Dealing with Subjective Uncertainty in Knowledge Based Systems

Fabio Campos
UFPE-DDesign-ProGames
MCT/Sebrae/Finep
Recife – Brazil
ffcc@ieee.org

Andre Neves
UFPE-DDesign-ProGames
MCT/Sebrae/Finep
Recife – Brazil
andremneves@gmail.com

Edvar Vilar Neto
UFPE-DDesign-ProGames
MCT/Sebrae/Finep
Recife – Brazil
edvar.vilar@gmail.com

Abstract

Knowledge based systems depend on algorithms able to relate the inputs of a system to a correct answer coming out of the knowledge-base. Practical systems shows that imperfect information will always get into the data-base and an imperfect knowledge-base will always exist, thus it is usual for a knowledge based system has to be able to model and deal with information imperfections.

The kind of information imperfections involved in knowledge-base systems can be modeled by the "subjective uncertainty", which with the "objective uncertainty" compose the dual nature of the uncertainty, a taxonomy first defined by Helton.

One of the formal models that deals with subjective uncertainty is the Mathematical Theory of Evidence, or Dempster-Shafer Theory. This theory provides a method for combining evidence from different sources without prior knowledge of their distributions, however, it has some pitfalls caused by the non natural embodiment of the uncertainty in the results.

In this paper we analyze the counter-intuitive behavior of the theory, identify the sources of the subjective uncertainty, and present a method of automatic embodiment of the uncertainty which overcomes the pitfalls, allowing its use in a broad range of situations.

1 Introduction

Knowledge based systems depend on algorithms able to relate the inputs of a system to a correct answer coming out of the knowledge-base, and both the inputs and the knowledge-base are subject to information imperfections caused by the lack of knowledge and the conflict.

If it were always possible to get perfect information from users and if the knowledge-base had a perfect modeling with respect to the information it provided (always providing precise answers for each question, without uncertainties and conflicts between its records), the establishment of a mapping between the data inputs of users, and the correct answers in the knowledge-base would not be a difficult task. However, practical reality shows that imperfect information will always get into the data-base and an imperfect knowledge-base will always exist, or else we would have just limited application systems, since they could only be used by specialist users who had total certainty about their applied inputs and whose knowledge-base would have been elaborated by specialists able to supply precise and non-conflicting answers for each possible question.

The kind of information imperfections involved in knowledge-base systems can be modeled by the "subjective uncertainty", which with the "objective uncertainty" compose the dual nature of the uncertainty, a taxonomy first defined by Helton [6].

Objective uncertainty corresponds to the "variability" that emerges from the stochastic characteristic of an environment. At least in principle, it cannot be reduced through additional investigation (although it can be better characterized) [5], [8].

Subjective Uncertainty is the uncertainty that comes from scientific ignorance, uncertainty in measurement, impossibility of confirmation or observation, censorship, or other knowledge deficiency. *A priori* is possible to reduce it through additional empiric efforts [5], [8].

Since objective uncertainty has already been extensively explored in works on classic probability, the decision making under subjective uncertainty is the subject of this article, which extends one of the formal models that deals with it, the Mathematical Theory of Evidence (or Dempster-Shafer Theory).

A key issue in dealing with knowledge representation is how to combine bodies of evidence from different sources, adequately modeling its subjective uncertainty. It is important to understand the sources of the subjective uncertainty embodied in the bodies of evidence themselves and in their combination.

The Theory of Evidence tries to do this but exhibits a

1-4244-1521-7/07/$25.00 ©2007 IEEE

counter intuitive behavior when the bodies of evidence to be combined have a high degree of conflict or when they are disjoint regarding the more believed hypothesis. This counter intuitive behavior limits the range of application of this theory, and, at the same time, leads to a potential disregard of hypotheses that otherwise could add information to the system.

In this work we present a new rule of combining bodies of evidence which is able to overcome these flaws, by the means of a meta-probability mass, named "Lateo". With this approach it becomes possible to eliminate the counter intuitive behavior of the original theory, therefore extending its range of application and better using the available information.

2 The Theory of Evidence

The Theory of Evidence, or Dempster-Shafer Theory, was introduced in the late seventies based on Dempster's works, extended by Shafer [9].

Unlike the Bayesian Theory, the Theory of Evidence does not need prior knowledge of the probability distribution, and it is able to assign probability values to sets of possibilities rather than to single events only. Another differential is that there is no need to divide all the probability among the events, once the remaining probability is assigned to the environment and not to the remaining events. These two differentials allow this theory to model more precisely the natural reasoning process on evidence accumulation, making it progressively more popular.

This formalism provides methods for combining the bodies of evidence carried by different sources, being the Dempster's Rule the de-facto method [4], although there are other rules differing basically in their normalization part [5]. The procedures adopted by all rules of combination, are independent of evidence order (exchangeability).

2.1 Frame of Discernment

A Frame of Discernment, or Environment, is a set of primitive hypotheses, denoted by Θ. It must:

- be exhaustive, in the sense of being complete, containing all possible primitive (atomic) solutions.
- have mutually exclusive primitive elements.

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 a given evidence, the Mass Function, m, assigns to each subset of Θ (i.e. to 2^Θ, the powerset of Θ), a number in the interval $[0,1]$, where 0 means no belief, and 1 means certainty.

2.3 Belief Function

The Belief Function, $\mathcal{B}el$, measures how much the information given by a source support the belief in a specified element as the right answer. The Belief Function for the element $\mathcal{A}$, $\mathcal{B}el(\mathcal{A})$, is given by:

$$\mathcal{B}el : 2^\Theta \rightarrow [0,1] \tag{1}$$

$$\mathcal{B}el(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} m(\mathcal{B}) \tag{2}$$

2.4 Dempster's Rule

The reasoning process over evidence accumulation needs a method for combining the independent evidence from different sources [11]. The method usually used to combine the bodies of evidence is the Dempster's Rule [4], [9]. Although there are other rules of combination, they differ basically in their normalization part [1], [7], being the procedures adopted by all rules independent of the evidence presentation order.

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

$$m_1 \oplus m_2(\mathcal{A}) = \mathcal{X} \sum_{\substack{\mathcal{B} \cup \mathcal{C} = \mathcal{A} \\ \mathcal{A} \neq \varnothing}} m_1(\mathcal{B}).m_2(\mathcal{C}) \ , \ \forall \mathcal{A} \subseteq \Theta \tag{3}$$

Where $m_1 \oplus m_2(\mathcal{A})$ denotes the combined effects of the mass functions m_1 and m_2 and $\mathcal{X}$ is the normalization constant, defined as $1/k$, where:

$$k = 1 - \sum_{\mathcal{A}_i \cap \mathcal{B}_j = \varnothing} m_1(\mathcal{A}_i).m_2(\mathcal{B}_j) \tag{4}$$

Or, likewise:

$$k = \sum_{\mathcal{A}_i \cap \mathcal{B}_j \neq \varnothing} m_1(\mathcal{A}_i).m_2(\mathcal{B}_j) \tag{5}$$

Example 1 *An examination question has as the possibilities of correct answer $\Theta = \{a,b,c,d,e\}$, considering $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{d\}$, and $E = \{e\}$, was asked to two people what was the probability of each answer to be the correct one. The first person answered:*

$$m_1(\mathcal{A}) = 0.23$$
$$m_1(\mathcal{B}) = 0.18$$
$$m_1(\mathcal{C}) = 0.28$$
$$m_1(\mathcal{D}) = 0.18$$
$$m_1(\mathcal{E}) = 0.13$$

Note that 100% of the belief was assigned to the elements of Θ, nothing being assigned to Θ itself.

The second person's opinion became the second evidence:

$$m_2(\mathcal{A}) = 0.27$$
$$m_2(\mathcal{B}) = 0.17$$
$$m_2(\mathcal{C}) = 0.21$$
$$m_2(\mathcal{E}) = 0.21$$
$$m_2(\Theta) = 0.14$$

Note that the second person preferred not stating anything about the possibility "d"; and as he did not divide 100% of his beliefs among the possibilities, the remaining (0.14*) was assigned to* Θ.

Using Dempster's combination rule, would result in:

$$m_3(\mathcal{A}) = 0.30$$
$$m_3(\mathcal{B}) = 0.17$$
$$m_3(\mathcal{C}) = 0.31$$
$$m_3(\mathcal{D}) = 0.08$$
$$m_3(\mathcal{E}) = 0.14$$

2.5 Weight of Conflict

It is the logarithm of the normalization constant, denoted by $Con(\mathcal{B}el_1, \mathcal{B}el_2)$, where:

$$Con(\mathcal{B}el_1, \mathcal{B}el_2) = log(\mathcal{X}) = log(\frac{1}{k}) \qquad (6)$$

The combination of bodies of evidence with a high weight of conflict can lead to counter intuitive, unreasonable, results by the Dempster's Rule.

3 Counter intuitive behavior of the combination rules

A classical problem [9], [10], [13] with the Combination Rules used until now is a counter intuitive result found when the evidence to be combined have a concentration of belief in elements 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 uncertainty (coming from the conflict among them), they can assign 100% of belief to the element less believed but common to the evidence.

Example 2 *Your car has broken and you called two auto mechanics to give their diagnostics.*

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

$$m_1(\{injection\}) = 0.99$$

$$m_1(\{ignition\}) = 0.01$$

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

$$m_2(\{belt\}) = 0,99$$

$$m_2(\{ignition\}) = 0,01$$

By the Dempster's Rule:

$$\Theta = \{injection, ignition, belt\}$$

$$m_3(\{belt\}) = 0$$
$$m_3(\{injection\}) = 0$$
$$m_3(\{ignition\}) = 1$$

That is, a 100% of belief on an electronic ignition problem, contradicting the intuition, and making some authors as [12] state as not advisable the combination of evidence with weight of conflict bigger than a certain value, as 0.5 (as a rule of thumb).

4 Analyzing the counter intuitive behavior

It is of great importance to analyze the kind of phenomenon portrayed on Example 2. By an epistemic point of view it should be a "confirmation effect" about the hypothesis upon which the opinions agreed, once both opinions came from specialists with the same degree of reliability. Thus the discordance concerning the hypothesis in which most belief were assigned, must, in fact, decrease the belief on these hypotheses, increasing the uncertainty about them, and at the same time, increasing the belief in the hypothesis in which they assigned a lesser degree of belief, but about which they agreed, although it is exaggerated a belief assignment of a 100% to the less believed, but common, hypothesis.

Thus, "specialists" agreeing about a hypothesis increase its degree of certainty, although do not make it "totally certain" (i.e. with an assignment of a 100% of belief) given the divergence about the more individually believed one. Corroborating this, the assignment of only a small portion of the individual belief to the common hypothesis decreases its intrinsic information value.

5 Sources of Subjective Uncertainty

It is possible to identify three sources of the subjective uncertainty coming from the evidence and their combination:

1. Explicit lack of knowledge.

2. Non-uniqueness of the assignment of belief and relative division of the belief among the hypotheses chosen.

3. Conflict among the evidence.

Now it will be explained each of these sources:

- Explicit lack of knowledge: it is the lack of knowledge explicitly assigned by the source of evidence when it does not want or does not has conditions, like enough valid information, to divide all its belief among the hypotheses or set of hypotheses. It is represented by the belief assigned to environment, Θ, vide Example 3.

  Example 3 *Example of the explicit lack of knowledge expressed in a mass function "m_1", that is, the lack of knowledge assigned to the environment. In this example, the explicit lack of knowledge is* 0.4.

$$m_1(\mathcal{A}) = 0.1$$
$$m_1(\mathcal{B}) = 0.15$$
$$m_1(\mathcal{C}) = 0.35$$
$$m_1(\Theta) = 0.4$$

- Non-uniqueness of the assignment of belief and relative division of the belief among the hypotheses chosen: if a source of evidence is not able to assign all its belief to just one singleton (atomic hypothesis) this splitting of belief indicates an amount of subjective uncertainty, which would be inversely proportional to the relative difference among the quantity of belief assigned to each hypothesis.

  It is possible to note from this former statement that no subjective uncertainty is embodied in an evidence if all the belief represented in this evidence is assigned to just one singleton (Example 4).

  Example 4 *Example of an evidence with no subjective uncertainty due to the non-uniqueness of the assignment of the belief.*

$$m_1(\mathcal{A}) = 1$$

  On the other hand imagine two bodies of evidence, one with all belief divided equally between two singletons, and another with 99% of its belief assigned to one hypothesis and 1% assigned to another one, the former body of evidence would have a higher amount of subjective uncertainty than the later one (Example 5).

  Example 5 *m_1 embodies a higher amount of subjective uncertainty than m_2.*

$$m_1(\mathcal{A}) = 0.5$$
$$m_1(\mathcal{B}) = 0.5$$

$$m_2(\mathcal{A}) = 0.99$$
$$m_2(\mathcal{B}) = 0.01$$

- Conflict among the evidence: if two or more evidence to be combined do not have their mass functions exactly equal the difference among the beliefs assigned by the diverse sources represents a conflict, and the amount of subjective uncertainty coming from the combination is proportional to the amount of conflict among the bodies combined (Example 6).

  Example 6 *The combination of the bodies m_1 and m_2 would result in a higher amount of subjective uncertainty than the combination of m_3 with m_4.*

$$m_1(\mathcal{A}) = 0.5$$
$$m_1(\mathcal{B}) = 0.5$$

$$m_2(\mathcal{A}) = 0.5$$
$$m_2(\mathcal{B}) = 0.5$$

$$m_3(\mathcal{A}) = 0.0$$
$$m_3(\mathcal{B}) = 0.1$$
$$m_3(\mathcal{C}) = 0.9$$

$$m_4(\mathcal{A}) = 1.0$$

With a new rule of evidence combination we can model all these sources of subjective uncertainty, extending the Theory of Evidence, correcting the counter intuitive effect of the original theory, and also embodying in the result the subjective uncertainty (by using a "measure" – so to speak – of the subjective uncertainty named "Lateo").

6 Our approach

The proposed rule derates the beliefs according to the degree of conflict between the evidence, assigning the remaining belief to the environment (and not to the common hypothesis) along with the uncertainty that would be assigned to the environment by the original Dempster's Rule [3]. This quantity of belief assigned to the environment constitutes a measure of the subjective uncertainty coming from the non knowledge or conflict among the evidence, being named "Lateo" and denoted by Λ, in allusion to its causes, once "Lateo" in Latin means "being hidden", "being out of sight", "be unknown".

This rule makes possible to combine evidence with most of their belief assigned to disjoint hypotheses, without the side effect of a counter intuitive behavior. It also allows the use of evidence with high values of conflict, making useful evidence otherwise useless.

For two bodies of evidence, this is accomplished by dividing the orthogonal sum, as in Dempster's Rule, by $(1 + log(1/k))$, that is, $(1 + \mathcal{C}on(\mathcal{B}el_1, \mathcal{B}el_2))$:

$$m_1 \Psi m_2(\mathcal{A}) = \frac{\mathcal{X} \sum_{\substack{\mathcal{B}\cap\mathcal{C}=\mathcal{A} \\ \mathcal{A}\neq\varnothing}} m_1(\mathcal{B}) \cdot m_2(\mathcal{C})}{1 + log(\frac{1}{k})} \text{ , } \forall \mathcal{A} \subset \Theta \quad (7)$$

The additional belief from the derating of the hypotheses is added to the initial environment belief, originating the Lateo:

$$\Lambda = (\mathcal{X}.m_1(\Theta).m_2(\Theta)) + 1 - \sum_{\substack{\mathcal{A}\subset\Theta \\ \mathcal{A}\neq\varnothing}} m_1 \Psi m_2(\mathcal{A}) \quad (8)$$

It can be noted that $(\mathcal{X}.m_1(\Theta).m_2(\Theta))$ is equal to $m_1 \oplus m_2(\Theta)$ by the Dempster's Rule, and the proposed rule adds to this belief a value proportional to the conflict and non assigned belief among the evidence.

The numeric value expressed by the Lateo represents a mobile mass of belief that in the absence of unknown belief and conflict among the evidence could be associated with any element or combination of elements of the frame of discernment.

6.1 Combining evidence with most of their beliefs assigned to disjoints hypotheses

The proposed approach solves the counter intuitive behavior of the original theory when combining evidence whose most belief is assigned to disjoint hypotheses, as it is illustrated by Example 7.

Example 7 *Applying our rule to the data from Example 2, we get:*

$$k = 0.0001$$
$$\mathcal{X} = 10,000$$
$$log(\mathcal{X}) = 4$$

And thus:

$$m_3(\{belt\}) = 0$$
$$m_3(\{injection\}) = 0$$
$$m_3(\{ignition\}) = 0.2$$
$$\Lambda = 0.8$$

As it can be seen, the reasoning is more naturally modeled once the belief in the command belt and in the fuel injection continue to be disregarded due to their disjunction, but the uncertainty is better represented, since 80% of the belief is assigned to the environment and not to a hypothesis in particular [2].

The Plausibility Function and the Belief Interval would be:

$$\mathcal{B}el(\{ignition\}) = 0.2$$
$$\mathcal{P}l(\{ignition\}) = 1$$
$$\mathcal{I}(\{ignition\}) = [0.2, 1]$$

This shows a much more realistic modeling of the problem, as the plausibility of the electronic ignition hypothesis continue to be 100%, while its belief is decreased to 20%. From an epistemological point of view it would not be appropriate a belief assignment of 100% to the electronic ignition simply because the mechanics disagreed about the most believed hypothesis, and agreed about the one with a low belief.

Note that in which regards a decision making process the original theory would suggest an "ignition" problem without uncertainty (Example 2), while our approach makes clear that the information collected is not enough to allow a reasonable decision, once the subjective uncertainty measure (the Lateo) is bigger than the knowledge available (that is, 80% Lateo against 20% "ignition").

6.2 Combining evidence with high degree of conflict

It should be noted that the proposed rule shows a better modeling even if the evidence combined do not show concentration of belief in disjoint elements, once whatever be the case it will decrease the beliefs assigned to the hypotheses proportionally to the weight of conflict between them, allowing the combination of evidence with a high degree of conflict, and modeling the uncertainty and/or inconsistence among the specialists/consultants.

Example 8 *Using Example 1 data, it can be seen that even a relatively high weight of conflict ($\mathcal{C}on(\mathcal{B}el_1, \mathcal{B}el_2) = 0.4965$), do not make any difference to an evidence combination by the Dempster's Rule, working the same way as if the evidence had no conflict at all:*

$$m_3(\mathcal{A}) = 0.30$$
$$m_3(\mathcal{B}) = 0.17$$
$$m_3(\mathcal{C}) = 0.31$$
$$m_3(\mathcal{D}) = 0.08$$
$$m_3(\mathcal{E}) = 0.14$$

However, applying the new rule we get an belief assignment of 33% of belief to the environment, and an accompanying decrease of each hypothesis' belief, denoting the uncer-

tainty from the conflict between the evidence:

$$m_3(\mathcal{A}) = 0.200$$
$$m_3(\mathcal{B}) = 0.114$$
$$m_3(\mathcal{C}) = 0.207$$
$$m_3(\mathcal{D}) = 0.053$$
$$m_3(\mathcal{E}) = 0.094$$
$$\Lambda = 0.332$$

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 evidence.

Regarding a decision making process, by the Dempsters Rule, one would choose hypothesis $\mathcal{A}$ or $\mathcal{C}$ as the correct answer, while our rule makes clear that no hypotheses should be chosen, as the value of Lateo summed to any hypothesis is enough to make this hypothesis the bigger one.

7 Conclusion

Although the Theory of Evidence is able to deal with subjective uncertainty, it shows two major flaws caused by the rules of evidence combination until now used:

1. A counter intuitive behavior when the evidence to be combined have a concentration of belief in elements disjoints between them, and a common element with low degrees of belief assigned to it.

2. A lack of an intrinsic representation of the subjective uncertainty, coming from the unknown or from the conflict among the evidence, becoming non advisable to combine evidence with a high weight of conflict.

A decision making process can be affected by these flaws leading to erroneous decisions. But it is possible to solve these two flaws, extending the application range of the Theory of Evidence, by the adoption of a proposed new rule of evidence combination. This rule corrects the counter intuitive effect, and embodies in the result the subjective uncertainty. This is accomplished by decreasing the beliefs proportionally to the degree of conflict between the evidence, and assigning the remaining belief to the environment instead of to the common hypothesis, resulting in a measure of the subjective uncertainty named "Lateo".

With the proposed rule it becomes possible to know the degree of subjective uncertainty involved in the combination of evidence, making clear the possibility of making reasonable decisions based in the evidence combined.

Additionally, the implementation of the Lateo introduces a number of interesting possibilities once it represents a measure of the subjective uncertainty, allowing:

- An indication of how much the numerical results obtained by the Dempster-Shafer Theory are distant from the numeric results obtained by the theories of precise probability.
- To know how much one can trust in the results for them to be used in a decision making process.
- An estimation of the level of confidence that one can have in the sources consulted regarding the solution of the given question.

References

[1] G. Bittencourt. *Inteligência Artificial Ferramentas e Teorias.* Editora da UFSC, Santa Catarina, 1998. ISBN 85-328-0138-2.

[2] F. Campos and S. Cavalcante. An extended approach for dempster-shafer theory. In *IEEE-IRI-03*, Las Vegas, USA, 2003.

[3] F. Campos and S. Cavalcante. A method for knowledge representation with automatic uncertainty embodiment. In *IEEE-NLP-KE-03*, Beijing, China, 2003.

[4] G. do Espírito Santo Borges. Sgmoo: Sistema gestor de métodos orientados a objetos baseado em conhecimento. Master's thesis, Departamento de Ciência da Computação - Universidade de Brasília, 1998.

[5] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. *Sand Report*, Jan. 2003. Unlimited Release.

[6] J. C. Helton. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation*, 57:3–76, 1997.

[7] A. Joshi, S. Shasrabudhe, and K. Shankar. Sensivity of combination schemes under conflicting conditions and a new method. In J. Wainer, A. Carvalho, et al., editors, *Advances in Artificial Inteligence*, Lectures Notes in Computer Science, pages 39–47. Springer-Verlag, 1995.

[8] K. Sentz and S. Ferson. Combination of evidence in dempster-shafer theory. *Sand Report*, Apr. 2002. Unlimited Release.

[9] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976. ISBN: 0-691-08175-1.

[10] P. Smets. Belief functions. In P. Smets, E. Mamdani, D. Dubois, and H. Prade, editors, *Non-Standard Logics for Automated Reasoning*. Academic Press, San Diego CA, USA, 1988. ISBN: 0-12-649520-3.

[11] R. Stein. The dempster-shafer theory of evidential reasoning. *Artificial Intelligence Expert*, Aug. 1993.

[12] J. Q. Uchoa, S. M. Panotim, and M. d. C. Nicoletti. Elementos da teoria da evidência de dempster-shafer. Tutorial do Departamento de Computação da Universidade Federal de São Carlos.

[13] L. A. Zadeh. Book review: A mathematical theory of evidence. *AI Magazine*, 5(3):81–83, 1984.