

On Fuzzy Focal Elements Combining

Ewa Straszecka^(✉)

Faculty of Automatic Control, Electronics and Computer Science,
Institute of Electronics, Silesian University of Technology,
16 Akademicka St., 44-100 Gliwice, Poland
`ewa.straszecka@polsl.pl`

Abstract. The Dempster-Shafer theory can be extended for fuzzy focal elements. When knowledge from different sources is combined, membership functions of the elements should be also joined. The paper suggests a combination of knowledge by means of the conjunction of data-driven membership functions. A discussion on an influence of the combination on the basic probability assignment is provided. The method can be helpful for medical knowledge transfer.

Keywords: Dempster-Shafer theory · Fuzzy sets · Diagnosis support

1 Introduction

The Dempster-Shafer theory of evidence (DST) [1, 5] may be extended for fuzzy focal elements [6]. The extension consists in defining focal elements by means of membership functions (mfs) that correspond to linguistic values. For instance if a variable is a ‘glucose level’ then the mf ‘high’ can be used in a premise of a diagnostic rule and this mf represents the medical symptom. A conclusions of the rule is a diagnose, for example ‘diabetes’. Thus, a set of focal elements is determined for each diagnosis. The basic probability assignment (bpa) which is next defined becomes an evaluation of the rule weights. Afterwards, the belief [1, 5] can be calculated for all considered diagnoses and compared to chose the diagnosis of the greatest belief value as the final conclusion. Such a use of the DST is convenient in medical diagnosis support, but fuzzy focal elements create not only new opportunities, but also new problems to solve. One of them is the combination of assignments.

Medical knowledge represented by focal elements and the bpa should be subjected to combination from different sources, i.e. an expert and a training database or from two experts, as well as from two databases [7]. The classical DST combination refers only to bpas, while mfs require an individual treatment [7]. The present paper suggests an approach for the fuzzy focal elements combination and makes an attempt to solve several related problems. Firstly, a conjunction is proposed for the combination. Secondly, a dependence between a similarity of combined focal elements and changes in resulting bpa are studied. Since similarity factors for a comparison of mfs and bpas are necessary, a choice of the factors

as well as an evaluation criteria for the combination effectiveness are proposed. Conclusions are driven as a result of multiple numerical simulations, which are close to a simplified task of medical diagnosis support and face usual difficulties of medical knowledge transfer. Theoretical background for the extension of the DST for fuzzy focal elements is given in [6, 7] and other works of the author. Because of the limited length of this paper, only the most necessary information is provided. For the same reason experiments are shortly summarized. Yet, every interested reader can easily build and investigate the proposed diagnostic model.

2 DST for Fuzzy Focal Elements

The bpa for fuzzy focal elements [6] is defined as:

$$m(f) = 0, \quad \sum_{\substack{s_i \in S, i=1, \dots, n \\ \eta_i > \eta_{BPA}}} m(s_i) = 1. \quad (1)$$

where η_{BPA} is the minimal level of precision for which a symptom is considered as carrying information. The symptom is defined by means of the $\mu(x)$ mf and η is the actual precision of a symptom found for the x^* data case, i.e. $\eta = \mu(x^*)$. For a premise including n symptoms $\eta = \min_{j=1, \dots, n} (\mu_j(x_j^*))$. The bpa can be found from data when mfs for symptoms are determined. To this end, the η_{BPA} threshold is assumed and a number of data cases that fit mfs better than the threshold ($\eta \geq \eta_{BPA}$) is counted. After normalization of the numbers for all symptoms the (1) conditions hold true [7].

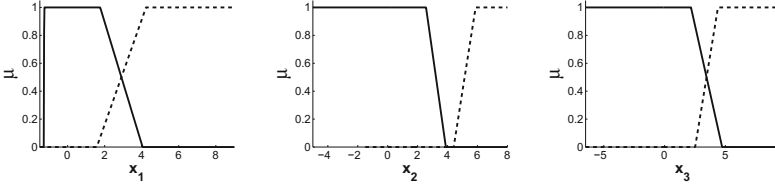


Fig. 1. Mfs for two competitive diagnoses: D_1 (solid line) and D_2 (dashed line).

Shapes of mfs are based on quartiles and intersection points of the training data distributions [6]. The quartile indicates the point for which $\mu(x) = 1$, while the intersection point determines the point of half membership ($\mu(x) = 0.5$). The intersection point is the x value for which theoretical distributions of x for two competitive diagnoses crosses [6]. When quartiles and the intersection point do not correspond each other then quartiles make the mfs with ‘steep’ slopes, e.g. with 100 coefficient gradient. The Fig. 1 shows mfs for two diagnoses (D_1 , D_2), each of them considering symptoms as low and high values of the variables x_1 , x_2 and x_3 . Training data distributions for x_2 are incompatible, so mfs shapes for this variable are ‘steep’.

3 A Model of a Diagnosis

Let us assume that the diagnosis is based on three symptoms: X_1 , X_2 and X_3 . The symptoms are numerical variables which low values are assigned to the D_1 diagnosis and high values to the D_2 . The symptoms X_2 and X_3 are correlated, since disorders of the human body are often related. Thus, four rules can be formulated for D_k , $k = 1, 2$:

$s_j^{(k)}$: IF X_j is $A_j^{(k)}$ then D_k , $j = 1, 2, 3$,
 $s_4^{(k)}$: IF X_2 is $A_2^{(k)}$ and X_3 is $A_3^{(k)}$ then D_k .

The linguistic values $A_j^{(k)}$, $j = 1, 2, 3$, for D_1 are represented by the ‘low’ mf $\mu^{(1)}(x_j)$, and for D_2 by the ‘high’ mf $\mu^{(2)}(x_j)$. Mfs are data driven using simulated data of numbers generated from the normal distribution: one-dimensional for X_1 and two-dimensional for X_2 and X_3 . In the present model it is assumed that X_1 – X_3 have different distributions for the D_1 diagnosis and the same distribution for the D_2 . Obviously, parameters of the distributions differ and each training data case is independently simulated. The bpas are calculated assuming $\eta_{bpa} = 0.5$. Bpas for different data sets are compared.

Similarity of bpas can be determined concerning a cardinality of focal elements. To this end, the Jaccard index matrix [2] is defined:

$$J^{(k)}(s_i, s_j) = \frac{|s_i^{(k)} \cap s_j^{(k)}|}{|s_i^{(k)} \cup s_j^{(k)}|}, i, j = 1, 2, 3; k = 1, 2. \quad (2)$$

The index matrix for the proposed focal elements and the both diagnoses is:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix}. \quad (3)$$

Let us assume that the diagnostic knowledge comes from two sources of information, particularly, the mfs originate from two populations. A combination of them could be the conjunction, since a cautious approach is presumed in medicine [7]. Let us use the simplest conjunction which is the minimum of mfs. Now, it should be investigated how differences in training data may influence variability of bpas. Various diagnostic situations are considered as it is described in the next section. Mfs of two sets will be compared, then combined and the result will be compared to the mf obtained for the data of both sets put into the one set. The bpas change along with mfs, so they will be compared in the same way. The comparisons should show how resistant are bpas for changes of knowledge, or – on the other hand, if we can tell irrelevant data from the changes of mfs or bpas.

4 Simulated Data

Five samples of data are simulated, each of them includes 100 data sets. Every data set contain 400 cases, i.e. 200 for each of two diagnoses. Each data sample is generated for different parameters of the normal distribution $N(\bar{x}, \sigma)$, where x - mean, σ - variance. When correlated data are necessary, data of two-dimensional distribution of the mentioned mean and variance are generated. Normality of data is verified by the Matlab® Liliefors test. The correlation coefficient maintains $r \geq 0.5$, except for $N(10, 5)$ sample, for which $r \geq 0.2$. In the present paper the following samples are used:

Sample 1 – is the ‘mixed sample’. Its single set includes 100 cases from distribution 1 and 100 from distribution 2, for each diagnosis. Distribution 1 for D_1 is: $N(1, 1)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 1$ and variances $\sigma_2 = 2, \sigma_3 = 3$ for x_2 and x_3 . Distribution 2 for D_1 is: $N(1, 2)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 1$ and variances $\sigma_2 = 3, \sigma_3 = 4$ for x_2 and x_3 . Distribution 1 for D_2 is: $N(5, 1)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 5$ and variances $\sigma_2 = \sigma_3 = 1$ for x_2 and x_3 . Distribution 2 for D_2 is: $N(5, 2)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 5$ and variances $\sigma_2 = \sigma_3 = 2$ for x_2 and x_3 . This sample should simulate combining knowledge from two similar, but not identical populations, for instance social groups of different living conditions or habits.

Sample 2 – is another ‘mixed sample’. A data set includes 200 data cases from one distribution for the D_1 and twice 100 cases from various distributions for the D_2 . Distribution for D_1 is: $N(1, 2)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 1$ and variances $\sigma_2 = 3, \sigma_3 = 4$ for x_2 and x_3 . Distribution 1 for D_2 is: $N(5, 2)$ for x_1 and two-dimensional normal distribution of the same means and variances for x_2 and x_3 . Distribution 2 for D_2 is: $N(10, 5)$ for x_1 and two-dimensional normal distribution of the same means and variances for x_2 and x_3 . This sample may simulate situation when irrelevant data are attached to the training set.

Sample 3 – is a ‘uniform sample’, its sets include 200 data cases from one distribution for the D_1 diagnosis and the same number from another distribution for the D_2 . Distribution for D_1 is: $N(1, 1)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 1$ and variances $\sigma_2 = 2, \sigma_3 = 3$ for x_2 and x_3 . Distribution for D_2 is: $N(5, 1)$ for x_1 , two-dimensional normal distribution of the same means and variances for x_2 and x_3 . Consistent knowledge from two sources is simulated by this sample. Variances are small and means are not close to each other which means that symptoms are quite significant and the diagnosis is easy.

Sample 4 – is the second ‘uniform sample’, with greater variances. Its structure is analogical to Sample 3. Distribution for D_1 is: $N(1, 2)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 1$ and variances $\sigma_2 = 3, \sigma_3 = 4$ for x_2 and x_3 . Distribution for D_2 is: $N(5, 2)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 5$ and variances $\sigma_2 = \sigma_3 = 2$ for x_2 and x_3 . The sample simulates consistent knowledge when symptoms are less significant and the diagnosis is more difficult.

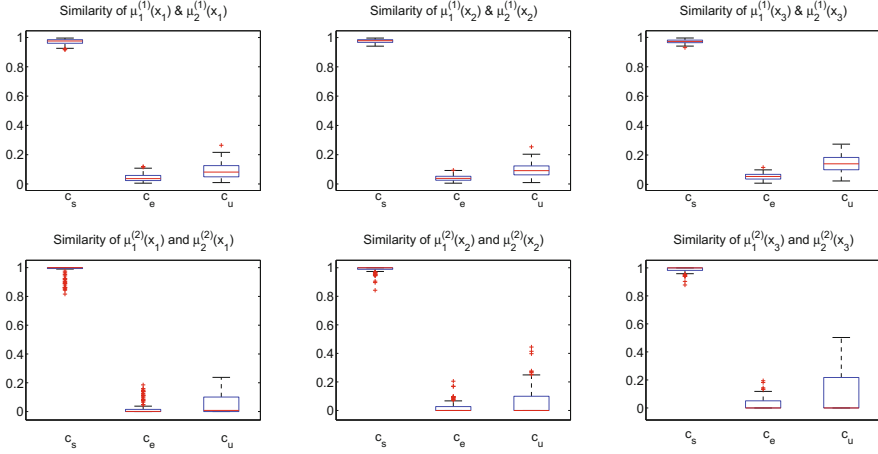


Fig. 2. A comparison of similarity factors for mfs.

Sample 5 – is the third ‘uniform sample’, yet the D_2 symptoms are very ambiguous with great variance and a distant mean. Its structure is the same as two previous samples. Distribution for D_1 is the same as for the Sample 4. Distribution for D_2 is: $N(10,5)$ for x_1 , two-dimensional normal distribution of means $\bar{x}_2 = \bar{x}_3 = 10$ and variances $\sigma_2 = \sigma_3 = 5$ for x_2 and x_3 . The sample simulate a diagnosis when symptoms of D_1 and D_2 have very different characteristics.

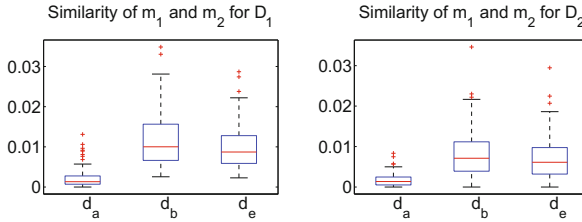


Fig. 3. A comparison of similarity factors for bpas.

5 Similarity Factors

Numerical experiments on mfs and bpas must be preceded by the choice of proper similarity factors. Let us first discuss factors of mfs comparison. The simplest is the maximum absolute distance:

$$c_u(\mu_1, \mu_2) = \max_x (|\mu_1(x) - \mu_2(x)|), \quad (4)$$

that is numerically calculated for multiple x points (e.g. $n = 300$). The distance represents rather a difference than a similarity, but the minimal distance cannot

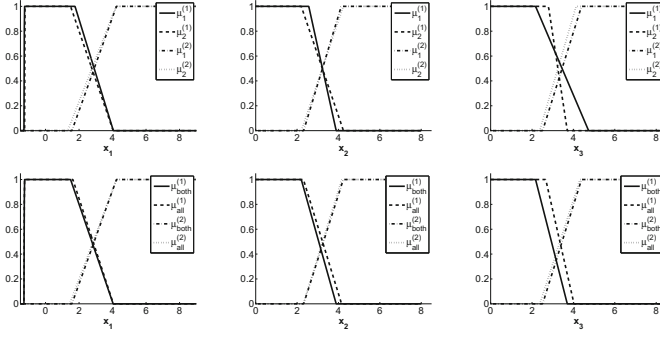


Fig. 4. Change of shape of mfs for two data sets of Sample 3.

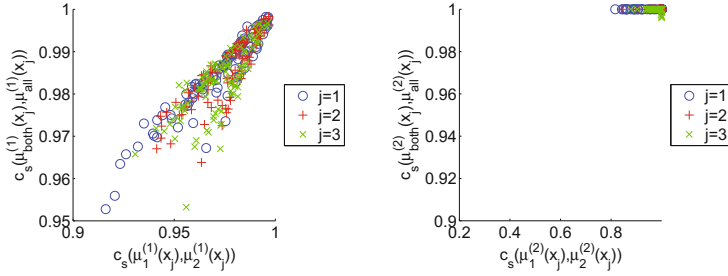


Fig. 5. Values of c_s for Sample 1,3,4 and 5.

be used for trapezoid mfs, as there must be a point in which compared mfs both equal 1 (see Fig. 4). The Euclidean distance is also applicable:

$$c_e(\mu_1, \mu_2) = \sqrt{\sum_{i=1}^n (\mu_1(x_i) - \mu_2(x_i))^2}. \quad (5)$$

Another possible similarity factor is [4]:

$$c_s(\mu_1, \mu_2) = \left| \min_x (\mu_1(x), \mu_2(x)) \right| + \left| \max_x (\mu_1(x), \mu_2(x)) \right| - 1, \quad (6)$$

calculated for the same points as (4) and (5).

In the Fig. 2 means, standard deviations, value intervals and outliers for similarity factors c_s , c_e and c_u for mfs of variables x_1 , x_2 and x_3 and two diagnoses are presented. The mfs for the D_1 are determined as in sample 2 and for the D_2 according to the second distribution of this sample. The diagram is obtained for 100 data sets. The factors behave similarly. It is shown that for great variance (for the $N(10,5)$ distribution) outliers of similarity factors appear. For the D_1 the factors c_s and c_e have small standard deviations. The former has greater values, so it is chosen for further research of the mf comparison.

Secondly, factors of evaluation the bpas similarity should be selected. Several factors can be used. The simplest is the distance [8]:

$$d_a(m_1, m_2) = \min(|\mathbf{m}_1 - \mathbf{m}_2|). \quad (7)$$

The Euclidean distance can be also used:

$$d_e(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'(\mathbf{m}_1 - \mathbf{m}_2)}. \quad (8)$$

The distance with the Jaccard index (2), (3) evaluates focal elements [3]:

$$d_b(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'\mathbf{B}(\mathbf{m}_1 - \mathbf{m}_2)}. \quad (9)$$

These distances have proved to be useful for combination of bpas [8], hence they are also applied for the present evaluation. Results of their performance tests are in the Fig. 3. Means, standard deviations, value intervals and outliers for the factors are presented. They are obtained for a comparison of two bpas (m_1 and m_2) calculated for the same data as mfs. The worst differentiation among various data sets is given by d_a , while d_b and d_e are comparable. It can be expected that if mf shapes are more sophisticated, d_b could work better than d_e . Thus, the d_b factor (9) will be used further on.

Table 1. Similarity μ_1 and μ_2

smp	Diagnosis D_1				Diagnosis D_2		
	par	$c_s(x_1)$	$c_s(x_2)$	$c_s(x_3)$	$c_s(x_1)$	$c_s(x_2)$	$c_s(x_3)$
1	\bar{x}	0.9339	0.9544	0.9456	0.9987	0.9968	1.0000
	std	0.0521	0.0364	0.0349	0.0117	0.0321	0.0000
2	\bar{x}	0.1738	0.1857	0.1954	0.2015	0.2300	0.2300
	std	0.3183	0.3402	0.3580	0.3711	0.4208	0.4208
3	\bar{x}	0.9555	0.9672	0.9568	0.9915	0.9978	0.9835
	std	0.0490	0.0308	0.0383	0.0664	0.0152	0.0580
4	\bar{x}	0.9473	0.9535	0.9505	1.0000	1.0000	0.9976
	std	0.0441	0.0393	0.0397	0.0000	0.0000	0.0243
5	\bar{x}	0.9721	0.9759	0.9718	0.9751	0.9880	0.9878
	std	0.0187	0.0137	0.0133	0.0479	0.0248	0.0220

6 Similarity of Focal Elements and Its Influence on Probability Assignment

The shapes of mfs influence the bpa, since it is calculated using frequency of occurrence of training data cases. However, the dependence is not straightforward as the slope of the mf concerns an interval of a variable domain in which

cases fall less frequently (the tail of the distribution) and one bpa value is influenced by the other focal elements. The Fig. 4 illustrates mfs $\mu_1^{(k)}(x_j)$ and $\mu_2^{(k)}(x_j)$ $k = 1, 2, j = 1, 2, 3$, obtained using two data sets of 100 elements each, $\mu_{both}^{(k)}$ found as minimum of the mfs and $\mu_{all}^{(k)}$ determined for the joined cases (200 elements). Differences between mfs of the individual samples are easy to notice, while between μ_{both} and μ_{all} are less significant. This observation allows to suppose that the conjunction can properly combine information given by mfs.

Table 2. Similarity μ_{all} and μ_{both}

smpl	Diagnosis D_1				Diagnosis D_2		
	par	$c_s(x_1)$	$c_s(x_2)$	$c_s(x_3)$	$c_s(x_1)$	$c_s(x_2)$	$c_s(x_3)$
1	\bar{x}	0.9670	0.9769	0.9718	0.9999	1.0000	1.0000
	std	0.0286	0.0190	0.0208	0.0010	0.0000	0.0000
2	\bar{x}	0.1820	0.2226	0.2257	0.2171	0.2216	0.2293
	std	0.3344	0.4079	0.4130	0.3985	0.4057	0.4196
3	\bar{x}	0.9802	0.9809	0.9749	1.0000	1.0000	0.9998
	std	0.0159	0.0175	0.0230	0.0000	0.0000	0.0013
4	\bar{x}	0.9736	0.9773	0.9742	0.9998	1.0000	1.0000
	std	0.0219	0.0213	0.0226	0.0022	0.0000	0.0000
5	\bar{x}	0.9848	0.9856	0.9826	1.0000	1.0000	0.9998
	std	0.0093	0.0077	0.0079	0.0000	0.0001	0.0007

The Table 1 show mean values (\bar{x}) and standard deviations (std) of $c_s(\mu_1^{(1)}(x_j), \mu_2^{(1)}(x_j))$, $j = 1, 2, 3$, calculated for D_1 and D_2 for 100 sets of data. The c_s values for the D_2 are very close to 1, since the variance of distribution for this diagnosis is quite low. The Sample 2 is an exception - this sample is simulated for the distribution of high variance and its c_s is low. In Fig. 5 the $c_s(\mu_1^{(k)}(x_j), \mu_2^{(k)}(x_j))$ vs. $c_s(\mu_{both}^{(k)}(x_j), \mu_{all}^{(k)}(x_j))$, $k = 1, 2, j = 1, 2, 3$, are depicted (except for Sample 2). In the left diagram it is noticeable, that a dependence between the two c_s is linear or better, i.e. the mf after combining ($\mu_{both}^{(1)}$) is at least as similar to the mf $\mu_{all}^{(1)}$ as $\mu_1^{(1)}$ to $\mu_2^{(1)}$. Few different values of $c_s(\mu_1^{(2)}(x_j), \mu_2^{(2)}(x_j))$ and $c_s(\mu_{both}^{(2)}(x_j), \mu_{all}^{(2)}(x_j))$ are obtained. Points in the right diagram represent multiple values and because of low sample variance the similarity is even better. Complete calculation results of $c_s(\mu_{both}^{(k)}(x_j), \mu_{all}^{(k)}(x_j))$, $k = 1, 2, j = 1, 2, 3$ are in Table 2. They confirm that similarity of mfs after combination is high for uniform mfs or mfs with low data variability. The standard deviation of c_s for the mfs is low. The opposite is observed for Sample 2.

The divergence between mfs influences differences of bpas, but a relation is not straightforward. The Table 3 presents $d_b(m_{both}, m_{all})$ related to $d_b(m_1, m_2)$. The mean values of the d_b are not very different, which may suggest that the

bpa is quite resistant to changes of knowledge. It is noticeable that the standard deviation of $d_b(m_{both}, m_{all})$ is low for the uniform samples and high for the mixed samples. Generally, this parameter for $d_b(m_{both}, m_{all})$ is lower in comparison to $d_b(m_1, m_2)$ for uniform samples and higher for mixed samples. The only exception is the standard deviation for D_2 and sample 1, but for this sample the mixed distributions do not vary much. The c_s and d_b values are not comparable, since they are completely different measures. Yet, if we study the \bar{x}/std coefficient for the measures using values from Tables 1, 2 and 3, we see that this coefficient is similar for sample 1, 3, 4 and 5 and very different for sample 2. Thus, the bpa is resistant for small changes of data samples, but it is changed significantly when irrelevant data are introduced.

Table 3. Similarity of bpas

		$d_b(m_1, m_2)$		$d_b(m_{both}, m_{all})$	
		D_1	D_2	D_1	D_2
1	\bar{x}	0.0135	0.0738	0.0082	0.0292
	std	0.0140	0.1438	0.0154	0.0718
2	\bar{x}	0.0135	0.0107	0.0137	0.0067
	std	0.0071	0.0066	0.0265	0.0137
3	\bar{x}	0.0171	0.0567	0.0115	0.0120
	std	0.0117	0.1229	0.0070	0.0162
4	\bar{x}	0.0119	0.0499	0.0134	0.0105
	std	0.0120	0.0926	0.0086	0.0074
5	\bar{x}	0.0117	0.0080	0.0045	0.0082
	std	0.0070	0.0058	0.0023	0.0040

7 Conclusions

In the paper a method of fuzzy focal elements combination is suggested. Tests performed on simulated data show that the conjunction of the elements defined as minimum of mfs is the proper operation of joining knowledge represented by the same linguistic values, but different fuzzy sets. Data simulated for the normal distribution confirm that mfs after combining provide functions more similar to the functions found for unified data, than the mfs to each other before combining. If the combined mfs originate from consistent sources of data, also the standard deviation of the bpa determined for combined focal elements is low. It indicates that the generalization of knowledge is correct. On the contrary, the great standard deviation of the resulting assignment appears when combined mfs derive from differently distributed data. This means that the focal elements and the assignment are sufficiently sensitive to a change of training data.

Three diagnostic situations were modeled to test the combination characteristics. The first was combining knowledge from the same source with different training data sets. The tests resulted with similar mfs and the similarity factor of basic probability values of low standard deviation. Thus, the method do not spoil the basic assignment when the sources of information are consistent.

The second was joining knowledge from various, but not very different sources. The combination outputs were similar mf and the similarity factor of bpa with the standard deviation not very different from the previous test. Hence, the method allows for some generalization of knowledge.

The third situation was modeled by different distributions and resulted with less similar mfs and the bpa similarity factors of high standard deviation. Therefore, knowledge from diverse sources should not be combined and fuzzy focal elements as well as the bpa should be build separately for each individual population. Moreover, if an assignment values after combination show high variance, it should be suspected that the combination was unjustified.

The proposed method of knowledge combination and its features presented for simulated data may facilitate medical knowledge transfer when diagnostic rules remain the same, but populations disclose slightly different characteristics.

Acknowledgement. This research was supported by statutory funds (BK-220/RAu-3/2016) of the Institute of Electronics, Silesian University of Technology.

References

1. Dempster, A.P.: A generalisation of Bayesian inference. *J. Royal Stat. Soc.* **30**, 205–247 (1968)
2. Han, D., Dezert, J., Yang, Y.: Evaluations of evidence combination rules in terms of statistical sensitivity and divergence. In: *Proceedings of the FUSION*, pp. 1–7 (2014)
3. Jousselme, A.-L., Maupin, P.: Distances in evidence theory: comprehensive survey and generalizations. *Int. J. Approximate Reasoning* **53**, 118–145 (2012)
4. Lee, S.H., Pedrycz, W.: Design of similarity measures for fuzzy sets on the basis of distance measure. *Int. J. Fuzzy Syst.* **11**(2), 67–72 (2009)
5. Shafer, G.: *A mathematical theory of evidence*. Princeton University Press, Princeton, New Jersey (1976)
6. Straszeka, E.: Combining uncertainty and imprecision in models of medical diagnosis. *Inf. Sci.* **176**, 3026–3059 (2006)
7. Straszeka, E.: Combining knowledge from different sources. *Expert Syst.* **27**(1), 40–52 (2010)
8. Straszeka, E.: Suggestions to make Dempster’s rule convenient for knowledge combining. In: Atanassov, K.T., et al. (eds.) *Novel Developments in Uncertainty Representation and Processing, Advances in Intelligent Systems and Computing* **401**, 17–27 (2016)

Uncertainty and Imprecision in Decision Making and
Decision Support: Cross-Fertilization, New Models and
Applications

Selected Papers from BOS-2016 and IWIFSGN-2016 held
on October 12-14, 2016 in Warsaw, Poland

Atanassov, K.T.; Kacprzyk, J.; Kałuszek, A.; Krawczak, M.;
Owsiński, J.; Sotirov, S.S.; Sotirova, E.; Szmidt, E.;
Zadrozny, S. (Eds.)

2018, XVIII, 350 p. 116 illus., Softcover

ISBN: 978-3-319-65544-4