

控制与决策

Control and Decision

基于规则约减与激活因子的扩展置信规则库推理模型

钟智昊, 龙江, 吴孟桐, 郭阳明

引用本文:

钟智昊, 龙江, 吴孟桐, 等. 基于规则约减与激活因子的扩展置信规则库推理模型[J]. 控制与决策, 2025, 40(5): 1695-1704.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1063>

您可能感兴趣的其他文章

Articles you may be interested in

基于Hessian正则的自适应损失半监督特征选择

Adaptive loss semi-supervised feature selection based on Hessian regularization

控制与决策. 2021, 36(8): 1862-1870 <https://doi.org/10.13195/j.kzyjc.2019.1510>

超启发式交叉熵算法求解模糊分布式流水线绿色调度问题

Hyper-heuristic cross-entropy algorithm for green distributed permutation flow-shop scheduling problem with fuzzy processing time

控制与决策. 2021, 36(6): 1387-1396 <https://doi.org/10.13195/j.kzyjc.2019.1681>

基于改进蚁群算法的水面无人艇路径规划

Path planning for unmanned surface vehicle based on improved ant colony algorithm

控制与决策. 2021, 36(4): 847-856 <https://doi.org/10.13195/j.kzyjc.2019.0839>

一种自适应拟牛顿-状态转移混合智能优化算法及应用

A hybrid state transition optimization algorithm based on adaptive quasinewton method and its application

控制与决策. 2021, 36(10): 2451-2458 <https://doi.org/10.13195/j.kzyjc.2020.0214>

基于互信息操作变量曲线参数化的间歇过程批内修正优化

Intra-batch correction optimization of batch process with manipulated variable trajectory parameterization based on mutual information

控制与决策. 2021, 36(1): 234-240 <https://doi.org/10.13195/j.kzyjc.2019.0825>

基于规则约减与激活因子的扩展置信规则库推理模型

钟智昊¹, 龙江², 吴孟桐¹, 郭阳明^{3†}

(1. 西北工业大学 软件学院, 西安 710072; 2. 西北工业大学 计算机学院, 西安 710072;
3. 西北工业大学 网络空间安全学院, 西安 710072)

摘要: 针对扩展置信规则库 (EBRB) 的规则冗余与激活规则一致性过低问题, 提出一种基于 Relief 算法框架的新型结构优化框架与激活因子的推理模型, 并应用于机器学习中的分类问题与回归问题. 首先, 基于 Relief 算法思想, 通过分析历史数据与其近邻的输入输出相关性, 赋予扩展置信规则不同的权重以识别关键规则, 并通过与近邻规则的融合实现规则约减; 然后, 在计算个体匹配度过程中引入激活因子, 通过离线优化策略确定激活因子取值, 以确保激活规则的一致性和有效性; 最后, 分别采用非线性函数与公共分类数据集对所提出方法与其他类型的 EBRB 模型在处理回归问题和分类问题时的表现进行对比, 结果验证了所提出模型的有效性和优越性.

关键词: 扩展置信规则库; Relief 算法; RRelief 算法; 规则约减; 激活因子

中图分类号: TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1063

引用格式: 钟智昊, 龙江, 吴孟桐, 等. 基于规则约减与激活因子的扩展置信规则库推理模型 [J]. 控制与决策, 2025, 40(5): 1695-1704.

An extended belief rule-based inference model based on rule reduction and activation factors

ZHONG Zhi-hao¹, LONG Jiang², WU Meng-tong¹, GUO Yang-ming^{3†}

(1. School of Software, Northwestern Polytechnical University, Xi'an 710072, China; 2. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China; 3. School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In response to the issues of rule redundancy and low activation rule consistency in the extended belief rule base (EBRB), this paper proposes an inference model based on a novel structural optimization framework under Relief algorithm framework and activation factor, which can be applied to classification and regression problems in machine learning. Specifically, based on the Relief algorithm, the model first assigns different weights to the extended belief rules by analyzing the relevance of historical data and its neighboring input and output to identify key rules, and achieves rule reduction by fusing with neighboring rules. Then, this paper introduces an activation factor in the process of calculating individual matching degrees and determines the value of the activation factor through offline optimization strategies to ensure the consistency and effectiveness of the activated rules. Finally, to verify the effectiveness and superiority of the proposed model, the comparison of the performance between the proposed method and some other types of EBRB models in the terms of regression and classification problems is conducted.

Keywords: extended belief rule base; Relief; RRelief; rule reduction; activation factor

0 引言

在 2006 年, Yang 等^[1]通过研究现有 if-then 规则方法的知识库结构和不确定知识表达方案, 设计了利用证据推理规则的广义规则库推理方法. 随后, 证据推理规则方法得到了众多学者的进一步研究^[2-6]. 然而, 因为 BRB 的规则数由参考值个数决定, 即 BRB

会生成所有参考值的组合作为规则, 所以存在组合爆炸问题. 为解决该问题, Liu 等^[7]将置信度嵌入每个规则的前项和后项中, 提出了扩展置信规则库 (EBRB), 它能够更好地捕获模糊、不完整、不确定和非线性信息. 目前, EBRB 已经广泛运用于海盗袭击事件风险预测^[8]、甲状腺结节诊断^[9]、桥梁风险评

收稿日期: 2024-09-04; 录用日期: 2024-11-14.

责任编辑: 刘向杰.

†通信作者. E-mail: yangming_g@nwpu.edu.cn.

本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

估^[10]等领域.

然而, EBRB 建模与推理方法仍存在以下两方面的不足: 一是随着数据增多, EBRB 系统的规模急剧膨胀, 这种膨胀往往伴随着信息冗余, 即大量规则蕴含重复或相似的信息, 最终导致 EBRB 的推理效率降低; 二是在传统 EBRB 中, 几乎所有规则都会被激活, 并参与到推理过程中, 使得激活规则的一致性过低, 从而导致 EBRB 推理精度降低. 为此, 国内外学者主要提出两种方案: 1) 减少 EBRB 系统的规模; 2) 提高激活规则的一致性. 为减少 EBRB 系统的规模, 最常见的方法是约减 EBRB 中的冗余规则. 例如: Yang 等^[11]引入数据包络分析来减少传统 EBRB 系统的低效规则. 实验表明, 从分类数据集 Iris、Seeds、Ecoli、Diabetes 和 Glass 中分别可以减少 23%、8%、22%、11% 和 9% 的规则. Zhang 等^[12]提出了一种基于密度聚类算法的规则约减方法, 将融合阈值内的规则进行融合, 并对融合后的 EBRB 进行参数训练. 杨隆浩等^[13]提出了一种通过聚类集成和激活因子改进的 EBRB 推理模型, 该模型先基于聚类集成思想对历史数据进行多次聚类分析, 再以簇为单位将所有历史数据生成扩展置信规则. Yang 等^[14]提出了一种用于多分类问题的微扩展的信念规则系统, 通过引入一种基于领域划分的规则约简方法, 获得了相对较好的时间复杂度和计算效率.

另一方面, 从提高激活规则的一致性角度, 常见的方法是约束激活规则的范围. 例如: 余瑞银等^[15]提出了一种基于 80/20 原则的激活方法, 在推理过程中只整合 20% 的激活规则. Calzada 等^[16]提出的动态规则激活方法以动态的方式选择一组合适的激活规则, 以解决 EBRB 系统中的不完整性和不一致性问题. Lin 等^[17]引入一种结构优化框架, 采用 Vantage-Point 树 (VP-树)、Multi-Vantage-Point 树 (MVP-树) 为 EBRB 规则构建索引, 并基于 K-Means 聚类方法选择合适的查询阈值. Zhu 等^[18]提出了一种最小中心距离规则激活方法, 该方法可以消除与输入数据无关的样本, 并选择和激活与输入数据高度相关的样本. Yang 等^[19]为每条规则定义并构造激活区域, 并提出了一种新的激活规则确定和权重计算方法. Yang 等^[20]将分解策略与重叠函数引入 EBRB, 增强了 EBRB 在多属性、多分类问题上的表现. Yang 等^[21]提出了一种用于处理不平衡分类问题的集成扩展置信规则库, 通过构建多样化的基础信念规则, 并采用基于信息增益的属性权重计算, 以及一致性优先的集成推理方案, 有效缓解了类别不平衡的影响.

综上所述, 上述研究成果在一定程度上对 EBRB 的结构与推理方法进行了优化, 但仍存在不足. 例如: 余瑞银等^[15]在推理中只激活权重排名前 20% 的规则, 但这种方法在计算每条规则的激活权重时, 仍然需要访问所有规则; Zhang 等^[12]采用密度聚类算法进行规则约减, 杨隆浩等^[13]采用聚类集成思想进行规则约减, Lin 等^[17]通过基于 K-Means 聚类的方法选择合适的查询阈值, 但均没有提及聚类阈值、聚类数等参数的来源, 其融合过程也无法说明融合规则的相关性; Calzada 等^[16]需要反复计算个体匹配度以使激活规则达到目标的一致性, 且 DRA 以类别作为衡量激活规则的一致性的依据, 无法应用于回归问题; Zhu 等^[18]以输入数据到类样本中心的距离作为半径构造超球面, 同样无法应用于回归问题.

针对当前研究成果中存在的不足, 本文提出一种基于规则约减和激活因子的 EBRB 推理方法, 简称 Relief-EBRB. 具体地, 首先基于 Relief 算法框架, 提出一种规则权重算法, 分别应用于分类问题和回归问题. 该算法通过分析规则与其近邻的输入输出相关性, 能够赋予对推理结果影响更大的规则以及更高的权重; 进一步, 以规则权重为依据, 将规则与其近邻中权重相同的规则进行融合, 从而实现规则约减; 然后, 在个体匹配度计算过程中引入激活因子, 采用交叉验证与参数优化算法确定激活因子取值, 同时通过对个体匹配度的修正限制激活规则的范围, 从而提高 EBRB 激活规则的一致性; 最后, 通过非线性函数拟合和多个常见的分类数据集验证所提优化方法的可行性.

1 传统 EBRB 推理模型

1.1 传统 EBRB 的表示与构建方法

传统扩展置信规则库第 k 条规则 R_k ^[7] 表示如下:

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

其中: M 为参考值个数; $U_i (i = 1, \dots, M)$ 为第 i 个前提属性; J_i 为 U_i 的参考值个数; D 为当前 EBRB 的结果属性, D 有 N 个参考等级 $D_n (n = 1, \dots, N)$; $\alpha_{i,j}^k$ 为规则 R_k 的第 i 个属性 U_i 在其第 j 个参考值 $A_{i,j}$ 上的置信度; β_n^k 为规则 R_k 在 D_n 上的置信度; $\theta_k (0 \leq \theta_k \leq 1)$ 为规则 R_k 的规则权重; $\delta_i (0 \leq \delta_i \leq 1)$ 为规则 R_k 中属性 U_i 的属性权重. EBRB 的构建流程简述如下.

step 1: 确定每个前提属性和结果属性的参考值

集合, 即确定前提属性的个数 M 、前提属性的参考值个数 $J_i (i = 1, \dots, M)$ 、参考值对应的效用值 $u(A_{i,j})$ 、结果属性的参考值个数 N 及其对应的效用值 $u(D_n)$ 。

step 2: 转换给定的输入输出为置信分布形式。假设 x_k 与 y_k 分别为第 k 组数据的输入输出, 其中 $x_k = (x_{k,1}, \dots, x_{k,M})$, 将每个前提属性与结果属性转换为如下所示的置信分布形式:

$$\begin{aligned} S(x_{k,i}) &= \{(A_{i,j}, \alpha_{i,j}^k), j = 1, \dots, J_i\}, \\ S(y_k) &= \{(D_n, \beta_n^k), n = 1, \dots, N\}. \end{aligned} \quad (2)$$

其中置信度计算方式^[7]如下所示:

$$\begin{cases} \alpha_{i,j}^k = \frac{u(A_{i,j+1}) - x_{k,i}}{u(A_{i,j+1}) - u(A_{i,j})}, \\ u(A_{i,j}) \leq x_{k,i} \leq u(A_{i,j+1}), \\ \alpha_{i,j+1}^k = 1 - \alpha_{i,j}^k, \\ \alpha_{i,t}^k = 0, t = 1, \dots, J_i, \text{ 且 } t \neq j, j + 1. \end{cases} \quad (3)$$

step 3: 生成扩展置信规则库。每个输入-输出对可以按照 step 2 生成一条规则, 第 k 组数据生成的规则 R_k 如式 (1) 所示, $k = 1, 2, \dots, L$ 。

1.2 传统 EBRB 的推理方法

假设有一条输入数据 $x = (x_1, \dots, x_M)$, $x_i (i = 1, \dots, M)$ 表示输入数据 x 在属性 U_i 上的属性值。传

$$\beta_n = \frac{\mu \times \left[\prod_{k=1}^L \left(\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k \right) - \prod_{k=1}^L \left(1 - \omega_k \sum_{i=1}^N \beta_i^k \right) \right]}{1 - \mu \times \left[\prod_{k=1}^L (1 - \omega_k) \right]}, \quad (7)$$

其中 β_j 为推理对结果 D_j 的置信度。 μ 计算如下:

$$\begin{aligned} \mu &= \left[\sum_{n=1}^N \prod_{k=1}^L \left(\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k \right) - \right. \\ &\quad \left. (N - 1) \prod_{k=1}^L \left(1 - \omega_k \sum_{i=1}^N \beta_i^k \right) \right]^{-1}. \end{aligned} \quad (8)$$

输入 x 的结果效用值如下所示:

$$f_U(x) = \sum_{s=1}^N u(D_s) \beta_s(x). \quad (9)$$

2 基于规则约减与激活因子的 EBRB 推理模型

针对现有 EBRB 推理模型研究中的局限性, 本文提出基于规则约减和激活因子的 EBRB 推理模型, 简称 Relief-EBRB 模型, 其框架结构如图 1 所示。

Relief-EBRB 的建模和推理流程如下。

step 1: 假设存在 L 条数据, 其中每条数据为 $x = (x_1, \dots, x_M)$, 根据 1.1 节描述的 EBRB 构建方

统 EBRB 的推理流程如下。

step 1: 转换输入数据形式。将每个属性 x_i 按式 (3) 转换为置信分布形式, 有

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}), j = 1, \dots, J_i\}, \quad (4)$$

其中 $\alpha_{i,j}$ 为 x_i 对 U_i 的参考值 $A_{i,j}$ 的置信度。

step 2: 计算个体匹配度。计算输入 x 与规则 R_k 在前提属性 U_i 上的个体匹配度

$$\begin{aligned} d_i^k(x_i, U_i) &= d_i^k = \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}, \\ S_i^k &= S^k(x_i, U_i) = \begin{cases} 1 - d_i^k, & d_i^k \leq 1; \\ 0, & d_i^k > 1. \end{cases} \end{aligned} \quad (5)$$

step 3: 计算激活权重。计算输入数据 x 在规则 R_k 上的激活权重

$$w_k = \frac{\theta_k \prod_{i=1}^M (S^k(x_i, U_i))^{\bar{\delta}_i}}{\sum_{j=1}^L \left[\theta_j \prod_{i=1}^M (S^j(x_i, U_i))^{\bar{\delta}_i} \right]}, \quad (6)$$

其中 $\bar{\delta}_i = \frac{\delta_i}{\max_{i=1, \dots, M} \{\delta_i\}}$ 。

step 4: 利用证据推理方法进行规则推理。证据推理解析算法如下所示:

法构建初始 EBRB。

step 2: 通过 2.1 节基于规则权重的规则约减方法约减规则, 并用约减后的规则构造新的 EBRB。

step 3: 根据 2.2 节激活因子确定方式的步骤计算激活因子 λ 的取值。

step 4: 根据 2.2 节基于激活因子的 EBRB 推理方法, 推理并输出结果 $f(x)$ 。

2.1 基于规则权重的规则约减方法

ReliefF 算法与 RReliefF(ReliefF for Regression) 算法是两种特征权重评估算法。本文基于这两种算法框架, 分别针对分类问题和回归问题提出 EBRB 规则推理重要性的权重计算方法, 随后以规则权重为依据进行规则约减, 旨在减少 EBRB 中的冗余规则对 EBRB 推理精度的影响。

2.1.1 分类 EBRB 规则权重计算算法

ReliefF 算法的基本思想是通过计算每个特征对样本之间差异的贡献来评估特征的重要性。如果这

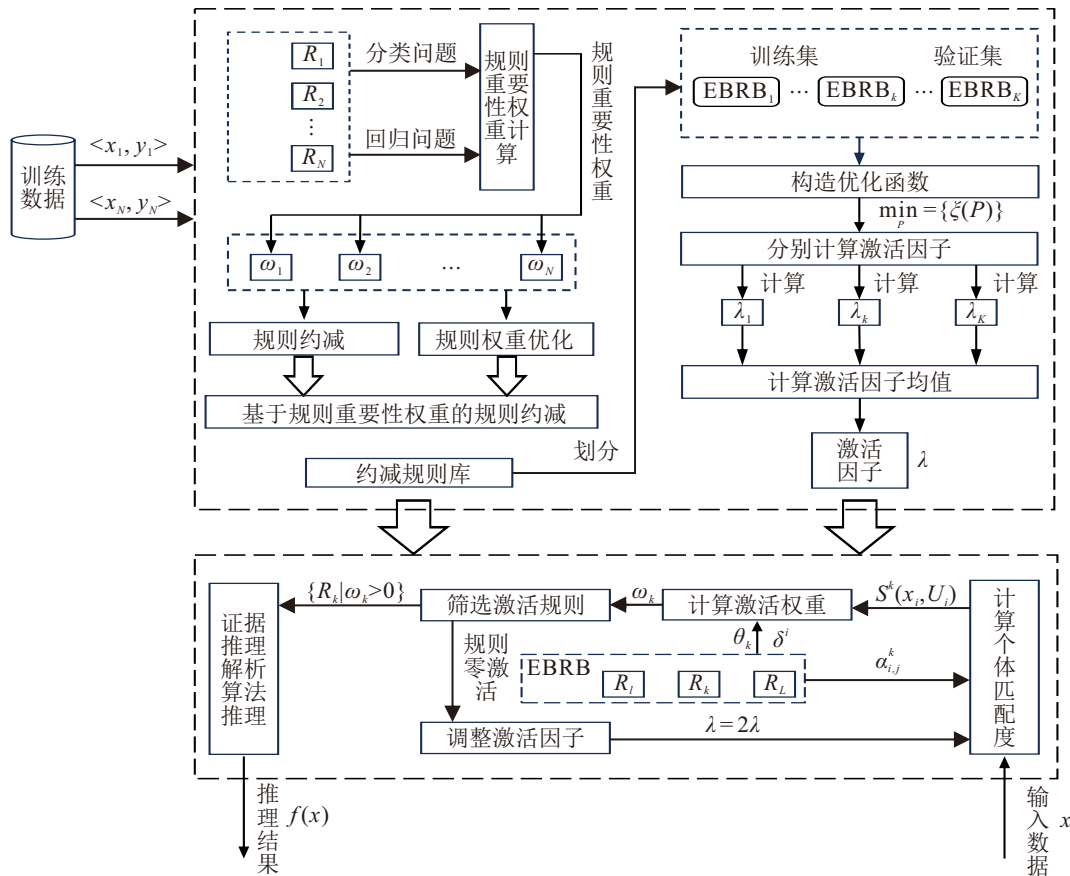


图1 基于规则约减与激活因子的 EBRB 模型框架

个特征对分类结果的影响较大, 则赋予其较大的权重. 在 ReliefF 算法中, 特征权重的估计是以下概率差的近似值^[22]:

$$W[i] = P(\text{在属性}i\text{上有差异} \mid \text{不同类近邻样本}) - P(\text{在属性}i\text{上有差异} \mid \text{同类近邻样本}). \quad (10)$$

其中 $W[i]$ 为特征 i 的权重, 其特征权重的更新公式如下:

$$W[i] = W[i] - \frac{\sum_{j=1}^A \text{diff}(R_i, H_{j,i})}{n \cdot A} + \sum_{c \neq \text{class}(R)} \sum_{j=1}^A \left[\frac{p(c)}{1 - p(\text{class}(R))} \cdot \frac{\text{diff}(R_i, M_{j,i}(C))}{n \cdot A} \right]. \quad (11)$$

这里: H_j 为同类样本, $M_j(c)$ 是类别为 c 的不同类样本, $\text{diff}(R_i, H_{j,i})$ 为样本 R 与样本 H_j 在属性 i 上的差异, n 为样本数, A 为临近样本数, $p(c)$ 为 c 类样本占总样本的比例.

在 ReliefF 算法框架下, 分类 EBRB 规则权重计算算法对规则权重的估计是以下概率差的近似值:

$$W[R_i] = P(\text{在输入上有差异} \mid \text{不同类近邻规则}) - P(\text{在输入上有差异} \mid \text{同类近邻规则}). \quad (12)$$

该算法执行过程中会遍历每条规则, 然后从该

规则的同类规则与不同类规则中分别寻找 A 个最近邻规则, 并计算这些近邻规则对于当前规则权重的平均贡献. 如果一个规则在同类临近样本和异类临近样本上的差异较大, 则认为该规则对分类有较大贡献, 相应地增加其权重. 算法伪代码如下.

算法 1 分类 EBRB 规则权重计算算法.

input: G : 运行次数, A : 近邻个数, L : 规则个数, x_s : 第 s 个规则输入, y_s : 规则 x_s 的类别, $p(c)$: c 类规则占总规则比例;
output: 规则权重平均值.

- 1: 初始化权重 $W(R_i)$ 均为 0
- 2: **for** 运行次数 $g = 1 : G$ **do**
- 3: **for** 每条规则 $l = 1 : L$ **do**
- 4: 当前规则 R_l ;
- 5: 在 R 的同类规则中找出 A 个近邻规则 H_j ;
- 6: 在 R 的异类规则中找出 A 个近邻规则 $M_j(C)$;
- 7: 计算 R 与 H_j 在输入上的 diff;
- 8: 计算 R 与 $M_j(C)$ 在输入上的 diff;
- 9: 计算本轮规则权重值 $W(R_{l,g})$;
- 10: **end for**
- 11: **end for**
- 12: 计算规则权重平均值 $W(R_i) = \frac{\sum_{g=1}^G W(R_{i,g})}{G}$.

算法 1 中, 规则 R_i 与 H_j 在输入上的差异为其在所有属性上的差异之和, 即

$$\text{diff}(R_i, H_j) = \sum_{i=1}^M \text{diff}(R_{i,i}, H_{j,i}). \quad (13)$$

其中规则 R_i 与规则 H_j 在属性 i 上的差异记为 $\text{diff}(R_{i,i}, H_{j,i})$. 对于离散属性, 有

$$\text{diff}(R_{i,i}, H_{j,i}) = \begin{cases} 0, & R_{i,i} = H_{j,i}; \\ 1, & R_{i,i} \neq H_{j,i}. \end{cases} \quad (14)$$

对于连续属性, 有

$$\text{diff}(R_{i,i}, H_{j,i}) = \frac{|R_{i,i} - H_{j,i}|}{\max(i) - \min(i)}. \quad (15)$$

最后, 规则 R_i 在第 g 轮运行中的权重计算公式为

$$W(R_{i,g}) = \frac{\sum_{c \neq \text{class}(R_i)} p(c) \sum_{j=1}^A \text{diff}(R_i, M_j(c))}{\sum_{j=1}^A \text{diff}(R_i, H_j)} \cdot \frac{1}{[1 - p(\text{class}(R_i))] \cdot L \cdot A} \quad (16)$$

经过 G 次迭代, 对规则重要性权重取均值以确定最终的规则重要性权重. 规则权重 $W(R_i)$ 的高低反映了规则 R_i 在推理中的作用大小.

2.1.2 回归 EBRB 规则权重计算算法

RReliefF 算法的基本思想与 ReliefF 算法类似, 但由于回归问题的特点, 其输出为连续的数值, 无法区分 2.1.1 节中的“同类近邻规则”与“不同类近邻规则”. RReliefF 通过引入一组概率来评估特征与连续数值输出的相关性, 从而确定特征的重要性^[23], 如下式所示:

$$P_{\text{diff}A} = P(\text{在属性 } i \text{ 上有差异} \mid \text{近邻样本}), \quad (17)$$

$$P_{\text{diff}C} = P(\text{在结果上有差异} \mid \text{近邻样本}), \quad (18)$$

$$P_{\text{diff}D} = P(\text{在结果上无差异} \mid \text{近邻样本}), \quad (19)$$

$$P_{\text{diff}C \mid \text{diff}A} = P(\text{在结果上有差异} \mid \text{在属性 } i \text{ 上有差异} \& \text{近邻样本}). \quad (20)$$

重写式 (10), RReliefF 算法对特征权重的估计值如下:

$$W[i] = P_{\text{diff}A \mid \text{diff}C} - P_{\text{diff}A \mid \text{diff}D}, \quad (21)$$

其中 $W[i]$ 表示特征 i 的权重. 运用贝叶斯规则, 可以将上式转化为

$$W[i] = \frac{P_{\text{diff}C \mid \text{diff}A} P_{\text{diff}A}}{P_{\text{diff}C}} - \frac{(1 - P_{\text{diff}C \mid \text{diff}A}) P_{\text{diff}A}}{1 - P_{\text{diff}C}}. \quad (22)$$

基于 RReliefF 算法框架, 回归 EBRB 的规则权重计算算法同样引入一组概率衡量规则与连续数值

形式的输出的相关性. 其中: 两条规则在结果上有差异的概率记为 N_{dC} ; 两条规则在输入上有差异的概率记为 N_{dA} ; 两条规则在输入上有差异且在结果上有差异的概率记为 $N_{dC \& dA}$. 该算法能评估规则与连续数值的结果属性的相关性, 并通过贝叶斯规则计算规则的重要性权重, 其伪代码如下.

算法 2 回归 EBRB 规则权重计算算法.

input: G : 运行次数, A : 近邻个数, L : 规则个数, x_s : 第 s 个规则输入, y_s : 规则 x_s 的输出;

output: 规则权重平均值.

1: 初始化权重 $W(R_i)$ 、 N_{dC} 、 N_{dA} 、 $N_{dC \& dA}$ 均为 0

2: **for** 运行次数 $g = 1 : G$ **do**

3: **for** 每条规则 $l = 1 : L$ **do**

4: 当前规则 R_l ;

5: 找出 R_l 的 A 个近邻规则 H_j ;

6: 计算 N_{dC} ;

7: 计算 N_{dA} ;

8: 计算 $N_{dC \& dA}$;

9: 计算本轮规则权重值 $W(R_{l,g})$;

10: **end for**

11: **end for** $\sum_{g=1}^G W(R_{l,g})$

12: 计算规则权重平均值 $W(R_i) = \frac{\sum_{g=1}^G W(R_{l,g})}{G}$.

算法 2 中, 这组概率可以用两个规则在输入、输出上的差异建模, 其计算方式如下所示:

$$N_{dC} = |f(R_i) - f(H_j)| \cdot d(l, j), \quad (23)$$

$$N_{dA} = \text{diff}(R_i, H_j) \cdot d(l, j), \quad (24)$$

$$N_{dC \& dA} = |f(R_i) - f(H_j)| \cdot \text{diff}(R_i, H_j) \cdot d(l, j). \quad (25)$$

其中: $\text{diff}(R_i, H_j)$ 由式 (13) 计算得出; $d(i, j)$ 为近邻规则 H_j 对规则 R_i 的影响因子, H_j 距离规则 R_i 越近对规则 R_i 的影响越大, 其计算方式如下:

$$d(l, j) = \frac{d_1(l, j)}{\sum_{k=1}^A d_1(l, k)}, \quad (26)$$

$$d_1(l, j) = e^{-\left(\frac{\text{rank}(R_i, H_j)}{\delta}\right)^2}, \quad (27)$$

$\text{rank}(R_i, H_j)$ 为 H_j 在规则 R_i 从近到远排序的近邻中的排名, δ 为使用者设定的常量.

基于式 (22), 规则 R_i 在第 g 轮运行的权重计算公式为

$$W(R_{i,g}) = \frac{N_{dC \& dA}}{N_{dC}} - \frac{N_{dA} - N_{dC \& dA}}{L - N_{dC}}. \quad (28)$$

G 次迭代后, 对规则重要性权重取均值. 规则权

重 $W(R_i)$ 越大, 规则 R_i 对推理的重要性越高.

2.1.3 基于规则权重的 EBRB 规则约减

规则权重 $W(R_i)$ 越大, 规则 R_i 对推理的重要性越高; 反之, 规则 R_i 对推理的重要性越低. 算法通过分析规则与其近邻, 得到规则的重要性权重. 因此, 若规则的重要性权重较低或与其近邻相等, 则可以认为该规则对于局部结果的影响程度不大或与其近邻相同, 可以将该规则与其近邻融合. 通过上述 EBRB 规则权重算法, 可以得到每条规则对于推理结果的重要性. 基于规则重要性权重的规则约减方法步骤如下.

step 1: 计算规则权重. 根据问题类型为分类问题或回归问题, 分别选择算法 1 或算法 2 计算 EBRB 的规则权重.

step 2: 规则约减. 根据问题种类进行规则约减: 对于回归问题, 根据 step 1 的计算结果, 将规则与其 A 近邻中规则权重相同的规则融合; 对于分类问题, 根据 step 1 的计算结果, 将规则与和他类别相同的 A 近邻中规则权重相同的规则融合, 具体如下:

$$R_{i,\text{new}} = \frac{\sum_{i=1}^A R_i}{A}. \quad (29)$$

step 3: 规则权重归一化. 将融合后的规则重要性权重归一化, 作为 EBRB 中新的规则权重, 有

$$W(R_{i,\text{new}}) = \frac{\sum_{i=1}^A W(R_i)}{A}. \quad (30)$$

不同于大多数规则约减方法, 本文提出的上述基于规则重要性权重的规则约减方法首先分析每条规则对于推理的重要性权重, 随后以规则权重为依据约减规则, 而非基于聚类的规则约减方法中仅以简单的欧氏距离作为规则约减的依据, 且保证了融合规则的相关性. 上述基于规则权重的规则约减方法, 可以很好地约减 EBRB 的规则, 且能够得到优化的规则权重.

2.2 基于激活因子的 EBRB 推理方法

本文通过引入激活因子, 缩放输入 x 与规则 R_k 在属性 U_i 上的距离 d_i^k , 同时设定阈值 1 限制激活规则的范围, 从而提高激活规则的一致性, 实现提高 EBRB 推理精度的目的.

假设有一条输入数据 $x = (x_1, \dots, x_M)$, $x_i (i = 1, \dots, M)$ 表示输入数据 x 在属性 U_i 上的属性值. 输入 x 与规则 R_k 在前提属性 U_i 上基于激活因子的个体匹配度计算如下:

$$d^k(x_i, U_i) = d_i^k = \sqrt{\frac{\sum_{j=1}^{J_i} (\alpha_{ij} - \alpha_{ij}^k)^2}{\lambda}}, \quad \lambda \in (0, 2].$$

$$S^k(x_i, U_i) = \begin{cases} 1 - d_i^k, & d_i^k \leq 1; \\ 0, & d_i^k > 1. \end{cases} \quad (31)$$

其中 λ 为激活因子. $\lambda \rightarrow 0$ 时, d_i^k 越大, 激活规则距离输入数据的距离越近; $\lambda = 1$ 时, 个体匹配度等同于修正的传统 EBRB; $\lambda = 2$ 时, $d_i^k \leq 1$, $S^k(x_i, U_i) \geq 0$, 所有规则均参与到推理过程中.

为了更好地展示激活因子对于激活规则的限制, 以 UCI 机器学习库中 Iris 分类数据集为例. 以花瓣长度、花瓣宽度两个属性作为输入, 将数据集中的样本在一个二维散点图中绘制出来, 分别用圆形、矩形、三角形表示 3 种不同类别的输出, 如图 2 所示. 图 2 中, 阴影区域内的规则为当前 EBRB 的激活规则, 假设存在一个输入数据 $x = (\text{花瓣长度}: 4.3, \text{花瓣宽度}: 1.56)$, 以菱形表示. 图 2(a) 中, 修正的传统 EBRB 会激活阴影区域的规则, 可见阴影区域内存在 virginica 与 versicolor 两种类别的规则. 图 2(c) 中, $\lambda = 0.88$, 小于图 2(a) 中的 $\lambda = 1$, 可见图 2(c) 中阴影区域缩小, 激活规则变少, 且阴影区域内仅存在一条 virginica 类别的规则, 激活规则的一致性得到了提高. 在图 2(d) 中, λ 进一步缩小至 0.21, 阴影区域内仅存在 versicolor 类别的规则, 激活规则一致性最高. 在图 2(b) 中, $\lambda = 2$, 由图 2(b) 可见所有规则均被激活, 参与到推理过程中.

基于交叉验证的激活因子取值确定方式如下.

step 1: 将初始 EBRB 随机划分为 K 组, 记为 $\text{EBRB}_k (k = 1, 2, \dots, K)$, 其中每组 EBRB 含有 L_k 条规则.

step 2: 构造优化模型, 确定 λ_k 取值. 将 EBRB_k 作为测试数据, 其余 $K - 1$ 组 EBRB 作为训练数据. 对于回归问题, 以平均绝对误差 (MAE) 为目标函数, 构造优化模型; 对于分类问题, 以分类准确率 (accuracy) 的倒数为目标函数, 构造优化模型, 有

$$\min_P = \{\xi(P)\}. \quad (32)$$

其中: 对于回归问题, $\xi(P) = \frac{1}{L_k} \sum_{t=1}^{L_k} |y_t - \hat{y}_t|$; 对于分类问题, $\xi(P) = \frac{L_k}{\text{TS}}$ (TS 表示正确预测的样本数).

step 3: 对于每一组 EBRB_k , 重复 step 2.

step 4: 计算激活因子取值

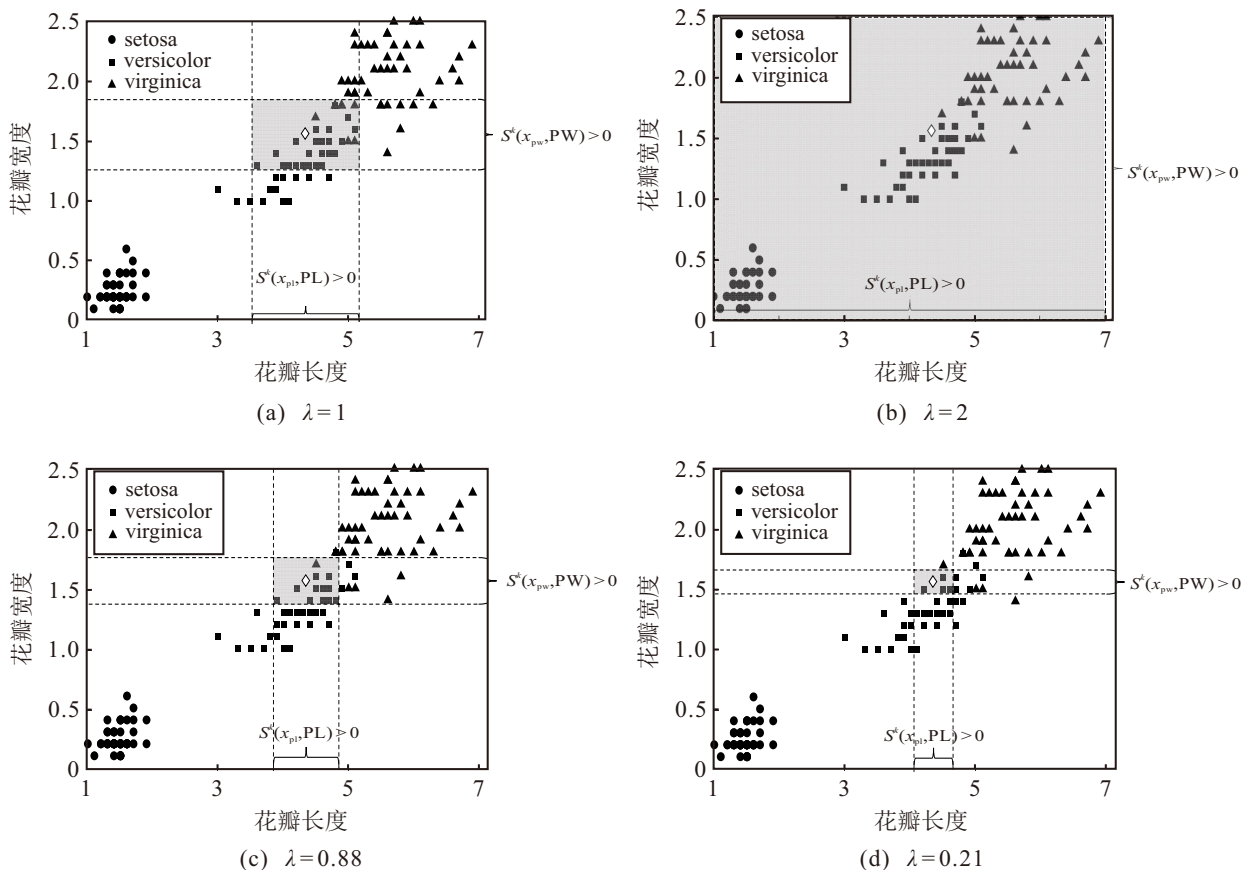


图2 不同λ下的激活规则

$$\lambda = \frac{\sum_{k=1}^K \lambda_k}{K} \tag{33}$$

经过上述 4 个步骤, 能够得到当前 EBRB 激活因子取值, 则基于激活因子的 EBRB 推理方法如下.

step 1: 计算个体匹配度. 假设有输入数据 $x = (x_1, \dots, x_M)$, 经由 1.2 节 step 1 将输入数据转换为置信分布形式后, 依据激活因子与式 (31) 计算输入与所有规则的个体匹配度.

step 2: 筛选激活规则. 根据 step 1 计算得到的个体匹配度, 依据式 (6) 计算激活权重, 随后筛选激活规则. 其中, 将激活权重大于 0 的规则参与到推理中, 即若 $\omega_k > 0$, 则 R_k 为激活规则.

step 3: 调整激活因子. 若激活规则数为 0, 则采用指数增长策略调整激活因子 λ , 并重新执行 step 2, 否则执行 step 4. 激活因子调整策略如下所示:

$$\lambda = 2\lambda. \tag{34}$$

即, 当激活规则数为 0 时, 每次更新将 λ 的值翻倍, 重新筛选激活规则. 这种方式能够以指数速度增大激活因子, 从而加速激活规则的筛选过程.

step 4: 推理并输出结果. 根据 step 2 筛选出的激活规则, 依据 1.2 节 step 4 的方法计算结果属性各参考值的置信度分布, 融合得到推理结果.

3 实验分析

3.1 回归问题

本节采用如下所示的非线性函数验证上述方法的有效性:

$$f(x) = x \cdot \sin(x^2). \tag{35}$$

将所提出的改进 EBRB 推理模型记为 Relief-EBRB, 并与 Liu-EBRB^[7]、VP/MVP-EBRB^[17]、CEAF-EBRB^[13] 进行比较.

在 $x \in [0, 3]$ 内均匀地选取 500 组输入-输出对作为训练数据, 为前提属性 x 设置 7 个参考值 $[0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0]$, 为结果属性 $f(x)$ 设置 6 个参考值 $[-2.5, -1.5, 0, 1.0, 2.5, 3.0]$. 在 x 的取值范围内均匀地选取 1000 组数据作为测试数据.

对于上述问题, 设置近邻个数 $A = 6$, 根据算法 2 计算规则权重, 随后进行规则约减, 得到约减后的 EBRB 规则 102 条. 基于第 2.2 节式 (32) 优化模型, 得到激活因子 $\lambda = 0.0099$. 表 1 总结了 Liu-EBRB、VP/MVP-EBRB、CEAF-EBRB 与 Relief-EBRB 的比较结果. 算法衡量标准为: 1) 平均绝对误差 (MAE): 所有预测误差的绝对值的平均值; 2) 均方误差 (MSE): 所有预测误差平方的平均值; 3) 规则数: EBRB 的规则数量.

表1 各 EBRB 在回归问题上的比较

	平均绝对误差	均方误差	规则数
Liu-EBRB	0.3543	0.2982	500
CEAF-EBRB	0.0520	0.0041	232
VP/MVP-EBRB	0.0232	0.0008	500
Relief-EBRB	0.0243	0.0014	102

各 EBRB 在回归问题上的比较见表 1. 由表 1 可知, Relief-EBRB、VP/MVP-EBRB、CEAF-EBRB 均明显优于 Liu-EBRB. 其中, CEAF-EBRB 与 Relief-EBRB 均对规则进行约减, 但 Relief-EBRB 约减后的规则数为 102 条, 少于 CEAF-EBRB 的 232 条, 且 Relief-EBRB 的平均绝对误差与均方误差分别为 0.024 3 和 0.001 4, 均小于 CEAF-EBRB 的 0.052 0 和 0.004 1. 同时, 虽然 VP/MVP-EBRB 的平均绝对误差与均方误差表现略优于 Relief-EBRB, 但 VP/MVP-EBRB 并没有进行规则约减, 仅对规则构造了索引, 因此规则数远多于 Relief-EBRB 的 102 条规则.

上述实验表明本文所提出的 Relief-EBRB 在规则约减方面具有显著优势, 在推理精度方面相比其他算法也有一定优势.

3.2 分类问题

为了验证本文提出的 Relief-EBRB 在分类问题

上的有效性, 本节采用 UCI 机器学习库中 9 个常见数据集比较 Liu-EBRB^[7]、VP/MVP-EBRB^[17]、CEAF-EBRB^[13]、DRA-EBRB^[16]、DEA-EBRB^[11]、M-EBRB^[18] 与 Relief-EBRB 在分类问题上的表现. 数据集的统计信息如表 2 所示.

表2 分类数据集信息统计

序号	数据集名称	样本数	属性数	类别数
1	Iris	150	4	3
2	Seeds	201	7	3
3	Transfusion	748	4	2
4	Diabetes	768	8	2
5	Ecoli	336	7	8
6	Yeast	1484	8	10
7	Glass	214	9	6
8	Wine	178	13	3
9	Cancer	683	9	2

在构造 Relief-EBRB 时, 采用十折交叉验证划分训练数据与测试数据, 采用以下两个指标衡量: 1) 准确率: 正确分类的样本占总样本比例; 2) 规则数: EBRB 中的规则数. Relief-EBRB 规则约减方法仍设置近邻个数 $A = 6$, 比较结果如表 3 所示.

表3 各 EBRB 在分类问题上的比较

	指标	Iris	Seeds	Transfusion	Diabetes	Ecoli	Yeast	Glass	Wine	Cancer	平均排名
Liu-EBRB	准确率	95.33(6)	91.33(7)	76.14(4)	73.44(5)	81.61(5)	45.61(5)	67.85(6)	81.16(4)	94.59(5)	5.22
	规则数	135(4)	186(4)	674(2)	692(4)	303(2)	1336(2)	193(3)	160(3)	615(3)	3
VP/MVP-EBRB	准确率	95.87(3)	92.57(4)	80.36(2)	72.59(6)	85.61(3)	58.15(3)	72.06(3)	—	—	3.43
	规则数	135(4)	186(4)	674(2)	692(4)	303(2)	1336(2)	193(3)	—	—	3
CEAF-EBRB	准确率	94.67(7)	93.33(3)	—	74.22(3)	—	—	—	97.22(2)	95.32(3)	3.6
	规则数	68(1)	94(1)	—	344(2)	—	—	—	79(1)	301(2)	1.4
DRA-EBRB	准确率	95.50(4)	92.02(5)	76.57(3)	71.17(7)	83.76(4)	54.13(4)	69.65(4)	83.76(3)	94.61(4)	4.22
	规则数	135(4)	186(4)	674(2)	692(4)	303(2)	1336(2)	193(3)	160(3)	615(3)	3
DEA-EBRB	准确率	95.40(5)	91.67(6)	—	74.17(4)	—	—	69.44(5)	—	—	5
	规则数	104(3)	174(3)	—	615(3)	—	—	175(2)	—	—	2.75
M-EBRB	准确率	97.80(1)	93.45(2)	—	75.17(2)	87.24(2)	58.62(2)	74.36(2)	—	95.79(2)	1.86
	规则数	135(4)	186(4)	—	692(4)	303(2)	1336(2)	193(3)	—	615(3)	3.14
Relief-EBRB	准确率	96.67(2)	93.81(1)	81.28(1)	85.81(1)	87.50(1)	66.31(1)	75.23(1)	98.31(1)	97.54(1)	1.11
	规则数	77(2)	105(2)	315(1)	299(1)	144(1)	575(1)	105(1)	86(2)	287(1)	1.33

由表 3 可知, Relief-EBRB 在各个分类数据集上均有较为不错的表现. 在规则数量方面, Relief-EBRB 在 Iris 数据集、Seeds 数据集、Wine 数据集上的规则数分别为 77、105、86, 略高于 CEAF-EBRB

的 68、94、79. 此外, Relief-EBRB 在其余数据集上的规则数分别为 315、299、144、575、105、287 条, 均少于其余 EBRB 模型的规则数. 其中, 在 Diabetes 数据集上, Relief-EBRB 在排名第 2 的 CEAF-EBRB 模

型 344 条规则的基础上减少了 45 条规则。由表 2 可知, Iris、Seeds、Wine 数据集分别为数据量最少的 3 个数据集。因此, 这 3 个数据集规则之间权重计算结果差异较大, 约减后的规则数相比 CEAF-EBRB 略高, 排名第 2。总之, 由表 3 可知, Relief-EBRB 在规则约减方面具有十分突出的表现, 能够有效约减冗余规则, 减小 EBRB 的规模。

在分类准确率方面, VP/MVP-EBRB、CEAF-EBRB、DRA-EBRB、DEA-EBRB、M-EBRB 与 Relief-EBRB 的平均表现均优于 Liu-EBRB, 可见各个 EBRB 模型均对 Liu-EBRB 进行了有效改进。在各个分类数据集上, Relief-EBRB 的准确率均有较为突出的表现。Relief-EBRB 在 Iris 数据集上的分类准确率为 96.67%, 略低于 M-EBRB 的 97.80%, 排名第 2。此外, Relief-EBRB 在其余数据集上的准确率均排名第 1。其中, Relief-EBRB 在 Diabetes 数据集和 Yeast 数据集上的表现最为突出: Relief-EBRB 在 Diabetes 数据集上的准确率为 85.81%, 高出第 2 名 M-EBRB 模型的准确率 10.64%; 在 Yeast 数据集上, Relief-EBRB 准确率为 66.31%, 相比排名第 2 的 M-EBRB 模型准确率高出 7.69%。可见 Relief-EBRB 能够很好地限制激活规则的范围, 提高激活规则的一致性, 从而得到较高的推理精度。

4 结论

本文基于 Relief 算法框架与激活因子提出了一种基于规则约减与激活因子的扩展置信规则库推理模型。与其他 EBRB 模型不同, 该模型首先通过分析规则与其近邻的相关性, 得到规则对于推理结果的重要性权重, 随后以此为依据约减冗余规则, 而非简单地以欧氏距离为依据进行规则融合。此外, 激活规则数过多会导致激活规则的一致性降低, EBRB 的推理精度也会随之降低。本文在规则约减的基础上采用激活因子限制激活规则的范围, 提高激活规则的一致性, 最终达到提高 EBRB 推理性能的目标。实验结果表明, 所提出的 EBRB 模型能够很好地识别关键规则并约减冗余规则, 同时在推理精度方面的表现也较为理想。

参考文献 (References)

- [1] 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.
- [2] Liu J, Chen S W, Martinez L, et al. A belief rule-based generic risk assessment framework[C]. *Atlantis Computational Intelligence Systems*. Paris: Atlantis Press, 2013: 145-169.
- [3] Liu J, Martinez L, Wang H, et al. An new belief rule base representation scheme and its generation by learning from examples[C]. *World Scientific Proceedings Series on Computer Engineering and Information Science*. Singapore: World Scientific, 2012: 1030-1035.
- [4] Liu J, Martinez L, Ruan D, et al. Optimization algorithm for learning consistent belief rule-base from examples[J]. *Journal of Global Optimization*, 2011, 51(2): 255-270.
- [5] 宋鑫涛, 常雷雷, 戴嘉栋, 等. 面向大型工程多安全指标的解析可追溯安全控制方法[J]. *控制与决策*, 2024, 39(1): 95-102.
(Song X T, Chang L L, Dai J D, et al. Analytical and traceable safety control method for major projects with multiple safety objectives[J]. *Control and Decision*, 2024, 39(1): 95-102.)
- [6] 廉正, 胡来红, 周志杰, 等. 基于置信规则库和弹道仿真的加速度计健康状态评估[J]. *控制与决策*, 2024, 39(9): 3143-3150.
(Lian Z, Hu L H, Zhou Z J, et al. Accelerometer health state assessment based on belief rule base and ballistic simulation[J]. *Control and Decision*, 2024, 39(9): 3143-3150.)
- [7] 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.
- [8] 吕靖, 齐海迪, 李宝德. 基于扩展置信规则库的海盗袭击事件风险预测[J]. *交通运输系统工程与信息*, 2022, 22(3): 247-254.
(Lv J, Qi H D, Li B D. Extended belief rule based system for risk prediction of pirate attacks[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2022, 22(3): 247-254.)
- [9] Fu C, Hou B B, Xue M, et al. Extended belief rule-based system with accurate rule weights and efficient rule activation for diagnosis of thyroid nodules[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(1): 251-263.
- [10] 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.)
- [11] Yang L H, Wang Y M, Lan Y X, et al. A data envelopment analysis (DEA)-based method for ruler education in extended belief-rule-based systems[J]. *Knowledge-Based Systems*, 2017, 123: 174-187.
- [12] 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.

- [13] 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.
(Yang L H, Ren T Y, Hu H B, et al. Extended belief rule base inference model based on clustering ensemble and activation factor[J]. Control and Decision, 2023, 38(3): 815-824.)
- [14] 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.
- [15] 余瑞银, 杨隆浩, 傅仰耿. 数据驱动的置信规则库构建与推理方法[J]. *计算机应用*, 2014, 34(8): 2155-2160.
(Yu R Y, Yang L H, Fu Y G. Data driven construction and inference methodology of belief rule-base[J]. *Journal of Computer Applications*, 2014, 34(8): 2155-2160.)
- [16] 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*, 2014, 27(4): 880-894.
- [17] Lin Y Q, Fu Y G, Su Q, et al. A rule activation method for extended belief rule base with VP-tree and MVP-tree[J]. *Journal of Intelligent & Fuzzy Systems*, 2017, 33(6): 3695-3705.
- [18] Zhu H Z, Xiao M Q, Yang L H, et al. A minimum centre distance rule activation method for extended belief rule-based classification systems[J]. *Applied Soft Computing*, 2020, 91: 106214.
- [19] Yang L H, Liu J, Wang Y M, et al. Extended belief-rule-based system with new activation rule determination and weight calculation for classification problems[J]. *Applied Soft Computing*, 2018, 72: 261-272.
- [20] Yang L H, Liu J, Wang Y M, et al. Enhancing extended belief rule-based systems for classification problems using decomposition strategy and overlap function[J]. *International Journal of Machine Learning and Cybernetics*, 2022, 13(3): 811-837.
- [21] Yang L H, Ren T Y, Ye F F, et al. An ensemble extended belief rule base decision model for imbalanced classification problems[J]. *Knowledge-Based Systems*, 2022, 242: 108410.
- [22] Kira K, Rendell L A. The feature selection problem: Traditional methods and a new algorithm[C]. *Proceedings of the Tenth National Conference on Artificial Intelligence*. Piscataway: IEEE, 1992: 129-134.
- [23] Robnik-Šikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression[C]. *Proceedings of the 14th International Conference on Machine Learning*. Piscataway: IEEE, 1997: 296-304.

作者简介

钟智昊 (1999-), 男, 硕士生, 主要研究方向为置信规则构建、博弈协同分配, E-mail: zhongzh@mail.nwpu.edu.cn;

龙江 (1994-), 男, 副教授, 博士, 主要研究方向为集群无人系统协同感知、决策和控制, E-mail: jlong@nwpu.edu.cn;

吴孟桐 (1999-), 男, 硕士生, 主要研究方向为无人机集群、无人机任务规划, E-mail: wumengtong@mail.nwpu.edu.cn;

郭阳明 (1978-), 男, 教授, 博士生导师, 主要研究方向为故障检测 / 诊断、预测与健康管理和可靠性优化设计, E-mail: yangming_g@nwpu.edu.cn.