

DOI:10.16356/j.1005-2615.2019.05.015

基于粗糙集特征选择的过拟合现象及应对策略

张文冬¹ 亓慧² 刘克宇¹ 杨习贝¹

(1. 江苏科技大学计算机学院, 镇江, 212003; 2. 太原师范学院计算机系, 太原, 030619)

摘要: 在粗糙集方法中, 利用向前启发式算法进行特征选择, 是一个逐步加入重要度最高的特征的过程, 直至满足所给定的约束条件。但使用这一策略选择出来的特征子集有可能产生过拟合现象。鉴于此, 设计了一种新的启发式算法, 其主要思想是借助交叉验证的方法对特征的重要度进行计算, 当过拟合出现时, 则采用截断式机制终止算法。使用邻域粗糙集模型, 在 UCI 数据集上将启发式算法与所提算法进行对比分析, 实验结果表明: 所提算法能够有效地降低过拟合的程度; 利用所提算法得到的特征子集能够带来更好的分类性能。

关键词: 特征选择; 启发式算法; 邻域粗糙集; 过拟合

中图分类号: TP391 **文献标志码:** A **文章编号:** 1005-2615(2019)05-0687-06

Over-Fitting and Its Countermeasure in Feature Selection Based on Rough Set

ZHANG Wendong¹, QI Hui², LIU Keyu¹, YANG Xibei¹

(1. School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212003, China;
2. Computer Science and Technology Department, Taiyuan Normal University, Taiyuan, 030619, China)

Abstract: In rough set theory, forward heuristic algorithm selects the most important feature in the process of feature selection until the given constraint is satisfied. However, the feature subset selected by such strategy may bring us over-fitting. To solve this problem, a new heuristic algorithm is designed. The importance of the feature is obtained by cross validation and then the early stopping is employed to terminate the algorithm when over-fitting occurs. Based on the neighborhood rough set, the new method is compared with the heuristic algorithm over several UCI data sets. The experimental results show that: the proposed algorithm can effectively reduce the degree of over-fitting, and the feature subset obtained by the new algorithm may offer better classification performances.

Key words: feature selection; heuristic algorithm; neighborhood rough set; over-fitting

作为求解特征选择问题的一种有效技术手段, 启发式策略^[1-2]受到了众多学者的广泛关注。因为这一策略是建立在贪心搜索基础上的, 所以其具有较高的时间效率。此外, 启发式算法具备相当的灵活性, 因为针对不同的特征选择目标, 仅需设计不同的特征重要度函数, 无需修改算法的整体框架。

特征重要度函数的设计与特征选择的语义解释是紧密相关的。例如, 在粗糙集理论中^[3-4], 若以近似逼近^[5]设计约束条件进行特征选择, 则特征的

重要度可以定义为特征子集的变化对近似逼近的影响程度; 若以获取较低的决策代价为目标^[6-8], 则特征的重要度可以定义为特征子集的变化对决策代价的影响程度。

虽然基于粗糙集的特征选择具有丰富的语义解释, 并已在众多领域得到了成功的应用^[9-12], 但这一方法依然有可能会带来过拟合现象。这种过拟合现象与学习任务中的过拟合是有一定区别的, 因为特征选择的目标是找到一个满足给定约束条件

基金项目: 国家自然科学基金(61572242, 61502211, 61503160)项目资助。

收稿日期: 2018-05-10; **修订日期:** 2018-06-30

通信作者: 亓慧, 女, 硕士, 副教授, E-mail: qihui@tynu.edu.cn。

引用格式: 张文冬, 亓慧, 刘克宇, 等. 基于粗糙集特征选择的过拟合现象及应对策略[J]. 南京航空航天大学学报, 2019, 51(5): 687-692. ZHANG Wendong, QI Hui, LIU Keyu, et al. Over-Fitting and Its Countermeasure in Feature Selection Based on Rough Set[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 687-692.

的特征子集或者进行特征排序^[13-16],而非训练出一个学习模型。所以特征选择带来的过拟合现象可以表现为在训练样本上所得到的特征子集满足所给的约束条件,而在测试样本上利用这一特征子集则有可能达不到约束条件。

在一般的学习任务中,降低过拟合的方法通常有 Early Stopping^[17],正则化^[18]等。依据 Early Stopping的基本思想:在每一个 epoch 结束时(一个 epoch 即对所有训练数据的一轮遍历)计算测试数据的分类精度,当分类精度不再提高时,就停止训练。鉴于此,在特征选择过程中,可以设计一种截断式的启发式搜索策略:利用交叉验证^[19-20]的方法对特征的重要度进行计算,并逐步加入重要度最高的特征,当过拟合发生时就立即停止搜索。

1 邻域粗糙集方法

在邻域粗糙集理论中,研究对象为一个邻域决策系统 $DS = \langle U, AT \cup D \rangle$,其中, $U = \{x_1, x_2, \dots, x_n\}$ 是由 n 个样本组成的非空有限集合,称为论域; AT 是所有条件特征集合; D 为决策特征。

定义 1 给定一个邻域决策系统 $DS, \forall x \in U, \forall \delta \in [0, 1], r(x, y)$ 为欧氏距离函数; $\forall B \supseteq AT, M$ 是由特征集合 B 所得到的相似度矩阵,则点集 $\delta_B(x) = \{y | y \in U : r(x, y) \leq \delta\}$ 表示 x 的 δ 邻域, δ 称为邻域半径。

给定论域 $U = \{x_1, x_2, \dots, x_n\}$, r_{ij} 表示样本 x_i 与 x_j 之间的欧氏距离。为了解决因半径过小而产生邻域信息量不足的问题^[21],可以采用邻域区间的表示方法。给定半径 $\delta, \forall x_i \in U, x_i$ 的邻域区间为

$$\text{Int}(x_i) = \min_{1 \leq j \leq n, j \neq i} r_{ij} + \delta \times \left(\max_{1 \leq j \leq n, j \neq i} r_{ij} - \min_{1 \leq j \leq n, j \neq i} r_{ij} \right) \quad (1)$$

式中: $\min_{1 \leq j \leq n, j \neq i} r_{ij}$ 表示与样本 x_i 的距离的最小值, $\max_{1 \leq j \leq n, j \neq i} r_{ij}$ 表示与样本 x_i 的距离的最大值。借助邻域区间, $\forall x_i \in U, \forall B \supseteq AT$, 其邻域为

$$\delta_B(x_i) = \{x_j \in U | x_j \neq x_i, r_{ij} \leq \text{Int}(x_i)\} \quad (2)$$

定义 2 给定一个邻域决策系统 $DS, \forall B \supseteq AT, \forall X \supseteq U$, X 的邻域下近似集与上近似集分别定义如下

$$\underline{B}(X) = \{x_i \in U : \delta_B(x_i) \supseteq X\} \quad (3)$$

$$\overline{B}(X) = \{x_i \in U : \delta_B(x_i) \cap X \neq \emptyset\} \quad (4)$$

$[\underline{B}(X), \overline{B}(X)]$ 即为 X 的邻域粗糙集。

定义 3 给定一个邻域决策系统 $DS, U/\text{IND}(D) = \{X_1, X_2, \dots, X_N\}$ 是由决策特征 D 所诱导的论域上的划分, $\forall B \supseteq AT$, 决策特征 D 相对于 B 在 U 上的近似质量可用 $\gamma_u(B, D)$ 表示, 其定义如下

$$\gamma_u(B, D) = \frac{\left| \bigcup_{i=1}^N \underline{B}(X_i) \right|}{|U|} \quad (5)$$

式中 $|X|$ 表示集合 X 的基数。 $\forall X_i \in U/\text{IND}(D), X_i$ 为决策类。

2 特征选择

基于粗糙集的特征选择是粗糙集理论研究中的核心问题之一。其本质是在考虑一些度量标准的前提下,删除数据中不相关或冗余的特征。以下将讨论邻域粗糙集框架下的特征选择方法。

2.1 传统启发式算法

定义 4 给定一个邻域决策系统 $DS, \forall B \supseteq AT, B$ 为 AT 的一个约简当且仅当

$$(1) \gamma_u(B, D) = \gamma_u(AT, D)$$

$$(2) \forall B' \subset B, \gamma_u(B', D) \neq \gamma_u(B, D)$$

由定义 4 可知,条件特征 AT 的约简 B 是一个能够保持近似质量不变的最小特征子集。根据定义 3 所示的近似质量,可进一步考察特征的重要度。

定义 5 给定一个邻域决策系统 $DS, \forall B \supseteq AT, \forall a_i \in AT - B$, 特征 a_i 在 U 上的重要度定义如下

$$\text{Sig}_u(a_i, B, D) = \gamma_u(B \cup \{a_i\}, D) - \gamma_u(B, D) \quad (6)$$

$\text{Sig}_u(a_i, B, D)$ 反映了当 a_i 被加入到 B 时近似质量的变化程度。若 $\gamma_u(B \cup \{a_i\}, D) = \gamma_u(B, D)$, 则表示特征 a_i 对于计算近似质量没有带来任何贡献, a_i 是冗余的; 若 $\gamma_u(B \cup \{a_i\}, D) > \gamma_u(B, D)$, 则表示加入特征 a_i 后可以提高近似质量, 从而降低不确定性程度。

基于上述特征重要度的定义,启发式搜索算法可以用来迭代地选择邻域决策系统中最重要特征子集。形式化的启发式算法(算法 1)步骤如下:

算法 1(启发式算法):

输入: 邻域决策系统 DS , 邻域半径 δ

输出: 选择出的特征子集 B

步骤 1 $B \leftarrow \emptyset$;

步骤 2 $\forall a_i \in AT$, 计算 $\text{Sig}_u(a_i, B, D)$, 其中 $\gamma_u(\emptyset, D) = 0$;

步骤 3 选择一个特征 a_j 满足

$$\text{Sig}_u(a_j, B, D) = \max\{\forall a_i \in AT : \text{Sig}_u(a_i, B, D)\};$$

若 $\text{Sig}_u(a_j, B, D) > 0$, 则转步骤 5, 否则转步骤 6

步骤 5 (1) $B = B \cup \{a_j\}$;

(2) $\forall a_i \in AT - B$, 计算 $\text{Sig}_u(a_i, B, D)$;

(3) 选择一个特征 a_j 满足

$$\text{Sig}_u(a_j, B, D) = \max\{\forall a_i \in AT - B : \text{Sig}_u(a_i, B, D)\};$$

步骤 6 输出特征子集 B

算法 1 的时间复杂度为 $O(|U|^2 \cdot |AT|^2)$ 。其中 $|U|$ 为样本个数, $|AT|$ 为条件特征个数。算法 1 每次把最接近目标近似质量的特征依次加入到被选的特征子集中去,直到近似质量不再随特征的增加而增大为止。然而,这种策略选择出来的特征子集有时会陷入局部最优或者过于优秀,从而导致所选的特征子集在测试集上出现过拟合。

例如,在 UCI 数据集中,取“Wine”这个数据集前 60% 的数据作为训练集来进行特征选择,剩下 40% 的数据当作测试集。半径 δ 取 0.1,经过计算,原始训练集的近似质量为 0.904 0;经特征选择后训练集的近似质量为 0.928 0;原始测试集的近似质量为 0.981 1;特征子集作用于测试集后的近似质量为 0.886 8。可见特征选择后虽然提高了训练集的近似质量,但当所选特征子集作用于测试集时,所得近似质量却不如原始测试集的近似质量。

2.2 截断式启发式算法

为了减轻由启发式算法进行特征选择所带来的过拟合现象,以下将在启发式算法中引入 Early Stopping 的基本思想。Early Stopping 实际上是采用交叉验证的方法,在特征选择进程中模拟过拟合现象,并在过拟合发生时停止搜索。鉴于此,所提出的截断式启发式算法(算法 2)步骤如下:

算法 2(截断式启发式算法):

输入: 邻域决策系统 DS , 邻域半径 δ , 交叉验证

折数 m , 特征个数 n

输出: 选择出的特征子集 B

步骤 1 将数据分成 m 份: U_1, U_2, \dots, U_m ;

步骤 2 $B \leftarrow \emptyset, \gamma_{\text{after}} \leftarrow \text{inf}, \gamma_{\text{before}} \leftarrow 0, k \leftarrow 1$;

步骤 3 若 $k = n$, 则转步骤 11;

步骤 4 若 $k \geq 2$, 则 $\gamma_{\text{before}} \leftarrow \gamma_{\text{after}}$;

步骤 5 For $p = 1$ to m

(1) 取 $U_p' = U - U_p$;

(2) $\forall a_i \in AT - B$, 计算

$\text{Sig}_{U_p'}(a_i, B, D)$;

End

步骤 6 $\forall a_i \in AT - B$, 计算平均重要度

$$\text{mean}_{a_i} = \frac{\sum_{p=1}^m \text{Sig}_{U_p'}(a_i, B, D)}{m};$$

步骤 7 选择一个特征 a_j 满足 $\text{Sig}_{U_p'}(a_j, B, D) = \max\{\text{mean}_{a_i} : \forall a_i \in AT - B\}$ 且 $B = BU\{a_j\}$;

步骤 8 For $p = 1$ to m

(1) 取 $U_p' = U_p$

(2) 计算 $\gamma_{U_p'}(B, D)$;

End

步骤 9

$$(1) \quad \gamma_{\text{after}} = \frac{\sum_{p=1}^m \gamma_{U_p'}(B, D)}{m}$$

(2) $k \leftarrow k + 1$;

步骤 10 若 $\gamma_{\text{before}} \geq \gamma_{\text{after}}$, 则删除步骤 7 加入的特征并转步骤 11, 否则返回步骤 3;

步骤 11 输出属性子集 B

算法 2 的时间复杂度为 $O(m \cdot |U|^2 \cdot |AT|^2)$; 其中 m 为交叉验证的折数, $|U|$ 为样本个数, $|AT|$ 是条件特征个数。

与文献[17]中的方法相比,所提出的截断式启发式算法重点是找到一个满足给定约束条件的特征子集或者进行特征排序,而非是为了训练出一个学习模型。简言之,截断式启发式算法可以有效的抑制在特征选择时出现的过拟合现象,这为进一步研究粗糙集特征选择提供了一个新的思路。

3 实验分析

为了验证截断式启发式算法的有效性,选取了 9 组 UCI 数据集进行实验分析,数据集信息的基本描述如表 1 所示。实验环境为 PC 机,双核 2.60 GHz CPU,4 GB 内存,Windows10 操作系统, MATLAB R2016a 实验平台。

表 1 数据集描述

Tab. 1 Data sets description

ID	数据集名称	样本 个数	特征 个数	类别 个数
1	Breast cancer Wisconsin (Diagnostic)	569	30	2
2	Breast tissue	106	9	6
3	Dermatology	366	34	6
4	Drug consumption	1 885	30	7
5	Ecoli	336	7	8
6	Forest	523	27	4
7	Statlog(Vehicle-Silhouettes)	846	19	4
8	Vertebral column	310	6	2
9	Wilt	4 839	5	2

在邻域粗糙集中,当半径较大时,近似质量会趋向于 0,故为了能更好地展现近似质量的变化,适当选取了 10 个不同的半径 δ , 值分别为 0.02, 0.04, ..., 0.20。

实验采用了 10 折交叉验证的方法。10 折交叉验证的具体过程是将实验数据中的样本平均分成 10 份,即 U_1, U_2, \dots, U_{10} , 第 1 次使用 $U_2 \cup U_3 \cup \dots \cup U_{10}$ 作为训练集求得约简 red_1 , 使用 U_1 作为测试集求得 red_1 的近似质量;第 2 次使用 $U_1 \cup U_3 \cup \dots \cup U_{10}$ 作为训练集求得约简 red_2 , 使用 U_2 作为测试集求得 red_2 的近似质量;依次类推,第 10 次使用 $U_1 \cup U_2 \cup \dots \cup U_9$ 作为训练集求得约简 red_{10} , 使用 U_{10} 作为测试集求得 red_{10} 的近似质量。

3.1 近似质量比较

本节分别在训练集和测试集上对比了使用算法 1 和算法 2 进行特征选择前后的近似质量,实验结果如图 1 所示。

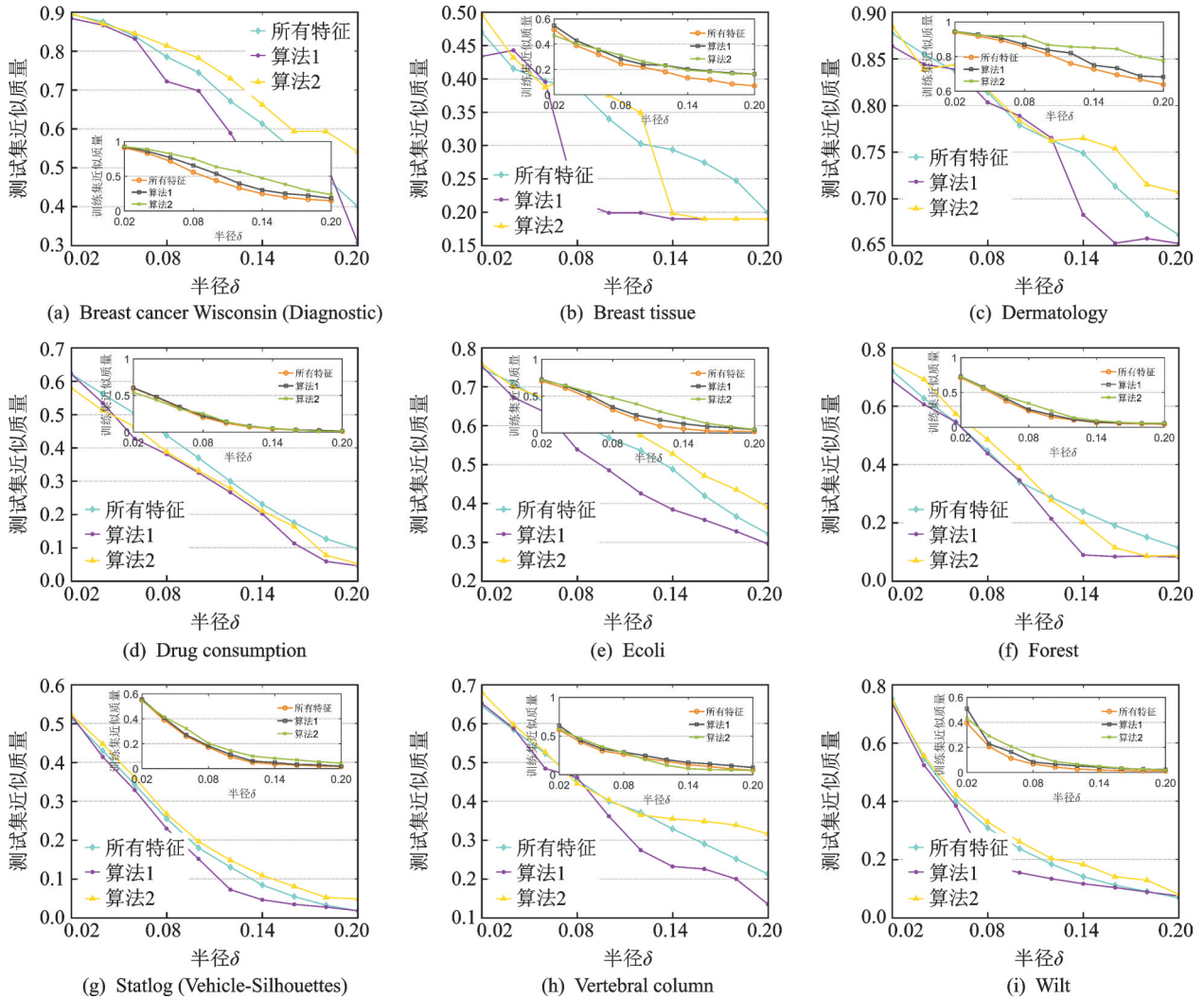


图1 近似质量

Fig. 1 Approximation qualities

观察图1, 不难发现, 利用算法1选择出来的特征子集在测试集上出现了过拟合, 因此可以得到如下结论:

(1) 利用传统启发式算法选择出的特征子集在训练集上都能获得较高的近似质量, 但当所选特征子集作用于测试集时, 却产生了较低的近似质量, 即出现了过拟合现象。

(2) 利用截断式启发式算法选择出的特征子集在训练集上依然满足约简所给定约束条件, 并且当所选特征子集作用于测试集时, 能够有效地降低甚至消除过拟合现象, 例如在数据集“Wilt”中, 截断式启发式算法就完全消除了过拟合现象。

3.2 分类精度比较

采用邻域分类器^[21](NEC), KNN分类器(K 的取值为5), SVM, 分类回归树(CART)来对比利用两种算法进行特征选择后在测试集上的分类精度, 结果是取文中使用10个给定半径在测试集上分类精度的平均值, 实验结果如表2所示。

根据表2所示的实验结果不难发现, 利用截断式启发式算法所得到的特征子集往往能够提供较好的分类性能。例如, 在KNN分类器上, 利用算法2得到特征子集所产生的平均分类精度比利用算法1得到特征子集所产生的平均分类精度高出了5%左右。

3.3 约简长度和约简时间比较

最后, 将两种算法所得到的特征子集的长度进行了对比, 同时亦对比了两种算法的时间消耗, 实验结果如表3所示。

观察表3不难发现, 利用截断式启发式算法所选择出的特征子集长度明显高于利用传统启发式算法所得到的特征子集长度。此外, 算法2的时间消耗也高于算法1的时间消耗。这主要是因为: (1) 截断式启发式算法在求解特征重要度时运用了交叉验证; (2) 利用截断式启发式算法往往会输出更多的特征, 因而其迭代次数较多。

表 2 分类精度

Tab. 2 Classification accuracies

ID	NEC		KNN		SVM		CART	
	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2
1	0.933 0	0.952 9	0.929 8	0.964 9	0.936 8	0.955 7	0.929 8	0.912 3
2	0.498 5	0.511 2	0.699 2	0.745 4	0.462 0	0.484 5	0.680 3	0.723 6
3	0.755 6	0.753 6	0.772 5	0.788 4	0.771 3	0.771 8	0.756 6	0.820 1
4	0.920 9	0.933 2	0.914 9	0.924 2	0.922 0	0.940 5	0.891 9	0.918 9
5	0.791 9	0.837 7	0.801 6	0.832 6	0.801 6	0.832 6	0.768 6	0.806 4
6	0.694 8	0.725 0	0.699 3	0.745 5	0.721 6	0.750 2	0.680 3	0.723 7
7	0.576 8	0.647 5	0.494 1	0.705 9	0.505 4	0.575 0	0.541 2	0.588 2
8	0.789 0	0.744 8	0.645 2	0.741 9	0.786 5	0.780 0	0.645 2	0.741 9
9	0.946 1	0.946 1	0.933 9	0.925 6	0.947 0	0.949 5	0.913 2	0.921 5
平均值	0.767 4	0.783 6	0.765 6	0.819 3	0.761 5	0.782 2	0.756 3	0.795 1

表 3 特征子集长度与时间消耗

Tab. 3 Lengths of feature subsets and time consumptions

ID	约简长度		约简时间/s	
	算法 1	算法 2	算法 1	算法 2
1	2.500 0	4.390 0	1.421 6	8.537 4
2	1.950 0	2.290 0	0.015 4	0.095 5
3	6.320 0	7.930 0	1.323 5	6.121 7
4	9.790 0	4.910 0	60.509 8	135.931 7
5	3.520 0	4.940 0	1.413 2	6.746 7
6	2.540 0	4.280 0	0.119 6	0.631 5
7	3.450 0	5.860 0	2.747 2	14.949 0
8	2.440 0	2.260 0	0.066 6	0.274 8
9	1.520 0	2.700 0	13.632 3	76.478 1

4 结 论

传统启发式算法根据所给约束条件选择出的特征子集有可能在测试集上出现过拟合现象。为了缓解这一现象,提出了一种截断式启发式算法:利用交叉验证的方法来进行特征选择,当过拟合发生时采用截断式的思想停止算法。实验结果表明,与传统的启发式算法相比,新算法能有效地降低过拟合现象的发生,此外,利用新算法所得到的特征子集亦能够提升分类器的分类性能。

在本文工作的基础上,笔者将进一步研究特征选择中过拟合现象的背后机理,力争完全消除过拟合现象并进一步提高特征选择的时间效率。

参考文献:

[1] 杨习贝, 颜旭, 徐苏平, 等. 基于样本选择的启发式属性约简方法研究[J]. 计算机科学, 2016, 43(1): 40-43.
 YANG Xibei, YAN Xu, XU Suping, et al. New heuristic attribute reduction algorithm based on sample selection [J]. Computer Science, 2016, 43(1): 40-43.

[2] XU S P, YANG X B, YU H L, et al. Multi-label learning with label-specific feature reduction[J]. Knowledge Based Systems, 2016, 104: 52-61.
 [3] PAWLAK Z. Rough sets-theoretical aspects of reasoning about data [M]. Dordrecht, Boston, London: Kluwer Academic Publishers, 1991.
 [4] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.
 [5] YANG X B, QI Y, YU H L, et al. Updating multigranulation rough approximations with increasing of granular structures[J]. Knowledge Based Systems, 2014, 64: 59-69.
 [6] ZHU H Y, WANG X Z. A cost-sensitive semi-supervised learning model based on uncertainty [J]. Neurocomputing, 2017, 251: 106-114.
 [7] JU H R, YANG X B, YU H L, et al. Cost-sensitive rough set approach[J]. Information Sciences, 2016, 355/356: 282-298.
 [8] DOU H L, YANG X B, SONG X N, et al. Decision-theoretic rough set: A multicost strategy [J]. Knowledge Based Systems, 2016, 91: 71-83.
 [9] KANEIWA K, KUDO Y. A sequential pattern mining algorithm using rough set theory [J]. International Journal of Approximate Reasoning, 2011, 52(6): 881-893.
 [10] MARTINEZ I G, PEREZ R E B. Making decision in case based systems using probabilities and rough sets [J]. Knowledge Based Systems, 2003, 16(4): 205-213.
 [11] HU Q H, CHE X J, ZHANG L, et al. Rank entropy based decision trees for monotonic classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(11): 2052-2064.
 [12] 王长宝, 杨习贝, 王平心. 粗糙集中的近似距离与粗糙距离[J]. 江苏科技大学学报, 2017, 31(2):

- 196-199.
WANG Changbao, YANG Xibei, WANG Pingxin. Approximate and rough distances in rough set [J]. Journal of Jiangsu University of Science and Technology, 2017, 31(2): 196-199.
- [13] HU Q H, PEDRYCZ W, YU D R, et al. Selecting discrete and continuous features based on neighborhood decision error minimization [J]. IEEE Transactions on Systems Man and Cybernetics Part B, 2010, 40(1): 137-150.
- [14] 杨习贝, 徐苏平, 戚勇, 等. 基于多特征空间的粗糙数据分析方法[J]. 江苏科技大学学报, 2016, 30(4): 370-373.
YANG Xibei, XU Suping, QI Yong, et al. Rough data analysis method based on multi-feature space [J]. Journal of Jiangsu University of Science and Technology, 2016, 30(4): 370-373.
- [15] 王熙照, 王婷婷, 翟俊海. 基于样例选取的属性约简算法[J]. 计算机研究与发展, 2012, 49(11): 2305-2310.
WANG Xizhao, WANG Tingting, ZHAI Junhai. An attribute reduction algorithm based on instance selection [J]. Journal of Computer Research and Development, 2012, 49(11): 2305-2310.
- [16] 李京政, 杨习贝, 窦慧莉, 等. 重要度集成的属性约简方法研究[J]. 智能系统学报, 2017, DOI: 10.11992/tis.201706080.
LI Jingzheng, YANG Xibei, DOU Huili, et al. Research on ensemble significance based attribute reduction approach [J]. CAAI Transactions on Intelligent Systems, 2017, DOI: 10.11992/tis.201706080.
- [17] BEINRUCKER A, DOGAN U, BLANCHARD G. Early stopping for mutual information based feature selection [C]//International Conference on Pattern Recognition. [S.l.]: [s.n.], 2013: 975-978.
- [18] NG A Y. Feature selection, L1 vs. L2 regularization, and rotational invariance [C]//Proceedings of the Twenty-first International Conference on Machine Learning. [S.l.]: [s.n.], 2004: 78-87.
- [19] JIANG G X, WANG W J. Error estimation based on variance analysis of k -fold cross-validation [J]. Pattern Recognition, 2017, 69: 94-106.
- [20] JIANG G X, WANG W J. Markov cross-validation for time series model evaluations [J]. Information Sciences, 2017, 375: 219-233.
- [21] HU Q H, YU D R, XIE Z X. Neighborhood classifiers [J]. Expert Systems with Applications, 2008, 34(2): 866-876.

(编辑:夏道家)