分 类 号 TP 18
收藏编号
学校代码 10386

密级 公开
学号 N160320067
编号

福州大学

学术型硕士研究生学位（毕业）论文

基于析取范式的置信规则库推理方法研究

学 科 专 业：计算机应用技术
研 究 方 向：智能决策与专家系统
研 究 生 姓 名：张婕
指 导 教 师 、职 称：傅仰耿 副教授
协 助 导 师 、职 称：
所 在 学 院：数学与计算机科学学院
答辩委员会主席签名：

二〇一九 年 一 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实施办法》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

基于析取范式的置信规则库推理方法研究

中文摘要

实际应用中常需要处理包括模糊不确定、不完整性、概率不确定等各类复杂的决策问题，解决该类问题常用的方法包括贝叶斯概率推理法、D-S（Dempster-Shafter）证据理论、模糊推理法、神经网络等，但这些方法往往仅能处理其中某类问题，无法同时处理具有多种不确定性的问题。为此，Yang 等人提出置信规则库 (Belief Rule-Base，BRB)推理系统，在传统规则库专家系统的基础上，引入分布式置信框架，使其具有处理各类不确定问题的能力。同时，为能更好地表达和处理知识，基于析取范式的置信规则库（Disjunctive Belief Rule-Base，DBRB）被引入用于解决实际问题。DBRB 采用析取的属性连接方式，能有效降低系统的规则数量，完成信息推理，因此，被广泛应用于多属性决策问题中。

然而，现有的 DBRB 研究仍存在许多问题。首先，现有的激活权重计算方法存在未考虑属性相关性和忽略激活方式变化的问题；其次，现有的 DBRB 构建方法存在过拟合和依赖主观因素的问题；最后，在分类问题上存在未合理利用数据信息完成规则库构建的问题。针对上述问题，本文提出基于容斥原理的析取置信规则库激活方法，利用聚类分析构建析取范式置信规则库，构建基于两阶段离散化（Two-Stage Discretization，TSD）的 DBRB 分类系统。具体工作如下：

（1）针对现有的 DBRB 激活权重计算方法未考虑激活方式的变化，且属性间存在互斥不相容的限制条件的问题。本文提出基于容斥原理的析取置信规则库激活方法，并在此基础上根据激活属性个数和重要性不同调整规则权重，获取新的激活权重计算方法。新的规则激活权重计算方式的有效性在输油管道检漏和桥梁风险评估中得到了验证。

（2）针对现有 DBRB 构建方法存在规则数过多导致过拟合、依赖专家经验等问题，本文采用聚类分析方法对数据输出结果进行聚类，分析数据的特征分布，从而获取构建 DBRB 所需的规则数、结果评价等级等参数，完成 DBRB 的构建。并完成输油管道检漏和桥梁风险评估的知识库构建证明本文方法的有效性。

（3）DBRB 具有较高的分类性能，为构建更合理有效的 DBRB 分类系统，本文提出基于两阶段离散化的 DBRB 分类系统构建方法。通过局部离散化和全局离散化方法，确定系统的特征属性、属性参考值以及规则数量等一系列参数，以完成 DBRB 的构建，并在 UCI 的公共分类数据集上验证了该方法的有效性。

关键词： 置信规则库；析取范式；容斥原理；聚类分析；离散化

Disjunctive belief rule-base inference methodology research

Abstract

In the application, we should deal with various complex decision-making problems frequently including fuzzy uncertainty, incomplete and probability uncertainty, etc. However, the common methods to deal with these problems include Bayesian inference, Dempster-Shafter theory, fuzzy reasoning and neural network, etc, which can only deal with one of the issues but not multiple uncertainties issues at the same time. Therefore, Yang et al. introduced the belief framework and proposed Belief Rule Based(BRB) on the basis of the traditional rule-base. The new rule-base system has the ability to deal with various uncertain problems. Meanwhile, in order to express and process knowledge better, Disjunctive Belief Rule Based(DBRB) was introduced to solve practical problems. DBRB can complete information reasoning effectively and reduce the number of rules in the system by using a disjunctive connection. Therefore, it is applied to the multiattribute decision problems widely.

However, there are many problems in existing DBRB studies. Firstly, the existing calculation methods of the activation weight for DBRB is based on probability theory but the correlation between attributes is not considered. In addition, the number and importance of attributes are more prominent in decision making because the activation mode is changed from rule activation to attribute activation. Secondly, there are some disadvantages including over-fitting and dependent on subjective factors in DBRB construction method. In the classification problem, data information is not used to determine the important parameters of the system. In view of the above problems, this paper proposes three methods: activation method for disjunctive belief rule base using the principle of inclusion and exclusion, Construction of Disjunctive Belief Rule-Base Based on Clustering, Constructing the DBRB classification system based on Two-Stage Discretization.

(1) For the existing calculation method of DBRB activation weight，it is not consider the change of activation method, and exists the problem of mutual exclusion and incompatibility between attributes. Based on combinatorial mathematics, this paper proposes a disjunctive confidence rule base activation method using the principle of inclusion and exclusion. The new formula adjusts rule weight according to the number and importance of the active attributes. The effectiveness of the new formula

has been verified in pipeline leak detection and bridge risk assessment.

(2) There are some issue in the DBRB construction methods, such as over-fitting and relying on expert experience. Therefore, this paper proposes to use the cluster for obtaining data distribution. Then, we can obtain the parameters such as rule number and consequence to completing the construction of DBRB. In the experiment, the rule-base for pipeline leak detection and bridge risk assessment was constructed to verify the effectiveness of the method.

(3) In order to build a more reasonable and effective DBRB classification system, this paper use two-stage discretization——local discretization and global discretization. TSD can determine the system's feature attribute, referential values of antecedent attribute and rule numbers jointly that can establishment DBRB completely. The effectiveness of the proposed method is verified on the UCI classification datasets.

Key words: Belief Rule-Base, Disjunctive Normal Form, Principle Of Inclusion And Exclusion, Clustering, Discretization

目 录

第一章 引言

1.1 课题背景与意义

知识就是力量，而专家系统的力量不仅来自于人类专家提供的知识内容，同时源于系统自身所能处理、运用以及学习的内容。自第一个专家系统 DENDRAL 于 1965 年问世以来，专家系统经过多年的发展已遍布各个研究领域，也被应用到人类的学习生活和工业生产中，在应用中得到更大的发展[1]。作为智能系统常见的一种形式，专家系统[2]是一种能够模拟人类专家决策方式，学习运用人类专家知识进行复杂问题处理，并在过程中不断学习更新自身知识库[3]的智能系统。随着信息技术的发展，待处理的问题呈现复杂化、数据量大的特点，而这些问题已不能单纯依靠人类专家人为处理，更多需要利用专家系统进行问题建模、知识表达和问题求解。

生产生活中，问题是复杂、多样且存在不确定性，智能系统为模仿人类处理信息的过程需要对信息进行融合，这是一个不确定性信息推理与决策的过程，而目前常用的方法有贝叶斯概率推理法、D-S（Dempster-Shafter）证据理论[4-5]、模糊理论[6]、神经网络[7]等。但是这些方法往往仅能处理某种特定的不确定性问题，无法对同时存在多种不确定性的问题进行求解。

因此，为了更好的处理各种不确定性问题，Yang 等人[8]于 2006 年提出基于证据推理的置信规则库推理方法（Belief Rule-Base Inference Methodology Using The Evidential Reasoning Approach，RIMER）。该方法基于 IF-THEN 规则库系统，通过引入置信框架，并设置相关的属性、权重和置信度等参数对传统的规则库系统进行扩展得到置信规则库（Belief Rule-Base，BRB），以完成对知识的表达。除了分布式的知识表达形式外，在推理机部分，还引入模糊理论、D-S 证据理论、决策理论[9-10]等，使 BRB 具有处理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数据的能力。其推理过程主要分为以下三个步骤：首先，将输入数据与相关规则进行知识匹配；其次，利用证据推理算法对匹配的规则进行合成；最后，对合成的信息进行转化，使其符合用户要求。

置信规则库不仅在知识的表达上使用了分布式框架表达知识的不确定性，同时在推理过程中考虑了数据的不确定性，且选用的推理算法具有不确定数据处理和学习能力，相比于传统的模糊逻辑推理和人工神经网络，能在参数数量少的情况下完成推理且过程简单易懂，具有一定的实用价值。因此，BRB 系统也已成功应用于各领域评估分析中，包括工程系统安全评估[11]、石墨成分分析[12]、输

油管道泄漏[13]、涡轮增压器可靠性预测[14]、武器装备军事能力评估[15]、软件评估[16]、图像纹理分类[17]、临床诊断[18]、智能交通信号灯[19]、消费预测等[20]，具有较高的实际应用价值。

作为 BRB 的另一种表达形式，基于析取范式的置信规则（Disjunctive Belief Rule-Base，DBRB）自 Yang 提出后，未有学者对其进行研究应用，直到 2016 年，Chang 等人[21]首次将 DBRB 应用到分类问题上，并成功解决 BRB 在多属性多参考值决策问题中，规则数量随属性、属性参考值数量增长呈现指数级增长的问题——“组合爆炸”。因此，在不影响系统推理性能的前提下，DBRB 通过改变属性的组合方式对 BRB 系统结构进行优化，减少系统中的规则数量和参考值数量，大大缩短规则遍历时间和推理复杂性。

1.2 研究现状

传统的 BRB 系统建立在专家知识上，BRB 中的参数均由领域专家或者决策者根据经验知识决定。但是，仅仅利用专家知识来确定 BRB 的参数是困难的，也是不准确的。决策过程中，系统参数的微小偏差，都将给最终决策带来不同程度的影响。因此，Yang 等[12]提出 BRB 的参数优化模型，在满足各种不等式、等式约束条件，通过动态调整 BRB 的参数后进行模型构建完成决策，以最小化系统输出和实际输出的误差为目标函数，选取合适的参数取值，提高 BRB 系统的推理能力；Chen 等[22]基于 Yang 提出的 BRB 参数学习模型，提出利用 MATLAB 自带的优化工具箱对初始的 BRB 系统进行参数训练，该方法在 fmincon 函数的基础上构建训练模型进行参数训练，但是算法的训练时间过长且复用性不高；Chang 等[23]提出利用梯度下降和二分法策略进行参数学习方法，该方法优于 fmincon 函数，但训练广度不足，算法的收敛速度和精度还有待提高；因此，Wu 等人[24]提出加速梯度下降的方法并在参数训练过程中考虑了更多参数；Zhou 等[25-26]基于贝叶斯推理和极大似然估计方法提出基于置信规则库的在线参数学习方法，能实时更新系统参数完成决策，但该方法依赖人为假定的概率分布，对实际应用范围有所局限；Su 和 Wang 等[27-28]相继提出基于群智能算法的参数学习方法以及基于专家干预的差分进化算法，引入群智能对 BRB 进行参数训练提高了 BRB 的推理准确性，但群智能学习过程需要反复迭代，收敛速度慢、效率低。除了在参数优化上的研究，为了使 BRB 能够更好的解决多属性决策问题，学者也开始关注 BRB 系统结构优化的问题。因为传统的 BRB 系统构建时，采用遍历组合的方式构建规则，系统需要覆盖所有的前提属性和属性的参考值的组合情况，因此，每增加一个属性参考值系统规则数呈指数级增长，当属性个数和参考

值数量过多时容易出现“组合爆炸”的问题。鉴于此，Chang 等[29]针对 BRB 的结构优化进行研究，基于多维尺度变换（Multidimensional Scaling，MDS）、灰靶理论（Grey Target，GT）、主成分分析（Principle Component Analysis，PCA）等特征提取的方法，提取系统特征属性以减少参与决策的属性数量；Yang 等[30]提出基于关联系数标准差融合的置信规则库的约简方法，但该方法依赖人为确定的关联属性阈值和评价矩阵，具有一定的局限性；Wang 等[31]提出利用粗糙集和密度聚类等方法对置信规则库结构做进一步研究。以上关于 BRB 的优化方法仅考虑了参数或结构的单方面优化，未从整体上进行 BRB 的研究探索且上述方法大多采用遍历组合的系统构建方法，一方面存在过拟合的问题，另一方面对于分类问题等多属性决策研究将造成系统推理效率低下等问题。因此，Zhou 等[32]提出基于“统计效用”规则库构建方法，随机生成规则，并计算每条规则的效用值，根据效用值大小动态增删系统中的规则完成系统构建；Wang 等[33]提出动态规则自适应的方法，通过对系统过拟合和欠拟合的判定，优化 BRB 系统的规则数量和参数设置。除了对 BRB 参数及结构的优化，Wang 和 Ye 等[34-35]也提出了 RIMER 推理过程中包括个体匹配度、属性权重、激活权重计算在内的公式改进方法，新的计算公式能带来推理准确率的明显提升。除了对传统 BRB 的研究，有学者发现对 BRB 进行扩展和改进得到新形式的 BRB，新的知识表达形式能被应用于解决实际问题，并获得较高的推理性能。例如，Liu 等[36]在置信规则的前件部分引入分布式置信框架，使前后件同时具有表达不确定信息的能力，从而提出扩展置信规则库(Extended Belief Rule-Base，EBRB)系统。EBRB 更形象的表达生活中的不确定信息且其采用数据驱动的方式将数据与规则一一对应，充分利用数据的已知信息，无需进行参数训练并具有较好的决策准确性，但当数据过于庞大时，系统构建的规则数量较多则在知识匹配阶段需要搜索大量的规则，因此，针对 EBRB 的上述问题许多专家学者也展开了研究讨论[37-40]。

虽然，现有针对 BRB 结构优化的研究在一定程度上已经能有效避免“组合爆炸”的问题，但是，在一定程度上仍造成系统的信息损失，影响决策。因此，为了不损失系统精度，Chang 等[21]采用 Yang 所提的基于析取范式的置信规则库推理系统（Disjunctive Belief Rule-Base，DBRB）从根本上解决 BRB 的“组合爆炸”问题。DBRB 将规则连接方式由合取‘∧’改成析取‘∨’，改变属性参考值的组合方式，每个属性参考值只需遍历一次，有效解决了因属性参考值增多规则呈指数增长的问题，对 BRB 的结构进行了优化并成功应用在分类问题中；Ye 等人[35]对于 DBRB 在分类问题上的应用提出当系统规则数设置为分类数时，DBRB 具有最佳的分类推力性能，并对个体匹配度公式进行改进，避免出现“零激活”的情况；Chang 等人[41]提出基于赤池信息量化标准的规则数和规则参数联合优化方

法，通过枚举指定的规则数量，并以最小训练误差为标准动态生成规则库系统；Yang 等人[42-43]利用霍夫丁不等式推导 BRB 系统的泛化误差，提出规则数量和系统参数的联合优化模型，该模型可同时用于 CBRB 和 DBRB 的构建，并在桥梁风险评估中得到应用，较其他方法取得了更高的推理准确度。

从 BRB 目前的研究现状可知，相比于传统的 BRB 和 EBRB，DBRB 在结构和规则匹配上具有一定的优势，用 BRB、EBRB、DBRB 共同表达相同的知识内容，DBRB 结构更简单且构建的规则数较少，在构建初期，能大规模的缩减系统冗余的规则，更适合非线性复杂系统的建模，同时减少系统参与训练的参数个数，能有效提高参数训练的效率和缩短规则匹配的时间。然而现有对于 DBRB 的研究仍存在许多不足：首先，由 BRB 到 DBRB 的转变，RIMER 的推理方法由规则激活变成属性激活，虽然 Chang 在应用 DBRB 解决分类问题时提出新的规则激活权重计算方式，但是由于被激活的规则对应的属性参考值不一定全部被激活，所以，需要重新考虑被激活规则的规则权重以及规则激活权重的计算问题；其次，现有 DBRB 的构建方式一般有两种，一种是基于专家给定参考值进行线性组合，另一种是随机生成初始规则库动态增删系统规则，因此，无法在构建规则库前先确定系统规模；最后，DBRB 的决策能力在规则数量相同的前提下，一般高于 BRB 和 EBRB，特别是在多属性决策中，如分类问题上。因此，如何根据实际应用场景有效的构建 DBRB 系统、确定系统参数，并应用于分类等问题也成为 DBRB 的研究重点之一。

1.3 论文主要内容

本文主要围绕 DBRB 在规则激活权重优化、DBRB 构建方法以及如何有效构建 DBRB 分类系统等问题进行研究讨论，主要研究内容可以分为以下三个方面：

第一、基于容斥原理的析取范式置信规则库激活方法。针对现有的 DBRB 激活权重计算方式忽略实际应用中属性的相关性和激活方式改变的问题，本文采用基于容斥原理的析取置信规则库激活方法，该方法基于容斥原理对 Yang 方法进行重定义，考虑属性间的相关性，并在此基础上根据激活属性个数和重要性不同调整规则权重，进而获取新的激活权重计算方法。新的激活权重计算方式能有效提高系统的推理精度。

第二、利用聚类分析构建析取范式置信规则库。现有的规则构建方法大多采用属性遍历组合或线性组合的方法，需要基于专家给定参考值或参考值个数的情况下才能确定系统规则数，对于大型系统无法准确提供所需的参考值相关参数，

则对系统构建造成影响。另外，采用动态的规则构建方法无法在构建初期确定系统规则数，需要动态增删规则，规则数量上下界不明显，且对选择的指标依赖性较大。为此，本文提出基于聚类分析的 DBRB 构建方法，利用聚类分析确定系统规则数量和结果评价等级，完成 DBRB 系统的构建。

第三、构建基于 TSD 的 DBRB 分类系统。采用传统的 BRB 解决分类问题，常由于属性和参考值数量过多，容易引起“组合爆炸”的问题。因此，本文提出基于两阶段离散化（Two-Stage Discretization，TSD）[44]的 DBRB 构建方法，通过对数据进行两次连续离散化处理，提取出特征属性和对应的参考值并用于 DBRB 构建，不仅能有效确定 DBRB 规则数量，同时能够确定系统重要参数并对 DBRB 进行结构优化，有效应用于分类问题。

1.4 论文组织结构

本文的主体内容由六个章节组成，其中，论文第三、四、五章为本文的主要工作。论文的第四、五章分别采用聚类和离散化的方法构建不同问题的 DBRB 系统，并在推理过程中采用第三章所提的激活权重计算方法完成推理，并与引言和基础知识共同构成论文主体内容，具体的组织结构如下所示：

第一章为论文的引言部分，主要介绍置信规则库（BRB）的研究背景和意义，讨论了置信规则库系统国内外的研究现状，包括扩展置信规则库（EBRB）和基于析取范式的置信规则库（DBRB），并着重介绍了 DBRB 目前的研究现状，以及本文的主要研究内容与组织结构。

第二章介绍了置信规则库（BRB）和基于析取范式的置信规则库（DBRB）的基础知识，包括知识的表达形式以及 RIMER 的详细推理过程，以及 BRB 和 DBRB 之间的异同点，为后面对 DBRB 的研究打下基础。

第三章对 DBRB 的激活权重计算进行研究，提出基于容斥原理的析取置信规则库激活方法。该激活权重计算方法不仅考虑到激活方式的变化，同时考虑属性间的相互联系，以容斥原理为理论背景完成激活权重的计算更符合实际应用。

第四章对基于析取范式的置信规则库的构建问题进行研究，提出利用聚类分析构建析取范式置信规则库，该方法通过对数据的输出结果进行聚类获取输出数据的特征分布情况，从而确定系统的规则数量，能够有效地确定系统规模，完成 DBRB 构建。

第五章针对置信规则库在分类问题上的应用进行研究，提出 DBRB 分类系统的构建方法。通过引入 TSD 方法获取系统特征属性和属性参考值，确定系统规则数量和重要参数，完成 DBRB 系统的构建，并选取 UCI 上的公共分类数据

集进行验证，均取得较为理想的分类效果。

论文的最后部分是对本文研究工作的总结与展望，对本文所做的研究工作进行概括，同时对未来该领域的工作做进一步的展望。

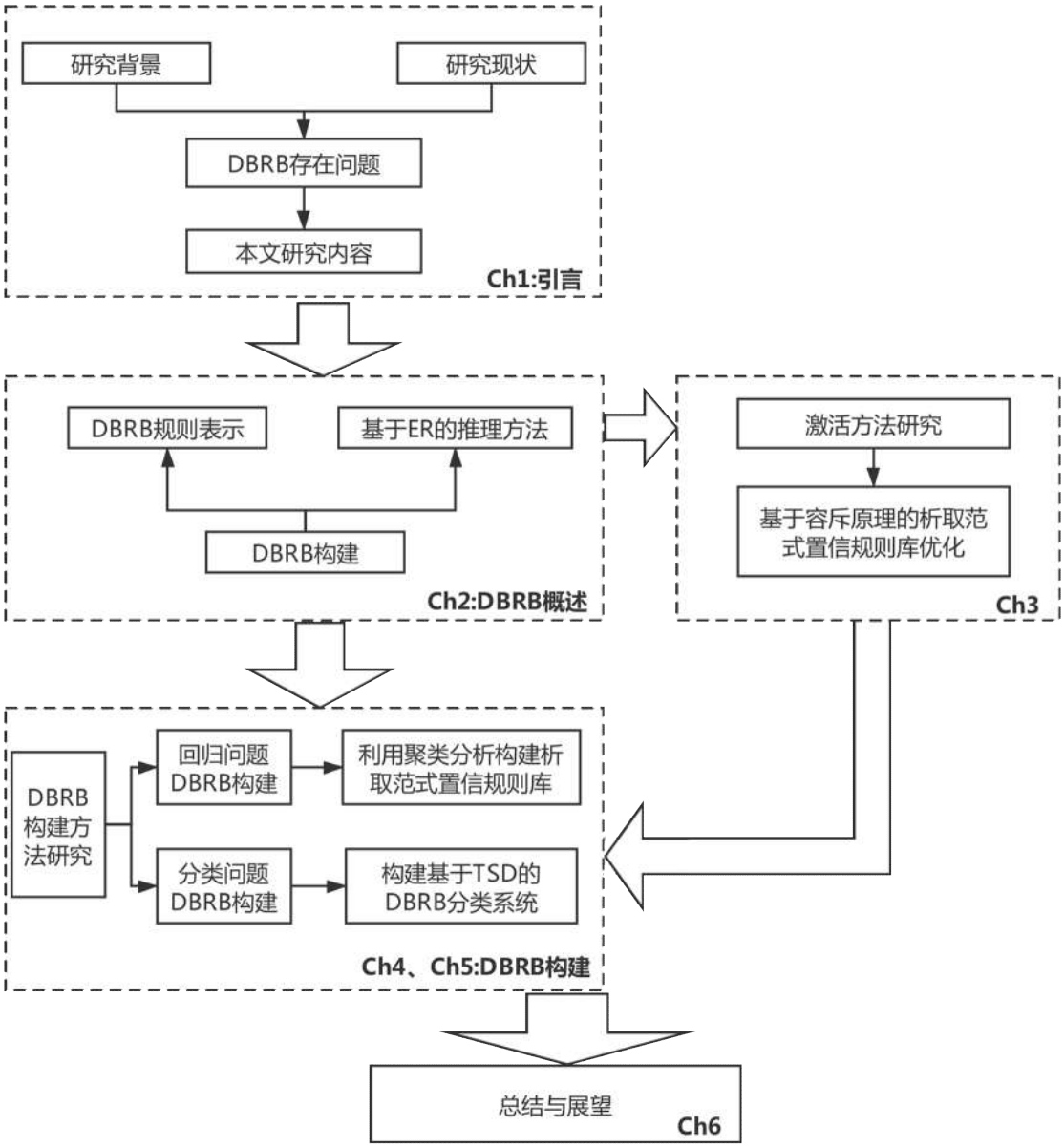

图 1-1 论文组织结构图

第二章 基于析取范式的置信规则库相关知识

2.1 引言

为有效解决不同环境下概率不确定性、模糊不确定性以及非线性特征数据的建模问题，Yang 等人[8]于 2006 年提出 RIMER 方法，该方法结合 D-S 证据理论[4-5]、决策理论[9]、模糊集理论[6]等，在传统的 IF-THEN 规则基础上，引入分布式框架，对结果部分进行扩展，使其具有表达和处理不确定性问题的能力。因此，许多专家学者对置信规则库（Belief Rule-Base，简称 BRB）展开研究讨论。根据知识的不同表达形式，置信规则库大致分为三类：基于合取范式的置信规则库（Conjunctive Belief Rule-Base，简称 CBRB）也就是传统的 BRB、基于析取范式的置信规则库（Disjunctive Belief Rule-Base，简称 DBRB）[8]和扩展置信规则库（Extended Belief Rule-Base，简称 EBRB）[36]。目前，三种方式的规则表达形式均被广泛应用在实际应用中。

本章将重点介绍 CBRB 和 DBRB 的相关知识，包括规则表达形式、规则构建以及推理方法。比较两者的共性和差异，为后续对基于析取范式的置信规则库的研究提供理论依据。

2.2 置信规则库

2.2.1 置信规则库的表示

IF-THEN 作为知识库框架的一种，通过收集专家或领域知识，以规则的形式，利用前件信息，推导出对用户有用的结果，其通用形式如下：

$$R_k : if\ A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k \text{ then } \mathrm{D}_k \qquad \text{公式（2-1）}$$

其中，$A_i^k (i = 1,2,\cdots,T_k)$ 表示第 i 个属性在第 k 条规则中的属性参考值，T_k 表示属性个数；$\mathrm{D}_k \left(D_k \in D, D=\{D_n; n=1,2,\cdots,N\} \right)$ 表示第 k 条规则的推理结果；属性间可以采用‘∨’或者‘∧’作为连接符。

采用该形式表达的知识，利用前件属性推导得到的是一个准确的结果，例如：如果下雨且心情不愉悦那么不去郊游。结合实际生活情况可以看出，通过下雨和心情愉悦程度两个条件，对人类的活动进行预测是一个不确定的事情，只可以推

断出，下雨且心情不愉悦，不郊游的概率较大，但不能获得确定的结论。

可以看出传统的 IF-THEN 规则在表达不确定、模糊、不完整的知识时存在一定的弊端，因此，Yang 等人[8]提出置信规则库的知识表示方法。该方法由传统的 IF-THEN 规则扩展而来，在结果部分引入分布式置信框架，同时，增加前提属性权重和规则权重表示属性和规则的重要性等，使 BRB 具有处理模糊、不确定、不完整数据的能力。具体规则表示如下：

$$R_{\mathrm{k}}:$$

$$if\ A_1^{\mathrm{k}} \wedge A_2^{\mathrm{k}} \wedge \cdots \wedge A_{T_{\mathrm{k}}}^{\mathrm{k}}\ then\left\{\left(D_1,\overline{\beta}_{1,k}\right),\left(D_2,\overline{\beta}_{2,k}\right),\cdots,\left(D_N,\overline{\beta}_{N,k}\right)\right\} \quad \text{公式（2-2）}$$

$$with\ a\ rule\ weight\ \theta_k\ and\ attribute\ weight\ \delta_{1,k},\delta_{2,k},\cdots,\delta_{T_k,k}$$

其中，$R_{\mathrm{k}}\left(k=1,2,\cdots,L\right)$ 表示第 k 条规则，L 表示系统中的规则总数；$D_j\left(j=1,2,\cdots,N\right)$ 表示规则的结果评价等级，该规则库系统共有 N 个结果评价等级；$\overline{\beta}_{j,k}\left(j=1,2,\cdots,N,k=1,2,\cdots,L\right)$ 表示第 k 条规则的输出结果在第 j 个评价等级上的结果置信度，$0\le\sum_{j=1}^{N}\overline{\beta}_{j,k}\le 1$。当 $\sum_{j=1}^{N}\overline{\beta}_{j,k}=1$ 时，表示该条规则包含完整的信息，否则表示信息不完整。$\theta_k\left(k=1,2,\cdots,L\right)$ 为第 k 条规则的权重，表示该条规则相对其他规则的重要性；$\delta_{i,k}\left(i=1,2,\cdots,T_k,k=1,2,\cdots,L\right)$ 表示第 i 个前提属性相对其他属性的重要程度，即属性权重[8]。此外，规则中属性的连接方式可以是“∧”或者“∨”，分别对应基于合取范式的置信规则库（Conjunctive Belief Rule-Base，简称 CBRB）、基于析取范式的置信规则库（Disjunctive Belief Rule-Base，简称 DBRB）。由于“∧”更常应用于知识的表达中，因此，通常研究中所指的 BRB 即为 CBRB。

2.2.2 置信规则库的推理方法

BRB 的推理方法基于 RIMER，该方法以证据理论（Evidence Reasoning，简称 ER）[45]为理论依据，将输入值所激活的规则进行合成，从而获取推理结果。其主要推理过程大致分为三个步骤：（1）激活权重计算；（2）置信度修正；（3）激活规则合成。

2.2.2.1 激活权重计算

BRB 可以处理带有各种不确定性的问题，实现复杂决策问题的建模。因此，对于输入信息的不确定性，需要在输入值和属性参考值之间做一定的转化，即输入信息的个体匹配度。

以数值型输入为例，假设存在输入信息 x，$x_i\left(i=1,2,\cdots,T_k\right)$ 为 x 的第 i 个分量，则对于输入 x 可以转化为如下分布式形式：

$$S(x)=\left\{\left(A_{i,j},\alpha_{i,j}\right),i=1,2,\cdots,T_k,j=1,2,\cdots,J_i\right\} \qquad \text{公式（2-3）}$$

其中，$\alpha_{i,j}\left(i=1,2,\cdots,T_k,j=1,2,\cdots,J_i\right)$ 表示输入分量 $x_i\left(i=1,2,\cdots,T_k\right)$ 对属性参考值 $A_{i,j}$ 的个体匹配度，J_i 表示第 i 个属性的参考值个数，计算方式如下：

$$\begin{cases}\alpha_{i,j}=\dfrac{A_{i,j+1}-x_i}{A_{i,j+1}-A_{i,j}}, & A_{i,j}\le x_i\le A_{i,j+1}\\ \alpha_{i,j+1}=\dfrac{x_i-A_{i,j}}{A_{i,j+1}-A_{i,j}}, & A_{i,j}\le x_i\le A_{i,j+1}\\ \alpha_{i,j}=0\ , & otherwise\end{cases} \qquad \text{公式（2-4）}$$

则，输入 x 对第 k 条规则的激活权重为：

$$\omega_k=\frac{\theta_k\prod_{i=1}^{M}(\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L}\theta_l\prod_{i=1}^{M}(\alpha_i^l)^{\bar{\delta}_i}},\ \bar{\delta}_{ik}=\frac{\delta_{ik}}{\max\limits_{i=1,2,\cdots,T_k}\{\delta_{ik}\}} \qquad \text{公式（2-5）}$$

其中，$\omega_k\in[0,1]$，$k=1,2,\cdots,L$，由于 BRB 以“∧”作为属性间的连接方式，所以个体匹配度间采用累乘的计算方式，$\bar{\delta}_{ik}$ 为归一化后的属性权重。

2.2.2.2 置信度修正

对于结果评价等级置信度 $\bar{\beta}_{j,k}\left(j=1,2,\cdots,N,k=1,2,\cdots,L\right)$，当 $\sum_{j=1}^{N}\bar{\beta}_{j,k}$=1 时，表示该条规则是完整的，否则，说明该条规则不完整。由于输入数据的模糊性和不完整性将影响结果的推理，因此，需要对结果评价等级置信度进行修正：

$$\beta_{j,k}=\bar{\beta}_{j,k}\frac{\sum_{i=1}^{T_k}\left(\tau(i,k)\sum_{t=1}^{J_t}\alpha_{it}\right)}{\sum_{i=1}^{T_k}\tau(i,k)} \qquad \text{公式（2-6）}$$

$$\tau(i,k)=\begin{cases}1\ ,\ A_i\in R_k(i=1,2,\cdots,T_k)\\ 0\ ,\ \text{otherwise}\end{cases}$$

当输入数据完整时，$\beta_{j,k}=\bar{\beta}_{j,k}$。

2.2.2.3 ER 推理合成

当规则被输入数据激活后，采用 ER 推理合成激活规则，从而获得推理结果。

对于每个结果置信度$\beta_{j,k}$，将其转换成对应的基本概率值，可得：

$$m_{j,k}=\omega_k\beta_{j,k} \quad \text{公式（2-7）}$$

$$m_{D,k}=1-\omega_k\sum_{j=1}^{N}\beta_{j,k} \quad \text{公式（2-8）}$$

$$\bar{m}_{D,k}=1-\omega_k \quad \text{公式（2-9）}$$

$$\tilde{m}_{D,k}=\omega_k\left(1-\sum_{j=1}^{N}\beta_{j,k}\right) \quad \text{公式（2-10）}$$

其中，$m_{j,k}$表示相对于D_j的基本概率设置；$\bar{m}_{D,k}$表示第 k 条规则由激活权重引起的未分配值；$\tilde{m}_{D,k}$表示由第 k 条规则评价结果不完整引发的基本概率值，且$m_{D,k}=\bar{m}_{D,k}+\tilde{m}_{D,k}$，表示相对于结果评价等级的基本概率分布。

随后，采用 Dempster 准则对激活规则进行合成，过程如下：

$$C_j=k\left[\prod_{l=1}^{L}\left(m_{j,l}+\bar{m}_{D,l}+\tilde{m}_{D,l}\right)-\prod_{l=1}^{L}\left(\bar{m}_{D,l}+\tilde{m}_{D,l}\right)\right] \quad \text{公式（2-11）}$$

$$\tilde{C}_D=k\left[\prod_{l=1}^{L}\left(\bar{m}_{D,l}+\tilde{m}_{D,l}\right)-\prod_{l=1}^{L}\bar{m}_{D,l}\right] \quad \text{公式（2-12）}$$

$$\bar{C}_D=k\prod_{l=1}^{L}\bar{m}_{D,l} \quad \text{公式（2-13）}$$

$$k^{-1}=\sum_{j=1}^{N}\prod_{l=1}^{L}\left(m_{j,l}+\bar{m}_{D,l}+\tilde{m}_{D,l}\right)-(N-1)\prod_{l=1}^{L}\left(\bar{m}_{D,l}+\tilde{m}_{D,l}\right) \quad \text{公式（2-14）}$$

$$\beta_j=\frac{C_j}{1-\bar{C}_D}(j=1,...,N) \quad \text{公式（2-15）}$$

$$\beta_D=\frac{\tilde{C}_D}{1-\bar{C}_D} \quad \text{公式（2-16）}$$

根据上述公式，可计算输入x对应的每个输出结果评价等级D_j的置信度β_j，相应的分布式输出如下：

$$f(x)=\{(D_j,\beta_j),j=1,2,...,N\} \quad \text{公式（2-17）}$$

上述迭代算法基于 ER 推理，但计算复杂，为了方便应用于工程系统中，Wang 和 Yang 等对该方法进一步归纳[46]，提出 ER 解析算法，该方法可对 BRB 中所有规则进行一次性组合，并得到 BRB 的最终输出$S(x)$:

$$S(x)=\{(D_j,\hat{\beta}_j),j=1,2,..,N\} \quad \text{公式（2-18）}$$

其中，$\hat{\beta}_j$ 表示输入 x 相对的输出结果评价等级 D_j 置信度，则

$$\hat{\beta}_j=\frac{\mu\times\left[\prod_{k=1}^{L}(\omega_k\beta_{j,k}+1-\omega_k\sum_{i=1}^{N}\beta_{i,k})-\prod_{k=1}^{L}(1-\omega_k\sum_{i=1}^{N}\beta_{i,k})\right]}{1-\mu\times\left[\prod_{k=1}^{L}(1-\omega_k)\right]} \quad \text{公式（2-19）}$$

$$\mu=\left[\sum_{j=1}^{N}\prod_{k=1}^{L}(\omega_k\beta_{j,k}+1-\omega_k\sum_{i=1}^{N}\beta_{i,k})-(N-1)\prod_{k=1}^{L}(1-\omega_k\sum_{i=1}^{N}\beta_{i,k})\right]^{-1} \quad \text{公式（2-20）}$$

当所求问题为数值型问题时，假设 D_j 对应的的效用值为 $\mu(D_j)$，则 BRB 系统的输出为：

$$f(x_t)=\sum_{j=1}^{N}\left(\mu(D_j)\beta_j\right)+\frac{(\mu(D_1)+\mu(D_N))}{2}\left(1-\sum_{j=1}^{N}\beta_j\right) \quad \text{公式（2-21）}$$

2.3 基于析取范式的置信规则库

2.3.1 基于析取范式的置信规则库表示

Yang 等人在文献[8]中提到 BRB 的属性连接方式可以有两种：“∧”或者“∨”，分别对应基于合取范式的置信规则库（Conjunctive Belief Rule-Base，简称 CBRB）和基于析取范式的置信规则库（Disjunctive Belief Rule-Base，简称 DBRB）。在专家给定参考值的情况下，CBRB 需要遍历所有参考值可能的组合情况进行规则库构建，该方法在多属性决策中容易因属性和参考值数量过多发生“组合爆炸”问题，而 DBRB 由于采用线性组合的方式，能有效减少系统规则数，但提出后近 10 年，未被应用到实际问题中。直到 2016 年， Chang 等人为 DBRB 提出新的激活权重计算方式并将其应用于分类问题中，得到突出的成效。因此，DBRB 开始被应用于工程系统中。其规则表达形式如下：

R_k：

$$\textit{if } A_1^k\vee A_2^k\vee\cdots\vee A_{T_k}^k \textit{ then}\left\{\left(D_1,\overline{\beta}_{1,k}\right),\left(D_2,\overline{\beta}_{2,k}\right),\cdots,\left(D_N,\overline{\beta}_{N,k}\right)\right\} \quad \text{公式（2-22）}$$

with a rule weight θ_k *and attribute weight* $\delta_{1,k},\delta_{2,k},\cdots,\delta_{T_k,k}$

对比公式（2-2）和公式（2-22）可知，除属性连接方式不同外，DBRB 的规则表达形式与 CBRB 相同，由属性前件和置信分布式结果构成，包括一系列

参数：属性参考值 $A_{i,j}$、结果评价等级 D_j、结果置信度 $\overline{\beta}_{j,k}$、规则权重 θ_k 以及属性权重 $\delta_{i.k}$。

2.3.2 基于析取范式的置信规则库构建

由于属性连接方式的不同，CBRB 和 DBRB 的规则构建方式也不同。如图 2-1 所示：

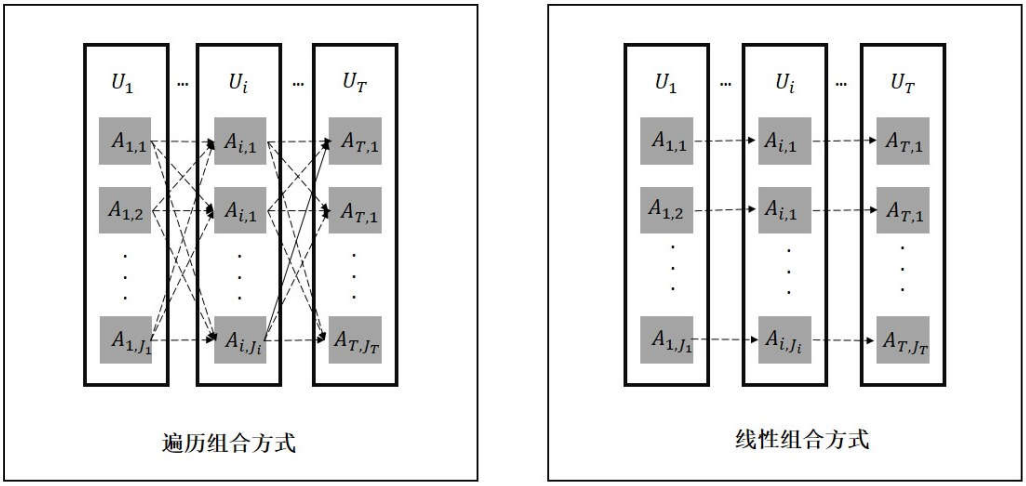

图 2-1 CBRB 和 DBRB 规则组合方式

当专家给定系统所需参数时，由于 CBRB 采用遍历组合的方式，属性参考值间一一组合，系统规则数：

$$\text{size}_{BRB}=\prod_{i=1}^{T}J_i \qquad \text{公式（2-23）}$$

由公式（2-23）可知，系统规则数随参考值数量增加呈现指数级增长。而对于 DBRB，由于属性间采用析取的连接方式，则系统无需重复遍历属性参考值，属性参考值采用线性组合的方式，系统的规则数由参考值个数最多的属性确定：

$$\text{size}_{BRB}=\max\left(J_i\right) \qquad \text{公式（2-24）}$$

例如，对于有 9 个属性的规则库系统，专家给定每个属性 5 个参考值，则 CBRB 的规则数将达到 5^9=1953125 个，而 DBRB 仅需构建 5 条规则即可完成规则库系统的构建。因此，按照专家给定参考值的方式构建规则库系统，相同参考值个数下，DBRB 构建的规则数量更少，能有效避免“组合爆炸”问题。

2.3.3 基于析取范式的置信规则库推理方法

与 CBRB 相同，DBRB 推理方法同样采用 RIMER 进行知识推理。不同的是，

由于属性连接方式的变化，在计算激活权重时，CBRB 采用累乘的方式进行计算，鉴于此，Chang 提出采用累加和的形式计算 DBRB 的激活权重：

$$\omega_k = \frac{\theta_k \sum_{i=1}^{M} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \theta_l \sum_{i=1}^{M} (\alpha_i^k)^{\bar{\delta}_i}}, \bar{\delta}_{ik} = \frac{\delta_{ik}}{\max\limits_{i=1,2,\cdots,T_k} \{\delta_{ik}\}} \quad \text{公式（2-25）}$$

由公式（2-5）和公式（2-25）可知，规则权重的计算基于概率论方法，因此，分别采用累乘和累加方法进行规则权重计算更符合规则表达形式。

除了规则表达和激活权重计算上的差异，DBRB 和 BRB 在推理过程中是相同的，均是采用 RIMER 方法对激活规则进行推理合成，并获取推理结果。

2.4 本章小结

本章主要对基于合取范式的置信规则库和基于析取范式的置信规则库的相关知识进行简要介绍。主要包括 CBRB 和 DBRB 的规则表达形式、规则构建、RIMER 推理方法的过程，以突出基于析取范式的置信规则库与传统基于合取范式的置信规则库的不同和优势，为后续章节对基于析取范式的置信规则库的研究内容提供理论依据。

第三章 基于容斥原理的析取置信规则库激活方法

3.1 引言

为有效解决多属性决策导致的“组合爆炸”问题，Chang 等人[21]首次将基于析取范式的置信规则用于解决实际问题，并提出新的激活权重计算方法。此后，关于 DBRB 的研究大多是在该公式基础上进行。如，Yang 等人[43]将属性参考值约束进行调整，提出 DBRB 动态参数优化模型，并将其应用到桥梁风险检测；Chang 等人[41]提出利用赤池信息量化准则作为 DBRB 的参数和结构联合优化模型的指标；Yang 等人[42]以最小训练误差为标准，通过霍夫丁不等式推导出 BRB 的泛化误差，作为 CBRB（以合取为连接方式的 BRB）和 DBRB 的规则库构建指标。上述关于 DBRB 的研究，激活权重计算方法均采用 Chang 等人[21]提出的累加和方法，该方法以概率论为理论基础，在 CBRB 激活权重计算的基础上，根据属性连接方法的不同将规则激活权重计算由累乘改成累加，通过累加和方式计算属性的联合影响因子从而获取规则激活权重，虽然形式简单，计算便捷，但仍存在以下两个问题：（1）属性连接方式变化导致激活方式由规则激活变成属性激活，激活规则中存在属性全激活与部分激活的情况，且激活属性重要性不同，对规则权重将产生不同程度的影响，而现有激活公式未体现该情况；（2）常用的激活权重计算方法以概率论为理论依据，当属性间以“∨”作为属性连接方式时，现有激活权重公式成立需满足前提属性间互斥不相容的条件，但实际应用领域研究中，属性间往往相互影响共同对决策产生作用。因此，累加和的激活权重计算方法在实际应用中存在一定的弊端。

针对上述问题，本章提出基于容斥原理的析取置信规则库激活方法。该方法在文献[8]的基础上进行改进，从容斥原理[47]的角度对属性间的相关关系进行解释，并在文献[8]公式的基础上，同时考虑属性间的相关关系和激活属性对规则合成的影响，新的激活权重计算方式更符合实际应用需求。为了说明方法的有效性，实验中选取输油管道检漏和桥梁风险评估两个经典的置信规则库实验研究案例对本章方法进行检验，从实验结果可知，基于容斥原理的析取置信规则库激活方法在规则数量较少的情况下，不仅不增加系统的时间复杂度，同时，相比现有方法能获得更高的推理准确性。

3.2 基于容斥原理的 DBRB 激活方法

现有的 DBRB 激活权重计算方法源于 CBRB，两者均以概率论为理论基础，因此需要满足属性间互斥不相容的原理，与实际应用场景相违背。而本章在 Yang 方法[8]的基础上进行改进，不仅解决属性间的相关关系问题，且考虑到激活方式变化，更符合实际应用需求。

3.2.1 DBRB 规则权重优化

假设仅考虑两个条件属性 A、B 的情况，每个属性有三个参考值分别为 $A_1, A_2, A_3, B_1, B_2, B_3$，构建的 BRB 中将包含 9 条规则，如图 3-1 中点所示。

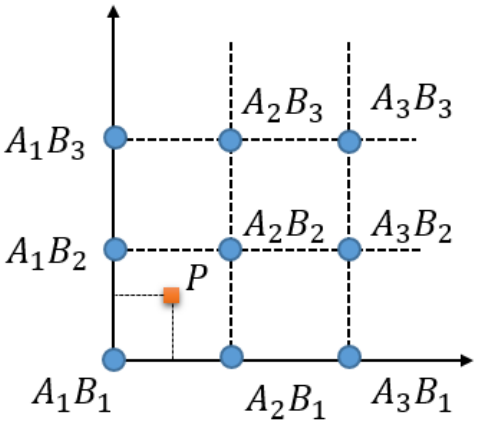

图 3-1 合取方式下的规则库

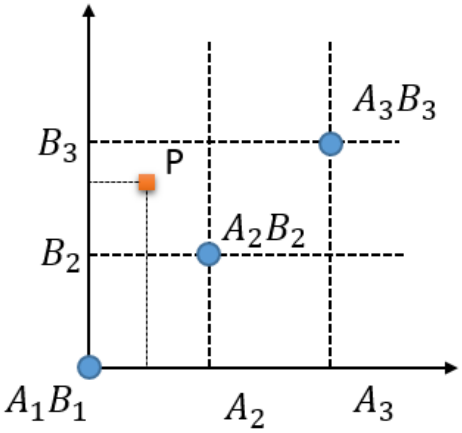

图 3-2 析取方式下的规则库

在 CBRB 中，对于输入值 $P(a,b)$，从图 3-1 可知，两个属性的输入值分别介于 A_1A_2、B_1B_2 之间，即对于 $P(a,b)$，存在 $A_1 \le a \le A_2$，$B_1 \le b \le B_2$。根据 RIMER

方法，规则 A_1B_1、A_1B_2、A_2B_1、A_2B_2 将同时被激活，因为 4 条规则的每个属性均被输入 P 激活。

而 DBRB 的激活情况如图 3-2 所示，相同情况下 CBRB 可以组合的规则共有 9 条，但 DBRB 因无需重复遍历属性参考值，因此，只需线性组合属性参考值，即可构建规则库系统。且由于属性间为“或”的组合关系，当输入值激活规则中的某一属性时，该条规则即被激活，即激活规则无需激活规则中的每个属性。假设构建的 DBRB 中包含 A_1B_1、A_2B_2、A_3B_3 三条规则，三条规则已包含所有属性参考值，则对于输入 $P(a,b)$，存在 $A_1 \le a \le A_2$，$B_2 \le b \le B_3$，在 CBRB 中，A_1B_2、A_1B_3、A_2B_2、A_2B_3 将被激活，而对于 DBRB，由于 A_1B_1 中 A_1 被激活，A_2B_2 中 A_2、B_2 被激活，A_3B_3 中 B_3 被激活，则对于该 DBRB，系统中的三条规则均被激活。不同的是，A_2B_2 中两个属性均被激活，而 A_1B_1、A_3B_3 中仅有一个属性被激活。因此，A_2B_2 被完全激活，可以得出，A_2B_2 在决策中发挥的推理作用相较于 A_3B_3、A_1B_1 更完整。对于规则权重 θ_k，如公式（2-25）所示，规则权重在计算激活权重时具有关键性作用。对于原始 BRB，规则激活时所有属性均被激活，则该条规则被完全激活，规则权重直接参与计算。而对于多属性的 DBRB，规则激活存在部分属性激活和完全属性激活的情况，则激活规则中的被激活属性个数、激活属性重要性都将造成该条规则对最终决策的影响，直接影响激活权重。因此，本章根据激活属性的个数和重要程度对参与激活权重公式计算的规则权重进行重新设计，即：

$$\theta_k' = \theta_k \frac{\sum_{j=1,\alpha_{k,j}\neq 0}^{M} \delta_{j,k}}{\sum_{i=1}^{M} \delta_{i,k}} \qquad \text{公式（3-1）}$$

公式（3-1）利用规则中激活属性的属性权重占比修正规则权重，有效区分规则的激活程度

3.2.2 DBRB 规则激活权重优化

在概率论中，当 $P(A \cap B) = P(A)P(B)$ 时，表示事件 A、B 相互独立，当 $A_1, A_2, \cdots, A_n$ 相互独立时，则有

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) \qquad \text{公式（3-2）}$$

古典概率具有规范性、非负性、有限可加性的特征，其中有限可加性表述为：假设事件 $A_1, A_2, \cdots, A_n$ 两两互不相容，则有：

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \qquad \text{公式（3-3）}$$

$$A \cup B = (1 - A) \cap B \cup A \qquad \text{公式（3-4）}$$

由上述概率论定义可知，激活权重计算公式（2-5）、（2-25）即是在公式（3-2）、公式 （3-3）的基础上演变的，因此需满足对应的条件，即保证属性间相对独立性或互不相容。而现实决策问题中，条件属性间的相关关系是复杂多变的，存在相互联系和影响。因此，Chang 等人[21]在激活权重计算上所提的方法，其可行性在现实问题研究中存在一定的局限性。

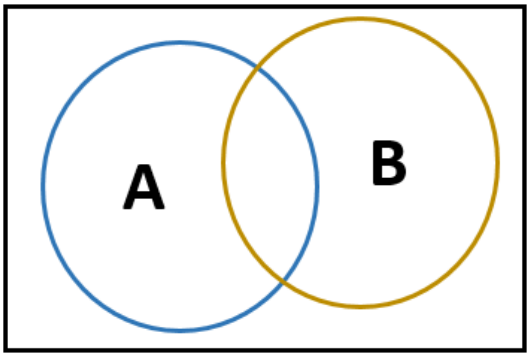

图 3-3 集合 A、B 相交

考虑到属性间可能存在相互联系，因此，从容斥原理的角度对激活权重公式重新定义。如图 3-3 所示，对于集合 A、B 并集的求解，若按式（3-3）进行计算，简单的相加将造成重叠部分的重复计算，同理，激活权重的计算若按照式（2-25）进行计算，则属性间相互影响的部分将被扩大。因此，当集合间存在重叠情况时，式（3-4）的计算方法能避免相交部分的重复计算，相应的属性间相互影响的情况与集合重叠情况类似，因此，采用多属性联合权重计算公式如下：

$$\sigma_{1,k} = h_{1,k} = \bar{\delta}_{1,k} \times \alpha_{1,k} \qquad \text{公式（3-5）}$$

$$\sigma_{i+1,k} = \sigma_{i,k} + (1 - \sigma_{i,k}) \times h_{i+1.k} \qquad \text{公式（3-6）}$$

$$\sigma_k = \sigma_{M,k} \qquad \text{公式（3-7）}$$

$$h_{j,k} = \bar{\delta}_{j,k} \times \alpha_{j,k} \qquad \text{公式（3-8）}$$

其中，$\sigma_{i+1,k} \in [0,1]$，$k = 1,2,\cdots,L$，$i = 1,2,\cdots,M-1$，表示第 k 条规则前 i+1 个属性的联合影响因子；$\alpha_{j,k}\left(k = 1,2,\cdots,L, j = 1,2,\cdots,M\right)$表示输入 x_j 相对于属性参考值 $A_{j,k}$ 的个体匹配度；$\bar{\delta}_{j,k}\left(k = 1,2,\cdots,L, j = 1,2,\cdots,M\right)$为对应属性权重归一化值：

$$\bar{\delta}_{i,k}=\frac{\delta_{i.k}}{\max\limits_{i=1,2,\cdots,M}\left\{\delta_{i.k}\right\}} \qquad \text{公式（3-9）}$$

新的规则激活权重公式由规则权重和联合影响因子共同组成：

$$\omega_k=\frac{\theta_k'\sigma_k}{\sum_{i=1}^{L}\theta_i'\sigma_i} \qquad \text{公式（3-10）}$$

该公式考虑到属性间可能存在相互联系，因此利用容斥原理的方法计算新加入属性与已计算属性共同作用下的影响因子，并排除已有属性对其的影响，从而避免属性间的重复影响被扩大，解决式（2-25）对属性互斥不相容的限制，更符合实际应用需求。

3.2.3 算法复杂度分析

针对所提方法进行算法复杂度分析，与实验部分相结合，共同说明本章方法在未提高复杂度的情况下可以获得更高的推理准确性（L：规则数；M：属性个数；N：结果评价等级个数；F：参数训练每次迭代的时间复杂度）。根据 RIMER 推理机制，假设对每个输入值获取个体匹配度的时间代价为$O(LM)$，可分析各个环节所需的时间代价，式（3-10）为$O(LM)$，式（2-19）为$O(LN)$。由于本章采用群智能算法进行迭代寻优，该优化算法根据迭代过程个体情况不同进行相应的优化，由于优化过程相同且时间分析较为复杂，故设每次迭代时间复杂度为$O(F)$，共迭代 P 次，则本章所提方法的算法复杂度为：

$$\begin{aligned}&\left(O(LM)+O(LM)+O(LN)+O(F)\right)*P\\&=O\left(P*\left(LN+LM+F\right)\right)\end{aligned} \qquad \text{公式（3-11）}$$

分析可知，采用式（3-10）或式（2-5）进行推理，具有相同的时间复杂度，说明与文献[21]方法相比，本章在优化激活权重计算的同时未提高算法的复杂度。

3.3 实验分析

为验证算法的有效性，本节选取输油管道检漏和桥梁风险评估进行实验室验证。输油管道检漏和桥梁风险评估作为 CBRB 和 DBRB 最经典的验证实验，被多次应用在不同方法中进行研究对比。同时，论文采用群智能算法作为求解 DBRB 参数训练模型的方法，设置种群规模 NP=100，迭代次数为 num=50000，运行 300 次取平均值作为最终结果。此外，实验环境为：Intel(R) Core i5-4570@

3.20GHz；4GB 内存；Windows 8 操作系统；算法使用 Visual C++实现。

3.3.1 输油管道检漏实验

输油管道检漏实验作为 BRB 最经典的研究对比实验已经被用于多个方法的研究验证。该实验以英国一条 100 多公里长的油气管道作为研究对象，通过管道的流量差异（FD）和平均压力差（PD）判断管道的泄漏大小（LS），从而检测管道的泄漏情况。

构建 BRB 系统时，根据专家经验，给定以下几个相关约束条件：（1）FD 包含 8 个参考值，$FD \in [-10,3]$；（2）PD 包含 7 个参考值，$PD \in [-0.042,0.042]$；（3）LS 包含 5 个值，$LS \in [0,8]$。利用专家给定的 FD 和 PD 初值，构建原始的 DBRB 系统，由公式（2-24）可知，系统共由 8 条规则构成，实验中，全部 2008 组数据作为测试数据并从中随机选取 500 组作为训练数据，然后用本章的方法计算激活权重，采用群智能的方法对构建的 DBRB 进行参数训练，以平均绝对误差（MAE）作为衡量指标，结果如图 3-4、图 3-5 所示：

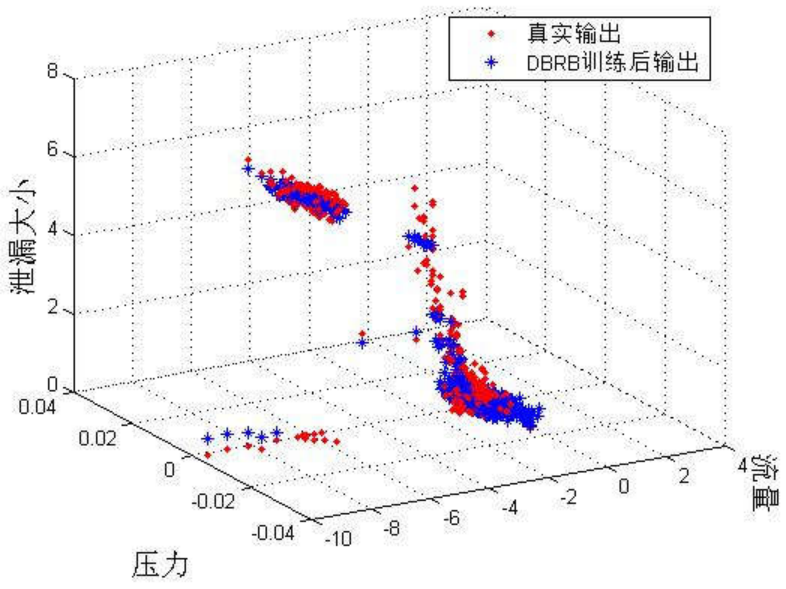

图 3-4 真实输出和 DBRB 训练后输出数据

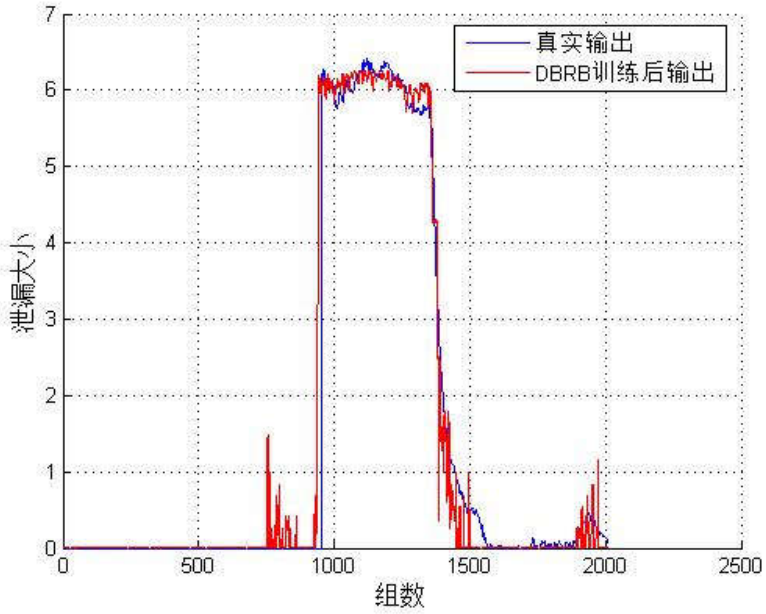

图 3-5 真实输出和 DBRB 训练后输出拟合效果

图 3-5 为训练后的 DBRB 输出与真实系统输出的拟合效果，从实验结果可看出，DBRB 系统的输出能较好的拟合真实数据。实验中，训练得到的 DBRB 系统性能最优可达到 MAE=0.1503，最差达到 MAE=0.1990，平均值为 MAE=0.1665。

为进一步说明本章方法的有效性，将改进后的方法与现有 BRB、EBRB、DBRB 方法进行比较，以平均绝对误差（MAE）作为主要衡量指标，规则数量和参数个数作为参考指标，比较不同结构、不同方法下 BRB、EBRB、DBRB 系统的推理性能。结果如表 3-1 所示：

表 3-1 不同类型方法下 BRB 推理性能的比较

方法	类型	规则数	参数个数	MAE
扩展置信规则库[36]	EBRB	900	18902	0.2169
加速梯度法[24]	BRB	56	358	0.1828
动态规则自适应[33]	BRB	6	43	0.2080
Chang-联合优化[41]	DBRB	5	40	0.1941
Yang-联合优化[42]	DBRB	3	22	0.1738
本章方法	**DBRB**	**8**	**70**	**0.1665**

从表 3-1 可知，与原有的 BRB、EBRB 相比，DBRB 可以在参数个数、规则数量较少的情况下获取更高的推理准确率，说明 DBRB 在不损失系统精度的情况下能有效的缩减系统规模；与现有的 DBRB 方法相比，本章采用容斥原理的激活权重计算方法，不仅能避免属性独立性限制，同时考虑了属性激活方式下，

激活属性对推理过程的影响，因此，相比 Chang-DBRB 和 Yang-DBRB，本章方法具有更高的推理准确性，相较于 Yang-DBRB，本章方法的平均绝对误差降低了 4%，说明本章所提方法的有效性。

3.3.2 桥梁风险评估实验

为验证本章方法具有一般性，除输油管道检漏实验，另外选取了桥梁风险评估实验对本章方法进行验证。桥梁作为陆路交通的重要枢纽，对社会生活经济发展具有重大影响。但由于老化、损坏、结构存在缺陷等原因，桥梁事故经常发生在生产生活中。因此，对桥梁进行准确的安全性评估对生产生活具有重要的现实意义和经济价值。

根据英国高速公路局提供的数据可知[48-50]，桥梁风险评估主要根据桥梁的安全性（SA）、功能性（FU）、可持续发展性（SU）以及环境因素（EN）对桥梁危险程度（RS）进行评估。以 SA、FU、SU、EN 作为前提属性并各设有 5 个参考等级，非常低（VL）、低（L）、中等（M）、高（H）、特高（VH）。每个属性的参考值等级语义相同，且具有相同的效用值范围，SA、FU、SU、$EN \in [-1,4]$，但每个属性对应的具体效用值可以是不同的；同样的，以 RS 作为评价结果，按专家经验 RS 设有五个等级，零风险（Z）、风险较低（S）、中等（M）、高风险（H）、特别危险（VH）。

$$\begin{aligned} RS &= \{Z,S,M,H,VH\} \\ &= \{0,25,50,75,100\} \end{aligned} \qquad \text{公式（3-12）}$$

本章以 SA、FU、SU、EN 作为前提属性，RS 作为决策属性，根据专家给定的参考值个数，构建由 5 条规则构成的 DBRB 系统，并以文献[43]中的初始化参数构建初始 DBRB 系统。从英国高速公路局提供的 23387 个桥梁结构维修工程中选取 506 组完整数据作为测试数据，并从 506 组数据中挑选最具代表性的 66 组数据进行学习。根据文献[48]所述，这 66 组数据经过特殊挑选，如果所构模型能对该 66 组数据进行较好的学习拟合，则对于其他桥梁项目的风险评估即能得到很好的预测，因此，以该 66 组数据作为训练数据，全部 506 组数据作为测试数据进行实验。用本章所提方法计算规则激活权重，并以平均绝对误差比例（MAPE）、均方根误差（RMSE）、相关系数（R）作为衡量指标，检验该方法的准确性。

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(f\left(x_t\right)-y_t\right)^2} \qquad \text{公式（3-13）}$$

$$MAPE=\frac{1}{T}\sum_{t=1}^{T}\left|\frac{f(x_t)-y_t}{y_t}\right| \qquad \text{公式（3-14）}$$

$$R=\frac{\sum_{t=1}^{T}\left(f(x_t)-\bar{f}\right)\left(y_t-\bar{y}\right)}{\sqrt{\sum_{t=1}^{T}\left(f(x_t)-\bar{f}\right)^2\cdot\sum_{t=1}^{T}\left(y_t-\bar{y}\right)^2}} \qquad \text{公式（3-15）}$$

$$\bar{f}=\frac{1}{T}\sum_{t=1}^{T}f(x_t)$$
$$\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t \qquad \text{公式（3-16）}$$

同时将本章方法与文献[42]、文献[48]中其他模型进行对比，如人工神经网络（ANN）、证据推理学习方法（ERL）、回归分析（MRA）以及其他 BRB 系统进行比较，结果如表 3-2 所示。

由表 3-2 可知，本章以及其他 BRB 模型在桥梁风险评估建模上均明显优于 ANN、ERL、MRA 等方法。与 Yang-BRB 比较，虽然本章方法在 MAPE 上未能优于 Yang-BRB，但在 RMSE 和 R 上均优于上述其他方法，说明本章方法具有更强的泛化能力；且 Yang-BRB 构建的系统规则数量达 48 条而本章方法仅通过 5 条规则即可获得上述推理结果，说明为达到相同精度的情况下，本章所需的推理系统结构更简单，需要的参数规模更小，性能更高。同时说明 DBRB 不仅能精简系统规模同时保证较高的推理准确性，说明本章方法的可行性。

表 3-2 DBRB 与其他模型的比较

模型	MAPE(%)	RMSE	R
BP-ANN	9.6294	4.1871	0.9834
ER1	22.4544	8.9255	0.9077
MRA3	18.5775	10.9527	0.8687
MRA8	23.9799	10.4653	0.8794
ER2	18.7808	11.2736	0.8918
MRA7	24.1941	10.3510	0.8796
MRA9	19.1456	11.3653	0.8904
Yang-BRB	**2.5999**	2.8060	0.9910
Yang-DBRB	6.7620	3.0916	0.9911
本章方法	4.8093	**2.5697**	**0.9943**

3.4 本章小结

本章针对现有的 DBRB 规则激活权重计算存在属性间互斥不相容的限制，引入容斥原理的方法，在此基础上根据激活属性个数和属性重要性对规则权重进行修正，从而优化现有激活权重计算方法。相比于现有的 DBRB 激活权重计算方法，本章方法可以避免属性间互斥不相容的限制条件，更具一般性，且考虑到规则激活方式的改变对推理造成影响，对规则权重进行修正，更具合理性和现实研究价值。在实验分析中，通过对输油管道检漏和桥梁风险评估两个案例对本章方法进行验证，充分说明本章所提方法的有效性，与其他方法相比，在不增加算法复杂度的同时采用本章方法可以获得更高的推理准确率。

第四章 利用聚类分析构建析取范式置信规则库

4.1 引言

第三章采用基于容斥原理的置信规则库激活方法，该方法考虑属性间相关关系以及激活属性对推理结果的影响，更符合实际应用需求，并得到了较高的推理性能，因此，该方法也将用在后续章节的 DBRB 推理中。但第三章采用的规则库构建方法，是领域专家根据经验给定属性参考值个数和范围，根据参考值进行线性组合构建的。然而，该方法过于依赖专家的经验知识，当系统规模较大时，专家无法给定合适的参考值或者个数完成系统构建。且，由于 DBRB 构建的规则数较少，则属性参考值的组合方式对 DBRB 的构建至关重要。现有的针对 DBRB 构建的研究大都采用动态生成规则的方法确定系统规模并构建规则库系统，如，利用问题分类数确定系统规则数量[35]，并随机生成初始化参数完成初始规则库系统的构建；或者利用赤池信息量化标准衡量 DBRB 的整体性能，动态构建 DBRB 系统；Yang 等人则利用霍夫丁不等式推导出 BRB 系统的泛化误差，用于衡量系统结构和参数的联合优化模型，完成 CBRB 和 DBRB 系统的构建。但上述方法仍存在以下几个问题：(1)采用分类结果数确定系统的规则数，仅适用于分类问题，无法处理结果评价等级不确定的回归问题；(2)仅采用一个衡量指标容易出现过拟合的现象；(3)现有方法的结果评价等级的确定依赖专家判断，不具有客观性，并没有合理利用已知数据。

针对上述问题，本章引入 k-means 聚类方法用于构建回归型问题 DBRB 推理系统。对输出结果进行聚类，获取结果的特征分布，同时获取结果评价等级和系统规则数。采用该方法，系统的规则数由样本数据直接确定，充分利用现有的数据确定系统规模以及部分参数，完成初始的规则库系统构建。该方法利用输出数据直接确定初始的系统，避免动态探索构建规则，更合理有效地构建规则库系统。本章同样选取输油管道和桥梁风险评估两个案例对所提方法进行验证以说明该方法的可行性。

4.2 聚类分析

聚类[51]是常用的一种“无监督”学习方法，该方法通过训练未标记样本，将相似样本归为一个簇，常用于研究样本的分类问题。常用的聚类方法包括：基于

区域的 k-means、k-modes；基于密度的 DBSCAN；基于网格的 STING(Statistical information Grid)和 CLIOUE(Clustering in quest)；基于模型的高斯混合模型聚类算法，另外还有模糊 C 聚类，基于粒度的聚类方法等[52-53]。其中，k-means 作为经典的聚类分析方法常被用于应用研究中。本章节将介绍基于 k-means 构建 DBRB 的方法。

4.2.1 k-means

k-means 是一种基于距离的聚类算法，该方法以距离作为样本的相似性度量。对于样本集，给定 K 个初始中心，将样本集划分为 K 个簇，各样本被划分到距离中心最近的簇中，并形成新的簇中心，不断划分更新簇的过程，即为 k-means 的聚类过程。算法一般选取欧氏距离衡量相似度，以误差平方和（SSE）作为聚类目标，随着迭代次数的增加，SSE 逐渐减小，直到样本划分不发生变化，聚类中心不更新。

$$d_{i,j}=\sqrt{\left(x_i-m_j\right)^2} \qquad \text{公式（4-1）}$$

$$SSE=\sum_{j=1}^{k}\sum_{x_i\in C_j}\left|x_i-m_j\right|^2 \qquad \text{公式（4-2）}$$

$$m_j=\frac{\sum_{x_i\in C_j}x_i}{\left|C_j\right|} \qquad \text{公式（4-3）}$$

其中，$d_{i,j}$ 表示样本点 x_i 到类中心 $m_j\left(j=1,2,\cdots,k\right)$ 的距离。随着迭代次数的增加，SSE 逐渐减小，簇内样本的相似性不断提高。具体算法流程：

算法 1：k-means 聚类算法

输入：数据集 $D=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$

输出：聚类结果 $C=\{C_1,C_2,\cdots,C_j,\cdots,C_k\}$

过程：

1、从 D 中随机选取 k 个不重复的样本作为初始质心 $\{m_1,m_2,\cdots,m_k\}$；

2、对每个样本 x_i，计算其到各质心 m_j 的距离 d_{ij}，并将其划入与中心距离最小的簇中；

3、待所有样本划分完毕，根据公式（4-3），更新簇质心；

4、重复 2、3 两步操作，直到 m_j 不发生变化。

4.2.2 最优 K 值确定

K-means 算法的关键在于如何确定最佳聚类数 K，K 值的选取将影响数据的整体分布，而实际应用中 K 值的选取往往根据行业经验确定，存在人为主观偏差，无法获取最优聚类。因此，本章选取误差平方和（SSE）以及轮廓系数（S）确定最优 K 值范围[54]。

对表 4-1 的一维数据，进行 $K=\{1,2,\cdots,12\}$ 的聚类，并获取 K-SSE，K-S 的关系，结果如图 4-1、图 4-2 所示。

表 4-1　一维数据

编号	数值	编号	数值	编号	数值	编号	数值
1	1	7	9	13	23	19	61
2	2	8	10	14	27	20	62
3	3	9	11	15	40	21	100
4	5	10	20	16	41	22	150
5	6	11	21	17	42	23	200
6	7	12	22	18	43	24	1000

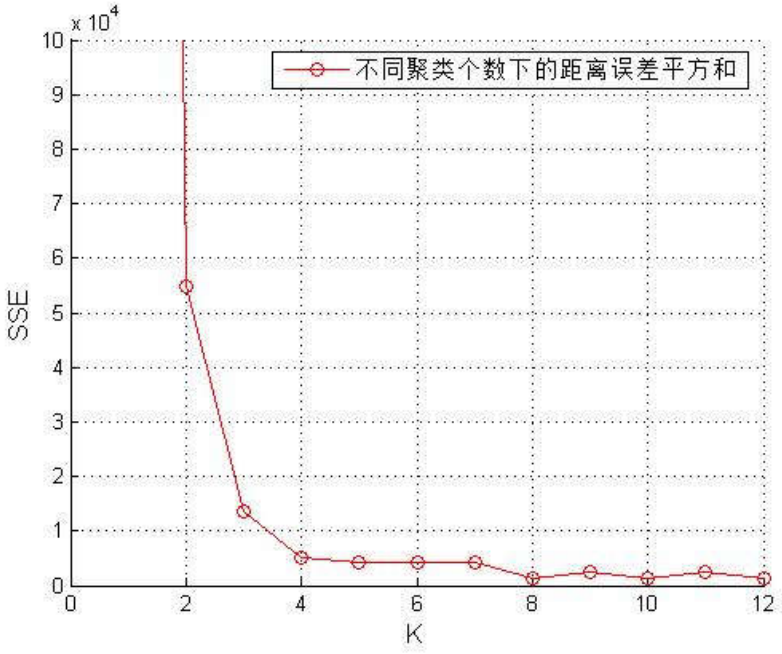

图 4-1 聚类数 K 与误差平方和（SSE）关系图

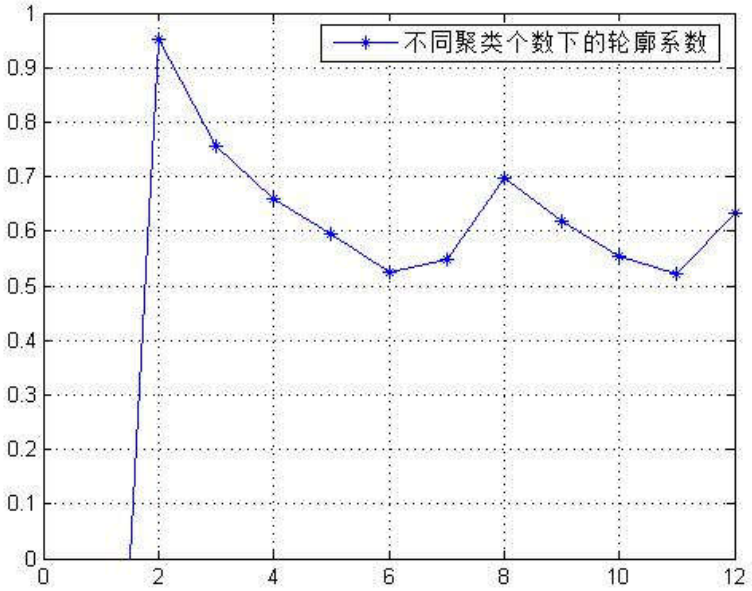

图 4-2 聚类数 K 与轮廓系数（S）关系图

4.2.2.1 误差平方和

误差平方和（SSE）用于衡量数据的类内误差，SSE 越小，表示簇内聚合度越高，聚类效果越佳。

如图 4-1 所示，随着 K 值的增大，数据划分的类别数增大，簇内的误差逐渐减小，且下降趋势逐渐减缓。说明，随着 K 的增大，误差平方和逐渐减小，但是簇的聚合度增大不明显，即增大 K 值，不能带来明显的聚合效果的提升。且，当 K 值不断增大到数据个数时，虽然 SSE=0，但是分类冗余，因此，SSE 不是单纯的越小越好。所以，在 K 值逐渐增大的过程中，存在最优 K 值，使得数据存在最佳聚类效果，簇内的聚合度高且不纯在多余的分类。如图 4-1，当 $K<4$ 时，SSE 下降幅度大；当 $K\geq 8$ 时，SSE 下降趋于平缓，且 SSE 时大时小，说明，K 的增大带来多余的分类则可以确定，当 $K\in[4,8]$ 时，存在最佳 K 取值，使得该场景下具有最佳的聚类效果。

4.2.2.2 轮廓系数

轮廓系数指通过比较当前样本与本簇样本和最近簇样本的平均距离，衡量聚类效果，即同时考虑类间距和类内距离，具体计算如下：

$$S_i=\frac{b_i-a_i}{\max\{a_i,b_i\}} \qquad \text{公式（4-4）}$$

$$a_i = \frac{\sum\sqrt{\left(x_i - x_j\right)^2}}{N_{C_l}}\left(x_i、\ x_j \in C_l, j = 1,2,\cdots,N_{C_l}\right) \qquad \text{公式（4-5）}$$

$$b_j = \frac{\sum\sqrt{\left(x_i - x_j\right)^2}}{N_{C_N}}\left(x_i \in C_l, x_j \in C_{N,} j = 1,2,\cdots,N_{C_N}\right) \qquad \text{公式（4-6）}$$

$$b_i = \min\left\{b_1,\cdots,b_j,\cdots,b_k \mid j \neq l, x_{\ i} \in C_l\right\} \qquad \text{公式（4-7）}$$

$$S = \frac{\sum S_i}{N} \qquad \text{公式（4-8）}$$

上述公式中，a_i 表示样本 x_i 与本簇内所有样本的类内平均距离，b_i 表示 x_i 与最近簇内所有样本的平均距离，S_i 表示样本 x_i 的轮廓系数，S 表示所有样本的平均轮廓系数。S 越大表示类间距越大，类内距离越小，聚合效果越好；反之，聚类效果较差。

从图 4-2 的 K-S 图中可知，当 K=2 时，轮廓系数最大，具有最佳的聚类效果；但是，在 K=2 时，误差平方和较高，说明聚类后的类内距离过大，K 不是最佳聚类数，其原因在于，当 K=2 时，样本离散度较大，导致轮廓系数过大。综合图 4-1 和图 4-2 可知，当 K=8 时具有最佳的聚类效果，从表 4-1 数据的划分可以证明。因此，最优 K 值的确定需要综合轮廓系数（S）和误差平方和（SSE）共同确定。

4.3 构建置信规则库

根据文献[35]可知，当 DBRB 系统规则数等于分类数时具有最佳的推理效果，对于聚类后的结果，每个簇可当成一个类别，则对于回归问题存在当规则数等于系统输出结果聚类数时，具有最佳的推理效果。因此，本章将系统的规则数设置为输出结果的最佳聚类数，并设置由聚类中心和结果评价等级的上下界共同构成系统的结果评价等级：

$$L = K \qquad \text{公式（4-9）}$$

$$N = K + 2 \qquad \text{公式（4-10）}$$

$$\begin{cases} D_{i+1} = m_i \quad i = 1,2,\cdots,K \\ D_1 = D_{\min} \\ D_{k+2} = D_{\max} \end{cases} \qquad \text{公式（4-11）}$$

其中，K 表示聚类数，N 为结果评价等级个数。规则库的具体构建流程如下：

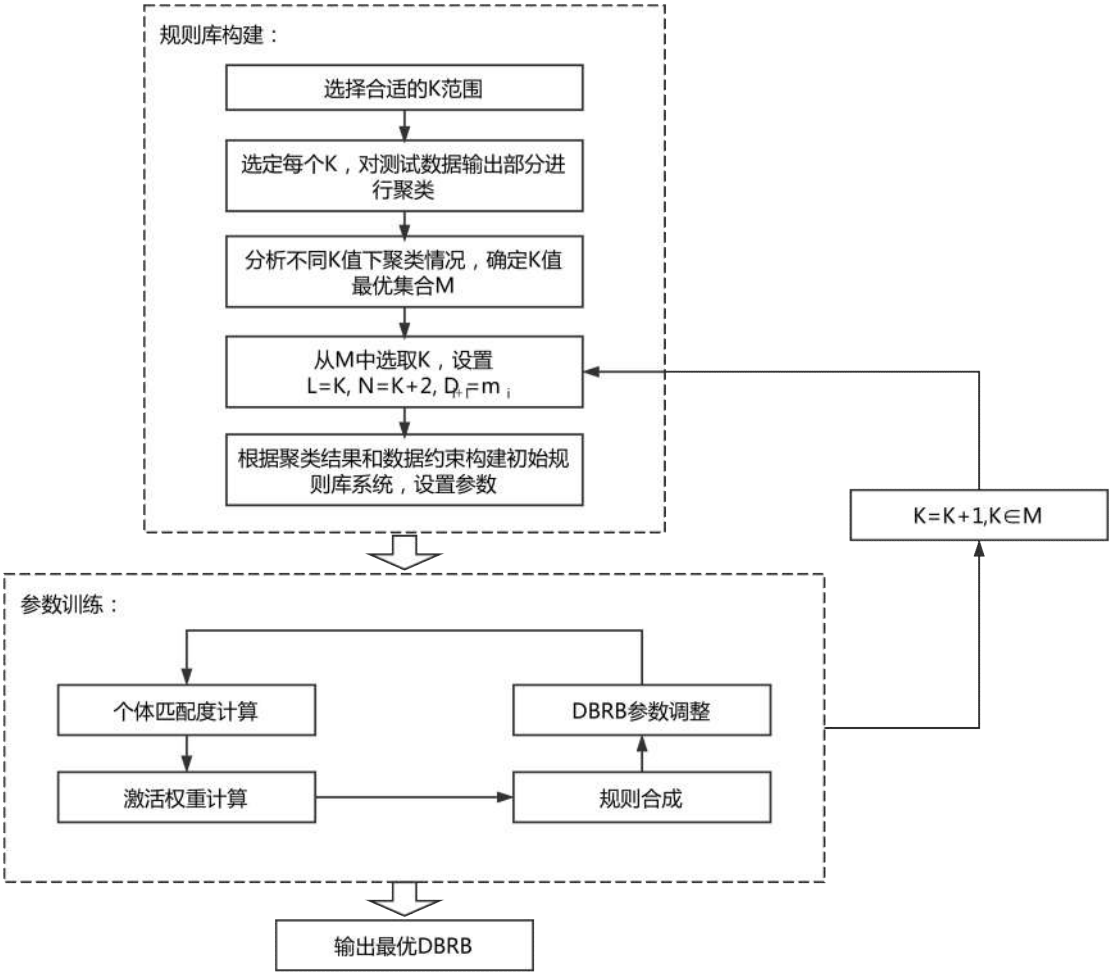

图 4-3 规则库构建算法流程图

步骤 1：根据数据规模，选定 K 取值范围，一般为 $K \in [1,10]$；

步骤 2：采用 k-means 聚类算法，对输出结果在不同 K 取值下进行聚类，并获得聚类误差平方和、类中心、轮廓系数。

步骤 3：根据步骤 2 的计算结果，构建 K-SSE、K-S 关系图，确定最优 K 值范围；

步骤 4：对最优 K 值范围内的每个 K 取值，根据公式（4-9）——公式（4-11），确定系统规模并构建规则库：

(1)根据属性参考值边界值限定，随机生成前提属性的初始参考值，并动态设置属性参考值的上下界：

$$\begin{cases} A_{i,j} = A_{i,\min} + random() * (A_{i,\max} - A_{i,\min}) \\ A_{i,s1} = A_{i,\min}, s1 = \arg\min_{j=1,\cdots,L} \{A_{i,j}\} \\ A_{i,s2} = A_{i,\max}, s2 = \arg\max_{j=1,\cdots,L} \{A_{i,j}\} \end{cases} \quad \text{公式（4-12）}$$

其中，$random()$ 表示生成 $[0,1]$ 之间的随机数；

(2)随机生成初始化属性权重和规则权重：

$$\theta_k = random() \qquad \text{公式（4-13）}$$

$$\delta_k = random() \qquad \text{公式（4-14）}$$

(3)根据式（4-11）设置结果评价等级；

(4)设置结果评价等级置信度：

$$\beta_{j,k} = \frac{random()_j}{\sum_{j=1}^{N} random()_j} \qquad \text{公式（4-15）}$$

(5)采用群智能方法，对初始化的 DBRB 系统进行参数训练，本章采用的群智能方法是改进后的蛙跳算法[55-57]；

(6)分析不同 K 值下的 DBRB 系统性能。

4.4 实验分析

本章实验中所用的数据和环境均与第三章相同，选用输油管道检漏和桥梁风险评估两个案例对本章方法构建的规则库系统的合理性进行检验。具体实验数据描述及实验环境详见第三章。

4.4.1 输油管道检漏实验

输油管道检漏实验以英国一条 100 多公里长的油气管道为背景，通过管道的流量差异（FD）和平均压力差（PD）推测管道泄漏情况。FD 和 PD 限制以及数据详细描述见第三章。

2008 组实验数据中的 500 组被选为训练数据，并对这 500 组训练数据的输出结果进行 $K=\{1,2,\cdots,10\}$ 的聚类，同时计算对应 K 值下的 SSE 和 S，结果如图 4-4、4-5 所示：

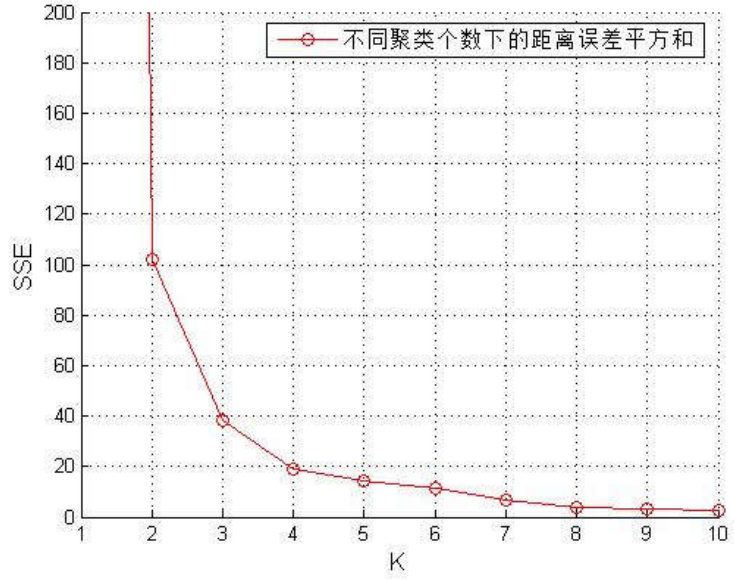

图 4-4 输油管道——K 与 SSE 关系图

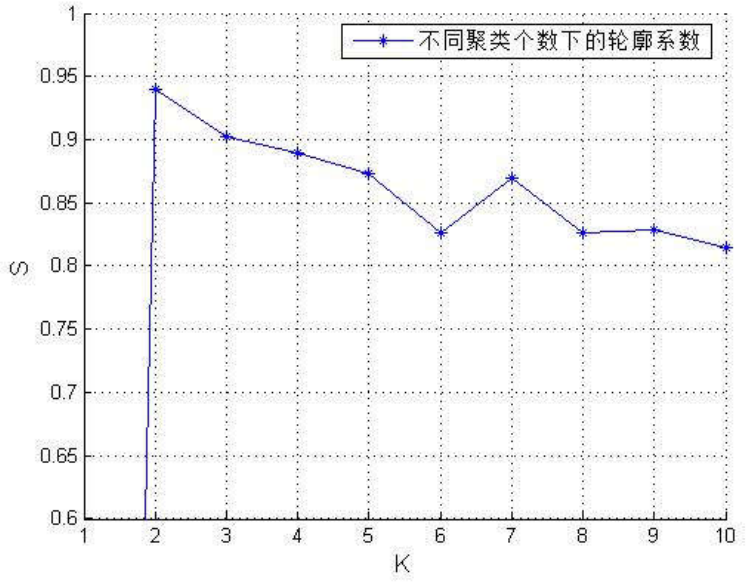

图 4-5 输油管道——K 与 S 关系图

从图 K-S 和 K-SSE 关系图可知，当 K=6 时，S 取得最小值，未取得理想的效果。当 $K \geq 6$ 时，虽然取得较小的 SSE，但相对于 K=5，SSE 的小幅度下降并没有带来轮廓系数的提升，且 K 的增大将导致系统规则数量增大，容易导致系统过拟合。因此，综合各方面因素可确定，最佳 K 值范围为 $K \in [3,5]$，而 K=4 时，具有最佳聚类效果。

根据步骤 4 和上文描述，分别取 $K=\{3,4,5\}$ 进行规则库构建，以平均绝对误差（MAE）、平均运行时间、参数个数作为 DBRB 的衡量指标，结果如表 4-2 所示：

表 4-2 输油管道——不同 K 取值下 BRB 性能比较

	MAE	TIME(S)	Parameter
K=3	0.171544	65.4630	26
K=4	0.168872	79.8660	38
K=5	0.168672	90.1850	59

从表 4-2 可知，在最优 K 值范围内，随着 K 值的增大，系统推理准确度不断提升，但运行时间和空间消耗同时增大。$K>4$ 时，规则的增加已经不能带来明显的推理精度提升，且空间和时间复杂度增加明显。因此，$K>4$ 带来的规则增加，无法为系统带来等价的精度提升。而 K=3 时，增加一条规则，推理精度上升 1.5%，效果明显。说明采用本章方法确定的最优 K 值范围是有效的。

其次，将最优 K 值范围下构建的规则库系统性能与其他方法进行比较，比较结构如表 4-3 所示。

从表 4-3 可知，本章方法构建的 DBRB 与现有方法下构建的 BRB、EBRB、DBRB 相比，能有效确定系统的规则数量，从而确定系统规模。且该方法下构建的 DBRB 系统相较于同等规则数量的 BRB、DBRB 具有较高的推理精确度，或增加少量规则的情况下可以带来较大的准确性的提升。且相较于第三章方法，本文在没有专家干预的情况下完成规则库构建和推理，虽然精度略低于本文第三章方法但更符合实际应用需求。说明采用本方法能有效确定系统规模独立完成 DBRB 系统构建。

表 4-3 输油管道——不同方法 BRB 推理性能比较

方法	类型	规则数	参数个数	MAE
扩展置信规则库	EBRB	900	18902	0.2169
加速梯度法	BRB	56	358	0.1828
动态规则自适应	BRB	6	43	0.2080
Chang-联合优化	DBRB	5	40	0.1941
Yang-联合优化	DBRB	3	22	0.1738
本文第三章方法	DBRB	8	70	**0.1665**
本章方法(K=5)	DBRB	5	59	0.1687

4.4.2 桥梁风险评估实验

桥梁风险评估通过对桥梁的安全性（SA）、功能性（FU）、可持续发展性（SU）

以及周边环境（EN）进行分析，共同确定桥梁的危险程度（RS）。有关桥梁安全性评估的数据，详见第三章实验部分。为了对构建的 DBRB 进行性能评估，本章同样采用 MAPE、RMSE、R 作为衡量 DBRB 系统性能的指标：

$$MAPE=\frac{1}{T}\sum_{i=1}^{T}\left|\frac{f(x_t)-y_t}{y_t}\right| \qquad \text{公式（4-16）}$$

$$RMSE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(f(x_t)-y_t\right)^2} \qquad \text{公式（4-17）}$$

$$R=\frac{\sum_{t=1}^{T}\left(f(x_t)-\bar{f}\right)\left(y_t-\bar{y}\right)}{\sqrt{\sum_{t=1}^{T}\left(f(x_t)-\bar{f}\right)^2\cdot\sum_{t=1}^{T}\left(y_t-\bar{y}\right)^2}} \qquad \text{公式（4-18）}$$

$$\bar{f}=\frac{1}{T}\sum_{t=1}^{T}f(x_t) \qquad \text{公式（4-19）}$$

$$\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t \qquad \text{公式（4-20）}$$

根据本章方法，对 66 组训练数据的风险评估部分进行单独聚类，选取 $K=\{1,2,\cdots,10\}$，其结果如图 4-6、4-7 所示。

从图 4-7 可知，当 $K=7$ 时，轮廓系数 S 最大，当 $K>7$ 时，轮廓系数逐渐减少，且随着 K 值增大本章方法下构建的 DBRB 规则数量将逐步增加，复杂度同时增加。当 $K\geq 4$ 时，平均绝对误差下降趋于缓慢，且 $K=4$ 与 $K=5$ 具有相近的轮廓系数，而 $K=5$ 时，聚类误差平方和更小。因此，采用上述两个指标可以确定 $K\in[5,7]$ 时，存在最佳聚类。

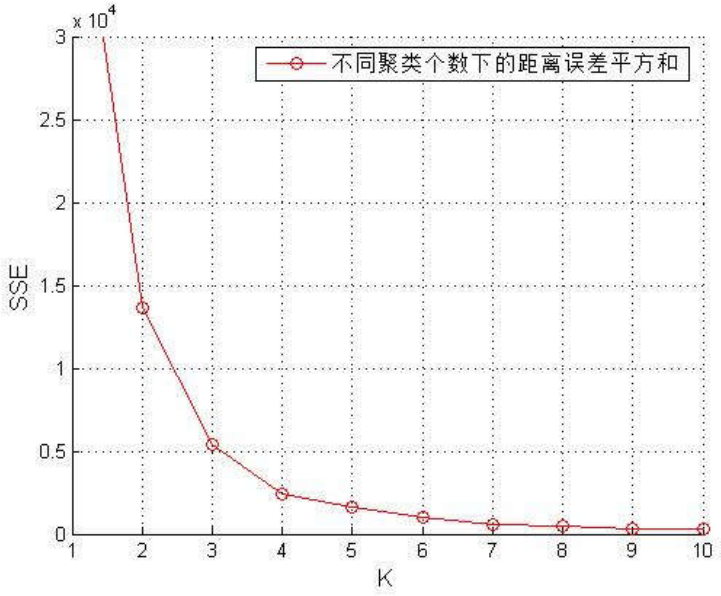

图 4-6 桥梁风险评估——K 与 SSE 关系图

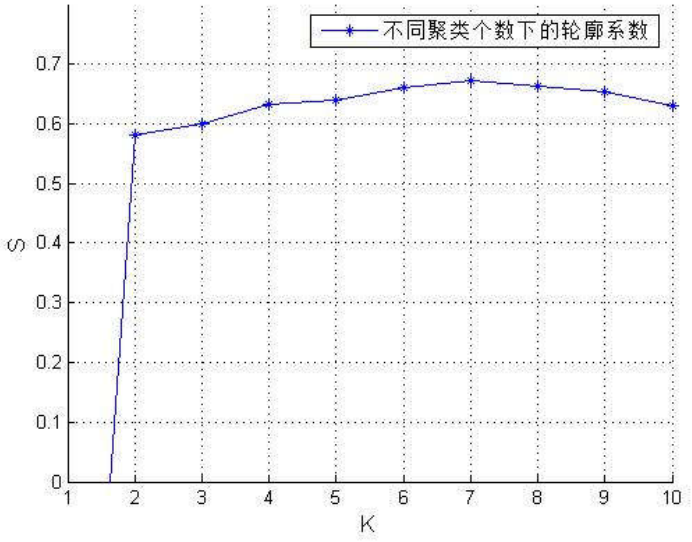

图 4-7 桥梁风险评估——K 与 S 关系图

因此，在不同 K 取值下，根据步骤 4.3 构建 DBRB 系统，并比较不同规则数量下的系统的推理性能、参数、以及运行时间，结果如表 4-4 所示。

从表 4-4 可以看出，当 K=8时，规则数增加并没有带来推理性能的增强，反而造成系统推理性能的大幅下降且带来了一定的时间和空间消耗。而当 $K \in [5,7]$ 时，构建的 DBRB 性能相近，因此，表 4-3 可说明采用轮廓系数和误差平方和可以准确的定位最佳聚类范围。

同样，将确定的最佳聚类效果下构建的 DBRB 系统与其他文献中所提方法进行比较，结果如表 4-5 所示。

表 4-4 桥梁风险评估——不同 K 取值下 BRB 性能比较

	MAPE(%)	RMSE	R（%）	TIME（s）	Parameter
K=4	5.7648	3.1703	99.1950	33.550	48
K=5	5.1051	2.3920	99.4064	44.468	64
K=6	4.8997	2.2297	99.6830	57.759	82
K=7	5.1966	2.2190	99.6641	78.325	102
K=8	5.5362	2.3135	99.6810	90.543	124

表 4-5　桥梁风险评估——不同方法 BRB 性能比较

模型	MAPE(%)	RMSE	R
BP-ANN	9.6294	4.1871	0.9834
ER1	22.4544	8.9255	0.9077
MRA3	18.5775	10.9527	0.8687
MRA8	23.9799	10.4653	0.8794
ER2	18.7808	11.2736	0.8918
MRA7	24.1941	10.3510	0.8796
MRA9	19.1456	11.3653	0.8904
Yang-BRB	**2.5999**	2.8060	0.9910
Yang-DBRB	6.7620	3.0916	0.9911
本文第三章方法	4.8093	2.5697	0.9943
本文方法（K=6）	4.8997	**2.2297**	0.**9968**

表 4-5 中所示为常用的桥梁风险评估模型以及部分 BRB 推理方法，其中，Yang-BRB 和 Yang-DBRB 方法在衡量桥梁风险时分别构建了 48 条和 5 条规则，相比于本章采用 6 条规则构建 DBRB 系统，Yang-BRB 采用传统的 BRB 构建方法且在有参数训练的情况下利用 48 条规则完成信息推理，更完整的描述数据信息，虽然在 MAPE 上优于其他模型但是其规则数量多参数数量大，需花费大量的时间进行规则训练。而本章在较少的时间和空间复杂度的基础上，同样获得了较高的推理性能，RMSE 和 R 均优于上述方法，且 MAPE 仅次于 Yang-BRB。说明本方法下构建的 DBRB 系统高效可行。

4.5 本章小结

本章针对现有 DBRB 方法不易在构建初期确定系统规模或过多依赖人为主观因素等问题，提出基于聚类分析构建 DBRB 系统的方法。该方法针对回归问题构建 DBRB 系统，利用聚类分析方法确定结果评价等级和系统的规则数。且该过程充分挖掘数据中的已知信息，完成初始 DBRB 的构建。最终，利用参数训练确定完整的 DBRB 系统。从输油管道和桥梁风险两个实验室中可以证明，使用该方法构建的 DBRB 系统能有效表达系统信息并获取更高的推理性能。

第五章 构建基于 TSD 的 DBRB 分类系统

5.1 引言

在第三章、第四章中，本文分别研究 DBRB 激活权重优化和回归问题中 DBRB 的构建。其中第四章 DBRB 的构建方法主要是通过聚类方法确定系统的规则数和结果评价等级从而完成 DBRB 的构建。但是，对于 DBRB 的构建除了需要确定规则数，其次还要确定前提属性参考值。因此，本章针对分类问题，采用两阶段离散化（Two-stage discretization，TSD）[44]方法确定属性参考值和系统规则数，完成规则库构建。

现有的 DBRB 构建方法分为两种：一种是采用专家提供的相关领域经验值进行组合生成规则；另一种是随机生成初始的属性参考值并进行参数训练后确定，如，Chang 等人[21]提出的在属性参考值上下界范围内随机生成初始值。前一种方法依赖专家的领域知识，当系统规模过于庞大时，无法人为确定参数，且人为确定易存在主观偏差，影响系统最终的推理性能。其次，采用随机生成属性参考值的方法，不同初始值将对参数训练造成不同的影响，存在不确定性，最终影响系统的推理性能。且分类问题前提属性数量过多，每增加一条规则，需同时增加属性参考值，规则权重、置信度等参数，系统参数训练时间将成倍增长。因此，上述两种方法均不适合用于解决分类问题。

针对上述情况，本章提出基于 TSD 的 DBRB 分类系统构建方法。该方法采用基于条件信息熵的 TSD 方法，对测试数据中前提属性数据进行局部离散化和全局离散化处理，不仅能减少系统的属性个数，同时可以在离散化后确定属性的参考值，从而减少系统待确定的参数数量。并在此基础上对参考值进行线性组合，通过参数训练方法确定规则权重和结果置信度，即可完成 DBRB 的构建。实验中选用 UCI 三个经典的分类数据集用于验证 DBRB 分类系统的性能并得到较高的分类准确率。

5.2 基于条件信息熵的 TSD 算法

在本小节中，将对 TSD 算法进行详细的介绍，首先是决策系统的基础知识介绍，然后对两阶段的离散化过程进行介绍，包括局部离散化和全局离散化的介绍及其相应的方法。

5.2.1 决策系统

一个完整的决策系统[58]由五部分构成：

$$S=(U,A,D,\{V_a|a\in A\cup D\},\{I_a\mid a\in A\cup D\}) \qquad \text{公式（5-1）}$$

其中，S 表示决策系统，该系统共由五个部分组成，U 表示事件集合，A 表示前提属性集合，D 表示决策属性集合，V_a 表示前提属性 a 的属性参考值集合，I_a 表示参考值间对应的函数关系。如表所示：

表 5-1 决策表 S

U	a_1	a_2	D
x_1	0.8	2.0	1
x_2	1.0	0.5	1
x_3	1.3	3.0	0
x_4	1.4	1.0	1
x_5	1.4	2.0	0
x_6	1.6	3.0	0
x_7	1.3	1.0	1

信息熵和条件信息熵[59]常被用于衡量属性的信息量大小，即属性的决策能力，其计算公式如下：

$$H(P)=-\sum_{i=1}^{n}p(X_i)\log(p(X_i)) \qquad \text{公式（5-2）}$$

$$H(Q\mid P)=-\sum_{i=1}^{n}p(X_i)\sum_{j=1}^{m}p(Y_j\mid X_i)\log(p(Y_j\mid X_i)) \qquad \text{公式（5-3）}$$

P，Q 分别表示前提属性和决策属性。$H(P)$ 为属性 P 下的信息熵，$H(Q\mid P)$ 表示属性 P 在 Q 结果下的条件信息熵。X_i、Y_i 为属性 P 和 Q 的所有可能取值。条件信息熵越小，说明属性对事件的分类能力越强。

5.2.2 基于断点的信息扩展处理

在介绍 TSD 前，本章先介绍基于断点的离散化信息处理。本章断点指对属性 a_i 所有参考值进行排序后，取相邻两个参考值中点即为断点，a_i 所有断点组成的集合即为属性 a_i 的断点集[60-62]：

$$C_{a_i}=\{c_i^{a_i}\mid a_i\in A,1\le i\le\|V_{a_i}\}$$ 公式（5-4）

$$c_i^{a_i}=\left(v_i^{a_i}+v_{i+1}^{a_i}\right)/2$$ 公式（5-5）

其中，C_{a_i} 表示属性 a_i 的断点集，$c_i^{a_i}$ 表示属性 a_i 在参考值区间$\left[v_i^{a_i}，v_{i+1}^{a_i}\right)$的断点。在确定每个属性的断点集后，对决策表 S 进行信息扩展：

$$SA=\{S_{a_i}\mid a_i\in A\}$$ 公式（5-6）

$$S_{a_i}=\left(a_i,c_i^{a_i}\right)(x)=\begin{cases}0,\ if\ a_i(x)\le c_i^{a_i}\\1,\ otherwise\end{cases}$$ 公式（5-7）

上述公式中，SA 表示决策表 S 的扩展信息表，SA 由多个 S_{a_i} 扩展信息表组成，每个 S_{a_i} 表示 S 根据 $c_i^{a_i}$ 依次展开的信息。以表 S 为例对属性 a_1 进行扩展，其断点集为 $C_{a_1}=\{0.9,1.15,1.35,1.5\}$，扩展的 S_{a_1} 如下：

表 5-2 扩展信息表 S_{a_1}

U	$c_1^{a_1}$	$c_2^{a_1}$	$c_3^{a_1}$	$c_4^{a_1}$	D
x_1	0	0	0	0	1
x_2	1	0	0	0	1
x_3	1	1	0	0	0
x_4	1	1	1	0	1
x_5	1	1	1	0	0
x_6	1	1	1	1	0
x_7	1	1	0	0	1

5.2.3 基于条件信息熵的 TSD 算法

两阶段离散化（Two-Stage Discretization，简称 TSD）顾名思义将离散化分为两个阶段——局部离散化和全局离散化。局部离散化对每个属性进行独立离散化操作并选取前 K 个最优断点参与第二阶段的离散化。全局离散化对所有属性及其局部离散化信息表中的新断点进行二次选择，形成最终的离散化方案。

5.2.3.1 局部离散化

在局部离散化（Cut shift local discretization，简称 CSLD）中，对于给定的 K 和每个扩展信息表S_{a_i}，以条件信息熵为衡量标准，迭代选出S_{a_i}中条件信息最小的前 K 个$c_i^{a_i}$作为S_{a_i}的关键断点，具体算法如下：

算法 2：CSLD

输入：K

输出：断点集合P

过程：

1、计算a_i的断点集合C_{a_i}；

2、根据C_{a_i}对S进行扩展得到S_{a_i}；

3、根据S_{a_i}计算每个$c_i^{a_i}$的条件信息熵；

4、选取具有最小条件信息熵的$c_i^{a_i}$加入到P_{a_i}。

5、以P_{a_i}中的断点为基础断点，逐一选取C_{a_i}中未被选中的断点$c_j^{a_i}$加入P_{a_i}中形成P_{a_i}'，根据P_{a_i}'对a_i进行扩展并计算条件信息熵CE_j；

6、选取CE_j最小的$c_j^{a_i}$加入到P_{a_i}中；

7、重复步骤 5、6 直到$\left|P_{a_i}\right|=K$。

8、输出 P，$P=\bigcup P_{a_i}$

经过 CSLD 处理的决策表，每个属性有 K 个断点被选取，K 个断点将属性参考值映射到$[0,K]$的整数，得到局部离散化扩展信息表。

5.2.3.2 全局离散化

在局部离散化的基础上对局部离散化信息表进行全局离散化（Scaling-based global discretization，简称 SBGD），局部离散化将属性的每个断点当成一个“属性”进行比较，并选取出条件信息熵最小的前 K 个断点；全局离散化综合局部离散化信息表中的所有属性和断点，进行二次选择。对局部离散化结果进行扩展得到扩展信息表S'，对S'中的每个“属性”不断选取信息量最大的“属性”加

入到断点集合P_s中，直到所选的断点集合能完全区分结果。具体算法步骤如下：

算法 3：SBGD

输入：S'，P

输出：断点集合P_s

过程：

1、根据S'计算每个c_i'（P中的所有断点）的条件信息熵；

2、选取具有最小条件信息熵的c_i'加入到P_s。

3、以P_s中的断点为基础断点，逐一选取P中未被选中的断点c_j'加入P_s中形成P_s'，根据P_s'中的所有属性计算条件信息熵CE_j；

4、选取CE_j最小的c_j'加入到P_s中；

5、重复步骤 3、4 直到CE_j=0。

全局离散化后被选定的所有属性和断点共同表示系统的决策信息。离散化后，系统中的数据是经过两次离散化后的压缩数据，虽然数据范围发生变化但仍保留有原始决策系统的信息，提高后期数据处理的效率。

5.3 构建基于 TSD 的 DBRB 分类系统

分类问题中属性数量较多，且每个属性取值多样化，采用 BRB 系统直接处理分类问题容易造成“组合爆炸”，而直接采用 DBRB 虽然能在规则数量上有所减少但是系统的属性参考值无法直接确定。因此，本章考虑采用 TSD 方法确定系统参数。TSD 算法获取的断点集合可以对决策表中的信息进行离散化处理且能有效减少系统中决策属性的个数，同时，TSD 算法压缩过的决策表可以准确的定位系统的关键属性及其对应的参考值。因此，采用 TSD 算法对原始信息进行离散化，既可以提取出特征属性同时可以确定属性参考值，且该属性和参考值能被直接用于构建 DBRB 分类系统，有效缩减 DBRB 系统规模且仍保留系统的决策性能。具体构建方法如下：

步骤 1：根据经验，选定合适的 K 值；

步骤 2：对数据进行 CSLD，获取局部离散化断点；

步骤 3：根据 CSLD 选取的断点对信息表进行拓展，得到扩展数据表。对扩展数据表进行 SBGD，当$CE_j = 0$或|P|=|P_S|时，结束离散化过程，获取全局离散化断点P_s。当$CE_j = 0$时，执行步骤 4，当|P|=|P_S|时，执行步骤 5；

步骤 4： 根据P_s中的断点对数据进行扩展，设P_s中断点对应的属性为特征属性，数据取值为属性参考值；

步骤 5：P_s对应属性为特征属性，P_s中与P相对应的断点为属性对应的参考值；

步骤 6：随机生成属性权重、规则权重、结果评价等级置信度；

$$\theta_k = random() \qquad \text{公式（5-8）}$$

$$\delta_k = random() \qquad \text{公式（5-9）}$$

$$\beta_{j,k} = \frac{random()_j}{\sum_{j=1}^{N} random()_j} \qquad \text{公式（5-10）}$$

步骤 7：对初始化 DBRB 进行参数训练，直到满足迭代条件或精度要求。

在构建 DBRB 分类系统时，本章将 SBGD 的终止情况分为两种，并根据 SBGD 的不同终止条件采用不同的方法构建 DBRB。原因在于，离散化操作后，数据的粒度变大将造成数据冲突，即相同前件数据指向不同的分类结果。因此，当离散化后产生冲突数据时，则不适合直接对原始数据进行离散化处理。所以，当数据发生冲突时，需要保留原始数据形态，同时选择对应 CSLD 阶段的断点作为属性参考值；当$CE_j = 0$时，说明离散化后的数据能够完全表示系统的分类信息，因此，SBDG 阶段被选中的属性作为 DBRB 的属性，对应的离散化后的属性取值为属性参考值，具体算法流程如图 5-1 所示。

5.4 实验分析

本章采用 TSD 方法对数据进行离散化处理，获取系统特征属性和属性参考值，并确定系统规则数量。在此基础上构建 DBRB 分类系统，以该方法构建的 DBRB 系统不仅能够减少前提属性个数，以更少的属性进行分类推理，其次能在构建初始 DBRB 时确定属性的参考值，减少参数学习的数量。另外，本章采用了第三章的激活权重计算方法对 DBRB 推理过程中的激活权重计算进行优化。实验选取了 3 个 UCI 上常用的分类数据集对本章构建的 DBRB 系统进行验证，并将结果与常用的分类方法进行对比。

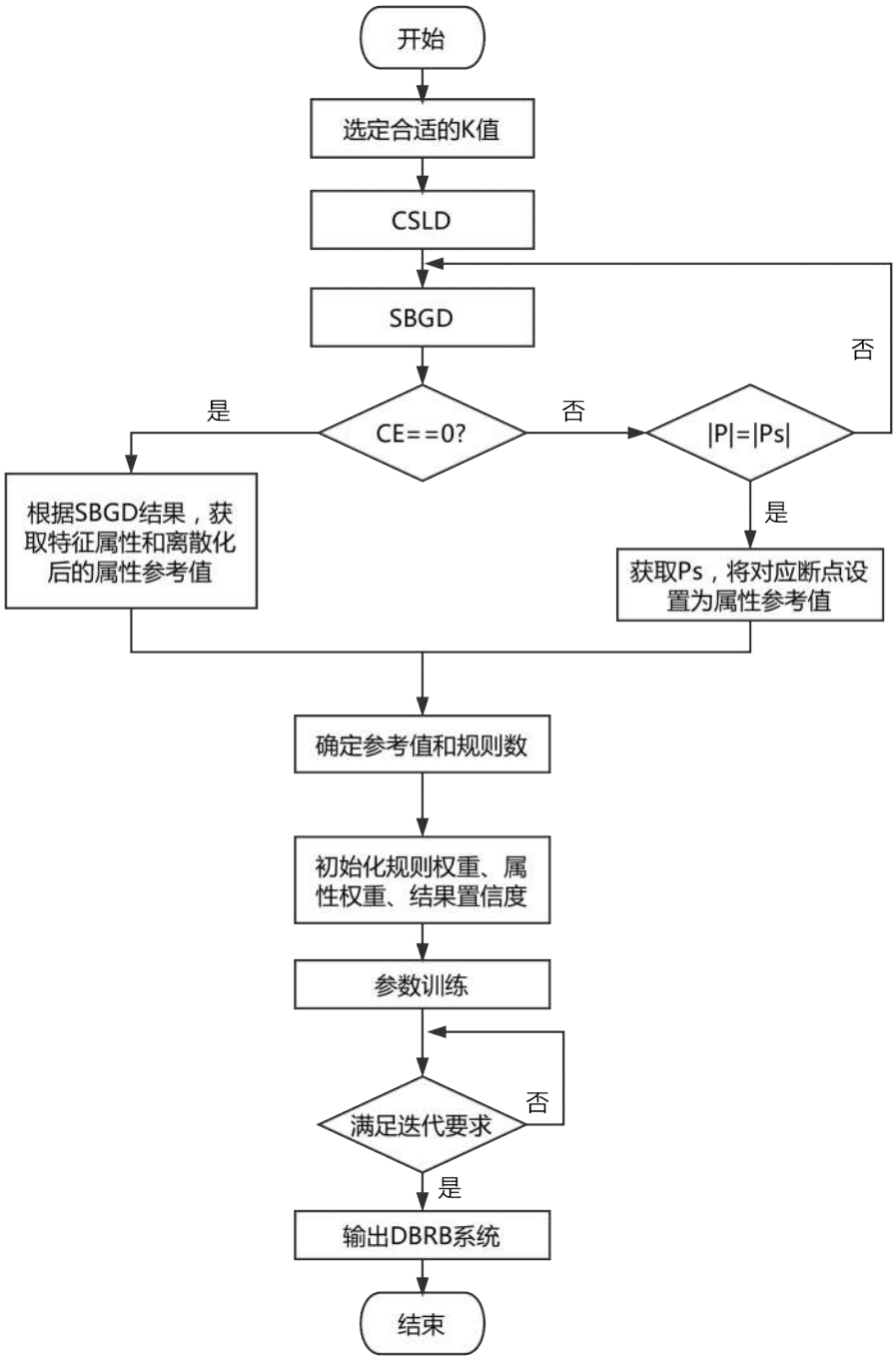

图 5-1 基于 TSD 的 DBRB 分类系统构建流程图

5.4.1 实验环境及数据

在对本章构建的 DBRB 分类系统性能进行验证时，选用了群智能算法作为 DBRB 参数训练的优化方法，设置种群规模 NP=50，迭代次数为 num=50000。此外，实验环境为：Intel(R) Core i5-4570@ 3.20GHz；4GB 内存；Windows 8 操作系统；算法使用 Visual Studio 2012 实现。

对于实验数据集，本章选用 UCI 公共数据上常用的 3 个分类数据集：鸢尾花 iris、玻璃 glass、酒 wine，对构建的 DBRB 分类器进行验证，详细数据信息见表 5-3。

表 5-3 实验数据

数据集	属性个数	分类数	数据量
Iris	4	3	150
Glass	9	7	214
Wine	13	3	178

5.4.2 实验设置

实验中，在 CSLD 阶段分别设置 $K_{\text{iris}}=2$、$K_{glass}=2$、$K_{wine}=2$ 对各个数据集中的数据进行局部离散化和全局离散化，其中 wine 使用离散化后的数据进行规则库构建，iris 和 glass 采用原始数据和断点集进行规则库构建。为详细说明构建方法，本章分别以 iris 和 wine 为例，分别介绍两种 DBRB 分类系统的构建方法。

Iris-DBRB 分类系统的构建步骤：

（1）获取 iris 的所有断点，并对数据进行扩展；

（2）选取 $K_{iris}=2$ 并在此基础上对 iris 扩展信息表进行 CSLD，获取局部断点集合以及局部离散化信息表；

（3）根据局部离散化扩展信息表进行扩展获取新的断点，对每个属性的每个断点进行 SBGD;由于 $CE_{iris}\neq 0$，SBGD 输出 CSLD 全部断点以及属性上下界作为属性参考值；

（4）根据上述参考值构建 DBRB 系统并对其余参数进行参数学习。

以上即为 Iris-DBRB 分类系统的构建过程。

Wine-DBRB 分类系统的构建方法与 Iris-DBRB 大致相同，不同点在于步骤（3），Wine-DBRB 的步骤（3）如下：

（3）根据局部离散化扩展信息表进行扩展获取新的断点，根据每个属性的局部断点集进行 SBGD，直到 $CE_{wine}=0$。根据选中的断点对应的 CSLD 断点，对数据进行扩展获取全局离散化扩展信息表。SBGD 输出被选中属性，以及全局离散化信息表中所有取值作为属性参考值。

5.4.3 实验结果分析

为验证本章构建的 DBRB 分类系统的可行性，本章采用十折交叉验证的方法对实验结果进行验证，取十次结果的平均值为最终的分类准确率。同时，将 TSD-DBRB 分类系统与文献[63]中列举的常用分类器进行对比，包括 C4.5、贝叶斯、模糊集等，并列举出各方法在三个分类数据集上的分类准确率，结果如表 5-4 所示：

表 5-4 不同方法的平均分类准确率对比结果

分类方法	分类准确率（%）		
	Iris	Wine	Glass
Naïve Bayes	96.00	96.75	42.9
C4.5	95.13	91.14	67.9
SMO	96.69	97.87	58.85
FGM	96.88	**98.36**	**69.17**
Fallahnezhad	97.46	97.88	56.5
Ye-DBRB	96.63	—	61.86
本章方法	**98.00**	98.33	67.27

从表 5-4 可知，与其他分类算法相比，本章所提方法在公共分类数据集上均取得了较高的分类准确率，其中，iris 和 wine 数据集均取得较高的分类正确率 ，glass 的分类正确率仅次于 FGM 和 C4.5，且相比于 FGM，本章方法所使用的参数数量远小于 FGM，说明本章方法构建的 DBRB 分类器准确有效。

5.5 本章小结

本章提出基于 TSD 的 DBRB 分类系统构建方法。该方法采用两段离散化方法 CLSD 和 SBGD 用于提取特征属性和属性参考值，构建 DBRB 分类系统。并在 UCI 公共分类数据集上进行实验，与经典的分类算法进行比较，从实验结果可以证明，本章方法在分类问题上能获取较高的分类准确率。

结论与展望

BRB 作为基于规则的专家系统中的一种，在传统的 IF-THEN 规则的基础上引入置信框架，使系统具有处理不完整性、模糊不确定性、概率不确定性以及非线性问题的能力。为了更好的表达知识以达到解决实际问题的目的，学者将 DBRB 用于解决各类问题并得到了较高的推理效果。然而，目前关于 DBRB 的研究仍处于起步阶段，DBRB 由于属性连接方式的变化影响激活权重的计算，且现有的规则库构建方法大多依赖人为因素，或者构建过程中系统规模存在不确定性。因此，本文基于容斥原理提出新的激活权重计算方法。同时并且为了有效构建规则库系统处理不同场景下的推理问题，本文结合聚类算法和离散化方法，分别提出利用聚类分析构建析取范式置信规则库和构建基于 TSD 的 DBRB 分类系统，两种 DBRB 的构建方法，在改进的激活权重计算方法下完成 DBRB 的推理，有效提升 DBRB 的推理性能。具体研究内容包括以下三点：

（1）传统的 BRB 主要采用合取的属性连接方式，因此，Yang 根据概率论选择了累乘的激活权重计算方式。而 DBRB 虽然是采用线性组合参考值的方式，但是累加和的激活权重公式无法表达属性间的相关关系。且由于激活方式由规则激活变成属性激活，属性的权重和激活个数将对激活权重计算产生更大的影响。因此，本文提出基于容斥原理的析取置信规则库激活方法。该方法基于容斥原理计算激活权重，不仅考虑了属性间的相关关系，避免属性间相互独立的限制，并根据激活属性的个数和重要性不同调整规则权重，优化基于析取范式的置信规则库激活方法。本文所提方法应用在输油管道检漏和桥梁风险评估实验中，结果表明，该方法具有更好的推理性能。

（2）目前 DBRB 的构建方法主要采用参考值组合或根据某类指标动态生成规则的方式，但是现有方法构建的规则库系统存在过拟合、主观因素太强、未合理利用已知数据信息等问题。鉴于此，本文提出利用聚类分析构建析取范式置信规则库。通过对样本数据的输出结果进行聚类分析，获取结果分布特征、规则数、结果评级等级等，完成规则库的构建，并成功应用于输油管道检漏和桥梁风险评估的推理中。

（3）决策过程中数据呈现数量多、多样性、复杂性的特点，如何从大量数据中抽取决策信息成为关键。因此，针对分类问题数据量大，属性多的问题，本文提出基于 TSD 的 DBRB 分类系统构建方法。通过对数据进行两次连续的离散化——CLSD 和 SBGD，获取特征属性和属性参考值。根据获取的特征值完成规则库构建。实验分析中，采用 UCI 上的公共数据集对构建的 DBRB 分类系统进

行验证，有效说明本文构建的分类系统具有较高的分类准确性。

本课题关于 DBRB 的研究仍有大量不足需待改善，未来的研究工作方向如下：

（1）目前只采用 k-means 对数据的输出结果进行聚类获取规则数量和结果评价等级，尚未考虑到前提属性部分。因此，接下来将考虑尝试多种不同的聚类方式对数据进行处理，同时，从数据整体性出发，通过对每条数据完整的观测获取构建规则库所需参数。

（2）采用离散化对数据进行处理时，容易造成数据的冲突问题，但目前 BRB、DBRB 等尚无法处理数据冲突问题，需要对冲突数据分别处理。其次，现有的数据大多是完整的，未存在缺失数据。因此，将考虑如何对 DBRB 进行拓展，使其具有表达和处理缺失数据、冲突数据的能力。

（3）目前针对 DBRB 的研究均是基于离线模型的研究，不具有处理实时性数据的能力，因此，考虑将 DBRB 与基于长短记忆神经网络相结合（LSTM）结合，使其具有处理动态预测和处理时间序列问题的能力，增强模型的学习能力。

参考文献

[1] 蔡自兴, 徐光祐. 人工智能及其应用[M]. 清华大学出版社, 2004.

[2] Durkin J, Durkin J. Expert Systems: Design and Development[M]. DBLP, 1994.

[3] Alexander M. Meystel, James Sacra Albus. Intelligent Systems: Architecture, Design, and Control[J]. Robotica, 2002, 20(6):2.

[4] Dempster A P. A Generalization of Bayesian Inference[J]. Journal of the Royal Statistical Society, 1968, 30(2):205-247.

[5] Shafer G. A mathematical theory of evidence[M]. Princeton university press, 1976.

[6] Zadeh L A, Klir G J, Yuan B. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers[M]. World Scientific, 1996.

[7] Haykin S. Neural networks: a comprehensive foundation[M]. Prentice Hall PTR, 1994.

[8] Yang J B, Liu J, Wang J, et al. Belief rule-base inference methodology using the evidential reasoning Approach-RIMER[J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 2006, 36(2):266-285.

[9] Hwang C L, Yoon K. Methods for Multiple Attribute Decision Making[M]// Multiple Attribute Decision Making. Springer Berlin Heidelberg, 1981:58-191.

[10] Huynh V N, Nakamori Y, Ho T B, et al. Multiple-attribute decision making under uncertainty: the evidential reasoning approach revisited[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2006, 36(4): 804-822.

[11] Wang Y M, Elhag T M S. A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks[J]. Expert Systems with Applications, 2007, 32(2):336-348.

[12] Yang J B, Liu J, Xu D L, 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.

[13] 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.

[14] 刘佳俊, 胡昌华, 周志杰, 等. 基于证据推理和置信规则库的装备寿命评估[J]. 控制理论与应用, 2015, 32(2): 231-23.

[15] Jiang J, Li X, Zhou Z, et al. Weapon system capability assessment under uncertainty based on the evidential reasoning approach[J]. Expert Systems with Applications, 2011, 38(11): 13773-13784.

[16] 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, Paris, 2013: 145-169.

[17] 方志坚, 傅仰耿, 陈建华. 纹理图像分类的置信规则库推理方法[J]. 应用科学学报, 2017, 35(5): 545-558.

[18] 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.

[19] Lin Y Q, Li M, Chen X C, et al. A Belief Rule Base Approach for Smart Traffic Lights[C]// International Symposium on Computational Intelligence and Design. IEEE, 2017:460-463.

[20] 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.

[21] Chang L, Zhou Z J, You Y, et al. Belief rule based expert system for classification problems with new rule activation and weight calculation procedures[J]. Information Sciences, 2016, 336(C):75-91.

[22] 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.

[23] 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(自然科学版), 2011, 32(1):154-157.

[24] 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.

[25] Zhou Z J, Hu C H, Yang J B, et al. Online updating belief-rule-base using the RIMER approach[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2011, 41(6): 1225-1243.

[26] Zhou Z J, Hu C H, Xu D L, et al. Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection[J]. Expert Systems with Applications, 2011, 38(4): 3937-3943.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] Zhou Z J, Hu C H, Yang J B, et al. A sequential learning algorithm for online constructing belief-rule-based systems[J]. Expert Systems with Applications, 2010, 37(2): 1790-1799.

[33] Wang Y M, Yang L H, Fu Y G, et al. Dynamic rule adjustment approach for optimizing belief rule-base expert system[J]. Knowledge-Based Systems, 2016, 96: 40-60.

[34] Wang Y M, Luo Y. Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making[J]. Mathematical and Computer Modelling, 2010, 51(1): 1-12.

[35] 叶青青, 杨隆浩, 傅仰耿,等. 基于改进置信规则库推理的分类方法[J]. 计算机科学与探索, 2016, 10(5):709-721.

[36] 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.

[37] AbuDahab K, Xu D, Chen Y. A new belief rule base knowledge representation scheme and inference methodology using the evidential reasoning rule for evidence combination[J]. Expert Systems with Applications, 2016, 51: 218-230.

[38] Yang L H, Wang Y M, Lan Y X, et al. A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems[J]. Knowledge-Based Systems, 2017, 123: 174-187.

[39] Alberto C, Liu J, Wang H, Anil K. A new dynamic rule activation method for extended belief rule-based systems [J]. IEEE Transactions on knowledge and data engineering, 2015，7(4):880-888.

[40] Lin Y Q, Fu Y G, Su Q, et al. A rule activation method for extended belief rule base with VP-tree and MVP-tree[J]. Journal of Intelligent & Fuzzy Systems, 2017, 33(6): 3695-3705.

[41] Chang L L, Zhou Z J, Chen Y W, et al. Belief rule base structure and parameter joint optimization under disjunctive assumption for nonlinear complex system modeling[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 48(9): 1542-1554.

[42] Yang L H, Wang Y M, Liu J, et al. A joint optimization method on parameter and structure for belief-rule-based systems[J]. Knowledge-Based Systems, 2018, 142: 220-240.

[43] Yang L H, Wang Y M, Chang L L, et al. A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model[J]. Computers & Industrial Engineering, 2017, 113: 459-474.

[44] Wen L Y, Min F, Wang S Y. A two-stage discretization algorithm based on information entropy[J]. Applied Intelligence, 2017, 47(1):1-17.

[45] Yang J B, Xu D L. Evidential reasoning rule for evidence combination[J]. Artificial Intelligence, 2013, 205(205):1-29.

[46] 周志杰. 置信规则库专家系统与复杂系统建模[M]. 科学出版社, 2011.

[47] Brualdi R A, Pearson. Introductory Combinatorics:International Edition[J]. Pearson Schweiz Ag, 2011.

[48] Wang Y M, Elhag T M S. A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks[J]. Expert Systems with Applications, 2007, 32(2): 336-348

[49] Elhag T M S, Wang Y M. Risk assessment for bridge maintenance projects: Neural networks versus regression techniques[J]. Journal of computing in civil engineering, 2007, 21(6): 402-409

[50] Wang Y M, Elhag T M S. Evidential reasoning approach for bridge condition assessment[J]. Expert Systems with Applications, 2008, 34(1): 689-699

[51] MacQueen J. Some methods for classification and analysis of multivariate observations[C]//Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967, 1(14): 281-297.

[52] SCOTT E L. Berkeley Symposium on Mathematical Statistics and Probability[M]. University of California Press, 1951, 23-67

[53] Forgy E W. Cluster analysis of multivariate data : efficiency versus interpretability of classifications[J]. Biometrics, 1965,21(3):41-52.

[54] Aloise D, Deshpande A, Hansen P, et al. NP-hardness of Euclidean sum-of-squares clustering[J]. Machine Learning, 2009, 75(2):245-248.

[55] Elbeltagi E, Hegazy T, Grierson D. Comparison among five evolutionary-based optimization algorithms[J]. Advanced Engineering Informatics, 2005, 19(1):43-53.

[56] Rahimi-Vahed A, Dangchi M, Rafiei H, et al. A novel hybrid multi-objective shuffled frog-leaping algorithm for a bi-criteria permutation flow shop scheduling problem[J]. International Journal of Advanced Manufacturing Technology, 2009, 41(11-12):1227-1239.

[57] Rahimi-Vahed A, Mirzaei A H. Solving a bi-criteria permutation flow-shop problem using shuffled frog-leaping algorithm[J]. Soft Computing, 2008, 12(5):435-452.

[58] Min F, Liu Q. A hierarchical model for test-cost-sensitive decision systems[J]. Information Sciences, 2009, 179(14):2442-2452.

[59] Wang G Y, Yu H, Yang D C. Decision table reduction based on conditional information entropy[J]. Chinese Journal of Computers, 2002, 25(7):759-766.

[60] Min F, Liu Q, Fang C. Rough sets approach to symbolic value partition[J]. International Journal of Approximate Reasoning, 2008, 49(3):689-700.

[61] Yao Y. A Partition Model of Granular Computing[M]// Transactions on Rough Sets I. Springer Berlin Heidelberg, 2004:232--253.

[62] Nguyen H S. Discretization Problem for Rough Sets Methods[C]// International Conference on Rough Sets and Current Trends in Computing. Springer-Verlag, 1998:545-552.

[63] Fallahnezhad M, Moradi M H, Zaferanlouei S. A Hybrid Higher Order Neural Classifier for handling classification problems[J]. Expert Systems with Applications, 2011, 38(1):386-393.

致谢

时间如白驹过隙，转瞬即逝，两年半的研究生生涯马上接近尾声。回首这段时光，研究生生活的每个片段在脑海中浮现，是一段充满读书声、欢笑声，有汗水、有收获，激励我不断向前的旅程。在论文即将完成之际，谨向在这段旅程中关心我、帮助我、陪伴我、鼓励我的人致以最诚挚的谢意！

首先，我要感谢我的导师：傅仰耿副教授。在研究生期间对我学术上的指导，从论文选题、完成实验到论文书写上给了我很多建议和帮助。在研究过程中，导师的细心交流指导让我受益颇多，指导我如何徜徉在学术的海洋。

其次，还要感谢实验室的其他老师，他们是吴英杰教授、王一蕾副教授和孙岚老师。感谢他们的辛勤付出，在研究生期间给予我很多帮助，让我能在一个良好的环境下学习，并给我很多锻炼成长的机会。

同时，感谢李佐勇老师，对我的论文书写进行指导和修改。感谢 BRB 小组的成员：杨隆浩学长、吴伟昆学长、方志坚学长、刘莞玲学姐、李敏学长、林燕清学姐、苏曼娜、陈楠楠、方炜杰、黄宏运、庄金惠、刘永裕。感谢大家每周汇报上的分享和讨论，以及在论文阅读上的交流和帮助，帮助我解决很多科研上的问题。

感谢实验室的小伙伴和舍友在学习和生活上的关心和帮助，一起学习、一起玩耍，陪伴我走过一段快乐的时光。

最后，感谢我的家人在背后默默的关心和支持，给予我无微不至的关怀，支持我、鼓励我，让我顺利完成学业。

个人简历

姓名：张婕

出生年月：1993.02.25

性别：女

籍贯：福建省泉州市

学习经历：

2016.9-2019.3：福州大学 数学与计算机科学学院 计算机应用技术 硕士研究生

2012.9-2016.7：福州大学 数学与计算机科学学院 计算机科学与技术 本科

在学期间的研究成果及发表的学术论文

在读期间已发表和录用的论文：

第一作者（2 篇）：

【1】Zhang Jie, Fu Yanggeng, Chen Nannan, Wu Yingjie. Activation Method for Disjunctive Belief Rule Base Using Principle of Inclusion and Exclusion[C]. 2018 international Conference on Cloud Computer, Big Data and Blockchain.（已录用）

【2】张婕, 傅仰耿, 巩晓婷. 利用聚类分析构建基于析取范式的置信规则库[J]. 福州大学学报.（已录用）

参与的科研项目及成果：

【1】 国家自然科学基金项目（71501047, 61773123）

【2】 福建省自然科学基金项目（2015J01248）