x \in S} m^g_{12}(x)}$$

where $nm^g_{12}(z)$ represents a normalized confidence degree of loss z, and S represents the focal losses whose confidence degrees are non-zero.

Other rules can be generalized using similar steps and thus are omitted. The obvious benefit of the generalized Dempster's rule is the freedom to choose the best operation based on the semantics of risk estimations and relevant losses, which could potentially improve the quality of combined risk estimations.

7. Conclusion

In this paper, a formal approach based on well developed t-norms is proposed to aggregate uncertain risk estimations from different sources. The logic explanations of existing combination rules are presented, based on which generalized combination rules are developed. In the future, we plan to devise new aggregation rules using the K-Algebra by considering additional factors. For instance, the size of the intersection of losses and the size of the union of losses should affect the combined risk estimations of these losses.

Acknowledgments

The work reported in this paper has been partially supported by the MURI award FA9550-08-1-0265 from the Air Force Office of Scientific Research.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272

References

[1] JASON Program Office, "HORIZONTAL INTEGRATION: Broader Access Models for Realizing Information Dominance," MITRE Corporation, McLean, Virginia 22102, Tech. Rep. JSR-04-132, 12 2004.

[2] P.-C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, and A. S. Reninger, "Fuzzy multi-level security: An experiment on quantified risk-adaptive access control," in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2007, pp. 222–230.

[3] Kari Sentz and Scott Ferson, "Combination of Evidence in Dempster-Shafer Theory," Sandia National Lab., Tech. Rep. SAND 2002-0835, April 2002.

[4] P. Hájek, *Metamathematics of Fuzzy Logic*, 1st ed., ser. Trends in Logic. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1998, vol. 4.

[5] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*, 1st ed., ser. Trends in Logic - Studia Logica Library. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000, vol. 8.

[6] S. Gottwald, *A Treatise on Many-Valued Logics*, 1st ed., ser. Studies in Logic and Computation, D. M. Gabbay, Ed. Baldock, Hertfordshire, England: Research Studies Press Ltd., 2001, vol. 9.

[7] G. Shafer, "Probability judgment in artificial intelligence," in *UAI*, L. N. Kanal and J. F. Lemmer, Eds. Elsevier, 1985, pp. 127–136.

[8] L. A. Zadeh, "Book review: A mathematical theory of evidence," *AI Magazine*, vol. 5, no. 3, pp. 81–83, 1984.

[9] R. R. Yager and L. Liu, Eds., *Classic Works of the Dempster-Shafer Theory of Belief Functions*, ser. Studies in Fuzziness and Soft Computing. Springer, 2008, vol. 219.

[10] M. Itoh and T. Inagaki, "Combination and updating for belief revision in the theory of evidence," in *SCAI*, 1997, pp. 71–82.

[11] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, "Constructing probability boxes and dempster-shafer structures," Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, Tech. Rep. SAND2002-4015, Jan 2003.

[12] D. Dubois and H. Prade, "Evidence, knowledge, and belief functions," *Int. J. Approx. Reasoning*, vol. 6, no. 3, pp. 295–319, 1992.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8272
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8272