

DOI: 10.11992/tis.201710012

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180409.0915.002.html>

基于 NSGA-II 的扩展置信规则库激活规则多目标优化方法

林燕清, 傅仰耿

(福州大学 数学与计算机科学学院, 福建 福州 350116)

摘 要: 针对扩展置信规则库 (extended belief rule base, EBRB) 系统在不一致的激活规则过多时推理准确性不高的问题, 引入带精英策略的快速非支配排序遗传算法 (NSGA-II), 提出一种基于 NSGA-II 的激活规则多目标优化方法。该方法首先将激活权重大于零的规则 (即激活规则) 进行二进制编码, 把最终参与合成推理的激活规则集合的不一致性以及激活权重和作为多目标优化问题的目标函数, 通过带精英策略的快速非支配排序遗传算法求解不一致性更小的激活规则集合, 从而降低不一致激活规则对于 EBRB 系统推理准确性的影响。为了验证本文方法的有效性和可行性, 引入非线性函数和输油管道检漏实例进行测试。实验结果表明, 基于 NSGA-II 的扩展置信规则库激活规则多目标优化方法能够有效提高 EBRB 系统的推理能力。

关键词: 扩展置信规则库; 不一致性; 激活规则; 多目标优化; NSGA-II 算法

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2018)03-0422-09

中文引用格式: 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.

英文引用格式: LIN Yanqing, FU Yanggeng. NSGA-II-based EBRB rules activation multi-objective optimization[J]. CAAI transactions on intelligent systems, 2018, 13(3): 422-430.

NSGA-II-based EBRB rules activation multi-objective optimization

LIN Yanqing, FU Yanggeng

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

Abstract: To address the low reasoning accuracy of extended belief rule-base (EBRB) systems with too many inconsistent activated rules, this paper introduces a fast elitist non-dominated sorting genetic algorithm (NSGA-II) and proposes a rule activation multi-objective optimization approach based on the NSGA-II algorithm. In this approach, binary coding is carried out for the activated rules whose activation weights are greater than zero. The inconsistent set of activated rules following synthetic reasoning and the sum of activation weights are taken as the objective function of the multi-objective optimization problem. Using the fast elitist non-dominated sorting genetic algorithm, the problem of a set of activation rules with a small inconsistency is solved, reducing the effect of the inconsistent activated rules on the reasoning accuracy of EBER systems. To validate the efficiency and feasibility of the proposed method, this paper introduces a nonlinear function and the proposed method was tested against the leak detection of an oil pipeline. The experimental results show that the rule activation multi-objective optimization approach based on NSGA-II can effectively improve the reasoning performance of EBRB systems.

Keywords: extended belief rule base (EBRB); inconsistency; activation rules; multi-objective optimization; NSGA-II algorithm

收稿日期: 2017-10-17. 网络出版日期: 2018-04-09.

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

通信作者: 傅仰耿. E-mail: ygf@qq.com

专家系统^[1]作为一种人工智能系统, 已经被广泛应用于图像处理、医疗检测、地质勘探、石油化工等领域。利用专家系统进行决策时, 需要先将有

的信息表示成知识,知识的表示形式有很多种,其中应用最广泛的是产生式规则,即 IF-THEN 规则。置信规则 (belief rule) 是在传统 IF-THEN 规则的结果部分加入置信分布转化而来的,由一组置信规则组成的集合称为置信规则库 (belief rule base, BRB)。置信规则对应的是系统的输入信息,而实际工程中包含的信息既可能是定性知识也可能是定量数据,同时,信息中往往存在着含糊或模糊不确定性、不精确性以及不完整性。为了能够有效利用带有这些不确定性的信息, Yang 等^[2]于 2006 年在 D-S 证据理论^[3-4]、决策理论^[5]、模糊理论^[6]和传统的 IF-THEN 规则库基础上提出了基于证据理论的置信规则库推理方法 (belief rule-base inference methodology using the evidential reasoning approach, RIMER)。RIMER 方法作为 BRB 系统的核心,能够对具有线性或非线性关系的输入和输出进行建模,具有处理各种不确定性的能力。目前, BRB 系统已被广泛应用到输油管道泄漏^[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。

利用 BRB 系统对实际问题进行建模时,要先确定其内部参数的取值,各个参数的不同取值会对 BRB 系统的推理能力造成一定的影响,传统的参数取值确定方法是专家根据经验给定其值,但当参数个数较多或系统较复杂时,专家很难确定参数取值。鉴于此, Yang 等^[10]最先提出一种通用的 BRB 参数学习模型,该模型以满足各个参数约束条件为基础,通过最小化由置信规则库推理得出的模拟输出值和真实数据的输出值二者的差值来训练 BRB 参数。在 Yang 的基础上, Chen 等^[13]提出了包括前提属性参考值在内的全局参数学习方法。不过以上两种参数学习方法均是建立在 MATLAB 工具自带的 FMINCON 函数的基础上,参数学习效率不高。鉴于此,常瑞等^[14]、吴伟昆等^[15]提出结合梯度下降法的 BRB 参数学习方法,这种方法虽然优于 FMINCON 函数,但其涉及的公式推导过于复杂,只能训练少量的参数,不适合用来学习大量的参数,且参数学习的收敛速度过慢。为了提高 BRB 系统参数学习的收敛速度和精度,苏群等^[16]、王韩杰等^[17]相继提出基于群智能算法的参数学习方法。以上参数学习方法均能够优化 BRB 参数取值,从而提高 BRB 的推理能力,但是,这些参数学习方法都是建立在 Yang 等^[10]提出的参数学习模型基础上的,无法避免重复的搜索过程,需要不断地进行迭代。此外,因为 BRB 系统构建时需要覆盖所有的前提属性的参考值,当前提属性和前提属性参考值过多时,会出现“组合爆炸”问题。为了解决该问题,专家学

者们提出了各自的降维方法: Zhou 等^[18]提出“统计效用”的概念,根据每条规则的统计效用值来确定是否删减该规则,如果某条规则的统计效用值越小,说明该规则对整体系统的贡献越小,就可以将其删除; Chang 等^[19]提出利用灰靶理论、多维尺度变换、主成分分析来选择关键前提属性,从而减少前提属性的数量;王应明等^[20]引入粗糙集理论,提出客观的规则约简方法,该方法不需要借助 BRB 以外的任何先验知识。这些 BRB 结构优化方法虽然可以在一定程度上避免组合爆炸问题,但是会降低 BRB 的推理能力。为了从根本上解决 BRB 参数取值以及组合爆炸问题, Liu 等^[21]对原有的置信规则进行扩展,在规则的前件部分也加入置信分布,提高其对含糊、不完整和不确定信息的表示能力,改进后的规则库被称为扩展置信规则库 (extended belief rule base, EBRB)。同时, Liu 等^[21]还提出了一种简单且有效直接利用训练数据生成初始扩展置信规则的规则生成机制,数据驱动型的规则生成机制可以详细表示出数据中包含的各种信息,不需要进行反复迭代的参数学习过程也不会造成组合爆炸问题。但正因为扩展置信规则是根据训练数据集得到的,数据集的质量对 EBRB 系统的推理能力影响要更大,相互矛盾、不一致的数据容易降低 EBRB 的推理准确性。在 EBRB 中,数据的不一致是指两条或多条规则的前提属性取值大致相同,但评价结果却完全不同或与专家知识相冲突。

针对 EBRB 系统中的数据不一致性的问题, Alberto 等^[22]提出动态规则激活方法,该方法通过调整改进后的相似性度量公式中的参数来选择不一致性最小的激活规则集合,不过,该方法中用于衡量激活规则集合不一致性的公式只考虑了规则数量对其的影响,没有考虑规则激活权重对其的影响。要减小激活规则集合之间的一致性,最简单的方法就是舍弃部分不一致性较高的激活规则,这部分激活规则不参与最终的合成推理过程,但这部分激活规则可能带有重要的信息,而且很难确定这部分激活规则的具体数量。为了减小激活规则集合的不一致性同时保留激活规则集合中的大部分重要信息,本文提出了基于 NSGA-II 的扩展置信规则库激活规则多目标优化方法,将激活规则的不一致性与激活权重和分别作为目标函数,通过求解多目标优化问题获得相对较优的激活规则集合并用于最终的合成推理。本文首先介绍扩展置信规则库和多目标优化问题的基础知识,然后介绍如何利用 NSGA-II 求解 Pareto 的最优解,最后通过非线性函数问题和输油管道检漏实例对所提方法进行实验验证,并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

扩展置信规则库由一系列扩展置信规则 (extended belief rule) 组成, 其中第 k 条扩展置信规则的表示为

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

with a rule weight θ_k and attribute weight δ_i

式中: (A, α^k) 表示扩展置信规则前件部分的置信分布, 可表示成 $\{(A_{i,j}, \alpha_{i,j}^k), j = 1, 2, \dots, J_i | i = 1, 2, \dots, T_k\}$, $A_{i,j}$ 表示第 i 个前提属性的第 j 个参考值, 且第 i 个前提属性的参考值总数为 J_i , $\alpha_{i,j}^k$ 表示第 k 条规则的第 i 个前提属性输入值相对该属性的第 j 个参考值 $A_{i,j}$ 的置信度, T_k 表示第 k 条规则前提属性总数; δ_i 表示第 i 个前提属性的属性权重; θ_k 表示第 k 条规则的规则权重; $\beta_j^k (j = 1, 2, \dots, N; k = 1, 2, \dots, L)$ 表示第 k 条规则第 j 个评价结果 D_j 的置信度, 每条置信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 $\sum_{j=1}^N \beta_j^k = 1$ 则称第 k 条规则是完整的, 否则称第 k 条规则是不完整的。

与 BRB 系统复杂烦琐的规则生成机制不同, Liu 等^[21]提出的规则生成机制简单且有效, 可直接将训练数据转化为扩展置信规则。假设 $U_i (i = 1, 2, \dots, T_k)$ 示第 k 个样本数据的第 i 个前提属性, 其输入值为 x_i , 首先决策者或专家需要将第 i 个前提属性的参考值 $A_{i,j}$ 与数值量建立起对应关系:

$$\gamma_{i,j} \text{ means } A_{i,j} \quad (2)$$

然后, 将输入 x_i 转化成式 (3) 所示的期望形式:

$$S(x_i) = \{(\gamma_{i,j}, \alpha_{i,j}), i = 1, 2, \dots, T_k, j = 1, 2, \dots, J_i\} \quad (3)$$

式中 $\alpha_{i,j}$ 的计算公式为

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

$$\alpha_{i,j+1} = 1 - \alpha_{i,j}, \gamma_{i,j} \leq x_i \leq \gamma_{i,j+1}, j = 1, 2, \dots, J_i - 1 \quad (5)$$

$$\alpha_{i,s} = 0, s \neq j, j+1, s = 1, 2, \dots, J_i \quad (6)$$

通过式 (4) ~ (6) 可得到 $\alpha_{i,j}$ 的具体取值从而生成规则的前件部分, 相应输出的评价结果置信分布可采用同样的方法产生。

1.2 扩展置信规则库推理方法

1.2.1 激活权重的计算

假设第 i 个前提属性取值 x_i 已经被表示成式 (3) 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 可通过两个置信分布的距离值来衡量, 因为 EBRB 前件部分的置信分布实质上是概率分布, 故 S_i^k 可借助式 (7) 所示的欧氏距离来计算:

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

则 S_i^k 计算方法为

$$S_i^k = 1 - d_i^k \quad (8)$$

第 k 条置信规则的激活权重可由式 (9) 得

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} (S_i^k)^{\delta_i}}{\sum_{l=1}^L \left[\theta_l \prod_{i=1}^{T_k} (S_i^l)^{\delta_i} \right]} \bar{\delta}_i = \frac{\delta_i}{\max\{\delta_i\}_{i=1,2,\dots,T_k}} \quad (9)$$

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。

1.2.2 激活规则的合成

用证据推理方法 (evidential reasoning, ER)^[23] 得到推理结果前, 要先按式 (10) ~ (13) 计算评价结果置信度的基本可信值:

$$m_{n,i} = \omega_i \beta_{n,i} \quad (10)$$

$$m_{H,i} = 1 - \omega_i \sum_{n=1}^N \beta_{n,i} = \bar{m}_{H,i} + \tilde{m}_{H,i} \quad (11)$$

$$\tilde{m}_{H,i} = \omega_i (1 - \sum_{n=1}^N \beta_{n,i}) \quad (12)$$

$$\bar{m}_{H,i} = 1 - \omega_i \quad (13)$$

式中: $n = 1, 2, \dots, N$ 。在此基础上通过 ER 解析公式^[24] 计算评价结果的基本可信值, 合成公式如式 (14) ~ (19):

$$C_n = k \left[\prod_{j=1}^L (m_{n,j} + \bar{m}_{H,j} + \tilde{m}_{H,j}) - \prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}) \right] \quad (14)$$

$$\tilde{C}_H = k \left[\prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}) - \prod_{j=1}^L \bar{m}_{H,j} \right] \quad (15)$$

$$\bar{C}_H = k \prod_{j=1}^L \bar{m}_{H,j} \quad (16)$$

$$k^{-1} = \sum_{n=1}^N \prod_{j=1}^L (m_{n,j} + \bar{m}_{H,j} + \tilde{m}_{H,j}) - (N-1) \prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}) \quad (17)$$

$$\beta_n = \frac{C_n}{1 - C_H}, n = 1, 2, \dots, N \quad (18)$$

$$\beta_H = \frac{\tilde{C}_H}{1 - C_H} \quad (19)$$

根据式 (14) ~ (19) 可得到式 (20) 所示的具有置信分布形式的 BRB 推理输出:

$$f(x) = \{(D_n, \beta_n), n = 1, 2, \dots, N\} \quad (20)$$

基于上述 ER 解析算法, Wang 等^[25] 进一步推导出了组合所有的置信规则的计算公式, 即

$$\beta_j = \frac{\mu \times \left[\prod_{k=1}^L (\omega_k \beta_{j,k} + 1 - \omega_k \sum_{i=1}^N \beta_{i,k}) - \prod_{k=1}^L (1 - \omega_k \sum_{i=1}^N \beta_{i,k}) \right]}{1 - \mu \times \left[\prod_{k=1}^L (1 - \omega_k) \right]} \quad (21)$$

$$\mu = \left[\frac{\sum_{j=1}^N \prod_{k=1}^L (\omega_k \beta_{j,k} + 1 - \omega_k \sum_{i=1}^N \beta_{i,k}) - (N-1) \prod_{k=1}^L (1 - \omega_k \sum_{i=1}^N \beta_{i,k})}{\sum_{j=1}^N \prod_{k=1}^L (\omega_k \beta_{j,k} + 1 - \omega_k \sum_{i=1}^N \beta_{i,k})} \right]^{-1} \quad (22)$$

1.3 问题提出

Liu 等^[21]提出的 EBRB 规则生成机制虽然简单且有效,但也使得 EBRB 系统的推理性能容易受训练数据的质量影响,由于训练数据生成的扩展置信规则库可能存在不一致的规则即规则相互矛盾的问题,这些不一致的规则会降低 EBRB 系统的推理准确性,尤其当这些不一致规则同时成为激活规则。在 EBRB 中,这些激活权重大于零的规则被称为激活规则,激活规则是用来进行 ER 合成推理的,即 EBRB 系统的推理结果就是依靠这些激活规则得到的。由此可见,激活规则对于最终推理结果的重要性,而相互矛盾的、不一致的激活规则会对最终的 BRB 系统推理结果造成一定的干扰,进而影响 BRB 系统的推理能力。为此,Alberto 等^[22]对式(8)进行改进,提出动态规则激活方法,通过不断重复的搜索过程以找到不一致性最小的激活规则集合,该方法能够有效减小激活规则之间的不一致性,但其参数取值需要不断迭代,而且参数增加和减小幅度也较难确定。此外,实际工程应用中,训练数据总数都比较多,当采用 Liu 等^[21]提出的规则激活方法时,多数规则的激活权重都会大于零,激活规则数量的增多,意味着规则间的不一致性增大。要减小激活规则之间的不一致性,最简单的方法就是尽可能多地减少激活规则的数量,不一致性较高的这部分激活规则不参与最终的合成推理过程,但这种方法不一定有效,因为激活规则数量一旦减少,原有激活规则集合中包含的信息就会减少,如果这些不参与最终合成推理过程的激活规则包含了原有激活规则中绝大部分重要信息,则 EBRB 的推理准确性也会受到一定程度的影响。在 EBRB 中,激活规则的激活权重代表激活规则的重要性,激活权重越大,说明该激活规则越重要,其中包含的重要信息越多。鉴于此,为了减小激活规则之间的不一致性,同时保留住原有激活规则集合的绝大部分重要信息,本文提出激活规则多目标优化方法,把激活规则之间的不一致性以及激活规则的激活权重总和作为优化目标,通过 NSGA-II 来求解较优的激活规则集合用于最终的合成推理。

2 基于 NSGA-II 的 EBRB 激活规则多目标优化方法

2.1 多目标优化问题

在实际应用问题中,所求解的优化目标通常包

含多个,且它们之间经常是相互矛盾、冲突的。也就是说,对于这一类问题,几乎找不到一个可以同时满足所有优化目标的最优解。一个由 m 个决策参数和 n 个目标变量组成的多目标优化问题的数学表达式^[26]为

$$\begin{aligned} & \text{Min/Max } y = F(x) = (f_1(x), f_2(x), \dots, f_n(x)) \\ & \text{sub to: } g_i(X) \leq 0, i = 1, 2, \dots, k_g \\ & h_i(X) = 0, i = 1, 2, \dots, k_h \\ & \text{where: } x = (x_1, x_2, \dots, x_m) \in X \subseteq R \\ & y = (y_1, y_2, \dots, y_m) \in Y \subseteq R \end{aligned} \quad (23)$$

式中: $x = (x_1, x_2, \dots, x_m)$ 表示 m 维的决策参数; $y = (y_1, y_2, \dots, y_n)$ 表示 n 维的目标变量; $F(x) = (f_1(x), f_2(x), \dots, f_n(x))$ 表示所有的目标函数; $g_i(X) \leq 0$ 表示所有的不等式约束条件; $h_i(X) = 0$ 表示所有的等式约束条件。

定义 1 (可行解) 如果存在一个决策参数 x 它满足所有不等式约束条件和等式约束条件,则称 x 为可行解。

定义 2 (可行解集合) 可行解集合是指所有可行解组成的集合,记作 $x_f (x_f \subseteq X)$ 。

定义 3 (Pareto 占优) 假设 $x_A, x_B (x_A, x_B \in X_f)$ 是多目标优化问题的两个可行解,若 x_A 相对 x_B 是 Pareto 占优 (或称 x_A 支配 x_B , 记作 $x_A > x_B$), 则 x_A, x_B 需要同时满足以下两个条件:

- 1) $\forall i = 1, 2, \dots, n, f_i(x_B) \geq f_i(x_A)$;
- 2) $\exists j = 1, 2, \dots, n, f_j(x_B) > f_j(x_A)$ 。

定义 4 (Pareto 最优解) 若多目标优化问题的一个可行解 $x_C (x_C \in X_f)$ 是 Pareto 最优解,则 x_C 需要满足条件 $\neg \exists x \in X_f : x > x_C$ 。

2.2 带精英策略的快速非支配排序遗传算法

2000 年 Kalyanmoy Deb 等^[27]首次提出了带精英策略的快速非支配排序遗传算法 (简称 NSGA-II), 该方法是众多求解多目标优化问题方法中应用最为广泛的一种。NSGA-II 算法是非支配排序遗传算法 (non-dominated sorting genetic algorithm, NSGA) 的改进, 运行速度更快, 复杂度更低, 且其求解的 Pareto 最优解集收敛性更好。算法首先随机产生一定规模数量的初始种群, 然后利用非支配排序方法对种群中所有个体进行分层, 接着执行遗传算法的选择、交叉和变异 3 个操作产生第一代子群。从第二代子群开始, 先将父代种群和子代种群中所有个体合并在一起, 然后利用快速非支配排序方法对其进行分层, 并计算每个非支配分层中所有个体的拥挤距离, 在此基础上从中选出较优的个体组成新的父代种群, 接着执行遗传算法的选择、交叉和变异 3 个操作产生下一代子群, 直至达到程序结束条件时终止, 算法的具体流程如图 1 所示。

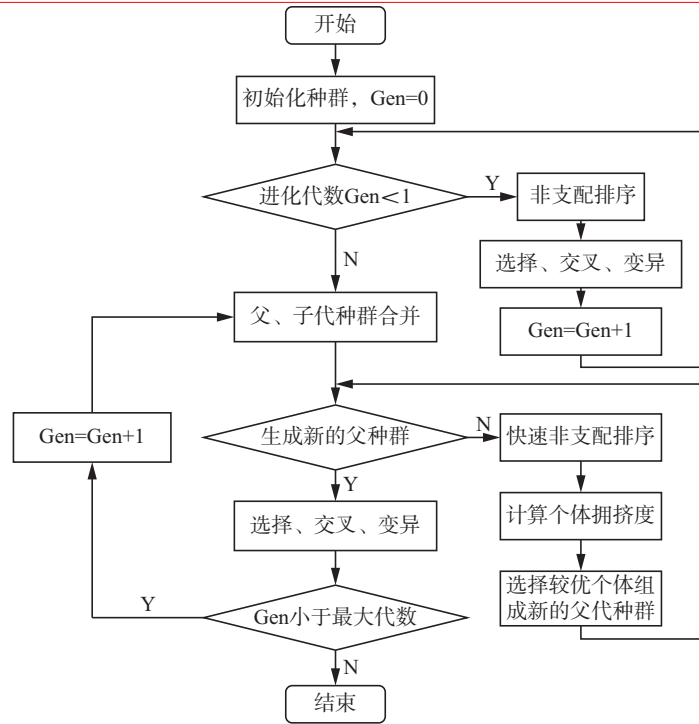


图 1 NSGA-II 算法流程

Fig. 1 The process of NSGA-II

NSGA-II 算法中个体的优劣之分主要由个体的两个属性取值来决定, 一个是其所在的非支配分层级别, 另一个是个体的拥挤距离。前者是通过快速非支配排序方法确定, NSGA-II 的快速非支配方法与 NSGA 的非支配排序方法相比, 计算复杂度从原先的 $O(MN^3)$ 减少至 $O(MN^2)$ (其中 M 为种群大小, M 为目标函数的个数)。计算处于同一个非支配分层的个体拥挤距离之前, 需要先对所有个体的拥挤距离进行初始化, 然后根据目标函数将其按照升序进行排序, 接着再计算每个个体的拥挤距离。详细的快速非支配排序算法以及个体拥挤距离的计算过程可参见文献[27]。

确定完每个个体所在的非支配分层级别以及拥挤距离之后, 就可以确定种群所有个体的优劣, 假设其中两个个体 i 和 j , 其所在非支配分层级别为 i_{rank} 和 j_{rank} , 拥挤距离为 i_{distance} 和 j_{distance} , 如果这两个个体满足以下两个条件中的一个条件:

- 1) $i_{\text{rank}} < j_{\text{rank}}$;
- 2) $i_{\text{rank}} = j_{\text{rank}}$ 并且 $i_{\text{distance}} > j_{\text{distance}}$;

则称个体 i 优于 j , 表示成 $i \prec_n j$ 。

2.3 基于 NSGA-II 的激活规则多目标优化方法

数据驱动型的 EBRB 规则数量等于训练数据集数量, 每条置信规则对应一个训练数据, 当训练数据过多时, 由式 (9) 计算得到的激活规则数量也会比较多, 但很多激活规则之间存在相互矛盾、不

一致的情况, 这些激活规则会对推理造成一定的干扰。为了减少不一致激活规则对 EBRB 推理准确性的影响, 本文提出基于 NSGA-II 的激活规则多目标优化方法, 因此, 多目标优化的两个目标函数分别为激活规则集合的不一致性以及激活权重和, 其中, 激活规则集合的不一致性用 Liu 等^[21]提出的方法来衡量, 假设 R_p, R_q 为扩展置信规则库中的两条规则, 二者的不一致性可通过前提属性相似度 SRA 和评价结果相似度 SRC 来衡量, 规则 p 和规则 q 的 SRA 和 SRC 计算公式如下:

$$\text{SRA}(R_p, R_q) = \min_{i=1}^T \left(1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j}^p - \alpha_{i,j}^q)^2} \right) \quad (24)$$

$$\text{SRC}(R_p, R_q) = 1 - \sqrt{\sum_{i=1}^N (\beta_i^p - \beta_i^q)^2} \quad (25)$$

根据文献[21], 规则 p 和规则 q 之间的一致性可根据式 (26) 计算得到:

$$\text{Cons}(R_p, R_q) = \exp\left(-\frac{(\text{SRA}(R_p, R_q)/\text{SRC}(R_p, R_q) - 1.0)^2}{1/\text{SRA}(R_p, R_q)^2}\right) \quad (26)$$

那么, 第 i 条规则的不一致性为

$$\text{Incons}(i) = \sum_{k=1, k \neq i}^l [1.0 - \text{Cons}(R_i, R_k)] \quad (27)$$

由此可得 NSGA-II 的优化目标为

$$\begin{aligned} \text{Min}F(R) &= (f_1(R), f_2(R)) \\ f_1(R) &= \sum_{k=1}^l \text{Incons}(k) \\ f_2(R) &= 1.0 - \sum_{k=1}^l \omega_k \end{aligned} \quad (28)$$

式中: R 表示最终参与ER合成推理的激活规则集合; l 表示参与ER合成推理的激活规则总数。

基于NSGA-II的激活规则多目标优化方法具体流程如图2所示,该方法根据激活规则集合的不一致性以及激活权重和来求解最终参与ER合成推理的激活规则集合 R 。本文方法首先计算每条扩展置信规则的激活权重,激活权重大于零的规则组成激活规则集合,然后对激活规则集合进行二进制编码,1表示该激活规则参与最终ER合成推理过程,0表示该激活规则不参与ER合成推理,不同的激活规则集合,其中被标识为1和0的规则不同,根据式(27)计算得到的规则不一致性以及激活权重和也不同,故接下来需要利用NSGA-II算法求解最优优化目标(式(28))的Pareto最优解集,这些Pareto最优解集中编码为1对应的激活规则之间的一致性既要最小同时激活权重和也要最大,然后从最优解集中选择一个合适的Pareto最优解,该最优解中编码为1对应的激活规则组成最终的激活规则集合并用于ER合成推理得出结果。本文方法的整体时间复杂度包括:1)EBRB系统查询激活规则复杂度 $O(MN)$; 2)NSGA-II算法中非支配排序方法复杂度 $O(MN^2)$; 3)NSGA-II算法中拥挤距离计算复杂度 $O(MN \log N)$; 4)NSGA-II算法中个体优劣排序复杂度 $O(N \log N)$ 。因此整体复杂度为 $O(MN^2)$ 。

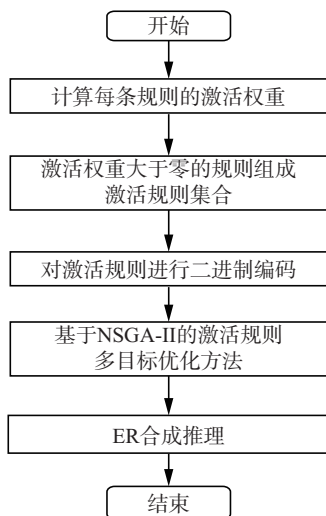


图2 激活规则多目标优化方法

Fig. 2 The activated rules multi-objective optimization approach

3 实例分析

本文引入非线性函数和输油管道检漏实例为研究对象以验证本文方法的有效性。实验中用到的NSGA-II算法的各个参数值分别为:个体交叉概率为0.9,个体变异概率为0.03,种群规模为100,进化代数为1000。实验环境为:Inter(R) Core(TM) i5-4570 CPU @ 3.20 GHz; 8 GB 内存; Windows 10 操作系统;算法实现平台 Visual Studio 2010。

3.1 非线性函数问题

为了验证本文方法的有效性,引入一个非线性函数作为基准测试函数进行测试以说明本文方法的性能。非线性函数的数学表达式为

$$f(x) = x \sin(x^2), 0 \leq x \leq 3 \quad (29)$$

在EBRB系统构建中,选取函数的自变量 x 作为前提属性,并从其定义域中均匀选择7个数值值作为其参考值,即{0,0.5,1.0,1.5,2.0,2.5,3.0},然后根据函数值设定评价结果等级数及相应效用值为{-2.5,-1.0,1.0,2.0,3.0}。该实验的测试数据是从 x 的定义域中均匀选取的500组数据,训练数据是从 x 的定义域中均匀选取的1000组数据,因此,构建的EBRB系统总共有500条规则,然后根据Liu等^[21]提出的方法和本文提出的方法构建EBRB系统,实验结果衡量的指标为系统的推理输出和真实输出之间的均方误差(mean squared error, MSE)、激活规则总数(Activated_rules)、参与合成推理的激活规则数(ER_rules),实验结果如表1所示。

表1 非线性函数问题实验结果
Table 1 The results on nonlinear function

衡量指标	Liu_EBRB	NSGA-II_EBRB
MSE	0.296 986	0.224 212
Activated_rules	160 043	160 043
ER_rules	160 043	79 843

从表1可以发现,NSGA-II_EBRB系统的推理准确性要比Liu_EBRB系统的推理准确性高,这主要是因为Liu_EBRB系统中用来参与ER合成推理的规则是所有激活权重大于零的激活规则,这部分规则的不一致性较高且会降低EBRB系统的推理能力,而NSGA-II_EBRB通过减少激活规则的不一致性,选择不一致性更小同时拥有原来激活规则中绝大部分信息的激活规则进行ER合成推理,最终合成推理的激活规则数只占原来激活规则总数的49.89%,这些激活规则不一致性较低,从而提高EBRB系统的推理能力。

3.2 输油管道检漏实例

为了验证本文所提方法的有效性,引入一个具体的实际问题——输油管道泄漏问题^[7]作为实例。该实例的研究对象为安装在英国的一条1 000多米长的输油管道,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FD、PD的参考值由专家根据经验给出,分别为{-10,-5,-3,0,1,2,3}和{-0.042,-0.025,-0.01,0,0.01,0.025,0.042},LS的评价等级为零、很小、中、高、很高,其数值效用值为{0,2,4,6,8}。

该实验的测试数据是发生泄漏的2 008组数据,训练数据按照一定比例分别从3个时间段随机选取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NSGA-II_EBRB系统,并将实验结果和Liu_EBRB系统相比较,衡量的指标为平均绝对误差(mean absolute difference, MAE)。

Liu_EBRB系统、NSGA-II_EBRB系统产生的推理输出和测试数据的真实输出对比如图3、4所示。分析图3、4可以发现,Liu_EBRB系统在 $PD \in [-0.02, 0.04]$, $FD \in [-10, 0]$ 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NSGA-II_EBRB系统推理输出总体上和真实输出相接近。这是因为,本文方法减少了不一致规则对于推理结果的影响,从而进一步提高了算法的推理性能。

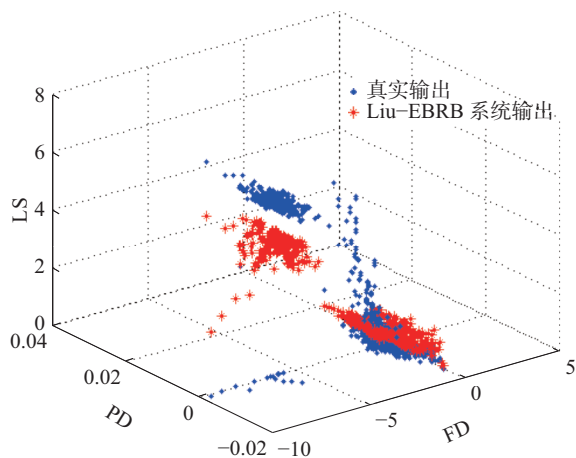


图3 Liu_EBRB输出和真实输出

Fig. 3 Liu_EBRB output and real output

为了进一步验证本文方法的有效性和可行性,表2列出了Liu_EBRB系统、BK_EBRB系统和

NSGA-II_EBRB系统产生的推理输出和测试数据的真实输出的MAE值,其中BK_EBRB系统是指根据文献[28]中的方法构建的EBRB系统。如表2所示,和Liu_EBRB系统的MAE相比,NSGA-II_EBRB系统比Liu_EBRB系统的MAE值减小了61.61%;和BK_EBRB系统相比,NSGA-II_EBRB系统要比BK_EBRB($\theta=0.7$)的MAE值减小56.92%;和BK_EBRB($\theta=0.4$)的MAE值相差无几。

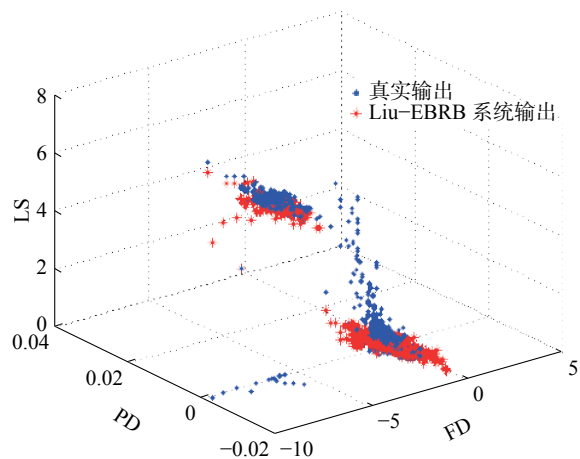


图4 NSGA-II_EBRB输出和真实输出

Fig. 4 NSGA-II_EBRB output and real output

表2 输油管道泄漏实例实验结果

Table 2 The results on pipeline leak detection

EBRB 系统类型	MAE
Liu_EBRB	0.626 240
BK_EBRB($\theta=1.0$)	0.626 240
BK_EBRB($\theta=0.7$)	0.558 087
BK_EBRB($\theta=0.4$)	0.231 400
NSGA-II_EBRB	0.240 443

4 结束语

数据驱动型的扩展置信规则库系统易受数据质量的影响,其推理能力常因不一致的数据而降低。因此,本文提出基于NSGA-II的扩展置信规则库激活规则多目标优化方法,通过NSGA-II来求解不一致性更小的激活规则集合,该方法既筛选出了不一致性更小的激活规则,同时又保留了原来激活规则集合中绝大部分信息,这些最终参与ER合成推理的激活规则一致性更高,更具代表性,能有效提高EBRB系统的推理能力。然而,本文方法还有许多需要改进的地方,如何从Pareto最优解集中选择一

个最合适的解,如何减少算法的复杂度等,这些都是将来的研究工作重点。

参考文献:

- [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 Jianbo, LIU Jun, WANG Jin, 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. Series B (methodological), 1968, 30(2): 205–247.
- [4] SHAFER G. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976.
- [5] HWANG C L, YOON K. Multiple attribute decision making: methods and applications a state of the art survey[M]. New York: Springer, 1981: 22–34.
- [6] ZADEH L A. Fuzzy sets[J]. Information and control, 1965, 8(3): 338–353.
- [7] ZHOU Zhijie, HU Changhua, YANG Jianbo, et al. Online updating belief rule based system for pipeline leak detection under expert intervention[J]. Expert systems with applications, 2009, 36(4): 7700–7709.
- [8] 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.
- [9] 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.
- [10] YANG Ying, FU Chao, CHEN Yuwang, et al. A belief rule based expert system for predicting consumer preference in new product development[J]. Knowledge-based systems, 2016, 94: 105–113.
- [11] JIANG Jiang, LI Xuan, ZHOU Zhijie, 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 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.
- [13] 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.
- [14] 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]. 华北水利水电学院学报, 2011, 32(1): 154–157.
Chang Rui, Zhang Su. An algorithm for training parameters in belief rule-bases based on the gradient methods with optimization step size[J]. Journal of north China institute of water conservancy and hydroelectric power, 2011, 32(1): 154–157.
- [15] 吴伟昆, 杨隆浩, 傅仰耿, 等. 基于加速梯度求法的置信规则库参数训练方法[J]. 计算机科学与探索, 2014, 8(8): 989–1001.
WU Weikun, YANG Longhao, FU Yanggeng, et al. Parameter training approach for belief rule base using the accelerating of gradient algorithm[J]. Journal of frontiers of computer science and technology, 2014, 8(8): 989–1001.
- [16] 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 application, 2014, 34(8): 2161–2165.
- [17] 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.
- [18] ZHOU Zhijie, HU Changhua, YANG Jianbo, et al. A sequential learning algorithm for online constructing belief-rule-based systems[J]. Expert systems with applications, 2010, 37(2): 1790–1799.
- [19] 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.
- [20] 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.
- [21] LIU Jun, MARTINEZ L, CALZADA A, et al. A novel be-

- lief rule base representation, generation and its inference methodology[J]. Knowledge-based systems, 2013, 53: 129–141.
- [22] 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, 7(4): 880–894.
- [23] YANG Jianbo. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties[J]. European journal of operational research, 2001, 131(1): 31–61.
- [24] WANG Yingming, YANG Jianbo, XU Dongling. Environmental impact assessment using the evidential reasoning approach[J]. European journal of operational research, 2006, 174(3): 1885–1913.
- [25] 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.
- [26] ZITZLER E. Evolutionary algorithm for multi-objective optimization: methods and applications[D]. Zurich: Swiss Federal Institute of Technology, 1999.
- [27] DEB K, AGRAWAL S, PRATAP A, et al. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA- II [C]//Proceedings of the 6th International Conference Paris Parallel Problem Solving from Nature PPSN VI. Kanpur, India, 2000.
- [28] 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.

作者简介:



林燕清, 女, 1992 年生, 硕士研究生, 主要研究方向为智能决策与专家系统。



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。