

分类号 TP18
收藏编号
学校代码 10386



密级 公开
学号 130320045
编号

福州大学

学术型硕士研究生学位（毕业）论文

置信规则库系统的优化方法研究

学 科 专 业： 计算机软件与理论

研 究 方 向： 智能决策技术

研 究 生 姓 名： 苏群

指 导 教 师、职 称： 王晓东 教授

所 在 学 院： 数学与计算机科学学院

答辩委员会主席签名：

二〇一五 年 十二 月

一 遵守学术行为规范承诺

本人已熟知并愿意自觉遵守《福州大学研究生和导师学术行为规范暂行规定》和《福州大学关于加强研究生毕业与学位论文质量管理的规定》的所有内容，承诺所提交的毕业和学位论文是终稿，不存在学术造假或学术不端行为，且论文的纸质版与电子版内容完全一致。

二 独创性声明

本人声明所提交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得福州大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。本人完全意识到本声明的法律结果由本人承担。

三 关于论文使用授权的说明

本人完全了解福州大学有关保留使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。（保密的论文在解密后应遵守此规定）

本学位论文属于（必须在以下相应方框内打“√”，否则一律按“非保密论文”处理）：

- 1、保密论文：☐ 本学位论文属于保密，在_____年解密后适用本授权书。
- 2、非保密论文：☐ 本学位论文不属于保密范围，适用本授权书。

研究生本人签名：_____

签字日期：20____年____月____日

研究生导师签名：_____

签字日期：20____年____月____日

置信规则库系统的优化方法研究

中文摘要

由于人类在决策过程中具有不可替代的作用,因此在工程实际中,综合使用定量信息和专家提供的不完整的或不精确的主观信息,对决策问题进行建模和分析是非常重要的。为有效利用带有各种不确定性的定量信息和定性知识, Yang 等基于 D-S 证据理论、决策理论和模糊理论,在传统 IF-THEN 规则中加入置信框架,提出了置信规则库推理方法(RIMER)。该方法具有对带有含糊或模糊不确定性、不完整性或概率不确定性以及非线性特征的数据进行建模的能力。基于该方法构建的专家系统被称为置信规则库专家系统,为了提升置信规则库专家系统的推理能力, Yang 等提出置信规则库优化模型对置信规则库中的参数进行训练以得到参数的准确取值, Liu 等在置信规则的前提属性部分加入置信框架使置信规则能更好地描述不确定信息,同时提出新的置信规则库系统推理方法,扩展的置信规则库系统未进行参数学习仍能具有较好的推理能力。然而,现有的研究均忽略了置信规则库系统性能表现,无论是参数学习还是扩展置信规则库系统效率和推理准确性方面均不理想,仍需对置信规则库系统方法进行优化。本文的工作内容如下:

(1)针对基于 MATLAB 优化工具箱中 fmincon 函数的参数学习方法运行效率和寻优能力不理想的问题,本文引入粒子群优化算法,提出新的参数训练方法。通过增加重新赋予粒子速度策略以及限制粒子在解空间内,在一定程度上克服粒子群算法易早熟收敛的问题并能求解具有约束条件的最优化问题,最后通过函数拟合和输油管道泄漏检测实例验证该方法的有效性。

(2)针对置信规则库系统尤其是扩展置信规则库系统在进行推理时,对每个输入均需遍历整个规则库计算激活权重,导致系统推理效率不理想的问题,提出了对无序存储的规则建立基于 BK 树的索引方法。系统在推理进行时,可以通过剪枝策略减少搜索规则数量,并对关键规则进行合成,提升扩展置信规则库系统的推理效率和能力,最后利用回归数据和分类数据进行实验验证。

(3)针对规模较大且复杂的置信规则库系统存在推理能力和泛化能力不理想的问题,本文提出基于 AdaBoost 方法训练多个置信规则库系统,然后选择合适方法对多个系统进行组合。最后通过实验验证集成的置信规则库系统具有更好的推理能力。

关键词: 置信规则库, 证据推理, 粒子群算法, BK 树, AdaBoost

Optimization methodology research of belief rule based systems

Abstract

Because humans play an irreplaceable role in the decision-making process, so in engineering practice, integrated use of incomplete or inaccurate quantitative information and subjective information provided by experts on modeling decision problems and analysing issues are very important. In order to effectively use various quantitative and qualitative knowledge with uncertainties, Yang et al. proposed a belief rule base inference methodology using the evidential reasoning (RIMER) approach. RIMER is based on D-S evidence theory, decision theory, fuzzy theory and used a belief structure in the traditional IF-THEN rules. RIMER has the ability to model by vague or fuzzy uncertainty, incomplete or non-linear characteristics and probability uncertainties data. The expert systems using the RIMER approach are called belief rule based systems. In order to enhance the performance of belief rule based systems, Yang et al. proposed optimization models for belief rule based systems to train the parameters of systems to obtain accurate parameters. Liu, et al. used a belief structure in antecedent attributes to enable belief rules better describe uncertainty information and proposed a new inference methodology. The extended belief rule based systems have good performance without parameters learning. However, existing studies ignored the efficiency of belief rule based systems, the accuracy and efficiency are not good enough both in parameters learning and extended belief rule based systems. So belief rule base inference methodology still needs optimizing. The work of this research is as follows:

(1) For the operating efficiency and optimization capability of parameters learning based on fmincon function of MATLAB optimization toolbox is not good, this paper proposes the new parameters training algorithm based on particle swarm optimization algorithm. By proposing the strategy reassigning particle velocity and restricting particles in the solution space. To some extent, this approach overcomes the premature convergence of particle swarm optimization problem and particle swarm optimization can be used in a constraint optimization problem. Finally validating the effectiveness of this approach through function fitting and pipeline leak detection example.

(2) Belief rule based systems especially extended rule based systems need to traverse the entire rule base to calculate the activation weight for each input during inference process. So the efficiency of the system is not perfect, this paper proposes establish index between disorderly stored rules based on BK tree. The number of search rules will be reduced by a pruning strategy during inference process. The reasoning efficiency and ability of belief rule based systems can be enhance by combining key rules. Finally, the experimental verification is carried out using regression data and classification data.

(3) For the reasoning ability and generalization ability of large-scale and complex belief rule based systems is not good, this paper presents an approach based on AdaBoost to train multiple systems and then use appropriate method for combining multiple systems. Finally, experimental verifications show ensemble belief rule based systems have better reasoning ability.

Key words: belief rule base, evidential reasoning, particle swarm optimization, BK-Tree, AdaBoost

目录

中文摘要.....	I
Abstract.....	II
目录.....	IV
第一章 引言.....	1
1.1 课题研究背景.....	1
1.2 国内外研究现状.....	2
1.3 课题研究的目的是和意义.....	3
1.4 课题研究的内容和组织结构.....	4
第二章 置信规则库专家系统推理方法.....	5
2.1 引言.....	5
2.2 置信规则库.....	5
2.2.1 置信规则库的表示.....	5
2.2.2 置信规则库推理方法.....	6
2.2.3 分层置信规则库.....	9
2.3 扩展置信规则库.....	10
2.3.1 扩展置信规则库的表示与构建.....	10
2.3.2 扩展置信规则库推理方法.....	11
2.4 置信规则库系统特点.....	12
2.5 本章结语.....	13
第三章 基于变速 PSO 的置信规则库参数学习方法.....	14
3.1 引言.....	14
3.2 置信规则库系统优化模型.....	15
3.2.1 BRB 优化模型.....	15
3.2.2 优化模型的求解.....	16
3.3 基于变速 PSO 的参数学习方法.....	17
3.3.1 变速粒子群算法.....	17
3.3.2 VPSO 参数学习算法.....	19
3.4 实验结果与分析.....	20
3.4.1 多峰函数实例.....	20
3.4.2 输油管道泄漏检测.....	22
3.5 本章结语.....	25
第四章 基于 BK 树的扩展置信规则库结构优化框架.....	27

4.1 引言.....	27
4.2 扩展置信规则库结构优化框架.....	27
4.2.1 Burkhard-Keller 树.....	28
4.2.2 基于 BK 树的扩展置信规则库构建.....	29
4.2.3 基于 BK 树的扩展置信规则库搜索.....	30
4.2.4 结构优化框架下的规则推理.....	31
4.3 实验分析.....	33
4.3.1 函数拟合问题.....	33
4.3.2 输油管道泄漏问题.....	35
4.3.3 分类数据集测试.....	37
4.4 本章结语.....	38
第五章 基于 AdaBoost 的置信规则库集成学习方法.....	40
5.1 引言.....	40
5.2 集成学习方法.....	40
5.2.1 集成学习的概念.....	40
5.2.2 集成学习器的构建与组合.....	41
5.2.3 问题提出.....	41
5.3 置信规则库集成学习方法.....	42
5.3.1 AdaBoost 算法.....	42
5.3.2 置信规则库系统集成学习.....	43
5.4 实验验证.....	45
5.4.1 实验数据集.....	45
5.4.2 实验结果.....	46
5.5 本章结语.....	47
第六章 总结和展望.....	48
6.1 工作总结.....	48
6.2 工作展望.....	49
参考文献.....	50
致谢.....	54
个人简历.....	55
在学期间的研究成果以及发表的学术论文.....	56

第一章 引言

1.1 课题研究背景

自第一个专家系统问世以来,经过几十年的研究和发展,专家系统理论和技术日趋成熟,已成为最活跃的人工智能应用领域之一,并被广泛应用在医疗检测、化学工程、语音识别、图像处理、金融决策、地质勘探、石油、军事等领域^[1-2]。

专家系统是一类具有人工智能的计算机程序,其内部含有大量的某个领域专家水平的知识与经验,能够利用人类专家的知识 and 进行推理的方法来解决那些需要人类专家处理的复杂决策问题^[3]。传统的专家系统通常仅基于完整的历史数据,而在工程实际中,客观信息(定量信息)和专家提供的不完整或不精确的主观信息(定性信息)在决策问题的建模和分析过程中起着重要作用^[4]。为了有效利用含有各种不确定性的定量信息和定性信息,实现复杂决策问题的建模, Yang 等^[5]提出了基于证据推理算法的置信规则库推理方法(belief rule-base inference methodology using the evidential reasoning approach, RIMER)。置信规则库推理方法具有处理各种不确定性的能力, Yang 等^[6]基于 RIMER 在 Windows 操作系统上开发了图形化置信决策系统,使得 RIMER 在工程实际中得到广泛的应用。目前, RIMER 已用于海面采油系统的安全性分析以及输油管道的泄漏和泄漏大小估计等工程实际中^[7-8]。

在置信规则库(belief rule base, BRB)系统中,包括规则权重、前提属性权重和置信度等参数,为了使 BRB 系统能够准确地反映系统的行为, Yang 等^[9]提出对 BRB 系统的参数进行训练的置信规则库系统优化学习模型,并利用 MATLAB 优化工具箱中的 fmincon 函数进行参数学习从而确定参数的具体取值。当决策问题较为复杂时, Yang 等^[5]提出可以通过构建自下而上的分层 BRB 系统,将复杂问题分解为多个子问题,从而进一步进行解决。但当 BRB 系统规模较大时, fmincon 函数全局搜索能力差并且效率不理想;在进行推理时,需要反复地遍历 BRB,导致 BRB 系统的推理效率差;并且仅凭单个 BRB 系统往往无法准确反映系统的行为。因此,为了使 BRB 系统具有更好的推理效率和推理准确性,有必要对置信规则库系统优化方法进行研究。

本文对置信规则库系统进行深入研究,并优化 BRB 系统。具体包括:结合群智能优化算法,更准确地设置 BRB 系统中参数的取值,提升 BRB 系统推理准确性;构建树形结构的索引,实现对 BRB 中规则的高效搜索,提高 BRB 系统推理效率;基于集成学习思想对多个 BRB 系统进行集成,使得 BRB 系统能更准确地反映系统的行为。

1.2 国内外研究现状

基于 D-S 证据理论^[10-11]、决策理论^[12]、模糊理论^[13]和传统 IF-THEN 规则库^[14]，Yang 等^[5]提出基于证据推理算法的置信规则库推理方法即 RIMER 方法。RIMER 方法具有对带有含糊或模糊不确定性、不完整或概率不确定性以及非线性特征的数据进行建模的能力^[5]。由 RIMER 方法为核心构建的系统通常被称为置信规则库（belief rule base, BRB）专家系统即 BRB 系统，在 BRB 系统中知识是以置信规则的方式进行表达，一系列的置信规则组成置信规则库，而在置信规则库又中含有规则权重、前提属性权重等参数。领域专家根据其经验和历史数据给定参数的取值，当 BRB 系统比较复杂时，专家无法准确给定参数的取值，影响 BRB 系统反映系统行为的能力。为此，Yang 等^[9]提出了用于训练 BRB 系统的优化模型，首先由专家构建初始的 BRB，再利用系统中对应的输入和输出信息，通过 MATLAB 工具箱中的 fmincon 函数进行参数学习，优化参数取值的精确性；在此基础上，Chen 等^[15]对前提属性候选值存在的约束进行分析，改进了 BRB 系统的优化模型，将前提属性候选值作为被训练的参数进行参数学习，并将原优化模型称为局部训练模型，改进后的优化模型称为全局训练模型；常瑞等^[16]提出基于梯度下降法和二分法的联合参数学习方法，与 fmincon 函数相比，该方法的收敛速度有明显的提升；Zhou 等^[17]基于期望极大估计算法提出在线的参数学习方法，为对于时效性有较高要求的复杂决策问题提供了在线建模方法。

Chen 等在文献[18]中指出，BRB 系统的置信规则数量随着前提属性数量的增长呈现指数形式的增长，从而产生“维度灾难”问题。现有研究学者对该问题主要有两种解决方法。一种是对置信规则库的结构进行学习，李彬等^[19]提出置信 K 均值聚类算法，按照预先设定的 K 值聚类得到前提属性候选值；Chang 等^[20]引入特征提取的方法提取关键的前提属性，约减前提属性数量；王应明等^[21]引入粗糙集和密度聚类方法对置信规则库结构进行进一步研究。Liu 等^[22]针对 BRB 系统遇到的“维度灾难”问题，提出了一种由样本数据驱动，表示、构建和推理的扩展置信规则库方法（extended belief rule base, EBRB）。该方法对置信规则的表示进行扩展，在前提属性部分引入分布式置信框架，初始 BRB 的规则根据训练数据生成，置信规则数量与训练数据数量一致，从而避免“维度灾难”问题并且产生规则的方法简单、高效。EBRB 根据分布式前提属性提出相应的推理方法，由数据驱动的初始置信规则库系统无需进行参数学习就能具有不错的推理能力；在此基础上 Alberto 等^[23]针对数据驱动的 EBRB 所遇到的数据不完整和不一致问题，提出一种动态规则激活方法，提升了扩展置信规则系统的鲁棒性和推理能力；余瑞银等^[24]对 EBRB 规则前件表示方式进行改进并完善规则权重计算方法，在计算

激活权重时引入 80/20 法则，提升 EBRB 系统推理准确性和效率。

置信规则库推理方法自诞生以来，引起越来越多研究学者的关注和研究并已被广泛应用于各种工程实际问题之中。Yang 等^[5]最早提出置信规则库推理方法并应用于石墨成分检测实例；Xu 等^[8]、Chen 等^[15]和 Zhou 等^[17]将参数学习方法应用于输油管道泄漏检测实例；Wang 等^[25]和 Yang 等^[26]将多层 BRB 系统应用于消费者行为偏好和产品目标预测；李彬等^[27]提出区间信息处理方法并应用于生产运作管理中；Tang 等^[28]改进前提属性权重赋值方式并应用于新产品开发中消费者感知风险分析；Liu 等^[29]利用置信规则库推理方法求解具有约束条件的非线性现金流组合优化问题；Kong 等^[30]构建基于 RIMER 的置信决策支持系统并应用于心脏性胸痛的临床风险评估；Liu 等^[22]对置信规则表示形式进行扩展将扩展置信规则库推理方法应用于软件缺陷预测；Zhou 等^[31]提出一种双层置信决策支持系统并应用于胃癌淋巴结转移诊断；Zhang 等^[32]针对专家主观标准可能存在不同提出 G-BRB 并应用于海上安全管理绩效评估管理。

1.3 课题研究的目的和意义

本课题的研究目的主要包括以下四个方面：

其一，对 BRB 系统和 EBRB 系统的规则表示、构建方法和推理过程进行研究，总结二者的异同点，分析存在的不足之处；

其二，研究群智能算法，结合 BRB 系统各参数具有的等式和不等式约束条件，提出具有较好效率和寻优能力的参数学习方法，适用于规模较大且复杂的 BRB 系统；

其三，为了提高 BRB 系统的推理效率，研究基于树的索引技术，结合置信规则的表示方式，对置信规则构建树形结构的索引；

其四，研究集成学习思想以及相关方法，构造多个不同的 BRB 系统并合理地组合，为 BRB 系统集成学习提供相关的理论支撑和指导。

本课题的研究意义主要包括：

（1）分析总结 BRB 系统和 EBRB 系统的原理内容，更直观的体现二者的异同之处，对发展基于该系统的相关方法具有一定指导和借鉴意义；

（2）将群智能算法应用于 BRB 系统的参数学习之中，提出具有良好寻优能力和收敛速度的方法，丰富了训练 BRB 系统参数的方法，使训练后的 BRB 系统具有更好的推理能力；

（3）通过对置信规则构建树形索引，使得 BRB 系统进行推理时可以高效地搜索激活规则并进行推理，提升 BRB 系统性能和适用性；

（4）通过集成学习方法 AdaBoost，构建具有互补能力的 BRB 系统并进行适

当整合，提升对复杂决策问题建模和反映系统行为的能力。

(5) 通过对 BRB 系统进行优化，改善规模较大且复杂的 BRB 系统的推理效率、准确性和泛化能力，使得 BRB 系统和分层 BRB 系统可以更好的应用于工程实际中。

1.4 课题研究的内容和组织结构

本文深入研究置信规则库的表示、构建和推理方法，针对现有方法存在的问题提出能够改进 BRB 系统性能的优化方法，研究内容共分为六个章节，具体如下：

第一章为引言，介绍课题——置信规则库系统的优化方法研究的研究背景、国内外研究现状、课题研究的目的并总结本文的研究内容和组织结构。

第二章介绍传统置信规则库推理方法和数据驱动的扩展置信规则库推理方法的理论基础，研究二者在规则的表示、规则库的构建和推理方法上的相同与不同之处，总结置信规则库推理方法的特点，为改进和优化提供了理论支撑。

第三章针对现有置信规则库优化模型求解方法存在的问题，提出基于变速粒子群的置信规则库参数学习方法。针对 BRB 系统中存在的线性和非线性约束条件以及传统粒子群算法求解最优化问题时容易出现早熟收敛的情况，提出了一种变速粒子群算法，在一定程度上克服了上述问题。通过一个多峰函数拟合问题和输油管道泄漏检测实例与现有的方法进行比较，验证所提方法的可行性和有效性。

第四章针对现有的置信规则库系统在进行推理时需要遍历导致效率不理想的问题，提出基于 BK 树的置信规则库优化框架。通过对置信规则建立基于度量距离的树形索引，在搜索激活规则时可以通过剪枝策略进行剪枝，减少搜索规则的规模，同时通过阈值设置只对关键规则进行组合，提升了置信规则库系统的推理效率和推理精度。通过石油管道泄漏检测实例、非线性函数拟合和分类数据构建扩展置信规则库系统进行实验验证。

第五章针对规模较大或复杂的置信规则库系统泛化能力和稳定性不理想的问题，提出基于 AdaBoost 算法的置信规则库系统集成学习方法。该方法应用 AdaBoost 迭代算法使得新的置信规则库系统能关注于之前系统错误分类的数据，并引入差分进化算法训练置信规则库子系统，保证置信规则库子系统的推理能力从而使得 AdaBoost 算法的能发挥作用。通过分类测试数据进行实验，与单个系统进行比较，验证该方法的可行性。

第六章对本研究的内容和成果进行总结，并提出现有工作存在的不足以及对未来的研究方向进行展望。

第二章 置信规则库专家系统推理方法

2.1 引言

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) 是 Yang 等^[5]基于 D-S 证据理论^[10-11]、决策理论^[12]、模糊理论^[13]和传统 IF-THEN 规则库^[14], 所提出的一种能够有效利用具有不确定性的各种类型信息, 对具有线性或非线性关系的输入和输出进行建模的方法。基于 RIMER 方法构建的专家系统即置信规则库 (belief rule base, BRB) 专家系统主要包括两个部分, 分别是对知识的表达以及对知识的推理。在 BRB 系统中, 知识的表达以置信规则库的方式来实现, 而知识的推理则是以证据理论中的证据推理 (evidential reasoning, ER) 算法实现^[33-34]。BRB 系统能够有效利用具有模糊不确定、不完整或概率不确定的信息而且有别于 SVM^[35]、神经网络^[36]等黑箱方法, BRB 系统是一种灰箱, 具有良好的可解释性、可理解性^[37]。

Liu 等^[22]基于 RIMER 方法对置信规则进行扩展, 将分布式置信框架引入前提属性部分, 扩展置信规则对含糊、不完整和不确定的知识具有更优秀、更全面的表示能力, 改进后的规则库被称为扩展置信规则库 (extended belief rule base, EBRB)。Liu 等提出一种简单且有效地利用样例数据生成初始扩展置信规则的方法, 由数据驱动的 EBRB 系统能够充分利用数据信息, 无需进行反复迭代的参数学习就能具有良好的推理能力。Liu 等还对数据驱动所引起的规则库中规则不一致问题进行处理, 提升 EBRB 系统的可解释性。

本章将对 BRB 系统和 EBRB 系统置信规则的表示、置信规则库的构建以及系统推理方法的理论基础进行介绍; 其次分析说明置信规则库推理方法与其他方法相比所具有的特点并对两种系统进行比较, 为对置信规则库系统进行改进和优化提供了理论指导和支撑作用。

2.2 置信规则库

2.2.1 置信规则库的表示

Yang 等^[5]基于传统 IF-THEN 规则, 对其结果部分进行扩展以分布式置信度的形式表示, 同时引入前提属性权重和规则权重, 提出了能够有效表达知识不确定性的置信规则。置信规则库是由一系列的置信规则构成, 第 k 条置信规则的表示形式如下:

$$R_k : \text{if } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k \text{ then } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \dots, (D_N, \beta_N^k)\} \\ \text{with a rule weight } \theta_k \text{ and attribute weights } \delta_1, \delta_2, \dots, \delta_{T_k} \quad \text{公式 (2-1)}$$

其中, $A_i^k (i=1,2,\dots,T_k, k=1,2,\dots,L)$ 表示第 k 条规则中第 i 个前提属性的参考值; T_k 表示第 k 条规则中前提属性的数量; L 表示规则库中置信规则的数量; N 表示评价结果的数量; θ_k 表示第 k 条规则的规则权重, 反映第 k 条规则在置信规则库中相对于其他规则的重要程度; δ_i 表示规则中第 i 个前提属性的权重, 反映规则中第 i 个前提属性相对于其他前提属性的重要度; $\beta_j^k (j=1,2,\dots,N; k=1,2,\dots,L)$ 表示第 k 条规则中第 j 个评价结果 D_j 的置信度, 如果 $\sum_{j=1}^N \beta_j^k = 1$ 则称第 k 条规则是完整的, 否则第 k 条规则是不完整的^[38]。

BRB 系统能够对具有非线性关系的输入和输出进行建模, 假设 BRB 系统的输入为 \mathbf{x} , $\mathbf{x} = (x_1, x_2, \dots, x_{T_k})$ 其中 $x_i (i=1,2,\dots,T_k)$ 表示第 i 个前提属性的输入值; 设 $\mathbf{A}^k = (A_1^k, A_2^k, \dots, A_{T_k}^k)$ 表示第 k 条规则参考值的集合, 设 $\mathbf{D} = (D_1, D_2, \dots, D_N)$ 表示评价结果参考值的集合。将 \mathbf{A}^k 视为 BRB 系统的输入, 评价结果 D_j 上的评价可表示为^[38]:

$$S(\mathbf{A}^k) = \{(D_j, \beta_j^k), j=1,2,\dots,N\} \quad \text{公式 (2-2)}$$

2.2.2 置信规则库推理方法

BRB 系统的规则推理是通过证据推理 (Evidential Reasoning, ER) 算法^[33-34]实现。主要思想是利用 ER 算法合成激活规则, 并由此得到 BRB 系统的最终输出。而在规则合成前还需计算每条规则的激活权重以及修正激活规则评价结果的置信度。其中第 k 条规则激活权重的计算公式为:

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^L \left[\theta_l \prod_{i=1}^{T_l} (\alpha_i^l)^{\bar{\delta}_i} \right]}, \quad \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,2,\dots,T_k} \{\delta_i\}} \quad \text{公式 (2-3)}$$

其中, α_i^k 表示 BRB 系统输入为 \mathbf{x} 时, 第 k 条规则中第 i 个参考值的个体匹配度。假设 BRB 系统第 i 个前提属性对应的输入为 x_i , x_i 为定量的数值型形式。决策者或专家先将 x_i 的参考值 $A_{i,j} (j=1,2,\dots,J_i)$ 与数值量 $\gamma_{i,j}$ 建立如下对应关系:

$$\gamma_{i,j} \text{ means } A_{i,j}, j = 1, 2, \dots, J_i \quad \text{公式 (2-4)}$$

假设专家对较大值的 $\gamma_{i,j}$ 具有更高的偏好程度。设 $\gamma_{i,1}$ 为最小值, γ_{i,J_i} 为最大值, 则输入 x_i 的期望形式表示如下:

$$S(x_i) = \{(\gamma_{i,j}, \alpha_{i,j}), i = 1, 2, \dots, T_k, j = 1, 2, \dots, J_i\} \quad \text{公式 (2-5)}$$

其中, $\alpha_{i,j}$ 的计算公式如下:

$$\alpha_{i,j} = \frac{\gamma_{i,j+1} - x_i}{\gamma_{i,j+1} - \gamma_{i,j}}, \gamma_{i,j} \leq x_i \leq \gamma_{i,j+1}, j = 1, 2, \dots, J_i - 1 \quad \text{公式 (2-6)}$$

$$\alpha_{i,j+1} = 1 - \alpha_{i,j}, \gamma_{i,j} \leq x_i \leq \gamma_{i,j+1}, j = 1, 2, \dots, J_i - 1 \quad \text{公式 (2-7)}$$

$$\alpha_{i,s} = 0, s \neq j, j+1, s = 1, 2, \dots, J_i \quad \text{公式 (2-8)}$$

因为 $\gamma_{i,j}$ 与前提属性参考值对应, 所以 x_i 的期望形式可以转化为以下形式:
 $S(x_i) = \{(A_{i,j}, \alpha_{i,j}), i = 1, 2, \dots, T_k, j = 1, 2, \dots, J_i\}$ 。又因为 $\alpha_i^k \in \{a_{i,1}, a_{i,2}, \dots, a_{i,J_i}\}$, 根据 $\alpha_{i,j}$ 可以得到 α_i^k 。RIMER 方法可以对各种类型的输入信息进行处理, 本研究所进行地实验均基于定量的数值型数据, 文献[5]中详细介绍了对其他形式输入信息的处理方式。

由于输入数据可能不完整, 因此还需对激活规则结果集的置信度作修正, 第 k 条规则的评价结果集中第 i 个评价等级上的置信度修正公式如下:

$$\bar{\beta}_i^k = \beta_i^k \frac{\sum_{t=1}^{T_k} \left(\tau(t, k) \sum_{j=1}^{J_i} \alpha_{t,j} \right)}{\sum_{t=1}^{T_k} \tau(t, k)}, \quad \text{公式 (2-9)}$$

$$\tau(t, k) = \begin{cases} 1 & U_t \in R_k (t = 1, 2, \dots, T_k) \\ 0 & \text{else} \end{cases}$$

其中, U_t 表示第 k 条规则的第 t 个前提属性, J_i 表示第 t 个前提属性参考值的数量, 依据 ER 算法首先需将结果集修正后的置信度 $\bar{\beta}_i^k$ 转换成基本可信值^[33-34]:

$$m_i^k = \omega_k \bar{\beta}_i^k \quad \text{公式 (2-10)}$$

$$m_D^k = 1 - \omega_k \sum_{i=1}^N \bar{\beta}_i^k \quad \text{公式 (2-11)}$$

$$\bar{m}_D^k = 1 - \omega_k \quad \text{公式 (2-12)}$$

$$\tilde{m}_D^k = \omega_k (1 - \sum_{i=1}^N \bar{\beta}_i^k) \quad \text{公式 (2-13)}$$

其中, m_i^k 表示对于第 i 个评价结果的基本可信值; m_D^k 表示对于评价结果参考值集合 $D = (D_1, D_2, \dots, D_N)$ 的基本可信值, 即未分配给任何评价结果 D_i 的基本可信值; m_D^k 由 \bar{m}_D^k 和 \tilde{m}_D^k 两部分组成, \bar{m}_D^k 表示由第 k 条规则的激活权重引起未分配给任何评价结果的基本可信值; \tilde{m}_D^k 则表示由第 k 条规则评价结果的不完整性引起的。若第 k 条规则绝对重要或评价结果是完整的, 则 $m_D^k = 0$ 。

然后, 根据 Dempster 准则对激活规则进行组合, 可以得到评价结果 $D_i (i=1, 2, \dots, N)$ 的基本可信值, 然后再转化为对应的置信度。Dempster 准则是一种迭代组合规则的方法, 令 $m_i^{(1)} = m_i^1$ 和 $m_D^{(1)} = m_D^1$, 对 L 条激活规则进行组合的具体公式如下:

$$m_i^{I(k+1)} = K^{I(k+1)} [m_i^{I(k)} m_i^{k+1} + m_i^{I(k)} m_D^{k+1} + m_D^{I(k)} m_i^{k+1}] \quad \text{公式 (2-14)}$$

$$m_D^{I(k)} = \bar{m}_D^{I(k)} \tilde{m}_D^k \quad \text{公式 (2-15)}$$

$$m_D^{I(k+1)} = K^{I(k+1)} [\tilde{m}_D^{I(k)} \tilde{m}_D^{k+1} + \tilde{m}_D^{I(k)} \bar{m}_D^{k+1} + \bar{m}_D^{I(k)} \tilde{m}_D^{k+1}] \quad \text{公式 (2-16)}$$

$$m_D^{I(k+1)} = K^{I(k+1)} [\tilde{m}_D^{I(k)} \tilde{m}_D^{k+1}] \quad \text{公式 (2-17)}$$

$$K^{I(k+1)} = [1 - \sum_{i=1}^N \sum_{\substack{t=1 \\ t \neq i}}^N m_i^{I(k)} m_i^{k+1}]^{-1}, k=1, 2, \dots, L-1 \quad \text{公式 (2-18)}$$

$$\hat{\beta}_i = \frac{m_i^{I(L)}}{1 - \bar{m}_D^{I(L)}}, i=1, 2, \dots, N \quad \text{公式 (2-19)}$$

$$\hat{\beta}_D = \frac{\tilde{m}_D^{I(L)}}{1 - \bar{m}_D^{I(L)}} \quad \text{公式 (2-20)}$$

其中, $\hat{\beta}_i$ 表示评价结果 D_i 对应的置信度, $\hat{\beta}_D$ 表示未设置给任意评价结果 D_i 的置信度。Wang 等基于上述的 ER 迭代算法提出 ER 解析算法, 详见文献[39]。

假设专家对评价结果 D_i 给定一个效用 $\mu(D_i)$, BRB 系统的输出 $S(x)$ 的期望效

用形式表示如下：

$$\mu(S(x)) = \sum_{i=1}^N \mu(D_i) \hat{\beta}_i \quad \text{公式 (2-21)}$$

其中， $\hat{\beta}_i$ 表示输出相对评价结果 D_i 的似然下限值，似然上限值为 $\hat{\beta}_i + \hat{\beta}_D$ 。

当评价不完整或者不精确时，Yang 等^[5]对最大效用、最小效用和平均效用定义如下：

$$\mu_{\max}(S(x)) = \sum_{i=1}^{N-1} \mu(D_i) \hat{\beta}_i + \mu(D_N)(\hat{\beta}_N + \hat{\beta}_D) \quad \text{公式 (2-22)}$$

$$\mu_{\min}(S(x)) = \mu(D_1)(\hat{\beta}_1 + \hat{\beta}_D) + \sum_{i=2}^N \mu(D_i) \hat{\beta}_i \quad \text{公式 (2-23)}$$

$$\mu_{\text{avg}}(S(x)) = \frac{\mu_{\max}(S(x)) + \mu_{\min}(S(x))}{2} \quad \text{公式 (2-24)}$$

2.2.3 分层置信规则库

面对复杂的决策问题，Yang 等^[5]基于自下而上的思想提出了分层 BRB 系统。其主要思想是，构建多层次的 BRB 系统，先对底层的系统进行推理，然后把底层系统的推理结果作为上层系统的输入。

文献[5]中给出了一个分层 BRB 系统实例，如图 2-1 所示。三个 BRB 子系统构成一个分层 BRB 系统，BRB 子系统分别为 BRB 子系统 1，BRB 子系统 2，和 BRB 子系统 3。其中，BRB 子系统 1 和 BRB 子系统 2 为底层 BRB 系统且相互独立，BRB 子系统 1 的输出和 BRB 子系统 2 的输出作为 BRB 子系统 3 的输入。多层 BRB 系统在进行推理时，自下而上依次进行推理。

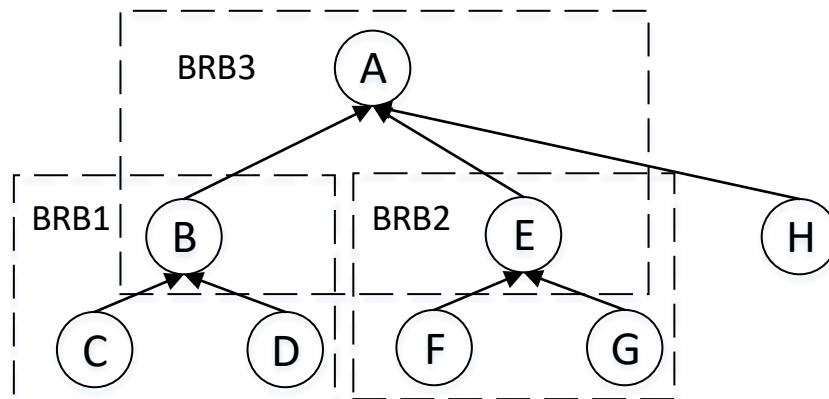


图 2-1 分层 BRB 系统

2.3 扩展置信规则库

2.3.1 扩展置信规则库的表示与构建

为了使置信规则能够对数据或知识中存在的 uncertainty 及不完整性时更加准确和全面地表示。Liu 等^[22]提出在规则的 IF 部分也引入分布式置信度框架，扩展置信规则库（EBRB）由一系列的扩展置信规则构成，第 k 条扩展置信规则的形式如下：

$$R_k: \text{if } \{A, \alpha^k\} \text{ then } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \dots, (D_N, \beta_N^k)\} \quad \text{公式 (2-25)}$$

其中， (A, α^k) 是以分布式置信度的形式进行表示的，EBRB 规则前件部分也可表示为 $\{(A_{i,j}, \alpha_{i,j}^k), j=1,2,\dots,J_i \mid i=1,2,\dots,T\}$ ， $A_{i,j}$ 表示第 i 个前提属性的第 j 个参考值，且参考值数量为 J_i ， T 表示前提属性的数量。

有别于 BRB 系统通常由领域专家根据经验构建初始 BRB，Liu 等^[22]提出一种简单高效的由数据驱动的构建 EBRB 系统的方法。假设第 k 个训练数据中第 i 个前提属性为定量数据以 x_i^k 表示，首先由专家或决策者将参考值 $A_{i,j} (j=1,2,\dots,J_i)$ 与数值量 $\gamma_{i,j}$ 建立起对应关系，假设专家对参考值的偏好程度满足 $\gamma_{i,j+1} > \gamma_{i,j}$ ，那么 x_i^k 可以等价地转换为分布式置信分布的期望形式：

$$E(x_i^k) = \{(\gamma_{i,j}, \alpha_{i,j}), i=1,2,\dots,T, j=1,2,\dots,J_i\} \quad \text{公式 (2-26)}$$

其中 $\alpha_{i,j}$ 的计算方法如下：

$$\alpha_{i,j} = \frac{\gamma_{i,j+1} - x_i^k}{\gamma_{i,j+1} - \gamma_{i,j}}, \gamma_{i,j} \leq x_i^k \leq \gamma_{i,j+1}, j=1,2,\dots,J_i-1 \quad \text{公式 (2-27)}$$

$$\alpha_{i,j+1} = 1 - \alpha_{i,j}, \gamma_{i,j} \leq x_i^k \leq \gamma_{i,j+1}, j=1,2,\dots,J_i-1 \quad \text{公式 (2-28)}$$

$$\alpha_{i,s} = 0, s \neq j, j+1, s=1,2,\dots,J_i \quad \text{公式 (2-29)}$$

通过公式 (2-27)、(2-28) 和 (2-29) 可以得到 $\alpha_{i,j}$ 的具体取值从而生成扩展置信规则的前件部分，与输入 x_i^k 相对应的输出 y^k 可采用同样的方法产生评价

结果的分布式置信分布形式。

扩展的置信规则能够表示更多的信息,更符合实际情况具有更好的灵活性和适应性。

2.3.2 扩展置信规则库推理方法

EBRB 系统的规则推理也是通过证据推理算法实现。第一步根据系统输入计算每条规则的激活权重,第二步通过 ER 方法组合激活规则。Liu 等^[22]基于扩展置信规则的表示形式提出一种新的计算规则激活权重的方法。因为 EBRB 规则的前件部分是置信分布的形式,所以 Liu 等也将输入数据 x_i 转化成置信分布的形式,然后用欧式距离度量第 i 个前提属性对应的输入数据 x_i 与其参考值的个体匹配度,具体公式如下:

$$S^k(x_i, A_i^k) = S_i^k = 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \quad \text{公式 (2-30)}$$

在计算得到输入与每个前提属性的个体匹配度之后就可以计算激活权重,具体公式如下:

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} (S^k(x_i, A_i))^{\bar{\delta}_i}}{\sum_{l=1}^L \left[\theta_l \prod_{i=1}^{T_l} (S^l(x_i, A_i))^{\bar{\delta}_i} \right]}, \quad \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,2,\dots,T_k} \{\delta_i\}} \quad \text{公式 (2-31)}$$

其中规则权重和前提属性权重是领域专家根据自身经验给定的。扩展置信规则是由数据生成,因此容易发生规则不一致的问题,所谓规则不一致是指两条规则具有相似的前件部分,但评价结果却大相径庭或者规则与专家知识相冲突^[5, 22]。可以发现规则不一致性与前提属性和评价结果有关,而二者在 EBRB 中均为置信度分布的形式,因此 Liu 等以类似公式 (2-27)、(2-28) 和 (2-29) 的方式计算规则间前提属性相似度 SRA 和评价结果相识度 SRC, 计算规则 p 和规则 q 的 SRA 和 SRC 具体方式定义如下:

$$\text{SRA}(R_p, R_q) = \min_i^T \left(1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j}^p - \alpha_{i,j}^q)^2} \right) \quad \text{公式 (2-32)}$$

$$\text{SRC}(R_p, R_q) = 1 - \sqrt{\sum_{i=1}^N (\beta_i^p - \beta_i^q)^2} \quad \text{公式 (2-33)}$$

根据文献[40]，规则 p 和规则 q 一致性度量方式如下：

$$\text{Cons}(R_p, R_q) = \exp\left(-\frac{(\text{SRA}(R_p, R_q)/\text{SRC}(R_p, R_q) - 1.0)^2}{(1/\text{SRA}(R_p, R_q))^2}\right) \quad \text{公式 (2-34)}$$

度量第 i 条规则和规则库中其他规则的不一致程度方式定义如下：

$$\text{Incons}(i) = \sum_{k=1, k \neq i}^L [1.0 - \text{Cons}(R_i, R_k)] \quad \text{公式 (2-35)}$$

考虑规则不一致性，Liu 等^[22]对规则权重进行修正，具体公式如下：

$$\theta_k^* = \theta_k + \lambda(1.0 - \text{Incons}(k) / \xi_{\text{Incons}}) \quad \text{公式 (2-36)}$$

$$\xi_{\text{Incons}} = \sum_{i=1}^L \text{Incons}(i) \quad \text{公式 (2-37)}$$

其中 λ 为影响因子，控制规则一致性水平。通过公式 (2-36) 对规则权重进行修正，增加一致性高的规则的权重使其在系统推理中发挥更大的作用，减少不一致规则的权重，减少这些规则对系统推理造成的影响。

表 2-1 中对 Yang 和 Liu 提出的 BRB 系统进行了简单比较。

表 2-1 Yang-BRB 系统与 Liu-EBRB 系统的比较

BRB 类型	Yang-BRB 系统	Liu-EBRB 系统
规则的产生	根据专家知识	根据对应的输入输出数据和专家知识
规则的表达	基于传统 IF-THEN 规则，在评价结果引入分布式置信度框架	基于传统 IF-THEN 规则，在前提属性和评价结果引入分布式置信度框架
规则的激活权重计算	根据前提属性参考值计算激活权重	根据前提属性参考值分布式置信度计算激活权重
规则的推理	ER 方法	ER 方法

2.4 置信规则库系统特点

置信规则库推理方法是 BRB 系统的核心，而 RIMER 方法是基于 D-S 证据理论

^[10-11]、决策理论^[12]、模糊理论^[13]和传统 IF-THEN 规则库^[14]，因此继承了这些方法的优点，置信规则库系统具有以下特点^[38]：

(1) 处理不确定数据的能力

置信规则引入置信框架，很好地对不确定信息进行表示，在进行推理时应用 ER 方法使得不确定性能够保留，从而在推理结果中含有更多的信息。

(2) 可解释的能力

置信规则基于传统 IF-THEN 的规则表示形式，根据输入数据进行分析推理，该过程与人类思考分析问题的过程相一致。诸如前提属性权重，规则权重等参数有明确的意义便于理解，使得人们可以对 BRB 系统推理过程和推理结果进行解释和说明，具有良好的可解释性。

(3) 可学习的能力

BRB 系统由规则库和推理机两部分组成，二者彼此独立。专家可以根据自身经验和历史数据对规则库中错误的置信规则进行修正，也可以根据需要扩充规则库。此外，BRB 系统还可以利用对应的输入输出数据进行监督学习，增强系统表现。

2.5 本章结语

本章主要介绍了置信规则库推理方法和扩展置信规则库推理方法的理论知识。主要包括置信规则如何对知识进行表示、如何构建置信规则库以及如何进行规则的推理，并对两种方法进行比较。分析说明了置信规则库系统具有处理不确定信息的能力、良好的可解释性和可学习能力。

然而现有方法在对较大数据规模或复杂的问题进行建模时仍存在问题，导致置信规则库系统的推理准确性、泛化能力和推理效率不理想，这些问题将在接下来的几个章节中进行具体说明并提出改进优化的方法。

第三章 基于变速 PSO 的置信规则库参数学习方法

3.1 引言

在 BRB 系统中, 置信规则的权重、前提属性权重和评价结果置信度等参数通常是凭借专家的领域知识和经验人为给定, 这些参数与 BRB 系统的推理能力息息相关, 规则权重、前提属性权重的微小变化可能会引起 BRB 系统推理表现巨大的变化, 参数取值的准确与否将影响 BRB 系统推理的准确性。但在复杂决策问题或规模较大的 BRB 系统中, 人为给定的方式难以保证参数取值的精确性, 导致 BRB 系统无法准确反映系统的行为。为解决这一问题, 提升 BRB 系统的推理能力, Yang 等^[9]提出了对 BRB 系统进行参数训练的优化模型, 并将参数学习过程转化为求解最优化模型问题, 通过 MATLAB 中 `fmincon` 函数进行求解。随后, Chen 等^[15]将前提属性候选值也作为被训练的参数进行参数学习, 并将原始参数学习方法应用在改进的优化模型中, 由于新模型下的参数学习方法对前提属性候选值进行训练, 提高了 BRB 系统的推理能力, 因此称其为 BRB 系统全局训练方法, 但新优化模型的求解方式依旧是使用 `fmincon` 函数, 当 BRB 规则数较多, 规模较大时, `fmincon` 函数的效率和寻找最优解的能力并不理想。于此同时, 有别于使用 `fmincon` 的参数学习方法相继被提出, 常瑞等^[16]基于梯度下降法和二分法提出联合的参数学习算法, 其收敛效率虽然明显的提升, 但由于算法中涉及复杂的公式推导, 因此该参数学习方法仅包含规则权重和置信度两个参数; Zhou 等^[17]依据期望极大估计法提出在线的参数学习方法, 但算法中不仅涉及复杂的公式推导, 同时算法有效性还需建立在人为假定的概率分布上, 不利于理解与实际应用。

为优化完善现有参数学习方法, 本章提出一种基于变速粒子群算法的置信规则库参数学习方法, 该算法的本质是利用粒子群算法求解最优化问题, 但传统的粒子群算法易早熟收敛, 陷入局部最优解。因此本章提出对失去飞行速度的粒子重新赋予速度的策略, 提升粒子群算法求解优化模型时最优解的精度。在实验分析中, 本章首先通过拟合多峰函数的实验验证改进的粒子群算法是否能有效避免传统粒子群算法遇到的早熟收敛问题。然后在输油管道检漏的实例中, 分析本章方法与现有参数学习方法, 对比各个参数学习方法的收敛精度和收敛时间。以下是本章的内容安排: 首先阐述 BRB 专家系统优化模型的理论知识, 并分析现有参数学习方法存在的问题; 然后介绍变速粒子群算法的原理和基于变速粒子群算法的参数学习方法; 接着引入多峰函数和输油管道检漏的仿真实例, 分析本章所提方法的有效性; 最后对本章的研究内容进行总结。

3.2 置信规则库系统优化模型

3.2.1 BRB 优化模型

在 BRB 专家系统中，系统的推理准确性受规则权重、前提属性权重和评价结果置信度等参数影响，这些参数的初值通常由专家根据个人经验或主观判断给定。因此当 BRB 系统过于复杂，参数数量较多时，人为方式难以准确地设定参数的取值。鉴此，Yang 等^[9]提出了对 BRB 系统的参数进行训练的优化模型，其过程如下图所示：

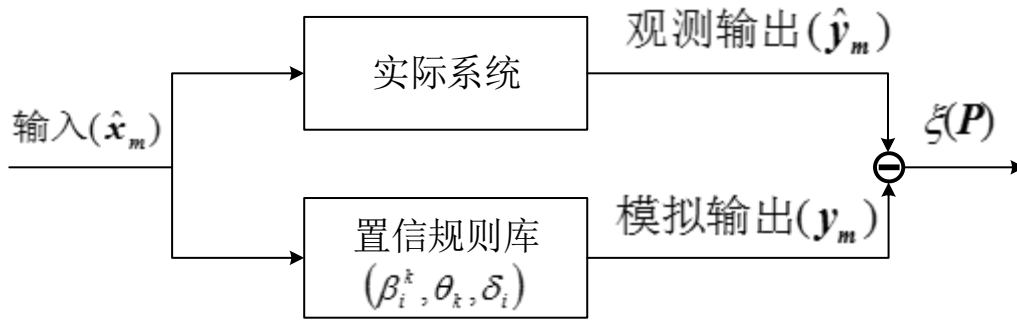


图 3-1 置信规则库优化模型

对于 BRB 的参数优化模型，其符号表达式可表示如下：

$$\begin{aligned} & \min \{\xi(\mathbf{P})\} \\ & s.t. A(\mathbf{P})=0, B(\mathbf{P})\geq 0 \end{aligned} \quad \text{公式 (3-1)}$$

其中， \mathbf{P} 表示由 $(\beta_j^k, \theta_k, \delta_i)$ 组成的参数向量， $\xi(\mathbf{P})$ 表示目标函数； $A(\mathbf{P})$ 表示等式约束条件； $B(\mathbf{P})$ 表示不等式约束条件。当 BRB 系统输入为 \hat{x}_m 时，通过目标函数比较 BRB 系统的模拟输出 y_m 与实际系统的观察输出 \hat{y}_m ，若目标函数的输出值接近 0 时，则说明 BRB 系统具有较好的推理准确性，其中目标函数可用平均绝对误差（Mean Absolute Error, MAE）表示，如公式（3-2）所示。

$$\xi(\mathbf{P}) = (1/M) \sum_{m=1}^M |y_m - \hat{y}_m| \quad \text{公式 (3-2)}$$

对于 BRB 优化模型中等式和不等式约束条件，Yang 等在文献[9]中给出如下的定义：

（1）任意一条规则中每个评价结果上的置信度均不得小于 0 或者大于 1，即：

$$0 \leq \beta_j^k \leq 1; j=1,2,\dots,N; k=1,2,\dots,L \quad \text{公式 (3-3)}$$

(2)若第 k 条规则是完整的,则该规则的所有评价结果的置信度之和等于 1,否则置信度之和小于 1,即:

$$\sum_{j=1}^N \beta_j^k \leq 1; j=1,2,\dots,N; k=1,2,\dots,L \quad \text{公式 (3-4)}$$

(3)规则权重需要归一化,归一化后规则权重的取值应在 0 到 1 之间,即:

$$0 \leq \theta_k \leq 1; k=1,2,\dots,L \quad \text{公式 (3-5)}$$

(4)前提属性权重需要归一化,归一化后前提属性权重的取值应在 0 到 1 之间,即:

$$0 \leq \delta_i \leq 1; i=1,2,\dots,T \quad \text{公式 (3-6)}$$

(5)偏好程度越高的评价结果,应具有越高的等级效用,即:

$$u(D_i) < u(D_j); i < j; i, j=1,2,\dots,N \quad \text{公式 (3-7)}$$

在此基础上,Chen 等对 Yang 等的 BRB 优化模型进行了改进,提出将前提属性的参考值也作为被训练的参数,当前提属性的参考值为定量数据类型时,需满足同一前提属性的相邻参考值差值不大于无穷小量 E_i^k , 即:

$$A_{i,j}^k - A_{i,j+1}^k \leq E_i^k; i=1,2,\dots,T; j=1,2,\dots,N-1 \quad \text{公式 (3-8)}$$

改进后的优化模型中待训练的参数向量表示如下:

$$P = P(\beta_j^k, \theta_k, \delta_i, A_{i,j}^k) \quad \text{公式 (3-9)}$$

3.2.2 优化模型的求解

Yang 等^[9]将 BRB 优化模型转化为一个最优化求解问题,并将这一过程称为参数学习。BRB 优化模型求解问题的本质就是一个具有等式和不等式约束条件的非线性最优化求解问题,传统的可用方法有^[41]:近似牛顿法、zoutendijk 可行方向法和 wolfe 简约梯度法等,但上述的求解方法均具有一定的困难和局限性。鉴此,Yang 等在文献[9]中提出利用 MATLAB 优化工具箱中的 fmincon 函数求解 BRB 的优化模型,进行参数学习。该方法有效地解决了 BRB 系统因参数初值不精确而引起的推理准确性不高的问题,但通过理论研究和实验分析不难发现基于 fmincon 函数的方法在求解复杂问题时,算法效率低且收敛速度慢,并且该方法受制于 MATLAB 导致程序可移植性较差。随着专家学者对参数学习研究的加深,有别于使用 fmincon 函数的参数学习方法被提出,常瑞等^[16]基于梯度下降法和二分法提出联合的参数学习算法,该算法的收敛速度较 fmincon 函数而言明显地提升,但由于算法中涉及复杂的公式推导,且在使用编程语言模拟偏导公式时具有一定的难度,因此该参数学习方法仅包含规则权重和置信度两个参数,相比于现有的 BRB 优化模型,其是一种退化的参数学习方法。

针对当前求解 BRB 优化模型中缺少高效、可行的方法,文章引入群智能算法中的粒子群算法^[42]进行参数学习。由于粒子群算法中涉及的参数较少,且易于实现,因此广泛应用于最优化问题求解。对于传统的粒子群算法,其在求解诸如 BRB 优化模型等复杂的非线性优化问题时,易早熟收敛而陷入局部最优解,主要原因是由于在算法迭代的后期粒子的飞行速度趋近于 0,失去搜索解空间寻找更优解的能力,最终导致无法达到理想的收敛精度。因此本章首先提出通过增加适当的策略使得能够让失去飞行速度的粒子跳出局部最优解从而获得全局范围内收敛精度更高的最优解,随后基于改进后的粒子群算法提出新的 BRB 参数学习的方法。

3.3 基于变速 PSO 的参数学习方法

3.3.1 变速粒子群算法

粒子群算法 (Particle Swarm Optimization, PSO)^[42]最初是受到鸟群捕食行为的启发而提出的算法。传统粒子群算法中,每个粒子都有一个由目标决定的适应值函数。对于一个规模为 N , 搜索空间维数为 D 的粒子群,第 i 个粒子的位置表示为 $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,D})$, 其飞行速度表示为 $\mathbf{v}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,D})$, 粒子根据自身保存的最优位置 $\mathbf{p}_i = (p_{i,1}, p_{i,2}, \dots, p_{i,D})$ 和群体中保存的最优位置 $\mathbf{p}_g = (p_{g,1}, p_{g,2}, \dots, p_{g,D})$ 更新飞行速度。其中速度和位置的更新公式如下所示:

$$v_{i,d}(t+1) = wv_{i,d}(t) + c_1r_1[p_{i,d} - x_{i,d}(t)] + c_2r_2[p_{g,d} - x_{i,d}(t)] \quad \text{公式 (3-10)}$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1) \quad \text{公式 (3-11)}$$

公式 (3-10)、(3-11) 中 $i=1,2,\dots,N$, $d=1,2,\dots,D$; w 表示惯性因子; c_1 和 c_2 表示学习因子; r_1 和 r_2 为 $[0, 1]$ 范围内的均匀随机数。在上述粒子群算法的基础上, 针对具有约束条件的最优化求解问题可以通过惩罚函数或将粒子的位置限制在约束条件的搜索空间里^[43]以保证解为可行解, 便可让传统粒子群算法具有求解具有等式和不等式约束条件的非线性优化问题的能力。为了便于理解, 以下称能够求解具有等式和不等式约束条件优化问题的粒子群算法为 PSO 算法。

PSO 算法在求解复杂优化问题时, 容易出现早熟收敛导致寻优能力不理想, 无法获得全局最优解的问题, 其主要原因是 PSO 算法中各个粒子是根据局部最优粒子的位置和全局最优粒子的位置决定当前粒子的飞行速度, 进而计算得到新的搜索位置。在不断搜索解空间的过程中, 粒子会不断地向最优解位置聚集, 导致种群失去多样性进而引起飞行速度逐渐趋近于 0, 即无法对自身位置进行调整, 造成获得的解通常为局部最优解。为解决该问题, 本章提出一种改进的 PSO 算法, 以下称为变速粒子群算法 (Variable Particle Swarm Optimization, VPSO)。

算法改进的实质针对寻优后期发生的粒子失去飞行速度的问题, 引入适当的策略让失去多样性的粒子重新获得飞行的速度, 以确保粒子能够重新调整自身位置, 进而跳出局部最优解搜索解空间中更理想的更优解。其中对粒子速度重新赋值的策略为首先对解空间的每一维度随机生成一个 $[0, 1]$ 范围内的数, 然后将生成的数与重新赋予速度概率进行比较, 若符合条件则在速度维度范围内生成一个均匀分布的新速度, 否则速度保持不变, 具体公式如下:

$$v_{i,d}(t+1) = \begin{cases} f(v) & \text{if } v_{i,d}(t+1) < eps \text{ and } rand[0,1] < P_v \\ v_{i,d}(t+1) & \text{else} \end{cases} \quad \text{公式 (3-12)}$$

$$f(v) = \frac{v_{\min}^d + v_{\max}^d}{2} + (v_{\max}^d - v_{\min}^d) * (rand[0,1] - 1) \quad \text{公式 (3-13)}$$

其中, v_{\min}^d 和 v_{\max}^d 分别表示解空间第 d 个维度上速度取值的下界和上界, P_v 是重新赋予速度概率 (P_v 取值为 0.9)。同时为了使算法能够应用于复杂的 BRB 系统, 采用限制粒子在搜索空间里的方式以求解具有约束条件的最优化问题。

VPSO 算法先根据公式 (3-10) 更新粒子速度, 其次根据公式 (3-12)、(3-13) 对更新后的速度进行处理, 最后根据公式 (3-11) 更新粒子位置, VPSO 算法的流程图如下:

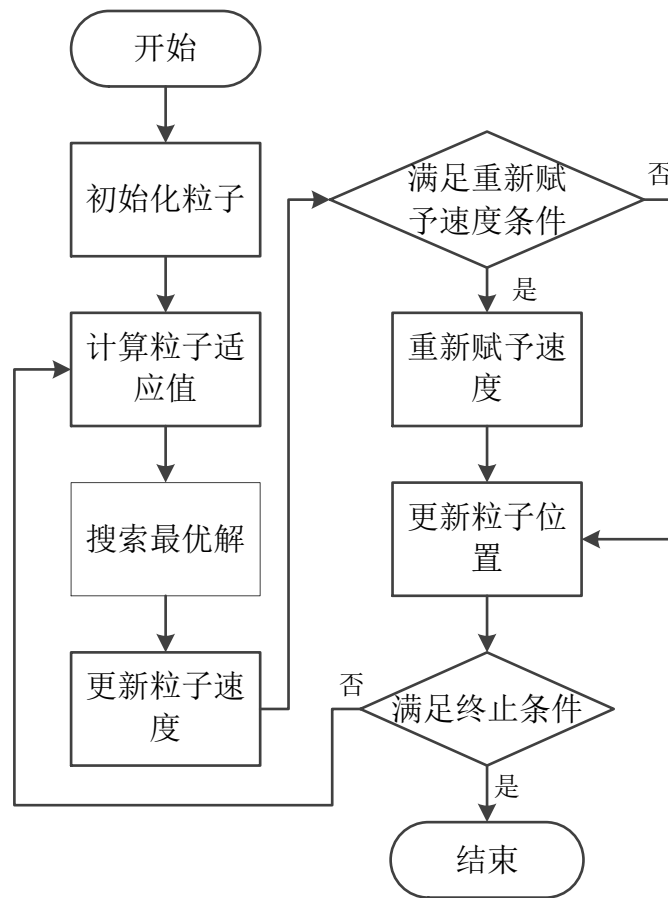


图 3-2 VPSO 算法的流程图

3.3.2 VPSO 参数学习算法

基于 VPSO 算法的 BRB 参数学习算法的步骤如下：

步骤一 初始化粒子群；在约束条件的范围内对粒子的位置和速度取随机初值，将其中一个粒子的位置设置为初始 BRB 训练的参数值，根据初始后的粒子求出局部最优解 pBest 和全局最优解 gBest。

步骤二 计算粒子适应值；通过目标函数依次计算每个粒子的适应值，其中目标函数可选取 BRB 系统的平均方差或平均绝对误差。

步骤三 搜索局部和全局的最优解；若某一粒子的适应值小于其本身所记录的局部最优解 pBest 的适应值，则用当前粒子的信息更新局部最优解 pBest，否则不做任何处理。若该粒子的适应值小于整个种群目前所记录的最优解 gBest 的适应值，则用当前粒子信息更新全局最优解 gBest，否则不做任何处理。

步骤四 更新粒子位置和速度；根据公式 (3-10)、公式 (3-11)、公式 (3-12) 和公式 (3-13) 计算粒子的新位置和速度。更新过程中，粒子各维度上的速度与位置必须根据约束条件进行调整，保证所有的粒子都在可行的搜索空间里。

其中，为使粒子拥有更好的搜索解空间的能力，惯性因子 w 由以下线性公式决定^[44]：

$$w = w_{\max} - \frac{(w_{\max} - w_{\min})}{IterTime} \times IterTime_n \quad \text{公式 (3-14)}$$

w_{\max} 和 w_{\min} 分别代表的最大惯性权重和最小惯性权重 $0 \leq w_{\min} \leq w_{\max} \leq 1$ ， $IterTime_n$ 为当前迭代次数， $IterTime$ 为预设的迭代次数。随着寻优动态变化的 w 能够保证粒子在迭代初期具有良好的搜索范围，容易发现新解，而在迭代后期获得更好的局部搜索能力。

步骤五 为克服传统粒子群早熟收敛的问题，当粒子某一维度上的速度小于预定阈值 eps 时，使用持种群中粒子多样性的策略调整粒子的速度，否则保持不变。

步骤六 若当前全局最优解满足收敛条件或迭代次数达到预设值，全局最优解 $gBest$ 即为所求的最优解，将 $gBest$ 的位置赋值给对应的 BRB 参数即得到训练后 BRB，则算法结束，否则执行步骤二。

3.4 实验结果与分析

为验证本章方法能够求解 BRB 优化模型获得较为准确的参数取值，提高 BRB 系统的推理准确性，本节引入多峰函数和输油管道泄漏检测两个实例。在以下实例分析中，收敛条件均设为迭代 10000 次。此外，实验环境为：Pentium(R) Dual-Core CPU E5400 @ 2.70GHz 2.69GHz；2GB 内存；Windows 7 操作系统；算法实现平台 Visual C++ 6.0。

3.4.1 多峰函数实例

Himmelblau 函数常被用于测试优化算法性能^[15]，因此本节将其作为对比基于 VPSO 参数学习方法和基于 PSO 参数学习方法的实例。其中 Himmelblau 函数定义如下：

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2, -6 \leq x, y \leq 6 \quad \text{公式 (3-15)}$$

构建拟合 Himmelblau 函数的 BRB 系统时，选择 x 和 y 作为前提属性，并假设两个前提属性具有 7 个相同的参考值 $\{-6, -4, -2, 0, 2, 4, 6\}$ ，评级结果集有 5 个评

价等级，且相对应的等级效用依次为 $\{0, 200, 500, 1000, 2200\}$ 。根据参考值生成 49 条置信规则，并以此组成初始的 BRB 系统。由初始 BRB 系统得到的函数拟合曲线图形如图 3-3 所示， x 表示在函数空间内均匀选取的点的编号。

图 3-3 中初始 BRB 系统对 Himmelblau 函数的拟合效果并不理想，因此使用本节方法对初始 BRB 系统进行参数学习。在参数学习中，设置粒子群粒子个数为 100，在 Himmelblau 函数中选取 169 个均匀分布的数据作为训练数据。训练后 BRB 系统拟合曲线的图形如图 3-4 所示。

由图 3-4 可以发现，通过本节方法对初始 BRB 进行参数学习，训练后的 BRB 系统具有更好的拟合复杂非线性多峰函数的能力。进一步分析本节方法 VPSO 参数学习方法的有效性，以基于 PSO 的参数学习方法作为比较对象，分别对用于拟合 Himmelblau 函数的初始 BRB 系统的参数进行训练，参数学习过程中 BRB 系统的模拟输出与函数的实际输出之间的 MAE 变化曲线如图 3-5 所示。基于 PSO 参数学习方法虽然迭代到 4000 多次时已趋于收敛，但可以发现其收敛的精度并不理想，而基于 VPSO 参数学习方法迭代到 8000 多次时收敛，其收敛的精度明显小于基于 PSO 的参数学习方法，综上可知基于 VPSO 参数学习方法能够克服早熟收敛的问题，更有效的确定 BRB 系统的参数取值，提升 BRB 系统的推理能力。

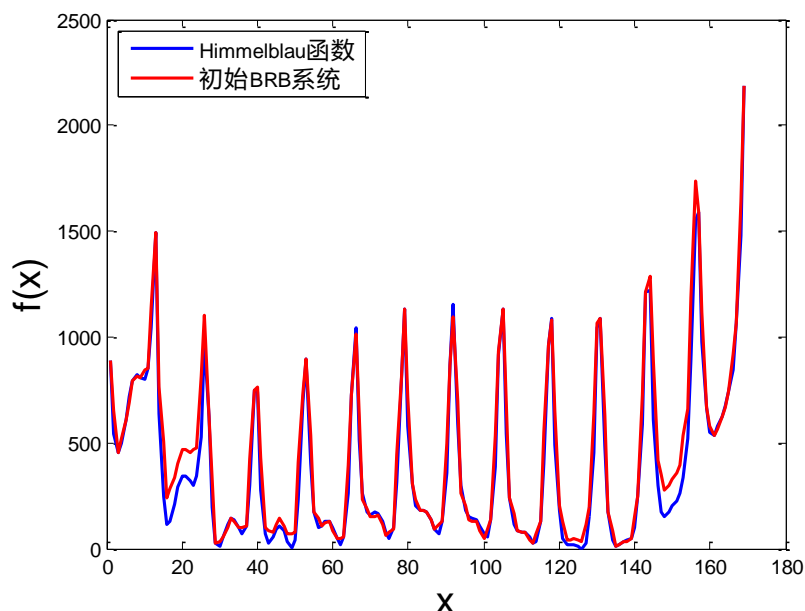


图 3-3 初始 BRB 系统对 Himmelblau 函数拟合效果

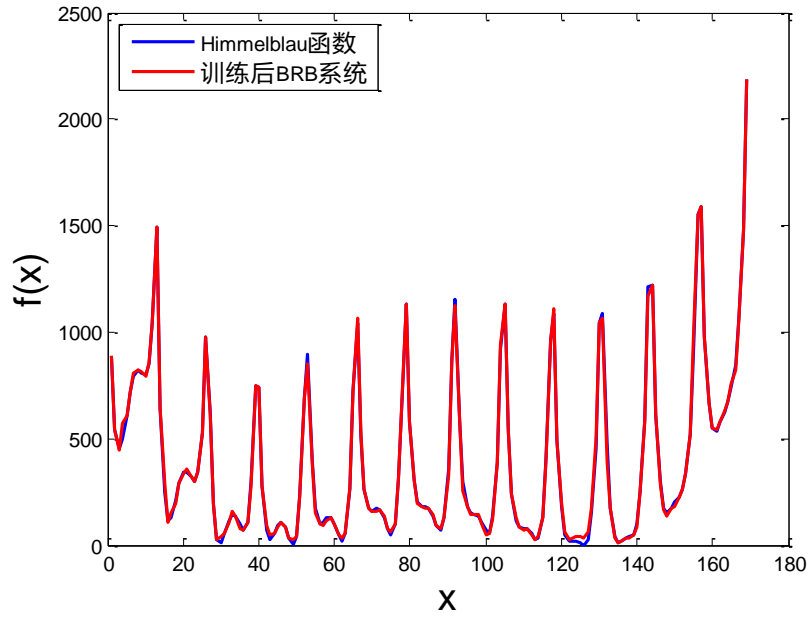


图 3-4 训练后 BRB 系统对 Himmelblau 函数拟合效果

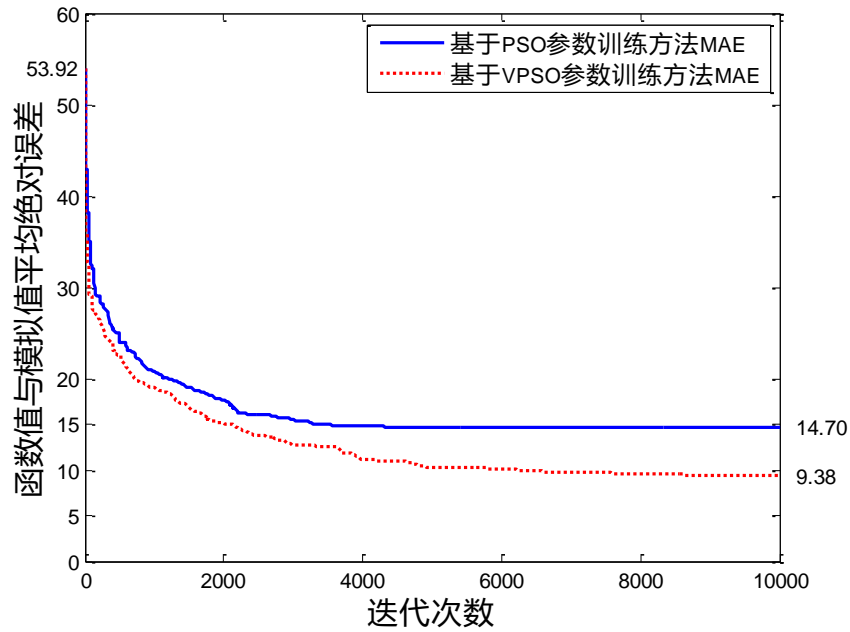


图 3-5 MAE 的变化曲线

3.4.2 输油管道泄漏检测

输油管道泄漏检测的实例主要用于检测 BRB 系统参数学习方法的有效性^[8, 15, 17]，因此将其作为对比本章方法与现有其他参数学习方法的实例。在该实例中，输油管道的漏洞大小 (Leak Size, LS) 主要根据两个因素判断，分别是输入和输出流量差 (Flow Difference, FD) 以及油液对管道产生的平均压力差

(Pressure Difference, PD)。因此将 FD 和 PD 作为 BRB 系统的输入，而 LS 作为输出。此外，选取 2008 组从正常情况到发生 25% 泄漏情况的数据作为测试数据。

对于初始 BRB 的构建，以 FD 和 PD 作为前提属性，并假设 FD 有 8 个参考值，分别为 $\{-10, -5, -3, -1, 0, 1, 2, 3\}$ ；PD 有 7 个参考值，分别为 $\{-0.042, -0.025, -0.01, 0, 0.01, 0.025, 0.042\}$ ；LS 则有 5 个评价结果等级，对应的等级效用值为 $\{0, 2, 4, 6, 8\}$ 。由前提属性的参考值可初始化得到由 56 条规则组成的 BRB 系统^[8]，其中利用初始 BRB 系统对输油管道进行检漏的结果如图 3-6 所示。

由图 3-6 可知初始 BRB 系统对输油管道实际漏洞大小的检测存在较大误差，无法准确反映真实情况，需要对初始 BRB 的进行参数学习。从 2008 组测试数据中随机选取 500 组数据作为训练数据，并使用基于 VPSO 的参数学习方法对其进行参数训练，粒子群粒子个数为 30。训练后的 FD 前提属性权重为 1.000，PD 前提属性权重为 0.845。训练后 BRB 系统对输油管道泄漏检漏的结果如图 3-7 所示，从图中可以发现训练后 BRB 系统的模拟输出与测试数据几乎吻合，能准确地反映真实情况，推理准确性获得了提升。

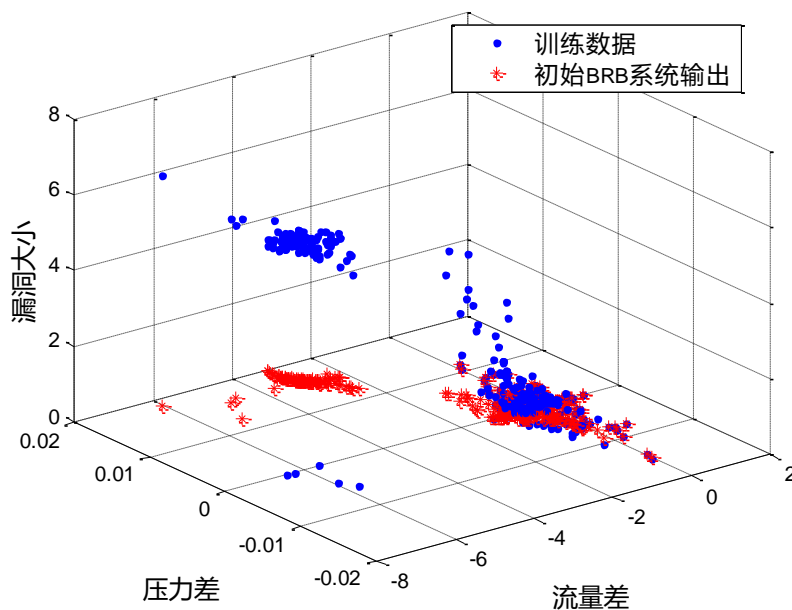


图 3-6 初始 BRB 系统输出和训练数据

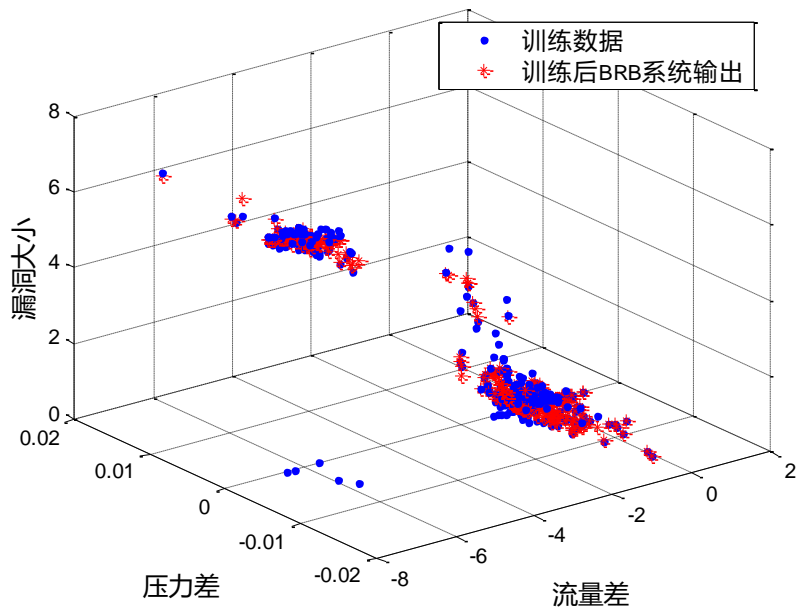


图 3-7 训练后 BRB 系统输出和训练数据

进一步分析经基于 VPSO 参数学习方法训练后的 BRB 系统, 将 2008 组数据作为测试数据对训练后的 BRB 进行测试, 图 3-8 给出了真实泄漏值和训练后的 BRB 系统模拟泄漏值的情况, 由图可知训练后的 BRB 系统具有良好的推理能力, 能够很好反映真实情况的动态特性, 并且未出现过拟合的现象。

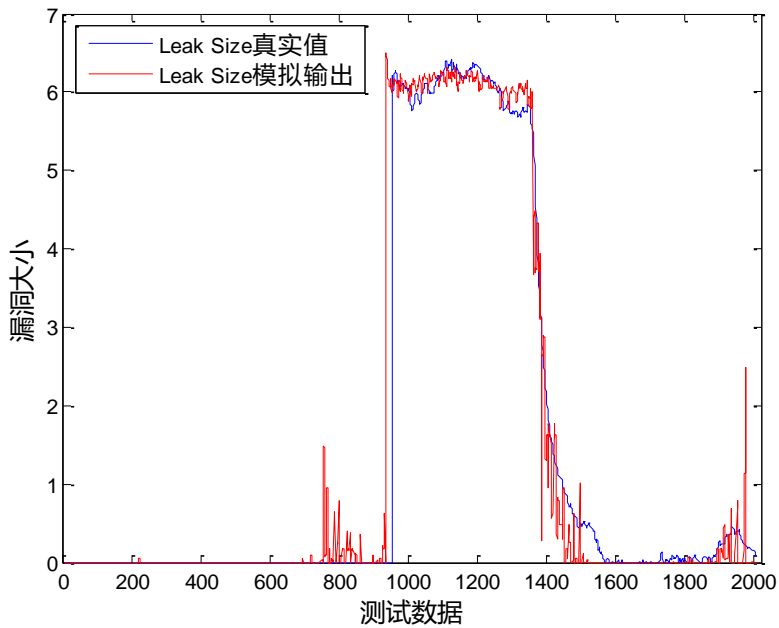


图 3-8 训练后 BRB 系统输出和测试数据

为对比本章方法与现有其他参数学习方法的性能,本章将 Xu 等^[8]和 Chen 等^[15]提出的参数学习方法作为比较对象,实验仍为输油管道泄漏检测实例。对比结果如表 3-1 所示,表 3-1 中 Xu-BRB 表示经 Xu 的方法训练所得的 BRB 系统;Chen-BRB 表示经 Chen 的方法训练所得的 BRB 系统,与本章所提出的基于 VPSO 参数学习方法相比二者都是运用 fmincon 函数求解 BRB 系统优化模型进行参数学习。与 Xu 方法相比 Chen 方法中新增前提属性参考值作为被训练的参数属于全局训练,从表中可以发现相比 Xu-BRB 的局部训练 Chen-BRB 获得到更优的 MAE,能更好的反映复杂系统的实际情况,但 Chen-BRB 训练参数更多,约束条件更为复杂,在利用 fmincon 函数求解时函数效率收敛速度相对较慢,耗时更多。PSO-BRB 表示基于 PSO 参数学习方法训练所得的 BRB 系统;VPSO-BRB 表示基于 VPSO 参数学习方法训练所得的 BRB 系统。通过对比表 3-1 中的 MAE 以及算法运行时间不难发现,PSO-BRB 系统的推理能力最不理想,获得的 MAE 比基于 fmincon 函数的 Xu-BRB 和 Chen-BRB 系统差。经本章所提方法训练所得的 BRB 系统,在一定程度上克服了传统 PSO 算法易早熟收敛以致寻优能力不理想的问题,其获得的 MAE 更优于其他参数训练方法训练所得的 BRB 系统,同时在算法的运行时间上比 Xu 和 Chen 的方法有明显的提升。

表 3-1 BRB 系统推理性能的比较

BRB 系统类型	MAE	运行时间/秒
初始 BRB	0.914578	—
Xu-BRB	0.171969	13449.060742
Chen-BRB	0.170341	21705.962351
PSO-BRB	0.197631	283.425000
VPSO-BRB	0.166478	290.150000

3.5 本章结语

本章针对现有参数训练方法在求解置信规则库优化模型中存在的收敛速度慢、可移植性差的缺点,提出了引入群智能算法中的粒子群算法求解该优化模型的解决方案,但传统的粒子群算法存在“早熟”及易陷入局部最优解的问题,因此本章首先对传统的粒子群算法进行了改进,提出一种速度重新赋值策略使粒子具有更好的寻优能力,同时针对等式和不等式约束条件采用对粒子位置进行约束的方法保证解的可行性,最后将其与置信规则库的参数训练有机的结合,提出了基于变速粒子群的置信规则库参数学习方法。理论分析和仿真实验结果表明,本

章所提的算法是有效可行的。

第四章 基于 BK 树的扩展置信规则库结构优化框架

4.1 引言

Liu 等^[22]将分布式置信框架引入置信规则的前件部分,并提出相应的扩展置信规则库(extended belief rule base, EBRB)系统表示、产生和推理的方法,该方法简单高效,而且 EBRB 系统在无需进行参数学习的情况下,也具有较好的推理准确性。Liu 等的方法,在推理效率方面仍存在瑕疵,主要体现在 EBRB 系统中规则均为无序存储状态,导致在对规则进行组合推理时需要遍历 EBRB 系统中所有规则以计算激活权重,当 EBRB 系统具有较多规则时,反复的遍历 EBRB 系统内规则将会导致推理效率低下。而在 BRB 系统中需要进行参数学习,通常该过程需要反复迭代以确定参数取值,所以需要反复进行推理,这势必增加算法的时间开销。推理效率不理想的问题将制约 BRB 系统和 EBRB 系统的实现与应用。因为扩展置信规则库具有较多的规则,更便于研究与实验分析,因此本章以 EBRB 系统作为研究对象。

BK 树(Burkhard-Keller tree, BK-tree)通过对数据构建树形索引结构进而可以对查询高效地搜索近邻数据^[45],被广泛应用于模式识别、文本和多媒体信息检索中^[46]。为了提高 EBRB 系统的推理效率以及组合更具代表性的规则进行推理决策,本章提出了一种基于 BK 树的结构优化框架。通过该结构优化框架可简单、高效地构建基于 BK 树的树形索引的 EBRB 系统,并对查询高效地搜索近邻规则,进而克服了传统 EBRB 系统在计算激活权重时需遍历整个 EBRB 的问题。结构优化框架的具体实现过程可概述为将 EBRB 内规则根据度量空间中彼此间的度量距离建立索引,在计算激活权重时利用索引对规则进行高效地搜索,再通过阈值设置的方式组合 EBRB 内关键规则,最终提升 EBRB 系统的推理效率和决策性能。此外,本章提出的基于 BK 树的结构优化框架有别于现有的参数学习方法,其并未改变 EBRB 系统的参数取值,因而可灵活的与任意的 EBRB 系统或其他具备置信框架的系统及方法相结合,达到提升系统或方法效率的目的。本章内容安排如下:首先介绍 BK 树数据结构的理论基础,其次在此基础上提出对扩展置信规则构建基于 BK 树的树形索引的方法,然后提出对具有树形结构索引的规则进行高效搜索的方法,并对该优化框架的有效性进行简单的分析,最后引入函数拟合问题、输油管道泄漏问题和多个分类数据集,通过与传统 BRB 系统及传统 EBRB 系统在推理效率和决策准确性方面比较,说明本章所提结构优化框架是切实可行的。

4.2 扩展置信规则库结构优化框架

扩展置信规则库方法能够简单高效地产生扩展置信规则，并避免了 BRB 系统存在的维度灾难问题^[18]。但在 EBRB 系统根据输入进行推理时，由于扩展置信规则以无序的方式存储，因此需要依次遍历 EBRB 中的所有规则以计算规则的激活权重，当规则数量较大时会导致 EBRB 系统的推理效率不理想。为了解决这一问题，本节提出一个基于 BK 树数据结构的 EBRB 结构优化框架，并与数据驱动的 EBRB 方法^[22]相结合，说明该框架的原理与作用。该结构优化框架先计算扩展置信规则间的度量距离，然后根据度量距离对原先无序存储的扩展置信规则建立基于 BK 树数据结构的索引，在搜索激活规则时通过设置的阈值减少搜索规则的数量并得到关键规则，最后使 EBRB 系统在推理决策时能够具有良好的效率和准确性。

4.2.1 Burkhard-Keller 树

Burkhard-Keller 树简称 BK 树，是由 Burkhard 和 Keller 提出的一种能高效地解决文章最优匹配问题的方法^[45]。该方法的主要思想是对数据在一个度量空间中建立树形数据结构的索引，并通过索引实现近邻数据搜索。设 X 为所有可能取值的集合， d 表示集合 X 的度量， $\forall x, y, z \in X$ ，对于一个度量空间 (X, d) ，应具有以下三个性质：

- (1) 非负性： $d(x, y) \geq 0$ ，且 $d(x, y) = 0$ 当且仅当 $x = y$
- (2) 对称性： $d(x, y) = d(y, x)$
- (3) 三角不等式： $d(x, z) \leq d(x, y) + d(y, z)$

BK 树具有一个特点，即一棵子树中的所有节点与父节点具有相同的度量距离。图 4-1 展示了一个三层 BK 树的结构图，其中第一层节点为根节点；第二层有 $m+1$ 个节点，表示可以将除根节点外的其他节点划分为 $m+1$ 棵子树，第 $i (i = 0, 1, \dots, m)$ 棵子树中的节点与根节点的度量距离都相同，要注意的是 i 并非一定等于度量距离，它可以是度量距离进行离散处理后对应的值，同样第三层子树中的节点与第二层的父节点也具有相同的度量距离。

基于度量空间的性质以及 BK 树的特点，文献[45]提出了一种有效的剪枝策略，可高效地实现多维空间中关键数据的搜索。需要注意的是，在常见的索引结构中，相似性查询有两种常用的方式，分别为 K 近邻查询和范围查询。对于一个询问 $q, q \in X$ ，搜索 BK 树的索引得到的结果数据集 Y 应满足 $d(q, x) \leq \theta_d, \forall x \in Y$ 。这表明结果数据集元素与询问的度量距离满足阈值 θ_d 范围的数据，可认为是一种范围查询方式。

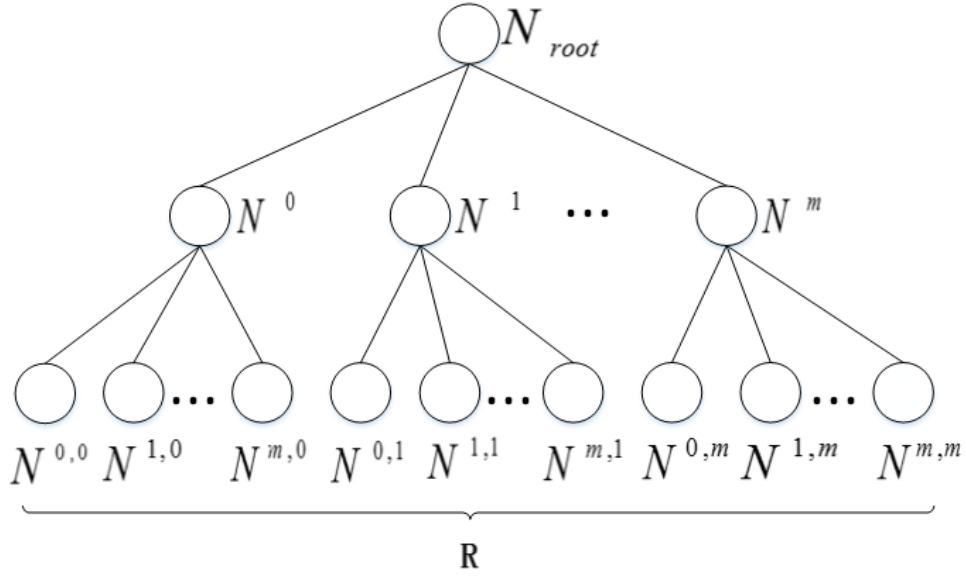


图 4-1 三层 BK 树结构图

4.2.2 基于 BK 树的扩展置信规则库构建

基于 BK 树构建 EBRB 是指基于 BK 树对扩展置信规则建立树形结构的索引。首先需要选择一个合适的度量空间对扩展置信规则进行度量。本章选择常见的欧氏距离作为度量距离并进行标准化，度量 R_p 和 R_q 两条规则的具体方法如下：

$$d(R_p, R_q) = \sqrt{\frac{1}{2 * T} \sum_{i=1}^T \sum_{j=1}^{J_i} (\alpha_{i,j}^p - \alpha_{i,j}^q)^2} \quad \text{公式 (4-1)}$$

其中，扩展置信规则的前提属性均为分布式置信度的形式，第 k 条规则的第 i 个前提属性表示为 $(A_{i,1}^k, \alpha_{i,1}^k), (A_{i,2}^k, \alpha_{i,2}^k), \dots, (A_{i,j_i}^k, \alpha_{i,j_i}^k)$ 。在对规则进行度量之后即可建立基于 BK 树的扩展置信规则间的索引，具体步骤如下：

步骤一，从扩展置信规则库 $R = \{R_1, R_2, \dots, R_L\}$ 中随意选择一条规则作为 BK 树的根结点即 R_{root} ，令 $R_n = R_{root} (n=1, 2, \dots, L)$ ， R_n 表示当前规则；

步骤二，计算 R_n 和集合中剩余规则的度量距离，将剩余的规则划分为 $m+1$ 个子集合 R^0, R^1, \dots, R^m ，其中 $d(R_n, r) = i, \forall r \in R^i$ ，依次对每个子集合执行步骤三；

步骤三，从集合 $R^i (i=0, 1, \dots, m)$ 中随意选择一条规则，建立其与 R_n 的索引并将其作为新的 R_n 。若子集合 R^i 的元素个数大于 1，即 $|R^i| > 1$ ，则执行步骤二，否则不做处理。

通过上述递归的算法步骤可以完成 BK 树的构建，从而将原本无序存储的扩展置信规则在一个度量空间中建立起树形结构的索引。假设现在有 5 条规则，规

则只有 1 个前提属性，前提属性有两个参考值，度量距离为欧氏距离，5 条规则在规则前件部分分别为：(0.4, 0.6)、(0.5, 0.5)、(0.3, 0.7)、(0.2, 0.8) 和 (0.6, 0.4)。图 4-2 给出了对这 5 条规则建立 BK 树形式索引的一种可能结构。

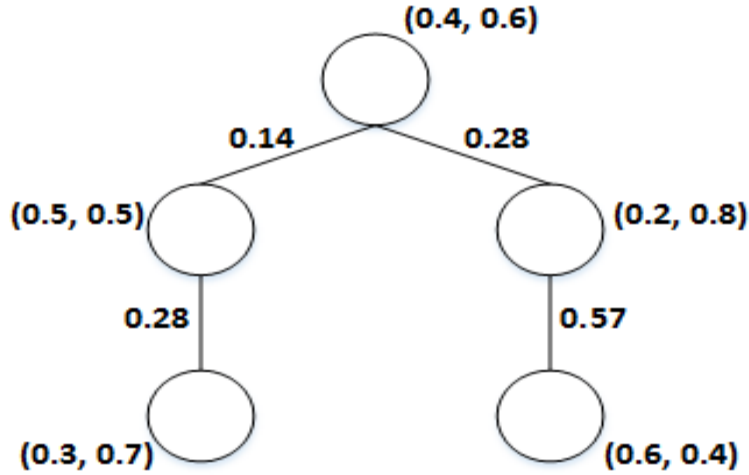


图 4-2 5 条规则索引结构图

4.2.3 基于 BK 树的扩展置信规则库搜索

构建完基于 BK 树的 EBRB 后，接着介绍基于 BK 树的 EBRB 搜索策略。假设输入数据为 x ，则在基于 BK 树的 EBRB 中搜索满足 $d(R_n, x) \leq \theta_d$ 的规则，并将其作为激活规则用于推理最终的决策结果，其中具体搜索步骤如下：

步骤一，设置阈值 θ_d ，计算当前规则 R_n 和 x 的度量距离，令 $d = d(R_n, x)$ 。若 $d \leq \theta_d$ 则说明 R_n 可能被激活。

步骤二，进行 $m+1$ 次判断， $m+1$ 表示以当前结点为根的子树个数，若满足剪枝策略 $|d_k - d| \leq \theta_d, k = 0, 1, \dots, m$ ，则将第 k 棵子树的根节点规则作为新的 R_n 执行步骤三，否则不做处理。其中， d_k 表示 R_n 和第 k 棵子树中规则的度量距离。

步骤三，计算当前规则 R_n 和 x 的度量距离更新 d ，若 $d \leq \theta_d$ ，则说明 R_n 可能被激活，然后执行步骤二。

在对基于 BK 树的 EBRB 进行激活规则搜索时，利用“三角形不等式性质”可以得到步骤二中的剪枝策略。根据剪枝策略只搜索可能满足 $d \leq \theta_d$ 条件的规则，减少了搜索规则的数量，从而提高 EBRB 系统推理的效率。以 4.3.2 节中的 5 条规则为例，现假设阈值 $\theta_d = 0.2$ ，输入 x 为 (0.45, 0.55)。根据公式 (4-1) 计算根节点规则和 x 的度量距离可得 $d = 0.07$ ，因此该规则可能被激活，然后根据剪枝策略可得左子树的结果为 $|0.14 - 0.07| = 0.07 < 0.2$ ，而右子树的结果为

$|0.28 - 0.07| = 0.21 > 0.2$ ，因此只对满足剪枝策略的左子树继续搜索而不对右子树进行搜索。随后计算节点 $(0.5, 0.5)$ 和 \mathbf{x} 的度量距离可得 $d = 0.07$ ，因此该规则可能被激活，然后根据剪枝策略可得子树的结果为 $|0.28 - 0.07| = 0.21 > 0.2$ ，可知子树中的规则与 \mathbf{x} 的度量距离不可能满足设置的阈值条件，所以无需对子树中的规则进行搜索。

4.2.4 结构优化框架下的规则推理

通过搜索得到可能被激活的规则集合之后，需计算规则的激活权重。根据公式(2-30)易知个体匹配度计算结果可能为负值，因此，本节提出改进的个体匹配度计算公式，改进后公式如下所示：

$$S^k(x_i, A_i^k) = S_i^k = 1 - \sqrt{\frac{1}{2} * \sum_{j=1}^{j_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \quad \text{公式(4-2)}$$

由公式(2-27)–(2-29)可将定量输入值 \mathbf{x} 转换为分布式形式，经公式(4-2)可得输入值 \mathbf{x} 与第 k 条规则的匹配度。此外，规则的不一致性极易影响 EBRB 系统的推理性能，由于在知识表示或获取时可能导致置信规则间存在不一致性，与此同时，当用历史数据产生置信规则时规则的不一致性还与噪声数据相关，因此规则推理时需采取适当的方法消除规则间的不一致性。本章采用文献[22]中方法对规则的一致性问题进行处理。

由公式(4-2)和公式(2-31)确定完激活规则后，接着根据 ER 方法对激活的规则进行组合，得到评价结果的置信度分布情况，再计算效用值得到最后 EBRB 系统的推理结果。为方便叙述，下文将基于 BK 树的结构优化框架和 EBRB 系统相结合的系统称为 BK-EBRB 系统，其中 BK-EBRB 系统进行构建和推理的流程如图 4-3 所示。

由图 4-3 可知结构优化框架是独立于扩展置信规则库和规则推理方法的框架，其没有改变 EBRB 系统的参数的取值，可见基于 BK 树的结构优化框架易与其他具备信度框架的决策模型相结合，对不同模型仅需确定合适的度量距离建立树形索引就能提升各方法搜索近邻数据的效率。

表 4-1 对 Liu-EBRB 系统与 BK-EBRB 系统的复杂度进行比较。从中可以发现 BK-EBRB 系统相比 Liu-EBRB 系统在构建系统时因为需要构建树形索引结构所以复杂度更高，而当查询激活规则时 BK-EBRB 系统无需对规则遍历，可以高效地搜索近邻规则，相比 Liu-EBRB 系统具有更低的复杂度，当询问个数越多时，BK-EBRB 系统的表现越好。通常 EBRB 系统仅需构建一次初始的规则库，而询问的次数则

比较多，因此将其视作询问密集型。所以在工程实际应用中 BK-EBRB 应能获得更好的推理效率。

表 4-1 Liu-EBRB 系统与 BK-EBRB 系统的复杂度比较

BRB 类型	Liu-EBRB 系统	BK-EBRB 系统
构建系统复杂度	n	$n \log_2 n$
查询激活规则复杂度	n	$n^\alpha (0 \leq \alpha \leq 1)$

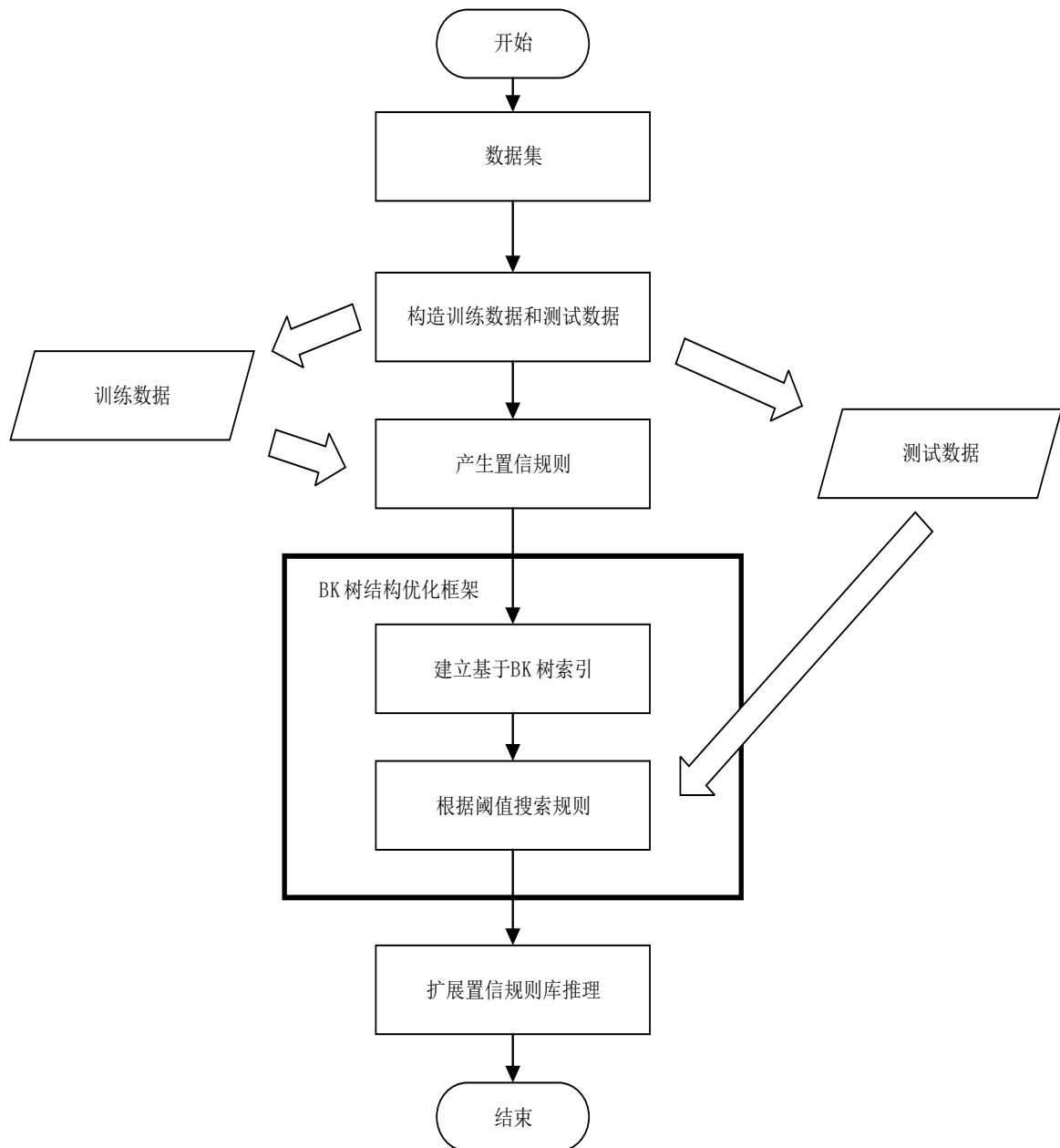


图 4-3 BK-EBRB 系统流程图

4.3 实验分析

为验证本章方法，本节引入非线性函数、输油管道泄漏两个实例以及多个分类数据集。实验环境为：Inter(R) Core(TM) i5-4570 CPU @ 3.20GHz 3.20GHz；4GB 内存；Windows 8 操作系统；算法实现平台 MATLAB R2012b 与 Visual Studio 2013。

4.3.1 函数拟合问题

文献[22]证明了 EBRB 系统是通用逼近器，可以逼近任意非线性映射。本节将通过一个非线性数学函数来检验 BK-EBRB 系统的推理性能和效率，并与 Yang 等^[9]的涉及局部参数学习的 BRB 系统和 Chen 等^[15]的涉及全局参数学习的 BRB 系统进行比较。为方便叙述，以下分别简称为 Yang-BRB 系统和 Chen-BRB 系统。

非线性数学函数如下所示：

$$f(x) = x \sin(x^2), 0 \leq x \leq 3 \quad \text{公式 (4-3)}$$

构建 EBRB 时， x 为前提属性并且具有 7 个参考值 $\{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ ，评价结果等级数目为 5，相对应的等级效用值依次为 $\{-2.5, -1, 1, 2, 3\}$ 。在 x 的取值范围内均匀地选择 500 个数值，并根据式公式 (4-3) 得到对应函数的真实值，再根据前文所提方法构建 BK-EBRB，其中阈值 θ_a 根据经验设定为 0.01，测试数据为在 x 的取值范围内均匀选择的 1000 组数据，评价指标为 MSE。

从图 4-4 中可以发现 Yang-BRB 系统的模拟输出与数学函数的真实输出存在明显的差距，拟合效果并不理想；图 4-5 中 Chen-BRB 系统的模拟输出与数学函数的真实输出差距不大，仅在极大极小值处存在明显欠拟合问题，整体上具有较好的拟合效果；图 4-6 中 BK-EBRB 系统的模拟输出与数学函数的真实输出差距不大，能够很好的拟合该数学函数。

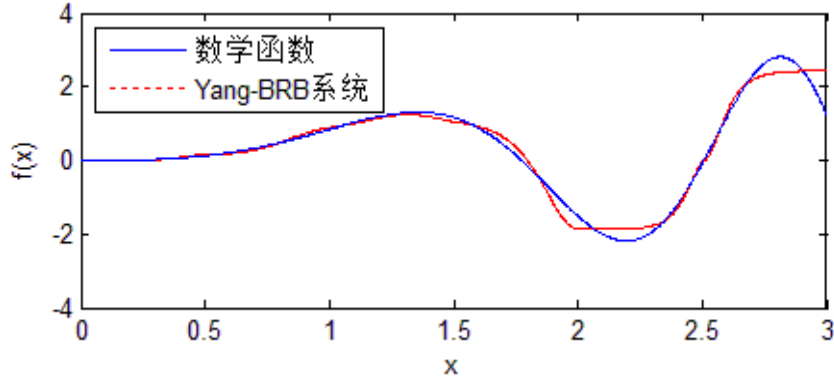


图 4-4 Yang-BRB 系统函数拟合图

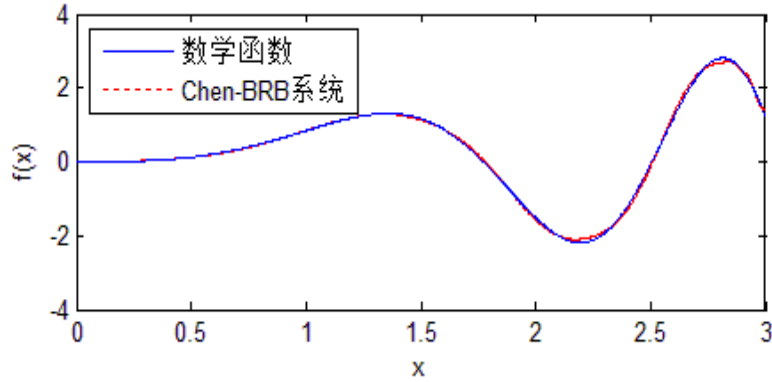


图 4-5 Chen-BRB 系统函数拟合图

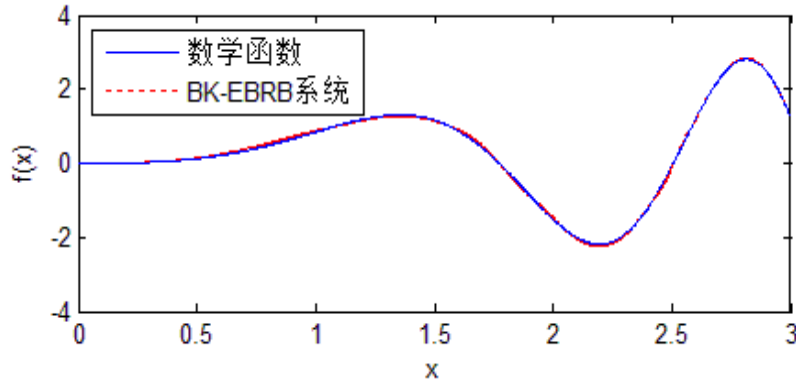


图 4-6 BK-EBRB 系统函数拟合图

表 4-2 对三种方法测试结果及运行时间进行了比较。其中，Yang-BRB 系统和 Chen-BRB 系统均运用 MATLAB 工具箱中的 `fmincon` 函数对 BRB 系统的参数进行学习，而 BK-EBRB 系统是将本章提出的 BK 树结构优化框架与数据驱动的 EBRB 方法相结合，并未进行参数学习。从表中可以发现 BK-EBRB 系统的模拟输出与真实值间的 MSE 为三种方法中最小的，在未进行参数学习情况下 BK-EBRB 系统也具有良好的推理能力。而参数学习是反复迭代的一个过程，需要大量的时间。因此，BK-EBRB 系统的运行时间最短，与其他两个方法相比极大的提高了系统的效率。

表 4-2 函数拟合 BRB 系统推理性能比较

BRB 类型	MSE	运行时间/秒 (s)
Yang-BRB	0.037129	394.2160
Chen-BRB	0.001996	695.6130
BK-EBRB	0.000814	0.00119

4.3.2 输油管道泄漏问题

一个具体的实际问题输油管道泄漏在本节中作为研究对象^[8,15,17]，通过使用输油管道泄漏的真实泄漏数据对本章提出的 BK-EBRB 系统性能进行验证。在该实际问题中，当输油管道发生泄漏时，输油管道中油液的流量和压力会发生变化。因此，选择输油管道的流量差 (Flow Difference, FD) 以及输油管道的平均压力差 (Pressure Difference, PD) 对泄漏大小 (Leak Size, LS) 进行估计。

在构造 EBRB 时，选取了 2008 组从无泄漏到发生 25% 泄漏状况的数据作为实验数据。因为 FD 和 PD 可以反映输油管道泄漏情况，所以系统的输入为 FD 和 PD ，而 LS 则为输出。其中，根据专家经验得到前提属性 FD 有 8 个参考值，分别为 $\{-10, -5, -3, -1, 0, 1, 2, 3\}$ ； PD 有 7 个参考值，分别为 $\{-0.042, -0.025, -0.01, 0, 0.01, 0.025, 0.042\}$ ；输出 LS 则有 5 个评价等级，分别为 $\{0, 2, 4, 6, 8\}$ 。

在 2008 组数据中，根据文献[15]的方法按照一定比例从三个时间段随机选择总共 1500 条数据作为训练数据，产生置信规则，然后按照 Liu 的方法^[22]和本章方法分别构造 Liu-EBRB 系统和 BK-EBRB 系统进行比较，并以平均绝对误差 (Mean Absolute Difference, MAE) 作为评价指标。

图 4-7 和图 4-8 分别将 Liu-EBRB 系统和 BK-EBRB 系统 ($\theta_a=0.4$, θ_a 即 θ_a) 产生的模拟输出与真实的数据进行比较。两种方法均根据训练数据产生 EBRB，后者引入基于 BK 树的优化框架。从图中可以发现 BK-EBRB 系统能较好的对输油管道泄漏情况进行检测，得到与真实值接近的结果。而当 $PD \in [-0.02, 0]$ 且 $FD \in [-10, -5]$ 时，Liu-EBRB 系统产生的模拟输出与真实值存在较大差距，BK-EBRB 系统的模拟输出则更接近真实的情况。其主要是因为 BK-EBRB 系统对置信规则建立基于 BK 树的索引，同时设置了的阈值，进而减少了激活规则的数量。另一方面，由于仅对关键规则进行组合，减少了不一致规则对最终结果的影响，提高了系统的推理能力。

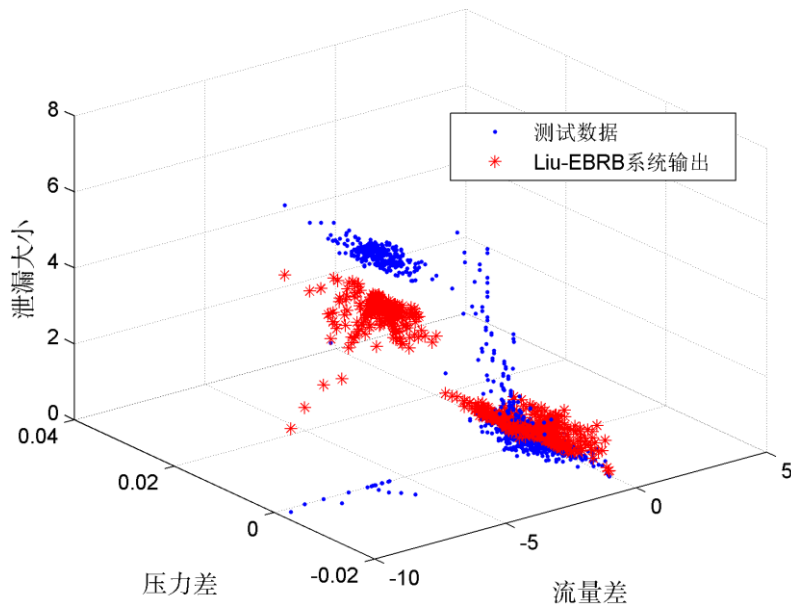


图 4-7 Liu-EBRB 系统输出和测试数据

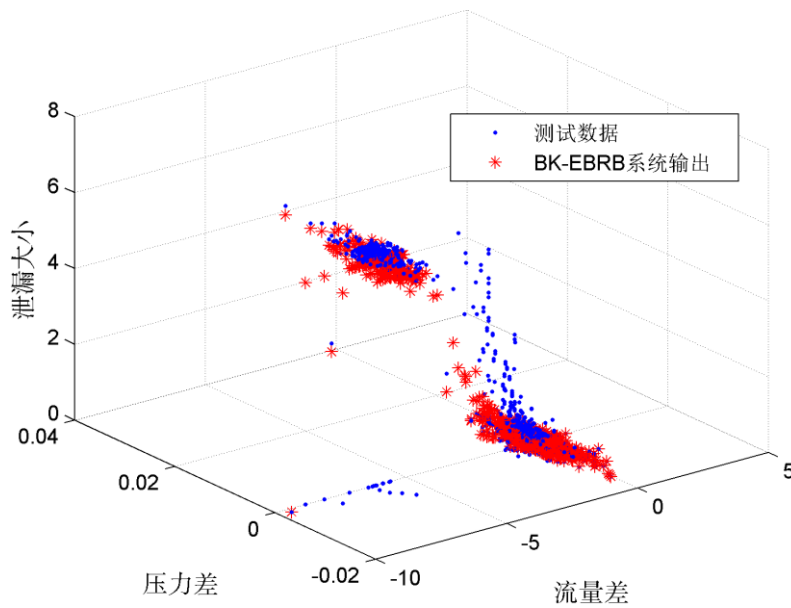


图 4-8 BK-EBRB 系统输出和测试数据

表 4-3 列出了 Liu-EBRB 系统和三个设置不同阈值的 BK-EBRB 系统产生的模拟输出与真实值间的 MAE 以及各自进行推理时搜索规则的次数。图 4-9 则以柱状图的形式更形象地对四个 EBRB 系统在规则推理时的搜索规则次数进行比较。可以发现当 BK-EBRB 系统的阈值设置为 1 时，BK-EBRB 系统和 Liu-EBRB 系统具有相同的 MAE 并且搜索规则的次数一致。这是因为 Liu-EBRB 系统需要对置信规则

进行完整地遍历，而当阈值为 1 时，剪枝策略没有发挥作用，BK-EBRB 系统仍需遍历所有的规则。当 BK-EBRB 系统的阈值小于 1 时，搜索规则次数减少。因为当阈值小于 1 时，剪枝策略将发挥作用，可以缩小搜索规则的范围，减少搜索规则的数量，从而提高 EBRB 系统的推理效率。由表 4-3 还可以发现相较于 Liu-EBRB 系统，BK-EBRB 系统获得了更小的 MAE 具有更好的推理能力。这是因为 BK-EBRB 系统通过阈值的设置，激活更为关键的规则进行组合，从而提升了 EBRB 系统的推理能力。

表 4-3 输油管道泄漏检测 EBRB 系统推理性能比较

BRB 类型	MAE	搜索规则次数
Liu-EBRB	0.626240	3,012,000
BK-EBRB ($\theta=1$)	0.626240	3,012,000
BK-EBRB ($\theta=0.7$)	0.558087	2,646,759
BK-EBRB ($\theta=0.4$)	0.231400	1,432,841

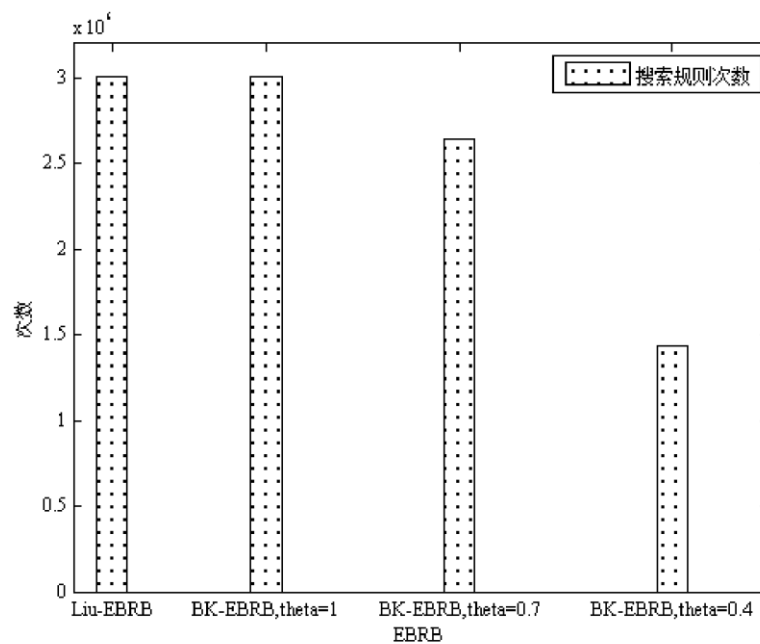


图 4-9 EBRB 系统进行推理搜索规则次数

通过输油管道泄漏实例对 Liu-EBRB 系统和 BK-EBRB 系统进行比较，表明基于 BK 树的结构优化框架在提高 EBRB 系统推理效率的同时也可使 EBRB 系统能够更准确反映系统的行为。

4.3.3 分类数据集测试

为了验证本章方法的有效性,从UCI上选择9个著名的分类数据集进行测试。通过5折交叉验证的方法,构造训练数据以及测试数据。每个前提属性都根据数据范围设置6个均匀分布的参考值,评价结果数与分类数一致,以此根据数据产生扩展置信规则。然后根据本章方法构造4个BK-EBRB系统,并分别设置阈值 θ 为1, 0.8, 0.6和0.4,度量距离为欧氏距离。实验结果如表4-4所示,从表中可以发现,在大部分数据集上随着阈值的减小,BK-EBRB系统的推理准确性获得了提高,具有更好的分类准确度,这表明通过阈值设置激活关键规则可以提高系统的推理能力。BK-EBRB系统在Ecoli、Knowledge和Yeast这三个数据集性能提升效果最为明显,而在Breast和Glass这两个数据集上,BK-EBRB系统的推理准确性呈现出先升高后降低的情况。阈值的设置会影响BK-EBRB系统的推理性能,合理、适合的阈值将会使系统具有更好的推理性能,反之将会降低系统的推理性能。从表中可以发现不同数据集相同阈值的推理能力不一致,应根据数据集的自身结构特点设置不同的阈值。阈值的设置可以通过枚举的方法对不同系统的推理性能进行比较,选择具有最优推理准确性系统对应的阈值取值,也可以将该问题视为一个最优化问题,并通过相关方法进行求解得到最优阈值取值。

表 4-4 分类数据集 BK-EBRB 系统推理性能比较

	BK-EBRB($\theta=1$)	BK-EBRB($\theta=0.8$)	BK-EBRB($\theta=0.6$)	BK-EBRB($\theta=0.4$)
Breast	71.70%	71.70%	72.64%	70.75%
Diabetes	73.28%	73.79%	73.79%	73.79%
Ecoli	79.76%	80.06%	82.44%	85.42%
Glass	64.95%	65.42%	66.36%	64.95%
Knowledge	78.91%	79.40%	81.89%	82.38%
Seeds	90.00%	90.00%	90.48%	90.48%
Transfusion	76.34%	76.47%	76.60%	77.01%
Vertebral	72.26%	72.58%	73.55%	73.55%
Yeast	45.42%	45.49%	46.16%	50.40%

4.4 本章结语

针对现有 BRB 系统和 EBRB 系统中因规则以无序的方式存储,导致在推理时需采用遍历规则库内所有规则的方式计算激活权重,从而导致系统推理效率不理想的问题。本章提出一种基于 BK 树的 EBRB 结构优化框架,其通过对规则建立基于 BK 树的索引结构减少推理时搜索规则库中规则的数量,另一方面,通过设定合适的阈值筛选出更具有代表性的规则用于规则组合,提高 EBRB 系统的推理效率。此外,该结构优化框架还易与其他具有置信框架的决策模型相结合,具有良

好的扩展性。实验分析中，通过在函数拟合问题和输油管道泄漏问题中与各类 BRB/EBRB 系统进行对比，验证了本章所提的方法能够提升 EBRB 系统的效率和决策性能，通过对多个分类数据集进行测试，进一步验证方法有效性并简单分析阈值设置方法。

第五章 基于 AdaBoost 的置信规则库集成学习方法

5.1 引言

置信规则库系统的推理准确性不仅与参数的取值息息相关还受到训练数据集的影响。BRB 系统的参数学习是一种监督学习,通过参数学习得到的参数取值能使 BRB 系统对训练数据具有较好的推理能力;EBRB 系统在构建初始规则库时,扩展置信规则是根据训练数据集而产生,训练数据集的规模和多样性会影响置信规则库系统的推理能力。不理想的训练数据集容易导致 BRB 系统对训练数据集具有良好的推理能力,而对测试数据集或其他数据无法给出准确结果,出现过拟合问题;理想的训练数据集则可使 BRB 系统具有良好的稳定性,此外置信规则间的一致性也会影响置信规则库系统的推理准确性。

为了使 BRB 系统具有良好的稳定性和泛化能力,本章提出一种基于 AdaBoost 算法的 BRB 系统集成学习方法。集成学习被认为是机器学习研究中的热门方向^[47],其主要思想是按照一定方法学习得到多个具有不同表现的模型,然后根据某种规则组合多个模型的结果,组合后的结果优于一个模型的结果。正因如此,集成学习受到许多研究学者的关注。Schapire 等^[48]提出对多个互补的模型进行组合的 Boosting 算法雏形;Freund 和 Schapire^[49]对训练实例赋予权重,根据结果对权重进行修正从而生成一系列互补的模型;Breiman^[50]提出 Bagging 方法并指出集成学习中集成模型稳定性的作用;周等^[51]对神经网络方法进行集成学习,提出选择集成的思想并证明该理论的有效性。

集成学习主要包括两个部分,其一是训练具有不同推理能力的模型;其二是按照某种规则对模型集群的结果进行有效组合。针对以上问题进行了相关的研究,本章的内容安排如下:首先介绍集成学习的理论基础并分析单个 BRB 系统可能存在的不足之处;然后介绍如何基于 AdaBoost 算法生成构建互不相同且具有互补能力的 BRB 系统以及组合 BRB 系统推理结果的方法;接着通过分类数据对所提集成学习方法进行实验验证,分析说明本章方法的有效性和可行性;最后总结本章的研究内容。

5.2 集成学习方法

5.2.1 集成学习的概念

从狭义上来说,集成学习是指对由一系列同样方法构建的模型进行组合的方法。常见的构建模型的方法如:决策树、神经网络和 SVM 等等。从广义上来说,

只要是对多个解决同一问题的模型进行组合就是集成学习。在集成学习研究的初期,研究学者主要是研究狭义的集成学习,随着大数据受到人们越来越多的重视,广义的集成学习被更多学者所接受。

5.2.2 集成学习器的构建与组合

集成学习过程分为两个步骤,步骤一是构造具有互补能力的多个学习器,步骤二是组合多个模型的结果得到最终决策。

集成学习效果的好坏与基本模型的多样性有关^[52],因此构建表现差异大的且能够互补的模型十分重要。广义的集成学习构建模型的方法不同,因此得到的模型也不尽相同;对于狭义的集成学习,构建不同的模型可以分为四种方法:

一、训练数据处理法,该方法在构建不同模型时,按照一定规则生成不同的训练数据,从而得到不同的模型,常见的方法有 Bagging、Boosting 等;

二、输入特征处理法,对于一个具有多个输入特征的问题,通过选取部分特征作为模型的输入从而获得不同的模型;

三、输出结果处理法,文献[53]中提出一种对输出结果进行编码,将多分类转换为二值分类;

四、参数处理法,常见的如决策树和神经网络等建模方法需要确定参数的取值,参数取值的不同,模型的输出结果也有明显的不同。

组合模型集群的预测结果意味着将不同的输出结果组合而产生一个决策。Xu 等在文献[54]中将分类问题的集成学习模型的输出形式归纳为三个层次,具体为:

- 一、抽象层次,每个模型只产生一个分类结果或分类子集,没有其他信息;
- 二、排序层次,每个模型产生所有可能的分类结果并按可能性排序;
- 三、度量层次,每个模型产生分类的度量值,如概率或置信度等形式。

在上述的模型输出形式层次中,层次越高所包含的信息越丰富,模型组合的方式也更灵活、丰富。常见的模型组合方式有:简单投票法^[54, 55]、贝叶斯投票法^[54, 56]、基于 D-S 证据理论的组合法^[54, 57]。

5.2.3 问题提出

Yang 等^[9]通过对应的输入输出数据进行参数学习,确定 BRB 系统参数的具体取值,Liu 等^[22]根据数据生成初始置信规则,在这两个过程中训练数据的质量对构建的 BRB 系统的性能起着至关重要的作用。数量较少或抽样不均衡的训练数据易使 BRB 系统出现过拟合、泛化能力差和稳定性不理想的问题。此外,Yang 和 Liu 均指出由数据生成的置信规则库可能存在不一致即置信规则相互冲突^[5, 15],

例如某些规则的规则前件部分十分相似，但在结论部分却具有明显的不同。值得注意的是，通过参数学习可能导致原先不存在不一致情况的规则库出现规则冲突的问题。不一致的置信规则库将会影响 BRB 系统对于某些数据进行推理的准确性。对于复杂的规模较大的决策问题，BRB 系统的推理性能与系统的参数取值密切相关^[38]，参数取值的细小差异都可能使 BRB 系统推理的结果出现明显差异。通过参数学习对复杂决策问题构建的单个 BRB 系统具有良好的泛化能力存在一定的困难。因此有必要对 BRB 系统进行优化提升其泛化能力和稳定性。

5.3 置信规则库集成学习方法

针对 5.2.3 节中 BRB 系统存在的问题，本节提出对一种对 BRB 系统进行集成学习的方法。该方法引入 AdaBoost 算法，在迭代过程中维护实例的权重，使得新生成的 BRB 子系统能够关注之前 BRB 子系统的输出与真实输出存在较大误差的实例；在构建 BRB 子系统时利用 DE 算法进行参数学习，保证 BRB 子系统的性能，从而能够生成一系列互补型的且推理能力较优的 BRB 子系统，以发挥模型组合作用。

5.3.1 AdaBoost 算法

AdaBoost (Adaptive Boosting) 算法是 Freund 和 Schapire^[49]提出的一种迭代算法，通常被应用于分类问题。该算法能够针对同一训练数据产生不同的分类器，其核心思想是首先赋予每个训练数据相同的权重，然后应用学习算法构建一个分类器，再根据该分类器的分类准确率对测试数据进行重新加权，增加错误分类数据的权值，减小正确分类数据的权值。按照该方式进行迭代可以生成一系列的分类器，并且新生成的分类器能够专注于错误分类的数据，从而互相补充^[52]。

AdaBoost 算法流程如算法 5-1 所示。

值得注意的是当分类器的加权误分类率大于等于 0.5 时，可以发现对错误数据进行权重更新后权重并未得到增大；若加权误分类率等于 0 则表示子分类器能够完全准确地进行分类，该分类器的权重趋近无穷大，在组合多个子分类器后相当于该完美子分类器进行分类，失去集成学习的意义。因此通过学习算法得到的分类器的加权误分类率应小于 0.5 且不等于 0，以保证分类器组合后能发挥提升的作用。

算法 5-1 AdaBoost 算法

Input L 组训练数据, 构建分类器数量 N , 学习算法 PL

Output 分类结果 $G(x)$

$$u^{(1)} = [\frac{1}{L}, \frac{1}{L}, \dots, \frac{1}{L}] // \text{初始化数据权重均为 } \frac{1}{L}$$

for $i = 1, 2, \dots, N$

$g_i = PL(D, u^{(i)}) //$ 应用学习算法 PL 构建基分类器

$$\varepsilon_i = \frac{\sum_{j=1}^L u_j^{(i)} [y_j \neq g_i(x_j)]}{\sum_{j=1}^L u_j^{(i)}} // \varepsilon_i \text{ 表示分类器的加权误分类率}$$

for $j = 1, 2, \dots, L //$ 更新数据权重

if $y_j \neq g_i(x_j)$

$$u_j^{(i+1)} = u_j^{(i)} \times \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} // \text{增大错误分类数据权重}$$

else

$$u_j^{(i+1)} = u_j^{(i)} \div \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} // \text{减小正确分类数据权重}$$

end for

$$\alpha_i = \ln\left(\sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}}\right) // \alpha_i \text{ 表示第 } i \text{ 个分类器的权重}$$

end for

$$\text{return } G(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i g_i(D)\right)$$

5.3.2 置信规则库系统集成学习

上一节中介绍了 AdaBoost 算法的基本原理。通过对训练数据权重的更新, 可以得到关键数据, 然后需要生成能够对权重较大的关键数据具有良好推理表现的 BRB 系统。本节引入差分进化 (Differential Evolution, DE) 算法, 作为学习算法用于求解新的 BRB 系统。

DE 算法是 Storn 等^[58]人于 1995 年提出的模拟生物进化的随机迭代算法, 基

于差分的简单变异操作使得 DE 算法具有良好的全局搜索能力和鲁棒性。

假设基于 DE 算法的置信规则库系统学习方法步骤如下：

步骤一：初始化规模为 NP 的种群 P ，设置迭代次数 T ，每个维度为 N 的个体可以表示为 $P_i^t = (x_1, x_2, \dots, x_N), i = 1, 2, \dots, NP, t = 1, 2, \dots, T$ ， P_i^t 表示初始种群中的第 i 个个体。

步骤二：为所有个体产生变异个体，具体公式如下：

$$VP_{i,j} = x_{r1,j} + F \times (x_{r2,j} - x_{r3,j}) \quad \text{公式 (5-1)}$$

其中 $VP_{i,j}$ 表示变异个体 VP_i 的第 j 维元素， $j = 1, 2, \dots, N$ 。 $r1, r2, r3$ 表示三个不同的种群个体， $r1 \neq r2 \neq r3$ 。 F 为缩放因子，通常取值为 2。

步骤三：对群众个体进行交叉重组，具体公式如下：

$$CP_{ij} = \begin{cases} VP_{i,j}, & \text{if } (rand() \leq CR) \vee (j = Ir) \\ P_{i,j}, & \text{otherwise} \end{cases} \quad \text{公式 (5-2)}$$

其中 $CP_{i,j}$ 表示交叉重组后个体 CP_i 的第 j 维元素， $j = 1, 2, \dots, N$ 。 $rand()$ 表示随机均匀生成 $[0, 1]$ 之间的数， CR 为交叉因子，通常取值为 0.5。 Ir 取值范围是 $[1, N]$ 的整数。

步骤四：更新目标函数计算个体适应值，选择具有较优适应值的个体成为新的种群个体。第 $t+1$ 代个体的生成方式如下：

$$P_i^{t+1} = \begin{cases} CP_i, & \text{if } (F(CP_i) < F(P_i^t)) \\ P_i^t, & \text{otherwise} \end{cases} \quad \text{公式 (5-3)}$$

步骤五，若迭代次数达到设定条件 T 时则返回具有最好适应值的个体，算法结束，否则跳转步骤二。

通过 DE 算法可以构建对不同权重的训练数据具有较好推理能力的 BRB 子系统，以进行集成。

基于 AdaBoost 算法的置信规则库集成学习方法流程如下图所示：

其中，根据算法 5-1 先对训练数据设置同样的权重，然后建立 BRB 系统，并利用 DE 算法对 BRB 系统的参数进行训练，具体过程与第三章中的基于变速 PSO 算法的参数学习过程类似。但需要注意的是，在该迭代的参数学习过程中，目标函数为算法 5-1 中分类器即 BRB 系统的加权误分类率，以此训练得到的 BRB 系统会对权重较高的数据更加关注。随后，根据训练得到的 BRB 系统对数据权重进行调整，重复上述过程训练对更新权重后的数据具有更好推理能力的 BRB 系统，最后组合所有 BRB 系统，得到最后结果。

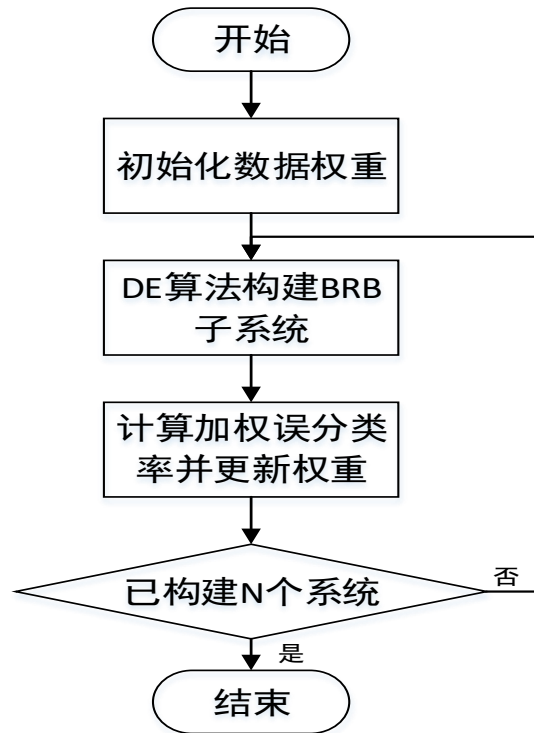


图 5-1 基于 AdaBoost 算法的置信规则库集成学习方法流程图

5.4 实验验证

本章提出了基于 AdaBoost 算法的置信规则库集成学习方法，因为 AdaBoost 算法主要被应用于分类问题，所以为了验证本章所提方法的有效性，本章选取了多个分类问题测试数据集进行实验验证。

5.4.1 实验数据集

本实验从 UCI 数据库中选择了 6 个分类问题测试数据集，数据集的详细信息如表 5-1 所示：

表 5-1 分类测试数据集

数据集	样例数	特征数	类别数
Breast	106	9	6
Ecoli	336	7	8
Glass	214	9	6
Haberman	306	3	2
Iris	150	4	3
Transfusion	748	4	2

本实验基于 EBRB 系统，引入 K 折交叉验证方法，在一定程度上克服了因训练数据集而引起的过拟合问题。首先应用 5 折交叉验证将数据集划分为训练数据和测试数据，由训练数据生成 EBRB 规则，构建初始的 EBRB 系统。其中每个前提属性参考值数量均设置为 5，参考值按照数据集中该属性对应数据的最大值和最小值进行均匀选择。

5.4.2 实验结果

应用本文所提方法对表 5-1 中的测试数据集进行实验，应用 AdaBoost 算法对训练数据的权重进行更新，然后通过 DE 算法对 EBRB 系统的规则权重进行训练，构建 20 个 EBRB 系统并进行组合，总共进行 10 次实验，实验结果取平均值，实验结果如表 5-2 所示。本实验中一共构建三种不同的 EBRB 系统用于分析比较算法的性能，其中 EBRB 系统是由本章所提方法所构建的第一个 EBRB 子系统；AdaEBRB_S 是由 20 个 EBRB 子系统组成的集成 EBRB 系统，该系统的组合方式是简单投票法；AdaEBRB_W 是由相同的 20 个 EBRB 子系统组成的集成 EBRB 系统，该系统的组合方式采用的是加权投票法，其中权重为每个 EBRB 子系统的系统权重。

从表中可以发现在 6 个不同的测试数据集上，两种进行集成学习的 EBRB 系统与单个 EBRB 系统相比均具有更好的分类准确性，其中在 Breast 和 Glass 两个测试数据集上提升效果较为明显，这表明本章所提的方法即构建 BRB 集成系统具有更好的推理准确性。对比简单投票的 AdaEBRB_S 集成系统和加权投票的 AdaEBRB_W 集成系统，可以发现 AdaEBRB_W 集成系统的分类准确性会略优于 AdaEBRB_S 集成系统，这是因为加权投票考虑了 EBRB 子系统的分类准确性，具有更好推理准确性的子系统在集成学习中可以发挥更多的作用。

表 5-2 基于 AdaBoost 算法的集成学习方法实验结果

数据集	EBRB	AdaEBRB_S	AdaEBRB_W
Breast	88.87%	93.58%	93.67%
Ecoli	91.52%	91.76%	92.32%
Glass	85.00%	86.21%	86.26%
Haberman	80.06%	80.13%	81.60%
Iris	98.20%	98.60%	98.80%
Transfusion	75.09%	75.48%	75.51%

图 5-2 为单次实验中 AdaEBRB_W 集成系统和 AdaEBRB_S 集成系统在测试数据集 Breast 上分类错误数。其中横轴表示参与集成的 EBRB 子系统数量，纵轴表示 EBRB 集成系统的分类错误数。从图中可以发现随着参与集成的 EBRB 子系统的数量增多，集成系统的分类错误数并不是单调递减，而是在整体趋势上呈现一种下降的趋势，其中，AdaEBRB_W 集成系统分类错误数变化较 AdaEBRB_S 集成系统相比波动较小。

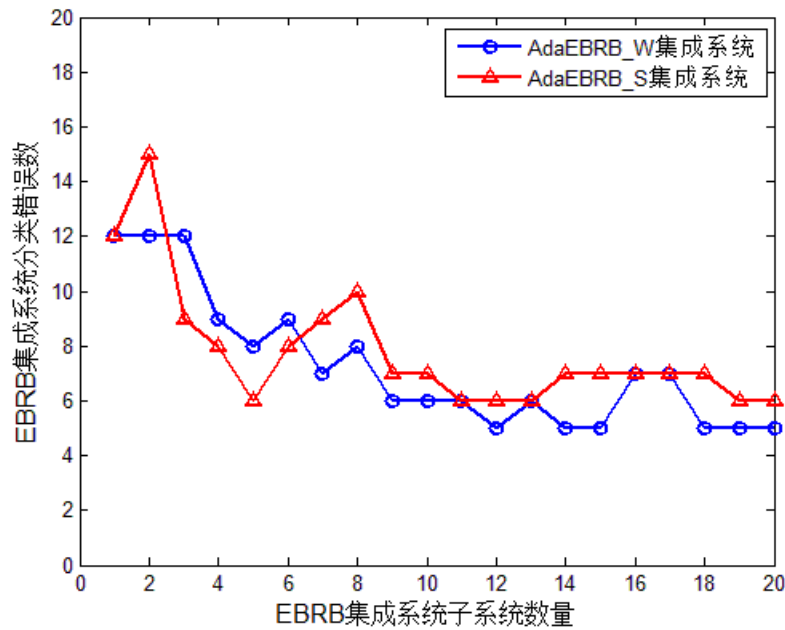


图 5-2 EBRB 集成系统分类情况

5.5 本章结语

本章针对单个置信规则库系统的推理准确性受到训练数据质量、规则一致性以及参数取值影响的问题，提出构建多个 BRB 子系统并对系统推理结果进行组合。因为最终系统的推理结果是由多个 BRB 子系统共同决定，只要多个 BRB 子系统能够相互补充，就能扬长避短，在一定程度上降低上述问题的影响而增强系统的推理表现。本章所提的方法基于 AdaBoost 算法，通过对训练数据权重的更新使得新生成的 BRB 子系统与旧的 BRB 子系统互不相同，同时应用 DE 算法对规则权重进行学习使得 BRB 子系统的推理能力能够满足用于集成的要求。在实验验证中通过分类测试数据集对所提方法进行实验并与单个统进行比较，实验结果表明本节方法能够提高 BRB 系统的推理准确性。

第六章 总结和展望

6.1 工作总结

置信规则库系统是基于证据理论算法的置信规则库推理方法所构建的专家系统，能够对模糊不确定性、概率不确定性的信息、数据进行表示，并建立输入和输出间的非线性模型。自从 2006 年杨剑波教授提出该方法以来，引起广大研究学者的关注。研究学者在理论和工程实际应用问题上对其展开了各种研究，已被成功应用于医疗诊断、输油管道泄漏检测、海面采油系统安全分析、生产运作管理等众多领域。

关于置信规则库专家系统的相关研究还不到十年的时间，还存在许多的不足之处。首先，为了使 BRB 系统具有较好的推理准确性需要根据优化模型以及对应的输入输出数据进行参数学习，该过程通常应用 MATLAB 工具箱中的 `fmincon` 函数，而面对参数较多、问题规模较大且约束条件复杂的问题时，该函数的寻优能力和效率都不够理想；其次，BRB 系统的规则是以无序的方式进行存储，BRB 系统在进行推理时，对每个数据均需对规则库的规则进行遍历，以计算每条规则的激活权重，然后对激活规则进行组合，对 BRB 系统而言，因为参数学习的本质是一个最优化问题，通常求解方法需要进行迭代，反复进行推理计算目标函数导致该过程效率低下，而对 EBRB 系统而言，EBRB 系统会存在大量的规则反复遍历规则库会影响系统推理效率；最后，现有的主要方法均是对复杂决策问题建立一个 BRB 系统，而 BRB 系统的推理能力不仅与领域专家的知识、经验相关还与训练数据相关，同时对于一个复杂的、规模较大的问题确定大量参数的取值使得 BRB 系统具有良好的推理能力、泛化能力和稳定性具有一定难度。

本课题对置信规则库系统优化方法开展深入地研究，研究成果和创新点具体如下：

(1) 提出基于变速 PSO 算法的置信规则库参数学习方法，在传统粒子群算法中引入变速策略以及限制粒子在解空间中，提出能够求解带有约束条件的最优化问题且具有更好寻优能力的 VPSO 算法，并将 VPSO 算法应用于参数学习。基于 VPSO 的参数学习方法具有良好的求解精度和时间效率，且利于移植更加灵活，丰富置信规则库系统参数学习方法。

(2) 提出基于 BK 树的置信规则库优化框架，根据规则的表达方式选择合适的度量距离对规则的相似度进行度量并构建基于 BK 树的树形索引，通过设定阈值仅对关键、重要的规则进行组合。该优化框架可以提高 BRB 系统的效率和推理准确性并具有良好的适用性和灵活性。

(3) 提出基于 AdaBoost 算法的置信规则库集成学习方法, 该方法对训练数据权重进行更新并应用 DE 算法学习新的 BRB 系统, 从而巧妙的生成一系列互补的 BRB 系统。通过该方法对 BRB 系统进行集成学习可以使 BRB 系统获得更好的推理能力。

6.2 工作展望

本课题对置信规则库优化方法的研究在很多方面仅为初步成果, 还存在许多不足和可以提高的地方。未来研究工作的方向如下:

(1) 除了参数学习可以提升 BRB 系统推理能力, BRB 系统的结构也对推理能力具有重要影响, 紧致、合理的置信规则库可以提升 BRB 系统的推理准确性和方法效率。如何对 BRB 系统进行有效的结构学习并将结构学习和现有的参数学习相结合是下一步的研究方向。

(2) 本研究基于一种树形的数据结构对置信规则建立索引, 实现了对规则的高效搜索, 除了基于树的索引技术, 还有基于哈希的索引技术。基于哈希的索引技术与基于树的索引技术相比对空间的需求更小, 更加灵活。所以考虑下一步研究如何对置信规则建立基于哈希方法的索引并根据输入实现近邻激活规则搜索。

(3) BRB 系统作为一种灰箱方法, 其可解释性和理解性也同样重要。现有的研究方法往往关注于 BRB 系统的推理准确性而忽略其可解释性。可解释性和推理准确性通常被认为是两个相互冲突的目标, 因此如何对 BRB 系统可解释性进行度量、如何构建 BRB 系统使其的可解释性和推理精度达到一个权衡稳定的状态将是下一步待解决的问题。

参考文献

- [1] Durkin J, Durkin J. Expert systems: design and development[M]. Prentice Hall PTR, 1998.
- [2] Meystel AM, Albus J S. Intelligent systems: architecture, design, and control[M]. John Wiley & Sons, Inc., 2000.
- [3] 蔡自兴. 人工智能及其应用[M]. 清华大学出版社, 1996.
- [4] Parsons S. Current approaches to handling imperfect information in data and knowledge bases[J]. Knowledge and Data Engineering, IEEE Transactions on, 1996, 8(3): 353-372.
- [5] Yang J B, Liu J, Wang J, et al. Belief rule-base inference methodology using the evidential reasoning approach-RIMER[J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2006, 36(2): 266-285.
- [6] Yang J B, Xu D L. Intelligent decision system via evidential reasoning[J]. IDSL, Cheshire, England, 1999.
- [7] Liu J, Yang J B, Wang J, et al. Engineering System Safety Analysis and Synthesis Using the Fuzzy Rule - Based Evidential Reasoning Approach[J]. Quality and Reliability Engineering International, 2005, 21(4): 387-411.
- [8] Xu D L, Liu J, Yang J B, et al. Inference and learning methodology of belief-rule-based expert system for pipeline leak detection[J]. Expert Systems with Applications, 2007, 32(1): 103-113.
- [9] Yang J B, Liu J, Xu D L, et al. Optimization models for training belief-rule-based systems[J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2007, 37(4): 569-585.
- [10] Dempster A P. A generalization of Bayesian inference[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1968: 205-247.
- [11] Shafer G. A mathematical theory of evidence[M]. Princeton: Princeton university press, 1976.
- [12] Wang C L, Yoon K S. Multiple attribute decision making[J]. Berlin: Spring-verlag, 1981.
- [13] Zadeh L A. Fuzzy sets[J]. Information and control, 1965, 8(3): 338-353.
- [14] Sun R. Robust reasoning: integrating rule-based and similarity-based reasoning[J]. Artificial Intelligence, 1995, 75(2): 241-295.
- [15] Chen Y W, Yang J B, Xu D L, et al. Inference analysis and adaptive training for belief rule based systems[J]. Expert Systems with Applications, 2011, 38(10): 12845-12860.

- [16] 常瑞, 王红卫, 杨剑波. 基于梯度法和二分法的置信规则库参数训练方法 [J][J]. 系统工程, 2007, 25: 287-291.
- [17] Zhou Z J, Hu C H, Yang J B, et al. Online updating belief rule based system for pipeline leak detection under expert intervention[J]. Expert Systems with Applications, 2009, 36(4): 7700-7709.
- [18] Chen Y W, Yang J B, Xu D L, et al. On the inference and approximation properties of belief rule based systems[J]. Information Sciences, 2013, 234: 121-135.
- [19] 李彬, 王红卫, 杨剑波, 等. 置信规则库结构识别的置信 K 均值聚类算法[J]. 系统工程, 2011, 29(5): 85-91.
- [20] Chang L, Zhou Y, Jiang J, et al. Structure learning for belief rule base expert system: A comparative study[J]. Knowledge-Based Systems, 2013, 39: 159-172.
- [21] 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.
- [22] 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.
- [23] Calzada A, Liu J, Wang H, et al. A New Dynamic Rule Activation Method for Extended Belief Rule-Based Systems[J]. Knowledge and Data Engineering, IEEE Transactions on, 2015, 27(4): 880-894.
- [24] 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.
- [25] Wang Y M, Yang J B, Xu D L, 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] Yang J B, Wang Y M, Xu D L, et al. Belief rule-based methodology for mapping consumer preferences and setting product targets[J]. Expert Systems with Applications, 2012, 39(5): 4749-4759.
- [27] 李彬. 置信规则推理方法及其在库存与生产运作管理中的应用[D]. 华中科技大学, 2012.
- [28] Tang D, Yang J B, Chin K S, et al. A methodology to generate a belief rule base for customer perception risk analysis in new product development[J]. Expert Systems with Applications, 2011, 38(5): 5373-5383.
- [29] Liu J, Chen S, Martinez L, et al. A belief rule-based generic risk assessment framework[M]//Decision Aid Models for Disaster Management and Emergencies. Atlantis Press, 2013: 145-169.
- [30] Kong G, Xu D L, Body R, et al. A belief rule-based decision support system for clinical risk assessment of cardiac chest pain[J]. European Journal of Operational Research, 2012, 219(3): 564-573.

- [31] Zhou Z G, Liu F, Jiao L C, et al. A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer[J]. Knowledge-Based Systems, 2013, 54: 128-136.
- [32] Zhang J, Yan X, Zhang D, et al. Safety management performance assessment for Maritime Safety Administration (MSA) by using generalized belief rule base methodology[J]. Safety Science, 2014, 63: 157-167.
- [33] Yang J B. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties[J]. European Journal of Operational Research, 2001, 131(1): 31-61.
- [34] Yang J B, Xu D L. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty[J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2002, 32(3): 289-304.
- [35] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [36] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the national academy of sciences, 1982, 79(8): 2554-2558.
- [37] Alonso J M, Magdalena L, González-Rodríguez G. Looking for a good fuzzy system interpretability index: An experimental approach[J]. International Journal of Approximate Reasoning, 2009, 51(1): 115-134.
- [38] 周志杰, 杨剑波, 胡昌华. 置信规则库专家系统与复杂系统建模[M]. 科学出版社, 2011.
- [39] Wang Y M, Yang J B, Xu D L. Environmental impact assessment using the evidential reasoning approach[J]. European Journal of Operational Research, 2006, 174(3): 1885-1913.
- [40] Jin Y, Von Seelen W, Sendhoff B. On generating FC 3 fuzzy rule systems from data using evolution strategies[J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 1999, 29(6): 829-845.
- [41] 非线性规划[M]. 高等教育出版社, 1990.
- [42] Kennedy J, Eberhart R C. A discrete binary version of the particle swarm algorithm[C]//Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1997, 5: 4104-4108.
- [43] Hu X, Eberhart R. Solving constrained nonlinear optimization problems with particle swarm optimization[C]//Proceedings of the sixth world multiconference on systemics, cybernetics and informatics. 2002, 5: 203-206.
- [44] Shi Y, Eberhart R. A modified particle swarm optimizer[C]//Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on. IEEE, 1998: 69-73.

- [45] Burkhard W A, Keller R M. Some approaches to best-match file searching[J]. Communications of the ACM, 1973, 16(4): 230-236.
- [46] Chávez E, Navarro G, Baeza-Yates R, et al. Searching in metric spaces[J]. ACM computing surveys (CSUR), 2001, 33(3): 273-321.
- [47] Ditterich T G. Machine learning research: four current direction[J]. Artificial Intelligence Magazine, 1997, 4: 97-136.
- [48] Schapire R E. The strength of weak learnability[J]. Machine learning, 1990, 5(2): 197-227.
- [49] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of computer and system sciences, 1997, 55(1): 119-139.
- [50] Breiman L. Bagging predictors[J]. Machine learning, 1996, 24(2): 123-140.
- [51] Zhou Z H, Wu J, Tang W. Ensembling neural networks: many could be better than all[J]. Artificial intelligence, 2002, 137(1): 239-263.
- [52] Witten I H, 威腾, Frank E, et al. 数据挖掘: 实用机器学习技术[M]. 机械工业出版社, 2006.
- [53] Dietterich T G, Bakiri G. Solving multiclass learning problems via error-correcting output codes[J]. Journal of artificial intelligence research, 1995: 263-286.
- [54] Xu L, Krzyżak A, Suen C Y. Methods of combining multiple classifiers and their applications to handwriting recognition[J]. Systems, man and cybernetics, IEEE transactions on, 1992, 22(3): 418-435.
- [55] Bahler D, Navarro L. Methods for combining heterogeneous sets of classifiers[C]//17th Natl. Conf. on Artificial Intelligence (AAAI), Workshop on New Research Problems for Machine Learning. 2000.
- [56] Dietterich T G. Ensemble methods in machine learning[M]//Multiple classifier systems. Springer Berlin Heidelberg, 2000: 1-15.
- [57] Al-Ani A, Deriche M. A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence[J]. Journal of Artificial Intelligence Research, 2002: 333-361.
- [58] Storn R, Price K. Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces[M]. Berkeley: ICSI, 1995.

致谢

两年半的研究生学习生涯转瞬即逝，回顾这段时光心中充满无限感慨。研究生期间我经历了成功与喜悦，也面对过各种磨练与挑战，最终获得了进步和成长。在论文完成的时候，谨向对我进行指导帮助的恩师以及一同学习生活的同学和朋友表示深深的谢意。

首先我要向我的导师王晓东教授和傅仰耿老师表示感谢，在论文的选题和研究内容以及方法细节上他们给予我许多重要的指导、帮助和启发。他们最初引领我进行学术研究，导师不仅有渊博的学术知识、严谨的治学态度和精益求精的钻研精神，还不断强调学术道德规范的重要性，使我能够恪守本心，专注于学术研究并不断努力奋斗。

感谢实验室的吴英杰副教授、王一蕾副教授和孙岚老师，是您们无私的奉献和努力，为我们提供了良好的实验室环境并营造浓厚的学术研究氛围。特别要感谢吴英杰副教授和傅仰耿老师，是您们给予我机会从而成为这个大家庭中的一员，在学习和研究中为了答疑解惑，在生活中传授我待人接物，为人处事的方法和道理，并给予我许多锻炼提高的机会。

感谢 BRB 小组的成员们。他们是杨隆浩、余瑞银学长；王韩杰、吴伟昆、方志坚、刘莞玲、李敏、林燕清和叶青青同学。每周的汇报讨论我们都一起分享自己所看所学所想，讨论自己遇到的困惑和难题；我们一起思考实验方法、互相进行实验验证，丰富了我的知识面并使我许多方法有了更全面、深刻的了解和认识。

感谢与我一同度过研究生生涯的实验室同学和舍友们。在这段短暂的时光里，我们一同求学、一同研究、一同生活，是你们让我的研究生生活变得丰富多彩，是你们给带给我来无尽的欢乐与许多难忘的回忆。

衷心感谢我的父母，在求学的日子里我们相聚胜少，感谢父母对我的支持和理解。感谢父母对我的爱、对我的关心和对我的帮助，是您们在我遇到困难挫折时带给我无限的动力和力量，让我勇往直前。愿您们永远健康快乐。

最后感谢我的母校福州大学对我的培养。

个人简历

姓名： 苏群
出生年月： 1991 年 2 月
性别： 男
籍贯： 福建福安

学习及工作经历：

2013. 09–2016. 03	福州大学数学与计算机科学学院	攻读硕士学位
2009. 09–2013. 06	福州大学数学与计算机科学学院	获学士学位

在学期间的研究成果以及发表的学术论文

在读期间已发表和录用的论文：

第一作者（2 篇）：

- 【1】 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.
- 【2】 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 计算机科学与探索 录用

参与的科研项目及成果：

- 【1】 国家杰出青年科学基金项目（70925004）
- 【2】 国家自然科学基金面上项目（71371053）
- 【3】 国家自然科学基金青年项目（61300026, 71501047）
- 【4】 福建省自然科学基金项目（2015J01248）
- 【5】 福建省教育厅科技项目（JA13036）