

DOI:10.11992/tis. 201705004

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170703.1854.018.html>

基于粗糙集相对分类信息熵和 粒子群优化的特征选择方法

翟俊海^{1,2}, 刘博³, 张素芳⁴

(1. 河北大学 河北省机器学习与计算智能重点实验室, 河北 保定 071002; 2. 浙江师范大学 数理与信息工程学院, 浙江 金华 321004; 3. 河北大学 计算机科学与技术学院, 河北 保定 071002; 4. 中国气象局 气象干部培训学院河北分院, 河北 保定 071000)

摘要: 特征选择是指从初始特征全集中, 依据既定规则筛选出特征子集的过程, 是数据挖掘的重要预处理步骤。通过剔除冗余属性, 以达到降低算法复杂度和提高算法性能的目的。针对离散值特征选择问题, 提出了一种将粗糙集相对分类信息熵和粒子群算法相结合的特征选择方法, 依托粒子群算法, 以相对分类信息熵作为适应度函数, 并与其他基于进化算法的特征选择方法进行了实验比较, 实验结果表明本文提出的方法具有一定的优势。

关键词: 数据挖掘; 特征选择; 数据预处理; 粗糙集; 决策表; 粒子群算法; 信息熵; 适应度函数

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2017)03-0397-08

中文引用格式: 翟俊海, 刘博, 张素芳. 基于粗糙集相对分类信息熵和粒子群优化的特征选择方法[J]. 智能系统学报, 2017, 12(3): 397-404.

英文引用格式: ZHAI Junhai, LIU Bo, ZHANG Sufang. A feature selection approach based on rough set relative classification information entropy and particle swarm optimization[J]. CAAI transactions on intelligent systems, 2017, 12(3): 397-404.

A feature selection approach based on rough set relative classification information entropy and particle swarm optimization

ZHAI Junhai^{1,2}, LIU Bo³, ZHANG Sufang⁴

(1. Key Lab of Machine Learning and Computational Intelligence, Hebei University, Baoding 071002, China; 2. College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China; 3. College of Computer Science and Technology, Hebei University, Baoding 071002, China; 4. Hebei Branch of Meteorological Cadres Training Institute, China Meteorological Administration, Baoding 071000, China)

Abstract: Feature selection, an important step in data mining, is a process that selects a subset from an original feature set based on some criteria. Its purpose is to reduce the computational complexity of the learning algorithm and to improve the performance of data mining by removing irrelevant and redundant features. To deal with the problem of discrete values, a feature selection approach was proposed in this paper. It uses a particle swarm optimization algorithm to search the optimal feature subset. Further, it employs relative classification information entropy as a fitness function to measure the significance of the feature subset. Then, the proposed approach was compared with other evolutionary algorithm-based methods of feature selection. The experimental results confirm that the proposed approach outperforms genetic algorithm-based methods.

Keywords: data mining; feature selection; data preprocessing; rough set; decision table; particle swarm optimization; information entropy; fitness function

特征选择^[1-2]是数据挖掘的重要预处理步骤, 通过删除不必要特征, 以达到降低算法复杂度, 提高算法的性能。特征子集的评价至关重要, 特征子

集的评价准则或启发式确定后, 特征选择问题从某种意义上来说就是在特征空间中搜索最优特征子集问题。研究人员提出了许多特征选择算法, 这些算法有的是从设计特征子集搜索的方法考虑的, 有的是从设计评价特征子集重要性的度量考虑的。

特征选择方法主要有筛选法(filter method)、封

收稿日期: 2017-05-07. 网络出版日期: 2017-07-03.

基金项目: 国家自然科学基金项目(71371063); 河北省自然科学基金项目(F2017201026); 浙江省计算机科学与技术重中之重学科(浙江师范大学)资助项目.

通信作者: 翟俊海. E-mail: mczjh@126.com.

装法(wrapper method)和嵌入法(embedded method)3种分类。筛选法独立于学习算法,它依据启发式(如依赖度、互信息、不一致性等)搜索整体最优解。因为这类特征选择算法独立于数据挖掘算法,所以通用性好,适应各种数据挖掘任务,研究也最为广泛。实际上,粗糙集领域中的属性约简方法就是这类方法的一种典型代表^[3-4]。Swiniarski等^[5]针对模式识别问题,系统地研究了特征选择的粗糙集方法;Jensen等^[6]提出了基于模糊粗糙集的方法,但是它在某些情况下不收敛;Bhatt等^[7]针对文献[6]中的问题,深入研究模糊粗糙集特征选择算法的收敛性,得出了非常有价值的结论;Jensen等^[8]研究了模糊粗糙集特征选择算法的可扩展性问题,并提出了相应的方法;钱宇华等^[9]提出了基于粗糙集正域逼近的属性约简(或特征选择)加速方法,该方法具有很高的计算效率,受到了广泛关注;胡清华等^[10]在邻域粗糙集框架下研究了异构特征选择问题,提出的特征选择算法可以处理实数值异构特征选择问题。

除了基于粗糙集知识的算法,其他特征选择问题代表性的工作还包括基于一致性的方法^[2,11-12]、基于互信息的方法^[13-15]及基于依赖度的方法等^[16-17]。

封装法以算法性能为标准衡量特征子集,由Kohavi等于1997年提出^[18]。多年来,一些专家学者沿着封装法的思路做出了大量研究;Sindhwani等将多层感知器和SVM相结合,提出了基于最大输出信息的特征选择算法^[19];Yang等提出了一种随机扰动特征选择方法,具有较好表示能力^[20]。封装法过于依赖学习算法,对于全部筛选出的特征子集,必须重复分类器训练的过程,虽然最终得到的特征子集比较理想,但复杂度太高,代价过大。

在嵌入法中,特征选择与分类器的构造同步完成。实际上,著名的ID3算法^[21]和CART算法^[22]都可以看作嵌入式特征选择方法。基于神经网络剪枝思想,Setiono等^[23]提出了一种特征选择算法。该算法通过计算删除某个特征(剪枝)对神经网络性能的影响来选择特征。在此基础上,Shen等提出了基于SVM的敏感度特征选择算法^[24]。而后,Perkins等提出了Grafting嵌入式特征选择算法^[25]。

粒子群算法由Kennedy提出^[26],逐渐被专家学者广泛接受和应用。但是直到2001年关于粒子群

算法的第一部专著Swarm Intelligence^[27]出版后,粒子群算法才得到广泛应用。为解决PSO算法在连续空间中的局限性,Kennedy等^[28]提出了离散二进制PSO算法。在此基础上,Chuang等^[29]提出了一种改进的二进制粒子群算法(IBPSO),有效简化了特征选择,并在医疗诊断基因表达领域取得了较好的效果;Chuang等^[30]基于Catfish机制,提出了一种二进制粒子群特征选择算法,可避免过早收敛的问题;Wang等^[31]提出了一种基于粗糙集的粒子群算法,具有较强的搜索能力能更快地寻找到最优解;Cervante等^[32]提出了基于信息熵的BPSO方法,测试结果非常理想。Liu等^[33]提出了一种集成邻域信息和粒子群算法的特征选择方法,结合粒子加权提高了算法的稳定性;Fong等^[34]面向高维和数据流格式的大数据,提出了一种基于粒子群算法的特征选择方法,提高测试精度的同时降低了处理时间^[34]。

本文选择相对分类信息熵作为适应度函数,用粒子群算法搜索最优特征子集,并与其他进化算法进行了实验比较,实验结果表明本文所提的方法更为有效。

1 基础知识

1.1 粒子群算法

粒子群算法是一种群体搜索行为。群体由若干个粒子 p_i 组成, p_i 包含位置信息 x_i^k 和速度信息 v_i^k 。位置信息 x_i^k 表示 p_i 在第 k 次迭代时其位置信息,速度信息 v_i^k 表示 p_i 在第 k 次迭代时移动的速度和方向。从一组初始粒子(初始种群)出发,通过一次次迭代,不断更新速度和位置信息,每个粒子不断向gbest和pbest靠近,最终找到整体最优解。方法过程如图1所示,具体步骤见第2节。

粒子群算法应用的关键是问题编码和粒子适应度评价标准。在问题编码上,显然二进制编码是最为合适的。在粒子适应度评价标准上,一个合适的适应度函数对特征子集的选择影响很大,因此适应度函数的设计是尤为关键的。

1.2 粗糙集

定义1 设 $DT=(U, A \cup D, V, f)$ 为知识涉及的决策表。其中, $U=\{x_1, x_2, \dots, x_n\}$ 为论域; $A=\{a_1, a_2, \dots, a_m\}$ 是描述对象的属性集合; $D=\{d\}$,即目标概念唯一; V 为属性值域; $f:U \times A \rightarrow V$ 为信息函数。

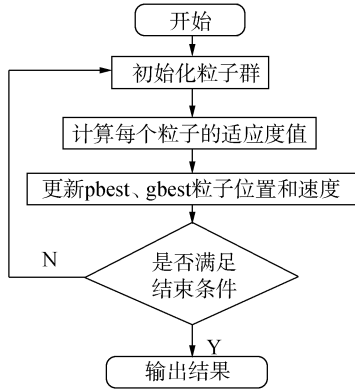


图1 PSO 算法流程图

Fig.1 The flowchart of particle swarm optimization algorithm

定义2 对于任意条件属性子集 $B \subseteq A$, B 的等价关系为

$$\text{IND}(B) = \{(x, y) \in U \times U \mid \forall b \in B, b(x) = b(y)\} \quad (1)$$

式中 $b(x)$ 和 $b(y)$ 分别是对象 x 和 y 在属性 b 上的属性值。

定义3 令 $X \subseteq U, B \subseteq A$, X 关于 B 的下近似为

$$\underline{BX} = \{x \mid [x]_B \subseteq X\} \quad (2)$$

定义4 令 $X \subseteq U, B \subseteq A$, X 关于 B 的上近似为

$$\overline{BX} = \{x \mid [x]_B \cap X \neq \emptyset\} \quad (3)$$

定义5 令 $X \subseteq U, B \subseteq A$, X 关于 B 的边界域为

$$\text{BN}(X) = \overline{BX} - \underline{BX} \quad (4)$$

2 基于粗糙集相对分类信息熵的粒子群特征选择方法

首先定义适用于特征选择问题的适应度函数。

对于形如 $\text{DT} = (U, A \cup D, V, f)$ 的决策表, $B \subseteq A$, B 和 D 对 U 的划分表示为

$$U/B = X_1 + X_2 + \cdots + X_m$$

$$U/D = Y_1 + Y_2 + \cdots + Y_k$$

定义6 划分 U/B 相对于 U/D 的概率分布为

$$p_{ij} = \frac{|X_i \cap Y_j|}{|X_i|} \quad (5)$$

定义7 划分 U/B 相对于 U/D 的相对分类信息熵为

$$H(U/B) = - \sum_{i=1}^m \sum_{j=1}^k \frac{|X_i|}{|U|} p_{ij} \log_2 p_{ij} \quad (6)$$

相对分类信息熵反映的不一致程度与其值成正比,因此值越小效果越好。本文所选择的适应度函数,即(6)式定义的相对分类信息熵,可通过定理1的证明来验证其合理性^[35]。

定理1 给定决策表 $\text{DT} = (U, A \cup D, V, f)$, 假设 $A_1, A_2 \subseteq A$ 且 $A_1 \subseteq A_2$, 则 $H(U/A_2) \leq H(U/A_1)$ 。

定理1保证了本文提出的方法的收敛性,即能够以很高概率找到整体最优解。

算法1 基于粒子群算法的特征选择算法。

输入 离散值决策表 $\text{DT} = (U, A \cup D, V, f)$, 群体大小 N 、惯性系数 w 、常量系数 c_1 和 c_2 、数据集维数 D 。

输出 $A' \subseteq A$ 。

- 1) 随机初始化大小为 N 的种群;
- 2) repeat;
- 3) for($i = 1; i \leq N; i = i + 1$);
- 4) 用式(6)计算种群中每一个个体的适应度值;
- 5) 更新历史最佳位置 best;
- 6) end for;
- 7) 更新全局最佳位置 gbest;
- 8) for($i = 1; i \leq N; i = i + 1$);
- 9) for($d = 1; d \leq D; d = d + 1$);
- 10) 更新速度信息 v_{id} ;
- 11) 更新位置信息 x_{id} ;
- 12) end for;
- 13) end for;
- 14) until(满足终止条件时);
- 15) 输出 A' 。

本文选择适用于离散值特征选择的二进制粒子群算法,问题的解可用 d 维 0-1 向量表示,其中 d 是属性个数。向量的第 i 个分量为 0 或 1 表示属性是否被选择。特征选择的过程分为以下几步。

1) 随机产生 N_0 个 0-1 向量(粒子)作为初始粒子群,其中 N_0 是种群的大小。

2) 利用式(6)计算所有粒子的适应度值。

3) 更新粒子历史最优点 pbest:将所有粒子的适应度值和其历史最佳点 pbest 做比对,择优更新 pbest。

4) 更新全局最优点 gbest:将所有粒子的适应度值和全局最优点 gbest 做比对,择优更新 gbest。

5) 根据上一步更新得到的 pbest 和 gbest,更新粒子的速度和位置。其中速度是根据上一步得到的 pbest 和 gbest 按照公式(7)进行更新:

$$v_i^{k+1} = wv_i^k + c_1 \text{rand}(\text{pbest} - x_i^k) + c_2 \text{rand}(\text{gbest} - x_i^k) \quad (7)$$

这里需要注意的是, p_i 的位置信息 x_i^k 中的每个 x_{id}^k 只能取 0 或 1,相应的速度信息也发生了改变,用

Sigmoid() 函数将 v_{id}^k 映射到 $(0,1)$ 区间上, 决定 x_{id}^k 在下一代位置分量为 1 的概率, 如公式(8):

$$x_{ij} = \begin{cases} 1, & \text{rand} < \text{Sigmoid}(v_{ij}) \\ 0, & \text{其他} \end{cases} \quad (8)$$

$$\text{Sigmoid}(v_{ij}) = \frac{1}{1 + e^{-v_{ij}}}$$

6) 若迭代次数已达到设定值, 终止并输出当前 gbest 对应的解; 否则, 令 $k=k+1$ 转入 2)。

粒子运动的步骤即全局最优解的求解过程, 通过不断计算每个粒子的适应度值, 粒子依靠自我认知能力和社会认知能力来搜索最优特征子集, 逐渐向全局最佳位置靠拢, 并最终得到理想结果。

3 实验结果及统计分析

在实验过程中, 选择规模大小和属性个数各异的 8 个离散 UCI 数据集进行实验。实验采用 Win7 操作系统 PC 机 (Intel Core i3 处理器, 内存大小 4 GB)。70% 的样例用作训练, 30% 的样例用作测试。实验研究发现, 当惯性系数取值在 $0.9 \sim 1.2$ 时, 粒子群算法在全局搜索和收敛能力上表现均较好^[36]。惯性系数和加速系数在常用范围内多次实验, 确定最终取值。惯性系数 w 取值为 1。加速系数 c_1 和 c_2 取 1.426 94, 种群数目统一设定为 20, 终止代数设定为 50。

首先选择与本文方法背景知识相关、时间复杂度相仿的方法进行比较, 即选用粗糙集与粒子群算法相结合的 RSPSO 算法^[31]和改进的二进制粒子群算法^[29]。在表 1 中给出本文方法与文献[31]和[29]两种方法实验对比结果, 主要对比两个方面的差异, 其中 #Fs 代表被选择的特征个数, Ac 代表测试精度。更直观的结果见图 2~9 所示。

表 1 实验结果比较

Table 1 The comparison of experimental results

| 数据集 | 本文方法 | | 文献[31]方法 | | 文献[29]方法 | |
|-------------|------|---------|----------|---------|----------|---------|
| | #Fs | Ac | #Fs | Ac | #Fs | Ac |
| Kr-vs-Kp | 12 | 0.973 3 | 15 | 0.954 5 | 13 | 0.979 2 |
| Voting | 10 | 0.655 8 | 10 | 0.608 5 | 11 | 0.651 2 |
| Mushroom | 8 | 0.882 9 | 10 | 0.813 6 | 8 | 0.876 2 |
| Nursery | 6 | 0.919 7 | 7 | 0.899 1 | 6 | 0.907 6 |
| Car | 5 | 0.877 3 | 5 | 0.857 8 | 5 | 0.882 1 |
| Soybean | 16 | 0.998 7 | 17 | 0.990 7 | 18 | 0.997 6 |
| Tic-Tac-Toe | 7 | 0.992 9 | 6 | 0.979 9 | 7 | 0.987 8 |
| SPECT | 12 | 0.984 2 | 11 | 0.886 8 | 13 | 0.980 5 |

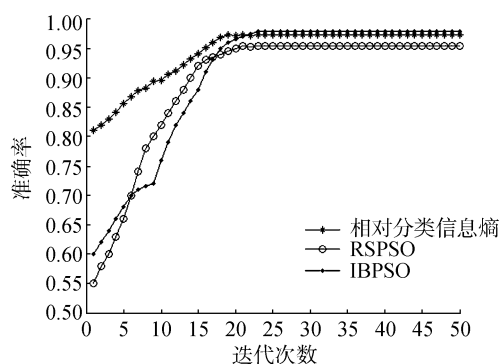


图 2 在 Kr-vs-Kp 数据集上的实验结果

Fig.2 The experimental results of data set Kr-vs-Kp

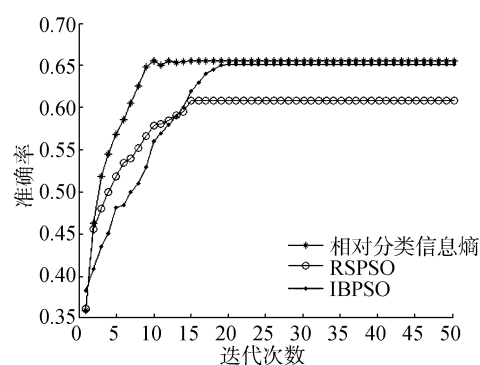


图 3 在 Voting 数据集上的实验结果

Fig.3 The experimental results of data set Voting

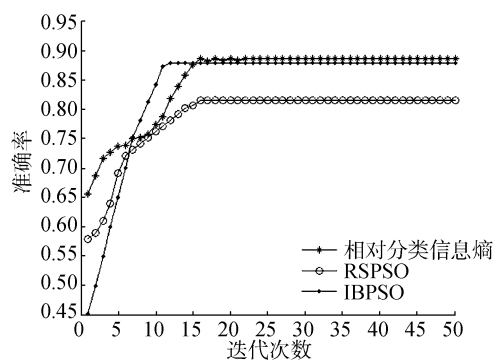


图 4 在 Mushroom 数据集上的实验结果

Fig.4 The experimental results of data set Mushroom

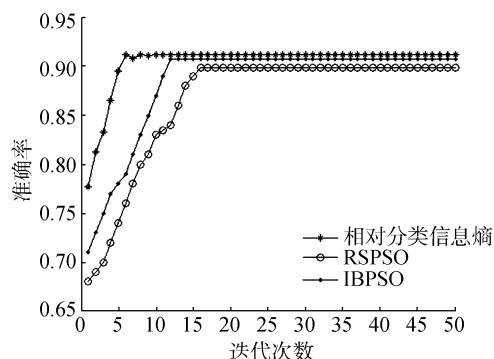


图 5 在 Nursery 数据集上的实验结果

Fig.5 The experimental results of data set Nursery

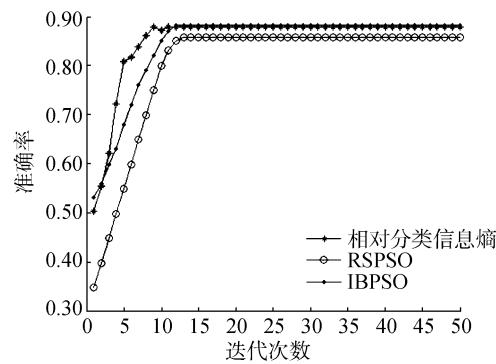


图6 在 Car 数据集上的实验结果

Fig.6 The experimental results of data set Car

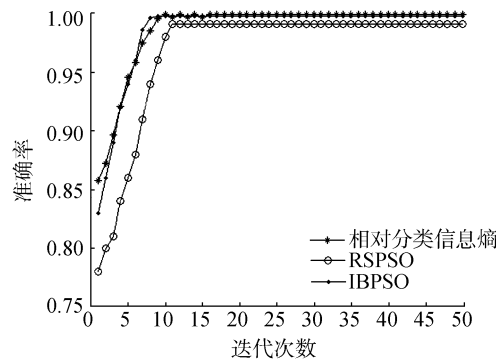


图7 在 Soybean 数据集上的实验结果

Fig.7 The experimental results of data set Soybean

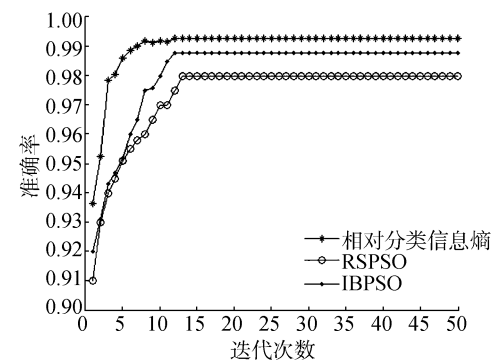


图8 在 Tic-Tac-Toe 数据集上的实验结果

Fig.8 The experimental results of data set Tic-Tac-Toe

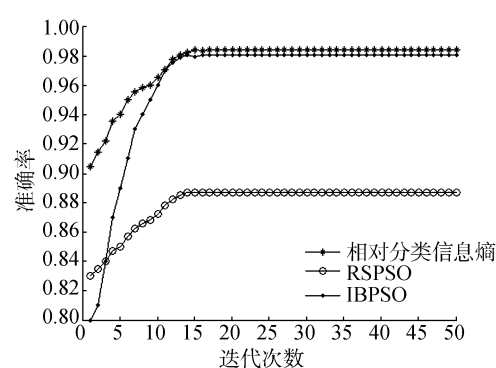


图9 在 SPECT 数据集上的实验结果

Fig.9 The experimental results of data set SPECT

通过上述实验,我们可以看出采用相对分类信息熵作适应度函数的粒子群算法与单纯粗糙集知识作为适应度函数的粒子群算法相比,在实验精度和筛选特征个数方面要明显占优。

接下来实验中,比较粒子群算法和遗传算法在处理特征选择上的优劣,对比方法为文献[35]中的采用相同适应度函数的遗传特征选择方法。从#Fs、Ac、 σ 和Time 4个方面进行了比较,结果列于表2和表3中。其中,#Fs表示选出的特征数,Ac表示测试精度, σ 表示分类标准差,Time表示运行时间。更直观的实验结果如图10~17所示。

表2 与遗传算法比较的实验结果(1)

| 数据集 | 本文方法 | | 文献[35]方法 | |
|-------------|------|---------|----------|---------|
| | #Fs | Ac | #Fs | Ac |
| Kr-vs-Kp | 12 | 0.973 3 | 13 | 0.951 6 |
| Voting | 10 | 0.655 8 | 12 | 0.622 4 |
| Mushroom | 8 | 0.882 9 | 10 | 0.837 1 |
| Nursery | 6 | 0.919 7 | 7 | 0.902 1 |
| Car | 5 | 0.877 3 | 5 | 0.862 8 |
| Soybean | 16 | 0.998 7 | 19 | 0.997 9 |
| Tic-Tac-Toe | 7 | 0.992 9 | 7 | 0.993 5 |
| SPECT | 12 | 0.984 2 | 15 | 0.976 9 |

表3 与遗传算法比较的实验结果(2)

| 数据集 | 本文方法 | | 文献[35]方法 | |
|-------------|-----------------------|------------|-----------------------|-------------|
| | σ | Time | σ | Time |
| Kr-vs-Kp | 2.27×10^{-3} | 415.257 6 | 4.75×10^{-3} | 508.584 7 |
| Voting | 4.62×10^{-3} | 24.316 6 | 7.15×10^{-3} | 26.047 9 |
| Mushroom | 1.08×10^{-4} | 85.121 8 | 6.68×10^{-3} | 96.008 6 |
| Nursery | 8.72×10^{-3} | 2311.143 8 | 1.04×10^{-2} | 2 653.861 2 |
| Car | 5.19×10^{-3} | 10.526 8 | 3.97×10^{-3} | 10.288 5 |
| Soybean | 2.01×10^{-3} | 35.173 6 | 2.59×10^{-3} | 64.110 1 |
| Tic-Tac-Toe | 3.68×10^{-3} | 12.163 5 | 6.02×10^{-3} | 19.982 8 |
| SPECT | 5.43×10^{-3} | 20.563 9 | 7.52×10^{-3} | 29.761 1 |

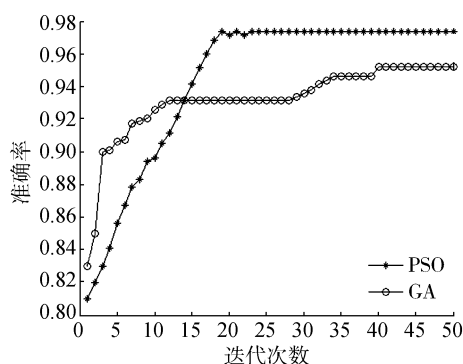


图 10 在数据集 Kr-vs-Kp 上的实验结果(PSO 与 GA 对比)

Fig.10 The experimental results of data set Kr-vs-Kp(PSO,GA)

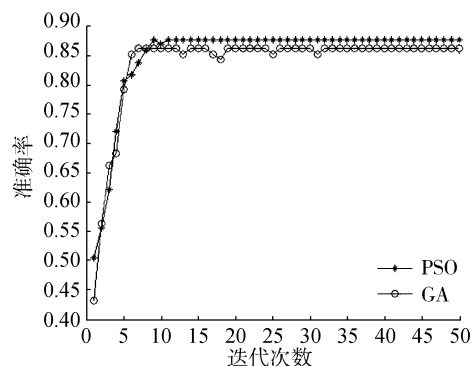


图 14 在数据集 Car 上的实验结果(PSO 与 GA 对比)

Fig.14 The experimental results of data set Car(PSO,GA)

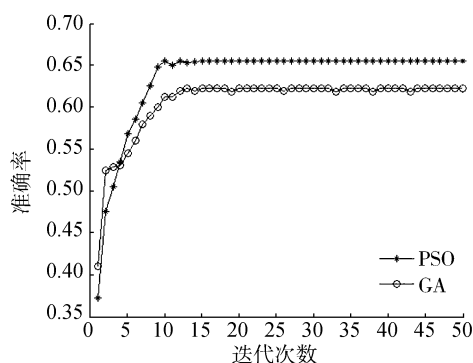


图 11 在数据集 Voting 上的实验结果(PSO 与 GA 对比)

Fig.11 The experimental results of data set Voting(PSO,GA)

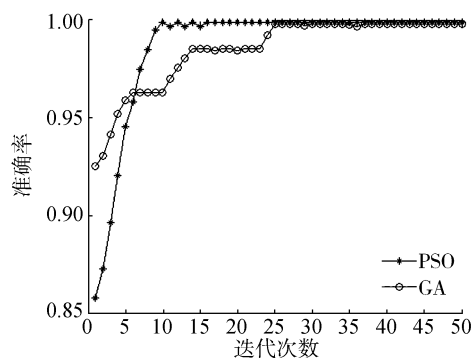


图 15 在数据集 Soybean 上的实验结果(PSO 与 GA 对比)

Fig.15 The experimental results of data set Soybean(PSO,GA)

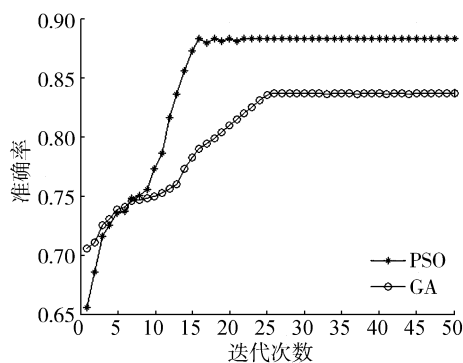


图 12 在数据集 Mushroom 上的实验结果(PSO 与 GA 对比)

Fig.12 The experimental results of data set Mushroom(PSO,GA)

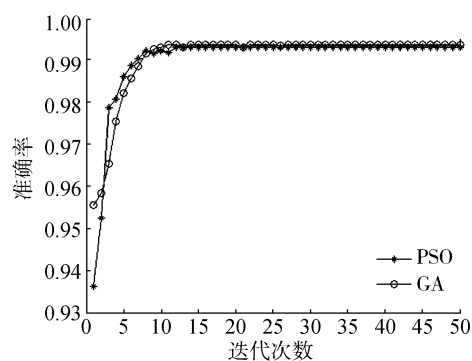


图 16 在数据集 Tic-Tac-Toe 上的实验结果(PSO 与 GA 对比)

Fig.16 The experimental results of data set Tic-Tac-Toe(PSO,GA)

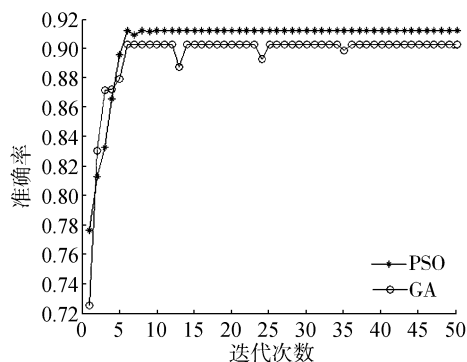


图 13 在数据集 Nursery 上的实验结果(PSO 与 GA 对比)

Fig.13 The experimental results of data set Nursery(PSO,GA)

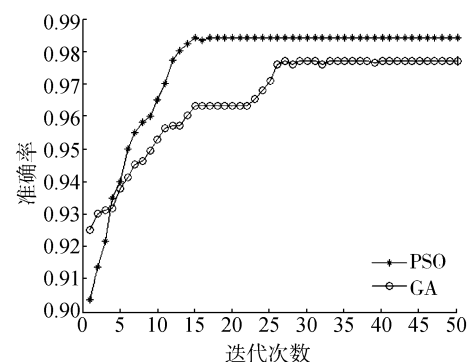


图 17 在数据集 SPECT 上的实验结果(PSO 与 GA 对比)

Fig.17 The experimental results of data set SPECT(PSO,GA)

从表2、表3、图10~11可以看出,用粒子群算法选出的特征子集在测试精度总体上优于遗传算法,且收敛速度比遗传算法有大幅提升,筛选出的特征个数和运行时间较遗传算法有一定程度的提高,两者分类标准差基本相当。其原因在于,粒子群算法在迭代的过程中具有一定的记忆功能,能够在保持自身优势的同时感知全局走向,且不需要进行较多的参数设置,省去了遗传和变异的过程,实现起来更容易。实验结果说明本文方法是行之有效的。

为进一步验证方法的有效性,由上述两种方法分别对每个数据集随机实验10次,得到的10组测试结果组成两个10维向量,记为 X_1 和 X_2 ,对 X_1 和 X_2 进行配对 t 检验, p 值如表4所示。

表4 配对 t 检验 p 值Table 4 p values of paired t -test

| 数据集 | P 值 |
|-------------|---------------------------|
| Kr-vs-Kp | $1.513\ 9\times 10^{-9}$ |
| Voting | $6.123\ 7\times 10^{-8}$ |
| Mushroom | $9.065\ 7\times 10^{-10}$ |
| Nursery | $2.638\ 1\times 10^{-7}$ |
| Car | $7.218\ 3\times 10^{-11}$ |
| Soybean | $1.026\ 2\times 10^{-4}$ |
| Tic-Tac-Toe | $8.820\ 9\times 10^{-12}$ |
| SPECT | $1.633\ 8\times 10^{-15}$ |

p 值均较小,显著性差异较大,进一步证明了方法的可行。

4 结论

本文选择相对分类信息熵作为适应度函数,用粒子群算法寻找最优特征子集,在解决离散值特征选择问题上效果较好,并与其他方法进行了比较。通过实验我们得出以下结论:1)粒子群算法相比遗传算法过程更简单,省去了复杂的参数设置,容易实现;2)粒子群算法拥有较好的测试精度,且在选择的特征个数和运行时间方面都有一定程度的提高;3)粒子群算法收敛较快,可以更早得到拥有更好表示能力的特征子集。

参考文献:

[1] GUYON I, GUNN S, NIKRAVESH M, et al. Feature extraction, foundations and applications[M]. Berlin: Springer, 2006.
 [2] DASH M, LIU H. Feature selection for classification[J]. Intelligent data analysis, 1997, 1: 131-151.

[3] PAWLAK Z. Rough sets[J]. International journal of information and computer sciences, 1982, 11: 341-356.
 [4] 苗夺谦, 李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008.
 [5] SWINIARSKI R W, SKOWRON A. Rough set methods in feature selection and recognition[J]. Pattern recognition letters, 2003, 24(6): 833-849.
 [6] JENSEN R, SHEN Q. Fuzzy-rough sets for descriptive dimensionality reduction[C]// IEEE International Conference on Fuzzy Systems, 2002. Fuzz-IEEE. 2002:29-34.
 [7] BHATT R B, GOPAL M. On fuzzy-rough sets approach to feature selection[J]. Pattern recognition letters, 2005, 26(7): 965-975.
 [8] JENSEN R, PARTHALAIN N M. Towards scalable fuzzy rough feature selection[J]. Information sciences, 2015, 323(C): 1-15.
 [9] QIAN Y H, LIANG J, PEDRYCZ W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory[J]. Artificial intelligence, 2010, 174(9/10): 597-618.
 [10] HU Q H, YU D R, LIU J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information sciences, 2008, 178(18): 3577-3594.
 [11] ALMUALIM H, DIETTERICH T G. Learning boolean concepts in the presence of many irrelevant features[J]. Artificial intelligence, 1994, 69(1/2): 279-305.
 [12] DASH M, LIU H. Consistency-based search in feature selection[J]. Artificial intelligence 2003(151):155-176.
 [13] BATTITI R. Using mutual information for selecting features in supervised neural net learning[J]. IEEE transactions on neural networks, 1994, 5(4): 537-549.
 [14] KWAK N, CHOI C H. Input feature selection by mutual information based on parzen window[J]. IEEE transactions on pattern analysis and machine intelligence, 2002, 24(12): 1667-1671.
 [15] ESTEVEZ P A, TESMER M, PEREZ C A, et al. Normalized mutual information feature selection[J]. IEEE transactions on neural networks, 2009, 20(2): 189-201.
 [16] SONG L, SMOLA A, GRETTON A, et al. Feature selection via dependence maximization[J]. Journal of machine learning research, 2012, 13:1393-1434.
 [17] HU Q H, ZHU Pengfei, LIU Jinfu, et al. Feature selection via maximizing fuzzy dependency[J]. Fundamenta informaticae, 2010, 98: 167-181.
 [18] KOHAVI R, JOHN G. Wrappers for feature subset selection[J]. Artificial intelligence, 1997, 97(1/2): 273-324.
 [19] SINDHWANI V, RAKSHIT S, DEODHARE D, et al. Feature selection in MLPs and SVMs based on maximum output information[J]. IEEE transactions on neural networks, 2004, 15(4): 937-947.
 [20] YANG Jianbo, SHEN Kaiquan, ONG Chongjin, et al. Fea-

- ture selection for MLP neural network: the use of random permutation of probabilistic outputs[J]. IEEE transactions on neural networks, 2009, 20(12): 1911–1922.
- [21] QUINLAN J R. Induction of decision trees [J]. Machine learning, 1986, 1: 81–106.
- [22] BREIMAN L, FRIEDMAN J H, RICHARD A S, et al. Classification and regression trees[M]. Belmont, CA: wadsworth international group, 1984.
- [23] SETIONO R, LIU H. Neural-network feature selector [J]. IEEE transactions on neural networks, 1997, 8(3): 654–662.
- [24] SHEN Kaiquan, ONG Chongjin, LI Xiaoping, et al. Feature selection via sensitivity analysis of SVM probabilistic outputs[J]. Machine learning, 2008, 70: 1–20.
- [25] PERKINS S, LACKER K, THEILER J. Grafting: fast, incremental feature selection by gradient descent in function space [J]. Journal of machine learning research, 2003 (3): 1333–1356.
- [26] KENNEDY J, EBERHART R. Particle swarm optimization [C]. IEEE International Conference on Neural Networks. Perth, Australia, 1995, 4: 1942–1948.
- [27] EBERHART R C, SHI Y H, KENNEDY J. Swarm Intelligence[M]. Massachusetts: Morgan Kaufmann, 2001.
- [28] EBERHART R C, KENNEDY J. A discrete binary version of the particle swarm algorithm [J]. IEEE conference on systems, 1997, 5: 4104–4109.
- [29] CHUANG L Y, CHANG H W, TU C J, et al. Improved binary PSO for feature selection using gene expression data[J]. Computational biology & chemistry, 2008, 32(1): 29–37.
- [30] CHUANG L Y, TSAI S W, YANG C H. Improved binary particle swarm optimization using catfish effect for feature selection[J]. Expert systems with applications, 2011, 38(10): 12699–12707.
- [31] WANG Xiangyang, YANG Jie, TENG Xiaolong, et al. Feature selection based on rough sets and particle swarm optimization [J]. Pattern recognition letters, 2007, 28(4): 459–471.
- [32] CERVANTE L, XUE B, ZHANG M, et al. Binary particle swarm optimisation for feature selection: a filter based approach[J]. Evolutionary computation, 2012, 41: 1–8.
- [33] LIU Qianjin, ZHAO Zhimin, LI Yingxin. Ensemble feature selection method based on neighborhood information and PSO algorithm[J]. Acta electronica sinica, 2016, 44(4): 995–1002.
- [34] FONG S, WONG R, VASILAKOS A. Accelerated PSO swarm search feature selection for data stream mining big data[J]. IEEE transactions on services computing, 2016, 9(1): 33–45.
- [35] 翟俊海, 刘博, 张素芳. 基于相对分类信息熵的进化特征选择算法[J]. 模式识别与人工智能, 2016, 29(8): 682–690.
- ZHAI Junhai, LIU Bo, ZHANG Sufang. Feature selection via evolutionary computation based on relative classification information entropy [J]. Pattern recognition and artificial intelligence, 2016, 29(8): 682–690.
- [36] SHI B Y, EBERHART R. A modified particle swarm optimizer[J]. IEEE world congress on computational intelligence, 1999, 6: 69–73.

作者简介:



翟俊海,男,1964年生,男,教授,中国人工智能学会粗糙集与软计算专业委员会委员,主要研究方向为机器学习。近几年主持或参与省部级以上项目10余项,获河北省自然科学三等奖1项,出版专著4部,发表论文70余篇。



刘博,男,1989年生,硕士研究生,主要研究方向为机器学习。



张素芳,女,1966年生,副教授,主要研究方向为机器学习。