

Distance Metrics of D Numbers

Liguo FEI^{a,1} and Yuqiang FENG^a

^a*School of Management, Harbin Institute of Technology, Harbin 150001, China*

Abstract. Belief function has always played an indispensable role in modeling cognitive uncertainty. As an inherited version, the theory of D numbers has been proposed and developed in a more efficient and robust way. Within the framework of D number theory, two more generalized properties are extended: (1) the elements in the frame of discernment (FOD) of D numbers do not required to be mutually exclusive strictly; (2) the completeness constraint is released. The investigation shows that the distance function is very significant in measuring the difference between two D numbers, especially in information fusion and decision. Modeling methods of uncertainty that incorporate D numbers have become increasingly popular, however, very few approaches have tackled the challenges of distance metrics. In this study, the distance measure of two D numbers is presented in cases, including complete information, incomplete information, and non-exclusive elements.

Keywords. D numbers, Distance metrics, Belief functions

1. Introduction

Belief function, proposed by Dempster [1] and developed by Shafer [2], is crucial in modeling a broad range of uncertainty. In recent years, this avenue of research has been considerable advancements triggered by increasing applications in information fusion [4] and decision making [5]. As an extended version, the D number theory was put forward by Deng [3], which inherits the advantages of the theory of belief function but at the same time had certain expansion mainly reflected in (1) the elements in FOD do not required to be mutually exclusive strictly (2) the completeness constraint is released.

Since D number theory was put forward, it has been widely used in various uncertain modeling problems, such as supplier selection [6], risk assessment [7], etc. In addition, its theory is gradually expanded and perfected [8]. The investigation shows that the distance function is very significant in measuring the difference between two D numbers. Modeling methods of uncertainty that incorporate D numbers have become increasingly popular, however, very few approaches have tackled the challenges of distance metrics. In [9], authors defined a distance function to measure the distance between two D numbers, which is a generalization of the distance measure of belief function. However, this distance function has high computational complexity and is not convenient and flexible to use. As a result, methods with high flexibility and case-by-case processing power are largely unexplored. In this study, the distance measure of two D numbers is presented in cases, including complete information, incomplete information, and non-exclusive elements.

¹Corresponding Author: Liguo Fei, School of Management, Harbin Institute of Technology, Harbin 150001, China; E-mail: feiliguohit@163.com.

2. D numbers

In this section, the basic concept of D numbers is recalled.

Definition 1 [3] Let a finite nonempty set Ω denote the problem domain. D numbers can be denoted as a mapping formulated by

$$D: \Omega \rightarrow [0, 1], \text{ with } D(\emptyset) = 0 \text{ and } \sum_{B \subseteq \Omega} D(B) \leq 1 \quad (1)$$

where \emptyset is the empty set and B is a subset of Ω . In addition, for a discrete set $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$, $b_i \in \mathbb{R}$, a special form of D numbers can be expressed by

$$D(\{b_1\}) = v_1, D(\{b_2\}) = v_2, \dots, D(\{b_i\}) = v_i, \dots, D(\{b_n\}) = v_n \quad (2)$$

or simply denoted as $D = \{(b_1, v_1), (b_2, v_2), \dots, (b_i, v_i), \dots, (b_n, v_n)\}$, where $v_i > 0$ and $\sum_{i=1}^n v_i \leq 1$. If $\sum_{B \subseteq \Omega} D(B) = 1$, the information is said to be complete; if $\sum_{B \subseteq \Omega} D(B) < 1$, the information is said to be incomplete.

3. Distance measure between two D numbers

In this section, we shall define the distance of D numbers from three perspectives. First, we consider the case where information is complete and elements in FOD are mutually exclusive. Second, we extend it to the case where information is incomplete. Finally, we discuss the case where elements in FOD are not mutually exclusive.

3.1. The case where the information is complete and the elements in FOD are mutually exclusive

Definition 2 Suppose d_1 and d_2 are two D numbers within the discrete set $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$, $b_i \in \mathbb{R}$, including n mutually exclusive and exhaustive hypotheses, and their information is complete, that is, $\sum_{B \subseteq \Omega} D(B) = 1$. The distance between d_1 and d_2 can be defined as follows.

$$d_{D\text{-number}}(d_1, d_2) = 1 - \frac{\langle \vec{d}_1, \vec{d}_2 \rangle}{|\vec{d}_1| \cdot |\vec{d}_2|} \quad (3)$$

where \vec{d}_1 and \vec{d}_2 are the vector forms of d_1 and d_2 , which can be denoted as $\vec{d}_1 = (v_1^1, v_2^1, \dots, v_n^1)^T$ and $\vec{d}_2 = (v_1^2, v_2^2, \dots, v_n^2)^T$. Therefore, the definition of distance measure in Eq.(3) can also be expressed as

$$d_{D\text{-number}}(d_1, d_2) = 1 - \frac{\sum_{i=1}^n v_i^1 \cdot v_i^2}{(\sum_{i=1}^n (v_i^1)^2)^{1/2} + (\sum_{i=1}^n (v_i^2)^2)^{1/2}} \quad (4)$$

where v_i^1 is the belief of b_i in d_1 and v_i^2 means the belief of b_i in d_2 .

An illustrative example is given to show the calculation of the distance between two D numbers defined above.

Example 1 Let Ω be a FOD with 2 linguistic constants, namely $\Omega = \{Good, Bad\}$, and they are mutually exclusive. Given two pairs of D numbers: $d_1(\{Good\}) = 1$, $d_1(\{Bad\}) = 0$; $d_2(\{Good\}) = 0$, $d_2(\{Bad\}) = 1$. The distance between the two D numbers can be calculated as below:

$$d_{D-number}(d_1, d_2) = 1 - \frac{1 \times 0 + 0 \times 1}{(1^2 + 0^2)^{1/2} + (0^2 + 1^2)^{1/2}} = 1 \tag{5}$$

Remark 1 From Definition 2 and Example 1, it can be seen some conditions are added to the defined distance measure that the elements in set Ω require mutually exclusive and the completeness is needed, that is, the above definition can work when D numbers is degenerated into D-S theory, but what if these limitations are exceeded? The further discussion will be conducted below.

As shown in Example 1, if the information is complete for the two D numbers, the reasonable result can be obtained, but what if the condition is released? Let's consider the following example.

Example 2 Let Ω be a FOD with 2 linguistic constants, namely $\Omega = \{Good, Bad\}$, and they are mutually exclusive. Given two pairs of D numbers: $d_1(\{Good\}) = 0.4$, $d_1(\{Bad\}) = 0.6$; $d_2(\{Good\}) = 0.2$, $d_2(\{Bad\}) = 0.3$. The distance between the two D numbers can be calculated as below:

$$d_{D-number}(d_1, d_2) = 1 - \frac{0.4 \times 0.2 + 0.6 \times 0.3}{(0.4^2 + 0.6^2)^{1/2} + (0.2^2 + 0.3^2)^{1/2}} = 0 \tag{6}$$

Apparently, d_1 and d_2 are two different D numbers, but their distance is 0 in above example, which is obviously unreasonable. This case shows that Definition 2 is invalid when information is incomplete for D numbers. To address this issue, the definition of distance measure between two D numbers is improved as follows.

3.2. The case where the information is incomplete and the elements in FOD are mutually exclusive

Definition 3 Suppose d_1 and d_2 are two D numbers within the discrete set $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$, $b_i \in R$, including n mutually exclusive and exhaustive hypotheses. The distance between d_1 and d_2 can be defined as follows.

$$d_{D-number}^+(d_1, d_2) = 1 - \frac{\langle \vec{d}_1^+, \vec{d}_2^+ \rangle}{|\vec{d}_1^+| \cdot |\vec{d}_2^+|} \tag{7}$$

where \vec{d}_1^+ and \vec{d}_2^+ are the vector forms of d_1^+ and d_2^+ , which can be denoted as $\vec{d}_1^+ = (v_1^1, v_2^1, \dots, v_n^1, v_{n+1}^1)^T$ and $\vec{d}_2^+ = (v_1^2, v_2^2, \dots, v_n^2, v_{n+1}^2)^T$. v_{n+1}^i represents the belief of the incomplete part in D number d_i , that is, if the information is complete of d_i , $v_{n+1}^i = 0$;

otherwise, $v_{n+1}^i = 1 - \sum_{i=1}^n v_i$. Also, the definition of distance measure in Eq.(7) can also be expressed as

$$d_{D\text{-number}}^+(d_1, d_2) = 1 - \frac{\sum_{i=1}^{n+1} v_i^1 \cdot v_i^2}{(\sum_{i=1}^{n+1} (v_i^1)^2)^{1/2} + (\sum_{i=1}^{n+1} (v_i^2)^2)^{1/2}} \tag{8}$$

An illustrative example, continuing the Example 2, is also given to show the calculation of the distance between two D numbers based on the definition above.

Example 3 (Continued Example 2). Let Ω be a FOD with 2 linguistic constants, namely $\Omega = \{Good, Bad\}$, and they are mutually exclusive. Given two pairs of D numbers: $d_1(\{Good\}) = 0.4$, $d_1(\{Bad\}) = 0.6$; $d_2(\{Good\}) = 0.2$, $d_2(\{Bad\}) = 0.3$. The distance between the two D numbers can be calculated as below: it is obvious the information of d_2 is incomplete, so \vec{d}_1 and \vec{d}_2 should be transformed by $\vec{d}_1^+ = (0.4, 0.6, 0)^T$ and $\vec{d}_2^+ = (0.2, 0.3, 0.5)^T$

$$d_{D\text{-number}}^+(d_1, d_2) = 1 - \frac{0.4 \times 0.2 + 0.6 \times 0.3 + 0 \times 0.5}{(0.4^2 + 0.6^2 + 0^2)^{1/2} + (0.2^2 + 0.3^2 + 0.5^2)^{1/2}} = 0.4151. \tag{9}$$

Compared with the result in Example 2, the distance measure above is obviously more reasonable.

Remark 2 It can be noticed by observing that in Definitions 2 and 3 a constraint is added that elements in the FOD Ω need to be mutually exclusive, which indicates that the distance between two D numbers can only be measured within this condition. However, according to the definition of D number, the elements in set Ω do not require mutually exclusive, so how to measure it in this case?

The following example is given firstly, which is another version of Example 1 by changing a condition.

Example 4 Let Ω be a FOD with 2 linguistic constants, namely $\Omega = \{Good, Bad\}$, and their relationship is shown in Fig.1, that means the two linguistic constants are not exclusive (as distinguished from Example 1). Given two pairs of D numbers: $d_1(\{Good\}) = 1$, $d_1(\{Bad\}) = 0$; $d_2(\{Good\}) = 0$, $d_2(\{Bad\}) = 1$. As the previous definition, the distance between d_1 and d_2 should be 0, however, in this case we consider it unreasonable because there is an intersection between elements Good and Bad, so if the condition elements in Ω are mutually exclusive cannot be satisfied, Definitions 2 and 3 fail to measure the distance between D numbers. To solve this problem, the further definition are developed as follows.

3.3. The case where the elements in FOD are not mutually exclusive

Definition 4 Suppose d_1 and d_2 are two D numbers within the discrete set $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$, $b_i \in R$. The distance between d_1 and d_2 can be defined as:

$$d_{D\text{-number}}(d_1, d_2) = 1 - \frac{\vec{d}_1 \cdot \underline{D} \cdot \vec{d}_2^T}{|\vec{d}_1| \cdot |\vec{d}_2| + \vec{d}_1 \cdot \underline{D} \cdot \vec{d}_2^T} \tag{10}$$

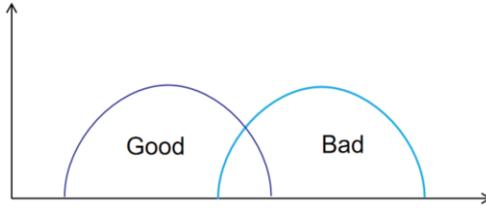


Figure 1. The relationship between two linguistic terms.

where \underline{D} is a $(n \times n)$ -dimensional matrix if the information of d_1 and d_2 is complete ;otherwise, it is a $(n + 1) \times (n + 1)$ -dimensional matrix whose elements are

$$\underline{D} = \begin{cases} \frac{A \cap B}{A \cup B} & A \neq B & A, B \in \Omega / \{\Omega, b_{n+1}\} \\ 1 & A = B & A, B \in \Omega / \{\Omega, b_{n+1}\}, \end{cases}$$

and \underline{D} is also a $(n \times n)$ -dimensional matrix if the information of d_1 and d_2 is complete ;otherwise, it is a $(n + 1) \times (n + 1)$ -dimensional matrix whose elements are

$$\underline{D} = \begin{cases} \frac{A \cap B}{A \cup B} & A \neq B & A, B \in \Omega / \{\Omega, b_{n+1}\} \\ 0 & A = B & A, B \in \Omega / \{\Omega, b_{n+1}\}. \end{cases}$$

if the information of d_1 and d_2 is incomplete, $A, B \in \{b_1, b_2, \dots, b_n, b_{n+1}\}$, which has been introduced in Definition 3.

The above definition considers another special property of D numbers, that is, the elements do not need to be mutually exclusive, which makes the distance measure between D numbers more perfect. To illustrate the calculation process of this definition, the issue proposed in Example 4 will be addressed below.

Example 5 (Continued Example 4). Let Ω be a FOD with 2 linguistic constants, namely $\Omega = \{Good, Bad\}$, and their relationship is shown in Fig. 1. Given two pairs of D numbers: $d_1(\{Good\}) = 1, d_1(\{Bad\}) = 0; d_2(\{Good\}) = 0, d_2(\{Bad\}) = 1$. Since elements in Ω do not mutually exclusive, so the distance measure between d_1 and d_2 depends not only on their belief distribution but also on the relationship among the elements. Firstly, matrix \underline{D} and \underline{D} can be obtained based on the relationship between "Good" and "Bad" in Fig. 1

$$\underline{D} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \quad \underline{D} = \begin{bmatrix} 0 & 0.2 \\ 0.2 & 0 \end{bmatrix}.$$

The distance can be measured based on Definition 4 as

$$d_{D-number}(d_1, d_2) = 1 - \frac{[1 \ 0] \cdot \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{1 + [1 \ 0] \cdot \begin{bmatrix} 0 & 0.2 \\ 0.2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}} = 0.8333. \tag{11}$$

It is obvious the above result is more reasonable.

4. Conclusion

In this work, we define the distance function between two D numbers. We employ some examples to demonstrate the use of distance function, and its rationality is illustrated by several cases. Our results show that it is valid and reasonable to define the distance function of D numbers according to different cases. In the following research, we will perfect the definition of the distance function of D numbers and further extend the theoretical results to applications, such as failure mode and effects analysis, risk analysis, multi-sensor data fusion.

References

- [1] Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* **38**(2), 325--339 (1967)
- [2] Shafer, G.: *A mathematical theory of evidence*, vol. 1. Princeton university press Princeton (1976)
- [3] Deng, Y.: D numbers: Theory and applications. *Journal of Information and Computational Science* **9**(9), 2421--2428 (2012)
- [4] Lin, Z., Zhang, Z., Liu, Y., Dezert, J., Pan, Q.: A new pattern classification improvement method with local quality matrix based on K-NN. *Knowledge-based Systems* **164**, 336--347 (2019)
- [5] Thierry, D.: Decision-making with belief functions: A review. *International Journal of Approximate Reasoning* **109**, 87--110 (2019)
- [6] Deng, X., Hu, Y., Deng, Y., Mahadevan, S.: Supplier selection using ahp methodology extended by d numbers. *Expert Systems with Applications* **41**(1), 156--167 (2014)
- [7] Lin, S., Li, C., Xu, F., Liu, D., Liu, J.: Risk identification and analysis for new energy power system in china based on d numbers and decision-making trial and evaluation laboratory (dematel). *Journal of Cleaner Production* **180**, 81--96 (2018)
- [8] Mo, H., Deng, Y.: A new aggregating operator for linguistic information based on d numbers. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **24**(06), 831-846 (2016)
- [9] Li, M., Hu, Y., Zhang, Q., Deng, Y.: A novel distance function of d numbers and its application in product engineering. *Engineering Applications of Artificial Intelligence* **47**, 61-67 (2016)