Suggestions to Make Dempster's Rule Convenient for Knowledge Combining

Ewa Straszecka

Abstract The paper deals with the Dempster's rule for the basic probability assignment combining from the point of view of a medical diagnosis support. The assignment determined on different sources of information is useful to establish symptoms weights, but often results of the combination are far from intuition. A modification of the Dempster's formula is proposed to make it possible to tune the resulting assignment according to the distance between the combined assignments. Properties of the proposed methods which are important for practical applications are shown on simulated data.

Keywords Dempster-Shafer theory of evidence · Knowledge combining

1 Introduction

The Dempster-Shafer theory of evidence [1] still remains one of the most important tools to represent and manage uncertainty in real-life problems of decision support. However, its flexibility is simultaneously an advantage and a drawback of applications. We benefit from neglecting dependence of focal elements as pieces of evidence, but the week point is a lack of indications in which way the basic probability assignment (bpa) for the elements should be determined. This is particularly crucial when the bpa expresses expert's knowledge, for instance in medical diagnosis. In this area a significance of a symptom is often roughly estimated and a combination of knowledge from different sources is necessary. Small changes of values of one bpa may be exaggerated after combination with another, also roughly determined bpa. Deficiencies of the classical Dempster's combination are pointed out in numerous papers, but no other general and satisfactory method is proposed so far, maybe except for the fuzzy rules aggregation, which however has its own disadvantages in

E. Straszecka (✉)
Institute of Electronics, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, 16 Akademicka St., 44-100 Gliwice, Poland
e-mail: ewa.straszecka@polsl.pl

K.T. Atanassov et al. (eds.), *Novel Developments in Uncertainty Representation and Processing*, Advances in Intelligent Systems and Computing 401,
DOI 10.1007/978-3-319-26211-6_2

applications [10]. Various concepts of changing Dempster's rule of combination are valuable and important because of their solid theoretical background, e.g. [3, 4, 9]. Still, there is no suggestion for smooth tuning of the resulting bpa, if we discover that it does not entirely follow our intuition.

The aim of this paper is to consider introducing a factor to the combination that will play a role similar to the fuzzifying coefficient in clustering. This means that changing the factor we could approach soft or severe estimation of symptom weights. Common points of the Dempster-Shafer theory and pattern recognition are noticed in [2]. Thus, it could be profitable to use experience of the latter old and highly developed domain to solve some application problems in knowledge combining. In the presented method a coefficient based on the similarity of bpas is used in combining. It is worth to notice that this concept can be widen for a number of similarity definitions which can open field for further research.

The present paper tries to make several suggestions about tuning the bpa, which hopefully are general enough to be of use in practical solutions. The proposed method is verified through simulations of bpas designed for a model of a medical diagnosis.

2 An Interpretation of Basic Probability Assignments as Sources of Diagnostic Knowledge

2.1 *The Classical Dempster's Combination and a Medical Diagnosis*

In medical diagnosis the bpa may represent weights of symptoms. It is particularly convenient since dependence of symptoms can be neglected and subsets of symptoms which are not distinct can be freely considered. In this case the classical bpa definition [7] is:

$$m_d(f) = 0, \sum_{S_i \in \mathbf{S}, i=1,\ldots,n} m_d(S_i) = 1, \tag{1}$$

where d denotes the diagnosis and S_i is its symptom—single or complex (i.e. a subset of symptoms). The false focal element f should be interpreted as the symptom that is never observed. Thus, the $m_d(S_i)$ is the weight of the rule: IF symptom is S_i THEN diagnosis is d.

Let us assume that the bpa is set by a physician using his experience. This is not rare and in many cases diagnostic indexes that are accepted as medical guidelines are not entirely based on statistic data, but also on subjective estimation [10]. It is also often that another physician has slightly different view on the diagnosis and make his own bpa. Then a combination of the bpas according to the classical Dempster's formula [1] is possible:

$$m_d(s_k) = \frac{\sum_{S_i \cap S_j = s_k} m_{d1}(S_i)m_{d2}(S_j)}{\sum_{S_i \cap S_j \neq f} m_{d1}(S_i)m_{d2}(S_j)}, \quad k = 1 \ldots n, \quad i = 1 \ldots n_1, j = 1 \ldots n_2. \tag{2}$$

It is obvious that the denominator in (2) stands for a normalization, such that (1) holds true. Let us denote it by:

$$m_d(s_k) = \left[\sum_{S_i \cap S_j = s_k} m_{d1}(S_i)m_{d2}(S_j) \right] \quad k = 1 \ldots n, i = 1 \ldots n_1, j = 1 \ldots n_2, \tag{3}$$

for simplicity. Usually, the symptoms in m_{d1} and m_{d2} are identically defined, so their number is stable: $n = n_1 = n_2$ and $s_i \equiv S_i$, $s_j \equiv S_j$ are identical in both sets of focal elements, but $s_i \cap s_j \neq \emptyset$, because complex symptoms usually include single symptoms which are individual focal elements in the same bpa. Let us notice, that if f is the symptom that is never observed, then the combination deficiency indicated by Zadeh [11], which occur when $m_1(a) = 0$; $m_1(b) = 0.1$; $m_1(c) = 0.9$ and $m_2(a) = 0.9$; $m_2(b) = 0.1$; $m_2(c) = 0$, is irrelevant for medical diagnosis. If $m_1(a) = 0$ then a should be a symptom that is never observed, so it does not occur in m_2, too. If a is a symptom that is not considered, for instance it is a result of a laboratory test which is not performed, its bpa value should not be equal to zero, though it could be low. Still, the classical Dempster's combination is too restrictive for applications for other reasons that illustrate the following example.

2.2 Example

Let us consider two bpas m_1 and m_2, defined for the following set of focal elements: $A = \{a, b, c, bc\}$. Assume at first that values of the bpas are set equal, i.e. 0.25, for the each element. Next let us change the bpa values, rising them by 0.05 for a and b as well as decreasing for c and ab, in m_1, while doing the opposite in m_2. Thus, at the end $m_1(a) = m_1(b) = 0.45$, $m_1(c) = m_1(bc) = 0.05$; $m_2(a) = m_2(b) = 0.05$, $m_2(c) = m_2(bc) = 0.45$. The combination is: $m_c(a) = [m_1(a)m_2(a)]$; $m_c(b) = [m_1(b)m_2(b) + m_1(b)m_2(bc) + m_2(b)m_1(bc)]$; $m_c(c) = [m_1(c)m_2(c) + m_1(c)m_2(bc) + m_2(c)m_1(bc)]$; $m_c(bc) = [m_1(bc)m_2(bc)]$; Results of calculations are in Fig. 1.

It can be noticed that though values of included bpas are at first equal, the resulting bpa shows big differences. We would prefer that $m_c(b)$ and $m_c(bc)$ would approach more the average value of the bpa than take extremely different values. Such phenomena usually contradict intuition. However, it is enough to add a modifying factor in the form of a power, i.e. $m_c^*(a) = \left[\left(m_1(a)m_2(a) \right)^\delta \right]$, for instance $\delta = 0.5$ in Fig. 1, to diminish changes of the final bpa.

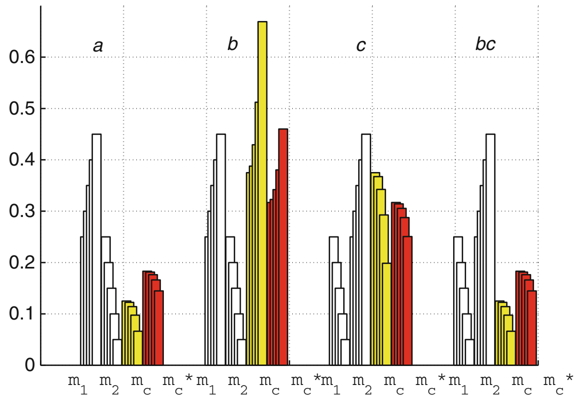

Fig. 1 Bpas resulting from the classical Dempster's formula (m_c) as well as from the modification (m_c^*), while gradual change of combined bpas (m_1 and m_2)

2.3 Tuning a Combination of Probability Assignments

If we choose a bpa parameter, we can allow for the classical Dempster's combination or its modification by the change of its value. It would be profitable to introduce a factor that can be decided according to the shape of combined bpas and in agreement with knowledge engineer's experience and intuition. Such a factor can be adjusted by investigation of similarity of the bpas. Let us consider tuning results of combination by means of a power coefficient, as it was suggested in Sect. 2.2. Thus:

$$m_c * (s_k) = \left[\left(\sum_{s_k = s_i \cap s_j} m_1(s_i) m_2(s_j) \right)^{\delta} \right] \tag{4}$$

It can be noticed, that for single focal elements indeed the combined bpas are both taken to the δ power, i.e. $\left[\left(m_1(s_i)\right)^{\delta} \left(m_2(s_i)\right)^{\delta} \right]$ which means that we simply moderate severe opinions of experts. The δ coefficient can be determined by means of similarity. Let us assume that the similarity is defined using a distance between the combined bpas. The distance can be formulated in several manners [6], for instance:

$$d(m_1, m_2) = \sqrt{(\mathbf{m_1} - \mathbf{m_2})' \mathbf{W} (\mathbf{m_1} - \mathbf{m_2})}, \tag{5}$$

where $\mathbf{m_1}$, $\mathbf{m_2}$ are vectors including bpa values. When $\mathbf{W}$ is equal to identity matrix (5) becomes the Euclidean distance d_e:

$$d_e(m_1, m_2) = \sqrt{(\mathbf{m_1} - \mathbf{m_2})'(\mathbf{m_1} - \mathbf{m_2})}. \tag{6}$$

If cardinality of focal elements is considered [5] then the Jaccard index [6] can be used, i.e. **W** elements are equal:

$$J(s_i, s_j) = \frac{\left|s_i \cap s_j\right|}{\left|s_i \cup s_j\right|} \tag{7}$$

In case of the example from Sect. 2.2, the Jaccard index matrix is:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix},$$

and the distance is:

$$d_b(m_1, m_2) = \sqrt{(\mathbf{m_1} - \mathbf{m_2})'\mathbf{B}(\mathbf{m_1} - \mathbf{m_2})}. \tag{8}$$

Additionally, a measure of distance is introduced:

$$d_a(m_1, m_2) = \min(\left|\mathbf{m_1} - \mathbf{m_2}\right|) \tag{9}$$

to find out what is an influence of the distance definition on the similarity coefficient.

Thus, the combination (4) can be used with δ:

$$\delta = \frac{k}{2 - d_{\#}},\ d < 2 \tag{10}$$

where $d_{\#}$ is d_e, d_b or d_a. The k parameter in (10) allows for precise tuning of the combination Fig. 4. The assumption $d < 2$ exclude the situation of extreme ("Zadeh's") conflict between the combined bpas. In the latter case, which the author thinks is irrelevant for practical problems of diagnosis support, the proposed modifications cannot be used.

There is a question why not to use the average [5] instead. The reason is that while we choose the type of the mean, for instance arithmetic, we cannot become "more strict" or "liberal" while combining distinct or similar bpas [8]. However, let us consider the average for a comparison of features expected from the combined bpa in applications. Then:

$$m_m(s_k) = \frac{1}{2}\left(m_1(s_k) + m_2(s_k)\right) \tag{11}$$

In this way four types of combinations are compared with the classical Dempster's rule: three based on distances (6)–(9) and δ (10), as well as the mean (11).

2.4 A Model of the Basic Probability Assignment in Medical Diagnosis

In medicine many symptoms are formulated by means of laboratory tests, for instance "test X result is low", "test Y result is normal". Let us use this formulation for the bpa determination. The point of norm (x_N) that indicates the symptom, is determined for ill (diagnosis 1) and healthy (diagnosis 2) populations (see Fig. 2). The norm point is set at the intersection of distributions of "ill" and "healthy". If patient's result is x, then $x < x_N$ means "test X result is low", while $x \geq x_N$ indicates "test X result is normal".

Let us assume that the diagnosis is based on values of three tests (variables): X, Y and Z. Tests Y and Z may be correlated, so they should be both considered as single focal elements as well as one complex focal element. The set of focal elements is described in Table 1 for the diagnosis 1. Results of the tests that are lower than the norm indicate diagnosis 1 (ill), which of course is a simplification. This assumption makes data simulation easier, but it does violate the generality of the diagnostic model. Symptoms for "healthy" (diagnosis 2) are formulated in analogy, and test results are "normal" when $x \geq x_N$, $y \geq y_N$, $z \geq z_N$.

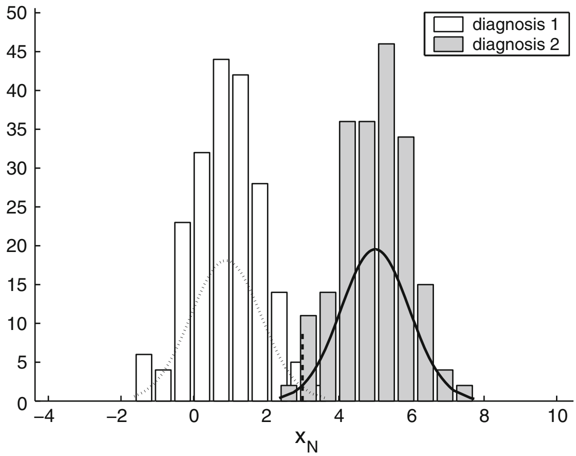

Fig. 2 An interpretation of laboratory test results: histograms and re-scaled normal distribution curves for patients with the diagnosis 1 (ill) and diagnosis 2 (healthy). The norm point is x_N

Table 1 Focal elements for diagnosis 1

Denotation	Heuristic meaning	Determination
s_1	"X is low"	$x < x_N$
s_2	"Y is low"	$y < y_N$
s_3	"Z is low"	$z < z_N$
s_4	"Y is low" and "Z is low"	$y < y_N$ and $z < z_N$

3 Simulations

Values of the X, Y, Z variables were generated by Matlab® normal distribution generator. Four kinds of normally distributed samples were simulated, with the mean ($\bar{\epsilon}$) and variance (σ), i.e. $N(\bar{\epsilon}_i, \sigma_i)$: $\epsilon_1 = 1$, $\sigma_1 = 1$, $\epsilon_2 = 1$, $\sigma_2 = 2$, $\epsilon_3 = 1$, $\sigma_3 = 3$, $\epsilon_4 = 5$, $\sigma_4 = 1$. Samples simulated for the X test were generated as not correlated with samples for Y and Z, while the latter were correlated. Each time the samples for diagnosis 2 (healthy) were characterized by the $N(\epsilon_4, \sigma_4)$ distribution (for all three tests). Samples for the diagnosis 1 (ill) had different distributions: $N(\epsilon_1, \sigma_1)$, $N(\epsilon_2, \sigma_2)$, $N(\epsilon_3, \sigma_3)$. Norm points were found from samples. From these samples basic probability assignments $m_{1d1}, m_{1d2}, m_{2d1}, m_{2d2}$ were calculated as normalized frequencies of occurrence of symptoms. Afterwards, m_{1d1}, and m_{2d1} were combined to obtain m_{d1}, while m_{1d2}, and m_{2d2} made m_{d2}. Calculations were performed for the same (200/200) and different (200/100) number of data in samples used for m_{1dj} and m_{2dj} $j = 1, 2$, determination. The cross-validation process was performed on 50 samples.

Combinations of m_{1dj} and m_{2dj} $j = 1, 2$ were performed to obtain m_c^* (4) with δ (10) for distances d_e (6), d_b (8), d_a (9) and the mean (11). It was necessary to define a criterion to estimate modifications of the classical combining. The modification results were compared to the bpa which was obtained if data from samples considered in one step of the cross-validation procedure were embedded in one dataset. This bpa was denoted as m_{all}. The criterion was the average of absolute differences between the result of a chosen modification and the m_{all} bpa, i.e.:

$$\begin{aligned} v_{ij\#} &= \textstyle\sum_{k=1}^{4} \left| m_{\#}(s_k) - m_{all}(s_k) \right| \\ v_{\#} &= \frac{1}{2500} \textstyle\sum_{i,j} v_{ij\#} \end{aligned} \tag{12}$$

where # denoted e, b, a, m or c.

4 Results

During calculations it was stated that equal or different number of data in samples had negligible influence on the value of the difference criterion (12) (changes were smaller than 0.001). Thus, in the following only results for the different number of data in samples are discussed. Results of combinations are denoted in Figs. 3, 4 and 5 by the following indexes: e, b, a, m, while the index c stands for the classical Dempster rule.

It is observable that bpas obtained as outcomes of the proposed modifications bring values of v_e, v_b and v_a within the $[v_m, v_c]$ interval. Differences among used types of modifications are hardly influenced by the choice of the sample characteristics. It is illustrated in the Fig. 3. The diagrams (a) and (b) were obtained for X

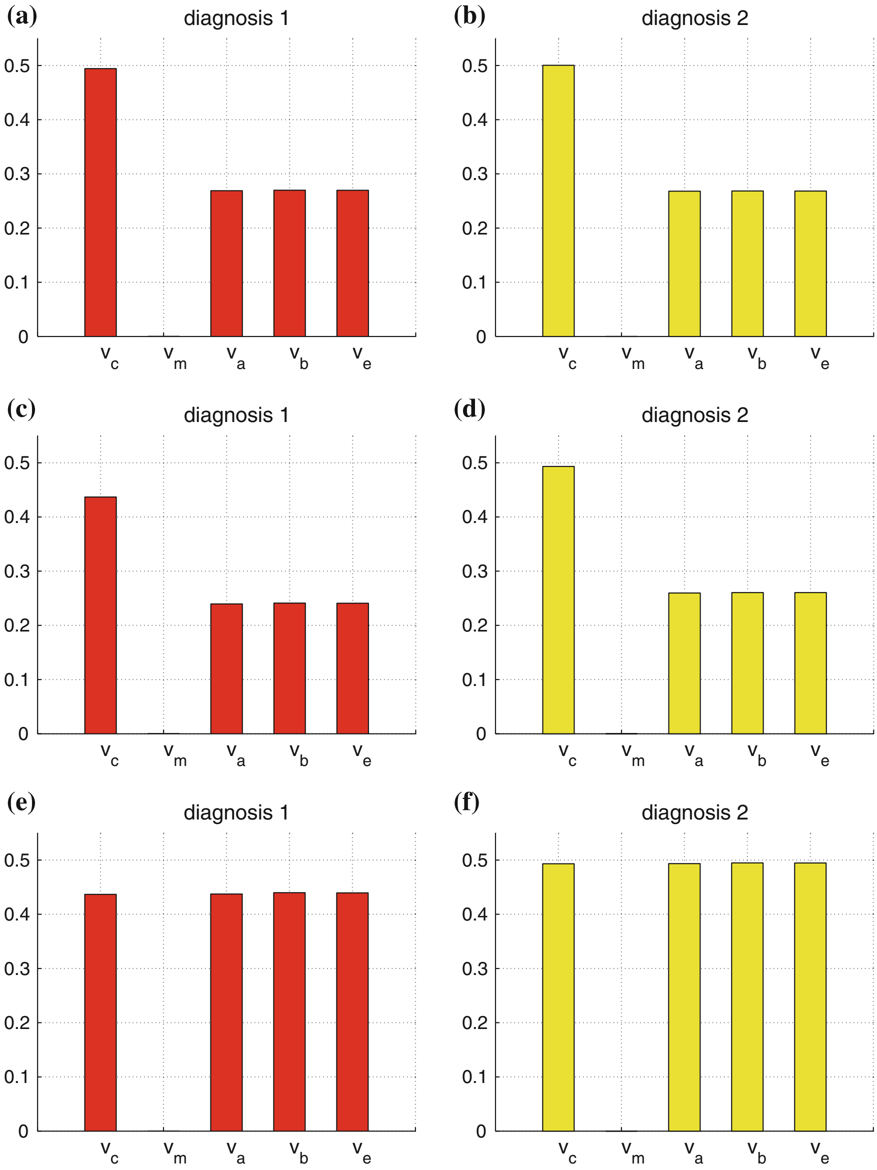

Fig. 3 The average difference (12) between the result of a chosen combination: $\mathtt{v}_{\mathtt{c}}$—classical, $\mathtt{v}_{\mathtt{m}}$—mean, $\mathtt{v}_{\mathtt{a}}$—minimum, $\mathtt{v}_{\mathtt{b}}$—Jaccard index, $\mathtt{v}_{\mathtt{e}}$—Euclidean; and the reference bpa—m_{all}. Diagrams **a**–**d** for $k = 1$, **e**, **f** for $k = 2$

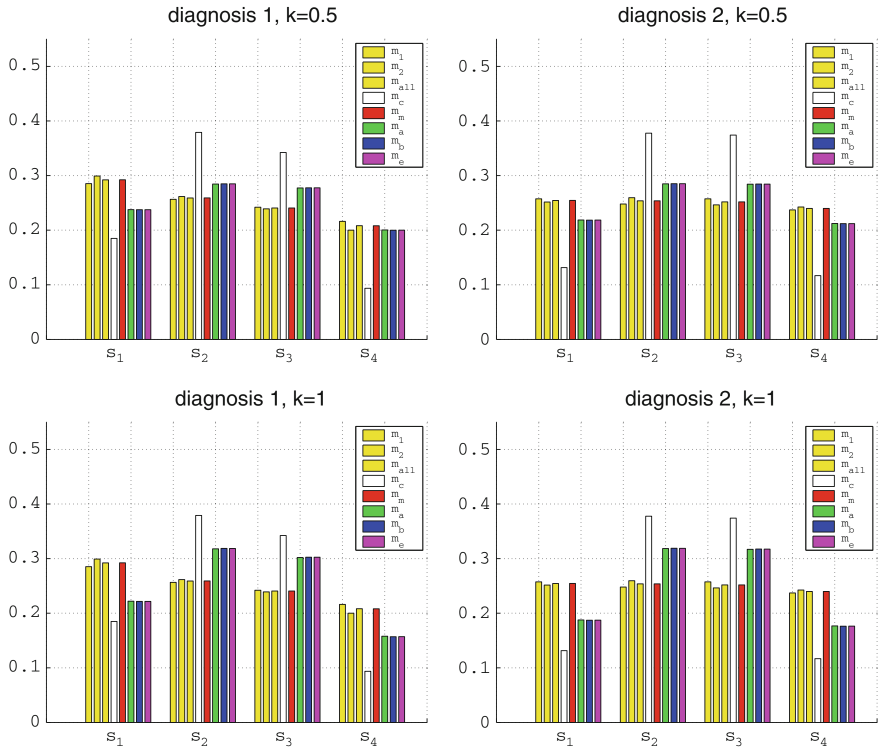

Fig. 4 Values of bpas obtained for different combination manners (part 1). Bars in the sequence denote: m_1, m_2, m_{all}, m_c, m_m, m_a, m_b, m_e

modeled by $N(\bar{\epsilon_3}, \sigma_3)$, Y by $N(\bar{\epsilon_2}, \sigma_2)$ and Z by $N(\bar{\epsilon_1}, \sigma_1)$, for the diagnosis 1 and X, Y, Z simulated by $N(\bar{\epsilon_4}, \sigma_4)$ for diagnosis 2. The diagrams (c) and (d) concern X modeled by $N(\bar{\epsilon_1}, \sigma_1)$, Y by $N(\bar{\epsilon_2}, \sigma_2)$ and Z by $N(\bar{\epsilon_3}, \sigma_3)$, for the diagnosis 1 and X, Y, Z simulated by $N(\bar{\epsilon_4}, \sigma_4)$ for diagnosis 2. Values of the $v_{\#}$ (12) are almost the same in the corresponding diagrams. This indicate that the method is universal enough to be used for a variety of diagnostic tests.

It is also effective, as the influence of the type of a modification as well as of its parameter k is visible. The v_m is close to zero (Fig. 3)—which is obvious for (11) definition. Yet, changing k in (10) we can tune the result of being closer to v_c or v_m. This is observable while comparing diagrams (a) and (e), as well as (b) and (f) in Fig. 3, determined for the same sample characteristics. Not only $v_{\#}$, but also values of bpas are regularly tuned along with the k (10) change. The Figs. 4 and 5 show values of bpas for different methods of combinations and various k. Sample characteristics for which bpas are calculated are the same as for diagrams (a), (b), (e), (f) in Fig. 3. For $k = 0.5$ the values of m_a, m_b and m_e are closer to m_m, while for $k = 2$ they are almost equal to m_c.

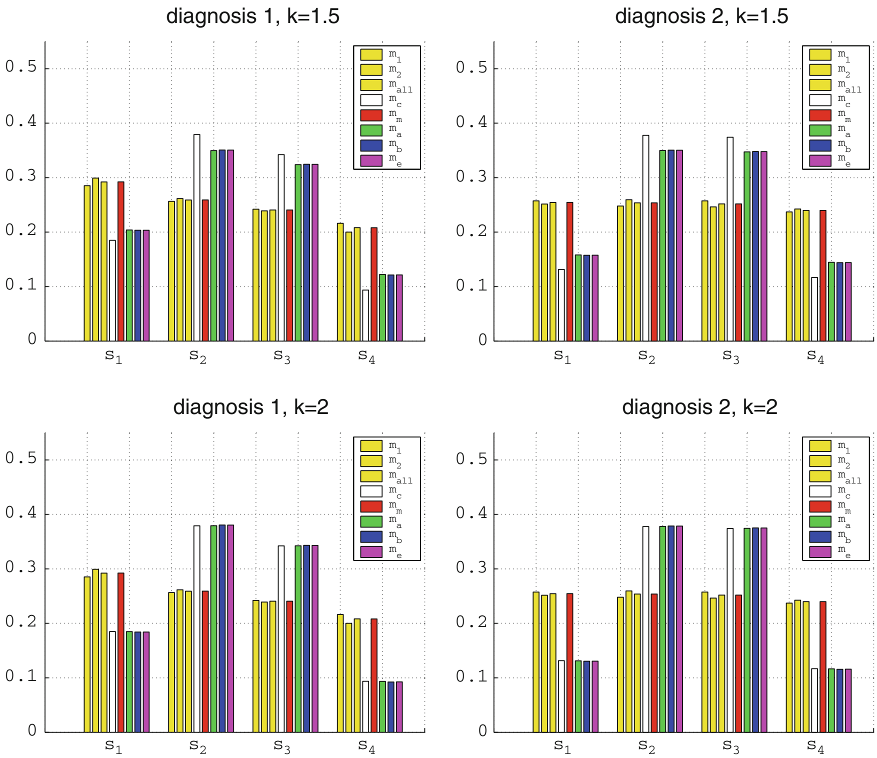

Fig. 5 Values of bpas obtained for different combination manners (part 2). Bars in the sequence denote: m_1, m_2, m_{all}, m_c, m_m, m_a, m_b, m_e

5 Discussion and Conclusions

In this paper modifications of the Dempster's rule for combining two bpas are proposed. Though the modifications do not eliminate "Zadeh's conflict", they can help to approach intuition while bpas combining. The idea comes from the author's experience from trials of diagnosis of combing information from different sources. In medical diagnosis support it is necessary to determine bpas on the basis of medical guidelines or expert's experience which are not very precise.

It should be noted that bpas combing is important not only from the point of view of calculation of the belief and plausibility of the diagnosis, but also as a tool for determination of symptoms weights. The right weights may help in planning diagnostic procedures, hence a moderate opinion is more valuable than two diverse judgments. Therefore, bpas combination is inevitable.

The presented modifications are based on the distance of combined bpas and generally aim at softening counter-intuitive effects while combining very similar or quite diverse bpas. A distance is chosen as a measure of similarity of the bpas. It is shown that the proposed concept works for several distance definitions. This may indicate that not only other manners of distance determination, but also other measures of similarity, not necessary based on distances, can be used to improve the Dempster's rule, by means of the introduced coefficient.

Simulations confirm that proposed methods make it possible to tune the resulting bpa towards mean or classical rule outcomes. Thus, it is possible to obtain results similar to the average, without explicit use of the mean. This allows for free change of final bpa to improve its conformity with intuition, whenever it is necessary.

The proposed modifications are numerically simple and do not show dependence on variables (symptoms) characteristics. Therefore, hopefully they can be used in many applications of medical diagnosis support.

References

1. Dempster, A.P.: A generalization of Bayesian inference. J. R. Stat. Soc. **30**(2), 205–247 (1968). Wiley, New Jersey
2. Denoeux, T.: A k-nearest neighbor classification rule based on the Dempster-Shafer theory. IEEE Trans. Syst. Man Cubernetics **25**(5), 804–813 (1995)
3. Denoeux, T.: Conjunctive and dijunctive combination of belief functions induced by nondistinct bodies of evidence. Artif. Intell. **172**, 234–264 (2008)
4. Dubois, D., Prade, H.: On the use of aggregation operations in information fusion process. Fuzzy Sets Syst. **142**, 143–161 (2004)
5. Han, D., Dezert, J., Yang, Y.: Evaluations of evidence combination rules in terms of statistical sensitivity and divergence. In: Proceedings of the 17th International Conference on Information Fusion (FUSION), pp. 1–7 (2014)
6. Jousselme, A.-L., Maupin, P.: Distances in evidence theory:comprehensive survaey and generalizatiuons. Int. J. Approx. Reason. **53**, 118–145 (2012)
7. Kacprzyk, J., Fedrizzi, M. (eds.): Advances in Dempster-Shafer Theory of Evidence. Wiley, New York (1994)
8. Murphy, C.K.: Combining belief functions when evidence conflicts. Decis. Support Syst. **29**(1), 1–9 (2000)
9. Smets, P.: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approx. Reason. **9**(1), 1–35 (1993)
10. Straszecka, E.: Measures of Uncertainty and Imprecision in Medical Diagnosis Support. Wydawnictwo Politechniki Slaskiej, Gliwice (2010)
11. Zadeh, L.A.: Review of mathematical theory of evidence by G. Shafer. AI Mag. **5**(3), 81–83 (1984)

http://www.springer.com/978-3-319-26210-9

Novel Developments in Uncertainty Representation and Processing
Advances in Intuitionistic Fuzzy Sets and Generalized Nets – Proceedings of 14th International Conference on Intuitionistic Fuzzy Sets and Generalized Nets
Atanassov, K.T.; Castillo, O.; Kacprzyk, J.; Krawczak, M.; Melin, P.; Sotirov, S.; Sotirova, E.; Szmidt, E.; De Tré, G.; Zadrozny, S. (Eds.)
2016, X, 400 p. 128 illus., 60 illus. in color., Softcover
ISBN: 978-3-319-26210-9