

引用格式:施启军,潘峰,龙福海,等.特征选择方法研究综述[J].微电子学与计算机,2022,39(3):1-8. [SHI Q J, PAN F, LONG F H, et al. A review of feature selection methods[J]. Microelectronics & Computer, 2022, 39(3): 1-8.]DOI:10.19304/J. ISSN1000-7180. 2021. 1033

特征选择方法研究综述

施启军,潘 峰,龙福海,李娜娜,苟辉朋,苏浩轲,谢雨寒
(贵州民族大学 数据科学与信息工程学院,贵州 贵阳 550025)

摘 要:在大数据时代,特征选择是对数据进行预处理的必要环节.特征选择作为一种数据降维技术,其主要目的是从原始数据中选择出对算法最有益的相关特征,降低数据的维度和学习任务的难度,提升模型的效率.现阶段,有关特征选择算法方面的研究已取得阶段性成效,但也面临着重大挑战,其中维度灾难就是特征选择与分类问题所面临的重大挑战.首先,介绍了特征选择算法的基本架构,依次描述了子集的生成、子集的评估、终止条件、结果验证四个过程;其次,综述了特征选择领域的研究方法及研究成果,对特征选择方法分别依据评价策略、搜索策略、监督信息进行分类阐述,并对这些传统方法进行比较,指出它们的优势和不足;最后对特征选择进行了总结,并对其未来的研究方向进行了展望.

关键词:机器学习;特征选择;评价策略;搜索策略;监督信息

中图分类号: TP301

文献标识码: A

文章编号: 1000-7180(2022)03-0001-08

A review of feature selection methods

SHI Qijun, PAN Feng, LONG Fuhai, LI Nana,
GOU Huipeng, SU Haozhou, XIE Yuhan

(School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China)

Abstract: In the era of big data, feature selection is a necessary part of data preprocessing. Feature selection is a data dimensionality reduction technology. Its main purpose is to select the most beneficial relevant features for the algorithm from the original data, reduce the dimensionality of the data and the difficulty of learning tasks, and improve the efficiency of the model. At this stage, the research on feature selection algorithms has achieved initial results, but it is also facing major challenges. Among them, the disaster of dimensionality is the major challenge faced by the feature selection and classification problem. First, the basic architecture of the feature selection algorithm is introduced, and the four processes of the generation of the subset, the evaluation of the subset, the termination condition, and the result verification are described in sequence. Second, the methods are classified according to evaluation strategy, search strategy, and supervision information respectively, and compare these traditional methods, and point out their advantages and disadvantages. Finally, the feature selection is summarized and its future research direction is prospected.

Key words: machine learning; feature selection, evaluation strategy; search strategy; supervision information

1 引言

随着大数据时代的到来以及人工智能的不断发展,在模式识别、机器学习、数据挖掘等众多领域,都

需要对含有大量数据的特征集进行预处理.对特征数据进行降维操作不仅能降低数据的维度,还能提高算法的性能.特征选择的主要目的是通过特征选择算法从原始特征集中选择出评估标准最佳的特征

子集,以便帮助研究人员把任务更好地进行分类与回归,提高算法的泛化能力.特征选择的优势在于它可以压缩学习算法搜索的空间、减少特征集合数量的大小.除此之外,特征选择不仅能提高数据分类的精确度和效率,还可以提升对模型学习的速度.

传统的特征选择方法对于传统的数据特征具有很好的选择效果,随着以数学特征为基础的评估措施、搜索技术、统计学等多学科的发展,更多的算法工具被运用到特征选择中,并取得了可观的研究成果.处于该领域的学术研究者开发了许多实用且有效的特征选择算法,但并不是所有的方法都是完美无缺的,在存在优势的同时也存在着相应的缺点.统计方法的缺点就在于计算过程相对复杂,数据具有一定的复杂性,而且数据特征会由于维度的增加呈指数倍趋势增长.因此,统计方法在现实使用中仍存在一定难度且不太实用.传统的计算方法已经不能适应当今时代的发展变化,应着眼于现实情况,引入更多不同的工具来更新传统算法的观念.此外,特征子集需要具有全局搜索的能力,所以需要引入粒子群优化算法、遗传算法等进化算法来解决在搜索最优特征子集过程中出现的问题.

2 特征选择算法的基本结构

对于一个由 N 个特征组成的对象,可以产生 2^N 个特征子集,特征选择就是从所生成的特征子集中选择出对特定任务更有益的最佳特征子集.最佳特征子集不仅要取得最少的特征数目,还要使得训练出的分类器的分类性能最好.特征选择算法的基本结构如图 1 所示.

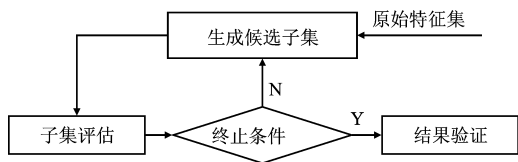


图 1 特征选择算法的基本结构图

Fig. 1 The basic structure diagram of feature selection algorithm

由图 1 可知,特征选择算法通常分为以下四个基本步骤:

(1)子集的生成:这一步骤是通过特定搜索策略搜索特征子集空间的过程,主要任务是为后续评价函数提供相应特征子集.所采用的搜索算法可以分

为三类,即全局搜索、启发式搜索以及随机搜索.

(2)子集评估:对子集的评估需要采取一系列的评价函数,其主要目的是对所选特征子集的好坏程度进行评价.可以把评估准则根据与学习算法的关联情况分成独立准则和关联准则两种.独立准则与学习算法是相互独立的,对所选特征子集的评价需要借助特征的内在特性,在以筛选器模型为基础的特征选择算法中有较多应用.用得比较多的独立准则有信息度量、距离度量、关联性度量以及一致性度量等.关联准则与学习算法是相互关联的,确定学习算法的同时利用学习算法的性能作为评价准则,在以封装器模型为基础的特征选择算法中用得比较多.这种准则的时间复杂度高且通用性不强,但所取得效果不容小觑.

信息度量:该度量准则的主要任务是计算特征的信息增益.信息增益的程度由信息熵的变化程度来度量,“熵”表示“一个系统的混乱程度”,熵越大,系统的不确定性越高.假设集合 D 中第 k 类样本所占比例为 p_k ,则 D 的信息熵为 $\text{Ent}(D)$,即

$$\text{Ent}(D) = - \sum_{k=1}^{|y|} p_k \log_2(p_k) \quad (1)$$

$\text{Ent}(D)$ 的值越小,则 D 的纯度越高.信息熵是对不确定性的一种描述,结合信息熵得出某个属性的信息增益,这里以计算属性 a 的信息增益为例,具体公式如式(2)所示.

$$\text{Gain}(D, a) = \text{Ent}(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} \text{Ent}(D^v) \quad (2)$$

距离度量:遵循同类样本距离小、异类样本距离大的原则.常用的有欧氏距离、马氏距离等.欧氏距离计算公式如式(3)所示.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

马氏距离计算公式如式(4)所示.

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^p \frac{(x_i - y_i)^2}{\sigma_i^2}} \quad (4)$$

式(4)中 σ_i 是 x_i 的标准差.

关联性度量:可以解释为通过一个已知变量去获取一个未知变量值的能力.在有监督特征选择算法中,着重考虑的是特征与类之间的关联性.而在无监督的特征选择方法中,主要考虑的是特征与特征之间的关联性.

一致性度量:该度量指标的主要目的是试图找到与全部数据特征集具有相同性能的最优特征子集.假如样本 1 与样本 2 不属于同一分类类别,但它们在特征 A 和 B 上的取值相同,特征子集{A,B}也不能选作为最终的特征子集.

(3)终止条件:与评价函数息息相关,指终止算法搜索过程所需要满足的具体要求.常用有执行时间、评价次数以及设置阈值等三种停止准则.

(4)结果验证:通过先验知识以及对验证集进行实验的结果来验证所选特征子集的有效性.

3 特征选择研究进展与评述

目前,对于特征选择方面的研究主要包括三类:基于特征子集评价策略的特征选择算法、基于搜索策略的特征选择算法、基于不同监督信息的特征选择方法.其中,每一类方法又有其不同的实现方法,如图 2 所示.

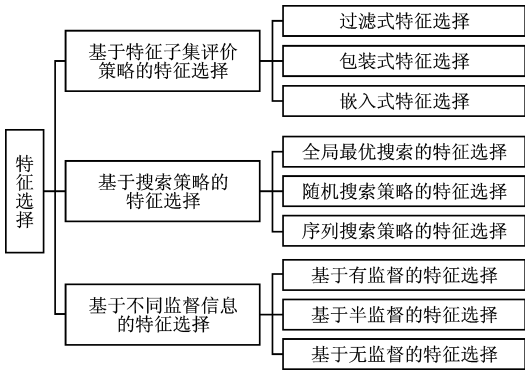


图 2 特征选择的方法
Fig. 2 Feature selection method

3.1 基于评价策略的特征选择

首先,Kira 等人给出了特征选择的定义,即在理想状况下寻找使得模型发挥效用的最小特征集合^[1].随后,Koller 等人再次给出特征选择的定义,即在特征与原始数据集分布相似的情况下,尽可能地找出较小的特征子集^[2].根据算法过程中子集评估准则和后续算法的结合方式可以把特征选择算法分为过滤式、包装式以及嵌入式三种类型.

(1)过滤式(Filter)特征选择

早期的特征选择算法大多属于过滤式特征选择方法,该方法在进行模型训练之前需要对数据集进行特征选择,两者之间是相互独立的关系.过滤式方法在评价所选特征的预测能力时,会借助一些基于

信息统计的启发式准则,所选特征子集根据评估准则的不同而不同,但是可以快速剔除噪声特征,其计算效率较高,通用性强.不一致性度量准则最先由 Almualliam 等人提出,随后他们提出一种被称为 FOCUS 的特征选择算法,该算法的计算复杂度相对较高且过于依赖搜索过程^[3]. Dash、Liu 等人提出了一种基于一致性的有效特征选择算法,该方法使用不一致度量去评估特征子集的优劣^[4]. Kwak 等人提出了一种基于 Parzen 窗口计算输入变量和类变量之间的互信息的新方法^[5]. Peng 等人提出一种两阶段的特征选择算法,该算法结合 mRMR(最小冗余最大相关准则)和其他更加复杂的特征选择器,在特征选择和分类准确性方面有了很大改进^[6]. Estévez 等人使用平均归一化互信息作为特征之间的冗余度的度量,提出了一种基于互信息的过滤式特征选择滤波方法,并将该方法与遗传算法相结合形成一种称为 GAMIFS 的混合过滤器方法^[7]. Brunato 等人提出了一种基于精确互信息的过滤式特征选择方法,他们的研究表明,在二分类和多分类任务中使用精确互信息识别的特征集进行实验具有显著的性能优势^[8]. 崔文岩等人将粒子群算法和支持向量机分类器引入到特征选择过程中,提出了一种基于粗糙集粒子群支持向量机的特征选择方法(RSPSO-SVM),该方法加速了迭代过程并缩短了筛选时间^[9]. 董红斌、滕旭阳等人提出了特征排序方法,并结合参数分析提出了一种基于关联信息熵度量的自适应特征子集选择方法,为之后的数据挖掘工作带来很好的借鉴意义^[10].

(2)包装式(Wrapper)特征选择

包装式(Wrapper)特征选择与过滤式方法有所不同,过滤式方法不需要考虑后续的学习器,而包装式是依赖于所选择的学习算法,特征子集的评价准则采用即将使用的学习器的性能.包装式方法最先由 Kohavi 等人^[11]提出.目前,有关包装式的方法成为特征选择领域的研究热点. Sindhvani 等人使用类标签和分类器之间的互信息作为目标函数,将多类支持向量机与多层感知器联系在一起,提出了一种基于最大输出信息的高效的包装式特征选择算法^[12]. Yang 等人提出了一种基于包装器的多层感知器(MLP)神经网络随机扰动的特征选择方法,该方法具有特别好的表示能力^[13]. 叶婷婷、刘明霞等

人将有效距离引入了特征选择,提出一种基于有效距离的多模态包装式特征选择方法,该方法对样例全局与局部的关系表现出较好的反映能力,从而对多模态数据的分类性能具有极大提升^[14]. Cao 等人提出了一种针对缺失数据进行分类的包装式特征选择方法,该方法从数据集中的缺失数据入手,对筛选丢失的数据进行分类操作,从而对分类器的测试精度有了很大提高. 包装式方法需要对学习器进行多次训练,计算开销相对过滤式方法较大^[15].

(3)嵌入式(Embedded)特征选择

嵌入式特征选择是将特征选择过程与学习器训练过程相互联系到一起,它是 Filter 和 Wrapper 组合式算法. 其基本思想是:首先用过滤式方法对特征进行预选,剔除冗余特征的同时降低数据集的维度,随后采用包装式方法从所预选出的特征子集中进行精选操作. 从根本意义上来说, ID3 算法^[16] 和 CART 算法^[17] 都属于嵌入式特征选择方法. Setiono 等人使用三层前馈神经网络来选择那些对

于区分给定输入模式集中对类最有用的输入属性,提出了一种基于前馈神经网络剪枝思想的特征选择算法^[18]. Shen 等人通过计算特征空间上具有和不具有该特征的 SVM 概率输出的绝对差值的合计值来评估特定特征的重要性,提出了一种基于 SVM 概率输出灵敏度分析的嵌入式特征选择算法^[19]. 周红标等人对多元序列预测建模过程中如何高效选择特征这一问题进行了深入研究,并将数据驱动型 KNN 用于对高维变量之间互信息的估计,提出了一种基于高维 KNN 互信息的特征选择算法,他们通过基于多层感知器神经网络预测模型的仿真实验验证了该方法的有效性^[20].

(4)三种基于评价策略方法的比较

前面章节概述了三种基于不同评价策略的特征选择方法,它们存在各自的优势与不足. 下面从使用效率、使用场景以及优缺点等几个方面对三种不同方法进行比较,如表 1 所示.

表 1 基于评价策略方法的比较

Tab. 1 Comparison of methods based on evaluation strategies

特征选择方法	过滤式方法	包装式方法	嵌入式方法
效率	较高	基于序列搜索的较高。基于随机搜索的较低。	较高
使用场景	适用于大规模数据集	不适合高维数据集	可处理高维数据集
优点	不依赖于特定分类器,算法的通用性强、复杂性低。特征子集冗余度较低。	特征子集分类性能通常更好,效率较高。	特征子集性能较好,效率较高。
缺点	算法的评估准则独立于特定的学习算法,所选的特征子集在分类精确率方面通常低于包装式方法。	通用性弱,计算复杂度很高。	依赖具体的学习算法,可能出现过拟合情况。

3.2 基于搜索策略的特征选择

除了根据评价策略对特征选择方法进行分类之外,还可以根据特征选择过程中所采用的搜索策略把特征选择算法进行分为全局最优搜索策略、随机搜索策略以及序列搜索策略的特征选择算法.

(1)基于全局最优搜索的特征选择

全局最优搜索指的是从原始特征集选择出最优的特征子集,也就是说所获得的特征子集能使给定的评估准则取得最优解,目前用得较多的有穷举法和分支定界法^[21]. 这两种方法只适用于低维度的特征集,时间复杂度随着特征维度的增加而增加.

(2)基于随机搜索策略的特征选择

这种方法需要先从原始特征集随机选取一个特征子集,紧接着有两种不同的处理方式,一种是将随机因素注入到序列搜索当中,将顺序性与随机性结合在一起,比如模拟退火算法^[22];另外一种是完全随机的,常被称为完全随机方法. 随机搜索方法不确定性强,每次选择的特征子集千差万别,但可防止算法陷入局部最优从而找到近似最优解. 经常用到的有粒子群优化算法(PSO)^[23]、蚁群算法(ACO)^[24]、遗传算法(GA)^[25]等.

(3)基于序列搜索策略的特征选择

序列搜索与全局搜索有所不同,不能保证最终获得的特征子集是最优的. 它由贪心爬山法演变而

来,可分为序列前向搜索(SFS)、序列后向搜索(SBS)以及双向搜索三类. 序列前向搜索是依次不停地选择单个特征加入原先为空集的特征子集当中,所选取的特征是使得评价函数取得最优值的特征,该过程是一个贪心算法的过程,只能不断添加特征而不能去除特征;序列后向搜索是不停地从特征全集中剔除相关性不大的特征,留到最后的特征能使评价函数取得最优解,其过程也属于贪心算法,只能不断筛选剔除相关性不大的特征而不能加入新特

征,其操作过程与序列前向搜索刚好相反;双向搜索是前向与后向搜索的结合搜索方式,既可增加特征,也可删减特征.

(4)三种搜索策略方法的比较

前面概述了基于三种不同搜索策略特征选择方法,它们存在各自的优势与不足. 下面从算法的使用效率、适应场景以及优缺点等几个方面对三种不同搜索策略的特征选择方法进行比较,如表 2 所示.

表 2 三种搜索策略方法的比较

Tab. 2 Comparison of three search strategy methods

搜索策略方式	全局搜索策略	随机搜索策略	序列搜索策略
效率	低	较高	高
使用场景	仅适用于低维度特征集	均适用	均适用
优点	可实现全局最优	时间复杂度比全局全局搜索低,可获得优于序列搜索的近似最优解。	时间复杂度最低
缺点	只在低维数据集中有使用价值	时间复杂度比序列搜索高	获得的特征子集仅是局部最优

3.3 基于不同监督信息的特征选择

特征选择过程中,监督信息扮演着重要角色. 根据监督信息的不同,可以把特征选择方法分为有监督特征选择、半监督特征选择和无监督特征选择^[26].

(1)有监督特征选择

有监督特征选择主要依赖于样本的类别信息,需要计算特征与类别的关系来选取最优特征子集. 目前,基于监督信息的特征选择方法大多都集中在监督特征选择上,通过测量特征与类别之间以及特征与特征之间的关系来确定特征子集. 在学术界,取得了很多有关监督特征选择的研究成果,例如 Wu X 等人提出一种被称为 Top- k 的正则化方法,用于回归和分类任务中的监督特征选择^[27].

(2)半监督特征选择

由于监督特征选择过程中获取足够的标记样本需要巨大成本,所以一些研究人员将注意力转移到了半监督特征选择上. 半监督特征选择方法主要考虑如何利用少量标记(具有类别信息)样本和大量未标记(不带类别信息)样本进行分类学习的问题,具有“小标签”和“高效”的突出成就,一直处于降维研究的前沿. Wang C 等人将用于监督任务的 ϵ -dragging 技术拓展到半监督任务中,提出了一种具有稀疏判别最小二乘回归的半监督特征选择,并在真实

数据集上验证了方法的优越性^[28]. Wu X 等人针对基于图的方法没有同时考虑异常值、噪声和所选特征冗余等多个因素的问题,提出了一种基于局部自适应和最小冗余的新型半监督特征选择方法,该方法根据数据情况灵活分配权重,减少离群点和噪声的影响,并且在特征映射矩阵中引入了高相似度惩罚机制,以提高区分度和低冗余度^[29].

(3)无监督特征选择

现有方法通常需要许多实例来进行特征选择. 然而,在实践中经常无法获得足够的实例. 无监督特征选择是一种用于降维的数据预处理技术,该方法不需要使用任何标签信息,直接对特征空间的样本进行聚类或无监督学习,以便对特征进行分组,并对特征重要性进行评估,根据特征的重要性程度进行特征选择. Kumagai A 等人使用具体的随机变量通过梯度下降来选择特征,提出一种基于少样本学习的无监督特征选择方法,他们通过实验证明了该方法优于现有的特征选择方法^[30].

(4)基于不同监督信息方法的比较

基于不同监督信息的特征选择方法各有各的优缺点,前面已经结合一些现有科研成果对有监督、半监督、无监督三种特征选择方法进行简要阐述. 下面对三种特征选择方法做一个比较,如表 3 所示.

表 3 基于不同监督信息特征选择方法的比较

Tab. 3 Comparison of Feature SelectionMethods Based on Different Supervision Information

特征选择方法	有监督特征选择方法	半监督特征选择方法	无监督特征选择方法
优点	不依赖于样本空间的分布	只需获取少量具有标签信息的样本,有效减少标注代价。	不适用任何标签信息,运算成本大幅度降低。
缺点	需要获得大量的具有标签信息的样本,且新应用的特征无法提前预测;运算成本高;无法对未知数据进行处理。	对有噪声干扰数据的处理效果不理想	训练样本的歧义性高

4 特征选择总结与展望

特征选择是机器学习领域的经典问题. 随着大数据时代的不断发展,新型应用层出不穷,所产生的数据呈指数倍增长,如何对庞大数据集进行特征选择一直是科研工作者所关注的热门话题之一. 首先,本文对特征选择算法的基本框架进行介绍,重点描述了特征选择的四个基本步骤,即子集的生成、子集的评估、终止条件以及结果验证;随后对特征选择领域的研究方法及研究成果进行阐述,从评估准则、搜索策略以及监督信息三个方面对特征选择算法进行分类,并对各种方法的子方法进行比较,分别指出它们的优势和不足.

现有的特征选择算法种类繁多,但缺陷依然不可避免,研究对象的日益复杂使得现有算法性能不佳,如何针对实际问题设计配套方案亟待解决,通过对特征选择算法的汇总分析,下面对特征选择所面临的一些挑战和技术手段进行分析展望.

如何对高维小样本数据进行高质量特征选择是特征选择领域最具挑战性的问题之一,目前的主要研究思路是使用多阶段特征选择算法,无法得到令人满意的特征选择结果. 基于评价策略的嵌入式方法和基于搜索策略的序列搜索策略方法,辅以重采样、增量式学习等技术手段,可能是较好的研究思路之一.

设计适应大数据环境下的特征选择算法具有很大的研究空间,而且应用价值大. 目前的一些工作利用深度神经网络技术来提高算法的鲁棒性和稳定性. 在高维大数据环境下,如何将深度学习技术与群体智能算法相结合并运用到特征选择领域,值得进一步探索.

参考文献:

[1] KIRA K, RENDELL L A. The feature selection problem: traditional methods and a new algorithm[C]//

Proceedings of the Tenth National Conference on Artificial Intelligence. San Jose, California: AAAI Press, 1992.

[2] KOLLER D, SAHAMI M. Toward optimal feature selection[C]//Proceedings of the Thirteenth International Conference on International Conference on Machine Learning. Bari Italy, 1996.

[3] ALMUALLIM H, DIETTERICH T G. Learning Boolean concepts in the presence of many irrelevant features[J]. Artificial Intelligence, 1994, 69 (1-2): 279-305. DOI: 10.1016/0004-3702(94)90084-1.

[4] DASH M, LIU H. Consistency-based search in feature selection[J]. Artificial Intelligence, 2003, 151(1-2): 155-176. DOI: 10.1016/S0004-3702(03)00079-1.

[5] 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. DOI: 10.1109/TPAMI.2002.1114861.

[6] PENG H C, LONG F H, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27 (8): 1226-1238. DOI: 10.1109/TPAMI.2005.159.

[7] 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. DOI: 10.1109/TNN.2008.2005601.

[8] BRUNATO M, BATTITI R. X-MIFS: exact mutual information for feature selection[C]//Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver: IEEE, 2016. DOI: 10.1109/IJCNN.2016.7727644.

[9] 崔文岩, 孟相如, 李纪真, 等. 基于粗糙集粒子群支持向量机的特征选择方法[J]. 微电子学与计算机, 2015, 32(1): 120-123.

CUI W Y, MENG X R, LI J Z, et al. Feature selec-

- tion based on RS-PSO-SVM[J]. *Microelectronics & Computer*, 2015, 32(1): 120-123.
- [10] 董红斌, 滕旭阳, 杨雪. 一种基于关联信息熵度量的特征选择方法[J]. *计算机研究与发展*, 2016, 53(8): 1684-1695. DOI: 10.7544/issn1000-1239.2016.20160172.
- DONG H B, TENG X Y, YANG X. Feature selection based on the measurement of correlation information entropy[J]. *Journal of Computer Research and Development*, 2016, 53(8): 1684-1695. DOI: 10.7544/issn1000-1239.2016.20160172.
- [11] KOHAVI R, JOHN G H. Wrappers for feature subset selection[J]. *Artificial Intelligence*, 1997, 97(1-2): 273-324. DOI: 10.1016/S0004-3702(97)00043-X.
- [12] 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): 937-948. DOI: 10.1109/TNN.2004.828772.
- [13] YANG J B, SHEN K Q, ONG C J, et al. Feature selection for MLP neural network: the use of random permutation of probabilistic outputs[J]. *IEEE Transactions on Neural Networks*, 2009, 20(12): 1911-1922. DOI: 10.1109/TNN.2009.2032543.
- [14] 叶婷婷, 刘明霞, 张道强. 基于有效距离的多模态特征选择[J]. *模式识别与人工智能*, 2016, 29(7): 658-664. DOI: 10.16451/j.cnki.issn1003-6059.201607010.
- YE T T, LIU M X, ZHANG D Q. Effective distance based multi-modality feature selection[J]. *Pattern Recognition and Artificial Intelligence*, 2016, 29(7): 658-664. DOI: 10.16451/j.cnki.issn1003-6059.201607010.
- [15] CAO T T, ZHANG M J, ANDREAE P, et al. A wrapper feature selection approach to classification with missing data[C]//*Proceedings of the 19th European Conference on the Applications of Evolutionary Computation*. Springer International Publishing. Porto, Portugal: Springer, 2016. DOI: 10.1007/978-3-319-31204-0_44.
- [16] QUINLAN J R. Introduction of decision trees[J]. *Machine Learning*, 1986, 1(1): 81-106.
- [17] BREIMAN L, FRIEDMAN J H, OLSHEN R A, et al. Classification and regression trees[J]. *Biometrics*, 1984, 40(3): 874. DOI: 10.2307/2530946.
- [18] SETIONO R, LIU H. Neural-network feature selector [J]. *IEEE Transactions on Neural Networks*, 1997, 8(3): 654-662. DOI: 10.1109/72.572104.
- [19] SHEN K Q, ONG C J, LI X P, et al. Feature selection via sensitivity analysis of SVM probabilistic outputs[J]. *Machine Learning*, 2008, 70(1): 1-20. DOI: 10.1007/s10994-007-5025-7.
- [20] 周红标, 乔俊飞. 基于高维 k-近邻互信息的特征选择方法[J]. *智能系统学报*, 2017, 12(5): 595-600. DOI: 10.11992/tis.201609020.
- ZHOU H B, QIAO J F. Feature selection method based on high dimensional k-nearest neighbors mutual information[J]. *CAAI Transactions on Intelligent Systems*, 2017, 12(5): 595-600. DOI: 10.11992/tis.201609020.
- [21] SOMOL P, PUDIL P, KITTLER J. Fast branch & bound algorithms for optimal feature selection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(7): 900-912. DOI: 10.1109/TPAMI.2004.28.
- [22] 张永波, 游录金, 陈杰新. 基于模拟退火的多标记数据特征选择[J]. *计算机工程与设计*, 2011, 32(7): 2494-2496. DOI: 10.16208/j.issn1000-7024.2011.07.084.
- ZHANG Y B, YOU L J, CHEN J X. Feature selection for multi-label data by using simulated annealing [J]. *Computer Engineering and Design*, 2011, 32(7): 2494-2496. DOI: 10.16208/j.issn1000-7024.2011.07.084.
- [23] 张翠军, 陈贝贝, 周冲, 等. 基于多目标骨架粒子群优化的特征选择算法[J]. *计算机应用*, 2018, 38(11): 3156-3160. DOI: 10.11772/j.issn.1001-9081.2018041358.
- ZHANG C J, CHEN B B, ZHOU C, et al. Feature selection algorithm based on multi-objective bare-bones particle swarm optimization[J]. *Journal of Computer Applications*, 2018, 38(11): 3156-3160. DOI: 10.11772/j.issn.1001-9081.2018041358.
- [24] 叶志伟, 郑肇葆, 万幼川, 等. 基于蚁群优化的特征选择新方法[J]. *武汉大学学报·信息科学版*, 2007, 32(12): 1127-1130. DOI: 10.3969/j.issn.1671-8860.2007.12.010.
- YE Z W, ZHENG Z B, WAN Y C, et al. A novel approach for feature selection based on ant colony optimization algorithm[J]. *Geomatics and Information Science of Wunan University*, 2007, 32(12): 1127-1130. DOI: 10.3969/j.issn.1671-8860.2007.12.010.

[25] WUTZL B, LEIBNITZ K, RATTAY F, et al. Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness [J]. PLoS One, 2019, 14(7): e0219683. DOI: 10.1371/journal.pone.0219683.

[26] LIU H, YU L. Toward integrating feature selection algorithms for classification and clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17 (4): 491-502. DOI: 10.1109/TKDE.2005.66.

[27] WU X X, CHENG Q. Top-k regularization for supervised feature selection[J]. arXiv:2106.02197, 2021.

[28] WANG C, CHEN X J, YUAN G W, et al. Semisupervised feature selection with sparse discriminative least squares regression[J]. IEEE Transactions on Cybernetics, 2021: 1-12. DOI: 10.1109/TCYB.2021.3060804.

[29] WU X P, CHEN H M, LI T R, et al. Semi-supervised feature selection with minimal redundancy based on local adaptive[J]. Applied Intelligence, 2021, 51 (11): 8542-8563. DOI: 10.1007/s10489-021-

02288-4.

[30] KUMAGAI A, IWATA T, FUJIWARA Y. Few-shot learning for unsupervised feature selection[J]. arXiv:2107.00816, 2021.

作者简介:

施启军 男,(1993-),硕士研究生.研究方向为智能计算、大数据分析 & 处理、机器学习.

潘 峰(通讯作者) 男,(1974-),博士,教授,硕士生导师.研究方向为大数据与机器学习、软件技术.

E-mail:panf@gzmu.edu.cn

龙福海 男,(1996-),硕士研究生.研究方向为智能计算、机器学习.

李娜娜 女,(1994-),硕士研究生.研究方向为优化理论、智能算法.

苟辉朋 男,(1996-),硕士研究生.研究方向为智能计算、机器学习.

苏浩轲 男,(1997-),硕士研究生.研究方向为智能算法、图像识别.

谢雨寒 女,(1997-),硕士研究生.研究方向为智能算法.