

基于改进相似性度量的扩展置信规则库规则激活方法

林燕清,傅仰耿

(福州大学 数学与计算机科学学院,福州 350116)

摘要: 针对扩展置信规则库(extended belief rule base, EBRB)系统在计算个体匹配度时会出现负值以及所有规则的激活权重都为零的问题,引入基于属性权重的欧氏距离,对传统的EBRB中相似性度量方法进行改进.此外,传统的规则激活方法激活的是所有激活权重大于零的规则,并没有考虑到规则之间的不一致性,而规则的不一致性会减弱EBRB系统的推理能力,因此将激活规则之间的一致性也考虑进来,提出一种基于改进相似性度量的规则激活方法.与传统的扩展置信规则激活方法相比,新的规则激活方法通过设置阈值来激活规则,这些被激活的规则不仅是激活权重大于零的规则,而且是不一致性最小的规则集合.最后,利用输油管道泄漏问题和多个公共分类数据集对新的规则激活方法的有效性进行验证,实验结果表明,基于改进的相似性度量的规则激活方法能够有效提高EBRB系统的推理准确性.

关键词: 扩展置信规则库;规则不一致;相似性度量;规则激活方法

中图分类号: TP391 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2018.01.003

引用格式: 林燕清,傅仰耿. 基于改进相似性度量的扩展置信规则库规则激活方法[J]. 中国科学技术大学学报, 2018,48(1):20-27.

LIN Yanqing, FU Yanggeng. A rule activation method for extended belief rule base based on improved similarity measures[J]. Journal of University of Science and Technology of China, 2018,48(1):20-27.

A rule activation method for extended belief rule base based on improved similarity measures

LIN Yanqing, FU Yanggeng

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116)

Abstract: When calculating negative individual matching degrees, there might appear negative values and all rules' activation weights may be equal to zero. To address this problem, this paper introduces the Euclidean distance which is based on attribute weights and improves the traditional similarity computational formula. In addition, the traditional rule activation method activates all rules whose activation weights are greater than zero without considering inconsistency which exists in the activated rules, since the inconsistency of activated rules will weaken the reasoning performance of EBRB systems. Hence, considering the inconsistency existing in the activated rules, a new rule activation method of EBRB based on improved similarity measures is proposed. Compared with traditional rule activation method in the EBRB, the proposed approach activates rules by setting thresholds. And these activated rules are not only greater than zero but also have the smallest inconsistency. Finally, the pipeline leak detection problem

收稿日期: 2017-05-18; 修回日期: 2017-06-22

基金项目: 国家自然科学基金(71501047,61773123);福建省自然科学基金(2015J01248)资助.

作者简介: 林燕清,女,1992年生,硕士生,研究方向:智能决策与专家系统,E-Mail: 765305442@qq.com

通讯作者: 傅仰耿,博士/副教授,E-Mail: ygfu@qq.com

and multiple public classification datasets have been employed to validate the efficiency of the new rule activation method. The experimental results show that the proposed method based on improved similarity measures can improve the reasoning accuracy of EBRB systems.

Key words: extended belief rule base; rule inconsistency; similarity measures; rule activation method

0 引言

专家系统^[1]通过利用存储在其内部的特定领域专家的知识或经验来解决该领域的高水平难题,这些难题本身通常包含各种不确定的信息,专家系统根据这些不确定信息转化成的规则进行逻辑推理得出结论。目前,专家系统已经广泛应用于医疗检测、化学工程、图像处理、石油、军事等领域^[1]。

近年来,置信规则库(belief rule base, BRB)^[1]作为一种专家系统的实现,受到了学术界越来越多的关注。BRB 由一组置信规则(belief rule)组成,置信规则是在传统的 IF-THEN 规则的结果部分加入置信分布发展得到的。BRB 系统的核心是基于证据推理算法的置信规则库推理方法(belief rule base inference methodology using the evidence reasoning approach, RIMER)^[2], RIMER 是 Yang 等^[2]基于 D-S 证据理论^[3-4]、决策理论^[5]、模糊理论^[6]和传统 If-Then 规则提出的能有效利用具有不确定性的各种类型信息,既可以处理定量信息也可以处理定性知识,具有对模糊不确定性或概率不确定性问题进行建模的能力。目前,BRB 系统已广泛应用于输油管道泄漏^[7-8]、石墨成分检测^[9]、消费者偏好预测^[10]、军事能力评估^[11]和出租车乘车概率预测^[12]等领域。

BRB 系统包含了前提属性权重、前提属性参考值、规则权重、结果置信度等参数,这些参数的取值对于 BRB 系统的推理能力都起到至关重要的作用,最初的 BRB 系统进行推理时用到的这些参数取值都是由专家根据先验知识或历史信息给出,但是对于复杂的 BRB 系统中的参数取值,专家很难给出其精确取值。为了解决 BRB 系统中参数取值的问题, Yang 等^[10]最先提出了通用的 BRB 参数学习模型并利用 Matlab 自带的 Fmincon 函数进行参数学习;在此基础上, Chen 等^[13]把前提属性参考值也考虑进来并提出了全局参数学习方法;不过以上两种参数学习方法均是建立在 Fmincon 函数的基础上,学习效率不高且十分耗时,鉴于此,常瑞等^[14]提出结合梯度下降法和二分法的参数学习方法,该方法

加快了参数学习的收敛速度,但涉及的公式推导过于复杂,不适合用来学习大量的参数;之后,基于群智能算法的参数学习方法相继提出,苏群等^[15]和王韩杰等^[16]分别将粒子群算法和差分进化算法引入到 BRB 的参数学习中以提高算法效率。这些参数学习方法需要不断迭代搜索,因此算法效率都不高。Liu 等^[17]提出了一种不需要进行参数学习的“新置信规则库”以及相应的规则生成机制,它是对 BRB 的扩展,故又被称作扩展置信规则库(extended belief rule base, EBRB),是将置信分布加入到置信规则的前件部分。EBRB 系统的规则生成机制属于数据驱动型,规则数取决于样本数据的数量,同时 EBRB 系统的推理能力也受样本数据质量的影响,数据的一致性、完整性显得尤为重要。在实际决策问题中,数据普遍存在不完整性和不一致性,一个数据记录对应于 EBRB 系统中的一条规则,因此在 EBRB 系统中,数据的不一致性是指两条或多条规则的前提属性取值大致相同,但是结果属性却完全不同,数据的不完整性是指由于激活规则的缺乏而无法进行推理得出结论。为了解决上述问题, Alberto 等^[18]提出了一种动态规则激活方法,通过调整参数来选择最合适的激活规则集合;苏群等^[19]提出构建基于 BK 树的树形索引对 EBRB 进行结构优化,通过设置阈值来减少搜索规则的数量; Yang 等^[20]提出多属性搜索框架来重建规则间的关系。

上述对 EBRB 系统的改进并没有有效克服因个体匹配度计算结果出现负值而导致没有规则被激活从而无法得出推理结果的问题。针对此问题,本文的规则激活方法改进了原本的相似性度量,新方法引入了基于属性权重的欧氏距离,使得个体匹配度计算值不会出现负值情况,而且在相似性度量中把前提属性权重考虑进来,为了同时解决决策问题中常见的规则不一致问题,本文在改进相似性度量的基础上提出相应的规则激活方法来选择不一致性最小的规则集合作为最终用来进行合成推理得出结论的规则集合,该方法主要通过设定规则激活权重阈值的方式来动态调整被激活的规则集合。本文方法不仅延续了 EBRB 简单有效的推理能力,而且有效

解决了个体匹配度负值、零激活规则、规则不一致性的问题。最后引入输油管道泄漏问题和分类数据集进行实验验证,通过将本文所提的方法和其他已有方法作对比,分析说明所提方法的有效性和可行性。

1 扩展置信规则库系统与问题提出

1.1 扩展置信规则生成

为了使置信规则能够对数据或知识中存在的 uncertainty 及不完整性有更加准确和全面的表示, Liu 等^[17]在 BRB 的基础上,将置信分布也加入到置信规则的前件部分,它提供了一种更加灵活、方便的方法来表示数据中的模糊性、不完整性和不确定性。

EBRB 由一组扩展置信规则 (extended belief rule) 组成,其中第 k 条扩展置信规则的表达形式如下:

$$R_k: \{A, \alpha^k\} \text{ then } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \dots, (D_N, \beta_N^k)\} \quad (1)$$

式中, (A, α^k) 表示 EBRB 前件部分的置信分布,也可表示成 $\{(A_{i,j}, \alpha_{i,j}^k), j=1, 2, \dots, J_i\} | i=1, 2, \dots, T_k\}$, $\alpha_{i,j}^k$ 表示第 k 条规则第 i 个前提属性相对第 j 个参考值的置信度, $A_{i,j}$ 表示第 i 个前提属性的第 j 个参考值,且第 j 个前提属性的参考值总数为 J_i , T_k 表示第 k 条规则前提属性总数, N 表示评价结果总数, L 表示规则总数, δ_i 表示第 i 个前提属性的属性权重,它权衡了该前提属性相对其他前提属性的重要程度, θ_k 表示第 k 条规则的规则权重, β_j^k 表示第 k 条规则第 j 个评价结果的置信度,每条规则的所有评价结果置信度需满足 $\sum_{j=1}^N \beta_j^k \leq 1$ 。如果

$\sum_{j=1}^N \beta_j^k = 1$, 则说明该规则是完整的,否则说明该规则是不完整的。

有别于 BRB 复杂繁琐的规则生成机制, EBRB 的规则直接由样本数据转化而来,不需要额外的信息,简单有效。EBRB 的规则生成机制属于数据驱动型,把样本数据转化成规则之前,需先把样本数据的输入转化成置信分布结构。样本数据的输入有很多类型,可以是数值型,也可以是语义型,文献^[17]有给出各种类型的输入对应的 $\alpha_{i,j}$ 确定方法,本文只简单介绍基于效用的输入信息转化方法。

假设 $U_i (i=1, 2, \dots, T_k)$ 表示第 k 个样本数据的第 i 个前提属性,其输入值为 x_i , $\gamma_{i,j}$ 表示由专家给出的第 i 个前提属性第 j 个参考值的效用值,假

设专家对参考值 $A_{i,j}$ 的偏好程度满足 $\gamma_{i,j+1} > \gamma_{i,j}$, 那么 x_i 可以等价地转化为如下的期望形式,即

$$E(x_j^k) = \{(\gamma_{i,j}, \alpha_{i,j}) | i=1, 2, \dots, T_k, j=1, 2, \dots, j_i\} \quad (2)$$

式中, $\alpha_{i,j}$ 的计算方法为

$$\alpha_{i,j} = \frac{\gamma_{i,j+1} - x_i^k}{\gamma_{i,j+1} - \gamma_{i,j}}, \gamma_{i,j} \leq x_i^k \leq \gamma_{i,j+1}, j=1, 2, \dots, J_i - 1 \quad (3)$$

$$\alpha_{i,j+1} = 1 - \alpha_{i,j}, \gamma_{i,j} \leq x_i^k \leq \gamma_{i,j+1}, j=1, 2, \dots, J_i - 1 \quad (4)$$

$$\alpha_{i,s} = 0, s \neq j, j+1, s=1, 2, \dots, J_i \quad (5)$$

通过上式可得到 $\alpha_{i,j}$ 的具体取值从而生成规则的前件部分,相应的输出评价结果置信分布可采用同样的方法产生。

1.2 扩展置信规则库的推理

BRB、EBRB 的推理方法核心都是 RIMER 方法,二者的区别是个体匹配度的计算公式不同。RIMER 方法主要包括两个步骤:根据系统输入值计算每条规则的激活权重和激活规则的合成。

假设第 i 个前提属性 U_i 取值 x_i 已经被表示成式(2)所示的形式,则 x_i 相对第 k 条规则的第 i 个前提属性的个体匹配度 S_i^k 可用两个置信分布之间的相似性来衡量,因为 EBRB 前件部分的置信分布实质上是概率分布,故 S_i^k 可借助如下欧氏距离来计算:

$$d_i^k - d^k(X_i, U_i) = \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \quad (6)$$

则 S_i^k 计算方法为

$$S_i^k = S^k(x_i, U_i) = 1 - d_i^k \quad (7)$$

接着,第 k 条规则的激活权重可按如下公式得到

$$\omega_k = \frac{\theta_k * \prod_{i=1}^{T_k} (S_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^L [\theta_l * \prod_{i=1}^{T_l} (S_i^l)^{\bar{\delta}_i}]}, \bar{\delta}_i = \frac{\delta_i}{\max_{i=1, 2, \dots, T_k} \{\delta_i\}} \quad (8)$$

式中, $0 \leq \omega_k \leq 1 (k=1, 2, \dots, L)$, $\sum_{i=1}^L \omega_i = 1$ 。如果 $\omega_k = 0$, 则说明第 k 条规则未被激活,如果所有规则的激活权重都为零,即所谓的零激活规则,则无法利用证据推理 (evidential reasoning, ER) 方法来合成规则并推导出结论,因为只有激活权重大于零的规则会被合成, ER 合成公式如下:

$$\beta_i(x) = \frac{\prod_{j=1}^L (\omega_j(x)\beta_{ij} + 1 - \omega_j(X)) - \prod_{j=1}^L (1 - \omega_j(x))}{\sum_{s=1}^N \left[\prod_{j=1}^L (\omega_j(x)\beta_{sj} + 1 - \omega_j(X)) - N * \prod_{j=1}^L (1 - \omega_j(x)) \right]} \quad (9)$$

假设 $\{(x_m, y_m) | m=1, 2, \dots, T\}$ 是一个样本数据的实际输入输出值, 根据上式可得到推理输出置信分布为

$$S(\hat{y}_m) = \{(D_j, \beta_j), j=1, 2, \dots, N\} \quad (10)$$

假设第 j 个评价结果的效用值为 $\mu(D_j)$, 则推理输出值的数值为

$$\hat{y}_m = \sum_{j=1}^N \mu(D_j) \beta_j(m) \quad (11)$$

EBRB 的推理能力就是由样本数据的实际输出 y_m 和推理输出 \hat{y}_m 之间的差距来衡量, 差值越小, 说明 EBRB 的推理能力越好。

1.3 现有规则激活方法的不足

现用 Liu 等^[17] 提出的 EBRB 构建方法对 UCI^[21] 中的 Mammographic 数据集进行建模,

$$S_3^1 = 1 - d_3^1 = 1 - \sqrt{\sum_{j=1}^5 (\alpha_{3,j} - \alpha_{3,j}^1)^2} = 1 - \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-1)^2} < 0.$$

因为 EBRB 中个体匹配度的取值范围为 $[0, 1]$, 所以只能将出现负数的个体匹配度赋值为 0. 一旦规则中有一个前提属性的个体匹配度出现零值, 通过式(8)计算得到的该规则的激活权重也会变为零, 当所有规则的激活权重都为零时, 将无法进行推理. 由此可见, 传统的规则激活方法中个体匹配度的计算容易出现负值的情况, 从而导致零激活规则问题. 同时, 传统的规则激活方法激活的是所有激活权重大于零的规则, 并没有考虑到规则之间的不一致性, 而规则之间的一致性会降低 EBRB 的推理能力. 鉴于此, Alberto 等^[18] 提出动态规则激活方法, 苏群等^[19] 提出用 BK 树对 EBRB 进行结构优化, Yang 等^[20] 提出多属性搜索框架. 这些方法能有效减少规则不一致性, 提高 EBRB 的推理能力, 但没有从根本上解决因个体匹配度计算而导致的零激活规则问题. 为此, 本文对传统的个体匹配度计算公式进行改进, 提出基于改进相似性度量的规则激活方法。

Mammographic 数据集共有 5 个前提属性, 从每个前提属性对应数据所有取值中均匀选取 5 个作为其属性参考值取值, 其中第 3 个前提属性的参考值分别为 $\{1, 1.5, 2, 2.5, 3\}$. Mammographic 数据集中有两个数据记录分别为

$$D_1 = \{4, 59, 2, 1, 3, 1\} \text{ 和 } D_2 = \{5, 80, 3, 5, 3, 1\},$$

假设 D_1 对应于 EBRB 中的第 1 条规则 R_1 , D_2 为系统输入. 通过式(3)、(4)、(5)可计算得 R_1 的第 3 个前提属性取值相对其属性参考值的置信度分别为 $\{0, 0, 1, 0, 0\}$, D_2 的第 3 个前提属性取值相对其属性参考值的置信度分别为 $\{0, 0, 0, 0, 1\}$, 通过式(6)、(7)可计算得系统输入 D_2 的第 3 个前提属性取值相对规则 R_1 的第 3 个前提属性取值的个体匹配度为

2 基于改进相似性度量的规则激活方法

2.1 改进的相似性度量

属性权重表示属性的重要程度, 拥有不同权重的两个属性, 即使取值相同, 与其他属性的距离也应该不同, 所以本文引入基于属性权重的欧氏距离代替式(6)中的普通欧氏距离, 定义如下:

$$d_i^k = d^k(x_i, U_i) \sqrt{\sum_{j=1}^{J_i} \delta_i^2 (\alpha_{i,j} - \alpha_{i,j}^k)^2} \quad (12)$$

为了避免个体匹配度出现负值的情况, 本文对传统的 EBRB 中相似性度量公式进行如下改进:

$$S_i^k = S^k(x_i, U_i) = \frac{1}{1 + d_i^k} \quad (13)$$

2.2 新规则激活方法

式(13)满足 $S_i^k \in (0, 1]$, 也就是说所有规则的激活权重都大于零, 按照传统的 EBRB 规则激活方法, 所有的 EBRB 规则都会被激活并用于最终合成得到推理结果, 但是规则之间的一致性会降低

EBRB 的推理能力. 鉴于此, 本文提出通过设置激活权重阈值的方式来降低激活规则之间的一致性. 不同的激活权重阈值对应不同的激活规则集合, 当激活权重阈值为零时, 所有规则因其激活权重大于零都会被激活, 此时的激活规则集合是所有的规则. 当激活权重阈值从零增加至某个值时, 部分激活权重小于设定的阈值就会被淘汰而不会被激活, 此时的激活规则集合就是所有规则的一个子集. 如果此时的激活规则集合不一致性比原先的规则集合不一致性小, 则选择当前的激活规则集合用于最终的合成推理. 本文提出的规则激活方法就是通过设置阈值的方式从所有规则中去掉部分规则以找到不一致性最小的规则子集. 最优的、规则不一致性最小的规则集合选择准则选用 Alberto 等^[18] 提出评判标准, 即

$$C(\Delta) = \frac{\max_{s=1,2,\dots,N} \{C_s\}}{|\Delta|} \quad (14)$$

$$C_s = |D_s; s = \arg(\max_{i=1,2,\dots,N} \{\beta_{ik}\})|$$

式中, Δ 表示激活权重大于设置的阈值的规则集合, $C(\Delta)$ 可解释为激活规则集合中评价结果相同且总数最多的规则数量与激活规则集合总数之比. 假设现在有 26 条规则被激活, 评价结果有 4 个, 评价结果相同的规则总数分别为 5, 3, 11, 7, 由式(14)可计算得 $C(\Delta) = \frac{11}{26}$. 算法如下:

算法 2.1 基于改进相似性度量的规则激活方法

```

Input 包含所有规则的扩展置信规则库 EBRB
Output 一组被激活的规则
t=0.0; /* 激活权重阈值变量 */
best_t=0.0; /* 保存最优的激活权重阈值 */
while t<=1
for each Rk of  $\phi$  do
    calculate  $\omega_k$  /* 计算所有规则的激活权重 */
    if  $\omega_k \geq$  threshold then
        /* 选择激活权重大于阈值的规则 */
         $\Delta_t = \Delta_t \cup R_k$ 
    end if
end for
if  $|\Delta_t| > 1$  then /* 计算激活规则集合的不一致性 */
    calculate C( $\Delta_t$ )
    if C( $\Delta_t$ ) > C( $\Delta_{best_t}$ ) then
        best_t=t /* 选择更优的阈值 */
    end if
end if
t=t+0.01 /* 迭代修改阈值变量 */

```

```

end while
for each Rk of  $\phi$  do
    calculate  $\omega_k$ 
    if  $\omega_k \geq$  best_t then
        /* 选择最优的激活规则集合 */
         $\Delta_{best_t} = \Delta_{best_t} \cup R_k$ 
    end if
end for

```

算法 2.1 给出了基于改进相似性度量的规则激活方法(SRA)的伪代码实现过程, 其中 Δ_t 表示当激活权重阈值设为 t 时, 激活权重大于 t 的规则集合, 此外激活权重的阈值增加幅度不一定要设为 0.01, 可根据具体应用来选择. 算法首先计算所有规则的激活权重, 然后选择激活权重大于设定的阈值的规则集合, 并计算此时的激活规则集合不一致性, 对于不同的阈值, 会有不同的激活规则集合不一致性, 选择不一致性最小的规则集合作为最终的激活规则用于后面的合成.

3 实例分析

本文引入输油管道泄漏问题和公共分类数据集为研究对象, 将本文提出的方法与现有的方法进行比较以验证该方法的有效性. 实验环境为: Inter(R) Core(TM) i5-4570 CPU @ 3.20 GHz 3.20 GHz; 8 GB内存; Windows 10 操作系统; 算法实现平台 Visual Studio 2010.

3.1 输油管道泄漏问题

为了验证本文所提方法的有效性, 引入一个具体的输油管道泄漏问题^[7] 作实例. 该实例的研究对象为安装在英国的一条 100 多千米长的输油管道, 当输油管道发生泄漏时, 管道的泄漏大小 (leak size, LS) 会随输油管道输入输出的流量差 (flow difference, FD) 和输油管道内的平均压力差 (pressure difference, PD) 变化而变化, 因此, 该实例的 EBRB 系统的前提属性为 FD 和 PD, 结果属性为 LS. 其中 PD、FD 的参考值由专家根据经验给出, 分别为 $\{-0.042 - 0.025, -0.010, 0, 0.010, 0.025, 0.042\}$ 和 $\{-10, -5, -3, 0, 1, 2, 3\}$, LS 的评价结果为 $\{0, 2, 4, 6, 8\}$.

该实验的测试数据是泄漏为 25% 时的 2 008 组数据, 训练数据则根据文献[13]的方法得到, 即按照一定比例分别从 3 个时间段各随机选取 500 组数据, 因此, 构建的 EBRB 系统总共有 1 500 条规则, 然后利用本文提出的方法构建 SRA_EBRB 系统,

并将实验结果和文献[19]列出的 Liu、Su 等的实验结果相比较,衡量的指标为平均绝对误差(mean absolute difference,MAE).

Liu_EBRB 系统、BK_EBRB 系统和 SRA_EBRB 系统产生的推理输出和测试数据的真实输出对比如图 1~3 所示.分析图 1~3 可以发现,Liu_EBRB 在 $PD \in [-0.02, 0.04]$, $FD \in [-10, 0]$ 附近产生的推理输出和测试数据的真实输出有较大差距,尤其是 $PD \in [-0.02, 0]$, $FD \in [-10, -5]$ 附近的数据.而 BK_EBRB 系统和 SRA_EBRB 系统推理输出总体上和真实输出相接近.此外,从图 2 可以看出,BK_EBRB 系统在 $PD \in [0.02, 0.04]$, $FD \in [-5, 0]$ 附近有个别数据的推理输出和真实输出仍有较小的差距.与 Liu_EBRB 及 BK_EBRB 系统相比,SRA_EBRB 系统产生的推理输出要更接近于测试数据的真实输出,这主要是因为 SRA_EBRB 系统激活的规则集合之间的不一致性要更小,减少了相互矛盾的规则对推理结果的影响,从而提高了 EBRB 系统的推理准确性.

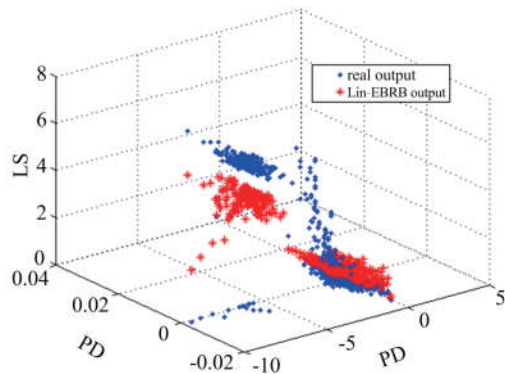


图 1 Liu_EBRB 输出和真实输出
Fig.1 Liu_EBRB output and real output

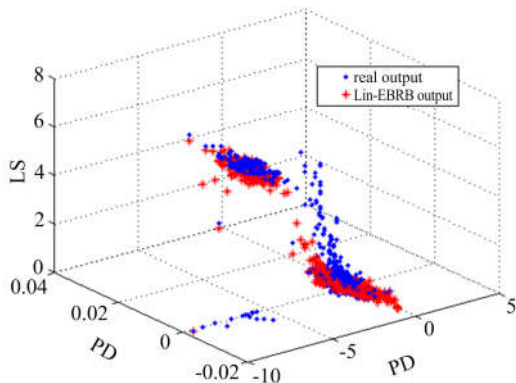


图 2 BK_EBRB 输出和真实输出
Fig.2 BK_EBRB output and real output

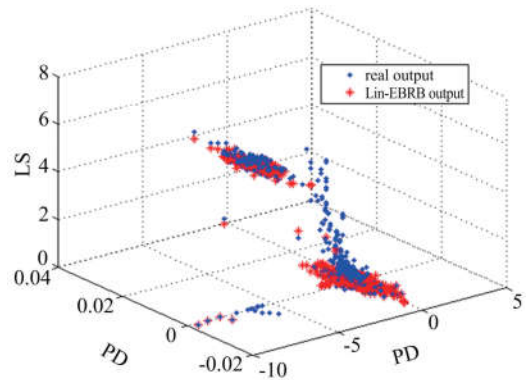


图 3 SRA_EBRB 输出和真实输出
Fig.3 SRA_EBRB output and real output

表 1 列出了 Liu_EBRB、BK_EBRB 及 SRA_EBRB 系统产生的推理输出和测试数据的真实输出的 MAE 值.表 1 中,4 种不同 EBRB 系统下的 MAE 分别为 0.626 240、0.558 087、0.231 400 和 0.205 822,其中 SRA_EBRB 系统比 Liu_EBRB 系统、BK_EBRB 系统 ($\theta = 0.7$ 和 $\theta = 0.4$) 的 MAE 值分别缩小了 67.13%、63.12%、11.05%,这表明本文提出的规则激活方法是可行有效的.

表 1 输油管道泄漏问题实验结果

Tab.1 The results on pipeline leak detection

Names	Liu_EBRB	BK_EBRB ($\theta=0.7$)	BK_EBRB ($\theta=0.4$)	SRA_EBRB
MAE	0.626 240	0.558 087	0.231 400	0.205 822

3.2 公共分类数据集测试

本文从 UCI^[21] 上选择了 10 个公共分类数据集进行测试,数据集的详细信息见表 2.为了克服因训练数据集而引起的过拟合问题,本文运用十折交叉验证法,并从数据集中每个前提属性对应数据所有取值中均匀选取 5 个作为其属性参考取值,由训练数据生成 EBRB 规则,构建初始的 SRA_EBRB 系统.为了说明本文方法的有效性,表 3 列出了与文献[18]中实验结果的对比,衡量指标为交叉验证的平均分类准确率和因个体匹配度为零而无法进行推理的平均样本数.

从表 3 可以发现,在 10 个不同的测试数据集上,运用本文方法能有效提高分类准确性,其中在 Contraceptive, Mammographic 和 Glass 三个测试数据集上提升效果较为明显,这表明本文所提方法具有更好的推理准确性,这是因为本文提出的方法使得具有更好推理准确性的规则集合,在 EBRB 推理过程中可以发挥更多的作用.

表 2 分类数据集信息

Tab.2 The basic information of classification datasets

Names	No. of classes	No. of attributes	No. of samples
Pima	2	8	768
Mammographic	2	5	830
Bupa liver	2	6	345
Wine	3	13	178
Iris	3	4	150
Seeds	3	7	210
Contraceptive	3	9	1473
Glass	7	9	214
Ecoli	8	7	336
Yeast	10	8	1484

表 3 分类数据集实验结果

Tab.3 The results on classification datasets

Names	WA	DRA+WA	ER	DRA+ER	SRA_EBRB
Pima	73.59	71.34	73.39	71.44	71.71
	0	0	0	0	0
Mammographic	77.61	78.67	78.39	77.64	82.53
	0.05	0	0.1	0	0
Bupa liver	67.81	65.72	67.65	64.90	70.46
	0	0	0	0	0
Wine	96.29	96.40	96.32	96.46	96.85
	0	0	0	0	0
Iris	95.10	95.63	95.20	95.50	94.80
	0	0	0	0	0
Seeds	84.83	92.14	87.04	92.02	91.24
	0	0	0	0	0
Contraceptive	30.47	35.69	32.17	36.41	48.46
	231.05	0	229.7	0	0
Glass	47.85	70.26	51.43	69.65	73.08
	3.55	0	3.35	0	0
Ecoli	19.53	83.75	33.72	83.76	84.85
	1	0	1	0	0
Yeast	31.20	54.15	30.91	54.13	56.85
	0	0	0	0	0

为了进一步验证本文方法的推理性能,将本文方法同文献[20]列出的其他分类算法相比较,衡量

指标为:分类准确率的平均值±标准差.由表 4 可知,同其他分类算法相比,本文提出的方法的分类准确率均较理想,其中 Glass 数据集的分类准确率要高于其他分类算法,而 Wine 数据集的分类准确率则仅次于 FGM 和 SMO 算法.

表 4 和其他分类算法比较结果

Tab.4 Comparison with other methods

Names	Wine	Iris	Glass
SRA_EBRB	96.85±0.004	94.80±0.60	73.08±0.008
KSBKT+SBR	96.63±0.71	95.67±0.68	70.09±1.00
KSKDT+SBR	96.52±0.79	95.73±0.68	69.72±1.40
EBRB	96.46±0.56	95.13±0.31	67.90±1.18
Naive Bays	96.75±2.32	96.00±0.30	42.90±1.70
C4.5	91.14±5.12	95.13±0.20	67.90±0.50
SMO	97.87±2.11	96.69±2.58	58.85±6.58
FGM	98.36±1.26	96.88±2.40	69.14±4.69

4 结论

为了解决现有的 EBRB 系统计算个体匹配度时易出现负值、零激活规则问题,本文对相似性度量方法进行改进并提出了基于改进相似性度量的规则激活方法,相比于现有的规则激活方法,本文所提方法能有效克服个体匹配度负值、零激活规则问题.此外,通过设定规则激活权重阈值的方式筛选出不一致性最小的规则集合用于规则合成,减少了相互矛盾的规则对于推理的影响,提高了 EBRB 的推理能力.为了验证本文方法的可行性和有效性,以输油管道泄漏问题和多个分类数据集作为实验实例,通过和现有的 EBRB 系统以及其他分类算法进行对比,验证了本文方法的合理有效性.在接下来的研究工作中,将进一步研究合理设置激活权重阈值和减少冗余规则问题,以期提出更合理有效的规则激活方法.

参考文献(References)

- [1] 周志杰, 杨剑波, 胡昌华, 等. 置信规则库专家系统与复杂系统建模[M]. 北京: 科学出版社, 2011.
- [2] YANG J B, LIU J, WANG J, et al. Belief rule-base inference methodology using the evidential reasoning approach-RIMER[J]. IEEE Transactions on Systems, Man and Cybernetics Part A-Systems and Humans, 2006, 36(2): 266-285.

- [3] DEMPSTER A P. A generalization of Bayesian inference [J]. *Journal of the Royal Statistical Society B*, 1968, 30(2): 205-247.
- [4] SHAFER G A. *Mathematical Theory of Evidence* [M]. Princeton: Princeton university press, 1979.
- [5] HWANG C L, YOON K S. Multiple attribute decision making: Methods and applications [J]. *Wuropean Journal of Operational Research*, 1981, 22(1): 22-34.
- [6] ZADEH L A. Fuzzy sets [J]. *Information and control*, 1965, 8(3): 338-353.
- [7] ZHOU Z J, HU C H, YANG J B, et al. Online updating belief rule based system for pipeline leak detection under expert intervention [J]. *Expert Systems with Applications*, 2008, 36(4): 7700-7709.
- [8] XU D L, LIU J, YANG J B, et al. Inference and learning methodology of belief-rule-based expert system for pipeline leak detection [J]. *Expert Systems with Applications*, 2007, 32(1): 103-113.
- [9] YANG J B, LIU J, XU D L, et al. Optimization models for training belief-rule-based systems [J]. *IEEE Transactions on Systems, Man and Cybernetics*, 2007, 37(4): 569-585.
- [10] YANG Y, FU C, CHEN Y W, et al. A belief rule based expert system for predicting consumer preference in new product development [J]. *Knowledge Based Systems*, 2016, 94(C): 338-353.
- [11] JIANG J, LI X, ZHOU Z J, et al. Weapon system capability assessment under uncertainty based on the evidential reasoning approach [J]. *Expert Systems with Applications*, 2011, 38(11): 13773-13784.
- [12] 杨隆浩, 蔡芷铃, 黄志鑫, 等. 出租车乘车概率预测的置信规则库推理方法[J]. *计算机科学与探索*, 2015, 9(8): 985-994.
- YANG L H, CAI Z L, HUANG Z X, et al. Belief rule-base inference methodology for predicting probability of taking taxi[J]. *Journal of Frontiers of Computer Science and Technology*, 2015, 9(8): 985-994.
- [13] CHEN Y W, YANG J B, XU D L, et al. Inference analysis and adaptive training for belief rule based systems [J]. *Expert Systems with Applications*, 2011, 38(10): 12845-12860.
- [14] 常瑞, 王红卫, 杨剑波. 基于梯度法与二分法的置信规则库参数训练方法 [J]. *系统工程*, 2007: 287-291.
- CHANG R, WANG H W, YANG J B. An algorithm for training parameters in belief rule-bases based on the gradient and dichotomy methods [J]. *Systems Engineering*, 2007: 287-291.
- [15] 苏群, 杨隆浩, 傅仰耿. 基于变速粒子群优化的置信规则库参数训练方法[J]. *计算机应用*, 2014, 34(8): 2161-2165.
- SU Q, YANG L H, FU Y G. Parameter training approach based on variable particle swarm optimization for belief rule base [J]. *Journal of Computer Application*, 2014, 34(8): 2161-2165.
- [16] 王韩杰, 杨隆浩, 傅仰耿, 等. 专家干预下置信规则库参数训练的差分进化算法[J]. *计算机科学*, 2015, 42(5): 88-93.
- WANG H J, YANG L H, FU Y G, et al. Differential evolutionary algorithm for parameter training of belief rule base under expert intervention [J]. *Computer Science*, 2015, 42(5): 88-93.
- [17] LIU J, MARTINEZ L, CALZADA A, et al. A novel belief rule base representation, generation and its inference methodology [J]. *Knowledge-Based Systems*, 2013, 53: 129-141.
- [18] CALZADA A, LIU J, WANG H, et al. A new dynamic rule activation method for extended belief rule-based systems[J]. *IEEE Transactions on Knowledge and data Engineering*, 2015, 7(4): 880-888.
- [19] 苏群, 杨隆浩, 傅仰耿, 等. 基于 BK 树的扩展置信规则库结构优化框架[J]. *计算机科学与探索*, 2016, 10(2): 257-267.
- SU Q, YANG L H, FU Y G, et al. Structure optimization framework of extended belief rule base based on BK-tree[J]. *Journal of Frontiers of Computer Science and Technology*, 2016, 10(2): 257-267.
- [20] YANG L H, WANG Y M, SU Q, et al. Multi-attribute search framework for optimizing extended belief rule-based systems [J]. *Information Science*, 2016: 370-371.
- [21] FRANK A, ASUNCION A. UCI machine learning repository[EB/OL]. [2017-04-16] <http://archive.ics.uci.edu/>.