Abstract- The modeling of epistemic knowledge is a necessity of most systems dealing with some sort of artificial reasoning. There are several formalisms able to mathematically model someone's degrees of belief. A very popular one is the Bayesian Theory, which is based on a prior knowledge of a probability distribution. Another model is the Theory of Evidence, or Dempster-Shafer Theory, which provides a method for combining evidences from different sources without prior knowledge of their distributions. In this latter method, it is possible to assign probability values to sets of possibilities rather than to single events only, and it is not needed to divide all the probability values among the events, once the remaining probability should be assigned to the environment and not to the remaining events, thus modeling more naturally certain classes of problems. There are some pitfalls however, in particular, the Dempster-Shafer Theory does not model well evidences with a high degree of conflict, and evidences with the more probable possibility disjoint but with a less probable possibility in common tend to bias the results toward the less probable hypothesis in an illogical way, assigning 100% probability to it.

In this paper we present an extension of Dempster-Shafer Theory that overcome the afore mentioned pitfalls, allowing the combination of evidences with higher degrees of conflict, and avoiding the excessive tendency toward the common possibility of otherwise disjoint hypothesis. This is accomplished by means of a new rule of evidences combination that embodies the conflict among the evidences, modeling naturally the epistemic reasoning.

Keywords- Dempster-Shafer Theory, belief functions, belief modeling, Theory of Evidence, reasoning representation.

1 Introduction

If it would be possible always to gather perfect information, it would be trivial to establish a mapping between the input data and the right knowledge base output answer. However, one usually has to deal with imperfect information otherwise our systems would be of very limited application, once they only could be utilized by specialists with total certainty in their inputs.
Table 1. Formal models to deal with information imperfections

<table>
<thead>
<tr>
<th>Formal Model</th>
<th>Probabilistic</th>
<th>Imprecise or Vague</th>
<th>Possibilistic</th>
<th>Uncertain</th>
<th>Inconsistent</th>
<th>Incomplete</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian Theory</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Sha76]</td>
</tr>
<tr>
<td>Theory of Evidence</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Zad85] [DP80]</td>
</tr>
<tr>
<td>Fuzzy Sets</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Paw92]</td>
</tr>
<tr>
<td>Rough Sets</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Ko83]</td>
</tr>
<tr>
<td>Reference Classes</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Zad85] [DP80]</td>
</tr>
<tr>
<td>Possibility Theory</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Be77]</td>
</tr>
<tr>
<td>Para-consistent Logic</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Be77]</td>
</tr>
<tr>
<td>4 Value Logic</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Be77]</td>
</tr>
<tr>
<td>Default Logic</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[Rei80]</td>
</tr>
<tr>
<td>Circumscription Logic</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td>[McC80]</td>
</tr>
</tbody>
</table>

Example 1: The frame of discernment of John’s possible grades in math is \( \Theta = \{a, b, c, d, e\} \).

2.2 Mass Function

The basic probability assignment, or Mass Function, assigns some quantity of belief to the elements of the Frame of Discernment.

Considering an evidence, the Mass Function, \( m \), assigns to each subset of \( \Theta \) (i.e. to \( 2^\Theta \), the powerset of \( \Theta \)), a number in the interval \([0,1]\), where 0 means no belief, and 1 means certainty. The sum of all assignments is equal to 1, meaning that the right hypothesis is in the Frame of Discernment. Also, 0 should be assigned to the empty set, once it is the representation of the false hypothesis. The probability not assigned to any subset of \( \Theta \), is named “non assigned belief”, \( m(\Theta) \), being in fact assigned to \( \Theta \), and not to the negation of the hypothesis that received some belief, as it would be in the Bayesian Theory.

Thus, \( m(A) \) is the measure of the belief assigned by a given evidence to \( A \), where \( A \) is any element of \( 2^\Theta \). As \( m(A) \) deals with the belief assigned to \( A \) only, and not to \( A \) subsets, none belief is forced by the lack of knowledge.

Summarizing:

\[
m: 2^\Theta \rightarrow [0, 1] \tag{1}
m(\emptyset) = 0 \tag{2}
m(A) \geq 0, \forall A \in 2^\Theta \tag{3}
\sum\{ m(A) | A \in 2^\Theta \} = 1 \tag{4}
\]

Example 2: Mass Function and frame of discernment for an evidence of Mary’s grade in philosophy:

\[
m(A) = 0.3
m(B) = 0.25
m(C) = 0.35
m(\Theta) = 0.1
\]

Note that the belief not assigned to the subsets is assigned to the environment.

2.3 Belief Function

The Belief Function, \( Bei \), measures how much the information given by a source support the belief in a specified element as the right answer, thus

\[
Bei: 2^\Theta \rightarrow [0, 1] \tag{5}
\]

The Belief Function to the element \( A \), \( Bei(A) \), is given by:

\[
Bei(A) = \sum_{B \in A} m(B) \tag{6}
\]

Example 3: (using the data from example 2)

\[
Bei(A) = 0.3
Bei(B) = 0.25
Bei(C) = 0.6
Bei(\Theta) = 1
\]

Note that the belief in \( C \) is the sum of the mass of belief of \( B \), 0.25, with the mass of belief of \( C \), 0.35, given that \( C \) contains \( B \); and the belief in \( \Theta \) is the sum of the mass of beliefs of the subsets.

2.4 Plausibility Function

The Upper Probability Function, or Plausibility Function, \( Pl \), measures how much the information given by a source does not contradict a specified element as the right answer, or in other words, how much we should believe in an element if all unknown belief is assigned to it. Thus, the Plausibility Function to the element \( A \), \( Pl(A) \), is:

\[
Pl: 2^\Theta \rightarrow [0, 1] \tag{7}
\]
\[ P(A) = \sum_{B \cap A = \emptyset} m(B) \]  

(8)

Example 4: (using the data from example 2)

For \( A \):

- \( A \cap A = A \neq \emptyset \quad m_1(A) = 0.3 \)
- \( B \cap A = \emptyset \quad m_1(B) = 0.25 \)
- \( C \cap A = \emptyset \quad m_1(C) = 0.35 \)
- \( \emptyset \cap A = A \neq \emptyset \quad m_1(\emptyset) = 0.1 \)
- \( P(A) = 0.4 \)

For \( B \):

- \( A \cap B = \emptyset \quad m_1(B) = 0.25 \)
- \( B \cap B = B \quad m_1(B) = 0.35 \)
- \( C \cap B = B \quad m_1(B) = 0.35 \)
- \( \emptyset \cap B = B \quad m_1(\emptyset) = 0.1 \)
- \( P(B) = 0.7 \)

And for \( C \):

- \( A \cap C = \emptyset \quad m_1(C) = 0.35 \)
- \( B \cap C = B \quad m_1(B) = 0.25 \)
- \( C \cap C = C \quad m_1(C) = 0.35 \)
- \( \emptyset \cap C = C \quad m_1(\emptyset) = 0.1 \)
- \( P(C) = 0.7 \)

2.5 Belief Interval

It is the interval

\[ I(A) = [\text{Bel}(A), P(A)] \]  

(9)

meaning the range of probability where we can believe in \( A \) without severe errors. The Belief Interval is as big as the uncertainty in \( A \).

Example 5: (using data from examples 3 and 4)

- \( I(A) = [0.3, 0.4] \)
- \( I(B) = [0.25, 0.7] \)
- \( I(C) = [0.6, 0.7] \)

2.6 Dempster's Rule

The reasoning process over evidences accumulation needs a method for combining the independent evidences from different sources [Stein93]. The method usually used to combine the evidences is the Dempster’s Rule [Sha76], [Gil98]. Although there are other rules of combination, they differ basically in their normalization part [JSS95], [Gui03], being the procedures adopted by all rules independent of the evidences order.

The Dempster's Rule is composed by an orthogonal sum and a normalization:

\[ m_3(A) = \frac{1}{k} \sum_{A \cap B = \emptyset} m_1(B) m_2(C) \]  

(10)

Where: \( m_1 \oplus m_2 \) denotes the combined effects of the mass functions \( m_1 \) and \( m_2 \); \( k \) is the normalization constant, defined as:

\[ \frac{1}{k} = \sum_{A \cap B = \emptyset} m_1(B) m_2(C) \]  

(11)

And

\[ k = \sum_{A \cap B = \emptyset} m_1(B) m_2(C) \]  

(12)

Example 6: an examination question has as the possibilities of correct answer \( \Theta = \{ a, b, c, d, e \} \), considering \( A = \{ a \}, B = \{ b \}, C = \{ c \}, D = \{ d \}, E = \{ e \} \), was asked to two people what was the probability of each answer be the correct one. The first person answered:

- \( m_1(A) = 0.23 \)
- \( m_1(B) = 0.18 \)
- \( m_1(C) = 0.28 \)
- \( m_1(D) = 0.18 \)
- \( m_1(E) = 0.13 \)

Note that 100% of the belief was assigned to the elements of \( A \), nothing being assigned to \( \Theta \) itself.

The second person's opinion became the second evidence:

- \( m_2(A) = 0.27 \)
- \( m_2(B) = 0.17 \)
- \( m_2(C) = 0.21 \)
- \( m_2(D) = 0.21 \)
- \( m_2(E) = 0.14 \)

Note that the second person preferred does not state anything about the possibility “d”; and as he did not divide 100% of his beliefs among the possibilities, the remaining was assigned to \( \Theta \).

Doing the combination by the Dempster's Rule, would result in:

- \( m_3(A) = 0.30 \)
- \( m_3(B) = 0.17 \)
- \( m_3(C) = 0.31 \)
- \( m_3(D) = 0.08 \)
- \( m_3(E) = 0.14 \)

2.7 Weight of Conflict

It is the logarithm of the normalization constant, denoted by \( \text{Cor}(\text{Bel}_i, \text{Bel}_j) \), where:

\[ \text{Cor}(\text{Bel}_i, \text{Bel}_j) = \log(X) \]  

(13)

If there is no conflict between \( \text{Bel}_i \) and \( \text{Bel}_j \), the sum of the beliefs will be 1 and

\[ \text{Cor}(\text{Bel}_i, \text{Bel}_j) = 0 \]

same wise, if there is nothing in common between the evidences,

\[ \text{Cor}(\text{Bel}_i, \text{Bel}_j) = \infty \]
Example 7: (using data from example 6)

\[
X = 3.1368
\]

\[
\text{Con} (\text{Beli}, \text{Bell}) = \log(X) = 0.4965
\]

The combination between evidences with a high weight of conflict can lead to illogical results by the Dempster’s Rule. Values bigger than 0.5 indicates evidences with more conflict than agreement.

3 Counter Intuitive Behavior of Combination Rules

A classic problem [Joa00], [Vic02] with the Combination Rules until now used is a counter intuitive result found when the evidences to be combined have a concentration of belief in an element disjoint between them, and a common element with low degrees of belief assigned to it. Because the rules do not include any intrinsic mean of belief derating, proportionally to the amount of conflict, they can assign 100% of probability to the element less believed but common to the evidences.

Example 8: your car has broken and you called two auto-mechanics to give their diagnostics.

The mechanic 1 gave his opinion of 99% of chance of a fuel injection problem (\{injection\}), and 1% of chance of an electronic ignition problem (\{ignition\}):

\[
m_1(\text{injection}) = 0.99
\]

\[
m_1(\text{ignition}) = 0.01
\]

The mechanic 2 assigned 99% of certainty to a command belt problem (\{belt\}), and 1% to an electronic ignition problem (\{ignition\}):

\[
m_2(\text{belt}) = 0.99
\]

\[
m_2(\text{ignition}) = 0.01
\]

By the Dempster’s Rule:

\[
\Theta = \{\text{injection}, \text{ignition}, \text{belt}\}
\]

\[
m_\Theta(\text{injection}) = 0
\]

\[
m_\Theta(\text{ignition}) = 0
\]

\[
m_\Theta(\text{belt}) = 1
\]

That is a 100% probability of an electronic ignition problem, contradicting the intuition, and making some authors as [Joa00] state as not advisable the combination of evidences with weight of conflict bigger than a certain value, as 0.5 (as a rule of thumb). Other authors, as [Vic02] complain about this problem and the absence of an embodied representation of evidences’ inconsistency and consequential uncertainty.

4 An Extended Approach for Dempster-Shafer Theory

It is of great importance to analyze the kind of phenomenon portrayed in Example 8. In it, there are two specialists, both with the same degree of reliability, thus the discordance concerning the hypothesis in which most probability were assigned, in fact, decreases the belief on these hypothesis, increasing the uncertainty about them, and at the same time, increasing the belief in the hypothesis in which they assigned a lesser degree of probability, but about which they agree.

Another example: imagine a problem with a frame of discernment having 11 hypothesis. We then ask to 10 people which one of these hypothesis would be the right answer. Each one of these 10 people assigned most of their belief to an hypothesis disjoint from the choice of the others, and little of their belief to a common hypothesis. Considering all people with the same reliability, the divergence about the more individually believed hypothesis, increase the uncertainty about them, at the same time increasing also the chance about the individually lesser believed one, once all the people agreed about it.

Thus, “specialists” agreeing about a hypothesis increase its degree of certainty, although it is exaggerated the assignment of 100% of belief to it, given the divergence about the more individually believed one. By the other side the assignment of only a small portion of the individual belief to the common hypothesis, decrease its intrinsic information value.

We can model this, extending the Theory of Evidence, by using a new rule of evidence combination, that not only correct this counter intuitive effect, but also embodies in the result the uncertainty coming from conflicting hypothesis.

5 Our approach

The proposed rule derates the beliefs according to the degree of conflict between the evidences, assigning the remaining belief to the environment, and not to the common hypothesis. It makes possible to combine evidences with most of their belief assigned to disjoint hypothesis, without the side effect of a counter intuitive behavior. It also makes possible the use of evidences with high values of conflict, which otherwise would be disregarded.

This rule automatically embodies in the result the uncertainty coming from conflicting hypothesis. For two evidences, this is accomplished by dividing the orthogonal sum, as in Dempster’s Rule, by \((1 + \log(1/k))\), i.e. \((1 + \text{Con}(\text{Beli}, \text{Bell}))\):
as it should happen in an epistemological model of reasoning.

Since the orthogonal sum is quasi-associative, if more than two evidences should be combined, one must first combine them by the Dempster's Rule, and thus divide the result by

\[ k_1 + k_2 + \ldots + k_n \]

(16)

(where \( k_1, k_2, \ldots, k_n \) are the \( k \) factor from each pair of evidences combination); finally, one must calculate the additional belief to be added to the initial environment belief (by equation 15).

Example 10: imagine the problem shown in example 8 but at this time two more mechanics are consulted, the first two saying the same thing and the others stating the following beliefs:

Mechanic 3 expects a plug problem with 80% of belief, a battery problem with 19%, and an ignition one with 1% of chance:

\[ m_3((\text{plug})) = 0.80 \]
\[ m_3((\text{battery})) = 0.19 \]
\[ m_3((\text{ignition})) = 0.01 \]

But the mechanic 4 bets on a 75% chance of an fuse problem, 20% of an lubrication problem, 4% of a gas problem, and 1% of an ignition problem:

\[ m_4((\text{fuse})) = 0.80 \]
\[ m_4((\text{lubrication})) = 0.19 \]
\[ m_4((\text{gas})) = 0.19 \]
\[ m_4((\text{ignition})) = 0.01 \]

Applying the Dempster's Rule one still gets the same counter intuitive behavior:

\[ m((\text{plug})) = 0 \]
\[ m((\text{battery})) = 0 \]
\[ m((\text{ignition})) = 1 \]

On the other side, applying the rule proposed here one would get:

The \( k \) factor from the combination between the evidences of mechanic 1 and mechanic 2 is:

\[ k_1 = 0.0001 \]

The factor between the prior result and the evidence from mechanic 3 is:

\[ k_2 = 0.01 \]
And, with the combinations between the partial result and the evidence from mechanic 4, we get:

\[ k_4 = 0.01 \]

By the summation of all \( k \):

\[ k_{\text{total}} = 0.0201 \]

And:

\[ 1 + \log \left( \frac{1}{k} \right) = 1.70 \]

Therefore:

\[
\begin{align*}
\mu(Belt) &= 0 \\
\mu(Injection) &= 0 \\
\mu(Plug) &= 0 \\
\mu(Battery) &= 0 \\
\mu(Fuse) &= 0 \\
\mu(Lubrication) &= 0 \\
\mu(Gas) &= 0 \\
\mu(Ignition) &= 0.37 \\
\mu(\theta) &= 0.63
\end{align*}
\]

It should be noted that the proposed rule shows a better modeling even if the evidences combined shows non disjoint elements, once whatever be the case it will decrease the beliefs assigned to the hypothesis proportionally to the weight of conflict between them, allowing the combination of evidences with a high degree of conflict, and modeling the uncertainty and/or inconsistence among the specialists/consultants.

Example 11: using Example 6 data, it can be seen that even a relatively high weight of conflict \((\text{Con}(Belt, Belt) = 0.4965)\), do not make any difference to an evidence combination by the Dempster's Rule, working the same way as if the evidences had no conflict at all:

\[
\begin{align*}
\mu(A) &= 0.30 \\
\mu(B) &= 0.17 \\
\mu(C) &= 0.31 \\
\mu(D) &= 0.08 \\
\mu(E) &= 0.14
\end{align*}
\]

However, applying the new rule we get a belief assignment of 33% of belief to the environment, and an accompanying decrease of each hypothesis' belief, denoting the uncertainty from the conflict between the evidences:

\[
\begin{align*}
\mu(A) &= 0.200 \\
\mu(B) &= 0.114 \\
\mu(C) &= 0.207 \\
\mu(D) &= 0.053 \\
\mu(E) &= 0.094 \\
\mu(\theta) &= 0.332
\end{align*}
\]

Note that the relative position among the elements stay intact, but their beliefs are reduced proportionally to the weight of conflict, as happens in the real world when we intuitively process our conflicting evidences.

This paper concentrates on the theory extensions. More examples can be found in [CC03].

6 Conclusions

The Theory of Evidence, with the rules of evidence combination until now used, shows two major flaws:

- a counter intuitive behavior when the evidences to be combined have a concentration of belief in an element disjoint between them, and a common element with low degrees of belief assigned to it.
- a lack of an intrinsic representation of evidences' conflict, becoming non advisable to combine evidences with a high weight of conflict.

It is possible to solve these two flaws, extending the application range of the Theory of Evidence, by the adoption of a new rule of evidence combination. This rule corrects the counter intuitive effect, and embodies in the result the uncertainty coming from conflicting hypothesis. This is accomplished by decreasing the beliefs proportionally to the degree of conflict between the evidences, and assigning the remaining belief to the environment instead of to the common hypothesis.

With the proposed rule becomes possible to combine evidences that have most of their belief assigned to disjoint hypothesis, without the side effect of a counter intuitive behavior, and also becomes viable to use evidences with high values of conflict, making useful evidences otherwise inutile.

7 Bibliography


[CC03] F. Campos and S. Cavalcante; “A Method for Knowledge Representation with Automatic Uncertainty Embodiment”; Accepted for presentation at the IEEE NLP-KE 2003, Beijing, China, 6pp., 2003


[Gi98] Borges, Gilene do Espirito Santo; “SGMOO: Sistema Gestor de Métodos Orientados a Objetos Baseado...

[Gui03] Bittencourt, Guilherme; artificial intelligence tutorial in the Web Site of the Automation and Systems Department of the Universidade Federal de Santa Catarina; 2003.

[Joa00] Uchoa, Joaquim Quinteiro; Panotim, Sônia Maria; Nicoletti, Maria do Carmo; “Elementos da Teoria da Evidência de Dempster-Shafer”; Tutorial do Departamento de Computação da Universidade Federal de São Carlos.


[Vic02] Lesser, Victor; “Slides of the Lecture at the CMPSCT”; Fall 2002.
