

控制与决策

Control and Decision

基于知识粒度特征的多目标粗糙集属性约简算法

李雪岩, 李学伟, 蒋君

引用本文:

李雪岩, 李学伟, 蒋君. 基于知识粒度特征的多目标粗糙集属性约简算法[J]. *控制与决策*, 2021, 36(1): 196–205.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0490>

您可能感兴趣的其他文章

Articles you may be interested in

基于改进多目标优化算法的分布式数据中心负载调度

Multi-objective optimization of energy and performance management in distributed data centers

控制与决策. 2021, 36(1): 159–165 <https://doi.org/10.13195/j.kzyjc.2019.0702>

基于负荷平衡的柔性预约决策

Flexible outpatient appointment decision model with loading balance

控制与决策. 2021, 36(1): 226–233 <https://doi.org/10.13195/j.kzyjc.2019.1690>

基于稀疏度阶数优化的杂波密度估计算法

A clutter density estimation algorithm by optimized sparsity order

控制与决策. 2020, 35(12): 2923–2930 <https://doi.org/10.13195/j.kzyjc.2019.0429>

基于树形结构无界存档的多目标粒子群算法

Multi-objective particle swarm optimization algorithm based on tree-structured unbounded archive

控制与决策. 2020, 35(11): 2675–2686 <https://doi.org/10.13195/j.kzyjc.2019.0276>

基于社交网络的双知识表达分类方法

Double knowledge representations based classification method from perspective of social networks

控制与决策. 2020, 35(11): 2653–2664 <https://doi.org/10.13195/j.kzyjc.2019.0141>

基于知识粒度特征的多目标粗糙集属性约简算法

李雪岩^{1†}, 李学伟¹, 蒋 君²

(1. 北京联合大学 管理学院, 北京 100101; 2. 北京交通大学 经济管理学院, 北京 100044)

摘要: 针对多知识粒度粗糙集在条件属性权重计算及约简过程中易忽略单个属性序列产生的等价划分的问题, 引入帕累托最优思想, 同时考虑基于等价关系的知识依赖分辨度以及属性的重要性程度, 将多粒度粗糙集属性约简问题转化为离散多目标优化问题. 针对该问题的结构设计具有集群智能优化思想及复杂网络拓扑结构的优化算法, 在算法中引入基于个体的非支配解集以平衡局部最优与全局最优的关系, 引入基于“均值-方差”的遗传算子增加种群多样性. 以 UCI 中的测试数据集作为算例构建粗糙集决策表进行优化计算, 引入多种智能算法进行性能比较, 依据约简结果, 利用多层感知机对数据集中的对象进行分类, 验证约简方法的有效性. 研究表明: 所提出方法具有更强的多目标属性挖掘性能; 基于帕累托最优思想的多目标属性约简方法能较好地综合知识分辨度与知识粒度建模方式的优点, 提升数据集的分类精度.

关键词: 知识粒度; 粗糙集; 多目标优化; 集群智能; 复杂网络; 多层感知机

中图分类号: O159; O224

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0490

开放科学(资源服务)标识码(OSID):



引用格式: 李雪岩, 李学伟, 蒋君. 基于知识粒度特征的多目标粗糙集属性约简算法[J]. 控制与决策, 2021, 36(1): 196-205.

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

LI Xue-yan^{1†}, LI Xue-wei¹, JIANG Jun²

(1. School of Management, Beijing Union University, Beijing 100101, China; 2. School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China)

Abstract: In the process of multi-granulation rough set's weight calculation and the condition attributes reduction, the equivalence partition produced by single attribute is usually ignored. Therefore, the attribute reduction problem of multi-granulation rough sets is transformed into the discrete multi-objective optimization problem by introducing the idea of Pareto optimality, in which both resolution of dependability based on equivalence relation and attributes' significance are taken into consideration. For this optimization problem, a swarm intelligent optimization algorithm with complex network based population structure is designed, in which the non-dominant solution set for individual is introduced to balance the local optimum and global optimum, and the mean-variance based genetic operator is also designed to increase the diversity of the population. The optimization is conducted based on the rough set decision tables which are obtained from the test data sets in UC Irvine Machine Learning Repository, some other multi objective intelligent algorithms are also introduced as comparison, then based on the reduction results, the multi-layer perceptron is introduced to classify the samples in data sets, and the validity of the proposed algorithm is verified. The results show that: 1) The algorithm proposed in this paper shows better performance in multi objective attribute reduction; 2) The algorithm of multi objective rough set attribute reduction combines the advantage of knowledge resolution and knowledge granularity well, and also improves the classification accuracy of the data sets.

Keywords: knowledge granularity; rough set; multi objective optimization; swarm intelligence; complex network; multi-layer perceptron

0 引言

自 1982 年波兰数学家 Pawlak 提出粗糙集理论^[1]以来, 该理论已在数据挖掘、模式识别、机器学习等领

域与其他方法得到了广泛结合. 近年来, 随着大数据以及智能计算技术的不断发展, 多源异构数据的处理需求不断增加^[2], 粗糙集的知识提取、特征选择功能

收稿日期: 2019-04-18; 修回日期: 2019-06-03.

基金项目: 中国国家铁路集团有限公司科技研究开发计划课题(K2019Z006).

责任编辑: 李登峰.

[†]通讯作者. E-mail: gongye1632006@163.com.

更加具有重要的理论及实践意义,知识粒度及相关的粒计算问题日益成为研究的热点。

就知识颗粒的度量而言,粗糙集的知识粒度有两个重要的研究维度,第1个维度从条件属性自身出发,重点在于刻画条件属性对论域内对象的划分所产生的权重问题,从而使计算结果更加接近实际。目前,对于传统的条件属性权重计算方式,已有不少研究从知识粒度、信息熵、模糊性等方面进行不断改进,如文献[3]将信息熵引入粗糙集属性重要程度计算,有效避免了权重为零的情形,提升了数据的分辨能力;文献[4]在信息熵的基础上引入模糊计算的概念作为属性筛选的标准;文献[5]将模糊集和优势粗糙集进行结合,构建了论域内对象关于属性集综合优势度的决策方法;文献[6]在粗糙集理论上提出了基于条件信息熵、粗糙度、精度的权重计算方法;文献[7]在总结了多数权重计算方式缺点的基础上将知识粒度理论引入特征权重计算。第2个维度则从粒计算的角度出发,采用条件属性序列对论域内的对象进行综合分类。以Qian等的研究^[8]为代表,提出了乐观多粒度粗糙集模型与悲观多粒度粗糙集模型,进而也产生了一系列多粒度条件下的属性约简研究,如文献[9]在多粒度粗糙集属性约简中设计了全新的粒化准则,提升了约简效率;文献[10]将直觉模糊粗糙集与多粒度粗糙集进行结合并提出相应的近似分布约简算法,得到了更加合理的约简结果;文献[11]将知识粒度的概念引入模糊邻域粗糙集模型,综合考虑属性依赖度和知识粒度两种不同的属性评估标准进行属性约简。

在属性约简方面,由于第1个维度上的研究大多是从单一条件属性的特征出发,属性权重的计算方式日趋完善。就知识粒度而言,目前已有研究针对不同等价划分得到相同下近似集的问题进行了基于知识分辨率的改进^[7,12],有效避免了不同等价划分得出相同属性权重的问题,确立了知识分辨率在属性权重计算中的重要性。然而,就上述第2个维度而言,在确定多粒度粗糙集的上、下近似集时,往往又需要对属性序列子集所产生的等价类进行进一步集合运算(如交、并)^[8],进而容易忽略知识分辨率的差异。可见,在基于条件属性序列的多粒度粗糙集中有效引入第1维度研究中基于知识分辨率的改进,既可以有效发挥多粒度粗糙集处理多源异构数据的优势,又可以使基于知识粒度的属性约简更加精确化。

进一步,属性约简问题可以转化为一个以条件属性重要性程度为目标函数、条件属性组合方式为变量的优化问题^[13]。本文针对上述问题,采用多目标优

化的思想对知识分辨率和多粒度粗糙集进行有机结合,避免了先验知识对属性约简的影响,在多粒度粗糙集属性约简的过程中同时考虑属性组合的重要性程度以及属性组合的知识分辨率,利用帕累托前沿建立属性组合重要性程度与知识分辨率之间的协调机制,形成一个基于离散变量的多目标优化问题。然而,对于基于离散变量的多目标优化问题,当变量数量较多时,则变为典型的NP-hard问题,一些传统的优化算法具有容易陷入局部最优、求解质量不稳定等缺点;因此,需要针对所求解问题的结构对算法进行设计和改进。近年来,随着多目标优化领域研究的不断深入,一方面,具有多智能体集群结构、交互性强的智能算法往往可以取得良好的优化效果^[14];另一方面,优化算子的提出不断倾向于全局最优与局部最优的协调演化^[15]。在这些算法中,MOcell^[16]与BBMOPSO^[14]分别代表了两种典型的改进思想,其中MOcell采用具有局部邻居交互特点的网络结构模拟种群分布,有效避免了算法过早收敛,但也会降低优秀信息在种群中的传播速度;而粒子群算法(PSO)本身具有收敛迅速、易过早陷入局部最优的缺点,BBMOPSO则是通过设计能够产生具有良好多样性后代的交叉算子提升优化的精确性,但由于粒子群所特有的飞行特点,依然存在早熟风险。因此,一些多目标优化算法也开始被引入粗糙集问题的多目标约简中:文献[17]基于正域、决策损失及交互信息3种约简原则建立了三目标粗糙集约简模型;文献[18]同时考虑最大化属性分类性能及最小化属性约简个数建立了双目标粗糙集属性约简模型,并采用粒子群算法进行求解。然而,目前对于多目标粗糙集属性约简问题的求解,依然以传统的优化算法为主。

本文基于既有研究,针对粗糙集属性权重问题设计了属性组合的知识依赖分辨率;在多粒度条件下,利用帕累托最优解思想将属性组合的知识依赖分辨率与属性组合的重要性程度进行有机结合,建立多目标属性约简模型;设计具有集群智能优化思想的复杂网络优化算法,引入基于个体的非支配解集以及基于“均值-方差”的遗传算子提升算法性能;在方法验证方面,依据约简结果,利用多层感知机对不同数据集内的对象进行分类,验证方法的有效性。

1 粗糙集与知识粒度

1.1 粗糙集

定义四元组 $DIS = \langle U, A, V, f \rangle$ 表示一个知识表达系统,令非空有限集合 U 表示论域,令非空有限集合 A 表示属性集合, $A = C \cup D$ 。其中: C 表示条

件属性集合, D 表示决策属性集合. 对于 $\forall a \in A$, 令 V_a 表示属性 a 的取值范围, 因此对于 $\forall a \in A$ 以及 $\forall x \in U$, 有 $f(x, a) \in V_a$, $f(x, a)$ 为信息函数, 表示 x 在属性 a 上的取值. 对于 $\forall a \in A$, 令 $\text{IND}(a)$ 表示基于属性的不可分辨关系, 则有 $\text{IND}(a) = \{(x, y) \in U^2, f(x, a) = f(y, a)\}$, 论域 U 内由属性 a 构成的等价类可以表示为 $[x]_a$ ($[x]_a = U/\text{IND}(a)$), 其中 $[x]_a = \{y | f(x, a) = f(y, a), \forall a \in A\}$.

对于知识表达系统 DIS , 令 $X \subseteq U$, R 表示论域 U 上的一个等价关系, 则粗糙集 X 在等价关系 R 上的上、下近似集可分别表示为^[1]

$$\underline{R}(X) = \bigcup \{x \in U | [x]_R \subseteq X\}, \quad (1)$$

$$\bar{R}(X) = \bigcup \{x \in U | [x]_R \cap X \neq \emptyset\}. \quad (2)$$

而粗糙集 X 在等价关系 R 上的边界域可以表示为^[1]

$$\text{Bn}_R = \bar{R}X - \underline{R}X, \quad (3)$$

其中下近似集 $\underline{R}(X)$ 也被定义为粗糙集 X 的正域, 表示为 $\text{POS}_R(X)$.

进一步, 不同的条件属性对于决策属性具有不同的重要性, 传统的粗糙集利用知识的依赖度来计算条件属性的权重, 决策属性对条件属性的依赖度可以表示为^[1,8]

$$\gamma_C(D) = \frac{\text{card}[\text{POS}_C(D)]}{\text{card}(U)}, \quad (4)$$

其中 $\text{card}(\cdot)$ 表示集合内元素基数, 条件属性 c 相对于决策属性 D 的重要性程度可表示为

$$\text{SIG}_c = \gamma_C(D) - \gamma_{C-c}(D), \quad (5)$$

$c \in C$. 在传统的粗糙集研究中, 属性约简则是通过从决策表中删除某些条件属性, 然后考察删除这些条件属性后决策表产生的信息处理能力变化来判断条件属性的重要程度. 这种信息处理能力的变化则由上述属性的依赖度以及属性的重要性程度来表示. 属性约简的目的, 是要找到一个约简后的条件属性组合 $\{c_i, c_j, c_k, \dots\}$, 使决策表在保持原有的分类能力不变的前提下, 提取出关键条件属性, 减小数据冗余, 即对于知识表达系统 $\text{DIS} = \langle U, A, V, f \rangle$, 有 $A = C \cup D, C' \subseteq C$. 如果对于任何 $c \in C'$, 有 $\gamma_{C'}(D) = \gamma_C(D)$ 并且 $\gamma_{C'-c}(D) \neq \gamma_{C'}(D)$, 则称 C' 为 C 的约简.

1.2 基于等价关系划分的知识依赖分辨率

近年来, 不断有研究发现, 基于式(4)中正区域集合基数与论域集合基数之比的条件属性重要性程度计算方式存在明显的局限性. 例如, 令 $U = \{a, b, c, d, e, f, g\}, X = \{a, b, d, e, g\}, R_1$ 与 R_2 分别表示两种

定义在 U 上的等价关系, 假设 R_1 与 R_2 所产生的等价划分分别为 $U/R_1 = \{\{a, b, d, e\}, \{c, f, g\}\}, U/R_2 = \{\{a, b\}, \{d, e\}, \{c, f, g\}\}$, X 关于 R_1 与 R_2 的正区域均为 $\{a, b, d, e\}$, 依据式(4)可得 R_1 与 R_2 相对应的属性重要性程度均为 $\text{card}[\text{POS}_{R_1}(D)]/\text{card}(U) = \text{card}[\text{POS}_{R_2}(D)]/\text{card}(U) = 4/7$. 由此可见, 虽然 R_1 与 R_2 相对应的属性重要性程度相同, 但 R_1 与 R_2 所产生的等价划分截然不同, 这种差异无疑会影响属性重要性程度(权重)的计算精度.

为了解决这一问题, 一些学者引入知识分辨率的概念对不同等价划分进行区分^[12].

定义1 令 R 为知识库 $\text{DIS} = \langle U, A, V, f \rangle$ 中的等价关系, $U/R = \{X_1, X_2, \dots, X_n\}$, 由等价关系 R 确定的知识粒度可以表示为

$$G(R) = \frac{\sum_i^n \text{card}(X_i)^2}{\text{card}(U)^2}. \quad (6)$$

定义2 令 $D(R)$ 表示由等价关系 R 确定的知识分辨率, 有

$$D(R) = 1 - G(R). \quad (7)$$

当等价关系 R 将论域内的每一个对象单独划分为一类时, $G(R)$ 达到最小值, 即 $G(R) = \frac{\text{card}(U)}{\text{card}(U)^2} = \frac{1}{\text{card}(U)}$; 当等价关系 R 将论域内的所有对象划分为一类时, $G(R)$ 达到最大值, 即 $G(R) = 1$, 因此 $\frac{1}{\text{card}(U)} \leq G(R) \leq 1$, 进而有 $0 \leq D(R) \leq 1 - \frac{1}{\text{card}(U)}$.

进一步, 结合式(4)、(6)和(7)可知, 条件属性特征的描述既要考虑属性的重要性程度, 又要考虑属性在论域内的知识分辨率. 因此, 本文将传统依赖度引入知识分辨率的构建, 使二者进行有机结合, 将考虑等价关系知识粒度情形下条件属性的知识依赖分辨率定义如下:

$$\gamma'_C(D) = \frac{\text{card}[\text{POS}_C(D)]}{\text{card}(U)} \left[1 - \frac{\sum_i^n \text{card}(\text{POS}_C^i(D))^2}{\text{card}(U)^2} \right]. \quad (8)$$

条件属性 c 相对于决策属性 D 的知识依赖分辨率可表示为

$$\text{RES}_c = \gamma'_C(D) - \gamma'_{C-c}(D). \quad (9)$$

由于 $0 \leq \text{card}[\text{POS}_C(D)]/\text{card}(U) \leq 1$, 则有 $0 \leq \gamma'_C(D) \leq 1 - 1/\text{card}(U)$. 可见, 上述知识依赖分辨率的引入可以在由式(4)确定的属性重要性程度(权重)的基础上进一步区分等价类划分的差异, 使属

性重要性程度的计算更加精确.

定理1 假设 E, F 为论域 U 上的等价关系, 若 $E \subseteq F$, 则 $\gamma'_E(D) \leq \gamma'_F(D)$.

证明 由于 $E \subseteq F$, 令 $E = F - \{f\}$, 对于粗糙集 X , 有 $\text{POS}_E(X) = \text{POS}_{F-f}(X)$ (条件属性越多, 对论域的分类越细, 正域内元素数量增加), 而依据正域的定义, 又有 $\text{POS}_{F-f}(X) \subseteq \text{POS}_F(X)$, 因此有 $\frac{\text{card}[\text{POS}_{F-\{f\}}(D)]}{\text{card}(U)} \leq \frac{\text{card}[\text{POS}_F(D)]}{\text{card}(U)}$; 进一步, 依据知识分辨度的定义, 有

$$\frac{\sum_i^n \text{card}[\text{POS}_{F-\{f\}}^i(D)]^2}{\text{card}(U)^2} \geq \frac{\sum_i^n \text{card}[\text{POS}_F^i(D)]^2}{\text{card}(U)^2},$$

即

$$1 - \frac{\sum_i^n \text{card}[\text{POS}_{F-\{f\}}^i(D)]^2}{\text{card}(U)^2} \leq 1 - \frac{\sum_i^n \text{card}[\text{POS}_F^i(D)]^2}{\text{card}(U)^2},$$

由式 (8), 有 $\gamma'_{F-\{f\}}(D) \leq \gamma'_F(D)$, 即 $\gamma'_E(D) \leq \gamma'_F(D)$. \square

1.3 基于属性序列的知识粒度

由 1.2 节可知, 知识分辨度的引入在一定程度上弥补了经典粗糙集的不足, 但在实际应用上仍存在诸多限制, 无论是传统的粗糙集还是利用知识分辨度对等价类划分的进一步精确化处理, 论域 U 上的对象 X_i 均是通过单个不可分辨关系进行粗糙表示, 即单粒度粗糙集模型, 而在实际应用中往往需要从多个角度处理问题, 相对于单粒度模型, 多粒度粗糙集模型被证明更容易发现隐含的知识. Qian 等^[8] 将基于单粒度的粗糙集拓展到由多个条件属性构成的多个等价类, 继而形成基于属性序列的多粒度粗糙集模型.

定义3 乐观多粒度粗糙集 (optimistic multi-granulation rough set). 对于知识表达系统 $\text{DIS} = \langle U, A, V, f \rangle$, $X \subseteq U$, 令 A_1, A_2, \dots, A_k 表示 A 的 k 个属性子集, 则在乐观粒度条件下, X 关于 A_1, A_2, \dots, A_k 的下、上近似集^[8] 分别表示为

$$\sum_{i=1}^k \underline{A}_i^o(X) = \{x \in U : [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \dots \vee [x]_{A_k} \subseteq X\}, \quad (10)$$

$$\sum_{i=1}^k \bar{A}_i^o(X) = \sim \sum_{i=1}^k \underline{A}_i^o(\sim X). \quad (11)$$

定义4 悲观多粒度粗糙集 (pessimistic multi-granulation rough set). 对于知识表达系统 $\text{DIS} =$

$\langle U, A, V, f \rangle$, $X \subseteq U$, 令 A_1, A_2, \dots, A_k 表示 A 的 k 个属性子集, 则在悲观粒度条件下, X 关于 A_1, A_2, \dots, A_k 的下、上近似集^[8] 分别表示为

$$\sum_{i=1}^k \underline{A}_i^p(X) = \{x \in U : [x]_{A_1} \subseteq X \wedge [x]_{A_2} \subseteq X \wedge \dots \wedge [x]_{A_k} \subseteq X\}, \quad (12)$$

$$\sum_{i=1}^k \bar{A}_i^p(X) = \sim \sum_{i=1}^k \underline{A}_i^p(\sim X). \quad (13)$$

其中: $[x]_{A_i} \subseteq X$ 表示由属性子集 A_i 导出的等价类, $\sim X$ 表示 X 的补集. 由式 (10)~(13) 不难发现, 由于乐观、悲观粗糙集中分别对由每个属性序列产生的等价划分统一进行了并集 (交集) 运算, 在计算决策属性对条件属性的依赖度时依然要沿用式 (4) 的方法^[9], 即计算正域与论域内元素基数之比.

综合分析 1.2 节与 1.3 节分别给出的知识分辨度及知识粒度的设定可见, 基于等价关系的知识分辨度对传统粗糙集的突破之处在于其在传统上近似集概念的基础上进一步明确量化了不同等价划分对每个属性权重的影响, 但是这一改进在不同属性序列构成的知识粒度中具有局限性. 这是由于, 就基于属性序列的知识粒度而言, 乐观 (悲观) 粒度粗糙集在计算属性序列的上下近似集时要对由每个属性序列产生的等价划分统一进行并集 (交集) 运算, 而并未考虑单个条件属性序列中基于等价划分的知识分辨度. 因此, 乐观 (悲观) 知识粒度对论域内等价划分的计算处理方式等同于进一步放大了 1.2 节开头所述的局限性, 从而会在属性约简过程中带来条件属性组合重要性程度计算结果的差异.

2 多目标属性约简优化模型

2.1 属性约简优化模型建立

属性约简问题可以转化为一个以不同关键条件属性组合方式为变量, 以条件属性的重要性程度为目标函数的 0-1 优化问题^[13], 即

$$\max \text{SIG}\{x_1, \dots, x_m\}, \quad x_i = 0, 1. \quad (14)$$

其中: x_i 为 0-1 变量, $x_i = 1$ 表示条件属性 c_i 为关键属性, 反之 $x_i = 0$ 表示条件属性 c_i 可以被约简. 结合式 (4)~(14) 可知, 当 m 值较大时, 寻找具有较高重要性的条件属性组合是一个典型的 NP-hard 问题. 而在基于属性序列知识粒度的量化范式下, 为了有效克服知识分辨度在多粒度粗糙集属性约简计算过程中产生的局限性, 本文将多目标优化的思想引入基于知识粒度的属性约简算法中, 则多目标属性约简问题可

以表示为

$$\begin{cases} f_1 = \max \text{RES}\{x_1, \dots, x_m\}, x_i = 0, 1; \\ f_2 = \max \frac{\text{SIG}\{x_1, \dots, x_m\}}{\sum_{i=1}^m x_i}, x_i = 0, 1. \\ \text{s.t. } \sum_{i=1}^m x_i < m. \end{cases} \quad (15)$$

其中: f_1 表示最大化知识依赖分辨率, f_2 表示最大化属性组合重要性程度的同时尽可能减少约简所含属性的数量. 进一步, 式(15)可以表示为

$$\begin{cases} \max F(X) = [f_1(X), f_2(X)]^T, \\ \text{s.t. } X \in S. \end{cases} \quad (16)$$

其中 S 表示所有条件属性组合的解空间.

2.2 属性约简帕累托最优解

由于两种知识粒度的构建方式存在一定差异, 上述多目标优化问题的各个子目标之间具有矛盾性, 本文通过引入帕累托最优解的思想对不同知识粒度之间的差异进行协调和折中处理. 帕累托最优解并非唯一, 而是由众多最优解组成的一个集合, 这个最优解集合, 被称为帕累托前沿. 帕累托最优思想有效地避免了对两种知识粒度构建方式的人为加权, 从而得到更加全面的非劣解. 依据多目标优化理论中对支配关系及帕累托前沿的描述^[15,19], 给出属性约简问题中的对应定义.

定义5 属性组合支配关系. 对于条件属性组合 $X_A, X_B \in S$, 如果有 $f_i(X_A) \geq f_i(X_B) (i = 1, 2, \dots, n)$ 且至少存在一个 $j \in \{1, 2, \dots, n\}$ 使得 $f_j(X_A) > f_j(X_B)$, 则称条件属性组合 X_A 支配条件属性组合 X_B , 记作 $X_A \prec X_B$.

定义6 非支配条件属性组合. 如果属性组合 $X_C \in S$ 并且 X_C 不被其他任何属性组合支配, 则称 X_C 为非支配条件属性组合.

定义7 属性约简的帕累托前沿. 由所有非支配条件属性组合计算得出的目标函数值集合在解空间中的表示称为属性约简的帕累托前沿(表示为PF), 即

$$\text{PF} = \{f_1(X_C), f_2(X_C)\}. \quad (17)$$

3 基于集群智能思想的多目标属性约简算法设计

3.1 针对问题结构进行的算法改进

式(15)所示的属性约简问题是一个典型的离散多目标优化问题, 相对于变量连续的多目标优化问题, 由于变量离散的优化问题在编码方式、算子设计等方面存在较大差异, 一些传统的启发式算法或遗传进化算法容易陷入局部最优. 近年来, 新的优化算法

不断被提出并在多种优化问题中取得了良好的效果, 例如人工蜂群算法、蛙跳算法、鱼群算法、元胞遗传算法、粒子群算法及其大量改进版本等. 这些新算法有一个共同点, 即在集群智能思想的基础上更加注重种群内个体间交互形式的改进, 从而避免算法早熟^[13-15,19]. 对于基于粗糙集属性约简的多目标0-1优化问题, 变量维度及解空间较大, 不容易实现性能良好的交叉算子且局部最优具有较大迷惑性. 因此, 本文以集群智能的思想作为框架, 通过设计具有良好的优秀个体信息交互功能的种群拓扑结构以及保持良好种群多样性的遗传算子来实现属性约简的精确化. 具体方案如下.

1) 建立具有复杂网络拓扑结构的种群空间.

为了便于算法迭代过程中优秀个体信息在种群中的传播, 本文在既有的一些基于空间拓扑结构的智能优化算法基础上, 引入复杂网络生成机制, 利用复杂网络的空间结构表征种群内个体之间的交互关系, 借助具有特定信息传播性能的网络结构提升优秀个体信息在种群内的传播效率.

2) 建立基于个体的非支配解存储解集.

传统的优化算法在迭代到一定步数后, 由于个体间的差异逐渐变小, 导致算法的收敛速度变慢, 收敛结果容易落入局部最优. 为了减小这一现象造成的影响, 本文建立基于个体的非支配解存储机制, 即在每一次迭代过程中, 除了更新全局最优解, 也同时为种群中的每一个个体存储其“邻居”(与该个体建立连接的个体)范围内的非支配解, 达到局部优秀信息与全局优秀信息的合理平衡.

3) “均值-方差”交叉算子.

2012年, Zhang等^[14]针对粒子群算法容易早熟的缺点, 设计了协调“个体最优”与“全局最优”之间关系的“均值-方差”交叉算子, 取得了较好的寻优效果, 因此, 为了进一步提升获取最优解的搜索效率和质量, 本文在交叉算子的设计中沿用这一方法.

基于上述改进, 将本文提出的基于集群智能思想的算法具体命名为多目标复杂网络优化算法(multi objective complex network algorithm, MOCNA).

3.2 算法具体步骤

step 1: 种群复杂网络拓扑结构建立. Watts等^[20]提出的复杂网络生长方式所产生的是环形复杂网络. 近年来, 一些研究表明, 具有网格结构的复杂网络在信息传播方面具有更好的性能^[21-22], 因此, 本文采用具有网格结构的种群复杂网络, 具体网络生成方式如下.

step 1.1: 将区域 $p \in [0, n], q \in [0, n]$ 设置为种群产生区域, p 与 q 均为整数, 每一个点 (p, q) 表示种群中的一个个体。

step 1.2: 每个个体 (p, q) 建立与邻居的连接, 本文采用 Moore 型邻居, 即每个个体首先与周围的 8 个个体建立连接, 分别是 $(p-1, q-1), (p-1, q), (p-1, q+1), (p, q-1), (p, q), (p, q+1), (p+1, q-1), (p+1, q), (p+1, q+1)$, 构成了一个规则网络。

step 1.3: 令 p_{cut} 表示种群网络的断点重连概率, $p_{\text{cut}} \in [0, 1]$, 为每一个个体 (p, q) 设置一个随机数 $p_{\text{rand}} \in [0, 1]$, 如果 $p_{\text{rand}} \leq p_{\text{cut}}$, 则随机断开 (p, q) 的一条连接, 然后随机建立 (p, q) 与其邻居范围外任意一个个体之间的连接, 可见, 随着 p_{cut} 的取值不断增大, 种群网络内的连接逐渐趋于随机(图1为本文复杂网络拓扑结构)。

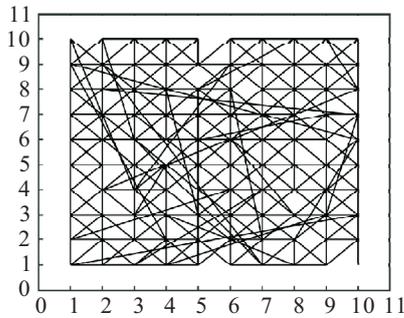


图1 种群复杂网络($p_{\text{cut}} = 0.5$)

step 2: 种群初始化. 网络中的每一个节点表示一个个体, 对于具有 m 个条件属性的约简问题, 个体 (p, q) 所代表的解可以表示为 $X_{p,q} = (X_{p,q,1}, X_{p,q,2}, \dots, X_{p,q,m}), F(X_{p,q}) = [f_1(X_{p,q}), f_2(X_{p,q})]$. 其中: $f_1(X_{p,q})$ 表示属性组合的知识依赖分辨度; $f_2(X_{p,q})$ 表示属性组合重要性程度; $X \in S, S_{p,q,i}$ 表示个体 (p, q) 选取属性 i 为关键属性, 本文采用实数编码, 令 $S_{p,q,i} \in [0, 1]$, 则有

$$S_{p,q,i} = \begin{cases} 0, & x_{p,q,i} \in [0, 0.5); \\ 1, & x_{p,q,i} \in [0.5, 1]. \end{cases} \quad (18)$$

step 3: 建立基于个体的非支配条件属性组合存储解集. 令 $\text{ND}_{p,q}$ 表示种群中的每个个体 (p, q) 的非支配条件属性组合解集, 依次将与个体 (p, q) 连接的每个“邻居”的目标函数值(用 $F(X_{i,j})$ 表示)加入 $\text{ND}_{p,q}$. 加入规则如下: 遍历 $\text{ND}_{p,q}$, 对于任意非支配条件属性组合 $\text{ND}_{p,q}$ 内的第 k 个解 $\text{ND}_{p,q}(k)$, 如果有 $F(X_{i,j}) \prec \text{ND}_{p,q}(k)$, 则在 $\text{ND}_{p,q}$ 内删去 $\text{ND}_{p,q}(k)$; 如果 $F(X_{i,j})$ 不被任何 $\text{ND}_{p,q}$ 内的解支配, 则将 $F(X_{i,j})$ 加入 $\text{ND}_{p,q}$.

step 4: 建立全局非支配条件属性组合存储解

集. 令 ND_g 表示全局非支配条件属性组合解集, 将种群中所有个体依据 $X_{p,q}$ 计算出的目标函数值 $F(X_{p,q})$ 依次加入 ND_g . 加入规则如下: 对于全局非支配条件属性组合 ND_g 内的解 $\text{ND}_g(k)$, 如果有 $F(X_{p,q}) \prec \text{ND}_g(k)$, 则在 ND_g 内删去 $\text{ND}_g(k)$; 如果 $F(X_{p,q})$ 不被任何 ND_g 内的解支配, 则将 $F(X_{p,q})$ 加入 ND_g .

step 5: 遗传操作. 在遗传操作过程中, 种群中的个体基于“拥挤距离”与“邻居”中的优秀个体以及全局范围内的优秀个体进行交互。

step 5.1: 计算 $\text{ND}_{p,q}$ 及 ND_g 中的拥挤距离. 拥挤距离用来衡量帕累托前沿上不同最优解之间的空间距离, 是衡量帕累托最优解分布均匀性的重要指标, 一般采用欧式距离计算^[16]. 以 $\text{ND}_{p,q}$ 为例(ND_g 同理)进行说明, 假设 $\text{ND}_{p,q}$ 存储了 K 组解, 首先按照目标函数值 $f_r(X_s)$ 对 $\text{ND}_{p,q}$ 中的个体进行降序排列, 得到 $f'_r(X_s)$, 其中 r 表示目标函数索引, 在本文两目标属性约简问题中 $r = 1, 2, s \in [1, K]$, 则 $\text{ND}_{p,q}$ 中目标函数 r 第 s 组解的拥挤距离为

$$\begin{cases} d_{s,r} = f'_r(X_{s+1}) - f'_r(X_{s-1}), & s \in (1, K); \\ d_{s,r} = \infty, & s = 1 \text{ or } s = K; \end{cases} \quad (19)$$

则第 s 组解的总体拥挤距离为 $d_s = \sum_{r=1}^2 d_{s,r}$.

step 5.2: 选择. 对于个体 (p, q) , 分别依据拥挤距离在 $\text{ND}_{p,q}$ 及 ND_g 内选择一个优秀个体, 选择方式采用原始遗传算法中常用的“轮盘赌”方法, 依然以 $\text{ND}_{p,q}$ 为例(ND_g 同理), 记 $\sum_{s=1}^K d_s$ 为 $\text{ND}_{p,q}$ 中的拥挤距离之和, 则依据轮盘赌规则, 第 s 组解所对应的个体被选中的概率为 $d_s / \sum d_s$, 在 $\text{ND}_{p,q}$ 中选出的个体记作 (p', q') , 在 ND_g 中选出的个体记作 (p'', q'') .

step 5.3: 交叉. 令交叉概率为 p_c , $X_{p,q}^*$ 表示邻居最优个体 (p', q') 与全局最优个体 (p'', q'') 进行交叉操作后的结果, 参考“均值-方差”交叉算子^[14], 则有

$$\begin{cases} \mu = \frac{X_{p',q'} + X_{p'',q''}}{2}, \\ \sigma = |X_{p',q'} - X_{p'',q''}|, \\ X_{p,q}^* = N(\mu, \sigma^2). \end{cases} \quad (20)$$

由式(20), 得到 $X_{p,q}^* = (x_{p,q,1}^*, x_{p,q,2}^*, \dots, x_{p,q,m}^*)$, 遍历 $X_{p,q}^*$ 中的 $X_{p,q,i}^*, X_{p,q,i}^*$ 先以概率 p_c 变换为 $x_{p',q',i}$, 后以概率 p_c 变换为 $x_{p'',q'',i}$.

step 5.4: 变异. 设置个体变异概率 p_m , 种群中依据概率 p_m 对交叉操作后的个体 (p, q) 进行变异操作, 对于参加变异操作的 $X_{p,q} = (s_{p,q,1}, s_{p,q,2}, \dots, s_{p,q,m})$ (通过式(18)转为0-1形式)中的每一个0-1变

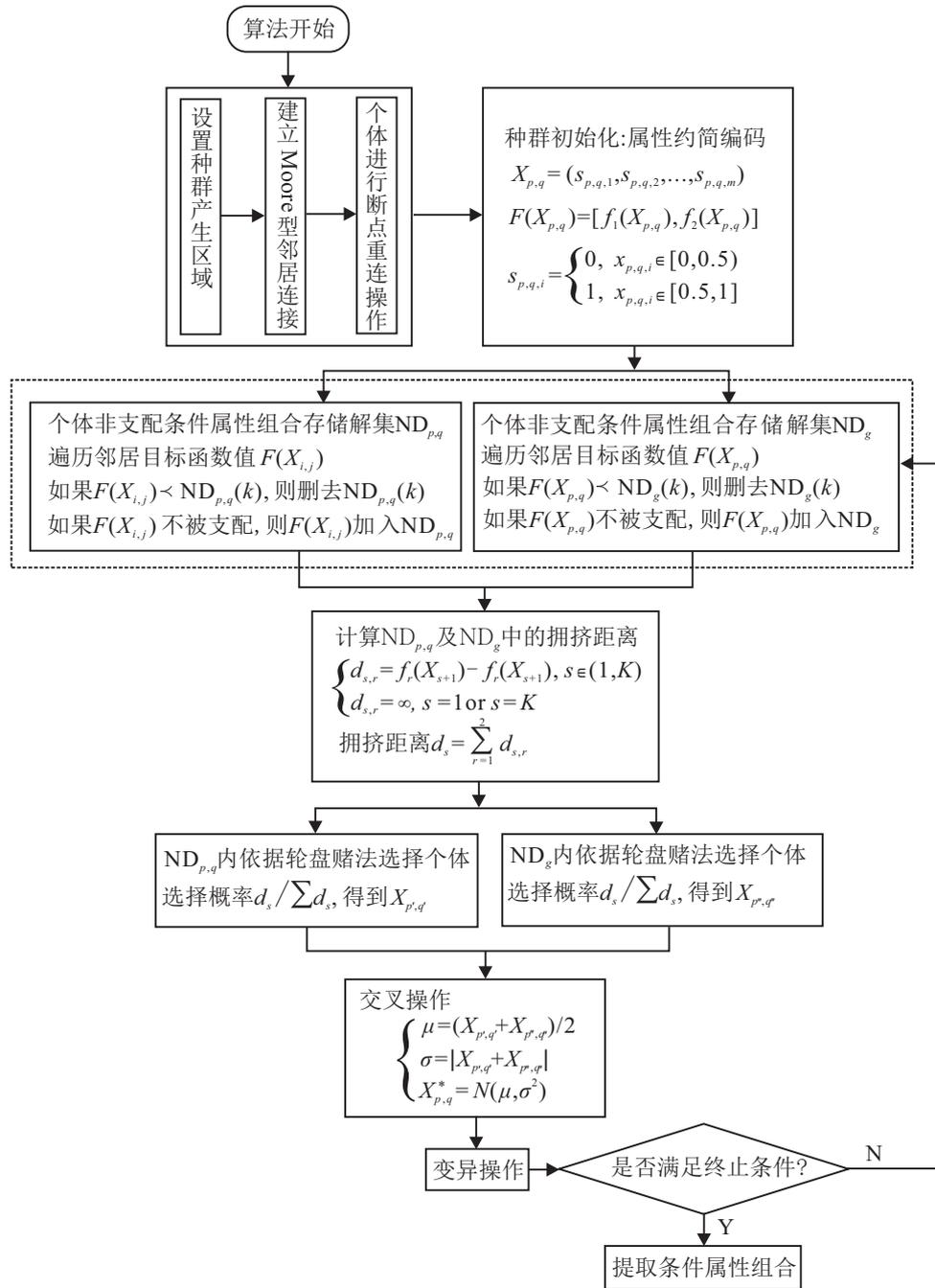


图2 算法流程

量,以概率 p_m 进行0变1(1变0)的操作.

step 6: 返回 step 3,直到算法满足终止条件(全局非支配解集不再发生变化).

算法整体流程如图2所示.

4 数据集测试与分析

4.1 测试算例

本文从 UCI 机器学习数据库中选取 5 个测试数据集进行属性约简计算,编程环境为 Windows10 操作系统,软件平台为 Matlab2018,各数据集内选取的属性个数与样本个数如表 1 所示,定义 3 和定义 4 中属性子集的构成设置为 $\{A_1, A_2, \dots, A_m\} = \{c_1,$

$c_2, \dots, c_m\}$. 此外,本文算法参数设置如下: $p_{cut} = 0.5, p_c = 0.7, p_m = 0.01$,种群数量 100,算法迭代 30 次(可得到稳定的全局非支配解集),同时引入已被证明具有良好性能的 MOcell^[16] 与 BBMOPSO^[14] 两种多目标优化算法进行对比,均设置与本文相同的种群数量及交叉变异概率. 选取 MOcell 与 BBMOPSO 两种算法的原因在于,MOcell 算法与 BBMOPSO 算法分别是空间结构智能算法与仿生智能算法领域中的经典代表,分别在种群空间拓扑结构以及交叉算子性能两个方面具有明显强于其他算法的优势,且分别对应于本文所提出改进的主要方面.

表1 测试数据集

数据集	选取属性数量(维度)	选取样本数量	类别数
wine	13	178	3
superconduct	22	300	3
wdbc	22	300	2
absenteeism	18	300	3
dermatology	35	366	3

4.2 性能指标

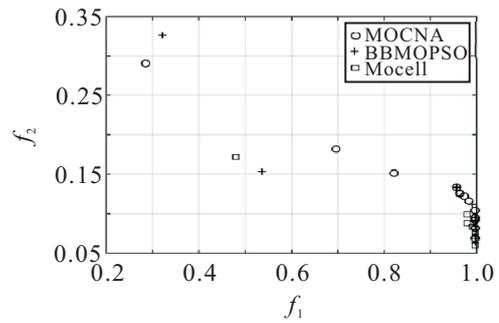
Hypervolume(超体积, 简称为HV)用来计算算法所获得的帕累托最优解集在目标域所覆盖的体积. 在多目标算法性能测试领域, 并非所有的多目标优化问题都能够获取绝对意义上的最优帕累托前沿, 而HV的计算不需要理论上最优帕累托前沿信息; 因此, HV被广泛用于衡量多目标优化算法解集的质量, 表达式如下:

$$HV = \text{volume}(\bigcup_{i=1}^Q v_i). \quad (21)$$

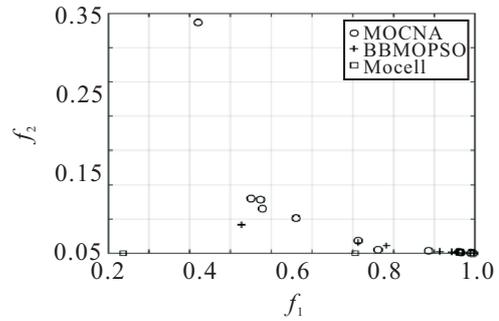
其中: Q 表示帕累托前沿上最优解的个数, v_i 表示由前沿上第*i*个点与参照点形成的超体积. 可见, HV越大, 表明算法所获得的解能够越广泛地分布在帕累托前沿上, 并且解的质量越好.

4.3 约简结果性能分析

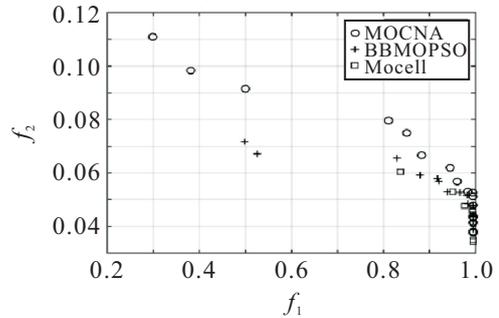
图3给出了不同方法在不同数据集条件下得出的具有代表性的属性约简帕累托前沿, 其中“乐观粒度”与“悲观粒度”分别对应于由定义3和定义4中的下近似集得到的属性组合约减. 由图3可见, 在针对5个测试数据集进行的约简中, 相对于BBMOPSO与MOcell算法, 本文的MOCNA算法获取了分布更加广泛的帕累托最优解. 进一步, 表2给出了各算法在不同数据集的约简中所得帕累托最优解的平均数量对比. 由表2可知, 在由5个数据集构成的10种属性约简情境下, 本文采用的MOCNA算法在8种属性约简情境下获取了最多数量的属性约简帕累托最优解, 具有较强的最优解搜索能力. 表3给出了各算法在不同数据集的约简中所得帕累托前沿的HV值, 结合表3中各算法获得的超体积计算结果可知, 无论是在乐观粒度下还是悲观粒度下, 本文的MOCNA算法在8种属性约简情境下获得了最高的HV值, 即所获得的属性约简帕累托前沿更加精确. 上述结果说明, 本文提出的方法在拓展种群多样性和提升最优解全局搜索能力方面具有更好的表现, 避免了属性约简结果落入局部最优. 综上可见, 本文提出的方法具有较好的属性约简性能.



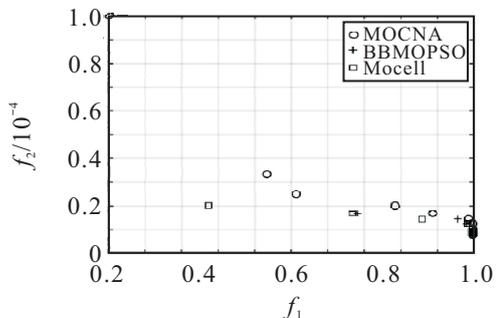
(a) wdbc 乐观粒度



(b) wdbc 悲观粒度



(c) dermatology 乐观粒度



(d) dermatology 悲观粒度

图3 属性约简帕累托前沿对比

表2 不同算法所获取帕累托最优解平均数量对比

测试数据集	属性序列知识粒度	MOCNA	BBMOPSO	MOcell
wine	乐观粒度	11	10	10
	悲观粒度	6	5	5
superconduct	乐观粒度	15	13	11
	悲观粒度	10	13	9
wdbc	乐观粒度	12	8	9
	悲观粒度	11	9	5
absenteeism	乐观粒度	14	14	13
	悲观粒度	7	8	7
dermatology	乐观粒度	15	14	9
	悲观粒度	10	7	7

表3 不同算法之间HV值对比

测试数据集	属性序列知识粒度	MOCNA	BBMOPSO	MOcell
wine	乐观粒度	2.033 0	1.649 2	1.710 8
	悲观粒度	0.145 8	0.145 8	0.145 8
superconduct	乐观粒度	0.290 3	0.269 5	0.214 8
	悲观粒度	0.010 1	0.010 2	0.005 6
wdbc	乐观粒度	1.248 1	0.828 7	0.827 6
	悲观粒度	0.265 0	0.045 3	1.0171e-04
absenteeism	乐观粒度	0.100 4	0.067 7	0.095 6
	悲观粒度	9.800 4e-05	9.805 3e-05	9.674 3e-05
dermatology	乐观粒度	0.936 3	0.796 4	0.490 6
	悲观粒度	1.192 9e-04	8.335 1e-05	5.942 0e-05

5 约简结果对比分析

为了进一步验证本文所提出约简方法的有效性,基于本文算法约简后的条件属性组合,采用基于多层感知机(multi-layer perceptron)的神经网络对上述5个数据集中的对象进行训练,利用训练好的神经网络进行分类测试(参数设置见表4),并与未考虑知识依赖分辨率时的约简结果(1.3节属性序列知识粒度条件下基于属性组合重要性程度的单目标约简)进行分类精度对比(见表5),分类精度统计采用十折交叉验证的方式进行.其中“单目标约简”表示仅考虑属性序列知识粒度的单目标约简结果,“约简1”到“约简5”表示将4.3节得到的属性约简帕累托前沿自右向左选取的5个代表性约简结果,可见,就5种约简结果在帕累托前沿上的位置而言,从“约简1”(右侧)到“约简5”(左侧),属性约简的结果发生从“知识依赖分辨率高(目标函数 f_1 值较大)”到“属性组合重要

表4 深层感知机参数设置

参数	取值
输入层神经元数量	约简后条件属性数量
隐含层数量	3
隐含层神经元数量	> 约简后条件属性数量
训练次数	500
learning rate	0.5
训练集数量	数据集的75%
测试集数量	数据集的25%

表5 不同约简结果对应的平均分类精度 单位: %

分类精度	单目标约简	约简1	约简2	约简3	约简4	约简5
wine 乐观粒度	41	81	73	72	36	55
wine 悲观粒度	54	81	35	30	63	27
superconduct 乐观粒度	65	77	69	79	63	66
superconduct 悲观粒度	70	91	86	93	81	92
wdbc 乐观粒度	79	83	79	49	49	45
wdbc 悲观粒度	45	81	65	80	57	78
absenteeism 乐观粒度	57	60	79	67	57	58
absenteeism 悲观粒度	55	67	59	76	64	55
dermatology 乐观粒度	86	94	94	93	91	86
dermatology 悲观粒度	60	94	90	96	95	83

性程度高(目标函数 f_2 值较大)”的转变.

由表5不难发现:1)无论是在乐观粒度条件下还是在悲观粒度条件下,相对于单目标情形下仅依据属性组合重要性程度进行的属性约简,引入知识依赖分辨率的多目标属性约简算法得到的条件属性组合明显可以得到更高的分类精度.在5个测试数据集的10个约简场景中,本文的多目标属性约简方法均得出了分类精度高于单目标约简的条件属性组合,其中在superconduct悲观粒度、wdbc悲观粒度、absenteeism乐观粒度、absenteeism悲观粒度、dermatology乐观粒度以及dermatology悲观粒度条件下,多目标属性约简方法所得到的条件属性组合,平均分类精度均不低于单目标属性约简.2)从“约简5”到“约简1”,其所代表条件属性组合的知识依赖分辨率不断提升,知识依赖分辨率的提升在一定程度上增加了条件属性组合的分类精度,就多目标属性约简的结果而言,在5个测试数据集的10个约简场景中,知识依赖分辨率最低的约简(“约简5”)均对应于较低的分类精度,其中在5个约简场景中分类精度最低;而知识依赖分辨率最高的约简(“约简1”~“约简3”)在10个约简场景中分类精度达到最高;可见,选取知识依赖分辨率较高的属性组合是提升数据分类精度的有效手段.上述结果也进一步说明了在基于属性序列知识粒度的粗糙集属性约简中引入多目标约简方法的必要性.

6 结论

本文针对多粒度粗糙集在属性约简过程中由运算方式产生的局限性,引入知识分辨率的概念,并将多粒度粗糙集属性约简问题转化为离散多目标优化问题,通过设计新的种群拓扑结构以及遗传算子形成了全新的属性约简优化算法,与两种常用优化算法的对比表明本文所提出的方法能够得到更加全面的约简结果,所得到的约简解集质量更优,具有较好的可行性与实际意义.进一步,相对于仅考虑属性重要性程度的单目标约简,本文的约简方法得到的属性集合具有更好的分类效果.本文研究的结果一方面可以帮助管理者增加属性约简的维度,实现不同数据维度下的科学决策(乐观、悲观),同时也可以帮助其在属性约简过程中依据优化得到的帕累托前沿定量地实现不同约简需求之间的取舍与协调,即通过帕累托前沿给出属性组合的知识依赖分辨率和重要性程度之间的权衡方案.但是,本文并未针对属性子集的构成层次作进一步讨论,在未来的研究中,可以将多粒度粗糙集的约简分为“子集层”和“属性层”进行双层

约简,在两个层次的属性约简中均引入知识依赖分辨度进行研究。

参考文献(References)

- [1] Pawlak Z. Rough sets[J]. International Journal of Parallel Programming, 1982, 11(5): 341-356.
- [2] Qian Y H, Liang X Y, Wang Q, et al. Local rough set: A solution to rough data analysis in big data[J]. International Journal of Approximate Reasoning, 2018, 97(6): 38-63.
- [3] 鲍新中, 张建斌, 刘澄. 基于粗糙集条件信息熵的权重确定方法[J]. 中国管理科学, 2009, 17(3): 131-135.
(Bao X Z, Zhang J B, Liu C. A new method of ascertaining attribute weight based on rough sets conditional information entropy[J]. Chinese Journal of Management Science, 2009, 17(3): 131-135.)
- [4] Zhang X, Mei C, Chen D, et al. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy[J]. Pattern Recognition, 2016, 56(1): 1-15.
- [5] 刘勇, 熊晓旋, 刘健. 优势直觉模糊粗糙集决策方法及其应用[J]. 系统管理学报, 2018, 27(3): 538-545.
(Liu Y, Xiong X X, Liu J. Decision making on dominant intuitionistic fuzzy rough set and its application[J]. Journal of Systems & Management, 2018, 27(3): 538-545.)
- [6] 滕书华, 鲁敏, 杨阿锋, 等. 基于一般二元关系的粗糙集加权不确定性度量[J]. 计算机学报, 2014, 37(3): 649-665.
(Teng S H, Lu M, Yang A F, et al. A Weighted uncertainty measure of rough sets based on general binary relation[J]. Chinese Journal of Computers, 2014, 37(3): 649-665.)
- [7] 施振铨, 陈世平. 基于粗糙集和知识粒度的特征权重确定方法[J]. 科技管理研究, 2018, 38(12): 248-253.
(Shi Z Q, Chen S P. A new method of knowledge characteristics weighting based on rough set and knowledge granulation[J]. Science and Technology Management Research, 2018, 38(12): 248-253.)
- [8] Qian Y H, Liang J, Yao Y, et al. MGRS: A multi-granulation rough set[J]. Information Sciences, 2010, 180(6): 949-970.
- [9] 马福民, 陈静雯, 张腾飞. 基于双重粒化准则的邻域多粒度粗糙集快速约简算法[J]. 控制与决策, 2017, 32(6): 1121-1127.
(Ma F M, Chen J W, Zhang T F. Quick attribute reduction algorithm for neighborhood multi-granulation rough set based on double granulate criterion[J]. Control and Decision, 2017, 32(6): 1121-1127.)
- [10] 万志超, 宋杰, 沈永良. 可变直觉模糊多粒度粗糙集模型及其近似分布约简算法[J]. 计算机应用, 2018, 38(2): 390-398.
(Wan Z C, Song J, Shen Y L. Variable intuitionistic fuzzy multi-granulation rough set model and its approximate distribution reduction algorithms[J]. Journal of Computer Applications, 2018, 38(2): 390-398.)
- [11] 任晓霞, 薛凡. 基于模糊邻域粗糙集的启发式属性约简算法[J]. 计算机工程与应用, 2018, 54(9): 47-53.
(Ren X X, Xue F. Heuristic attribute reduction algorithm based on fuzzy neighborhood rough set[J]. Computer Engineering and Applications, 2018, 54(9): 47-53.)
- [12] 梁吉业, 钱宇华. 信息系统中的信息粒与熵理论[J]. 中国科学, 2008, 38 (12): 2048-2065.
(Liang J Y, Qian Y H. Information granule and entropy theory in information system[J]. Science China, 2008, 38(12): 2048-2065.)
- [13] Luan X Y, Li Z P, Liu T Z. A novel attribute reduction algorithm based on rough set and improved artificial fish swarm algorithm[J]. Neurocomputing, 2016, 174(1): 522-529.
- [14] Zhang Y, Gong D W, Ding Z. A bare-bones multi-objective particle swarm optimization algorithm for environmental/economic dispatch[J]. Information Sciences, 2012, 192(6): 213-227.
- [15] Luo J, Liu Q, Yang Y, et al. An artificial bee colony algorithm for multi-objective optimization[J]. Applied Soft Computing, 2017, 50(1): 235-251.
- [16] Nebro A J, Durillo J J, Luna F. MOCELL: A Cellular genetic algorithm for multiobjective optimization[J]. International Journal of Intelligent Systems, 2009, 24(7): 726-746.
- [17] Wang L, Li W, Jia X, et al. A multi-objective attribute reduction method in decision-theoretic rough set model[C]. International Conference on Knowledge Science, and Management. Cham, 2017: 117-128.
- [18] Cervante L, Xue B, Shang L, et al. A multi-objective feature selection approach based on binary PSO and rough set theory[C]. Evolutionary Computation in Combinatorial Optimization. Berlin: Springer, 2013, 7832: 25-36.
- [19] Luo J P, Yang Y, Liu Q Q, et al. A new hybrid memetic multi-objective optimization algorithm for multi-objective optimization[J]. Information Sciences, 2018, 448(6): 164-186.
- [20] Watts D J, Strogatz S H. Collective dynamics of small world networks[J]. Nature, 1998, 393(6684): 440-442.
- [21] Schimit P H T, Santos B O, Soares C A, Evolution of cooperation in Axelrod tournament using cellular automata[J]. Physica A, 2015, 437(11): 204-217.
- [22] Ruan Y H, Li A W. A new small-world network created by cellular automata[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 456(8): 106-111.

作者简介

李雪岩(1987—), 男, 讲师, 博士, 从事管理科学与复杂系统决策理论、数据分析的研究, E-mail: gongye1632006@163.com;

李学伟(1962—), 男, 教授, 博士生导师, 从事管理科学与复杂系统决策理论、大数据关键技术等研究, E-mail: Xuwei@buu.edu.cn;

蒋君(1991—), 男, 博士生, 从事运输经济、经济物理的研究, E-mail: 982104919@qq.com.