A Review on Strategies for Combining Conflicting Dogmatic Beliefs

Zarna S. Parmar
parmarzarna176@gmail.com

Prof. Vrushank Shah
vrushankshah.ec@iite.edu.in

Abstract

The consensus operator provides a method for combining possibly conflicting beliefs within the Dempster-Shafer belief theory, and represents an alternative to the traditional Dempster's rule. In everyday discourse dogmatic beliefs are expressed by observers when they have a strong and rigid opinion about a subject of interest. Such beliefs can be expressed and formalised within the Demspter- Shafer belief theory. This paper describes how the consensus operator can be applied to dogmatic conflicting opinions, i.e. when the degree of conflict is very high. It overcomes shortcomings of Dempster's rule and other operators that have been proposed for combining possibly conflicting beliefs.

Key words: Dempster's rule, belief, conflict, consensus operator, subjective logic

1 Introduction

Ever since the publication of Shafer's book A Mathematical Theory of Evidence. there has been continuous controversy around the so-called

Dempster's rule. The purpose of Dempster's rule is to combine two conflicting beliefs into a single belief that reflects the two conflicting beliefs in a fair and equal way. Dempster's rule has been criticised mainly because highly conflicting beliefs tend to produce counterintuitive results. The problem with Dempster's rule is due to its normalisation which redistributes conflicting belief masses to nonconflicting beliefs, and thereby tends to eliminate any conflicting characteristics in the resulting belief mass distribution.

2 Dempster's rule

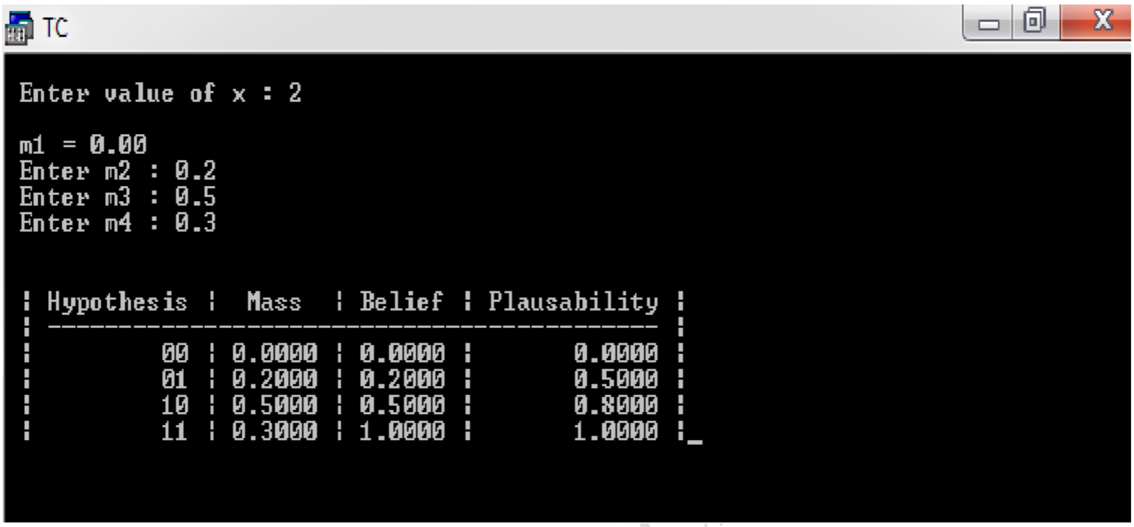
Dempster's rule allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The theory was first developed by Arthur P. Dempster and Glenn Shafer.
Dempster–Shafer theory is a generalization of the Bayesian theory of subjective probability; whereas the latter requires probabilities for each question of interest, belief functions base degrees of belief (or confidence, or trust) for one question on the probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ depends on how closely the two questions are related. Put another way, it is a way of representing epistemic plausibilities but it can yield answers that contradict those arrived at using probability theory.
Often used as a method of sensor fusion, Dempster–Shafer theory is based on two ideas: obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence. In essence, the degree of belief in a proposition depends primarily upon the number of answers (to the related questions) containing the proposition, and the subjective probability of each answer. Also contributing are the rules of combination that reflect general assumptions about the data.
In this formalism a degree of belief (also referred to as a mass) is represented as a belief function rather than a Bayesian probability distribution. Probability values are assigned to *sets* of possibilities rather than single events: their appeal rests on the fact they naturally encode evidence in favor of propositions.
Dempster–Shafer theory assigns its masses to all of the non-empty subsets of the entities that compose a system.

2.1 Belief and plausibility

Shafer's framework allows for belief about propositions to be represented as intervals, bounded by two values, *belief* (or *support*) and *plausibility*:

$$\text{belief} \leq \text{plausibility}.$$

Belief in a hypothesis is constituted by the sum of the masses of all sets enclosed by it (*i.e.* the sum of the masses of all subsets of the hypothesis). It is the amount of belief that directly supports a given hypothesis at least in part, forming a lower bound. Belief (usually denoted *Bel*) measures the strength of the evidence in favor of a set of propositions. It ranges from 0 (indicating no evidence) to 1 (denoting certainty). *Plausibility* is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. It is an upper bound on the possibility that the hypothesis could be true, *i.e.* it "could possibly be the true state of the system" up to that value, because there is only so much evidence that contradicts that hypothesis. Plausibility (denoted by *Pl*) is defined to be *Pl(s)=1-Bel(~s)*. It also ranges from 0 to 1 and measures the extent to which evidence in favor of *~s* leaves room for belief in *s*. For example, suppose we have a belief of 0.5 and a plausibility of 0.8 for a proposition, say "the cat in the box is dead." This means that we have evidence that allows us to state strongly that the proposition is true with a confidence of 0.5. However, the evidence contrary to that hypothesis (i.e. "the cat is alive") only has a confidence of 0.2. The remaining mass of 0.3 (the gap between the 0.5 supporting evidence on the one hand, and the 0.2 contrary evidence on the other) is "indeterminate," meaning that the cat could either be dead or alive. This interval represents the level of uncertainty based on the evidence in your system.



The null hypothesis is set to zero by definition (it corresponds to "no solution"). The orthogonal hypotheses "Alive" and "Dead" have probabilities of 0.2 and 0.5, respectively. This could correspond to "Live/Dead Cat Detector" signals, which have respective reliabilities of 0.2 and 0.5. Finally, the all-encompassing "Either" hypothesis (which simply acknowledges there is a cat in the box) picks up the slack so that the sum of the masses is 1. The belief for the "Alive" and "Dead" hypotheses matches their corresponding masses because they have no subsets; belief for "Either" consists of the sum of all three masses (Either, Alive, and Dead) because "Alive" and "Dead" are each subsets of "Either". The "Alive" plausibility is $1 - m$ (Dead) and the "Dead" plausibility is $1 - m$ (Alive). Finally, the "Either" plausibility sums $m(\text{Alive}) + m(\text{Dead}) + m(\text{Either})$. The universal hypothesis ("Either") will always have 100% belief and plausibility —it acts as a checksum of sorts.

2.2 Combining beliefs

Beliefs from different sources can be combined with various fusion operators to model specific situations of belief fusion, e.g. with *Dempster's rule of combination*, which combines belief constraints that are dictated by independent belief sources, such as in the case of combining hints or combining preferences. Note that the probability masses from propositions that contradict each other can be used to obtain a measure of conflict between the independent belief sources. Other situations can be modeled with different fusion operators, such as cumulative fusion of beliefs from independent sources which can be modeled with the cumulative fusion operator.

Dempster's rule of combination is sometimes interpreted as an approximate generalisation of Bayes' rule. In this interpretation the priors and conditionals need not be specified, unlike traditional Bayesian methods, which often use a symmetry (minimax error) argument to assign prior probabilities to random variables (*e.g.* assigning 0.5 to binary values for which no information is available about which is more likely). However, any information contained in the missing priors and conditionals is not used in Dempster's rule of combination unless it can be obtained indirectly—and arguably is then available for calculation using Bayes equations.

Dempster–Shafer theory allows one to specify a degree of ignorance in this situation instead of being forced to supply prior probabilities that add to unity. This sort of situation, and whether there is a real distinction between *risk* and *ignorance*, has been extensively discussed by statisticians and economists.

2.3 Formal definition

Let X be the *universal set*: the set representing all possible states of a system under consideration. The power set

$$2^X$$

is the set of all subsets of X, including the empty set $\varnothing$. For example, if:

$$X = \{a, b\}$$

then

$$2^X = \{\varnothing, \{a\}, \{b\}, X\}.$$

The elements of the power set can be taken to represent propositions concerning the actual state of the system, by containing all and only the states in which the proposition is true.

The theory of evidence assigns a belief mass to each element of the power set. Formally, a function

$$m : 2^X \to [0, 1]$$

is called a *basic belief assignment* (BBA), when it has two properties. First, the mass of the empty set is zero:

$$m(\varnothing) = 0.$$

Second, the masses of the remaining members of the power set add up to a total of 1:

$$\sum_{A \in 2^X} m(A) = 1$$

The mass $m(A)$ of A, a given member of the power set, expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to A but to no particular subset of A. The value of $m(A)$ pertains *only* to the set A and makes no additional claims about any subsets of A, each of which have, by definition, their own mass.

From the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense), and is bounded by two non-additive continuous measures called belief (or support) and plausibility:

$$\operatorname{bel}(A) \le P(A) \le \operatorname{pl}(A).$$

The belief bel(A) for a set A is defined as the sum of all the masses of subsets of the set of interest:

$$\operatorname{bel}(A) = \sum_{B \mid B \subseteq A} m(B).$$

The plausibility pl(A) is the sum of all the masses of the sets B that intersect the set of interest A:

$$\operatorname{pl}(A) = \sum_{B \mid B \cap A \neq \varnothing} m(B).$$

The two measures are related to each other as follows:

$$\operatorname{pl}(A) = 1 - \operatorname{bel}(\overline{A}).$$

And conversely, for finite A, given the belief measure bel(B) for all subsets B of A, we can find the masses m(A) with the following inverse function:

$$m(A) = \sum_{B \mid B \subseteq A} (-1)^{|A-B|} \operatorname{bel}(B)$$

where $|A - B|$ is the difference of the cardinalities of the two sets.

It follows from the last two equations that, for a finite set X, you need know only one of the three (mass, belief, or plausibility) to deduce the other two; though you may need to know the values for many sets in order to calculate one of the other values for a particular set. In the case of an infinite X, there can be well-defined belief and plausibility functions but no well-defined mass function.

3 Dempster's rule of combination

The problem we now face is how to combine two independent sets of probability mass assignments in specific situations. In case different sources express their beliefs over the frame in terms of belief constraints such as in case of giving hints or in case of expressing preferences, then Dempster's rule of combination is the appropriate fusion operator. This rule derives common shared belief between multiple sources and ignores *all* the conflicting (non-shared) belief through a normalization factor. Use of that rule in other situations than that that of combining belief constraints has come under serious criticism, such as in case of fusing separate beliefs estimates from multiple sources that are to be integrated in a cumulative manner, and not as constraints. Cumulative fusion means that all probability masses from the different sources are reflected in the derived belief, so no probability mass is ignored.

Specifically, the combination (called the joint mass) is calculated from the two sets of masses m_1 and m_2 in the following manner:

$$m_{1,2}(\varnothing) = 0$$

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C)$$

where

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C).$$

K is a measure of the amount of conflict between the two mass sets.

4 The Consensus Operator

4.1 The Opinion Space

The consensus operator is not defined for general frames of discernment, but only on binary frames of discernment. If the original frame of discernment is larger than
binary it is possible to derive a binary frame of discernment containing any element A and its complement A through simple or normal coarsening. After normal coarsening, the relative atomicity of A is equal to the relative cardinality of A in the original frame of discernment. An *opinion* basically consists of a bba on a (coarsened) binary _ with an additional *relative atomicity* parameter that enables the computation of the probability expectation value (or pignistic belief) of an opinion.

4.1.1 Opinion

Let Θ be a (coarsened) binary frame of discernment containing sets A and $\bar{A}$, and let m^X be the (coarsened) bba on Θ held by X. Let b_A^x = $mX(A)$, $d_A^x = mX(\bar{A})$ and $u_A^x = mX(\Theta)^4$ be called the belief, disbelief and uncertainty components respectively, and let a_A^x represent the relative atomicity of A. Then X's opinion about A, denoted by w_A^x, is the ordered tuple:

$$w_A^x = (b_A^x, d_A^x, u_A^x, a_A^x)$$

The belief, disbelief and uncertainty components of an opinion represent exactly the same as a bba, so the following equality holds:

$$b_A + d_A + u_A = 1 \; ; A \in 2^\Theta$$

Opinions have an equivalent represention as beta probability density functions (pdf) denoted by beta (œ; ß) through the following bijective mapping:

$$(b_A, d_A, u_A, a_A) \longleftrightarrow$$

$$\text{beta}\left(\frac{2b_A}{u_A} + 2a_A, \;\; \frac{2d_A}{u_A} + 2(1 - a_A)\right).$$

This means for example that an opinion with $u_A = 1$ and $a_A = 0{:}5$ which maps to beta (1; 1) is equivalent to a uniform pdf. It also means that a dogmatic opinion with $u_A = 0$ which maps to beta ($b_A\mathscr{M}$; $d_A\mathscr{M}$) where $\mathscr{M} \rightarrow$ OO is equivalent to a spike pdf with infinitesimal width and infinite height. Dogmatic opinions can thus be interpreted as being based on an infinite amount of evidence.

4.2.2 The Consensus Operator

The Consensus Operator defined below is derived from the combination of two beta pdfs. More precisely, the two input opinions are mapped to beta pdfs and combined, and the resulting beta pdf mapped back to the opinion space again. The Consensus Operator can thus be interpreted as the statistical combination of two beta pdf
Let $w_A^x = (b_A^x, d_A^x, u_A^x, a_A^x)$ and $w_A^y = (b_A^y, d_A^y, u_A^y, a_A^y)$ be opinions respectively held by agents X and Y about the same element A, and let $k = u_A^x + u_A^y - u_A^x u_A^y$. when $u_A^x, u_A^y \rightarrow 0$ the relative dogmatism between w_A^x and w_A^y is defined by $\mathscr{Y}_A^{X/Y}$ so that $\mathscr{Y}_A^{X/Y} = u_A^y / u_A^x$. Let $w_A^{x,y} = (b_A^{x,y}, d_A^{x,y}, u_A^{x,y}, a_A^{x,y})$ be the opinion such that:

$$\text{for } \kappa \neq 0 : \quad \begin{cases} b_A^{X,Y} = (b_A^X u_A^Y + b_A^Y u_A^X)/\kappa \\ d_A^{X,Y} = (d_A^X u_A^Y + d_A^Y u_A^X)/\kappa \\ u_A^{X,Y} = (u_A^X u_A^Y)/\kappa \\ a_A^{X,Y} = \frac{a_A^X u_A^Y + a_A^Y u_A^X - (a_A^X + a_A^Y) u_A^X u_A^Y}{u_A^X + u_A^Y - 2 u_A^X u_A^Y} \end{cases},$$

$$\text{for } \kappa = 0 : \quad \begin{cases} b_A^{X,Y} = (\gamma_A^{X/Y}\, b_A^X + b_A^Y)/(\gamma_A^{X/Y} + 1) \\ d_A^{X,Y} = (\gamma_A^{X/Y}\, d_A^X + d_A^Y)/(\gamma_A^{X/Y} + 1) \\ u_A^{X,Y} = 0 \\ a_A^{X,Y} = (\gamma_A^{X/Y}\, a_A^X + a_A^Y)/(\gamma_A^{X/Y} + 1) \end{cases}.$$

Then $w_A^{x,y}$ is called the consensus opinion between w_A^x and w_A^y, representing an imaginary agent [X, Y]'s opinion about A, as if that agent represented both X and Y .

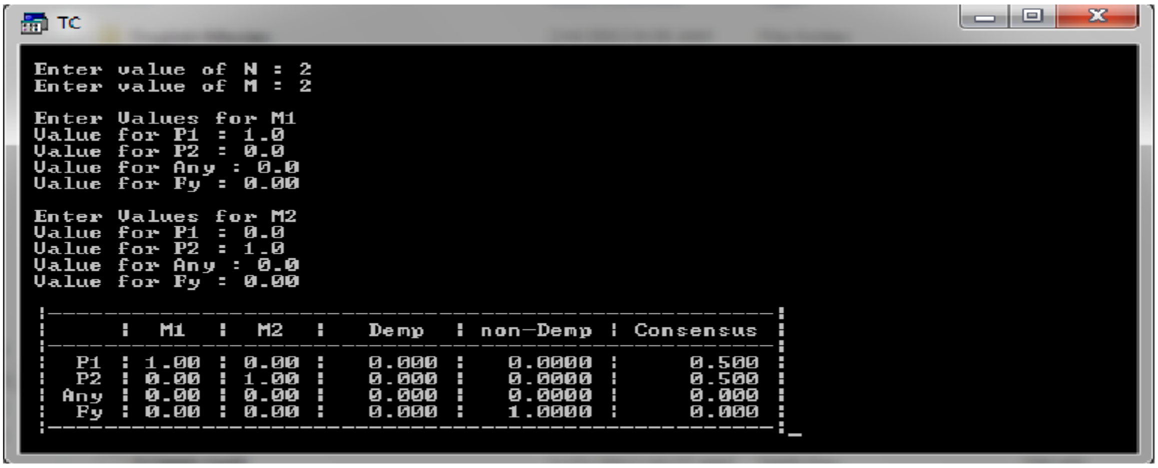
The consensus operator is commutative, associative and non-idempotent. Associativity in case k = 0. In case of two totally uncertain opinions (i.e. u = 1) it is required that the observers agree on the relative atomicity so that the consensus relative atomicity for example can be defined as $a_A^{x,y} = a_A^x$.

5 Comparison of Combination Rules

In this section, we present several examples in order to compare the performance of the different rules described in the previous sections.

5.1 Example 1 : One Dogmatic Belief

Let m1 and m2 represent two distinct pieces of evidence about the states in Θ = {P1, P2}. In this example, we suppose that we want to combine a dogmatic belief function m1 with a non-dogmatic one m2. Table 1 presents these two bba's with the results obtained by the previously presented operators.



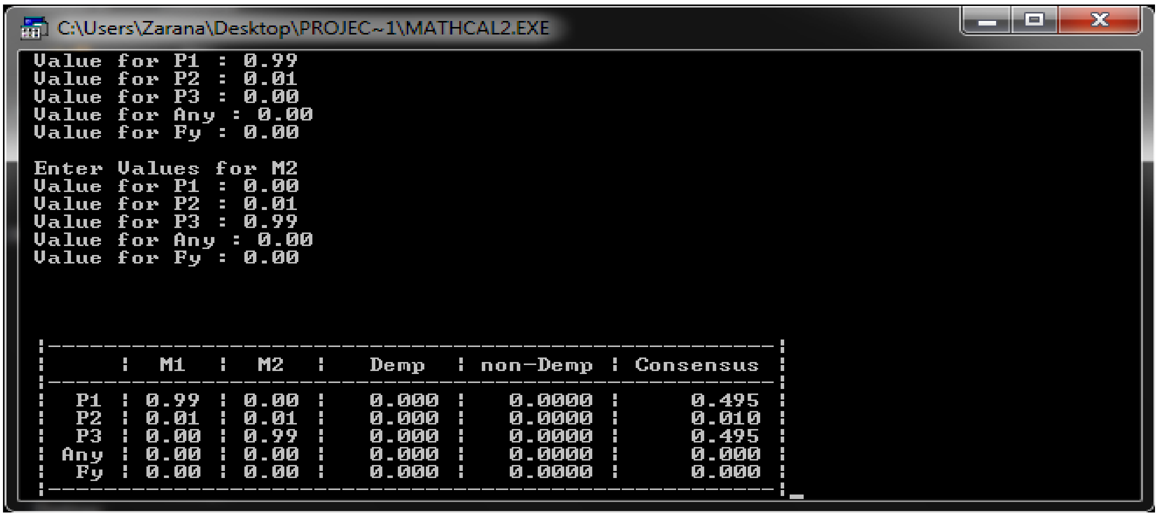
	M1	M2	Demp	non-Demp	Consensus
P1	1.00	0.00	0.000	0.0000	0.500
P2	0.00	1.00	0.000	0.0000	0.500
Any	0.00	0.00	0.000	0.0000	0.000
Fy	0.00	0.00	0.000	1.0000	0.000

5.2 Example 2 : Zadeh's Example

Suppose that we have a murder case with three suspects; Peter, Paul and Mary and two witnesses M1 and M2 who give highly conflicting testimonies. Table 2 gives the witnesses' belief masses in Zadeh's example and the resulting belief masses after applying Dempster's rule, the non-normalised rule and the consensus operator.

5.3 Example 3 : Zadeh's Example Modified

By introducing a small amount of uncertainty in the witnesses testimonies , we get the output as shown below.

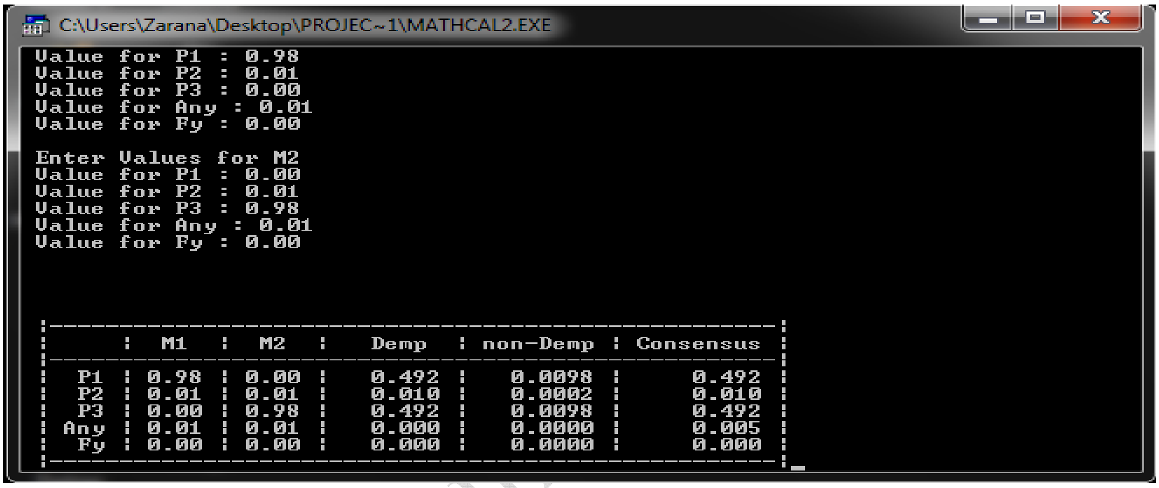


C:\Users\Zarana\Desktop\PROJEC~1\MATHCAL2.EXE

Value for P1 : 0.99
Value for P2 : 0.01
Value for P3 : 0.00
Value for Any : 0.00
Value for Fy : 0.00

Enter Values for M2
Value for P1 : 0.00
Value for P2 : 0.01
Value for P3 : 0.99
Value for Any : 0.00
Value for Fy : 0.00

	M1	M2	Demp	non-Demp	Consensus
P1	0.99	0.00	0.000	0.0000	0.495
P2	0.01	0.01	0.000	0.0000	0.010
P3	0.00	0.99	0.000	0.0000	0.495
Any	0.00	0.00	0.000	0.0000	0.000
Fy	0.00	0.00	0.000	0.0000	0.000



C:\Users\Zarana\Desktop\PROJEC~1\MATHCAL2.EXE

Value for P1 : 0.98
Value for P2 : 0.01
Value for P3 : 0.00
Value for Any : 0.01
Value for Fy : 0.00

Enter Values for M2
Value for P1 : 0.00
Value for P2 : 0.01
Value for P3 : 0.98
Value for Any : 0.01
Value for Fy : 0.00

	M1	M2	Demp	non-Demp	Consensus
P1	0.98	0.00	0.492	0.0098	0.492
P2	0.01	0.01	0.010	0.0002	0.010
P3	0.00	0.98	0.492	0.0098	0.492
Any	0.01	0.01	0.000	0.0000	0.005
Fy	0.00	0.00	0.000	0.0000	0.000

Conclusion

In this paper we have focused on the problem of combining highly conflicting and dogmatic beliefs within the belief functions theory. The opinion metric described here provides a simple and compact notation for beliefs in the Shaferian belief model.We have presented an alternative to Dempster's rule which is consistent with probabilistic and statistical analysis, and which seems more suitable for combining highly conflicting beliefs as well as for combining harmonious beliefs, than Dempster's rule and its non-normalised version. The fact that a binary focused frame of discernment must be derived in order to apply the consensus operator puts no restriction on its applicability. The resulting beliefs for each event can still be compared and can form the basis for decision making.

Refrences:

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

[2] A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.

[3] A. Jøsang and V.A. Bondi. Legal Reasoning with Subjective Logic. *Artificial Reasoning and Law*, 8(4):289–315, winter 2000.

[4] A. Jøsang. An Algebra for Assessing Trust in Certification Chains. In J. Kochmar, editor, *Proceedings of the Network and Distributed Systems Security Symposium (NDSS'99)*. The Internet Society, 1999

[5] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*,66:191–234, 1994..