

# 基于 FDR 的证据理论改进算法

侯庆山, 邢进生

(山西师范大学 数学与计算机科学学院, 山西 临汾 041000)

**摘要:**证据理论的合成规则作为一项重要的研究课题,是样本分类及决策的关键制约因素。大多数融合方法随着特征数量的增加,特征间关联随之增强,融合过程也变得更为复杂,进而导致证据融合结果不够理想。因此,提出一种基于特征降维的证据理论改进算法,该算法主要包括两方面:首先,对原始数据集进行特征降维(feature dimension reduction, FDR),降低数据集中样本特征之间的关联性,进而优化融合结果;其次,对降维后的数据集进行特征融合,由于原始数据样本特征数量的下降,融合过程也变得更为简单,进一步通过计算样本的基本概率分配(basic probability assignment, BPA),得出样本分类结果。实验表明,基于特征降维的证据理论改进算法相较于其他融合算法,融合过程更为简单,融合效果较为良好,经过 Instacart 数据集测试,最终的平均类型识别率为 94%。

**关键词:**证据理论;组合规则;BPA;样本分类;特征降维

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2020)06-0059-06

doi:10.3969/j.issn.1673-629X.2020.06.012

## Improved Algorithm of Evidence Theory Based on Feature Dimension Reduction

HOU Qing-shan, XING Jin-sheng

(School of Mathematics and Computer Science, Shanxi Normal University, Linfen 041000, China)

**Abstract:** As an important research topic, the synthesis rule of evidence theory is the key constraints of sample classification and decision making. In most fusion methods, as the number of features increases, the correlation among features increases, and the fusion process becomes more complex, leading to unsatisfactory results of evidence fusion. Therefore, an improved algorithm of evidence theory based on feature dimension reduction is proposed, which mainly includes two aspects. Firstly, feature dimension reduction is performed on the original data set to reduce the correlation between the sample features in the data set, so as to optimize the fusion results. Secondly, feature fusion is carried out for the data set after dimension reduction. Due to the decline in the number of features of the original data samples, the fusion process becomes simpler. Further, the classification results of the samples are obtained by calculating the basic probability assignment of the samples. Experiments show that compared with other fusion methods, the proposed algorithm has a simpler fusion process and a better fusion effect. Through the Instacart data set test, the final average type recognition rate is 94%.

**Key words:** evidence theory; evidence combination rule; BPA; sample classification; feature dimension reduction

## 0 引言

证据理论<sup>[1]</sup>是模糊决策的有效工具,在目标识别、决策、优化<sup>[2-3]</sup>等问题中得到了广泛的应用。为了简化样本在特征融合过程中的复杂性,往往在特征融合前对原始样本的特征进行简单的筛选、降维<sup>[4]</sup>等一系列处理。此外,样本特征权重分配对实验结果的影响较大,对此可以采用这样的方法:在特征融合时,对于相互矛盾的样本特征,可以分配较低权重或分配权重为零。但是,当数据样本特征过多时,这样的处理方式

往往是过于复杂的。

为了减小证据冲突的影响,文献[5]提出了一种基于将证据交集部分转化为联合部分的组合规则。但是,随着冲突的增加,该方法的性能变得越来越差。文献[6]提出了一种基于计算所有证据平均值的证据组合方法,但并没有考虑证据权重对于融合结果的影响,将所有的权重设定为同一值。因此,为了更好地确定每个证据的权重,引入证据相似性来计算每个证据的冲突程度,模糊度较大的证据由于包含的信息少而赋

收稿日期:2019-07-11

修回日期:2019-11-13

基金项目:山西省软科学基金资助项目(2011041033-03)

作者简介:侯庆山(1995-),男,硕士研究生,研究方向为数据挖掘、计算智能;通讯作者:邢进生(1964-),男,教授,博士,研究方向为计算智能、数据挖掘、计算机视觉。

予较小的权重值,这样的证据权重分配更为合理,因此样本分类更准确。文献[7]提出了一种基于文献[8]模糊度测量方法的证据组合规则,文献[8]提出了一种改进的基于证据距离测量的平均组合方法,每个证据主体(body of evidence, BOE)的权重都被考虑在内。但文献[8]的组合规则不能有效地处理高度冲突的证据组合。针对这一问题,文献[7]提出了一种新的加权平均证据组合方法,一方面利用证据的距离,另一方面利用不确定性度量来确定证据主体的权重,可以得到合理的组合结果。但随着样本特征增多,利用不确定性度量来确定证据主体的权重变得困难。文献[9]提出了另一种基于信息熵的证据组合规则,该方法能有效地处理高冲突证据,具有较好的收敛性能,但信息熵的确定需要经过复杂的计算,随着样本的特征数量及分类增多,这种方式的可行性越来越低。文献[10]提出了一种基于证据相似性和惩罚函数的组合规则,该方法将证据分为可信证据和不可信证据两部分。然后,用一种新的信息熵来度量证据的信息量。最后,得到每个证据的权值,并在使用 Dempster 组合规则之前对证据进行修改,该方法能有效地处理冲突证据,具有较好的收敛性,但对证据进行划分时,由于划分方法存在不足,容易对证据划分错误。文献[11]基于不一致测量值的冲突证据组合方法,提出了一种测量两种证据冲突的新方法,通过选择折现系数对矛盾证据进行修正,但改进后组合规则的融合效果提升不够明显。文献[12-13]提出了两种证据理论的组合方法,前者基于证据距离和模糊偏好,后者基于证据相似性和信念函数熵,然而,这两种方案仍然存在相似性冲突,样本特征的融合过程仍然复杂。

由于相似性碰撞现象<sup>[14]</sup>以及原始样本特征之间存在关联性,导致现有的证据融合规则过于复杂。此外,现有的融合规则存在无法有效降低冲突证据权重<sup>[6]</sup>的缺陷。在分析已有算法的基础上,文中提出一种改进的证据融合算法。在样本融合之前进行特征筛选,删除对分类影响较小的特征,简化初始样本特征数量,从而降低融合过程的复杂性。相较于其他融合算法在复杂数据集上的融合过程及融合效果,该算法的过程较为简单,融合效果更好。

## 1 基础理论

### 1.1 D-S 证据理论

Dempster 和 Shafer 提出的证据理论是一种从不确定信息中进行决策的有效工具。它在许多领域有着广泛应用。有关 D-S 证据理论的相关定义如下:

定义1:证据理论的基础。在实际应用中,通常  $2^\Omega$  是在空间  $\Omega$  上的所有子集的集合,  $2^\Omega = \{\emptyset, \{\theta_1\},$

$\{\theta_2\}, \dots, \{\theta_n\}, \{\theta_1, \theta_2\}, \dots, \{\theta_1, \theta_2, \dots, \theta_n\}\}$ 。在 D-S 理论中,数学基本概率分配<sup>[15-16]</sup>函数,也称为质量函数,满足在  $2^\Omega \rightarrow [0, 1]$  上的映射,满足式(1)。

$$\sum_{A \subseteq \Omega} m(A) = 1; m(\emptyset) = 0 \quad (1)$$

定义2:似然函数。似然函数满足式(2):

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - \bar{A}; m(A) \leq Pl(A) \quad (2)$$

其中,  $\bar{A}$  表示  $A$  的补集,是基于  $m(A)$  和  $Pl(A)$  的,  $[m(A), Pl(A)]$  表示  $A$  可能为真的概率取值范围。值得注意的是,构成证据体的所有 BPA 都应均为正值。焦点元素的所有 BPA 将构成证据的主体如下:

$$m: m(A), m(B), m(C), \dots, m(AB), \dots, m(\theta)$$

定义3:基于定义2,  $m(A)$  或  $m(B)$  是代表  $A$  或  $B$  的质量函数,根据 Dempster 提出的组合规则可以组合一组证据,满足式(3)和式(4):

$$m(A) = \begin{cases} 0, A = \emptyset \\ \frac{\sum_{A_i \cap A_j = A} m_1(A_i) m_2(A_j)}{C}, A \neq \emptyset \end{cases} \quad (3)$$

$$C = 1 - \sum_{A_i \cap A_j = \emptyset} m_1(A_i) m_2(A_j) \quad (4)$$

$m$  是  $m_1$  和  $m_2$  的组合结果。当证据数量大于2时,扩展组合规则如式(5)和式(6):

$$m(A) = \begin{cases} 0, A = \emptyset \\ \frac{\sum_{A_{i1} \cap A_{i2} \cap \dots \cap A_{in} = A} m_1(A_{i1}) m_2(A_{i2}) \dots m_n(A_{in})}{C}, A \neq \emptyset \end{cases} \quad (5)$$

$$C = 1 - \sum_{A_{i1} \cap A_{i2} \cap \dots \cap A_{in} = \emptyset} m_1(A_{i1}) m_2(A_{i2}) \dots m_n(A_{in}) \quad (6)$$

运用上述公式可以实现证据融合,但是当某一证据的支持度较小或者不被支持时,组合结果往往是不理想的。为此,一方面可以通过引入证据权值减少此类证据的影响,另一方面可以在样本融合前进行特征筛选,删除支持度小的特征。

### 1.2 特征选择

在样本特征中往往存在较高的相关度,导致在计算上产生了较大消耗;此外,部分特征对预测结果往往存在不好的影响,甚至是负影响。因此针对样本特征进行选择,即从样本的所有特征中选择部分特征作为新的样本特征,特征值在选择前后可以改变或者不改变,但是选择后的特征维数必然会降低。常用的特征选择方法主要包括以下两种:

(1)均方差分析法。

该方法一般设置某一特定的方差阈值,删除低于阈值的特征;若方差阈值未指定,则保留所有非零方差特征,删除样本中具有相同值的特征。均方差分析法往往适用于样本特征较少的数据集,利用方差对样本

数据集进行分析时,与极差和标准差相比较,方差的计算结果能够将样本特征的波动性放大,可以准确地观察出某一样本特征对于样本分类结果的影响程度。不足之处在于,利用方差对样本数据集进行分析要处理数据集上的全部数据,随着数据量和样本数的增多,计算过程将变得复杂。因此方差分析常常运用在数据量或样本特征量较少的数据集上。

## (2) 主成分分析法。

主成分分析法<sup>[17]</sup>是一种分析及简化数据集的方法,其目的是降低数据维数,尽可能地降低原始数据的复杂度并使数据损失尽可能的小。利用主成分分析法可以消减回归分析或者聚类分析中特征的数量。在高维数据集中,许多特征之间通常是线性相关的,利用主成分分析可以减小相关特征的个数,降低样本特征间的相关性。但是,对高维数据集进行主成分分析时,首先要确保提取的前几个主成分的贡献率之和应当达到较高水平;其次,对于降维后的样本特征,特征含义一般具有模糊性。在样本特征数量较少时,一般不采用此类方法,因此只有在高维数据集中才使用主成分分析法。

## 2 基于特征降维的证据理论的改进算法

原始样本数据集进行特征降维之后,减少了相关特征的个数,对降维数据生成新的BPA。对于简单数据集,采用方差分析法,选择方差大于阈值的特征。对于复杂数据集,采用主成分分析法,对样本特征进行选择和降维处理。算法过程描述如下:

输入:原始数据集矩阵

输出:降维数据集矩阵

算法步骤描述如下:

(a) 将初始数据按行列排成  $n$  行  $m$  列的矩阵  $X$ ;

(b) 将矩阵  $X$  的每一行进行零均值化处理;

(c) 求出矩阵的协方差矩阵  $C = \frac{1}{m}XX^T$ , 解它的特征值及对应的特征向量;

(d) 将特征向量按对应特征值的大小按行排列成矩阵,取前  $K$  行组成矩阵  $P$ ;

(e)  $Y = PX$  即为降维到  $K$  维之后的数据,对所得数据集样本抽样;

(f) 对抽样样本进行聚类,标注样本特征以及样本所属类型;

(g) 生成区间数模型,针对数据集的某一特征,取所有样本中的最大值以及最小值分别作为所求区间的上界以及下界;

(h) 求解某一测试样本特征值与区间数的距离。设  $A = [a_1, a_2]$  和  $B = [b_1, b_2]$  是两个区间数,则区间

距离<sup>[18]</sup>可用式(7)计算:

$$d(A, B) = \frac{1}{2} (|a_1 - b_1| + |a_2 - b_2|) \quad (7)$$

(i) 求解测试样本特征值与区间数模型之间的相似度。设  $A = [a_1, a_2]$  和  $B = [b_1, b_2]$  是两个区间数,相似度计算可采用式(8):

$$S(A, B) = \frac{1}{1 + \alpha d(A, B)} \quad (8)$$

其中,  $\alpha (\alpha > 0)$  为支持度系数。

(j) 对求解的相似度进行归一化处理,将归一化之后的结果作为最终BPA。

算法流程如图1所示。

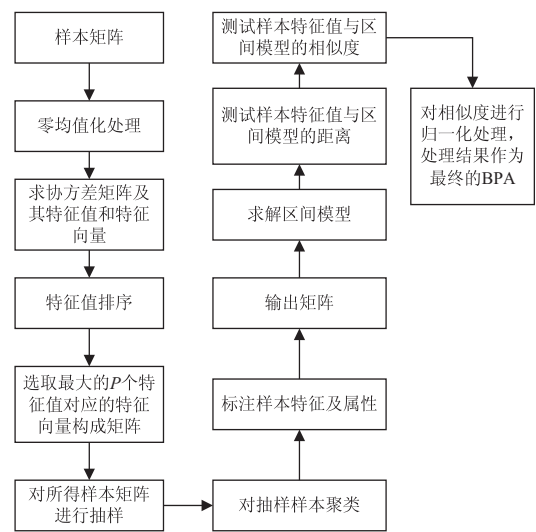


图1 算法流程

特征选择和降维是改进算法中的关键环节,为了便于理解改进算法中包含的降维算法部分,举例如下:

原始数据集矩阵  $X$ :

$$\begin{pmatrix} 1 & 1 & 2 & 4 & 2 \\ 1 & 3 & 3 & 4 & 4 \end{pmatrix}$$

去均值:

$$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$

协方差矩阵:

$$C = \frac{1}{5} \begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$$

特征值:

$$\lambda_1 = 2, \lambda_2 = \frac{2}{5}$$





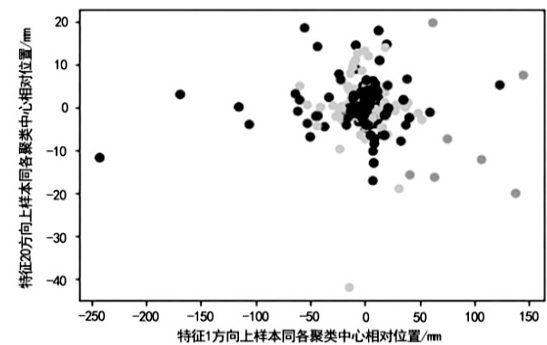


图 6 聚类分布

(5)对以上聚类结果进行分析,计算聚类数据集的轮廓系数为 0.612 7,因而可以判定聚类效果良好,轮廓系数可以采用式(9)进行计算:

$$sc_i = \frac{b_i - a_i}{\max(b_i, a_i)} \tag{9}$$

其中,  $i$  为已聚类数据中的样本,  $b^i$  为  $i$  到其它簇的所有样本的平均距离,  $a^i$  为  $i$  到本身簇的距离平均值,最终通过计算所有样本点的轮廓系数平均值对聚

类效果进行分析评定。如果  $sc_i$  小于 0,说明  $a^i$  的平均距离大于最近的其它簇;相反的  $sc_i$  的值越大,说明  $a^i$  的平均距离小于最近的其他簇,轮廓系数的值是介于  $[-1,1]$ ,越趋近于 1 代表数据集的内聚度以及分离度都相对良好。

3.4 对降维后的数据集生成 BPA 进行分类识别

(1)用降维后的样本数据作为验证数据库,选择 5 种用户类型的 400 个样本,每种类型各占 80 个,取样本中的最大值以及最小值来构造区间数模型,如表 2 所示。

(2)选取 100 个测试样本,5 类用户各占 20 个,将它们作为类别未知的测试样本。

(3)通过计算测试样本与区间数模型之间的相似度,求解出各特征值最终的 BPA。进一步通过 D-S 组合规则进行特征融和,产生最终的融合结果。

(4)对融合结果进行分析,测试样本的类型由融合结果决定,哪种类型的 BPA 的值最大,那么这个 BPA 对应的类别就是待测样本的类别。

表 2 区间数模型

| 特征    | 类型 0         | 类型 1         | 类型 2         | 类型 3         | 类型 4         |
|-------|--------------|--------------|--------------|--------------|--------------|
| 特征 1  | [-24.2,6.3]  | [-4.9,7.8]   | [-11.9,7.3]  | [-19.2,25.3] | [-14.2,9.3]  |
| 特征 2  | [-16.8,9.5]  | [-11.2,14.0] | [-29.4,15.7] | [-4.59,2.9]  | [-9.2,10.1]  |
| 特征 3  | [-14.2,7.9]  | [-13.8,11.9] | [-28.1,10.6] | [-16.3,9.4]  | [-7.4,11.0]  |
| 特征 4  | [-19.7,10.8] | [-12.0,10.8] | [-9.49,16.9] | [-9.1,17.7]  | [-11.4,16.3] |
| 特征 5  | [-28.2,14.3] | [-12.6,15.5] | [-7.21,12.6] | [-27.2,19.7] | [-16.0,18.5] |
| 特征 6  | [-22.6,8.3]  | [-25.7,22.4] | [-20.3,11.5] | [-24.0,14.0] | [-18.9,14.4] |
| 特征 7  | [-30.2,9.1]  | [-26.3,16.0] | [-19.3,8.30] | [-18.5,14.2] | [-6.7,16.7]  |
| 特征 8  | [-4.2,7.8]   | [-20.8,16.3] | [-1.80,19.5] | [-26.9,17.6] | [-12.9,13.5] |
| 特征 9  | [-14.8,14.1] | [-13.5,15.2] | [-1.59,5.01] | [-13.5,16.8] | [-22.8,10.1] |
| 特征 10 | [-27.9,16.3] | [-19.2,10.4] | [-12.8,0.35] | [-8.46,13.5] | [-11.0,17.6] |
| 特征 11 | [-28.2,14.8] | [-4.52,15.6] | [-5.56,8.57] | [-13.7,7.78] | [-1.62,4.60] |
| 特征 12 | [-21.3,5.9]  | [-1.69,12.1] | [-14.3,1.58] | [-6.89,13.4] | [-25.7,12.8] |
| 特征 13 | [-13.2,8.4]  | [-25.9,6.01] | [-26.0,0.51] | [-1.01,5.32] | [-26.0,7.10] |
| 特征 14 | [-34.2,6.6]  | [-11.4,17.9] | [-21.9,5.71] | [-14.7,10.2] | [-21.4,9.50] |
| 特征 15 | [-13.8,19.3] | [-8.43,22.8] | [-16.7,7.32] | [-13.0,4.18] | [-15.8,5.86] |
| 特征 16 | [-17.3,18.5] | [-29.1,4.96] | [-2.56,13.4] | [-15.8,6.33] | [-10.2,3.40] |
| 特征 17 | [-13.8,12.7] | [-25.3,9.26] | [-19.7,1.72] | [-24.9,14.9] | [-25.5,14.3] |
| 特征 18 | [-19.2,11.8] | [-0.13,11.1] | [-15.8,11.1] | [-17.2,10.3] | [-13.6,7.64] |
| 特征 19 | [-22.4,18.1] | [-16.7,20.4] | [-12.6,18.7] | [-21.3,6.68] | [-27.7,11.2] |
| 特征 20 | [-11.9,14.6] | [-28.9,14.5] | [-15.5,12.7] | [-9.18,10.0] | [-21.6,19.7] |
| 特征 21 | [-26.9,16.0] | [-13.0,10.6] | [-14.5,18.8] | [-9.83,16.5] | [-7.3,9.6]   |
| 特征 22 | [-26.2,21.2] | [-4.10,14.1] | [-23.1,6.73] | [-24.2,8.66] | [-3.3,19.5]  |
| 特征 23 | [-3.63,15.0] | [-12.3,12.0] | [-18.4,17.3] | [-12.6,13.2] | [-2.5,16.9]  |
| 特征 24 | [-16.2,15.9] | [-25.9,12.8] | [-10.3,16.3] | [-7.73,16.7] | [-25.5,12.1] |
| 特征 25 | [-6.96,13.6] | [-11.8,19.1] | [-14.2,16.5] | [-18.9,15.7] | [-23.4,9.3]  |

续表 2

| 特征    | 类型 0          | 类型 1          | 类型 2          | 类型 3          | 类型 4          |
|-------|---------------|---------------|---------------|---------------|---------------|
| 特征 26 | [-29.8, 10.7] | [-2.92, 7.39] | [-25.2, 16.2] | [-14.0, 14.3] | [-12.0, 19.2] |
| 特征 27 | [-26.9, 16.0] | [-4.10, 14.1] | [-19.4, 16.4] | [-22.7, 14.2] | [-18.2, 12.4] |

### 3.5 实验结果分析

为了检验改进算法在 Instacart 数据集上的分类效果,选取支持度系数  $\alpha$  的值为 5,在表 2 所示的区间数模型的条件下,对 500 个待测样本的数据集进行测试,包括 400 个已知分类的训练样本以及 100 个未知类型的测试样本。实验得出最终的平均类型识别率为 94%,对 0 类型以及 3 类型用户的识别率达到 96%,1 类型以及 4 类型的用户识别率达到 92%,2 类型用户识别率达到 94%。

## 4 结束语

由于传统证据融合算法在特征融合时过于复杂,样本间存在相似性碰撞,在复杂数据上融合效果不够理想,证据之间存在冲突等问题。文中提出了一种基于特征降维的证据理论改进算法,并采用 Instacart 数据集对改进算法进行检验。与其他相关算法相比,改进算法简化了证据融合过程中的复杂性,降低了原始样本中特征关联对于融合结果的影响。由实验结果可知,改进算法对于分类问题的平均准确率为 94%。由于主成分分析存在降维后特征含义不够明确,证据融合过程中如何将权重合理分配给相关特征等问题,所以下一步工作可以针对特征降维方法以及权重的合理分配做出改进,进一步提高证据融合的效果。

### 参考文献:

- [1] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [2] JIANG W, ZHAN J. A modified combination rule in generalized evidence theory[J]. Applied Intelligence, 2017, 46(3): 630-640.
- [3] DU W S, HU B Q. Attribute reduction in ordered decision tables via evidence theory[J]. Information Sciences, 2016, 364-365: 91-110.
- [4] ZHAO W, DU S. Spectral-spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach[J]. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(8): 4544-4554.
- [5] DUBOIS D, PRADE H. Representation and combination of uncertainty with belief functions and possibility measures[J]. Computational Intelligence, 1988, 4(3): 244-264.
- [6] MURPHY C K. Combining belief functions when evidence conflicts[J]. Decision Support Systems, 2000, 29(1): 1-9.
- [7] 韩德强, 邓 勇, 韩崇昭, 等. 基于证据距离与不确定度的证据组合法[J]. 红外与毫米波学报, 2011, 30(5): 396-400.
- [8] HAN D, DENG Y, LIU Q. Combining belief functions based on distance of evidence[J]. Decision Support Systems, 2005, 38(3): 489-493.
- [9] WANG J, XIAO F, DENG X. Weighted evidence combination based on distance of evidence and entropy function[J]. International Journal of Distributed Sensor Networks, 2016, 12(7): 3218784.
- [10] JIANG W, ZHUANG M, QIN X, et al. Conflicting evidence combination based on uncertainty measure and distance of evidence[J]. Springerplus, 2016, 5(1): 1217.
- [11] ZHAO Y, JIA R, SHI P. A novel combination method for conflicting evidence based on inconsistent measurements[J]. Information Sciences, 2016, 367: 125-142.
- [12] XIAO F. A novel evidence theory and fuzzy preference approach-based multi-sensor data fusion technique for fault diagnosis[J]. Sensors, 2017, 17(11): E2504.
- [13] XIAO F. An improved method for combining conflicting evidences based on the similarity measure and belief function entropy[J]. International Journal of Fuzzy Systems, 2018, 20(4): 1256-1266.
- [14] DE OLIVEIRA L G, ALMEIDA-FILHO A T. A multicriteria approach for analysis of conflicts in evidence theory[J]. Information Sciences, 2016, 346-347: 275-285.
- [15] 文成林, 周 哲, 徐晓滨. 一种新的广义梯形模糊数相似性度量方法及在故障诊断中的应用[J]. 电子学报, 2011, 39(3A): 1-6.
- [16] JIANG W, YANG Y, LUO Y, et al. Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers[J]. International Journal of Computers Communications & Control, 2015, 10(3): 333-347.
- [17] BISHOP C M. Neural networks for pattern recognition[J]. Agricultural Engineering International the Cigr Journal of Scientific Research & Development Manuscript Pm, 1995, 12(5): 1235-1242.
- [18] 陈春芳, 余克弟. 基于区间数距离的多属性决策方法[J]. 统计与决策, 2013(19): 69-71.
- [19] KRISHNA K, MURTY M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 1999, 29(3): 433-439.
- [20] LIKAS A, VLASSIS N, VERBEEK J. The global k-means clustering algorithm[J]. Pattern Recognition, 2003, 36(2): 451-461.