

DOI: 10.11992/tis.201809022

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20181227.1144.002.html>

# 利用 MISA 多目标优化的置信规则库分类算法

林锦, 胡家琛, 刘莞玲, 吴英杰

(福州大学 数学与计算机科学学院, 福建 福州 350116)

**摘 要:** 现有基于置信规则库的分类系统的分类准确率和效率受到系统参数设置以及规则库结构合理性的影响。为了寻找到最佳的参数值和最优的规则库结构, 本文结合多目标免疫系统算法 (multiobjective immune system algorithm, MISA) 提出利用 MISA 多目标优化的置信规则库分类算法。该方法融合特征属性约简思想和差分进化算法思想建立训练模型, 采用多目标免疫系统算法对系统复杂度和分类准确率进行多目标优化, 从而寻找到分类模型的最优解。在实验分析中, 首先将本文提出的置信规则库多目标分类系统 MISA-BRM 和置信规则库分类系统的实验结果进行对比, 从复杂度和准确率两个维度说明本文方法的有效性。同时还将本文方法与现有的其他分类方法进行比较, 验证本文方法的可行性和有效性。实验结果表明, 本文方法能够有效地对基于置信规则库的分类系统的准确率和复杂度进行多目标优化。

**关键词:** 置信规则库; 分类系统; 多目标优化; 多目标免疫系统算法; 帕累托优化; 差分进化; 自适应网格; 特征属性约简

中图分类号: TP18 文献标志码: A 文章编号: 1673-4785(2019)05-0982-09

中文引用格式: 林锦, 胡家琛, 刘莞玲, 等. 利用 MISA 多目标优化的置信规则库分类算法 [J]. 智能系统学报, 2019, 14(5): 982-990.

英文引用格式: LIN Jin, HU Jiachen, LIU Wanling, et al. Belief rule base classification algorithm using MISA multi-objective optimization[J]. CAAI transactions on intelligent systems, 2019, 14(5): 982-990.

## Belief rule base classification algorithm using MISA multi-objective optimization

LIN Jin, HU Jiachen, LIU Wanling, WU Yingjie

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

**Abstract:** The efficiency and accuracy of a belief rule base (BRB) classification system are limited to the determination of the systematic parameters and the structure of the rule base. In this study, we propose the usage of a BRB classification algorithm, i.e., the multi-objective immune system algorithm (MISA), along with multi-objective optimization to determine the optimal parameters and the structure of the rule base. This method simplifies the characteristic attributes using a differential evolution algorithm to develop a training model and subsequently uses MISA to optimize the systematic complexity and the classification accuracy for identifying an optimal solution for the classification model. In the experiment, we initially compare the results of the BRM-based MISA (MISA-BRM) and those of the BRB classification system with respect to their complexity and accuracy to present the benefits of our method. Further, we compare the results with those of the existing classification methods to verify the feasibility and availability of the proposed method. The experimental results denote that the proposed method can effectively optimize the accuracy and complexity of the BRB classification system.

**Keywords:** belief rule base (BRB); classification system; multi-objective optimization; MISA; Pareto optimal; differential evolution; adaptive mesh; feature attribute reduction

收稿日期: 2018-09-13. 网络出版日期: 2018-12-28.

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

通信作者: 刘莞玲. E-mail: [380509981@qq.com](mailto:380509981@qq.com).

数据分类研究是数据挖掘领域的一个重要分支, 它主要通过已知类别的数据集构建出各类别的特征空间, 再将未知类别的数据映射到该特征

空间中,从而来确定其所属类别<sup>[1]</sup>。目前,数据分类主要应用于机器学习、模式识别、遗传算法、神经网络、统计学、数据库、专家系统等领域。模糊规则分类系统(fuzzy rule-based classification system, FRBCS)<sup>[2-4]</sup>广泛地用于求解分类问题。FRBCS能够建立清晰的语义模型,具有良好的可解释性,能够定性定量地处理专家信息。此后, Yang等<sup>[5]</sup>提出了基于证据推理的置信规则库推理方法,该方法以IF-THEN规则库、模糊理论、D-S证据理论和决策理论等为基础,有效地利用不精确或不完整的信息对复杂决策问题进行建模; Jiao等<sup>[6]</sup>在置信函数框架上扩展模糊规则库分类系统(FRBCS),提出置信规则库分类系统(belief rule base classification system, BRBCS)。置信规则库分类系统采用数据驱动方式由数据集自动生成置信规则,建立特征空间与类空间之间的不确定关系,利用基于置信函数理论的置信推理方法(belief reasoning method, BRM)对查询模式进行分类推理。但是传统的置信规则库分类系统的推理性能极为依赖于内部参数取值,因此刘莞玲等<sup>[7]</sup>在BRBCS的基础上提出基于差分的置信规则库分类方法(DEBRM)。该方法能够有效解决BRBCS中参数优化的问题,使分类系统不需要依赖于增加模糊区间划分数量来提高分类准确度。但该方法随着前提属性个数的增加,规则数量呈指数式增长,就会导致置信规则库复杂度增加。因此本文针对Liu等提出的DEBRM分类方法存在无法兼顾准确率和复杂度的问题进行改进,提出置信规则库分类系统的多目标优化模型。

多目标优化问题(multi-criterion optimization problems)涉及多个目标的极大化或极小化,同时这些目标之间往往是互相冲突的<sup>[8]</sup>,一个目标性能的优化可能导致另一个目标性能的下降。多目标的最优解不是一个特定的解,而是一组Pareto最优解,这些解没有优劣之分<sup>[9-10]</sup>。目前用于解决多目标优化问题的智能优化算法有:传统遗传算法、群集智能算法(粒子群、蚁群)、人工免疫算法、神经网络等<sup>[11]</sup>。本文拟在DEBRM分类系统中,结合多目标免疫系统算法(MISA),提出利用MISA的置信规则库分类方法,通过特征属性约简来降低复杂度并使得准确率尽可能高;通过对系统的分类准确率和复杂度进行多目标优化,从而得到一组均衡解供决策者选择。在实验中选用Iris、Wine、Cancer、Glass这4个经典的公

共测试数据集来验证本文方法的有效性。

## 1 相关概念

### 1.1 基于DE的置信规则库分类方法

置信规则库分类系统(BRBCS)由模糊规则库分类系统(FRBCS)在置信函数框架上扩展得来,采用数据驱动方式构建规则库,规则库的规则条数即结构复杂度取决于前提属性个数和模糊区间划分数。BRBCS的置信规则结构:

$$\begin{cases} \text{If } x_1 \text{ is } A_1^q \wedge x_2 \text{ is } A_2^q \wedge \cdots \wedge x_p \text{ is } A_p^q, \\ \text{then } C_q = \{(\omega_1, \beta_1^q), (\omega_2, \beta_2^q), \dots, (\omega_m, \beta_m^q)\} \end{cases} \quad (1)$$

式中:  $X = (x_1, x_2, \dots, x_p)$  表示  $p$  维模式特征向量;  $A^q = (A_1^q, A_2^q, \dots, A_p^q)$  表示前提属性,将第  $p$  个特征划分为  $n$  个模糊子集,  $A_p^q$  表示模糊子集,  $A_p^q \in \{A_{p,1}, A_{p,2}, \dots, A_{p,n_p}\}$ 。规则权重  $0 \leq \theta^q \leq 1$ ,  $q = 1, 2, \dots, Q$ ;  $Q$  是规则的总数;  $\delta_1, \delta_2, \dots, \delta_p$  表示属性权重,  $0 \leq \delta_1, \delta_2, \dots, \delta_p \leq 1$ ;  $\beta_m^q$  表示第  $m$  个评价结果的置信度,当  $\sum_{i=1}^m \beta_i^q < 1$  时,表示评价结果可能是不完整,本文规则的结果置信度和为 1。

传统基于置信规则库参数学习的分类系统(BRBCS)准确率受参数值约束,因此Liu等<sup>[7]</sup>提出基于DE的置信规则库推理的分类方法(DEBRM),该方法主要包括基于DE的参数训练和置信推理过程。

### 1.2 多目标优化基本原理

一个具有  $n$  个决策变量,  $m$  个目标函数,  $k$  个约束条件的多目标优化问题可以描述为

$$\begin{cases} \min y = f(x) = (f_1(x), f_2(x), \dots, f_m(x)) \\ g_j(x) \leq 0, j = 1, 2, \dots, l \\ h_j(x) \leq 0, j = l+1, l+2, \dots, k \end{cases} \quad (2)$$

式中:  $x = (x_1, x_2, \dots, x_n)$  为  $n$  维决策向量,  $x \in X$ ,  $X$  表示决策空间,  $X \subset \mathbf{R}^n$ ;  $m$  维目标向量  $y = (f_1(x), f_2(x), \dots, f_m(x)) \in Y \subset \mathbf{R}^m$ ,  $Y$  为目标集合;  $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$  表示映射函数。

**定义 1** Pareto 支配。若  $a, b \in X_f$  满足约束条件的可行解,当且仅当  $\forall i \in \{1, 2, \dots, m\}, f_i(a) \leq f_i(b)$  且  $\exists i \in \{1, 2, \dots, m\}, f_i(a) < f_i(b)$ , 则称  $a$  支配  $b$ , 记作  $a < b$ 。

**定义 2** Pareto 最优解。当且仅当  $\neg \exists b \in X_f$  使得  $b < a$ , 决策向量  $a \in X_f$  被称为  $X$  上的 Pareto 最优解。

**定义 3** Pareto 最优解集。即包含所有最优解的集合, 定义为  $P^* = \{a | \neg \exists b \in X_f, b < a\}$ 。

## 2 MISA-BRM 分类系统

在分类系统中,增加前提属性个数可以提高系统推理的准确率,但同时也增加了分类系统的复杂度,所以复杂度和准确率是一组相互冲突的目标。如何从训练数据集中找到使准确率尽可能大、复杂度尽可能小的折中方案具有重要意义,因此本文提出利用 MISA 多目标优化的置信规则库分类方法。该方法对数据集进行特征属性约减,并采用差分进化方法进行参数学习来获取最优结构参数,同时对经典的 MISA 算法进行适应性改进,对分类系统的准确率和复杂度进行多目标优化,从而保证分类系统的规则数尽可能少(即系统结构复杂度尽可能低)的同时,又能保证系统分类的准确率尽可能高。

最大化分类准确率即最小化分类错误率,所以本文的两个目标函数为分类错误率  $f_1(\theta)$  和置信规则库规则条数  $f_2(\theta)$ ,  $\theta$  代表置信规则库中的参数:前提属性个数、模糊区间划分数、前提属性权重、规则权重、结果置信度,  $f_1$  中参数涉及以上所列参数,  $f_2$  中参数涉及前提属性个数、模糊区间划分数。因此本文的多目标优化问题描述为

$$\min y = f(\theta) = (f_1(\theta), f_2(\theta))$$

### 2.1 改进型 MISA 多目标优化算法

MISA 算法是 Coello 等<sup>[12]</sup>根据免疫系统中抗体的克隆选择原则提出的。

#### 2.1.1 改进型 MISA 算法原理

改进型 MISA 算法将群体中的所有个体均视作“抗体”。依据克隆选择原则,选择群体中的非支配解进行克隆,根据它们在解空间的拥挤程度来决定克隆的数目,并对克隆的个体进行高概率变异和交叉操作,生成新的抗体群<sup>[12]</sup>。本文对 Coello 等提出的 MISA 针对基于置信规则库分类问题做适应性改进,并在个体变异中增加接种疫苗的变异操作。改进型 MISA 与原始型 MISA 的一个主要区别在于确定待克隆抗体的复制个体数的方法不同,改进型 MISA 根据待克隆抗体所处网格密度及支配情况确定最大可复制数,再根据抗体群所能容纳的最大个体数确定最终的复制数量,确定待克隆抗体复制个体数的具体方法如下:

##### 1) 确定待克隆抗体最大可复制个数

按照个体在外部种群(外部种群中个体互为非支配解)中所处网格的密度及支配关系分配等级,按照等级确定最大的克隆数。分4个等级:所处网格密度小于平均网格密度,等级为4,最多可

复制4个个体;所处网格密度等于平均网格密度,等级为3,最多可复制3个个体;所处网格密度大于平均网格密度,等级为2;个体被支配且被选为待克隆抗体则等级为1,最多可复制1个个体。

##### 2) 确定待克隆抗体的复制个数

抗体群设置最多可容纳个体数。将抗体群可容纳个体数视为复制名额个数,具体过程如下:

①待克隆抗体按照优先级:被支配数少>所处网格密度小>支配其他个体数多排成一系列纵队;

②从队头到队尾依次查看待克隆抗体的等级,当待克隆抗体等级大于0且抗体群可容纳个体数大于0时,待克隆抗体获得1个复制名额,每分配出1个名额,获得名额的待克隆抗体等级减1,抗体群可容纳数减1;当抗体群可容纳个体数等于0时结束分配;

③走到队尾整个抗体群的抗体等级累加和大于0,返回执行②。

除了待克隆抗体的克隆个体数确定方式不同外,改进型 MISA 采用二进制染色体编码,抗体长度等于最大前提属性个数,取值为0代表删除这个前提属性,取值为1代表选中这一前提属性。确定前提属性后构建规则库,规则库编码采用实数编码,一个规则库作为一个个体;无交叉操作,变异操作包括单点变异和接种疫苗,具体的提取疫苗和接种疫苗方法说明如下。

1) 提取疫苗:疫苗群体限制最大个数,抗体被选为疫苗的条件为被支配数为0,支配数大于0。根据 Pareto 的定义,能够支配其他个体的个体不是属性最少也不是属性最多的个体,而是属性个数居中的个体,如选取的特征属性个数为5,受它支配的解是属性个数大于5且分类错误率大于它的个体,这一定程度上说明这5个属性中可能具有区分性较高的特征属性或特征属性组合。

2) 接种疫苗操作:随机从免疫群体中随机选取一个个体,从疫苗染色体中所有值为1的染色体单位中随机选取1个单位,再从所有值为0的染色体单位中随机选取1个单位,将这2个单位植入个体中。植入个体后,有4种可能情况:个体特征属性减少或增加一个;个体特征属性完全不变;个体属性特征个数不变但选取的特征属性改变;有可能使个体具有高区分性的特征属性,或删除不必要的属性。

#### 2.1.2 外部种群及自适应网格算法

外部种群用于保存 Pareto 非劣解,外部种群中的个体两两之间不存在支配与被支配关系。为了使非支配解尽可能均匀地分布于 Pareto 前沿,



外部种群采用 Knowles 等<sup>[13]</sup>提出的自适应网格法进行更新,详见算法1。

#### 算法1 自适应网格算法

输入 抗体群中所有非支配解  $Q$ , 外部种群 EP(external population);

输出 Pareto 最优前沿。

1) 输入抗体群中所有非支配解组成的集合  $Q$ ;

2) 循环遍历  $Q$ ;

3) 判断  $Q_i$  是否被外部种群中任一个个体支配: 如果  $\exists a_{nd} \in EP: a_{nd} < Q_i$  ( $a_{nd}$  表示 EP 中支配  $Q_i$  的个体), 那么删除  $Q_i$ , 否则执行 4);

4) 判断  $Q_i$  是否支配 EP 中任一解: 如果  $\exists a_d \in EP: Q_i < a_d$  ( $a_d$  表示 EP 中所有被  $Q_i$  支配的个体), 那么删除  $a_d$ , 将  $Q_i$  加入 EP, 否则执行 5);

5) 判断外部种群个体数是否达到最大值: 如果外部种群个体数达到所能容纳的最大值, 那么执行 6), 否则将  $Q_i$  加入 EP;

6) 按照所给划分分数构建自适应网格: 如果  $|N_{Q_i}| + 1 \geq \max N$  ( $|N_{Q_i}|$  表示  $Q_i$  所在网格未加入  $Q_i$  的密度,  $\max N$  表示所有网格的密度最大值, 这里所说的密度其实是指网格包含的个体数), 那么删除  $Q_i$ , 否则去掉档案中密度最大的网格中的任意一个解, 加入  $Q_i$ ;

7) 循环结束。

#### 2.1.3 改进型 MISA 算法流程

改进型 MISA 算法的算法流程的主要部分和原始 MISA 算法相同: 生成抗体群, 更新自适应网格, 选择待克隆抗体, 确定待克隆抗体的复制数量。但改进型 MISA 在变异操作中加入了接种疫苗操作, 因此增加了疫苗提取的过程, 具体流程如算法 2~4 所示。

#### 算法2 疫苗提取及更新算法

输入 抗体群;

输出 更新后的疫苗群体。

1) 选取出抗体群中被支配数为 0, 支配数大于 0 的个体, 将选出的个体按照支配个体数降序排列形成群体  $P$ ;

2) 疫苗提取及更新: 如果疫苗群体为空, 在符合群体  $P$  个体数和疫苗群体个体数最大值的约束条件下, 按队列先后顺序将群体  $P$  中个体复制到疫苗群体中; 否则, 将群体  $P$  中个体和疫苗群体中个体合并起来, 并按照支配个体数降序排列形成群体  $P'$ , 在符合群体  $P'$  个体数和疫苗群体个体数最大值的约束条件下, 按队列先后顺序将群体  $P'$  中个体复制到疫苗群体中。

#### 算法3 加入接种疫苗的多目标变异算法

输入 被选中将要进行克隆的个体;

输出 经过变异/接种疫苗生成新抗体群。

1) 将待克隆抗体按照可克隆个数进行克隆, 生成新的初始抗体群  $V$ ;

2) 循环遍历  $V$ ;

3) 随机选取  $[0, 1]$  的一个值  $r$ ;

4) 如果  $r < pc$ , 那么进行变异: 对  $x \in V$  的个体随机选取一位进行翻转;

5) 否则进行疫苗接种: 随机选取一个疫苗  $w$ ; 随机选取  $w$  中染色体单位值为 0 的片段  $w_0$ ; 随机选取  $w$  中染色体单位值为 1 的片段  $w_1$ ; 对  $x \in V$  的个体植入  $w_0$ 、 $w_1$ ;

6) 循环结束。

利用 MISA 多目标优化的置信规则库分类算法有机融合了算法 1、2、3、MISA 和 DE 参数训练方法。多目标免疫系统算法的选择进化机制是整个算法的核心, 使得差分训练规则库参数和特征属性选择组合两个优化过程相互促进, 最终获得一组 Pareto 最优解。DE 参数训练方法对根据抗体构建的规则库进行参数优化, 得到不同的特征属性选择组合构建的置信规则库的近似最优分类性能, 进而采用 MISA 的选择机制和自适应网格进行抗体选择, 将分类性能好的特征属性组合保留在外部种群中, 同时也将区分性高的特征属性保留下来, 在抗体进化的时候, 采用变异增加抗体群的多样性, 疫苗提取、更新和接种疫苗算法使得区分性高的特征属性在抗体群中扩散, 让抗体能够得到更优的分类性能, 迭代更新外部种群, 不断优化外部种群中抗体的特征属性选择组合、分类性能。

算法4 利用 MISA 多目标优化的置信规则库分类算法

输入 无;

输出 Pareto 最优前沿。

1) 初始化抗体群, 初始化外部种群为空;

2) 对由每个抗体构建得来的置信规则库采用 DE 参数训练方法进行参数训练;

3) 计算每个抗体的适应值, 进而判断每个抗体是否是 Pareto 解;

4) 将得到的非支配个体复制到外部种群 (调用算法 1);

5) 按一定的标准选取待克隆抗体;

6) 按一定的标准确定待克隆抗体的复制数量;

7) 抽取并更新疫苗 (调用算法 2);

8) 如果达到 MISA 算法最大迭代次数则 end, 否则进行 9);

9) 克隆抗体, 并对个体进行变异或疫苗接种 (调用算法 3);

10) 返回到 2)。

## 2.2 MISA-BRM 分类模型

在 MISA-BRM 分类方法中, 将前提属性的选取编码作为抗体, 根据抗体构建出初始的规则库, 将置信推理方法 BRM 作为推理机, 结合差分进化算法对待优化参数进行学习矫正, 进而计算抗体在复杂度、分类错误率两个目标函数上的适应值, 然后采用 MISA 的选择进化机制, 在系统的复杂度、分类错误率两个维度上寻找 Pareto 解。算法的流程如图 1 所示。

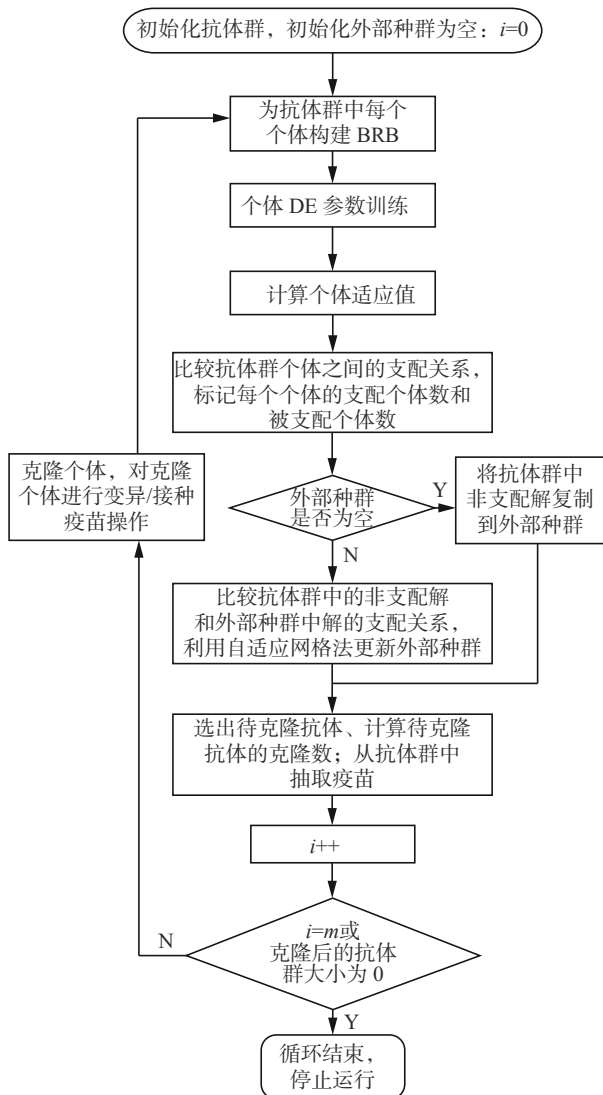


图 1 MISA-BRM 模型

Fig. 1 MISA-BRM model diagram

## 3 实验分析

### 3.1 实验环境

实验环境的基本信息: Intel(R) Core(TM) i7-6700 CPU @3.40 GHz 处理器, 8 核, 16 GB 内存,

Window10 操作系统。

### 3.2 数据集参数设置

#### 3.2.1 实验数据集

在实验部分本文选用 University of California at Irvine 分校网页中获取的公共测试集来验证算法有效性。数据集主要包括鸢尾花特征数据 Iris、红酒化学成分特征数据 Wine、乳腺癌数据 Cancer 和玻璃类型数据 Glass。表 1 列举了这 4 个测试数据集中前提属性数量、类别数量和数据集大小的信息。

其中 Cancer 原数据集中存在缺失数据, 实验中将缺失的 16 条数据删除, 保留 683 条数据。

表 1 测试数据的基本信息

Table 1 Basic information regarding the test data

数据集	前提属性数量	类别数量	数据量
Iris	4	3	150
Wine	13	3	178
Cancer	9	2	683
Glass	9	6	214

#### 3.2.2 参数设置

本文置信规则库的模糊区间划分数  $K=3$ 。本文所采用的数据集的数据量及前提属性个数不同, 因此实验中针对不同数据集设置不同的参数, 表 2、表 3 列举了 4 个测试数据集实验的参数设置信息。

表 2 DEBRM 参数设置信息

Table 2 The parameter setting information for DEBRM

DEBRM 参数	数据集			
	Iris	Wine	Cancer	Glass
$F$	0.5	0.5	0.5	0.5
CR	0.9	0.9	0.9	0.9
种群大小/个	100	50	50	50
迭代次数	3 000	3 000	3 000	3 000

表 3 MISA 参数设置信息

Table 3 The parameter setting information for MISA

MISA 参数	数据集			
	Iris	Wine	Cancer	Glass
变异概率	0.5	0.5	0.5	0.5
接种疫苗概率	0.5	0.5	0.5	0.5
疫苗群最大值	3	7	5	5
外部种群大小	20	20	20	20
抗体群最大值	15	15	15	15
迭代次数	8	8	8	8

### 3.3 数据集实验分析

实验中针对每组数据集,采用十折交叉验证方法对提出的算法进行测试,数据集划分方法采用分层抽样方法。实验中用总的数据集(训练集+测试集)的错误率进行 Pareto 适应值计算。针对每一数据集,取 10 组实验中在整个数据集上获得的最高准确率值最大的一组的实验结果来展示,采用图表方式展示实验结果。用图 2~5 分析整个数据集的 Pareto 最优解集分布情况,图中的点是外部种群中保存的 Pareto 非劣解,图中的曲线是外部种群中所有 Pareto 最优解对应的目标向量组成的 Pareto 最优曲线;用表格数据分析数据集的 Pareto 最优解集及其对应的训练集和测试集的实验结果。

#### 3.3.1 Iris 数据集实验分析

在 Iris 数据集中,每个类别的数据分布均匀。本文方法解决 Iris 数据集分类问题的 Pareto 最优曲线如图 2 所示。对 Iris 的训练数据集、测试数据集及整个数据集上的分类准确率如表 4 所示。

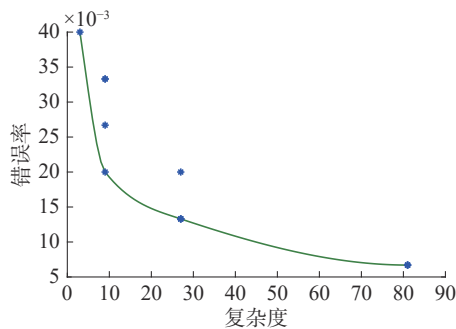


图 2 Iris 数据集的 Pareto 最优曲线

Fig. 2 The Pareto optimal curve for the Iris dataset

表 4 Iris 数据集上 MISA-BRM 的 Pareto 最优解集

Table 4 The Pareto optimal solution set of MISA-BRM for the Iris dataset

前提属性个数	规则条数	训练集准确率/%	测试集准确率/%	总数据准确率/%
4	81	99.26	100	99.33
3	27	98.52	100	98.67
2	9	97.78	100	98.00
1	3	95.56	100	96.00

通过以上实验结果分析可知,在 Iris 总数据集上, Pareto 最优解集中 4 种不同的前提属性个数的分类准确率差距较小,其中 4 个前提属性的分类结果与 3 个前提属性分类的准确率仅相差 0.66%。相同的前提属性个数,选取不同的前提属性构建规则库分类系统得到的分类准确率不同。通过实验可知,针对 Iris,本文方法通过对分类系统进行多目标优化可以在满足较高准确率情况下构建出复杂性更低的分类模型。

为了进一步说明本文方法的有效性,将本文方法与 BRBCS、DEBRM 分类方法解决 Iris 数据集分类问题时构建的分类系统的复杂度和准确率进行对比,如表 5 所示。通过表 5 分析可知,传统的 BRBCS 规则条数有两千多条,而本文方法将规则条数约简到二十多条,且本文方法在整个数据集的分类准确率比传统的 BRBCS 高。相比于 DEBRM,规则约简了 54 条,测试数据集的分类准确率与 DEBRM 持平,训练数据集的分类准确率低了 1.48%,总的数据集分类准确率低了 1.33%,尽管该方法牺牲了一定的准确率,但其降低了系统复杂度,可以为决策者根据实际需要提供更多的选择。

表 5 Iris 数据集结构复杂度和分类准确率对比

Table 5 Comparison of the structural complexity and classification accuracy with respect to the Iris dataset

方法	规则条数	训练集准确率/%	测试集准确率/%	数据集准确率/%
BRBCS	$7^4$	96.67	96.67	96.67
DEBRM	$3^4$	100	100	100
MISA-BRM	$3^3$	98.52	100	98.67

#### 3.3.2 Wine 数据集实验分析

在 Wine 数据集中,各类别的数据量分布较不均匀。本文方法用于解决 Wine 数据集的 Pareto 最优,曲线如图 3 所示。

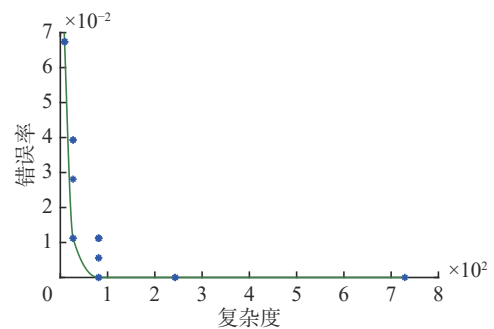


图 3 Wine 数据集的 Pareto 最优曲线

Fig. 3 The Pareto optimal curve of the entire Wine dataset

对 Wine 的训练数据集、测试数据集及整个数据集上的分类准确率如表 6 所示。

表 6 Wine 数据集上 MISA-BRM 的 Pareto 最优解集

Table 6 The Pareto optimal solution set of MISA-BRM with respect to the Wine dataset

前提属性个数	规则条数	训练集准确率/%	测试集准确率/%	总数据准确率/%
6	729	100	100	100
5	243	100	100	100
4	81	100	100	100
3	27	98.38	94.12	98.88
2	9	94.41	82.35	93.26



通过图3及表6分析可知,在Wine数据集上,取其中4个前提属性能获得100%的分类准确率,相比于取13个前提属性规则条数由上万条减少为几十条,且分类准确率得到提高;通过实验可知,针对Wine数据集前提属性个数在4~6,系统能够获得较好的分类效果。

为了进一步说明本文方法的有效性,将本文方法与DEBRM分类方法解决Wine数据集分类问题时构建的分类系统的复杂度和准确率进行对比,如表7所示。

表7 Wine数据集结构复杂度和分类准确率对比  
Table 7 Comparison of the structural complexity and classification accuracy with respect to the Wine dataset

方法	规则条数	训练集准确率/%	测试集准确率/%	总数据准确率/%
DEBRM	$3^{13}$	99.44	99.44	99.44
MISA-BRM	$3^4$	100%	100%	100%

通过表7分析可知,相比于Liu等提出的DEBRM分类系统,基于MISA多目标优化的置信规则库分类系统的规则数量由 $3^{13}$ 约简到 $3^4$ ,约简了1 594 242条,不仅极大降低了系统复杂度,且系统分类准确率达到100%。可见,系统分类准确率与系统复杂性并不呈正比关系,有时规则越多,推理准确率反而越低。

### 3.3.3 Cancer数据集实验分析

Cancer数据集总的的数据量较大,同样各类别数据分布较不均匀。本文方法用于解决Cancer数据集的Pareto最优曲线,如图4所示。对Cancer训练数据集、测试数据集及整个数据集上的分类准确率如表8所示。

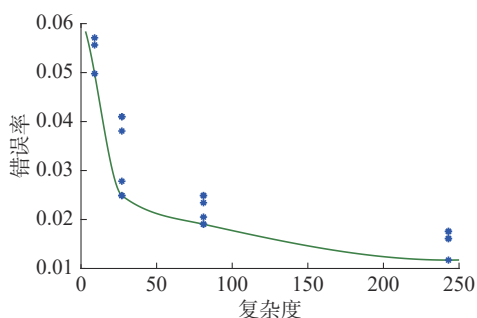


图4 Cancer数据集的Pareto最优曲线

Fig. 4 The Pareto optimal curve of the entire Cancer dataset

表8 Cancer数据集上MISA-BRM Pareto最优解集  
Table 8 The Pareto optimal solution set of MISA-BRM using the Cancer dataset

前提属性个数	规则条数	训练集准确率/%	测试集准确率/%	总数据准确率/%
5	243	99.02	97.06	98.83
4	81	98.21	97.06	98.10
3	27	97.56	97.06	97.51
2	9	94.96	95.59	95.02

通过图4和表8分析可知,Cancer数据集上取4~5个前提属性时在训练数据集上的分类准确率相差0.81%,测试数据集的分类效果相同。可知,针对Cancer数据集,该模型在保证系统复杂度尽可能低的情况下,还能有效保证系统的分类准确。

为了进一步说明本方法的有效性,将本文方法与BRBCS、DEBRM分类方法解决Cancer数据集分类问题时构建的分类系统的复杂度和准确率进行对比,如表9所示。

表9 Cancer数据集结构复杂度和分类准确率对比  
Table 9 Comparison of the structural complexity and classification accuracy with respect to the Cancer dataset

方法	规则条数	训练集准确率/%	测试集准确率/%	总数据准确率/%
BRBCS	$5^9$	99.82	99.74	99.20
DEBRM	$3^9$	99.82	99.54	99.56
MISA-BRM	$3^5$	99.02	98.06	98.83

通过表9分析可知,相比于BRBCS、DEBRM,在Cancer总数据集上,MISA-BRM分类方法的准确率分别降低0.37%、0.73%。但系统的规则条数则约简了上万条。因此,当对系统复杂度要求较高时,MISA-BRM在准确率和复杂度两个标准的均衡下有较好的效果。

### 3.3.4 Glass数据集实验分析

Glass数据集有6个类别,各类别数据分布不均匀,且有的类别的数据量占总的数据量的比率小。本文方法用于解决Glass数据集的Pareto最优曲线,如图5所示。

对Glass的训练数据集、测试数据集及整个数据集上的分类准确率如表10所示。

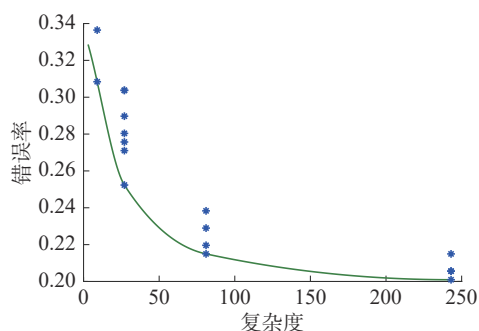


图 5 Glass 数据集的 Pareto 最优曲线

Fig. 5 The Pareto optimal curve of the entire Glass dataset

表 10 Glass 数据集 MISA-BRM 的 Pareto 最优解集

Table 10 The Pareto optimal solution set of MISA-BRM with respect to the Glass dataset

前提属性 个数	规则 条数	训练集 准确率/%	测试集 准确率/%	总数据 准确率/%
5	243	80.21	77.27	79.91
4	81	78.65	77.27	78.50
3	27	75.52	68.18	74.77
2	9	70.83	54.55	69.16

将本文方法与 BRBCS、DEBRM 分类方法解决 Glass 数据集分类问题时构建的分类系统的复杂度和准确率进行对比, 如表 11 所示。

表 11 Glass 数据集结构复杂度和分类准确率对比

Table 11 Comparison of the structural complexity and classification accuracy with respect to the Glass dataset

方法	规则 条数	训练集 准确率/%	测试集 准确率/%	总数据 准确率/%
BRBCS	5 <sup>9</sup>	94.77	71.84	90.18
DEBRM	3 <sup>9</sup>	83.04	72.09	80.84
MISA-BRM	3 <sup>5</sup>	80.21	77.27	79.91

分析表 10 和表 11 可知: 本文方法在前提属性个数为 4 个或 5 个时系统能够获得较好的分类效果; 本文方法在 Glass 数据集上的分类准确率比 BRBCS 和 DEBRM 分别降低了 14.56%、2.83%, 但规则条数分别减少了 1952882 条、19440 条, 虽然牺牲了一定的分类准确率, 但规则数极大降低。对于 Glass 测试数据集不仅极大约简了规则条数且分类准确率得到提高。

综上, 通过 Iris 数据集、Wine 数据集、Cancer 数据集、Glass 数据集的实验结果分析可知: 本文采用属性约简思想、差分参数训练方法, 并结合改进型 MISA 算法对置信规则库分类系统的复杂度和准确度进行多目标优化是有效可行的。

### 3.4 与其他分类方法的准确率比较

为了进一步说明本文方法的有效性, 本节对比了 MISA-BRM 与现有其他分类方法<sup>[14]</sup>用于解决 Iris 数据集、Cancer 数据集、Wine 数据集、Glass 数据集分类问题时的准确率, 详见表 12 和表 13。

表 12 MISA-BRB 和其他分类方法的比较

Table 12 Comparison with other classification methods %

分类方法	数据集			
	Iris	Cancer	Wine	Glass
Naïve Bayes	96.30	96.10	99.07	44.60
C4.5	95.33	94.80	96.26	68.40
AMO	99.27	98.48	99.98	65.43
Fuzzy-Gain Measure	99.28	99.04	99.62	73.83
Fallahnezhad	99.77	98.34	100	64.08
BRB-based-total	100	98.77	97.19	68.24
MISA-BRM-total	98.67(3)	98.83(5)	100(4)	79.91(5)

表 13 MISA-BRB 和其他建模方法的比较

Table 13 Comparison with other modeling methods %

分类方法	数据集名称			
	Iris	Wine	Glass	Cancer
NSGA-II-FRBCS	96.67	98.88	71.50	—
混合烟花-差分进化	98.7	98.3	—	98.1
MISA-BRM	98.67	100	79.91	98.83

对比表 12 中数据, MISA-BRB 分类方法在 Iris 数据集上准确率比 Naïve Bayes、C4.5 方法高, 在 Cancer 数据集上准确率仅比 Fuzzy-Gain Measure 方法低; 在 Wine 数据集上, 得到 100% 的最好结果; 在 Glass 数据集上, 相比于其他 6 中分类方法, MISA-BRB 的准确率最高。

在表 13 中, NSGA-II-FRBCS<sup>[15]</sup>分类方法、基于 Pareto 混合烟花-差分进化算法<sup>[16]</sup>(表中称混合烟花-差分进化)都是以准确率、解释性为目标, 进行多目标优化的模糊规则库分类方法。从表 13 可得, 针对 Iris 数据集、Wine 数据集和 Glass 数据集, 利用 MISA 多目标优化的置信规则库分类方法比 NSGA-II-FRBCS 准确率更高; 利用 MISA 多目标优化的置信规则库分类方法和基于 Pareto 混合烟花-差分进化算法相比, 在 Cancer 数据集上准确率略高, 在 Wine 数据集上准确率较高。

## 4 结束语

本文首次针对提高置信规则库分类系统准确



率和降低系统复杂度的多目标优化问题进行建模。算法采用基于特征属性序列编码的方法进行特征搜索,使得在规则数量减少的情况下尽可能提高分类准确率。此外,对经典的MISA算法进行改进,加入基于Pareto支配关系的个体疫苗提取和接种疫苗操作,提出MISA-BRM模型。实验结果表明,该算法能找到一组具有不同准确率和复杂度的Pareto最优解集,方便决策者根据实际需要进行选择。

## 参考文献:

- [1] 魏茂胜. 数据挖掘中的分类算法综述 [J]. [网络安全技术与应用](#), 2017(6): 65–66.  
WEI Maosheng. Overview of classification algorithms in data mining[J]. [Network security technology and application](#), 2017(6): 65–66.
- [2] TSAKIRIDIS N L, THEOCHARIS J B, ZALIDIS G C. DECO3R: a differential evolution-based algorithm for generating compact fuzzy rule-based classification systems[J]. [Knowledge-based systems](#), 2016, 105: 160–174.
- [3] PRUSTY M R, JAYANTHI T, CHAKRABORTY J, et al. Performance analysis of fuzzy rule based classification system for transient identification in nuclear power plant[J]. [Annals of nuclear energy](#), 2015, 76: 63–74.
- [4] LIUBICH L D, KOVALEVSKA L M, LISYANY M I, et al. TGF- $\beta$ 1 expression by glioma C6 cells in vitro[J]. [Experimental oncology](#), 2017, 39(4): 258–263.
- [5] YANG Jianbo, LIU Jun, WANG Jin, et al. Belief rule-based inference methodology using the evidential reasoning Approach-RIMER[J]. [IEEE transactions on systems, man, and cybernetics-part a: systems and humans](#), 2006, 36(2): 266–285.
- [6] JIAO Lianmeng, PAN Quan, DENÈUX T, et al. Belief rule-based classification system: Extension of FRBCS in belief functions framework[J]. [Information sciences](#), 2015, 309: 26–49.
- [7] 刘莞玲, 王韩杰, 傅仰耿, 等. 基于差分进化算法的置信规则库推理的分类方法 [J]. [中国科学技术大学学报](#), 2016, 46(9): 764–773.  
LIU Wanling, WANG Hanjie, FU Yanggeng, et al. Belief rule based inference methodology for classification based on differential evolution algorithm[J]. [Journal of University of Science and Technology of China](#), 2016, 46(9): 764–773.
- [8] MIRJALILI S, SAREMI S, MIRJALILI S M, et al. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization[J]. [Expert systems with applications](#), 2016, 47: 106–119.
- [9] HOU Yan, WU Naiqi, ZHOU Mengchu, et al. Pareto-optimization for scheduling of crude oil operations in refinery via genetic algorithm[J]. [IEEE transactions on systems, man, and cybernetics: systems](#), 2017, 47(3): 517–530.
- [10] 范立云, 周伟, 刘鹏, 等. 高速电磁阀动态响应的多目标优化 [J]. [哈尔滨工程大学学报](#), 2018, 39(1): 53–59.  
FAN Liyun, ZHOU Wei, LIU Peng, et al. Multi objective optimization of dynamic response of high speed solenoid valve[J]. [Journal of Harbin Engineering University](#), 2018, 39(1): 53–59.
- [11] 童晶. 多目标优化的 Pareto 解的表达与求取 [D]. 武汉: 武汉科技大学, 2009.  
TONG Jing. Expressing and obtaining Pareto solutions in multi-objective optimization problem[D]. Wuhan: Wuhan University of Science and Technology, 2009.
- [12] COELLO C A, CORTÉS N C. Solving multiobjective optimization problems using an artificial immune system[J]. [Genetic programming and evolvable machines](#), 2005, 6(2): 163–190.
- [13] KNOWLES J, OATES M, CORNE D. Advanced multi-objective evolutionary algorithms applied to two problems in telecommunications[J]. [BT technology journal](#), 2000, 18(4): 51–65.
- [14] FALLAHNEZHAD M, MORADI M H, ZAFERAN-LOUEI S. A hybrid higher order neural classifier for handling classification problems[J]. [Expert systems with applications](#), 2011, 38(1): 386–393.
- [15] 李继东. 遗传模糊分类系统构建中规则获取和解释性优化的关键技术研究 [D]. 昆明: 云南大学, 2011.  
LI Jidong. Research on key techniques of rule acquisition and interpretive optimization in the construction of genetic fuzzy classification system[D]. Kunming: Yunnan University, 2011.
- [16] 刘冲. 基于烟花算法的模糊建模方法研究 [D]. 郑州: 郑州大学, 2016.  
Liu Chong. Research on fuzzy modeling method based on fireworks algorithm [D]. Zhengzhou: Zhengzhou University, 2016.

## 作者简介:



林锦, 女, 1995 年生, 硕士研究生, 主要研究方向为智能图像处理、智能系统。



胡家琛, 男, 1995 年生, 硕士研究生, 主要研究方向为图像识别、决策理论方法、数据挖掘与机器学习。



刘莞玲, 女, 1991 年生, 硕士研究生, 主要研究方向为决策理论方法、数据挖掘与机器学习、智能系统。