

控制与决策

Control and Decision

基于聚类集成和激活因子的扩展置信规则库推理模型

杨隆浩, 任天宇, 胡海波, 叶菲菲, 王应明

引用本文:

杨隆浩,任天宇,胡海波,叶菲菲,王应明. 基于聚类集成和激活因子的扩展置信规则库推理模型[J]. *控制与决策*, 2023, 38(3): 815–824.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1091>

您可能感兴趣的其他文章

Articles you may be interested in

[一种邻域粒 \$K\$ 均值聚类方法](#)

A neighborhood granular K -means clustering method

控制与决策. 2023, 38(3): 857–864 <https://doi.org/10.13195/j.kzyjc.2021.1553>

[基于密度峰值聚类理念的大群体应急模糊决策模型](#)

A large group emergency fuzzy decision-making method based on theory of clustering by fast search and find of density peaks

控制与决策. 2022, 37(12): 3307–3313 <https://doi.org/10.13195/j.kzyjc.2021.0503>

[基于多约束条件的机器人抓取策略学习方法](#)

A learning method of robotic grasping strategy based on multi-constraint conditions

控制与决策. 2022, 37(6): 1445–1452 <https://doi.org/10.13195/j.kzyjc.2020.1716>

[CART决策树方法在煤电厂节能降耗中的应用](#)

Application of CART decision tree model in reducing coal consumption in coal power plant

控制与决策. 2021, 36(5): 1232–1238 <https://doi.org/10.13195/j.kzyjc.2019.1272>

[基于SRCSAC评价框架挖掘的跨语言查询译后扩展](#)

Cross language query post-translation expansion based on the SRCSAC evaluation framework mining

控制与决策. 2020, 35(11): 2787–2796 <https://doi.org/10.13195/j.kzyjc.2018.1647>

基于聚类集成和激活因子的扩展置信规则库推理模型

杨隆浩¹, 任天宇¹, 胡海波³, 叶菲菲^{2†}, 王应明¹

(1. 福州大学 经济与管理学院, 福州 350116; 2. 福建师范大学 文化旅游与公共管理学院, 福州 350117;
3. 香港理工大学 电子及资讯工程学系, 香港)

摘要: 规则约减和规则激活是扩展置信规则库 (EBRB) 推理模型优化研究中的两个重要方向. 然而, 现有研究成果大多存在方法参数确定主观性强和计算复杂度高等不足. 为此, 通过引入聚类集成和激活因子提出改进的 EBRB 推理模型, 称为 CEAF-EBRB 模型. 该模型先基于聚类集成对历史数据进行多次的数据聚类分析, 再以簇为单位将所有历史数据生成扩展置信规则; 同时, 通过激活因子修正个体匹配度计算公式以及离线的方式计算激活因子取值, 以确保高效地激活一致性的规则. 最后, 在非线性函数拟合、模式识别、医疗诊断等常见问题中验证了所提 CEAF-EBRB 模型的可行性和有效性, 从而为决策者提供更准确的决策支持.

关键词: 扩展置信规则库; 规则约减; 规则激活; 聚类集成; 激活因子

中图分类号: TP18 文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1091

引用格式: 杨隆浩, 任天宇, 胡海波, 等. 基于聚类集成和激活因子的扩展置信规则库推理模型 [J]. 控制与决策, 2023, 38(3): 815-824.

Extended belief rule base inference model based on clustering ensemble and activation factor

YANG Long-hao¹, REN Tian-yu¹, HU Hai-bo³, YE Fei-fei^{2†}, WANG Ying-ming¹

(1. School of Economics and Management, Fuzhou University, Fuzhou 350116, China; 2. School of Cultural Tourism and Public Administration, Fujian Normal University, Fuzhou 350117, China; 3. Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China)

Abstract: Rule reduction and rule activation are two important directions in the studies of improving extended belief rule base (EBRB) inference models. However, most of these studies are still suffering challenges, such as strong subjectivity of parameters determination and/or a high computational complexity. For this reason, this paper proposes an improved EBRB inference model, which is called CEAF-EBRB model, based on the clustering ensemble and activation factor. The CEAF-EBRB model performs multiple data clustering analyses on historical data based on the clustering ensemble firstly, and then generates extended belief rules from all historical data in the unit of clusters. Meanwhile, the activation factor is used to modify the calculation of individual matching degrees and then effectively activate consistent rules after using an offline way to initialize the activation factor. Finally, the feasibility and effectiveness of the CEAF-EBRB model are verified through solving non-linear function fitting, pattern recognition, and medical diagnosis. The proposed model can provide a more accurate decision support for decision-makers.

Keywords: extended belief rule base; rule reduction; rule activation; clustering ensemble; activation factor

0 引言

IF-THEN 规则是众多知识表达方式中最为普遍的方式之一. 众多研究表明, 其他类型的知识表达方式均可以等同或转换为 IF-THEN 规则^[1]. 因此, 基于 IF-THEN 规则的规则库系统得到国内外学者的广泛关注^[2-5], 并先后提出了模糊规则库系统和置信规则

库系统等. 在规则库系统的最新研究中, Liu 等^[6]在模糊规则库系统和置信规则库系统的基础上又进一步提出了扩展置信规则库 (EBRB) 推理模型, 该模型不仅继承了模糊规则库系统中 Wang-Mendel 建模方式的高效性, 还沿用了置信规则库系统中通过置信度分布表示多种不确定信息的置信规则表示框架. 目

收稿日期: 2021-06-23; 录用日期: 2021-12-30.

基金项目: 国家自然科学基金项目 (72001043, 61773123); 教育部人文社科项目 (20YJC630188); 福建省自然科学基金项目 (2020J05122); 福建省社会科学规划项目 (FJ2019C032).

责任编辑: 李勇建.

†通讯作者. E-mail: 13075810934@126.com.

前,EBRB推理模型已被成功用于桥梁风险评估^[5]、输油管线检漏^[6]、基础数据分类^[7]等实际问题。

为了提高EBRB推理模型的建模精度,国内外学者围绕约减冗余规则和选取一致性激活规则展开了大量研究工作,相关的研究成果包括:

1) 针对约减冗余规则的相关研究成果:为了鉴别EBRB中规则之间的相对有效性, Yang等^[8]先通过定义每条规则的贡献度构建EBRB的决策单元矩阵,再引入数据包络分析模型计算每条规则的相对效率值,从而通过剔除效率较低的规则达到精简无效规则的目的。随后,该规则约减方法与参数优化模型相结合被应用到桥梁风险评估问题中^[5]。Zhang等^[9]通过具有噪音的密度聚类(DBSCAN)算法对EBRB中的规则进行聚类分析,再将同一个簇中所有的规则统一生成一条新规则,实验结果表明,基于DBSCAN算法进行规则约减,不仅能够降低规则数量,还能够提高决策准确性; Yang等^[10]借鉴了模糊规则库系统中分割域的概念,将EBRB推理模型的输入域分割成若干个分割域,再将位于同一个分割域的规则合并成一条新的规则,目前该规则约减方法的有效性已在大数据分类问题中得到验证。

2) 针对选取激活规则的相关研究成果:为了避免EBRB推理模型激活不一致的规则和规则零激活的情形, Calzada等^[11]提出了动态激活规则方法,其核心思想是对个体匹配度的计算公式增设幂乘运算,再通过反复计算个体匹配度达到筛选激活规则的目的。相似的研究还包括:引入树形数据结构^[12]或哈希算法^[13]优化规则间索引关系,再在给定阈值的前提下基于规则索引框架筛选激活规则;基于数据包络分析模型提出一致性分析^[14]的规则激活方式等。最近,根据置信度分布之间的距离满足区间 $[0, \sqrt{2}]$, Yang等^[10]提出了以阈值1作为激活规则一致性的区分边界,即当输入数据与规则在属性上的距离都小于1时,当前规则才会被激活。

上述的研究成果在一定程度上完善了EBRB推理模型的构建和推理过程,但还需要注意以下几点: 1) 现有的EBRB规则约减方法仍存在计算复杂度高和对初始参数取值敏感等不足,如何更高效地完成规则约减且最终结果不受初始参数取值不同的影响还有待进一步探讨; 2) 现有的EBRB规则激活方法为保证激活规则的一致性,倾向于激活尽可能少的规则,弱化了EBRB推理模型的泛化能力。同时,在规则推理过程中通过迭代的方式筛选激活规则势必会降低推理效率。因此,优化EBRB推理模型的构建和推理过程仍然是亟待解决的重要问题和挑战。

针对上述问题和挑战,本文基于聚类集成和激活因子提出改进的EBRB推理模型,即CEAF-EBRB模型。首先,通过引入聚类集成的思想,利用DBSCAN算法在不同初始参数设定下对历史数据进行数据聚类分析;再根据 K 中心点聚类算法将多次聚类分析结果集成为唯一的聚类分析结果,并以簇为基本单位将所有的历史数据转换为EBRB中的规则;然后,通过提出激活因子的概念,对专家干预下的个体匹配度计算公式进行修正,围绕修正的个体匹配度计算公式对EBRB中规则进行交叉验证和参数学习,以确定激活因子的具体取值,其中由于该取值确定过程是与规则推理过程相分离的,有效地避免了现有规则激活方法中需要在规则推理过程中迭代式地筛选激活规则;最后,本文通过在非线性函数拟合问题和多个常见的源于模式识别与医疗诊断的数据分类问题中进行方法比较和灵敏度分析,验证所提CEAF-EBRB模型的可行性和有效性,并进一步拓宽EBRB推理模型的应用范畴。

1 EBRB推理模型和问题提出

1.1 EBRB的表示和构建方法

EBRB是由扩展置信规则组成的规则库,通常包含有 M 个前提属性 $U_i (i = 1, 2, \dots, M)$ 及相应的 J_i 个参考等级 $A_{i,j} (j = 1, 2, \dots, J_i)$ 和结果属性 D 及相应的 N 个参考等级 $D_n (n = 1, 2, \dots, N)$ 。由此,EBRB中第 $k (k = 1, 2, \dots, L)$ 条扩展置信规则可表示如下:

$$R_k: \text{If } U_1 \text{ is } \{(A_{1,j}, \alpha_{1,j}^k); j = 1, 2, \dots, J_1\} \wedge \dots \wedge \\ U_M \text{ is } \{(A_{M,j}, \alpha_{M,j}^k); j = 1, 2, \dots, J_M\}, \\ \text{Then } D \text{ is } \{(D_n, \beta_n^k); n = 1, 2, \dots, N\}, \\ \text{with } \theta_k \text{ and } \{\delta_1, \delta_2, \dots, \delta_M\}. \quad (1)$$

其中: $\alpha_{i,j}^k$ 和 β_n^k 分别表示第 k 条规则在参考等级 $A_{i,j}$ 与 D_n 上的置信度, θ_k 表示第 k 条规则的规则权重, δ_i 表示第 i 个前提属性的属性权重。EBRB的构建流程可总结为如下步骤。

step 1: 初始化关键参数。根据先验知识确定EBRB中关键参数的取值,这些关键参数包括: M 个前提属性的权重 δ_i 、 M 个前提属性对应的效用值 $u(A_{i,j})$ 和结果属性对应的效用值 $u(D_n)$ 。

step 2: 生成置信度分布。假设有 L 组历史数据 $\langle \mathbf{x}_k, y_k \rangle$ 。其中: $\mathbf{x}_k = (x_{k,1}, x_{k,2}, \dots, x_{k,M})$ 表示第 k 个输入向量, $x_{k,i}$ 表示第 i 个前提属性上第 k 个输入数据; y_k 表示结果属性上第 k 个输出数据。通过基于效用的信息转换方法^[15]将 L 组历史数据转换为置信

度分布,其中第 i 个前提属性的置信度分布如下:

$$S(x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k); j = 1, 2, \dots, J_i\}. \quad (2)$$

其中

$$\alpha_{i,j}^k = \frac{u(A_{i,j+1}) - u(x_{k,i})}{u(A_{i,j+1}) - u(A_{i,j})}, \alpha_{i,j+1}^k = 1 - \alpha_{i,j}^k, \\ u(A_{i,j}) \leq x_{k,i} \leq u(A_{i,j+1}); \quad (3)$$

$$\alpha_{i,t}^k = 0, t = 1, 2, \dots, J_i, t \neq j, j + 1. \quad (4)$$

同理,结果属性的置信度分布为

$$S(y_k) = \{(D_n, \beta_n^k); n = 1, 2, \dots, N\}. \quad (5)$$

step 3: 计算规则权重. 在生成 L 条扩展置信规则的置信度分布后,根据Yang等^[10]的研究,将每一条规则的规则权重设置为1.

1.2 基于EBRB的推理方法

在构建完EBRB后,EBRB推理模型便能依据新的输入数据进行规则推理,以预测其相应的输出,其中基于EBRB的推理流程总结为如下3个步骤.

step 1: 计算个体匹配度. 假设输入数据的向量为 $\mathbf{x} = (x_1, x_2, \dots, x_M)$. 首先,利用式(3)和(4)将输入数据向量中的每一个输入值转化为置信度分布

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}); j = 1, 2, \dots, J_i\}; \quad (6)$$

接着,可计算第 k 条扩展置信规则在第 i 个前提属性上的个体匹配度

$$S^k(x_i, U_i) = 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}. \quad (7)$$

step 2: 计算激活权重. 通过step 1所得到的个体匹配度 $S^k(x_i, U_i)$,以及属性权重 δ_i 和规则权重 θ_k ,可以算得第 k 条规则的激活权重

$$w_k = \frac{\theta_k \prod_{i=1}^M S^k(x_i, U_i)^{\bar{\delta}_i}}{\sum_{l=1}^L \theta_l \prod_{i=1}^M S^l(x_i, U_i)^{\bar{\delta}_i}}, \bar{\delta}_i = \frac{\delta_i}{\max_{j=1,2,\dots,M} \{\delta_j\}}. \quad (8)$$

step 3: 合成激活规则. 利用证据推理(ER)方法的解析公式,将所有被激活规则在结果属性上的置信度进行合成,即

$$\beta_n = \frac{\prod_{k=1}^L (w_k \beta_n^k + 1 - w_k) - \prod_{k=1}^L (1 - w_k)}{\sum_{i=1}^N \prod_{k=1}^L (w_k \beta_i^k + 1 - w_k) - N \prod_{k=1}^L (1 - w_k)}. \quad (9)$$

根据合成所得的置信度 β_n ,可以将关于输出数据 x 的

推理结果表示成

$$f(\mathbf{x}) = \{(D_n, \beta_n); n = 1, 2, \dots, N\}. \quad (10)$$

1.3 现有研究的不足之处

EBRB的构建需要根据训练数据一对一生成扩展置信规则,因此当可用数据不断增加时,规则也会无限增加,最终不仅增多EBRB中的冗余规则或由噪音数据转换生成的噪音规则,同时也会降低EBRB推理模型的推理效率和准确性. 针对该问题,现有的研究中已提出多种应对方法,包括基于数据包络分析的规则约减方法和基于聚类分析的规则约减方法. 虽然这些方法在一定程度上能够精简冗余规则和噪音规则,但仍存在如下不足之处:

1) 基于数据包络分析的规则约减方法具有较高的计算复杂度. 考虑到数据包络分析的本质是一个线性规划模型,因此在分析EBRB中每条规则的有效性时,均需要求解一个线性规划问题,且该线性规划问题的复杂程度与规则数量成正比,这导致当EBRB中的规则数量很大时,该规则约减方法需要耗费大量的时间进行规则有效性的筛选.

2) 基于聚类分析的规则约减方法易受人为主观因素的影响. 该类规则约减方法中涉及的聚类算法思想主要包括网格聚类分析和密度聚类分析:前者在进行规则聚类时需要专家事先设定网格的划分;后者虽然能够依据数据集的离散程度自动识别数据集的分布,但所使用的算法为DBSCAN算法,其核心参数同样需要专家事先给定,且不同的参数设置对聚类结果具有较大的影响.

相比于优化EBRB的构建过程,优化推理过程同样是该领域的研究热点之一,该研究核心主要是如何通过激活规则筛选方式,避免EBRB推理模型激活不一致的规则和规则零激活的情形. 针对这一问题,现有研究中通过修正个体匹配度的计算公式提出了一定数量的规则激活方法,其中修正方式主要有专家干预和幂乘运算两种方式. 而这些规则激活方法存在如下不足之处:

1) 基于专家干预计算个体匹配度易受人为主观因素的影响. 该类计算方式中常用的专家干预策略是引入约束条件,如式(10)通过设定阈值1修正个体匹配度的计算,进而保证部分规则因激活权重等于0而无法被激活. 虽然引入阈值1能够对激活规则进行一定程度的筛选,但如何针对不同的输入数据给定有效的阈值还有待进一步分析.

$$S^k(x_i, U_i) = \begin{cases} 1 - d_i^k, & d_i^k \leq 1; \\ 0, & d_i^k > 1. \end{cases}$$

$$d_i^k = \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}. \quad (11)$$

2) 基于幂乘运算计算个体匹配度具有较高的计算复杂度. 该类计算方式主要应用于动态规则激活方法中, 核心思想是通过设定不同的幂乘数反复计算个体匹配度, 以筛选出具备一致性的激活规则. 考虑到每次的计算过程均会涉及到每条扩展置信规则的全部前提属性, 因此当EBRB中含有大量的扩展置信规则或前提属性时, 基于幂乘运算计算个体匹配度的计算过程势必十分低效.

综上, 针对现有研究的不足, 本文拟提出改进的EBRB推理模型, 旨在更高效且更少受主观因素影响的情形下约减无效规则和激活一致性规则.

2 改进的EBRB推理模型

本节基于聚类集成提出新的EBRB构建方法并基于激活因子提出新的EBRB推理方法, 进而组成改进的EBRB推理模型.

2.1 基于聚类集成的EBRB构建方法

聚类集成由Strehl等^[16]于2002年提出, 核心思想为: 通过对数据集进行多次的聚类分析, 以确定数据集的不同聚类结果, 再将这些聚类结果通过特定方法组合成一个最终聚类结果. 相比于传统的聚类分析, 聚类集成在算法性能、鲁棒性、稳定性等方面均表现出了更优的性能. 因此, 本节通过引入聚类集成改进EBRB推理模型的扩展置信规则生成方法, 降低EBRB中冗余规则和不可靠规则对EBRB推理模型决策准确性的影响.

聚类集成可以分成基础聚类簇生成和基础聚类簇集成两个阶段^[17], 为在EBRB构建过程中实现这两阶段, 分别以DBSCAN密聚类算法和K中心点聚类算法作为基础算法, 提出改进的EBRB构建方法, 相应的流程如图1所示.

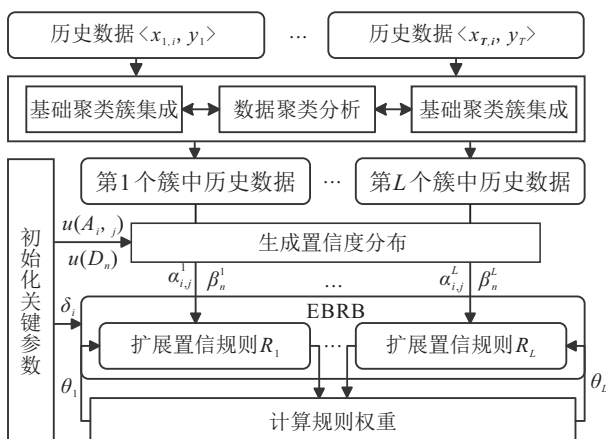


图1 改进的EBRB构建流程

根据图1, 改进的EBRB构建流程有如下步骤:

step 1: EBRB构建的数据聚类分析. 假设有T组训练数据 $\langle \mathbf{x}_k, y_k \rangle (k = 1, 2, \dots, T)$, 其中 $\mathbf{x}_k = (x_{k,1}, x_{k,2}, \dots, x_{k,M})$. 通过事先给定的数据间最小欧氏距离eps对T组训练数据进行如下的聚类分析:

step 1.1: 任意选择一组训练数据 $\langle \mathbf{x}_k, y_k \rangle$ (该训练数据尚未被分派到任一簇中), 并以其作为中心数据计算它与其他训练数据 $\langle \mathbf{x}_t, y_t \rangle$ (这些数据同样未被分配到任一簇中) 的欧氏距离

$$d(\mathbf{x}_k, \mathbf{x}_t) = \sqrt{\sum_{i=1}^M (x_{k,i} - x_{t,i})^2}. \quad (12)$$

step 1.2: 将满足条件 $d(\mathbf{x}_k, \mathbf{x}_t) \leq \text{eps}$ 的训练数据 $\langle \mathbf{x}_t, y_t \rangle$ 组成一个簇, 并将训练数据 $\langle \mathbf{x}_t, y_t \rangle$ 作为新的中心数据, 执行 step 1.1.

当不存在训练数据满足 $d(\mathbf{x}_k, \mathbf{x}_t) \leq \text{eps}$ 时, 将 step 1.1 和 step 1.2 遍历过的训练数据分派到同一个簇中, 并重新执行 step 1.1, 直到所有的训练数据均被分派到任一簇中.

step 2: EBRB构建的基础聚类簇生成. 通过先验知识给定数据间最小欧氏距离的取值区间 $[\text{lb}_{\text{eps}}, \text{ub}_{\text{eps}}]$ 和最小欧氏距离的更新步长stp, 并将eps初始化为 lb_{eps} . 通过反复更新最小欧氏距离, 即 $\text{eps} = \text{eps} + \text{stp}$, 对训练数据进行如 step 1 所示的聚类分析. 假设共计进行S次聚类分析, 其中第s ($s = 1, 2, \dots, S$) 次聚类分析中簇的数量为 L_s , 各个训练数据之间的关联关系表示为

$$C^s = \begin{bmatrix} c_{1,1}^s & c_{1,2}^s & \dots & c_{1,T}^s \\ c_{2,1}^s & c_{2,2}^s & \dots & c_{2,T}^s \\ \vdots & \vdots & \ddots & \vdots \\ c_{T,1}^s & c_{T,2}^s & \dots & c_{T,T}^s \end{bmatrix}. \quad (13)$$

其中: $c_{i,j}^s \in \{0, 1\}$ 表示第s次聚类分析中第i组训练数据与第j组训练数据之间的关联关系, 当 $c_{i,j}^s = 1$ 时, 第i组训练数据与第j组训练数据处于同一个簇中; 否则这两组训练数据在不同的簇中.

step 3: EBRB构建的基础聚类簇集成. 首先, 根据S个聚类分析矩阵 C^s 计算第i组训练数据与第j组训练数据处于同一个簇的概率

$$p_{i,j} = \frac{\sum_{s=1}^S c_{i,j}^s}{S}. \quad (14)$$

然后, 根据S次聚类分析的簇数量 L_s 计算EBRB中扩展置信规则的数量

$$L = \left[\frac{\sum_{s=1}^S L_s}{S} \right]. \quad (15)$$

接着,以 $p_{i,j}$ 作为任意两组数据之间的相似性和 L 作为聚类分析中簇的数量,对 T 组训练数据进行如下的聚类分析:

step 3.1: 任意选取 L 组训练数据作为中心数据,再根据剩余训练数据与这 L 组数据的相似性最大值,将这些训练数据以中心数据为中心分派成 L 个簇;

step 3.2: 对于 L 个簇中的所有训练数据,以当前数据与其他数据的相似性之和最大为标准,选取出每个簇的新中心数据,再执行 step 3.1;

step 3.3: 重复执行 step 3.1 和 step 3.2,直到各个簇的中心数据不再改变.

step 4: EBRB 构建的规则生成. 假设经 step 3 的聚类分析后,第 k 个簇中有 T_l 个训练数据,记为 $\langle \mathbf{x}_l^k, y_l^k \rangle$,其中 $\mathbf{x}_l^k = (x_{l,1}^k, x_{l,2}^k, \dots, x_{l,M}^k)$,则由第 k 个簇中第 l 个数据 $\langle \mathbf{x}_l^k, y_l^k \rangle$ 经式(3)和(4)转化置信度分布 $S(x_{i,j}^k) = \{(A_{i,j}, \alpha_{i,j}^{k,l}); j = 1, 2, \dots, J_i\}$ 和置信度分布 $S(y_n^k) = \{(D_n, \beta_n^{k,l}); n = 1, 2, \dots, N\}$. 接着,当 T_l 个训练数据全部转换为置信度分布后,计算与第 k 个簇相关的扩展置信规则的置信度

$$\alpha_{i,j}^k = \frac{\sum_{l=1}^{T_l} \alpha_{i,j}^{k,l}}{T_l}, \quad i = 1, 2, \dots, M, \quad j = 1, 2, \dots, J_i; \quad (16)$$

$$\beta_n^k = \frac{\sum_{l=1}^{T_l} \beta_n^{k,l}}{T_l}, \quad n = 1, 2, \dots, N. \quad (17)$$

最后,当 L 个簇中训练数据均转换为扩展置信规则后,将这些规则的规则权重赋值为1.

对于上述基于聚类集成的 EBRB 构建方法,需要说明的是:虽然新构建方法能够降低不可靠规则和冗余规则对于 EBRB 推理模型的影响,但规则约减后可能导致无法激活规则的情形,因此本文提出基于激活因子的 EBRB 推理方法.

2.2 基于激活因子的 EBRB 推理方法

针对修正的个体匹配计算公式,其在计算输入数据与规则之间相似性时,引入了固定值1筛选一致性激活规则,虽然能够有效提升 EBRB 推理模型的准确性,但考虑到固定值1的使用具有一定的主观性,以及无法根据不同的使用场景调整一致性激活规则的筛选过程.为此,通过将固定值1替换成激活因子 λ ,给出关于新个体匹配计算公式的定义.

定义1 (基于激活因子的个体匹配度计算公式) 假设第 k 条扩展置信规则中第 i 个前提属性 U_i 的置信度分布为 $\{(A_{i,j}, \alpha_{i,j}^k); j = 1, 2, \dots, J_i\}$,以及在第 i 个前提属性输入数据 x_i 的置信度分布为 $\{(A_{i,j}, \alpha_{i,j}); j = 1, 2, \dots, J_i\}$,则在第 i 个前提属性上关于第 k 条扩展置信规则和输入数据的个体匹配度计算公式定义如下:

$$S^k(x_i, U_i) = \begin{cases} \lambda - d_i^k, & d_i^k \leq \lambda; \\ 0, & d_i^k > \lambda. \end{cases}$$

$$d_i^k = \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}, \quad \lambda \in [0, \sqrt{2}]. \quad (18)$$

为能够结合 EBRB 中扩展置信规则之间的一致性确定 λ 的取值,本节首先提出基于交叉验证的激活因子取值确定方法,再根据 λ 的取值提出基于激活因子的 EBRB 推理方法,相应的流程如图2所示.

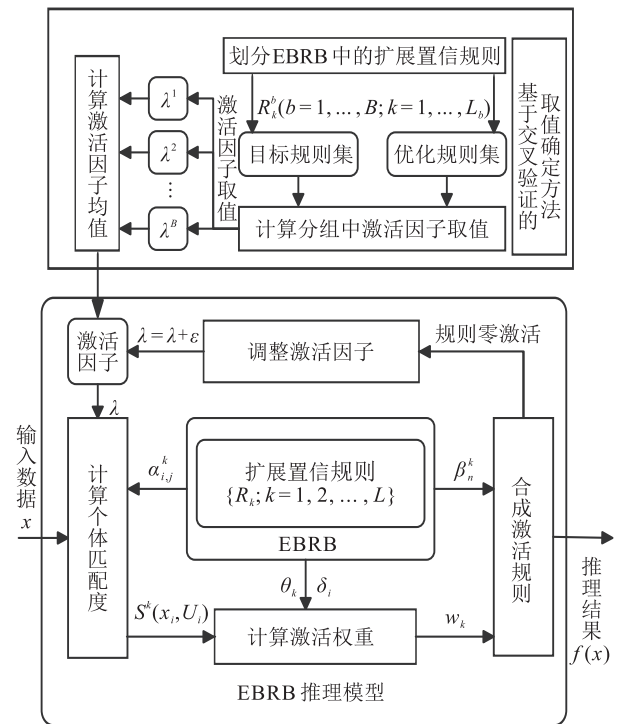


图2 改进的EBRB推理流程

根据图2,基于交叉验证的激活因子取值确定方法的具体步骤如下.

step 1: 划分 EBRB 中的扩展置信规则. 对于 EBRB 中的 L 条扩展置信规则 R_k ,通过交叉验证的思路将这 L 条扩展置信规则划分成 B 组,其中第 b 个分组中 L_b 条扩展置信规则记为 R_b^k .

step 2: 计算分组中激活因子取值. 假设以第 b 组扩展置信规则作为目标规则集,剩余的 $B - 1$ 组扩展置信规则作为优化规则集. 首先,通过2.1节中的

step 3 调整 $B - 1$ 组扩展置信规则的规则权重;然后,通过如下优化模型确定第 b 个分组中激活因子 λ^b 的取值:

$$\min \sum_k^{L_b} |f(R_k^b) - g(R_k^b)|; \text{ s.t. } 0 \leq \lambda^b \leq \sqrt{2}. \quad (19)$$

其中: $f(R_k^b)$ 表示以 R_k^b 的前提属性置信度分布作为输入值时,经基于激活因子的 EBRB 推理方法所得的推理输出; $g(R_k^b)$ 表示 R_k^b 的实际输出,对于回归问题,有

$$g(R_k^b) = \sum_{n=1}^N u(D_n) \beta_n^{b,k}, \quad (20)$$

对于分类问题,有

$$g(R_k^b) = D_t, \quad t = \arg \max_{n=1,2,\dots,N} \{\beta_n^{b,k}\}, \quad (21)$$

这里 $\beta_n^{b,k}$ 表示第 R_k^b 在 D_n 上的置信度.

优化模型 (20) 和 (21) 中: $\bar{\beta}_n^{b,k}$ 表示以 R_k^b 的前提属性置信度分布作为输入值时,经基于激活因子的 EBRB 推理方法所得的在 D_n 上的推理置信度; $\beta_n^{b,k}$ 表示 R_k^b 在 D_n 上的实际置信度.

step 3: 计算激活因子的均值. 依据 step 2 中计算激活因子取值的方式,依次将 B 个组中其中一个分组的扩展置信规则作为目标规则集,其余分组的扩展置信规则作为优化规则集,计算 B 个激活因子的取值 λ^b ,再通过如下公式计算得激活因子 λ 取值:

$$\lambda = \frac{\sum_{b=1}^B \lambda^b}{B}. \quad (22)$$

结合上述经 3 个步骤求得的激活因子取值,本节进一步提出如下基于激活因子的 EBRB 推理方法.

step 1: 根据定义 1 计算激活权重. 假设数据向量为 $x = (x_1, x_2, \dots, x_M)$, 依据 1.2 节中的 step 1 和 step 2 计算 EBRB 中 L 条扩展置信规则 R_k 的激活权重,记为 w_k ,其中以定义 1 所示的式 (18) 计算各条扩展置信规则在每个前提属性的个体匹配度.

step 2: 调整激活因子确定激活规则. 根据 step 1 计算所得的激活权重 w_k ,筛选关于输入数据 x 的激活规则,即:当 $w_k > 0$ 时,第 k 条扩展置信规则被选为激活规则.同时,为避免出现规则零激活的情形,当通过激活权重无法筛选出激活规则时,需要通过如下公式调整激活因子,并重新执行 step 1,即

$$\lambda = \lambda + \varepsilon, \quad (23)$$

其中 $\varepsilon > 0$ 表示激活因子的调整步长.

step 3: 合成激活规则和预测输出值.对于由 step 2 获得的激活规则,依据 1.2 节中的 step 3 将所有

激活规则在结果属性上的置信度分布经 ER 解析算法进行合成,以获得关于输入数据 x 的预测置信度分布 $\{(D_n, \beta_n); n = 1, 2, \dots, N\}$.

2.3 基于聚类集成和激活因子的 EBRB 推理模型

为了确保 EBRB 推理模型能够高效且稳定地进行规则约减和规则激活,根据 2.1 节提出的 EBRB 构建方法和 2.2 节提出的 EBRB 推理方法,本节进一步提出基于聚类集成和动态激活的 EBRB 推理模型,记为 CEAF-EBRB 模型. 在应用 CEAF-EBRB 模型解决实际问题时步骤如下:

step 1: 针对实际问题收集 T 组历史数据集 $\langle \mathbf{x}_k, y_k \rangle$, 其中 $\mathbf{x}_k = (x_{k,1}, x_{k,2}, \dots, x_{k,M})$, 再通过 2.1 节中基于聚类集成的构建方法构建 EBRB.

step 2: 假设 EBRB 中共有 L 条扩展置信规则,则由这 L 条规则经 2.2 节中基于交叉验证的激活因子取值确定方法,计算激活因子 λ 的取值.

step 3: 对于实际问题中任意的输入数据 $\mathbf{x} = (x_1, x_2, \dots, x_M)$, 通过 λ 的取值和 2.2 节中的 EBRB 推理方法,计算 CEAF-EBRB 模型的推理输出 $f(\mathbf{x})$.

通过对比第 2 节中介绍的传统 EBRB 推理模型可知,本文提出的 CEAF-EBRB 模型具有如下优点:

1) CEAF-EBRB 模型采用聚类集成的方式对数据进行聚类分析,能够降低初始参数的不同对聚类结果的影响,以及延续聚类算法的高效性.

2) 激活因子能够有效避免激活规则不一致性问题和规则零激活问题.同时,激活因子的取值采用离线方式确定,能够提高规则推理效率.

3 实验分析

3.1 非线性函数拟合

本节采用如下非线性函数验证本文所提方法在解决回归问题上的有效性:

$$y = x \cdot \sin x + x \cdot \cos x^2. \quad (24)$$

为了构建 CEAF-EBRB 模型,以输入变量 x 作为前提属性 U_1 ,并设置 5 个参考等级 $A_{1,j} (j = 1, 2, \dots, 5)$,在定义域 $[0, \pi]$ 内给定均匀间隔的效用值;以输出变量 y 作为结果属性 D ,并设置 5 个参考等级 $D_n (n = 1, 2, \dots, 5)$ 以及在值域范围 $[-3, 4]$ 内给定均匀间隔的效用值.此外,通过随机方式在定义域内分别选取 200 个数据作为训练数据并选取 1000 个数据作为测试数据,设定:最小欧氏距离取值区间 $[\text{lb}_{\text{eps}}, \text{ub}_{\text{eps}}] = [0.0026, 0.0036]$,最小欧氏距离 $\text{eps} = 0.0026$,最小欧氏距离更新步长 $\text{stp} = 0.0001$,激活因子确定过程中参数 $B = 10$,激活因子的更新

步长 $\varepsilon = 0.1$.

根据上述问题描述和初始参数设定,共对所给数据集进行了11次数据聚类分析;由这11次数据聚类分析的结果,通过基础聚类集成共得99个簇,即CEAF-EBRB模型中共有99条扩展置信规则;随后基于这些规则以及式(19)所示的优化模型,获得当前最优的激活因子取值 $\lambda = 0.105$. 对于1000个测试数据,可以通过2.3节中的step 3逐个推理得到相应的输出结果 $f(x)$.

为了检验经上述步骤所得模型的有效性,引入初始EBRB推理模型(Liu-EBRB)、仅基于聚类集成改进的EBRB推理模型(CE-EBRB)、仅基于激活因子改进的EBRB推理模型(AF-EBRB)作比较分析,并采用平均绝对误差、均方误差和规则数量作为评价指标. 表1中根据曲线拟合情况进一步给出在各个评价指标上的计算值.

表1 各类EBRB推理模型的指标值

模型类型	平均绝对误差	均方误差	规则数
Liu-EBRB	0.6417	0.8866	200
CE-EBRB	0.5948	0.6606	99
AF-EBRB	0.0575	0.0054	200
CEAF-EBRB	0.0520	0.0041	99

由表1可知,虽然CE-EBRB模型的曲线拟合效果并不理想,但相较于EBRB模型而言,CE-EBRB模型不仅具有更低的均方误差和平均绝对误差,同时规则数量还明显少于Liu-EBRB模型,这说明在曲线拟合问题上规则约减的必要性和基于聚类集成的EBRB构建方法的有效性. 而在考虑基于激活因子的EBRB推理方法后,AF-EBRB模型和CEAF-EBRB模型均具有更低的均方误差和平均绝对误差,这说明了筛选激活规则的必要性和所提推理方法的有

效性,其中在这4个模型的准确性和规则数量的比较中,CEAF-EBRB模型具有最低的均方误差和平均绝对误差,以及最少的规则数量,这说明通过所提的构建方法和推理方法不仅能够精简EBRB中无效规则或噪音规则的数量,而且能够最大程度地提升EBRB推理模型的准确性.

3.2 UCI数据集分类

为了进一步验证CEAF-EBRB模型的有效性,本节引入UCI数据库中常见的6个分类数据集,具体的数据统计信息如表2所示. 在构建CEAF-EBRB模型时,采用十折交叉验证划分训练数据集和测试数据集,并给定:最小欧氏距离取值的区间通过训练数据数量与EBRB中规则数量的比率为50%确定,且 $stp = 0.0005$ 、 $B = 10$ 、 $\varepsilon = 0.1$.

表2 UCI数据集的信息统计

序号	名称	样本数	属性数	类别数
1	Seeds	210	7	3
2	Iris	150	4	3
3	Wine	178	13	3
4	Diabetes	768	8	2
5	Cancer	683	9	2
6	Thyroid	215	5	3

1) 与现有EBRB推理模型的性能比较.

为了比较CEAF-EBRB模型与其他EBRB推理模型之间的性能差异,除了引入Liu-EBRB模型和AF-EBRB模型外,还引入现有研究中常用的基于动态激活规则的EBRB模型^[11](DRA-EBRB)和基于数据包络分析的EBRB模型^[8](DEA-EBRB),以及基于K均值与激活因子的EBRB模型(KAF-EBRB). 比较中涉及的评价指标包括:分类准确率与规则数量. 表3给出了各个EBRB推理模型在6个分类数据集中指标值.

表3 与其他EBRB推理模型的综合比较分析

模型类型	指标	Iris	Seeds	Wine	Diabetes	Cancer	Thyroid	平均排名
Liu-EBRB	准确率	95.33	91.33	81.16	73.44	94.59	81.47	4.42
	规则数	135	189	160.2	691.2	614.7	193.5	4.50
DRA-EBRB	准确率	95.50	92.02	83.76	71.44	94.61	97.18	3.00
	规则数	135	189	160.2	691.2	614.7	193.5	4.50
AF-EBRB	准确率	91.33	92.86	95.49	72.56	94.59	91.17	3.92
	规则数	135	189	160.2	691.2	614.7	193.5	4.50
KAF-EBRB	准确率	94.00	91.43	94.41	73.83	95.02	94.44	3.67
	规则数	67.5	94.5	80.1	345.6	307.35	96.75	1.83
DEA-EBRB	准确率	95.40	91.67	—	74.17	—	—	2.67
	规则数	104	173.88	—	615.2	—	—	3.00
CEAF-EBRB	准确率	94.67	93.33	97.22	74.22	95.32	94.46	1.67
	规则数	68.1	94	79.1	343.9	301.1	96.1	1.17

由表3可知,相比于其他EBRB推理模型,CEAF-EBRB模型在各分类数据集上都具有良好的性能,相应的比较分析如下。

首先,就分类准确率而言,与激活因子相关的AF-EBRB、KAF-EBRB和CEAF-EBRB模型均高于Liu-EBRB模型,其中CEAF-EBRB模型是这些模型中分类准确性最高的EBRB模型。同时,相比于DRA-EBRB模型和DEA-EBRB模型,CEAF-EBRB模型在Iris和Thyroid之外其他的4个分类数据集中均具有更高的分类准确性。而在分类准确性的平均排名中,CEAF-EBRB模型拥有高于其他EBRB推理模型的排名。

其次,就规则约减数量而言,CEAF-EBRB模型在各个分类数据集上均表现出优于其他EBRB推理模型的性能,例如:在Diabetes数据集上,Liu-EBRB、DRA-EBRB和AF-EBRB模型的平均规则数量均为691.2;DEA-EBRB模型的平均规则数量为615.2;KAF-EBRB模型具有更少的平均规则数量,为345.6;而CEAF-EBRB模型则具有最少的平均规则数量,为343.9。可见,本文所提构建方法能够尽可能约

减EBRB中冗余规则或噪音规则。

此外,为了验证定义1与式(11)在激活规则时的区别,以规则零激活的次数为评价指标,在分类数据集进行数据统计,相应的结果如表4所示。

表4 不同规则激活方式在各数据集上的零激活次数

规则激活方式	Iris	Seeds	Wine	Diabetes	Cancer	Thyroid
式(11)	0.1	0	2.2	1.7	21.7	0.4
定义1	0	0	0	0	0	0

由表4可见,基于式(11)的规则激活方式均遇到了规则零激活的情况。而基于定义(1)的规则激活方式在各数据集上均未出现规则零激活的情形,由此可见本文通过引入激活因子能够有效克服基于专家干预的个体匹配度计算方法的缺陷。

2) 与传统机器学习算法的性能比较。

为了进一步论证CEAF-EBRB模型的有效性,本节还与传统的机器学习方法进行比较分析,包括决策树(CART)算法、K近邻(KNN)、支持向量机(SVM)、人工神经网络(ANN)、朴素贝叶斯(NB),具体的实验结果如表5所示。

表5 与传统机器学习算法的比较分析

数据集	CART	KNN	SVM	ANN	NB	CEAF-EBRB
Iris	93.33 (3.5)	93.33 (3.5)	92.00 (5)	97.33 (1)	84.67 (6)	94.67 (2)
Seeds	87.14 (3)	85.24 (4)	83.80 (5)	70.00 (6)	88.10 (2)	93.33 (1)
Wine	84.31 (3)	63.10 (4)	55.56 (5)	42.22 (6)	96.11 (2)	97.22 (1)
Diabetes	66.56 (5)	68.13 (4)	69.20 (3)	58.40 (6)	73.20 (2)	74.22 (1)
Cancer	94.14 (6)	96.64 (3)	97.08 (2)	97.22 (1)	96.06 (4)	95.32 (5)
Thyroid	91.04 (3)	87.62 (4)	74.08 (6)	80.48 (5)	95.78 (1)	94.46 (2)
平均排名	3.92	3.75	4.33	4.17	2.83	2

由表5可知,相较于传统的机器学习算法,CEAF-EBRB模型在Seeds、Wine和Diabetes数据集上均取得了最高的分类准确性,分别为93.33%、97.22%和74.22%。虽然在Iris、Cancer和Thyroid数据集上,ANN算法和NB算法的分类准确性高于CEAF-EBRB模型,但CEAF-EBRB模型在这些数据集上仍取得了第2和第5的分类准确性。在平均排名上,CEAF-EBRB模型高于CART算法、KNN算法、SVM算法、ANN算法和NB算法。

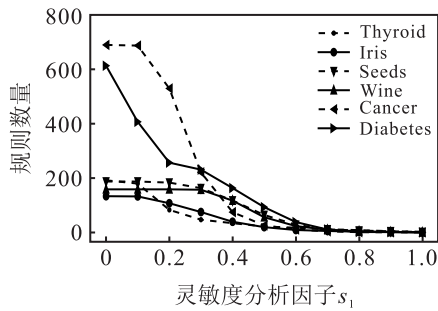
3.3 模型参数的灵敏度分析

为了分析参数初始取值对CEAF-EBRB模型的影响,本节使用表2中的分类数据集对CEAF-EBRB模型构建与推理过程中的两个关键参数数据间最小欧氏距离 ϵ 和激活因子 λ 进行灵敏度分析。

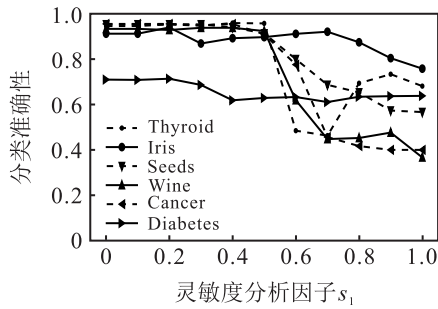
1) 关于数据间最小欧式距离 ϵ 灵敏度分析。

在数据间最小欧氏距离的灵敏度分析中,假设基于激活因子的EBRB推理方法中激活因子 $\lambda = 1$,以及基于聚类集成的EBRB构建方法中 ϵ 是通过 $\epsilon = s_1 \times \max_{k=1,2,\dots,T} \{ \min_{t=1,2,\dots,T} \{ d(\mathbf{x}_k, \mathbf{x}_t) \} \}$ 进行赋值。其中: $d(\mathbf{x}_k, \mathbf{x}_t)$ 表示数据 \mathbf{x}_k 与 \mathbf{x}_t 之间的欧氏距离, s_1 表示灵敏度分析因子。灵敏度分析时,在灵敏度分析因子的取值范围 $[0, 1]$ 内依次增大 s_1 的取值,并依此比较CEAF-EBRB模型中的规则数量和分类准确性,相应的实验结果如图3所示。

由图3可知,随着 s_1 逐渐增大,CEAF-EBRB模型中规则数量逐渐减少,其主要原因是随着 s_1 增大,数据聚类分析能够将彼此距离更大的数据划分到同一个簇中,由于每一个簇都对应一条规则,规则数量会



(a) 不同eps取值下的规则数量



(b) 不同eps取值下的推理准确率

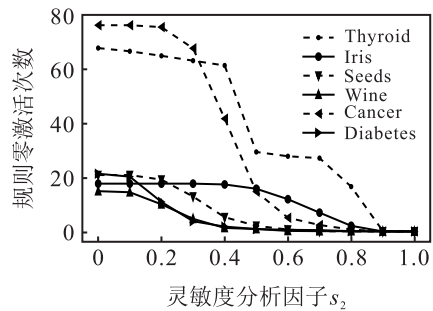
图3 关于eps的灵敏度分析

随着 s_1 增大而逐渐减少. 同时, CEAF-EBRB 模型的准确性随 s_1 增大先维持不变再逐渐降低, 这是因为 CEAF-EBRB 模型在规则推理过程中能够通过激活因子对激活规则进行有效地筛选, 以确保激活规则集合的一致性, 因此 CEAF-EBRB 模型在规则未被大量约减的情形下能够基本维持准确性不变, 当规则被过度约减后, 会因此无法激活有效的规则, 进而导致准确性也随之降低.

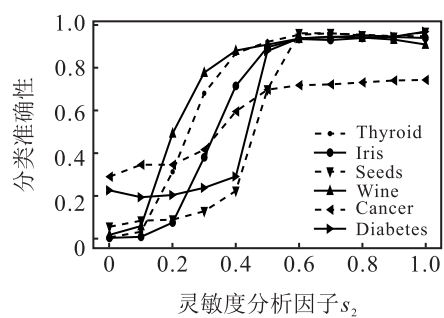
2) 关于激活因子 λ 的灵敏度分析.

在激活因子 λ 的灵敏度分析中, 假设基于聚类集成的 EBRB 构建方法中的参数取值设定与 3.2 节中参数取值设定相一致, 以及激活因子是通过 $\lambda = s_2 \times \sqrt{2}$ 进行赋值, 且在规则推理过程中不通过调整激活因子的取值处理规则零激活问题, 其中 s_2 表示灵敏度分析因子. 在灵敏度分析时, 在灵敏度分析因子的取值范围 $[0, 1]$ 内依次增大 s_2 的取值, 并依此比较 CEAF-EBRB 模型中的规则零激活次数和分类准确性, 相应的实验结果如图 4 所示.

由图 4 可知: 当灵敏度分析因子 $s_2 = 0$ 时, CEAF-EBRB 模型中 $\lambda = 0$, 其会导致任意的输入数据都无法激活规则, 从而无法产生相应的推理结果, 因此 6 个分类数据集的分类准确率均为 0; 同时, 随着 s_2 逐渐增大, CEAF-EBRB 模型中规则零激活次数逐渐减少且分类准确性逐渐增加, 其中在 Seeds、Iris 和 Thyroid 数据集上, CEAF-EBRB 模型的准确率还呈现出先逐渐增大再降低的趋势, 这是因为当激活因子



(a) 不同激活因子取值下的规则零激活次数



(b) 不同激活因子取值下的推理准确率

图4 关于激活因子的灵敏度分析

增大时, CEAF-EBRB 模型会激活更多的规则, 而更多的激活规则往往意味着激活规则集合中存在不一致性. 通过上述灵敏度分析可知: 在应对不同实际问题时, 需要选取合适的激活因子取值以优化 CEAF-EBRB 模型的性能.

4 结论

本文基于聚类集成与激活因子提出了 CEAF-EBRB 模型. 该模型首先通过聚类集成综合多次数据聚类分析的结果约减无效规则; 再通过基于交叉验证的方式确定激活因子的取值, 以及将激活因子用于规则激活过程提高推理准确率. 实验结果表明, 本文所提模型在函数拟合、模式识别、医疗诊断等问题中不仅能够减少规则数量和避免规则零激活情形, 还能确保理想的推理准确性.

在未来的研究中, 还可以再融入特征选择和参数学习, 以进一步提升 EBRB 推理模型的性能. 同时, 结合医疗诊断方面的实验分析, 未来还可以应用 CEAF-EBRB 模型对医院的临床数据进行建模.

参考文献 (References)

[1] Sun R. Robust reasoning: Integrating rule-based and similarity-based reasoning[J]. Artificial Intelligence, 1995, 75(2): 241-295.

[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] Chang L L, Fu C, Zhu W, et al. Belief rule mining using the evidential reasoning rule for medical diagnosis[J]. *International Journal of Approximate Reasoning*, 2021, 130: 273-291.
- [4] Zhou Z J, Feng Z C, Hu C H, et al. Aeronautical relay health state assessment model based on belief rule base with attribute reliability[J]. *Knowledge-Based Systems*, 2020, 197: 1-13.
- [5] 杨隆浩, 叶菲菲, 王应明. 基于扩展置信规则库联合优化的桥梁风险评估[J]. *系统工程理论与实践*, 2020, 40(7): 1870-1881.
(Yang L H, Ye F F, Wang Y M. Bridge risk assessment based on extended belief rule base with joint optimization[J]. *Systems Engineering—Theory & Practice*, 2020, 40(7): 1870-1881.)
- [6] 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.
- [7] Zhu H Z, Xiao M Q, Zhao X, et al. A structure optimization method for extended belief-rule-based classification system[J]. *Knowledge-Based Systems*, 2020, 203: 26-40.
- [8] Yang L H, Wang Y M, Lan Y X, et al. A data envelopment analysis(DEA)-based method for rule reduction in extended belief-rule-based systems[J]. *Knowledge-Based Systems*, 2017, 123: 174-187.
- [9] Zhang A, Gao F, Yang M, et al. A new rule reduction and training method for extended belief rule base based on DBSCAN algorithm[J]. *International Journal of Approximate Reasoning*, 2020, 119: 20-39.
- [10] Yang L H, Liu J, Wang Y M, et al. A micro-extended belief rule-based system for big data multiclass classification problems[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(1): 420-440.
- [11] 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, 27(4): 880-894.
- [12] Fu Y G, Zhuang J H, Chen Y P, et al. A framework for optimizing extended belief rule base systems with improved Ball trees[J]. *Knowledge-Based Systems*, 2020, 210: 1-11.
- [13] 刘莞玲, 肖承志, 傅仰耿. 哈希索引的扩展置信规则库推理方法[J]. *西安电子科技大学学报*, 2019, 46(2): 145-151.
(Liu W L, Xiao C Z, Fu Y G. Extended belief rule base inference method based on the Hash index[J]. *Journal of Xidian University*, 2019, 46(2): 145-151.)
- [14] Yang L H, Wang Y M, Fu Y G. A consistency analysis-based rule activation method for extended belief-rule-based systems[J]. *Information Sciences*, 2018, 445/446: 50-65.
- [15] Yang J B, Xu D L. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty[J]. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 2002, 32(3): 289-304.
- [16] Strehl A, Ghosh J. Cluster ensembles-A knowledge reuse framework for combining multiple partitions[J]. *Journal of Machine Learning Research*, 2002, 3(3): 583-617.
- [17] Zhou Z H. *Ensemble Methods: Foundations and Algorithms*[M]. London: Taylor & Francis, 2012: 135-155.

作者简介

杨隆浩(1990—), 男, 副教授, 博士, 从事规则库推理、证据推理等研究, E-mail: more026@hotmail.com;

任天宇(1998—), 男, 硕士生, 从事置信规则库推理、数据挖掘等研究, E-mail: lixvm9862@163.com;

胡海波(1979—), 男, 副教授, 博士, 从事数据安全、图数据库等研究, E-mail: haibo.hu@polyu.edu.hk;

叶菲菲(1991—), 女, 讲师, 博士, 从事环境治理成本预测、数据挖掘等研究, E-mail: 13075810934@126.com;

王应明(1964—), 男, 教授, 博士生导师, 从事证据推理、数据包络分析等研究, E-mail: ymwang@fzu.edu.com.

(责任编辑: 闫妍)