ALPHA-INTEGRATION OF MULTIPLE EVIDENCE

Heeyoul Choi[1], Anup Katake[2], Seungjin Choi[3] and Yoonsuck Choe[1]

[1]Dept. of Computer Sci. and Eng., Texas A&M Univ., TX
{hchoi,choe}@cs.tamu.edu
[2]Starvision Technologies, Inc., TX
akatake@starvisiontech.com
[3]Dept. of Computer Sci., Pohang Univ. of Sci. and Tech., Korea
seungjin@postech.ac.kr

ABSTRACT

In pattern recognition, data integration is a processing method to combine multiple sources so that the combined result can be more accurate than a single source. Evidence theory is one of the methods that have been successfully applied to the data integration task. Since Dempster-Shafer theory as the first evidence theory can be against our intuitive reasoning with some data sets, many researchers have proposed different rules for evidence theory. Among all these rules, the averaging rule is known to be better than others. On the other hand, α-integration was proposed by Amari as a principled way of blending multiple positive measures. It is a generalized averaging algorithm including arithmetic, geometric and harmonic means as its special case. In this paper, we generalize evidence theory with α-integration. Our experimental results show how our proposed methods work.

Index Terms— Evidence Theory, Data Integration

1. INTRODUCTION

In pattern recognition, data integration has been applied to multiple sources to achieve improved accuracy than that based on a single source of information because one sensor might not be good enough to provide unambiguous information. Some data integration algorithms have been proposed (see [1] and references therein) such as Bayesian inference [2], evidence theory (or Dempster-Shafer theory) [3, 4], clustering algorithms and neural networks [5]. Kernel-based integration methods are also used for integration [6, 7]. However, there is no consensus on which method is more universally applicable [1].

Evidence theory (ET) is a mathematically well defined theory for handling conflicts between different bodies of evidence. It is conceptually the same as Bayesian theory except that it uses epistemic (subjective) uncertainty [8]. The advantages of ET include its flexibility in theory and easy implementability. After the initial introduction of ET by Dempster and Shafer [3, 4], it has been improved [9, 10] because in some cases the original Dempster-Shafer (D-S) combination rule is against our intuitive reasoning. Many researchers have proposed different rules to address this issue, and recently, some effective averaging rules have emerged, and in these rules, how to assign the weights becomes an important issue [9, 11, 12]. However, they still use the D-S combination rule after the averaging.

On the other hand, α-integration has been proposed for stochastic model integration by Amari [13]. Given a number of stochastic models in the form of probability distributions, it finds out the optimal integration of the sources in the sense of minimizing the α-divergence. Many artificial neural models such as the mixture (or product) of experts model [14, 15] can be considered as a special case of α-integration.

In this paper, we generalize ET with α-integration in two ways: (1) the D-S combination rule is interpreted as a special case of α-integration and generalized, and (2) the previous averaging methods for ET are generalized. Note that both generalization approaches can be applied at the same time. Our experimental results show how our proposed methods work. We use the same data sets as the previous averaging methods used.

2. REVIEW OF PREVIOUS WORK

2.1. Evidence theory

As the first evidence theory, Dempster and Shafer proposed the D-S combination rule [3, 4]. Here, we give a brief review of it. For details, see [10] and references therein.

Let Θ be a set of hypotheses, and m be a basic belief assignment (BBA) which is a function from a subset of Θ to $[0, 1]$ with the following properties.

$$
\begin{aligned}
m(\phi) &= 0, \\
\sum_{A \subseteq \Theta} m(A) &= 1.
\end{aligned}
\tag{1}
$$

When two evidence bodies m_1 and m_2 are given, the D-S combination rule for $\tilde{m}(A)$ is defined by

$$
\tilde{m}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K},
\tag{2}
$$

where

$$
K = \sum_{B \cap C = \phi} m_1(B) m_2(C).
\tag{3}
$$

Here, K indicates basic probability associated with conflict. This can be easily expanded to more than two evidence bodies.

As pointed out in [9], in some cases the D-S combination rule is against our intuitive reasoning. For example, when only one evidence has 0 belief but all others have 1 belief, still the combination is 0. To overcome this weakness, ET has been improved in some directions, and the averaging approach is known to be better than others [10].

In [9], Murphy proposed an averaging rule to avoid the nonintuitive combination in ET. When there are N evidence bodies, Murphy's rule first calculates the average of each hypothesis for the evidence and applies the D-S combination rule with the averages $N-1$ times. That is, Eq. (2) is modified as follows.

$$\tilde{m}(A) = \frac{\sum_{B \cap C = A} \bar{m}(B)\bar{m}(C)}{1 - \bar{K}}, \tag{4}$$

where

$$\bar{K} = \sum_{B \cap C = \phi} \bar{m}(B)\bar{m}(C). \tag{5}$$

Here, $\bar{m}(B)$ and $\bar{m}(C)$ are the averages of evidence for B and C, respectively. Here, all bodies of evidence have the same importance with the same weight in calculating the average, which is not always the case. Instead of a simple averaging rule, some other researchers have tried a weighted sum of evidence bodies [11, 12].

2.2. α-Integration

Here we provide a brief overview of α-integration, more details on which can be found in [13]. Let us consider two positive measures of random variable x, denoted by $m_1(x) > 0$ and $m_2(x) > 0$ for $i = 1, 2$. α-*mean* [13] is a one-parameter family of means, defined by

$$\widetilde{m}_\alpha(x) = f_\alpha^{-1}\left(\frac{1}{2}\{f_\alpha(m_1(x)) + f_\alpha(m_2(x))\}\right), \tag{6}$$

where $f_\alpha(\cdot)$ is a differentiable monotonic function given by

$$f_\alpha(z) = \begin{cases} z^{\frac{1-\alpha}{2}}, & \alpha \neq 1, \\ \log z, & \alpha = 1. \end{cases} \tag{7}$$

α-mean includes various commonly used means as its special case: for $\alpha = -1, 1, 3, \infty$ or $-\infty$, α-mean becomes arithmetic mean, geometric mean, harmonic mean, minimum, or maximum, respectively. The value of the parameter α (which is usually specified in advance and fixed) reflects the characteristics of the integration. As α increases, α-mean resorts more to the smaller one between $m_1(x)$ and $m_2(x)$ (more optimistic), while as α decreases, the larger one is considered with more weight [13].

α-mean can be generalized to the weighted α-mixture of M positive measures $m_1(x)$, ..., $m_M(x)$ with weights $\boldsymbol{w} = [w_1, w_2, \ldots, w_M]$, which is referred to as α-*integration* of $m_1(x)$, ..., $m_M(x)$ with weights $\boldsymbol{w}$ [13].

Definition 1 (α-integration) *The α-integration of $m_i(x)$, $i = 1, \ldots, M$, with weights w_i is defined by*

$$\widetilde{m}(x) = f_\alpha^{-1}\left(\sum_{i=1}^{M} w_i f_\alpha(m_i(x))\right), \tag{8}$$

where $w_i > 0$ for $i = 1, \ldots, M$ and $\sum_{i=1}^{M} w_i = 1$.

Given M positive measures, $m_i(x)$, $i = 1, \ldots, M$, the goal of integration is to seek their weighted average $\widetilde{m}(x)$ that is as close to each of the measures as possible, while how close two positive measures are is evaluated using a divergence measure. It was shown by Amari that α-integration $\widetilde{m}(x)$ is optimal in the sense that the risk function

$$\mathcal{J}_\alpha[\widetilde{m}(x)] = \sum_{i=1}^{M} w_i D_\alpha[m_i(x) \| \widetilde{m}(x)] \tag{9}$$

is minimized, where $D_\alpha[m_i(x) \| \widetilde{m}(x)]$ is the α-*divergence* of $\widetilde{m}(x)$ from the measures $m_i(x)$ [13].

3. α-INTEGRATION OF MULTIPLE EVIDENCE

In this section, we propose a generalized evidence theory, α-*Integration of Multiple Evidence*, by applying α-integration to evidence theory in two ways. First, we interpret the D-S combination rule as a special form of α-integration and generalize it with α-integration. Second, α-integration is directly applied to the previous averaging approaches for ET.

3.1. α-integration on the D-S combination rule

The reason the D-S combination rule has nonintuitive results for some cases is that the combination rule is like an AND operator. So, once one evidence with 0 belief comes up, then the combined belief is always 0 no matter how the other evidence bodies are. It is because the combination rule is based on a multiplication operator. Here, we interpret the combination rule as a kind of averaging methods and propose a generalized method which can avoid the same problem as the original D-S combination rule has.

The D-S combination rule in Eq. (2) can be rewritten as follows.

$$\tilde{m}(A) = \frac{\sum_{B \cap C = A} m_{12}(BC)^2}{1 - K}, \tag{10}$$

where

$$m_{12}(BC) = \sqrt{m_1(B)m_2(C)}. \tag{11}$$

This is the geometric mean of $m_1(B)$ and $m_2(C)$.

Now using the fact that the geometric mean is a special case of α-integration with $\alpha = 1$, we generalize it as follows.

$$\tilde{m}(A) = \frac{\sum_{B \cap C = A} m_{\alpha,12}(BC)^2}{1 - K}, \tag{12}$$

where

$$m_{\alpha,12}(BC) = f_\alpha^{-1}\left[f_\alpha(m_1(B)) + f_\alpha(m_2(C))\right]. \tag{13}$$

Eq. (12) armed with Eq. (13) is a more generalized combination rule, which can avoid the nonintuitive problem by choosing the α value carefully, which is confirmed in section 4.

3.2. α-integration on the averaging rules

In this section, we also generalize the previous averaging methods with α-integration. By applying α-integration, the previous averaging methods which are arithmetic means with different weights can be replaced with

$$\bar{m}_\alpha(A_i) = f_\alpha^{-1}\left(\sum_{j=1} w_j f_\alpha(m_j(A_i))\right), \tag{14}$$

where w_j are obtained as in the previous averaging methods. Even though in experiment we used one example of the averaging rules, the probabilistic weight method in [12], it can be applied to any other averaging methods.

After calculating the α-integration of all the hypotheses, we apply the D-S combination rule $N-1$ times as other averaging methods do. With the new averaging rule in Eq. (14), the combination rule in Eq. (4) is modified as follows.

$$\tilde{m}_\alpha(A) = \frac{\sum_{B \cap C = A} \bar{m}_\alpha(B)\bar{m}_\alpha(C)}{1 - \bar{K}_\alpha}, \tag{15}$$

where

$$\bar{K}_\alpha = \sum_{B \cap C = \phi} \bar{m}_\alpha(B) \bar{m}_\alpha(C). \qquad (16)$$

Note that we can apply α-integration to both the D-S combination rule in the previous section and the averaging rule here at the same time. We show how the generalized rules work in section 4.

4. EXPERIMENTS

In order to show the useful behavior of our method, we carried out experiments with two different data sets used in two previously published papers: (a) the data set in [16] (Data 1) and (b) the data set in [11] (Data 2). Actually both data sets are for target recognition systems where there is one true target (for both cases, hypothesis A is the true target) with multiple evidence. We tested our proposed methods with different α values. Note that the D-S combination rule and the previous averaging methods are special cases with specific α values. For the comparison between the previous averaging methods, see [12].

Xu's data (Data 1): Xu's method is not an averaging approach but is based on grey relational analysis (GRA) [16]. They proposed their method as an alternative to averaging method such as Murphy's and Yong's. We tested this data to show how our method works based on the different α values.

Table 1. Evidence of Data 1

Belief	m_1	m_2	m_3	m_4	m_5	m_6	m_7
$m(A)$	0.53	0	0.54	0.57	0.55	0.55	0.56
$m(B)$	0.15	0.9	0.06	0.3	0.2	0.1	0.15
$m(C)$	0.32	0.1	0.4	0.13	0.25	0.35	0.29

One of the belief tables used in [16] is in Table 1. The hypothesis A is the real target, so we focus on the belief assigned to A. Note that the second BBA for the hypothesis A is zero which is seriously conflicted with other evidence. Intuitively, the hypothesis A should have a dominant belief after combination and the hypotheses B and C should go close to zero. Also, the influence of the second evidence is expected to be decreased as evidence bodies are added.

Table 2. Evidence of Data 2

Belief	m_1	m_2	m_3	m_4	m_5
$m(A)$	0.5	0	0.55	0.55	0.6
$m(B)$	0.2	0.9	0.1	0.1	0.1
$m(C)$	0.3	0.1	0	0	0
$m(A,C)$	0	0	0.35	0.35	0.3

Yong's data (Data 2): The belief table used in [11] is in Table 2. As mentioned earlier, D-S theory can have a set of elements as one hypothesis. In this table, there are 3 elements (A, B and C) and 4 hypotheses with ($\{A, C\}$) in addition to the 3 elements. There are 5 evidence bodies and the second one is seriously conflicted as in the previous data.

4.1. α-integration on the D-S combination rule

In Fig. 1, we can see that the original D-S combination rule which is a special case of our proposed method with $\alpha = 1$ is not the best way to combine the evidence bodies. Rather, when α is around -1, the combination result is more desirable.

Note that the results for hypothesis A with $\alpha = 1, 2, 3$ are zero after evidence 2 no matter how high other BBAs are, because it ignores all the conflicting evidence which can be interpreted as an AND operation as mentioned in [10]. Since α-integration is a monotonically decreasing function, once the result of a certain α value is

the minimum, then the bigger α values have the same minimum results as shown in Fig. 1. On the other hand, in the cases of $\alpha < 1$, the results are not against our intuitive reasoning.
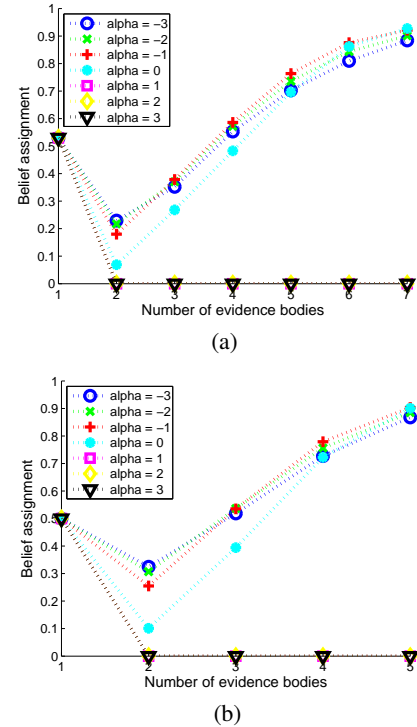

(a)

(b)

Fig. 1. The D-S combination rule is generalized with several α values. The belief assignments of the hypothesis A with different α values are shown for (a) Data 1 and (b) Data 2.

4.2. α-integration on an averaging rule

When an averaging rule is generalized with α-integration, the belief assignments of the hypothesis A are similar to Fig. 1. Fig. 2 shows the belief value of the hypothesis C which is expected to be decreased as the evidence bodies are added. In Fig. 2a, smaller values for α than -1 are generally better and in Fig. 2b, roughly speaking, $\alpha = 0$ seems better than others. That is, the best α value depends on each situation and we can wisely choose the value to get a more desirable result from multiple evidence. Also, note that even though we used only integer values for α, any real number would be possible.

4.3. α-integration on both

In this section, we applied α-integration to both the D-S combination rule and an averaging rule. In Fig. 3, to make the figures clearer, we tested two α values for the averaging rules and six α values for the D-S combination rule. Fig. 3 shows the converged belief assignment after combining all the evidence. We can see with different α values, the results are changing. That is, this generalized method gives us more room to improve the performance by fitting the α value. Note that the previous averaging method is a special case with $\alpha = -1$ (arithmetic mean) for the averaging rule and $\alpha = 1$ (geometric mean) for the combination rule. With different α values, we could get better results.
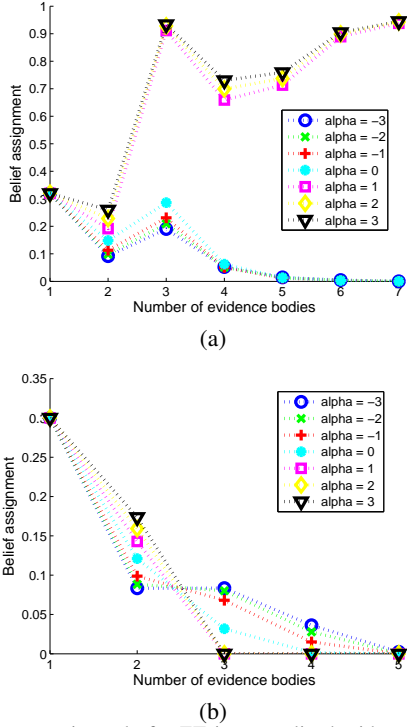
(a)

(b)

Fig. 2. One averaging rule for ET is generalized with several α values. The belief assignments of the hypothesis C with different α values for the averaging rule are shown for (a) Data 1 and (b) Data 2.

(a)

(b)

Fig. 3. The converged belief assignments of the hypothesis A from several α to both the D-S combination rule and the averaging rule for (a) Data 1 and (b) Data 2

5. CONCLUSION

We interpreted the D-S combination rule and the averaging methods for ET as a special case of α-integration and proposed generalized methods. We could apply both generalization methods at the same time. We can choose the α value for each data set to get a more desirable result that is more like human intuition. Even though the algorithms were tested with two synthetic data sets, we can apply the proposed methods to real target recognition problems without any difference.

In this paper, we discussed two generalization methods assuming the α values were given manually. So, a promising future direction is to develop algorithms to find out the α value automatically from each data set.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 369–376, 1997.

[2] J. Pearl, *Proabilistic Inference in Intelligent Systems*, Morgan Kaufmann, 1999.

[3] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of Statistics*, vol. 28, pp. 325–339, 1967.
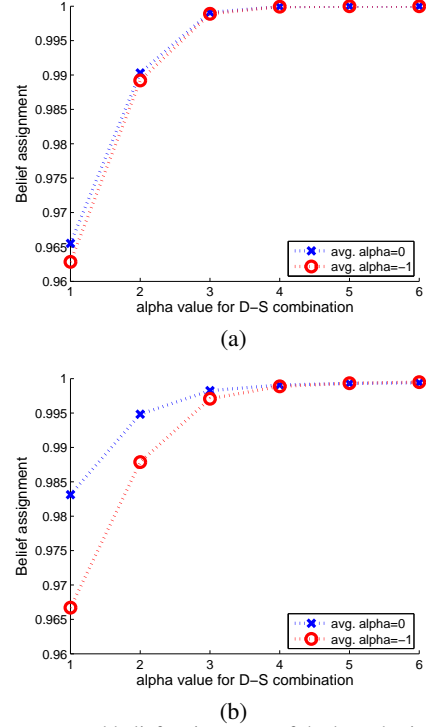
[4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, Princeton University Press, 1976.

[5] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[6] G. R. G. Lanckriet, M. Deng, N. Cristianini, M. I. Jordan, and W. S. Noble, "Kernel-based data fusion and its application to protein function prediction in yeast," in *Proc. Pacific Symposium on Biocomputing (PSB)*, Big Island, HI, 2004.

[7] H. Choi, S. Choi, and Y. Choe, "Manifold integration with Markov random walks," in *Proc. Association for the Advancement of Artificial Intelligent (AAAI)*, Chicago, IL, 2008.

[8] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, pp. 131–163, 1997.

[9] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision Support Systems*, vol. 29, pp. 1–9, 2000.

[10] K. Sentz and S. Ferson, "Combination of evidence in Dempster-Shafer theory," Tech. Rep. Sandia report SAND2002-0835, Albuquerque, NM, 2002.

[11] D. Yong, S. WenKang, Z. ZhenFu, and L. Qi, "Combining belief functions based on distance of evidence," *Decision Support Systems*, vol. 38, pp. 489–493, 2004.

[12] H. Choi, A. Katake, S. Choi, Y. Kang, and Y. Choe, "Probabilistic combination of multiple evidence," in *Proc. Int'l Conf. Neural Information Processing*, Bangkok, Thailand, 2009.

[13] S. Amari, "Integration of stochastic models by minimizing α-divergence," *Neural Computation*, vol. 19, pp. 2780–2796, 2007.

[14] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, pp. 79–81, 1991.

[15] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, pp. 1771–1800, 2002.

[16] G. Xu, W. Tian, L. Qian, and X. Zhang, "A novel conflict reassignment method based on grey relational analysis (GRA)," *Pattern Recognition Letters*, vol. 28, pp. 2080–2087, 2007.