Hierarchical Proportional Redistribution for bba Approximation

Jean Dezert
Deqiang Han
Zhunga Liu
Jean-Marc Tacnet

Originally published as Dezert J., Han D., Liu Z., Tacnet J.-M., *Hierarchical proportional redistribution for bba approximation*, in Belief 2012, Compiègne, May 2012, and reprinted with permission.

Abstract. Dempster's rule of combination is commonly used in the field of information fusion when dealing with belief functions. However, it generally requires a high computational cost. To reduce it, a basic belief assignment (bba) approxima-tion is needed. In this paper we present a new bba approximation approach called hierarchical proportional redistribution (HPR) allowing to approximate a bba at any given level of non-specificity. Two examples are given to show how our new HPR works.

1 Introduction

Dempster-Shafer Theory (DST), also called Theory of Evidence [10], has been widely used in many applications, e.g., information fusion, pattern recognition and decision making [11]. Although it is appealing in uncertainty modeling, while appearing more controversial for consistent reasoning, the high computational cost remains problematic which is often raised against its use [11]. To resolve such a problem, three major types of approaches have been proposed.

The first is to propose efficient procedures for performing exact computations [1, 8]. The second is composed of Monte-Carlo techniques [9]. The third is to approximate a belief function to a simpler one. The papers of Voorbraak [13], Dubois and Prade [5] are seminal works of this type. Other representative works include $k-l-x$ [3] and k-additive belief function [2, 6]. Denœux uses hierarchical clustering to implement the inner and outer approximation [3].

In this paper, we propose a new method called hierarchical proportional redistribution (HPR) to approximate any general basic belief assignment (bba) at a given level of non-specificity [4], up to the ultimate level 1 corresponding to a Bayesian bba [10]. The level of non-specificity can be controlled by the users through the adjustment of the maximum cardinality of remaining focal elements. For the approximated bba obtained by HPR, the maximum cardinality of the focal elements is k. Thus HPR can be considered as a generalized k-additive belief approximation. Some examples are given to show how our proposed HPR method works, and to compare it with other approximations.

2 Basics of Dempster-Shafer Theory (DST)

In DST [10], the frame of discernment (FoD) is a set Θ of mutual exhaustive and exclusive elements. $m(.) : 2^{\Theta} \to [0,1]$ is a basic belief assignment (bba), also called mass function, if it satisfies

$$\sum_{A\subseteq\Theta} m(A) = 1,\ m(\emptyset) = 0. \tag{1}$$

Belief function (Bel) and plausibility function (Pl) are defined as

$$Bel(A) = \sum_{B\subseteq A} m(B) \quad \text{and} \quad Pl(A) = \sum_{A\cap B\neq\emptyset} m(B). \tag{2}$$

Suppose that $m_1, m_2, ..., m_n$ are n bba's, then Dempster's rule of combination is defined by

$$m(A) = \begin{cases} 0,\ A = \emptyset \\ \dfrac{\sum\limits_{\cap A_i = A} \prod\limits_{1\leq i\leq n} m_i(A_i)}{\sum\limits_{\cap A_i \neq \emptyset} \prod\limits_{1\leq i\leq n} m_i(A_i)},\ A \neq \emptyset \end{cases} \tag{3}$$

This rule is used in DST to combine pieces of evidence expressed by bba's. As referred above, Dempster's combination has high computational cost and three types of approaches have been proposed to reduce it. We prefer belief approximation approaches [2, 3, 6, 12] since they both reduce the computational cost of the combination and allow to deal with smaller-size focal elements, which is more intuitive for human to catch the meaning and interpret fusion results [2].

3 Two bba Approximation Approaches

1) $k-l-x$ **approximation:** This was proposed by Tessem [12]. The simplified bba obtained by $k-l-x$ approach satisfies: a) keep no less than k focal elements; b) keep no more than l focal elements; c) the mass assignment to be deleted is no greater than x. In $k-l-x$, the focal elements of a original bba are sorted by their masses. Such an algorithm chooses the first p focal elements such that $k \leq p \leq l$ and such that the sum of the masses of these first p focal elements is no less than $1-x$. The deleted masses are redistributed to the other focal elements through a normalization.

2) k-additive belief function approximation: Given $m(.): 2^{\Theta} \to [0,1]$, one kind of k-additive belief function [2, 6] induced by the mass $m(.)$ is defined by

$$\begin{cases} m_k(B) = m(B) + \sum_{A \supset B, A \subseteq \Theta, |A|>k} \frac{m(A)\cdot|B|}{\mathcal{N}(|A|,k)}, & \forall |B| \leq k \\ m_k(B) = 0, & \forall |B| > k \end{cases} \tag{4}$$

where $B \subseteq \Theta$ and

$$\mathcal{N}(|A|,k) = \sum_{j=1}^{k} \binom{|A|}{j} \cdot j = \sum_{j=1}^{k} \frac{|A|!}{(j-1)!(|A|-j)!} \tag{5}$$

is the average cardinality of the subsets of A of size at most k. For k-additive belief approximation, the maximum cardinality of available focal elements is no greater than k. Other bba approximation methods can be found in related references.

4 Hierarchical Proportional Redistribution Approximation

In this paper we propose a new bba approximation approach called hierarchical proportional redistribution (HPR), which provides a new way to reduce step-by-step the mass committed to uncertainties. Ultimately an approximate measure of subjective probability can be obtained if needed, i.e. a so-called Bayesian bba in [10]. Our proposed procedure can be stopped at any step in the process and thus it allows to reduce the number of focal elements of a given bba in a simple manner to diminish the size of the core [10] of a bba. Thus we can reduce the complexity (if needed) when applying also some complex rules of combinations. By using HPR, we can obtain approximate bba's at any different non-specificity level that we want. Let us first introduce two new notations for convenience and conciseness:

1. Any element of cardinality $1 \leq k \leq n$ of the power set 2^{Θ} will be denoted $X(k)$ by convention. For example, if $\Theta = \{A,B,C\}$, then $X(2)$ can denote the following partial uncertainties $A \cup B$, $A \cup C$ or $B \cup C$, and $X(3)$ denotes the total uncertainty $A \cup B \cup C$.
2. The proportional redistribution factor (ratio) of width s involving elements X and Y of the powerset is defined by (for $X \neq \emptyset$ and $Y \neq \emptyset$)

$$R_s(Y,X) \triangleq \frac{m(Y) + \varepsilon \cdot |X|}{\sum_{\substack{Y \subset X \\ |X|-|Y|=s}} m(Y) + \varepsilon \cdot |X|} \tag{6}$$

   where ε is a small positive number introduced here to deal with particular cases where $\sum_{\substack{Y \subset X \\ |X|-|Y|=s}} m(Y) = 0$.
   By convention, we will denote $R(Y,X) \triangleq R_1(Y,X)$ when we use the proportional redistribution factors of width $s=1$, as we use in this paper for this HPR method.

The HPR is a step-by-step (recursive) proportional redistribution of the mass $m(X(k))$ of a given uncertainty $X(k)$ (partial or total) of cardinality $2 \leq k \leq n$ to all the least specific elements of cardinality $k-1$, i.e., to all possible $X(k-1)$, until $k=2$ is reached. The proportional redistribution is done from the masses of belief committed to $X(k-1)$ as done classically in DSmP transformation. The "hierarchical" masses $m_h(.)$ are recursively (backward) computed as follows. Here $m_{h(k)}$ represents the approximate bba obtained at the step $n-k$ of HPR, i.e., it has the maximum focal element cardinality of k.

$$\begin{aligned} &m_{h(n-1)}(X(n-1)) = m(X(n-1)) + \sum_{\substack{X(n)\supset X(n-1),\\ X(n),X(n-1)\in 2^{\Theta}}} [m(X(n)) \cdot R(X(n-1),X(n))]; \\ &m_{h(n-1)}(A) = m(A), \forall |A| < n-1 \end{aligned} \tag{7}$$

$m_{h(n-1)}(\cdot)$ is the bba obtained at the first step of HPR ($n-(n-1)=1$), the maximum focal element cardinality of $m_{h(n-1)}$ is $n-1$.

$$\begin{aligned} m_{h(n-2)}(X(n-2)) = {} & m(X(n-2)) \\ & + \sum_{\substack{X(n-1)\supset X(n-2)\\ X(n-2),X(n-1)\in 2^{\Theta}}} [m_{h(n-1)}(X(n-1)) \cdot R(X(n-2),X(n-1))] \\ m_{h(n-2)}(A) = {} & m_{h(n-1)}(A), \forall |A| < n-2 \end{aligned} \tag{8}$$

$m_{h(n-2)}(\cdot)$ is the bba obtained at the second step of HPR ($n-(n-2)=2$), the maximum focal element cardinality of $m_{h(n-2)}$ is $n-2$.

This hierarchical proportional redistribution process can be applied similarly (if one wants) to compute $m_{h(n-3)}(.)$, $m_{h(n-4)}(.)$, ..., $m_{h(2)}(\cdot)$, $m_{h(1)}(\cdot)$ with

$$\begin{aligned} &m_{h(2)}(X(2)) = m(X(2)) + \sum_{\substack{X(3)\supset X(2)\\ X(3),X(2)\in 2^{\Theta}}} [m_{h(3)}(X(3)) \cdot R(X(2),X(3))] \\ &m_{h(2)}(A) = m_{h(3)}(A), \forall |A| < n-2 \end{aligned} \tag{9}$$

$m_{h(2)}(\cdot)$ is the bba obtained at the first step of HPR ($n-2$), the maximum focal element cardinality of $m_{h(2)}$ is 2.

Mathematically, for any $X(1) \in \Theta$, i.e. any $\theta_i \in \Theta$ a Bayesian belief function can be obtained by HPR method in deriving all possible steps of proportional redistributions of partial ignorances in order to get

$$m_{h(1)}(X(1)) = m(X(1)) + \sum_{\substack{X(2)\supset X(1)\\ X(1),X(2)\in 2^{\Theta}}} [m_{h(2)}(X(2)) \cdot R(X(1),X(2))] \tag{10}$$

In fact, $m_{h(1)}(\cdot)$ is a probability transformation, called here the Hierarchical DSmP (HDSmP). Since $X(n)$ is unique and corresponds only to the full ignorance $\theta_1 \cup \theta_2 \cup \ldots \cup \theta_n$, the expression of $m_h(X(n-1))$ in Eq.(9) just simplifies as

$$m_{h(n-1)}(X(n-1)) = m_h(X(n-1)) + m(X(n)) \cdot R(X(n-1),X(n)) \tag{11}$$

For the full proportional redistribution of the masses of uncertainties to the elements least specific involved in these uncertainties, no mass is lost during the step-by-step hierarchical process and thus at any step of HPR, the sum of masses is kept to one.

5 Examples

5.1 Example 1

Let's consider the following bba over $\Theta = \{\theta_1, \theta_2, \theta_3\}$:

$$m(\theta_1) = 0.10, \quad m(\theta_2) = 0.17, \quad m(\theta_3) = 0.03, \quad m(\theta_1 \cup \theta_2) = 0.15,$$
$$m(\theta_1 \cup \theta_3) = 0.20, \quad m(\theta_2 \cup \theta_3) = 0.05, \quad m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30.$$

We apply the HPR with $\varepsilon = 0$ in this example because there is no mass of belief equal to zero. It can be verified that the result obtained with small positive ε parameter remains (as expected) numerically very close to what is obtained with $\varepsilon = 0$.

• **Step 1:** The first step of HPR consists in redistributing back $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$ committed to the full ignorance to the elements $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ only, because these elements are the only elements of cardinality 2 that are included in $\theta_1 \cup \theta_2 \cup \theta_3$. Applying the Eq. (8) with $n = 3$, one gets when $X(2) = \theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_1 \cup \theta_2$ the following masses.

$$m_{h(2)}(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_2) + m(X(3)) \cdot R(\theta_1 \cup \theta_2, X(3)) = 0.15 + (0.30 \cdot 0.375) = 0.2625$$

because $R(\theta_1 \cup \theta_2, X(3)) = \frac{0.15}{0.15+0.20+0.05} = 0.375$.
Similarly, one gets

$$m_{h(2)}(\theta_1 \cup \theta_3) = m(\theta_1 \cup \theta_3) + m(X(3)) \cdot R(\theta_1 \cup \theta_3, X(3)) = 0.20 + (0.30 \cdot 0.5) = 0.35$$

because $R(\theta_1 \cup \theta_3, X(3)) = \frac{0.20}{0.15+0.20+0.05} = 0.5$, and also

$$m_{h(2)}(\theta_2 \cup \theta_3) = m(\theta_2 \cup \theta_3) + m(X(3)) \cdot R(\theta_2 \cup \theta_3, X(3)) = 0.05 + (0.30 \cdot 0.125) = 0.0875$$

because $R(\theta_2 \cup \theta_3, X(3)) = \frac{0.05}{0.15+0.20+0.05} = 0.125$.

• **Step 2** Now, we go to the next step of HPR principle and one needs to redistribute the masses of partial ignorances $X(2)$ corresponding to $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ back to the singleton elements $X(1)$ corresponding to θ_1, θ_2 and θ_3. We use Eq. (10) for doing this as follows:

$$\begin{aligned} m_{h(1)}(\theta_1) &= m(\theta_1) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_1, \theta_1 \cup \theta_2) \quad + m_h(\theta_1 \cup \theta_3) \cdot R(\theta_1, \theta_1 \cup \theta_3) \\ &\approx 0.10 + (0.2625 \cdot 0.3703) + (0.35 \cdot 0.7692) = 0.10 + 0.0972 + 0.2692 = 0.4664 \end{aligned}$$

because $R(\theta_1, \theta_1 \cup \theta_2) = \frac{0.10}{0.10+0.17} \approx 0.3703$ and $R(\theta_1, \theta_1 \cup \theta_3) = \frac{0.10}{0.10+0.03} \approx 0.7692$
Similarly, one gets

$$\begin{aligned} m_{h(1)}(\theta_2) &= m(\theta_2) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_2, \theta_1 \cup \theta_2) + m_h(\theta_2 \cup \theta_3) \cdot R(\theta_2, \theta_2 \cup \theta_3) \\ &\approx 0.10 + (0.2625 \cdot 0.6297) + (0.0875 \cdot 0.85) = 0.17 + 0.1653 + 0.0744 = 0.4097 \end{aligned}$$

because $R(\theta_2, \theta_1 \cup \theta_2) = \frac{0.17}{0.10+0.17} \approx 0.6297$ and $R(\theta_2, \theta_2 \cup \theta_3) = \frac{0.17}{0.17+0.03} = 0.85.$ and also

$$\begin{aligned} m_{h(1)}(\theta_3) &= m(\theta_3) + m_h(\theta_1 \cup \theta_3) \cdot R(\theta_3, \theta_1 \cup \theta_3) + m_h(\theta_2 \cup \theta_3) \cdot R(\theta_3, \theta_2 \cup \theta_3) \\ &\approx 0.03 + (0.35 \cdot 0.2307) + (0.0875 \cdot 0.15) = 0.03 + 0.0808 + 0.0131 = 0.1239 \end{aligned}$$

because $R(\theta_3, \theta_1 \cup \theta_3) = \frac{0.03}{0.10+0.03} \approx 0.2307$ and $R(\theta_3, \theta_2 \cup \theta_3) = \frac{0.03}{0.17+0.03} = 0.15$
Hence, the result of final step of HPR is:

$$m_{h(1)}(\theta_1) = 0.4664, \quad m_{h(1)}(\theta_2) = 0.4097, \quad m_{h(1)}(\theta_3) = 0.1239.$$

We can easily verify that $m_{h(1)}(\theta_1) + m_{h(1)}(\theta_2) + m_{h(1)}(\theta_3) = 1$.

To compare HPR with the approach of $k-l-x$, we set the parameters of $k-l-x$ to obtain bba's with equal focal element number with HPR at each step. In Example 1, for HPR at first step, it can obtain a bba with 6 focal elements. Thus we set $k=l=6, x=0.4$ for $k-l-x$ to obtain a bba with 6 focal elements. Similarly, for HPR at second step, it can obtain a bba with 3 focal elements. Thus we set $k=l=3, x=0.4$ for $k-l-x$. Based on HPR and $k-l-x$, the results are shown in Table 1.

Table 1 Experimental results of Example 1.

Focal elements	$m_{h(k)}(\cdot)$ - approximate bba			$m(\cdot)$ obtained by $k-l-x$	
	$k=3$	$k=2$	$k=1$	$k=l=6$	$k=l=3$
θ_1	0.1000	0.1000	0.4664	0.1031	0.0000
θ_2	0.1700	0.1700	0.4097	0.1753	0.2573
θ_3	0.0300	0.0300	0.1239	0.0000	0.0000
$\theta_1 \cup \theta_2$	0.1500	0.2625	0.0000	0.1546	0.0000
$\theta_1 \cup \theta_3$	0.2000	0.3500	0.0000	0.2062	0.2985
$\theta_2 \cup \theta_3$	0.0500	0.0875	0.0000	0.0515	0.0000
$\theta_1 \cup \theta_2 \cup \theta_3$	0.3000	0.0000	0.0000	0.3093	0.4478

5.2 *Example 2*

Let's consider $\Theta = \{\theta_1, \theta_2, \theta_3\}$, and the bba $m(\theta_3) = 0.7$ and $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$. Here, the masses of all the focal elements with cardinality size 2 equal to zero. For HPR, when $\varepsilon > 0$, $m(\theta_1 \cup \theta_2 \cup \theta_3)$ will be divided equally and redistributed to $\{\theta_1 \cup \theta_2\}$, $\{\theta_1 \cup \theta_3\}$ and $\{\theta_2 \cup \theta_3\}$. Because the ratios are (taking for example $\varepsilon = 0.001$)

$$R(\theta_1 \cup \theta_2, X(3)) = R(\theta_1 \cup \theta_3, X(3)) = R(\theta_2 \cup \theta_3, X(3)) = \frac{0.00 + 0.001 \cdot 3}{(0.00 + 0.001 \cdot 3) \cdot 3} = 0.3333$$

In this case, HPR cannot work directly when $\varepsilon = 0$. This shows the necessity for the use of $\varepsilon > 0$. The bba's obtained from $\text{HPR}_{\varepsilon=0.001}$ and $k-l-x$ are listed in Table 2.

From the results of Examples 1 & 2, we can see that based on $k-l-x$, the users can control the number of focal elements but cannot control the maximum cardinality of focal elements. Although based on $k-l-x$, the number of focal elements can be reduced, the focal elements with big cardinality might also be kept. This is not good for further reducing computational cost. But with the proposed HPR method, users can easily control both the non-specificity of approximated bba's and the focal element's size.

Table 2 Experimental results of Example 2 ($\varepsilon = 0.001$)

Focal elements	$m_{h(k)}(\cdot)$ - approximate bba			$m(\cdot)$ obtained by $k-l-x$	
	$k=3$	$k=2$	$k=1$	$k=l=6$	$k=l=3$
θ_1	0.0000	0.0000	0.0503	0.0000	0.0000
θ_2	0.0000	0.0000	0.0503	0.0000	0.0000
θ_3	0.7000	0.7000	0.8994	0.7000	0.7000
$\theta_1 \cup \theta_2$	0.0000	0.1000	0.0000	0.0000	0.0000
$\theta_1 \cup \theta_3$	0.0000	0.1000	0.0000	0.0000	0.0000
$\theta_2 \cup \theta_3$	0.0000	0.1000	0.0000	0.0000	0.0000
$\theta_1 \cup \theta_2 \cup \theta_3$	0.3000	0.0000	0.0000	0.3000	0.3000

5.3 *Example 3*

In this work, an approximation method 1 (giving $m_1(.)$) is considered better than a method 2 (giving $m_2(.)$) if both conditions are fulfilled: 1) if the distance between $m_1(.)$ and original bba $m(.)$ is smaller than the distance between $m_2(.)$ and original bba $m(.)$, i.e. $d(m_1,m) < d(m_2,m)$; 2) if the approximate non-specificity value $U(m_1)$ is closer (and lower) to the true non-specificity value $U(m)$ than $U(m_2)$. We have used Jousselme's distance [7] which has been proved recently to be a strict distance metric because it is commonly used in applications. The Non-specificity [4] is given by $U(m) = \sum_{A\subseteq\Theta} m(A)\log_2|A|$. In this example, we make a comparison between HPR (method 1) and k-additive approach (method 2). We have taken $\Theta = \{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\}$ and generated randomly 30 bba's using the algorithm given in [7]. We compute and plot $d(m_1,m)$, $d(m_2,m)$, $U(m)$, $U(m_1)$ and $U(m_2)$ for several levels of approximation. The results are shown in Fig. 1 and indicate clearly the superiority of HPR over the k-additive approach.

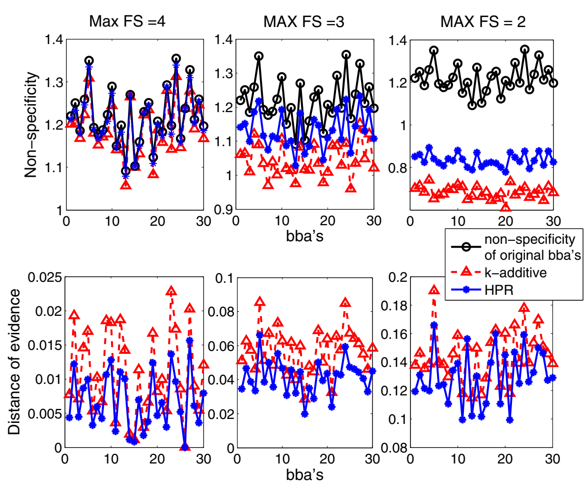

Fig. 1 Results for the Example 3. Comparison of k-additive belief function approximation with HPR approximation method. (FS means Focal element Size)

6 Conclusions

In this paper, a novel bba approximation called HPR has been proposed as an interesting alternative approach to two classical ones. With this HPR, the non-specificity degree can be easily controlled by the users. Our example show its behavior and advantage in comparisons with other well-known bba approximation approaches. HPR has a low computational cost compared with *k*-additive approach, which will be discussed in a more detailed paper in future. In further works, we will also compare our proposed HPR with more bba approximation approaches available in the literature. In this paper, we have used only the distance of evidence and the non-specificity criteria, which in fact are not enough, or comprehensive to evaluate efficiently bba approximations. So in future, we will try to propose more efficient evaluation criteria to evaluate and design better bba approximations (if possible).

References

1. Barnett, J.A.: Computational methods for a mathematical theory of evidence. In: Proceedings of IJCAI 1981, Vancouver, pp. 868–875 (1981)
2. Burger, T., Cuzzolin, F.: The barycenters of the k-additive dominating belief functions and the pignistic k-additive belief functions. In: Workshop on Theory of Belief Functions, Brest, France, pp. 1–6 (2010)
3. Denœux, T.: Inner and outer approximation of belief structures using a hierarchical clustering approach. International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems 9, 437–460 (2001)
4. Dubois, D., Prade, H.: A note on measures of specificity for fuzzy sets. International Journal of General Systems 10, 279–283 (1985)
5. Dubois, D., Prade, H.: An alternative approach to the handling of subnormal possiblity distributions. Fuzzy Sets and Systems 24, 123–126 (1987)
6. Grabisch, M.: Upper approximation of non-additive measures by *k*-additive measures - the case of belief functions. In: Proc. of the 1st Int. Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium (1999)
7. Jousselme, A.-L., Maupin, P.: Distances in evidence theory: Comprehensive survey and generalizations. International Journal of Approximate Reasoning 53, 118–145 (2011)
8. Kennes, R.: Computational aspects of the M*ö*bius transform of graphs. IEEE Transactions on SMC 22, 201–223 (1992)
9. Moral, S., Salmerón, A.: A Monte Carlo Algorithm for Combining Dempster-Shafer Belief Based on Approximate Pre-computation. In: Hunter, A., Parsons, S. (eds.) ECSQARU 1999. LNCS (LNAI), vol. 1638, pp. 305–315. Springer, Heidelberg (1999)
10. Shafer, G.: A Mathematical Theory of Evidence. Princeton University, Princeton (1976)
11. Smets, P.: Practical uses of belief functions. In: Lskey, K.B., Prade, H. (eds.) Uncertainty in Artificial Intelligence 15 (UAI 1999), Stockholm, Sweden, pp. 612–621 (1999)
12. Tessem, B.: Approximations for efficient computation in the theory of evidence. Artificial Intelligence 61, 315–329 (1993)
13. Voorbraak, F.: A computationally efficient approximation of Dempster-Shafer theory. Int. J. Man-Machine Studies 30, 525–536 (1989)