An Approximation of Basic Assignment Probability in Dempster-Shafer Theory

Ion IANCU

Faculty of Mathematics and Computer Science,
Department of Computer Science,
University of Craiova, Romania
i_iancu@yahoo.com

Abstract. The computational complexity of reasoning with the Dempster-Shafer theory of evidence is one of the main criticism of this formalism. A possibility to overcame this difficulty is to reduce the number of focal elements; various algorithms have been suggested with this goal. We propose a combination of three such algorithms in order to obtain a new one.
Keywords: basic assignment probability, mass function, Dempster rule
Math. Subject Classification 2000: 64T30; 03B42

1 Introduction

Information can be incomplete, imprecise, vague, contradictory or deficient in many ways. In general, these types of deficiencies of information produce different types of uncertainty. It is difficult to avoid uncertainty when attempting to make models of the real world. Uncertainty is inherent to natural phenomena and it is impossible to create a perfect representation of reality. Classic mathematics deals with ideal worlds where perfect geometric figures exist and can verify various conditions. The works of Zadeh in fuzzy sets [9] and Dempster [2] in belief functions have been represented new possibilities for working with uncertain information. Belief functions offer a non Bayesian method for quantifying subjective evaluations by using probability. In 1970s, it was further development by Shafer, whose book [5] remain a classic in belief functions, the so called Theory of Evidence. These theory has been also called the Dempster-Shafer Theory (DST). In the 1980s, the scientific community working in Artificial Intelligence got involved in using DST in various applications and the process of mathematical formalization have been continued and it continue today, too. There are many publications and congresses dedicated to uncertainty and its related fields. The increasing of computational power has offered new possibilities of working with uncertain information in various areas of science and technology: mathematics, engineering, medicine, business, social science. The computational of complexity of reasoning within the DST is one of the main criticism that accompanies this method. Orponen [4] shows that the combination of two basic probability assignment (bpa's) using Dempster's rule is #P complete (the class #P is functional analogue of the class NP of decision problems). To overcame this difficulty, various approximation methods have been suggested. Because the number of focal elements heavily influences the computational complexity of combination various independent pieces of evidence, these algorithms give approximations of bpa values by removing focal elements and/or redistributing the corresponding numerical values. In this paper we propose a new method obtained by combining three known algorithms.

2 Approximation algorithms

Dempster-Shafer Theory aim to provide a theory of partial belief. It attempts to overcame the representational deficiencies within the probability theory as well as to provide some mechanisms for making inferences from the available evidence. The frame of discernment Ω is the set of mutually exclusive and exhaustive propositions of interest. An important role in Dempster-Shafer theory is played by the basic probability assignment (bpa) or the mass function, denoted by m.

Definition 1. *The basic probability assignment associated with a frame of discernment Ω is a function*

$$m : 2^{\Omega} \rightarrow [0, 1]$$

that assigns a numerical value to each subset of Ω and satisfies the following properties:

$$m(\emptyset) = 0$$

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

Definition 2. *Subsets $A \subseteq \Omega$ with $m(A) > 0$ are called the focal elements.*

Two bpa's m_1 and m_2 can be combined using the Dempster's rule [2]

$$(m_1 \oplus m_2)(A) = \frac{\sum_{A_1 \bigcap A_2 = A} m_1(A_1) m_2(A_2}{\sum_{A_1 \bigcap A_2 \neq \emptyset} m_1(A_1) m_2(A_2)}.$$

Given a frame of discernment of size $|\Omega| = N$, a bpa m can have up to 2^N focal elements all of which have to be represented explicitly to capture the complete information encoded in m. It results that the combination of two bpa,s requires computation of up to 2^{N+1} intersections. Orponen [4] showed that the combination of various pieces of evidence using Dempster's rule has a #P complexity. Reducing the number of focal elements of the bpa's under consideration while retaining the essence of the information is an important problem for Dempster-Shafer theory. The most important algorithms known in the literature in order to solve this problem are the following.

2.1 The Bayesian approximation

This approximation [8] reduces a given bpa m to a probability distribution m_B

$$m_B(A) = \begin{cases} \dfrac{\sum_{S/A\subseteq S} m(S)}{\sum_{C/C\subseteq \Omega} |C| m(C)} & if \quad |A| = 1 \\ 0 & otherwise \end{cases}$$

Example 1 *Let m be a bpa over the frame of discernment $\Omega = \{a, b, c, d, e\}$ with the values*

$$m(A) = \begin{cases} 0.33 & if \quad A = \{a, b\} \\ 0.3 & if \quad A = \{a, b, c\} \\ 0.27 & if \quad A = \{b, c, d\} \\ 0.06 & if \quad A = \{c, d\} \\ 0.04 & if \quad A = \{d, e\} \end{cases}$$

Applying the Bayesian approximation to m yields the following result

$$m_B(A) \approx \begin{cases} 0.245 & if \quad A = \{a\} \\ 0.35 & if \quad A = \{b\} \\ 0.245 & if \quad A = \{c\} \\ 0.143 & if \quad A = \{d\} \\ 0.015 & if \quad A = \{e\} \end{cases}$$

This example shows that the Bayesian approximation is not reasonable in the cases when the number of focal elements of the input bpa is $\leq |\Omega|$.

2.2 The k-l-x method

The basic idea of this approximation [7] is to incorporate into the approximation m_{klx} only at least k and at most l focal elements with the highest values in the original bpa and having the sum of the m-values at least $1 - x$, where $x \in [0, 1)$. Finally, the values from the approximation are normalized in order to guarantee the basic properties of bpa.

Example 2 *For the bpa m given in the previous example and the values $k = 2, l = 3$ and $x = 0.1$ the following result is obtained*

$$m_{klx}(A) = \begin{cases} 11/30 \approx 0.366 & if \quad A = \{a, b\} \\ 1/3 \approx 0.333 & if \quad A = \{a, b, c\} \\ 3/10 = 0.3 & if \quad A = \{b, c, d\} \end{cases}$$

2.3 Summarization

This method [3] works likewise as klx. Let k be the number of focal elements to be contained in the approximation m_S of a given bpa m. M denotes the set of the $k-1$ subsets of Ω with the highest value in m. Then m_S is given by

$$m_S(A) = \begin{cases} m(A) & if \quad A \in M \\ \sum_{A' \subseteq A, A' \notin M} m(A') & if \quad A = A_0 \\ 0 & otherwise \end{cases}$$

where A_0 is defined as

$$A_0 = \bigcup_{A' \notin M, m(A') > 0} A'.$$

Example 3 *For the bpa m from the Example 1 and $k = 3$, m_S has the following values*

$$m_S(A) = \begin{cases} 0.33 & if \quad A = \{a, b\} \\ 0.3 & if \quad A = \{a, b, c\} \\ 0.37 & if \quad A = \{b, c, d, e\} \end{cases}$$

2.4 The D1 approximation

Let m a bpa to be approximated and k the desired number of focal elements of the approximated bpa m_D. The following notations are usefully:

a) M^+ is the set of $k-1$ focal elements of m with the highest values

$$M^+ = \{A_1, ..., A_{k-1} \subseteq \Omega / \forall A \notin M^+ : m(A_i) \geq m(A), i = 1, 2, ..., k-1\}$$

b) M^- is the set containing all other focal elements of m:

$$M^- = \{A \subseteq \Omega / m(A) > 0, A \notin M^+\}.$$

Given a focal element $A \in M^-$ of m the collection M_A of supersets of A is computed; if M_A is empty (i.e. M^+ contains no supersets of A) then the set $M_A^{'}$ is computed:

$$M_A^{'} = \{B \in M^+ / |B| \geq |A|, B \cap A \neq \emptyset\},$$

where $|A|$ represents the cardinality of the set A.
The ideas of the D1 algorithm are [1]:

i) all the members of M^+ are kept as focal elements of m_D;

ii) for every $A \in M^-$, the value $m(A)$ is distributed uniformly among the members of M_A with the smallest cardinality;

iii) if M_A is empty then the value $m(A)$ is shared among the smallest members of $M_A^{'}$ and the value to be assigned to a focal element depends on the size of its intersection with A.

iv) the procedure of distribution masses is invoked recursively until all of $m(A)$ are assigned to the members of M^+ or the set $M_A^{'}$ becomes empty. In this case, the remaining value is assigned to Ω which thus becomes a focal element of m_D.

The approximation m_D of a bpa with n focal elements can be computed in time $O(k(n-k))$.

Example 4 *For the bpa m from the Example 1 and $k = 3$, the algorithm $D1$ yields the following values*

$$m_D(A) = \begin{cases} 0.33 & if \quad A = \{a, b\} \\ 0.51 & if \quad A = \{a, b, c\} \\ 0.16 & if \quad A = \{a, b, c, d, e\} \end{cases}$$

3 A mixed algorithm

An analysis of the approximation of the original bpa is made in [1] and the conclusion is: the "best" approximation algorithm with respect to decision making does not exist. However, the $k-l-x$, $D1$ and Bayesian approximations yield definitely better results that the summarization does.

In this paper we propose a new algorithm obtained as a combination between the $k-l-x$, summarization and $D1$ algorithms. Let m be the bpa to be approximated; this combination is constructed in three steps, to obtain a new approximation m_M:

S1) Given the parameters k, l, x, having the same signification as in the $k-l-x$ method, we keep at least k and at most l focal elements, from the original bpa, with the sum of m-values at least $1-x$; let M be the set of these focal elements.

S2) The set M is considered as the set M^+ from the $D1$ algorithm. The focal elements not included in the set M at the step S1 play the role of M^- set in the $D1$ algorithm.

S3) The components of all focal elements $A \in M^-$ not distributed among the members of M_A or $M_A^{'}$ are included in a new focal element; the m_M-value of this set is computed as sum of m-values of its components. This idea is used by the summarization method to construct the set A_0.

Example 5 *For the bpa m from the Example 1 and $k = 3$, $l = 2$ and $x = 0.1$, the algorithm M yields the following values:*

Step S1): Removing $\{c, d\}$ and $\{d, e\}$ from m, the constraints concerning the number of focal elements and the numerical mass deleted are satisfied. Thus, the following approximation is obtained

$$m_{M1}(A) = \begin{cases} 0.33 & if \quad A = \{a, b\} \\ 0.3 & if \quad A = \{a, b, c\} \\ 0.27 & if \quad A = \{b, c, d\} \end{cases}$$

Steps S2 and S3: The Step S2 is applied with the following sets parameters:

$$M^+ = \{A_1, A_2, A_3\}, M^- = \{A_4, A_5\}$$

$$m_{M2}(A_1) = m_{M2}(\{a,b\}) = m_{M1}(\{a,b\}) = 0.33$$

$$m_{M2}(A_2) = m_{M2}(\{a,b,c\}) = m_{M1}(\{a,b,c\}) = 0.3$$

$$m_{M2}(A_3) = m_{M2}(\{b,c,d\}) = m_{M1}(\{b,c,d\}) = 0.27$$

$$m_{M2}(A_4) = m_{M2}(\{c,d\}) = m(\{c,d\}) = 0.06$$

$$m_{M2}(A_5) = m_{M2}(\{d,e\}) = m(\{d,e\}) = 0.04.$$

The set $A_3 \in M^+$ is the unique superset of $A_4 \in M^-$, such that the value of A_3 is increased by 0.06. Furthermore A_3 covers half of the elements of A_5 which adds up another $0.4/2 = 0.02$ to m_M value of A_3. The rest is assigned to $\{e\}$, the set constructed in the Step S3. The approximation m_M of the original bpa m is:

$$m_M(A) = \begin{cases} 0.33 & if \quad A = \{a,b\} \\ 0.3 & if \quad A = \{a,b,c\} \\ 0.35 & if \quad A = \{b,c,d\} \\ 0.02 & if \quad A = \{e\} \end{cases}$$

An analysis of the error measure associated to an approximation algorithm can be made using the *pignistic probability* P induced by a bpa that can be considered the standard function for decision making in Dempster-Shafer Theory [6]. It is given by

$$P(\{x\}) = \sum_{A/x \in A \subseteq \Omega} \frac{m(A)}{|A|}.$$

The error quantifies the maximal deviation in the pignistic probability induced by an approximated bpa. Let P_0 be the pignistic probability induced by the original version of a bpa m and $P_{m'}$ the one induced by the approximation m'. Then the error measure is defined as

$$Error(m') = \sum_{A \subseteq \Omega} |P_0(A) - P_{m'}(A)|.$$

For the approximations from the previous examples, we obtain:

$$Error(m_S) = 0.3225,$$

$$Error(m_{klx}) = 0.19,$$

$$Error(m_D) = 0.466,$$

$$Error(m_M) = 0.186.$$

One observes that the best result is given by our approximation m_M. We notice that, in all experiments the mixed algorithm gave better results than the D1 algorithm from which it is derived. If we increase the number of focal elements of the approximation algorithms the error decreases, because a greater number of focal elements from the original version of the bpa and the approximated bpa coincide.

The mixed algorithm can be implemented as follows

input: m, k, l, x; **output**: m_D

P1) S:= focal elements of bpa m, sorted in decreasing order w. r. t. m -values

P2) keep the focal elements of m that satisfied the condition
$(nf < l)$ and $((nf < k)$ or $(tmass < 1 - x))$
where nf and $tmass$ are the number and the total mass of these focal sets

P3) $M^+ :=$ the sets given by P2
$M^- := S \setminus M^+$
$m_M(A) := m(A)\ \forall A \in M^+$
$R := \emptyset$
$m_M(R) := 0$
for all $A \in M^-$
do
$M_A := \{B \in M^+ / A \subset B\}$
if $M_A \neq \emptyset$
then
$M'_A := \{B \in M_A / |B|$ is minimal in $M_A\}$
for all $B \in M'_A$
do
$$m_M(B) := m_M(B) + \frac{m(A)}{|M'_A|}$$
end do
else
$N_A = \{B \in M^+ / |B| \geq |A|, A \cap B \neq \emptyset\}$
if $N_A = \emptyset$
then
$R := R \cup A$
$m_M(R) := m_M(R) + m(A)$
else
$N'_A := \{B \in N_A / |B|$ is minimal in $N_A\}$
$N'_A = \{B_1, ..., B_n\}$
for all $a \in A$ let $n_1 =$ the number of the sets $B_i \in N'_A$:
$a \in A \cap B_i$

$$m_M(B_i) := m_M(B_i) + \frac{m(A)}{|A| \cdot n_1}$$

for all $b \in A$ with $b \notin (A \cap \cup_{i=1}^{n} B_i)$
do
$R := R \cup \{b\}$
$$m_M(R) := m_M(R) + \frac{m(A)}{|A|}$$
end do
end if
end if
end do

4 Conclusions

Starting from an original basic assignment probability m, we propose an approximate version obtained as a combination of three very known approximation algorithms. From the examples presented, it results that in many cases the proposed algorithm gives better results.

References

[1] **M. Bauer**, Approximations for Decision Making in the Dempster-Shafer Theory of Evidence, UAI 1996, 73-80
[2] **A. Dempster**, Uper and lower probabilities induced by multivalued mapping, Annals of Mathematical Statistics, 38, 1967, 325-339
[3] **J. Lowrance, T. Garvey, T. Strat**, A Framework for Evidential-Reasoning Systems, In Proc. of the 5th Nat. Conf. of the American Association for Artificial Intelligence, 1986, 896-903
[4] **P. Orponen**, Dempster's Rule of Combination is # P-complete, Artificial Intelligence, 44, 1990, 245-253
[5] **G. Shafer**, A Mathematical Theory of Evidence, Princeton Univ. Press, NJ, 1976
[6] **P. Smets**, Belief Functions versus Probability Functions, in B. Buchon, L. Saitta, R. Yager (Eds.): Uncertainty and Intelligent Systems, vol. 313 of Lecture Notes in Computer Science, Springer, 1988, 17-24
[7] **B. Tessem**, Approximations for efficient computation in the theory of evidence, Artificial Intelligence, 61, 1993, 315-329
[8] **F. Voorbraak**, A Computationally Efficient Approximation Dempster-Shafer Theory, Int. Journal of Man-Machine Studies, 30, 1989, 525-535
[9] **L. A. Zadeh**, Fuzzy Sets as a Basis for a Theory of Possibility, Fuzzy Sets and Systems, 1(1), 1978, 3-28