

DOI: 10.11992/tis.201906046

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190829.1518.008.html>

基于改进规则激活率的扩展置信规则库推理方法

陈楠楠¹, 巩晓婷², 傅仰耿^{1,2}

(1. 福州大学 数学与计算机科学学院, 福建 福州 350116; 2. 福州大学 决策科学研究所, 福建 福州 350116)

摘 要: 数据驱动的扩展置信规则库系统, 是在传统置信规则库的基础上利用关系数据来生成规则, 使用该方法构建规则库简单有效。然而, 该方法激活的规则存在不一致与不完整, 并且该方法无法处理零激活的输入。鉴于此, 本文提出基于改进规则激活率的扩展置信规则库方法, 通过高斯核改进个体匹配度计算方法, 权衡激活规则的一致性与完整性, 并利用 k 近邻思想解决规则零激活问题。最后, 本文选取非线性函数拟合实验和输油管道检漏实验来检验所提方法的效率和准确度。实验结果表明该方法既保证了扩展置信规则库系统的推理效率, 也提高了推理结果的精度。

关键词: 置信规则库; 数据驱动; 证据推理; 个体匹配度; k 近邻思想; 零激活; 一致性; 完整性

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1179-10

中文引用格式: 陈楠楠, 巩晓婷, 傅仰耿. 基于改进规则激活率的扩展置信规则库推理方法 [J]. 智能系统学报, 2019, 14(6): 1179-1188.

英文引用格式: CHEN Nannan, GONG Xiaoting, FU Yanggeng. Extended belief rule-based reasoning method based on an improved rule activation rate[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1179-1188.

Extended belief rule-based reasoning method based on an improved rule activation rate

CHEN Nannan¹, GONG Xiaoting², FU Yanggeng^{1,2}

(1. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China; 2. Decision Sciences Institute, Fuzhou University, Fuzhou 350116, China)

Abstract: The data-driven extended belief rule-based system uses relational data to generate rules based on the traditional belief rule base. Using this method to build a rule base is simple and effective. However, the rules activated by this method are inconsistent and incomplete, and this method cannot handle none-activated inputs. Therefore, this paper proposes an extended belief rule-based method, based on an improved rule activation rate. This method improves upon the individual matching degree calculation method through gauss kernels, balances the consistency and completeness of activation rules, and solves the problem of non-activation of rules based on the idea of k -nearest neighbors. Finally, this paper selects a nonlinear function fitting experiment and an oil pipeline leak detection experiment to test the efficiency and accuracy of the proposed method. Experimental results showed that the proposed method not only ensures efficiency, but also improves the accuracy of the extended belief rule-based system.

Keywords: belief rule base; data driven; evidence reasoning; individual matching degree; k -nearest neighbors; none activation; consistency; completeness

为了建模实际问题中信息存在的不完整性、不确定性与模糊性, Yang 等^[1]提出了基于证据理

论的置信规则库推理方法 (belief rule-based inference methodology using the evidence reasoning, RIMER)。该方法将传统 IF-THEN 规则同决策理论^[2]、模糊集理论^[3]、D-S 证据理论^[4-5]等理论相结合, 使其具有处理不完整、不确定、模糊信息的能

收稿日期: 2019-06-24. 网络出版日期: 2019-08-29.

基金项目: 国家自然科学基金项目 (61773123); 福建省自然科学基金项目 (2019J01647).

通信作者: 傅仰耿. E-mail: ygf@qq.com.

力。以 RIMER 为基础构建的专家系统称为置信规则库系统^[6], 它将置信规则库作为载体来表达知识, 并且利用证据推理 (evidence reasoning, ER) 算法实现知识的推理。置信规则库系统已经成功应用于输油管道检漏^[7]、出租车乘车概率预测^[8]、消费者偏好预测^[9-10] 等方面。

早期的置信规则库 (belief rule base, BRB) 系统需要根据领域知识人为设定系统参数, 无法推广到大规模规则库的构建上, 因此一些学者提出了梯度下降优化、差分进化、变速粒子群优化等基于参数学习的优化方法^[11-15]。参数学习的引入虽然提高了规则库的推理精度, 但是大多数参数学习方法需要反复迭代得出最终结果, 无法保证推理的效率。由于 BRB 的规则数量与条件属性及条件属性候选值呈指数相关, BRB 的规模易出现“组合爆炸”, 于是一些学者利用主成分分析、关联系数标准差融合、粗糙集约减等方法对 BRB 的结构进行优化^[16-18], 这有助于提高规则库的推理效率, 但要求条件属性具有可约减性。

因此, Liu 等^[19] 提出了基于数据驱动的扩展置信规则库 (extended belief rule base, EBRB) 系统, 它既不需要进行繁复的参数训练, 又能更好解决“组合爆炸”问题。EBRB 系统在传统 BRB 系统的基础上, 扩展了规则的前提属性部分, 引入类似于 BRB 系统中结果属性的置信分布形式, 使得规则库的条件部分对模糊性、不确定性信息具有更强的表达能力, 由此引发了众多学者的关注并产生了一系列研究成果。Yang 等^[20] 通过分析规则的不完整性和不一致性, 提出了适用于规则推理过程的激活规则筛选方法。林燕清等^[21] 利用 NSGA-II 多目标优化智能算法来寻找最佳激活规则子集, 改进了 EBRB 系统的整体效果。为了解决 EBRB 系统中每条规则无序存储导致推理效率低下的问题, 苏群等^[22] 提出对 EBRB 系统规则库构建 BK 树索引, 提高了 EBRB 系统的推理性能; 在此基础上, Yang 等^[23] 针对不同条件属性维度的规则库提出基于多属性搜索框架的 EBRB 系统, 增强 EBRB 系统在多场景下的适用性; Lin 等^[24] 提出了基于 VP、MVP 索引结构的 EBRB 系统, 并且通过聚类算法实现了索引参数的自动化选择。

采用 Liu 等^[19] 所提方法来构建 EBRB 系统, 相比传统 BRB 系统, 无需训练大量参数, 同时能取得较好的推理效果。但仍存在以下问题: 1) EBRB 系统的激活规则存在规则的不一致性与

不完整性问题, 影响了 EBRB 系统的推理效率与推理精度; 2) 当输入数据与所有规则的激活权重都较低时, 推理出现异常, 无法得出结果, 即规则零激活问题。Calzada 等^[25] 提出了动态规则激活方法, 通过调节激活规则的一致性与完整性来提高 EBRB 系统的推理精度, 但是未从本质上解决规则零激活问题, 并且推理效率受算法迭代次数影响。林燕清等^[26] 提出改进的个体匹配度计算方法来解决规则零激活问题, 但同时加剧了激活规则的不一致性, 每一次输入都要通过反复迭代选取最优子集的方式来降低不一致性, 限制了 EBRB 系统的推理效率。

为解决这些问题, 本文提出基于改进规则激活率的扩展置信规则库方法, 主要贡献有: 1) 在 EBRB 系统中, 输入数据与规则库的相似性度量会直接影响到系统的推理过程, 因此本文引入基于高斯核的动态个体匹配度计算方法, 以提高 EBRB 系统的推理能力; 2) 本文利用 k 近邻方法, 对产生零激活的输入数据进行二次处理, 在保证系统效率的前提下, 解决零激活问题; 3) 本文对 EBRB 系统激活规则的一致性与完整性进行讨论, 结合新的个体匹配度计算方法, 通过控制规则激活率来平衡激活规则的一致性与完整性, 提高 EBRB 系统的推理性能。

1 EBRB 专家系统

1.1 EBRB 表示

EBRB 系统中的置信规则格式如式 (1) 所示:

$$\begin{aligned}
 R_k : & \text{IF } U_1 \text{ is } \{(A_{11}, \alpha_{11}^k), (A_{12}, \alpha_{12}^k), \dots, (A_{1J_1}, \alpha_{1J_1}^k)\} \wedge \\
 & U_2 \text{ is } \{(A_{21}, \alpha_{21}^k), (A_{22}, \alpha_{22}^k), \dots, (A_{2J_2}, \alpha_{2J_2}^k)\} \wedge \dots \wedge \\
 & U_T \text{ is } \{(A_{T1}, \alpha_{T1}^k), (A_{T2}, \alpha_{T2}^k), \dots, (A_{TJ_T}, \alpha_{TJ_T}^k)\} \\
 & \text{THEN } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \dots, (D_N, \beta_N^k)\} \\
 & \text{with a rule weight } \theta_k \text{ and attribute weights} \\
 & \delta_1, \delta_2, \dots, \delta_T; \\
 & \text{s.t. } \sum_{j=1}^N \beta_j^k \leq 1, \sum_{j=1}^{J_i} \alpha_{ij}^k \leq 1
 \end{aligned} \tag{1}$$

式中: k 表示当前描述的是第 k 条规则; U_i 表示第 i 个属性; (A_{ij}, α_{ij}^k) 为属性 U_i 的置信分布形式; A_{ij} 表示第 i 个属性的第 j 个参考值; α_{ij}^k 为其对应的置信度; T 表示属性总数; J_i 表示第 i 个属性的参考值的个数; θ_k 表示第 k 条规则的规则权重; δ_i 表示第 i 个属性的属性权重; D_j 表示第 j 个结果属性参考值, β_j^k 为 D_j 对应的置信度; N 表示结果

属性参考值的总数。

1.2 EBRB 规则构建

有别于 BRB 系统的规则库构建方式, EBRB 系统中的规则可依据数据生成。假设有 L 条数据, 且每条数据有 T 个条件属性与一个结果属性:

$$\{(x_1^k, x_2^k, \dots, x_T^k; y^k) | k = 1, 2, \dots, L\}$$

以下给出 EBRB 系统的规则生成步骤:

1) 根据领域专家的经验得到^[1], 或者通过模糊隶属函数^[27]确定每个条件属性参考值 $\{A_{ij}, i = 1, 2, \dots, T, j = 1, 2, \dots, J_i\}$ 和结果属性参考值 $\{D_j, j = 1, 2, \dots, N\}$ 。

2) 利用 1) 确定的条件属性参考值和结果属性参考值, 将训练数据的输入 X 以及输出 y 分别转化为对应的置信分布形式。本文针对数值型数据给出置信分布转化方法:

首先, 考虑生成规则库条件属性的置信分布。对第 k 条数据的输入部分 $X^k = (x_1^k, x_2^k, \dots, x_T^k)$, 考虑将第 i 个分量 x_i^k 转化成如下置信分布形式:

$$E(x_i^k) = \{(A_{ij}, \alpha_{ij}^k), j = 1, 2, \dots, J_i\} \quad (2)$$

令 γ_{ij} 表示属性参考值 A_{ij} 对应的数值, 且保证 $\{\gamma_{i(j+1)} > \gamma_{ij}, j = 1, 2, \dots, J_i - 1\}$ 。则 α_{ij}^k 的计算公式如下:

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

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

$$\alpha_{it}^k = 0, t = 1, 2, \dots, j - 1, j + 2, \dots, J_i \quad (5)$$

同理, 根据第 k 条数据的输出值 y^k , 我们也可计算得到规则的结果属性的置信分布形式:

$$E(y^k) = \{(D_j, \beta_j^k), j = 1, 2, \dots, N\}$$

3) 利用 2) 的方法, 本文可将数据 $(x_1^k, x_2^k, \dots, x_T^k; y^k)$ 转化成如式 (1) 所示的规则, 从而得到初步的规则库。

4) 确定 EBRB 中每条规则的权重以及条件属性权重。由于 EBRB 的每条规则都由数据生成的, 因此规则权重的设定需要考虑到数据质量引起的规则之间的冲突与不一致, 将不一致性指标^[19]引入规则权重的计算可以缓解规则的冲突性。

1.3 EBRB 推理机制

EBRB 系统规则库生成之后, 即可进行 EBRB 推理。给定一个 T 维输入数据 $X = (x_1, x_2, \dots, x_T)$, 根据式 (2)~(5) 可得输入对应的置信分布形式:

$$E(x_i) = \{(A_{ij}, \alpha_{ij}), i = 1, 2, \dots, T, j = 1, 2, \dots, J_i\}$$

由此, 可计算第 k 条规则与该输入关于第 i 个条件属性的个体匹配度:

$$d_i^k = \sqrt{\sum_{j=1}^{J_i} (\alpha_{ij} - \alpha_{ij}^k)^2} \quad (6)$$

$$S_i^k = 1 - d_i^k \quad (7)$$

第 k 条规则的激活权重计算公式如下:

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

$$\text{s.t. } 0 \leq \omega_k \leq 1 (k = 1, 2, \dots, L), \sum_{i=1}^L \omega_i = 1$$

接下来对激活权重不为零的规则进行 ER 合成, 并获得推理结果。首先将式 (1) 中的 β_j^k 转化为对应的基本概率值:

$$m_j^k = \omega_k \beta_j^k$$

$$m_D^k = 1 - \omega_k \sum_{j=1}^N \beta_j^k$$

$$\bar{m}_D^k = 1 - \omega_k$$

$$\bar{m}_D^k = \omega_k \left(1 - \sum_{j=1}^N \beta_j^k \right)$$

式中: m_j^k 表示第 j 个结果属性参考值的基本可信度; \bar{m}_D^k 表示规则激活权重未分配给任何结果属性参考值的基本可信度; \bar{m}_D^k 表示结果属性参考值的不完备性所导致的基本可信度。根据以上公式可推导出 $m_D^k = \bar{m}_D^k + \bar{m}_D^k$ 。

接着对所有的规则进行 ER 合成, 得到结果属性参考值 D_j 的置信度:

$$C_j = t \left[\prod_{l=1}^L (m_j^l + \bar{m}_D^l + \bar{m}_D^l) - \prod_{l=1}^L (\bar{m}_D^l + \bar{m}_D^l) \right]$$

$$\tilde{C}_D = t \left[\prod_{l=1}^L (\bar{m}_D^l + \bar{m}_D^l) - \prod_{l=1}^L \bar{m}_D^l \right]$$

$$\bar{C}_D = t \prod_{l=1}^L \bar{m}_D^l$$

$$t^{-1} = \sum_{j=1}^N \prod_{l=1}^L (m_j^l + \bar{m}_D^l + \bar{m}_D^l) - (N - 1) \prod_{l=1}^L (\bar{m}_D^l + \bar{m}_D^l)$$

$$\beta_j = \frac{C_j}{1 - \bar{C}_D}, j = 1, 2, \dots, N$$

$$\beta_D = \frac{\tilde{C}_D}{1 - \bar{C}_D}$$

当最终的输出结果要求为单一数值时, 可以通过计算规则库的期望效用值来获取最后结果:

$$f(x_i) = \sum_{j=1}^N (\mu(D_j) \beta_j) + \frac{(\mu(D_1) + \mu(D_N))}{2} \left(1 - \sum_{j=1}^N \beta_j \right)$$

2 EBRB 激活方法优化

2.1 一致性与完整性问题

EBRB 属于数据驱动的置信规则库, 因此规则库质量会受数据质量的影响。当被激活的规则

中存在冲突规则,或者包含大量与输入相关度低的规则时,证据推理的效果会受影响,EBRB系统存在规则不一致性问题。相反,当被激活的规则中只包含少量规则,一些相关度高的规则未被激活时,同样也会影响最终的推理结果,即EBRB系统存在规则不完整性问题。

被激活的规则范围越大,越容易造成规则不一致性问题;被激活的规则范围越小,则规则不完整性问题会突显。因此,为了达到更好的推理效果,需要对规则的一致性与完整性进行权衡。

2.2 个体匹配度计算方法改进

根据1.3节中EBRB系统的推理框架,可以发现式(7)中个体匹配度 S_i^k 需要满足如下条件: 1) 非负性,即 $S_i^k \in [0, 1]$,其中 $S_i^k = 0$ 表示输入数据与规则完全不匹配, $S_i^k = 1$ 表示输入数据与规则完美匹配; 2) 单调性,输入数据与规则的相似度越高,对应的个体匹配度的值越大。

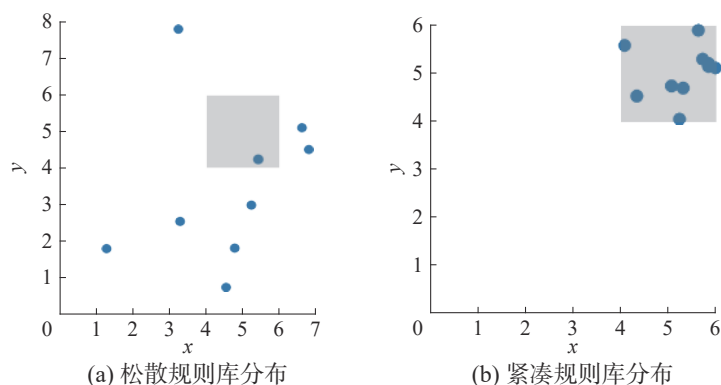


图1 静态个体匹配度计算方法的问题

Fig. 1 Problem of static individual matching calculation method

为了解决上述问题,本文将高斯核函数作为新的个体匹配度计算公式,通过引入参数 σ 来控制规则激活率,对规则的一致性与完整性进行权衡:

$$S_i^k = \exp(-(d_i^k)^2/(2\sigma^2)), \sigma > 0 \quad (9)$$

易知式(9)中 $S_i^k \in [\exp(-1/\sigma^2), 1]$,符合非负性;且 S_i^k 随 d_i^k 的减小而增大, d_i^k 越小说明个体匹配度越高,因此 S_i^k 满足单调性。

如图2所示,函数 S_1 、 S_2 分别来自文献[19, 26],函数 S_3 为本文所提方法。观察可知, S_3 相对于 S_1 与 S_2 ,有如下优点: 1) 函数 S_1 在 $d \in [1, \sqrt{2}]$ 时取值小于0,会使EBRB无法正常工作,而函数 S_3 的取值始终不小于零; 2) 函数 S_2 取值区间为 $[1/(1+\sqrt{2}), 1]$,无法保证激活规则的一致性,而函数 S_3 取值区间为 $[0, 1]$,因此激活规则的一致性可

结合式(1)中 α 的约束条件,可知式(7)中 $S_i^k \in [1-\sqrt{2}, 1]$,不符合非负性条件,EBRB系统存在隐患,不具备良好的鲁棒性。此外,式(7)中个体匹配度计算方法是静态的,无法控制规则激活率以适应不同EBRB系统对激活规则的一致性与完整性要求。本文以二维数据为例,利用Calzada等^[25]所提方法,将生成的规则库映射到二维空间中,如图1所示,其中结点代表规则库的分布,图1(a)的规则库分布密度较小,图1(b)的规则库分布密度较大。阴影区域表示利用式(7)计算得到的某一特定输入对应的激活域,位于激活域内的规则将会被激活。观察图1(a),10条规则仅有1条规则位于激活域内,易造成规则不完整性问题;相反,图1(b)中10条规则全部位于激活域内,易造成规则不一致性问题。可见,静态的个体匹配度计算方法无法适应不同分布下的EBRB系统。

以得到保证; 3) 函数 S_1 、 S_2 都是静态的,无法适应基于不同数据分布构建的规则库,而如图3所示,函数 S_3 能够通过调整参数 σ ,对个体匹配度计算方法进行调整,从而使其适应不同分布的规则库。

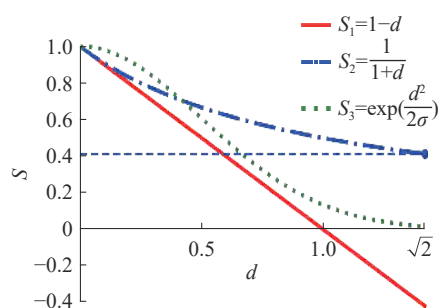
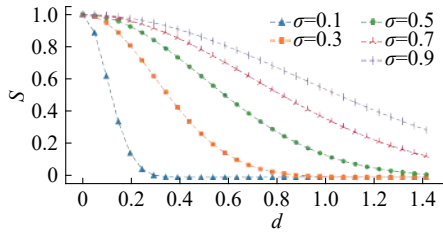


图2 不同个体匹配度计算方法对比

Fig. 2 Comparison of different individual matching methods

图3 对应不同 σ 参数的函数 S_3 Fig. 3 Function S_3 corresponding to different σ parameters

2.3 规则零激活处理方法

观察式(8),可知个体匹配度连乘得到规则激活权重,因此如果有一个属性的个体匹配度计算结果为零,则对应规则的激活权重为零。当所有规则激活权重都为零时,EBRB系统瘫痪,无法得到对应输出结果。

为了解决上述问题,林燕清等^[26]提出新的个体匹配度计算方法:

$$S_i^k = 1/(1+d_i^k) \quad (10)$$

由式(1)和式(6)可知式(10)中 S_i^k 的取值范围为 $[1/(1+\sqrt{2}), 1]$,虽然可以解决“零激活”问题,但 S_i^k 最小值也会超过0.4,这导致激活规则集合存在严重的不一致性,特别是在规则的条件属性较少的情况下。因此,林燕清等^[26]通过反复迭代选取最优激活子集的方式来降低不一致性,但这导致系统的推理效率得不到保障。

分析以上问题可以发现,主要原因是因为没有将发生零激活的输入与正常的输入分开处理。因此本文提出针对规则零激活的二次处理方法,结合 k 近邻思想^[28],在权衡规则的一致性与完整性的前提下,解决“零激活”问题。具体实现如算法1所示,其中参数 t 通过对训练数据进行多则交叉的方式获取。

算法1 零激活输入二次处理算法

输入 产生零激活的输入 $X=(x_1, x_2, \dots, x_T)$;

输出 激活权重前 t 大的规则集合 Rules2。

1) $W = []$ /*数组初始化为空*/

2) for R_k in Rules do

3) for U_i in R_k do

4) calculate similarity of (x_i, U_i) /*依据式(9),

σ 取较大值,保证规则全激活*/

5) end for

6) calculate w_k /*式(8)*/

7) $W.append(w_k)$

8) end for

9) sort(W) (descending)

10) threshold = $W[t]$

11) Rules2 = []

12) for R_k in Rules do

13) if $w_k > \text{threshold}$ do

14) Rules2.append(R_k)

15) end if

16) end for

17) return Rules2

2.4 EBRB推理方法改进

以第1节的EBRB框架为基础,结合第2.2和2.3节基于规则激活率优化的个体匹配度计算方法以及零激活输入二次处理算法,本节将介绍改进后的扩展置信规则库的推理过程。

图4为改进后的扩展置信规则库的推理过程,具体步骤描述如下。

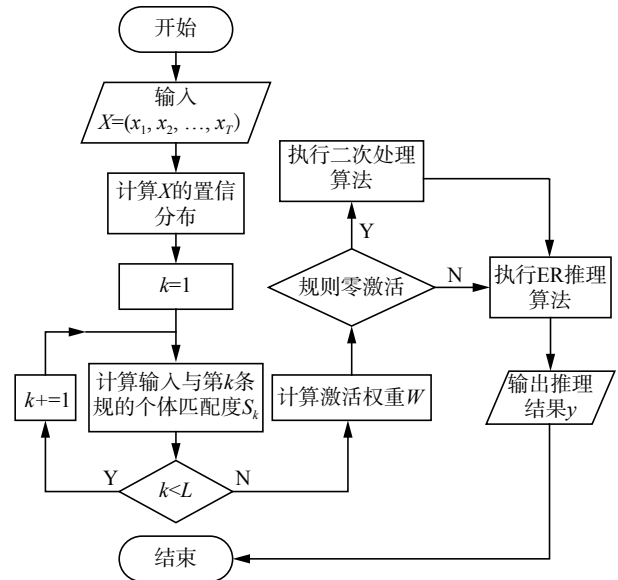


图4 改进的EBRB推理流程

Fig. 4 Improved EBRB inference process

1) 给定输入 $X=(x_1, x_2, \dots, x_T)$,首先根据式(2)~(5)计算得 X 的置信分布表示:

$$\alpha = \{\alpha_{ij}, i = 1, 2, \dots, T, j = 1, 2, \dots, J_i\}$$

2) 根据式(9)得到该输入与第 k 条规则对应的每个条件属性的个体匹配度:

$$S^k = \{S_1^k(x_1, U_1), S_2^k(x_2, U_2), \dots, S_T^k(x_T, U_T)\}$$

3) 循环执行2),得到输入 X 与规则库中所有规则的个体匹配度;

4) 以步骤2)、3)得到的结果为基础,按照公式(8)计算出每条规则对应的激活权重,如下所示:

$$W(X) = \{W_1(X), W_2(X), \dots, W_L(X)\}$$

5) 如果步骤4)中出现规则零激活问题,则执行步骤6);否则执行步骤7);

6) 执行2.3节提出的二次处理算法,重新计算个体匹配度,并且只选择激活权重前 t 大的规

则进行 ER 推理;

7) 运行 ER 算法融合所有激活规则, 得到 EBRB 系统的输出结果。

2.5 时间复杂度分析

本文提出的基于改进规则激活率的 EBRB 系统的时间复杂度主要体现在规则库构建与规则推理两个部分。其中规则库构建部分, 假设条件属性的个数为 T , 每个条件属性的候选值个数为 J , 规则总数为 L , 结果属性参考值数为 N , 则每条规则生成的时间为 $O(TJ)$, 规则初始权重调节时间为 $O(L^2TJ)$, 因此规则库总的构建时间为 $O(L^2TJ)$ 。

EBRB 系统推理部分, 将输入转变为对应的置信分布形式需要 $O(TJ)$; 计算个体匹配度的时间需要 $O(LTJ)$, 规则权重的计算需要 $O(LT)$, 若出现规则零激活, 需执行二次处理算法, 处理时间变为 $O(LTJ+L\log L)$ 。由于只有激活权重超过零的规则才会进入 ER 推理, 这里假设有 αL 条规则进入 ER 推理, 其中 $\alpha \in [0, 1]$, 则有 EBRB 的推理时间为 $O(N\alpha L)$ 。因此本文改进的 EBRB 系统处理每一条输入数据的时间复杂度为 $O(L(TJ+\alpha N+\log L))$ 。

与文献 [19] 对比可知本文方法的规则库构建部分复杂度与其相同。规则推理部分, Liu^[19]的时间复杂度为 $O(L(TJ+N))$ 。当输入未发生零激活时, 本文方法的推理复杂度为 $O(L(TJ+\alpha N))$, 要优于 Liu^[19]的方法。当输入发生零激活时, Liu^[19]的方法会出现异常, 而本文以较小时间代价, 解决了零激活问题。

3 实验与结果

本文首先介绍函数仿真实验, 研究参数 σ 对 EBRB 系统的影响, 接着选取输油管道检漏作为实验案例, 与其他方法进行对比分析, 验证本文方法的有效性。实验环境为: Intel(R) Core i7-6700@ 3.40 GHz; 16 GB 内存; Windows 10 操作系统; 算法在 Python3.6 环境下编写。

3.1 非线性函数拟合

林燕清等^[19]表明 EBRB 系统可以拟合任意非线性函数, 为了验证本文方法的效果, 将通过一个常用非线性函数进行测试, 其数学表达式为

$$f(x) = x \sin(x^2), 0 \leq x \leq 3 \quad (11)$$

构建 EBRB 系统的过程中, 令 x 为条件属性, 并且在区间 $[0, 3]$ 内均匀选取 7 个点当作参考值, 分别为 $\{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ 。函数值 $f(x)$ 作为输出结果, 其评价等级效用值设为 $\{-2.5, -1, 1, 2, 3\}$ 。在区间 $[0, 3]$ 中均匀选择 500 个点作为训练数据中

x 的取值, 计算出 $f(x)$ 并且加入高斯随机数噪声作为训练数据的输出结果, 训练数据如图 5 所示。利用训练数据构建出 EBRB 系统规则库之后, 在区间 $[0, 3]$ 中均匀选择 1 000 个点作为测试数据, 计算 EBRB 系统的输出与真实值 $f(x)$ 之间的损失, 用均方误差 (mean square error, MSE) 作为评价依据。

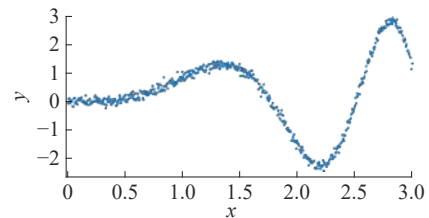


图 5 训练数据集

Fig. 5 Training data set

对式 (9) 中的 σ 取不同的值, 观察改进后的 EBRB 系统对非线性函数的拟合效果, 实验结果如表 1 所示, 其中平均处理时间是指处理一个输入需要耗费的平均时间。从表 1 中, 可观察到随着 σ 的减小, MSE 呈现出先下降后上升的趋势。 σ 从 1 变化到 0.05 的过程中, MSE 下降是因为当 σ 较大时, 平均激活规则数较多, 这容易引起规则的不一致性与不相关性问题, 因此随着 σ 的减小, 平均激活规则数也随之减小, 规则的一致性与相关性得到一定保障, EBRB 系统的预测精度提升; 当 σ 从 0.05 变化到 0.002 时, 平均激活规则数的过度减小引起了规则的不完整性问题, EBRB 系统的预测能力也逐渐降低。同时可以观察到随着平均激活规则数的降低, 平均处理时间也跟随下降, 这是因为进入证据推理环节的规则变少导致推理的时间减少。

表 1 不同 σ 值函数拟合效果

Table 1 Function fitting effects of different σ

参数 σ	平均激活规则数	平均处理时间/ms	MSE
1	500	5.6	1.271 665
0.5	500	5.7	0.557 641
0.1	65	4.0	0.006 046
0.05	31	3.7	0.003 678
0.01	7	3.6	0.007 308
0.002	2	3.6	0.008 809

图 6 为表 1 对应的函数拟合曲线对比图。当 σ 较大时, EBRB 系统的预测结果未能拟合式 (11), 但随着 σ 的减小, 预测结果在不断逼近式 (11); 当 σ 位于区间 $[0.05, 0.1]$ 时, 预测结果已经能够基本拟合式 (11), 且可以看出虽然 $\sigma = 0.05$ 时的 MSE

更小,但 $\sigma = 0.1$ 时曲线要更加光滑,因此可以判断出预测结果已经过拟合;随着 σ 的进一步减小,

预测结果出现锯齿状,且MSE也开始回升,EBRB系统过拟合现象更加明显。

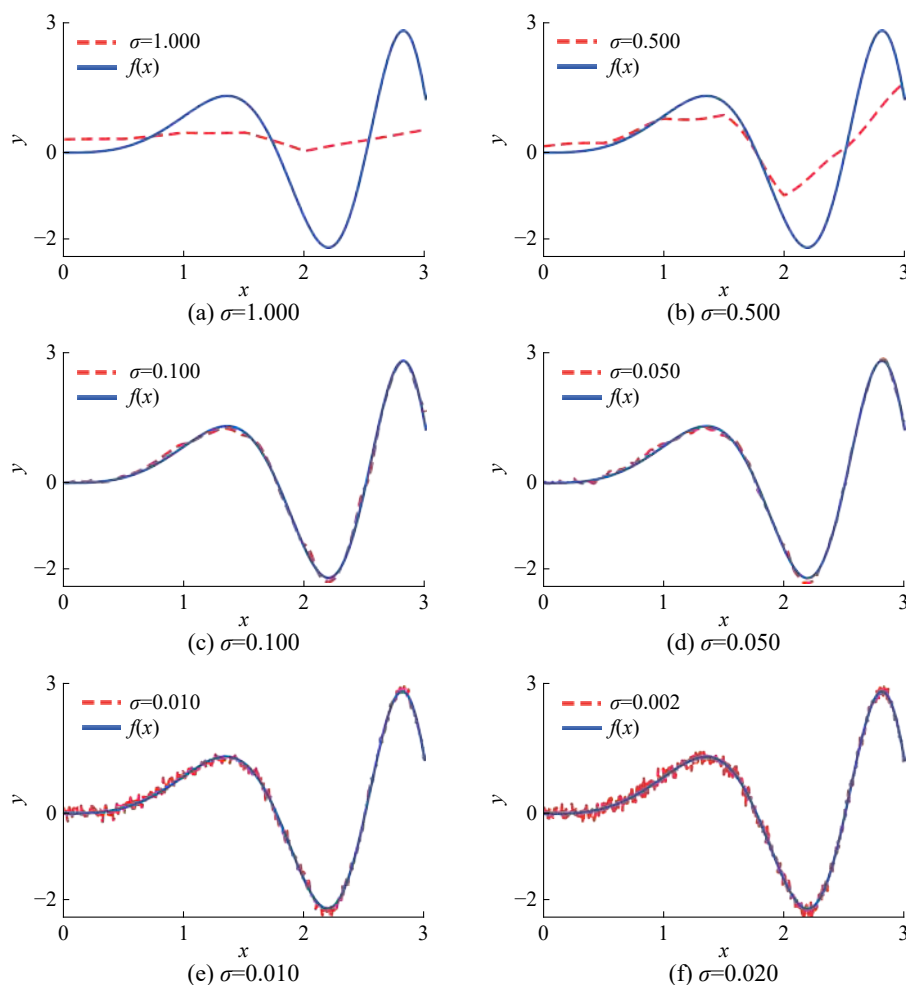


图6 不同 σ 值函数拟合曲线

Fig. 6 Function fitting curve corresponding to different σ

为了验证本文所提个体匹配度计算方法的有效性,本文将其与文献[19,26]中的个体匹配度计算方法进行对比。为了保证对比结果的可靠性,本文对文献[19,26]的方法也引入零激活输入二次处理算法,并且将个体匹配度小于零的值上调为零。

表2列出了3种方法的函数拟合输出结果,其中EBRB-S₁、EBRB-S₂分别代表文献[19,26]中个体匹配度计算方法构建的EBRB系统,EBRB-S₃代表本文所提方法。从表2中可得出EBRB-S₃的MSE值要远低于EBRB-S₁、EBRB-S₂,并且EBRB-S₁方法中出现个体匹配度小于零的输入有1 000条,EBRB-S₁中所使用的个体匹配度计算方法存在较大缺陷。EBRB-S₃方法不仅精度高,而且运行效率良好,其平均处理时间比EBRB-S₁降低了9.1%,比EBRB-S₂降低了33.3%。

表2 不同个体匹配度的函数拟合结果

Table 2 Function fitting results based on different individual matching methods

EBRB方法	平均激活 规则数	个体匹配度 (<0)	平均处理 时间/ms	MSE
EBRB-S ₁	160	1 000	4.4	0.297 247
EBRB-S ₂	500	0	6	1.289 319
EBRB-S ₃	65	0	4	0.006 046

图7为表2对应的函数拟合曲线,从中可观察到EBRB-S₂基本未拟合函数,主要是因为其对应的个体匹配度计算方法引起的规则不一致与不相关程度较大;EBRB-S₁的拟合效果要比EBRB-S₂好,但在函数的极值部分,拟合效果仍不理想;只有EBRB-S₃基本拟合了曲线的走势,并且曲线的走势也比较光滑。

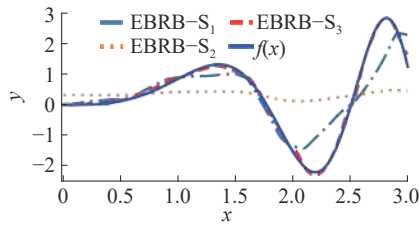


图7 不同个体匹配度的函数拟合曲线

Fig. 7 Function fitting curve based on different individual matching methods

通过非线性函数拟合实验,验证了 EBRB-S₃ 系统的重要性质。接下来将通过实际生活中的输油管道检漏实验进一步验证本文方法的有效性。

3.2 输油管道检漏

输油管道检漏实验的研究对象为英国一条长达一百多公里的输油管道,领域专家通过管道的入口与出口流量差异 (flow difference, FD) 以及输油管道内的平均压力差 (pressure difference, PD) 这两个因素来检测输油管道发生泄漏的大小 (leak size, LS), 因此该实验以 FD 和 PD 作为 EBRB 系统的两个条件属性,泄露大小 LS 作为输出结果。根据领域专家的经验给出两个条件属性参考值分别为 $A_{FD} = \{-10, -5, -3, -1, 0, 1, 2, 3\}$ 、 $A_{PD} = \{-0.042, -0.025, -0.010, 0.000, 0.010, 0.025, 0.042\}$, 设定输出结果属性参考值为 $D_{LS} = \{0, 2, 4, 6, 8\}$ 。

为了同文献 [19,25-26] 进行比较,本文选择出 500 组数据作为训练数据,然后利用本文所提方法构建 EBRB-S₃ 规则库。最后用构建好的系统预测 2 008 组测试数据,并将实验结果的平均绝对误差 (mean absolute error, MAE) 作为评价依据,同其他的 EBRB 系统进行对比。

表 3 为 3.1 节所提到的 EBRB-S₁、EBRB-S₂、EBRB-S₃ 3 种 EBRB 系统的输油管道检漏实验结果,EBRB-S₃ 的 MAE 值相比 EBRB-S₁ 下降了 32.5%, 并且远小于 EBRB-S₂。这表明本文所提个体匹配度计算方法能够提高 EBRB 系统的性能。观察 3 种方法的平均处理时间,与表 2 相比差距变小,主要原因是输油管道检漏实验的规则库条件属性数量与参考值数量都比非线性函数拟合多,因此置信分布转化、个体匹配度计算等初始化时间占了主要部分,激活规则数量变化导致的时间差异被弱化。

表 3 不同个体匹配度的输油管道检漏实验结果

Table 3 Test results of oil pipeline leak detection based on different individual matching methods

EBRB方法	平均激活规则数	平均处理时间/ms	MAE
EBRB-S ₁	276	8.6	0.246 045
EBRB-S ₂	500	9.1	1.450 974
EBRB-S ₃	10	8.7	0.166 041

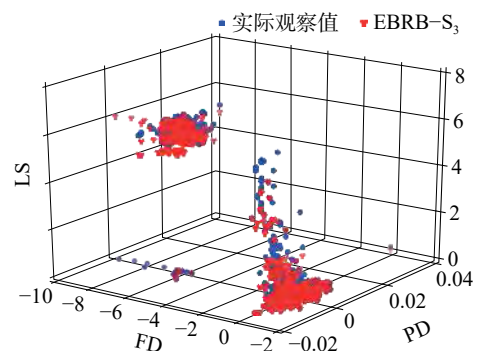
表 4 列出了 Liu-EBRB 系统^[19]、DRA-EBRB 系统^[25]、SRA-EBRB 系统^[26]以及本文方法的输出结果与真实结果之间的 MAE 值,系统处理每条输入需要花费的平均时间、出现零激活的输入个数 (Failed)。其中 EBRB-S₃ 的 MAE 值最小,比排名第二小的 SRA-EBRB 降低了 19.7%。并且 EBRB-S₃ 的平均处理时间仅次于 Liu-EBRB,比 SRA-EBRB 降低了 62.3%,比 DRA-EBRB 降低了 87.0%,这主要是因为 SRA-EBRB 与 DRA-EBRB 都是基于迭代的方法,需要反复扫描 EBRB 系统,因此平均处理时间相对较长。Liu-EBRB 出现了一条引起规则零激活的输入,而 EBRB-S₃ 对每一条输入都能够给出合理输出,这主要归功于 EBRB-S₃ 引入了零激活输入二次处理算法,从根本上解决规则零激活问题。综上所述,EBRB-S₃ 在保证系统运行效率的前提下,提高了 EBRB 系统的推理精度,增强了 EBRB 系统的鲁棒性。

表 4 与其他方法进行对比

Table 4 Compares with other methods

EBRB方法	MAE	平均处理时/ms	Failed
Liu-EBRB	0.246 029	5.7	1
DRA-EBRB	0.221 253	66.8	0
SRA-EBRB	0.206 692	23.1	0
EBRB-S ₃	0.166 041	8.7	0

图 8 为 EBRB-S₃ 的预测值与真实观测值的三维对比图,可以发现 EBRB-S₃ 能够得到接近真实值的结果。

图8 EBRB-S₃与真实输出三维图Fig. 8 3D diagram of EBRB-S₃ and real output

4 结束语

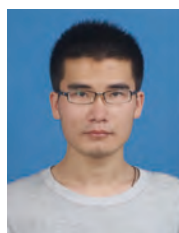
本文针对现有 EBRB 系统激活规则的不一致性与不完整性,以及无法处理“零激活”输入的问题,提出了基于改进规则激活率的扩展置信规则库方法。该方法将高斯核函数作为新的个体匹配度计算方法,对规则的激活率进行调整,保证 EBRB 系统激活规则的一致性与完整性的权衡;对于出现规则零激活现象的输入,本文提出了基于 k 近邻思想的二次处理算法,解决了规则零激活问题。为了验证本文方法的有效性,通过非线性函数拟合和输油管道检漏两个实验与其他 EBRB 系统进行了对比分析,研究结果表明本文所提方法取得了预期效果。

参考文献:

- [1] YANG Jianbo, LIU Jun, WANG Jin, et al. Belief rule-based 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] HWANG C L, YOON K. Methods for multiple attribute decision making[M]//HWANG C L, YOON K. Multiple Attribute Decision Making. Berlin, Heidelberg: Springer, 1981: 58–191.
- [3] ZADEH L A, KLIR G J, YUAN Bo. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers[M]. River Edge, NJ: World Scientific, 1996.
- [4] DEMPSTER A P. A generalization of Bayesian inference[J]. *Journal of the royal statistical society: series B (methodological)*, 1968, 30(2): 205–247.
- [5] ROTA G C. A mathematical theory of evidence: G. Shafer, Princeton University Press, 1976, 297 pp[J]. Shafer, Princeton University Press, 1977, 24(3): 341.
- [6] 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.
ZHOU Zhijie, YANG Jianbo, HU Changhua, et al. Belief rule base expert system and complex system modeling[M]. Beijing: Science Press, 2011.
- [7] XU Dongling, LIU Jun, YANG Jianbo, 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.
- [8] 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 Longhao, CAI Zhiling, HUANG Zhixin, 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.
- [9] WANG Yingming, YANG Jianbo, XU Dongling, et al. Consumer preference prediction by using a hybrid evidential reasoning and belief rule-based methodology[J]. *Expert systems with applications*, 2009, 36(4): 8421–8430.
- [10] YANG Jianbo, WANG Yingming, XU Dongling, et al. Belief rule-based methodology for mapping consumer preferences and setting product targets[J]. *Expert systems with applications*, 2012, 39(5): 4749–4759.
- [11] YANG Jianbo, LIU Jun, XU Dongling, et al. Optimization models for training belief-rule-based systems[J]. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, 2007, 37(4): 569–585.
- [12] CHEN Yuwang, YANG Jianbo, XU Dongling, et al. Inference analysis and adaptive training for belief rule based systems[J]. *Expert systems with applications*, 2011, 38(10): 12845–12860.
- [13] 常瑞, 张速. 基于优化步长和梯度法的置信规则库参数学习方法 [J]. *华北水利水电大学学报*, 2011, 32(1): 154–157.
CHANG Rui, ZHANG Su. An algorithm for training parameters in belief rule-bases based on gradient methods with optimization step size[J]. *Journal of North China Institute of Water Conservancy and Hydroelectric Power*, 2011, 32(1): 154–157.
- [14] 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 Hanjie, YANG Longhao, FU Yanggeng, et al. Differential evolutionary algorithm for parameter training of belief rule base under expert intervention[J]. *Computer science*, 2015, 42(5): 88–93.
- [15] 苏群, 杨隆浩, 傅仰耿, 等. 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 Qun, YANG Longhao, FU Yanggeng, et al. Parameter training approach based on variable particle swarm optimization for belief rule base[J]. *Journal of computer applications*, 2014, 34(8): 2161–2165.
- [16] CHANG Leilei, ZHOU Yu, JIANG Jiang, et al. Structure learning for belief rule base expert system: a comparative study[J]. *Knowledge-based systems*, 2013, 39: 159–172.
- [17] 杨隆浩, 王晓东, 傅仰耿. 基于关联系数标准差融合的置信规则库规则约简方法 [J]. *信息与控制*, 2015, 44(1): 21–28, 37.
YANG Longhao, WANG Xiaodong, FU Yanggeng. Rule reduction approach to belief rule base using correlation coefficient and standard deviation integrated method[J].

- Information and control, 2015, 44(1): 21–28, 37.
- [18] 王应明, 杨隆浩, 常雷雷, 等. 置信规则库规则约简的粗糙集方法 [J]. 控制与决策, 2014, 29(11): 1943–1950.
- WANG Yingming, YANG Longhao, CHANG Leilei, et al. Rough set method for rule reduction in belief rule base[J]. Control and decision, 2014, 29(11): 1943–1950.
- [19] LIU Jun, 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.
- [20] YANG Longhao, WANG Yingming, FU Yanggeng. A consistency analysis-based rule activation method for extended belief-rule-based systems[J]. Information sciences, 2018, 445-446: 50–65.
- [21] 林燕清, 傅仰耿. 基于 NSGA-II 的扩展置信规则库激活规则多目标优化方法 [J]. 智能系统学报, 2018, 13(3): 422–430.
- LIN Yanqing, FU Yanggeng. NSGA-II-based EBRB rules activation multi-objective optimization[J]. CAAI transactions on intelligent systems, 2018, 13(3): 422–430.
- [22] 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 Qun, YANG Longhao, FU Yanggeng, 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.
- [23] YANG Longhao, WANG Yingming, SU Qun, et al. Multi-attribute search framework for optimizing extended belief rule-based systems[J]. Information sciences, 2016, 370/371: 159–183.
- [24] LIN Yanqing, FU Yanggeng, SU Qun, et al. A rule activation method for extended belief rule base with VP-tree and MVP-tree[J]. Journal of intelligent and fuzzy systems, 2017, 33(6): 3695–3705.
- [25] CALZADA A, LIU Jun, WANG Hui, 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.
- [26] 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.
- [27] JIAO Lianmeng, PAN Qun, DENOEU T, et al. Belief rule-based classification system: extension of FRBCS in belief functions framework[J]. Information sciences, 2015, 309: 26–49.
- [28] COVER T M, HART P E. Nearest neighbor pattern classification[J]. IEEE transactions on information theory, 1967, 13(1): 21–27.

作者简介:



陈楠楠, 男, 1993 年生, 硕士研究生, 主要研究方向为智能决策与专家系统。



巩晓婷, 女, 1982 年生, 讲师, 主要研究方向为不确定多准则决策、信息隐藏技术。参与国家自然科学基金项目 3 项、福建省自然科学基金项目 2 项和教育部高等学校博士学科点专项科研基金项目 1 项。发表学术论文 10 余篇。



傅仰耿, 男, 1981 年生, 副教授, 博士, CCF 会员, CAAI 会员, 主要研究方向为决策理论与方法、数据挖掘、机器学习、智能系统。主持国家自然科学基金项目 1 项、福建省自然科学基金项目 2 项。获国家发明专利授权 2 项, 发表学术论文 30 余篇。