

特征选择算法研究综述

梁伍七, 王荣华, 刘克礼, 李 斌

(安徽广播电视大学 信息与工程学院, 合肥 230022)

摘要:模式识别领域中,特征选择作为预处理模块的关键步骤,特征选择函数用来降低特征空间的维数,提高分类器的分类性能。首先分析了特征选择的主要过程,从不同视角探讨了特征选择的分类方法,然后分析了基于 SVM 进行特征选择的作用,对基于 SVM 的特征选择算法进行了归纳总结,最后分析对比了基于 SVM 的三类特征选择的优缺点,指出理论研究和实际应用中的研究热点和应用发展方向。

关键词:模式识别;文本分类;特征选择;支持向量机;Wrapper 方法

中图分类号:TP391

文献标志码:A

文章编号:1008-6021(2019)04-0085-07

一、引言

特征选择在文本分类、文本检索、基因分析和药物诊断等场合有广泛应用,是模式识别领域的研究热点之一。例如,自动文本分类是指按照给定的分类体系,依据文本的内容自动进行文本所属类别判别的过 程,是一种有监督的学习过程。自动文本分类在信息过滤、信息检索、搜索引擎和数字图书馆等领域有广泛应用。分类系统主要包括数据预处理、文档分词、特征表示、特征选择、文本表示、分类器选择和训练以及分类结果评价等过程。1975 年,文献[1]提出向量空间模型(Vector Space Model, VSM),文档被表示成特征空间中的一个向量。文本分类中文本表示方法通常使用向量空间模型,采用词干抽取和去停用词处理后,特征词向量空间的维数虽有所降低,但特征空间的维数仍然是不可接受的。对于分类器来说,高维特征空间既增加了分类的时间复杂度和空间复杂度,也影响分类精度。

文本分类系统中高维特征空间的降维通常有两种方法,即特征选择(feature selection)和特征抽取(feature extraction)。二者都是在分类之前,针对原始特征的不足,降低特征维数,提高分类器的分类性能。特征抽取也称作特征重参数化(feature re-parameterization),通过对原始特征进行组合或者

变换,新的低维空间是原来特征的一个映射。特征选择作为文本分类预处理模块的关键步骤,任务是从原始特征空间中选择最重要的特征组成特征子集,从而实现特征空间降维。

二、特征选择分类

特征选择的过程包括四个主要环节^[2-3],包括子集生成(subset generation)、子集评估(subset evaluation)、停止准则(stopping criterion)和结果验证(result validation),如图 1 所示。生成子集的过程是一个搜索过程,根据特定的搜索策略得到候选特征子集。对每个候选子集,根据评价准则进行评价。若新子集的评价结果优于以前最好的子集,则将其更新为当前最优子集。子集生成和子集评价不断循环,直至满足给定的停止准则,最后对最优特征子集进行结果验证。特征选择在数据挖掘、模式识别和机器学习等多个领域内均得到了广泛的研究。根据不同的标准,特征选择分类方法也有所不同,典型的分类标准及其分类方法如下。

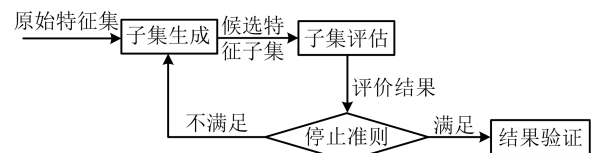


图 1 特征选择的四个主要环节

收稿日期:2019-06-12

基金项目:安徽省高校自然科学基金项目“中文文本分类特征选择和分类算法研究与实现”(项目编号:KJ2016A111)

作者简介:梁伍七(1969—),男,安徽怀宁人,副教授,硕士。研究方向:网络信息安全和数据挖掘。

(一) 基于搜索策略进行分类

搜索过程需要考虑两个基本问题^[3-4]:一是确定搜索起点。可以从一个空的集合开始,然后不断添加特征,也可以从一个完整的集合开始,然后不断移除特征,或者从两端开始,然后同时加入和移除特征;二是确定搜索策略。根据搜索策略的不同,特征选择可分为完全搜索(complete search)、随机搜索(random search)和启发式搜索(heuristic search)等。

完全搜索方法:完全搜索需要遍历特征空间中所有可能的特征组合,它能够找到性能最好的子集。对于具有 n 个特征的数据集,存在 2^n 个候选子集,这个搜索空间是指数级的,可以使用分支定界(branch and bound)算法^[5]等启发式方法来缩小搜索空间。

随机搜索方法:该方法首先随机选择特征子集,后续采用两种方式进行:第一种在传统的序列搜索中注入随机因素,称为概率随机方法,例如,随机开始爬山算法(random-start-hill-climbing)和模拟退火算法(simulated annealing)^[6]等;另一种称为完全随机方法,候选子集的产生方式是完全随机的,例如,Las Vegas 算法^[7]等。这些方法中,随机过程有助于避免在搜索空间中陷入局部最优,但其能否搜索到最优结果取决于可用的资源。

启发式搜索方法:启发式搜索可以消除组合爆炸,根据特定的启发式规则设计次优搜索策略,利用启发信息来引导搜索,可以得到近似最优解。常用的算法包括^[3,8]序列前向选择(SFS, sequential forward selection)、序列后向选择(SBS, sequential backward selection)和双向选择(BDSS, bidirectional selection)等。该方法运算速度较快,对于包含 n 个特征的集合,搜索时间往往低于 $O(n^2)$ 。

(二) 根据评价准则进行分类

每一个生成的候选子集都必须使用一个评价准则来进行评价。根据评价准则是否独立于学习算法,可以分为过滤式(Filter)、封装式(Wrapper)和混合式(Hybrid)三种类型。

(1) 过滤式^[4]:过滤式特征选择作为一种数据预处理过程,评价准则和学习算法无关,可以快速排除不相关的噪声特征,计算效率高。基于过滤式的评价函数包括距离度量(distance measure)^[9-10]、信息度量(information measure)^[11]、相关性度量(correlation measure)^[12]和一致性度量(consistency measure)^[8]

等,评价准则不同得到的结果子集也不同。

(2) 封装式^[13]:封装式和所使用的学习算法有关,特征选择算法作为学习算法的组成部分,将学习算法的性能作为衡量特征优劣的标准。在特征选择过程中直接用所选特征来训练分类器,根据分类器在验证集上的性能作为特征重要程度的评价标准,封装式能选出更适合特定学习算法的特征。对每个候选子集,分类器都需要重新训练,故该方法在速度上比过滤式要慢,优点是所选的优化特征子集的规模相对要小很多。基于启发式搜索策略的封装器方法是目前较实用的特征选择研究热点。

(3) 混合式^[14]:综合利用过滤式与封装式的优势,提出混合式方法来处理大规模的数据集。最理想的情况是和过滤式的时间复杂度相近,和封装式的算法性能相似。混合式方法的处理过程是,先使用过滤式基于数据集本身固有的特性快速进行特征选择,保留少量特征,减少进一步搜索的特征规模,然后再用封装式方法进一步优化,得到分类性能最优的特征子集。

(三) 根据不同的监督信息进行分类

监督信息是文本内容的重要组成部分,文本分类系统中,通常将类别信息作为监督信息。特征选择过程中,监督信息起着重要的指导作用。可以基于不同的监督信息,对特征选择方法进行分类。

(1) 根据训练集中给定类别标签样本和未给定类别标签样本的相对数量,可分为有监督特征选择(supervised feature selection)、半监督特征选择(semi-supervised feature selection)和无监督特征选择(unsupervised feature selection)^[3-4],反映了选择特征过程中对类别信息的依赖程度。

监督式特征选择:利用类别信息进行指导,通过计算特征与类别之间的关系,选择最具类别区分力的特征子集^{[2]131-156}。关于特征选择的研究最初大多聚焦于监督式特征选择,通过度量特征之间和特征与类别之间的相互关系来确定特征子集。

半监督式特征选择:半监督式学习主要考虑如何利用少量的具有类别信息的样本和大量的不带类别信息的样本进行分类学习的问题。Miller 等人提出了一种分类器架构和学习算法,算法可以有效利用未标记数据提高学习算法的性能^[15]。随着半监督学习的发展,半监督式特征选择的研究也越来越受到

重视。

无监督式特征选择: 没有类别信息指导, 通过对特征空间的样本进行聚类或无监督学习对特征进行分组, 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。文献[16]考虑了无监督学习的特征选择问题, 提出了一种新的算法, 能够识别嵌入在高维空间中支持复杂结构的信息特征, 算法通过一个目标函数将其表示为一个优化问题, 并用迭代法求解。

(2) 根据类别数目, 可分为二元特征选择(binary feature selection)和多类特征选择(multi-class feature selection)^[3]。当某个样本数据可以同时属于多个类别, 称为多类特征选择, 也称为多标签特征选择(multi-label feature selection)。根据类别的组织方式, 多类问题可以分为平铺结构和层次化结构。平铺结构的各个类别间关系是平等的, 若类别间的关系不是独立的, 而是具有某种复杂的关系, 可利用层次化特征选择进行处理^[17]。

三、基于 SVM 的特征选择

1963年, 文献[18]在解决模式识别问题时提出了基于支持向量机(support vector machine, SVM)方法。1995年, 文献[19]正式提出统计学习理论, 并提出用广义分类面来解决线性不可分问题, 据此构成了SVM的理论基础。SVM已被证明是具有最小化分类误差和最大化泛化能力的强有力的分类工具, 它建立在统计学习理论的VC维理论和最小化经验风险以及结构风险的线性组合原理基础上^[20-21]。但SVM是作为模式识别领域中逐步发展而来的分类工具, 标准的SVM并不能进行特征选择。文献[22]基于梯度最小化泛化边界方法来减少特征, 指出当存在不相关的特征时, 标准SVM分类的性能会大大降低。这一结论导致了众多学者对基于SVM的特征选择算法进行研究。

根据评价准则是否和学习算法相关, 特征选择可以分为过滤式(Filter)、封装式(Wrapper)和混合式(hybrid)三类^[3]。基于SVM的特征选择算法, 本质就是在特征选择过程中融入SVM学习算法, 因此基于SVM的特征选择算法也可以划分成三类: 基于SVM的Wrapper特征选择、基于SVM的Embedded特征选择和基于SVM的混合特征选择。

(一) 基于 SVM 的 Wrapper 特征选择

基于SVM的Wrapper特征选择利用SVM分类器的性能作为特征选择的评价准则。文献[23]提出了一种基于SVM的Wrapper特征选择算法, 即SVM-RFE(SVM recursive feature elimination)特征选择算法。SVM-RFE的算法流程是^[21]: 利用当前数据集训练SVM分类器, 得到分类器的参数; 计算权重向量; 根据特征排序准则, 计算所有特征的排序准则得分; 移除得分最小的特征; 上述过程多次迭代直至特征集中剩余最后一个特征。SVM-RFE方法是一个序列后向选择的过程, 目标是在 d 个特征中找出大小为 r 的特征子集, 使得SVM分类的学习性能最优。该方法首先训练SVM分类器, 得到分类器的参数, 利用参数信息对特征进行递归移除, 在特征选择的过程中存在参数不确定问题。

文献[24]针对SVM-RFE方法的不足提出了改进算法。在SVM-RFE方法的基础上, 由支持向量机理论的泛化误差界推导出评分准则, 每次迭代过程中移除最小得分的特征; 文献[25]提出了增强的SVM-RFE特征选择算法, 采用最小冗余和最大相关MRMR作为评分准则。与SVM-RFE算法相比, 在多数基因数据集上, 算法选择的特征个数较少。针对SVM-RFE算法中SVM参数难以确定的问题, 文献[26]采用粒子群算法搜索SVM的参数。

和SVM算法一样, SVM-RFE最初设计是用来解决两类基因选择问题的。文献[27]针对不同的多类SVM框架, 推广了SVM-RFE算法, 用来解决多类基因特征选择问题。文献[28]针对多类分类问题, 使用标准两类SVM-RFE算法来排序问题, 该算法能够较好地解决多类问题。

SVM-RFE是一种简单有效的特征选择算法, 已在许多领域得到应用, 基于SVM-RFE理论框架的特征选择方法得到众多研究者的关注, 算法以SVM分类器的性能作为特征重要性的评价准则, 优点是所选的特征子集的规模相对较小, 缺点是算法的时间复杂度较高。文献[29]提出了一种基于SVM的Wrapper特征选择算法, 算法采用序列后向选择, 特征排序准则使用验证子集的错分样本个数, 每一轮迭代过程中, 将特征所引起的错分样本个数最少的那个特征移除, 最后得到最优的特征子集。

(二)基于 SVM 的 Embedded 特征选择

与一般的 Embedded 方法类似,基于 SVM 的 Embedded 特征选择,特征选择过程融于学习过程中,但该类方法的学习算法依据 SVM 理论^[21]。

文献[30]通过构造非线性 SVM,提出了一种新的特征选择算法 RFSVM。该算法的基本思想:在 SVM 思想的基础上,通过一个正的参数 σ 来加权特征抑制项 $e^T E e$, 该问题可以转化为一个混合整数规划求解问题。算法的目标函数为:

$$\min_{u, \gamma, y, s, E} v e^T y + e^T s + \sigma e^T E e$$

$$\text{s. t. } \begin{cases} D(K(AE, EA^T)u - e\gamma) + y \geq e \\ -s \leq u \leq s \\ y \geq 0 \\ E = \text{diag}(0 \text{ or } 1) \end{cases} \quad (1)$$

其中, A 表示数据集矩阵, K 表示核函数, v 为正常数, u, s 和 γ 为分类器的待求量, e 表示全 1 的向量, y 表示正的变量, E 表示对角元素为 0 或 1 的对角矩阵。(1)式是一个混合整数规划问题,混合整数规划问题是一个 NP 难问题。可以通过固定 E , 将(1)式变为一个线性规划问题,通过求解得到 (u, γ, y, s) 的解,计算目标函数,上述过程多次迭代,直到函数值小于某一设定的阈值。该算法收敛于局部最小值,最终选择的特征个数最少。

针对多类别分类特征选择的问题,文献[31]提出了一种基于 SVM 的 Embedded 方法 L1MSVM,该算法通过加入 l_1 范数,修改标准 SVM 的目标函数,为自适应特征选择计算整个正则化解决方案路径。该算法的目标函数为:

$$\min_{w_c, b_c, c=1, \dots, k} \sum_{i=1}^n L(f, z_i) + \lambda \sum_{c=1}^k \|w_c\|_1$$

$$\text{s. t. } \sum_{c=1}^k w_c = 0, \sum_{c=1}^k b_c = 0 \quad (2)$$

其中, λ 是非负的正则化参数, $L(f, z_i) = \sum_{c \neq y_i} [f_c(x_i) + 1]_+$ 是损失函数, $f_c(x_i)$ 是决策函数, $f_c(x) = w_c^T x + b_c, c = 1, \dots, k, x_i$ 和 y_i 分别表示第 i 个样本及其类别且 $y_i \in \{+1, -1\}$ 。该算法的特征选择与模型训练同时进行,在选择出最优的特征子集的同时,还得到了线性分类函数。

l_1 范数 SVM 是标准 l_2 范数 SVM 的变体,在处理高维问题和冗余噪声特征时, l_1 范数 SVM 比 l_2 范数 SVM 有一些优势,但 l_1 范数 SVM 不适合处理存

在强相关特征组合的场合。为了较好地解决这一问题,文献[32]提出特征选择算法 DrSVM(doubly regularized support vector machine)。该算法混合了 l_2 范数和 l_1 范数,在标准 SVM 的目标函数加入 l_1 范数,算法的优点是能够同时移除或选择一组强相关的重要特征。DrSVM 算法等价于解决下面的问题:

$$\min_{\beta_0, \beta} \sum_{i=1}^n [1 - y_i (\beta_0 + \beta x_i^T)]_+ + \frac{\lambda_2}{2} \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \quad (3)$$

其中, λ_1 和 λ_2 均是调整参数, $[1 - z]_+ = \max(1 - z, 0)$, (x_i, y_i) 表示训练数据, y_i 表示第 i 个样本的类别, $y_i \in \{+1, -1\}$ 。 l_1 范数的作用是来进行特征选择,而 l_2 范数的作用是同时选择一组强相关的特征。文献[32]研究表明, l_2 范数倾向于产生强相关的特征,且这些特征对应的系数几乎相等,称这种现象为分组效应。

DrSVM 算法是基于 SVM 的 Embedded 特征选择算法,和基于 SVM 的 Embedded 特征选择算法 l_1 -SVM 相比,二者相同点是特征选择过程和学习过程融于一体,在学习过程中自动地进行特征选择;不同点是 DrSVM 算法能够同时选择或舍弃一组强相关的特征,而 l_1 -SVM 不考虑特征间的相关性,只能从强相关的特征中选择一个^[21]。DrSVM 算法适用于维数高于训练样本数的场合,而 l_1 -SVM 选择的特征数目不超过训练样本数。

基于 SVM 的 Embedded 特征选择算法,特征选择过程作为组成部分嵌入到学习算法里,这类算法效率较高,得到的学习算法有较好的性能。但如何基于标准 SVM 算法,来构造算法的目标函数是目前算法研究的热点。

(三)基于 SVM 的混合特征选择算法

基于 SVM 的混合特征选择先使用 Filter 算法快速进行特征选择,然后利用 Wrapper 特征选择算法进一步细化,得到更为有效的特征子集。

文献[33]提出基于 SVM 的混合特征选择算法 FS_SFS(Filtered and Supported Sequential forward search),和传统的采用序列前向选择的 Wrapper 方法比较,该算法有两个重要的特性来减少计算时间。该算法首先使用 Filter 算法对原始特征进行预处理,然后对预处理后的特征子集,利用 Wrapper 算法进

一步细化。该算法提出了一种新的评分准则,该准则既考虑了单个特征的区分能力,又考虑了特征之间的相关性,从而有效地过滤非本质特征。

文献[34]提出的基于 SVM 的混合特征选择算法 F_SSFS(F-score and Supported Sequential Forward Search),该算法结合了 F-score 和序列前向选择,结合了 Filter 方法和 Wrapper 方法各自的优势来选择最优特征子集,该算法可应用于股票市场预测。考虑到 Filter 算法和 Wrapper 算法各自的优缺点,即 Filter 方法的计算成本低,但分类可靠性不足,而 Wrapper 方法具有较高的分类精度,但需要很大的计算能力,文献[35]将二者整合成一个序列搜索算法,用以提高所选择特征对于分类算法的性能。该算法添加预选择步骤,以提高特征选择的效率,利用 ROC(receiver operating characteristic) 曲线作为搜索策略,利用 SVM 作为分类器,算法在生物数据分类上得到了很好的应用。

文献[36]提出了一种新的基于队列智能算法的特征选择与支持向量机(SVM)模型选择混合方法 SVM_SACI。方法将自适应队列智能(SACI)算法与 SVM 集成,形成了一种新的混合方法,用于同时进行特征选择和 SVM 模型选择,在多个数据集上的检验结果表明,SACI 在支持向量机分类精度和降维方面优于其他启发式方法。针对样本量远小于数据集特征个数的小样本可能带来的奇异性问题,文献[37]提出了一种 $0 < Q < 1$ 的稀疏 L Q-范数最小二乘支持向量机(L Q-范数 LS-SVM),它同时进行特征选择和预测。与传统的 LS-SVM 不同,L Q-范数 LS-SVM 最大限度地减少了 L Q-范数的权重,从而有效地实现

了特征选择。

文献[38]提出了一种基于蝗虫优化算法(GOA)和 SVM 的混合方法,该方法对支持向量机模型的参数进行优化,同时找到最佳特征子集。在多个低维和高维数据集上实验结果表明,该方法在分类精度上优于其他方法,同时最小化了所选特征的数目。文献[39]提出了一种基于 SVM 方法的特征选择分类问题的 MILP 模型,利用精确算法和启发式算法,分析了该模型的不同求解方法。通过在多个数据集的检验和经典分类方法的比较,对模型进行了验证。文献[40]提出了一种新颖的进化算法(lion 算法)和 SVM 的混合方法,利用 lion 算法选择高维数据集的特征子集,解决分类问题。特征选择过程识别并删除无关/冗余特征,以减少特征维数,从而提高分类的效率和准确性。

四、结语

本文阐述了自动文本分类中特征选择的算法框架,探讨了特征选择的分类方法,介绍了 SVM 用于特征选择的意义,对基于 SVM 的特征选择算法进行了归纳总结,重点总结了基于 SVM 的 Wrapper 特征选择、Embedded 特征选择以及混合特征选择算法,分析对比了基于 SVM 的三类特征选择的优缺点。SVM 算法的效率和性能主要取决于内核类型及其参数,在处理高维数据集时,用于 SVM 模型的特征子集选择是影响分类精度的另一个重要因素。如何利用已有的算法或提出新颖的算法用于高维数据集的特征选择并训练 SVM 模型的参数,是基于 SVM 特征选择的研究热点和分类应用发展方向。

参考文献:

- [1] SALTON G, WONG A, YANG C S. A Vector Space Model for Automatic Indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.
- [2] DASH M, LIU H. Feature Selection Methods for Classifications[J]. Intelligent Data Analysis: An International Journal, 1997, 1(3): 131-156.
- [3] 王博. 文本分类中特征选择技术的研究[D]. 长沙:国防科学技术大学, 2009: 24-27.
- [4] 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.
- [5] NARENDRA P M, FUKUNAGA K. A Branch and Bound Algorithm for Feature Subset Selection[J]. IEEE Transactions of Computation, 1977, 26(9): 917-922.

- [6] DOAK J. An Evaluation of Feature Selection Methods and Their Application to Computer Security[R]. Department of Computer Science, University of California at Davis, 1992:131-139.
- [7] BRASSARD G, BRATLEY P. Fundamentals of Algorithms[M]. New Jersey: Prentice Hall, 1996:377.
- [8] LIU H, MOTODA H. Feature Selection for Knowledge Discovery and Data Mining[M]. Boston: Kluwer Academic, 1998:446.
- [9] ALMUALLIM H, DIETTERICH T G. Learning Boolean Concepts in the Presence of Many Irrelevant Features[J]. Artificial Intelligence, 1994(69):279-305.
- [10] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000:322.
- [11] BASSAT B M. Pattern Recognition and Reduction of Dimensionality[J]. Handbook of Statistics, 1982(9):773-791.
- [12] HALL M A. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning[C]. // Proceedings of 17th International Conference of Machine Learning, San Francisco, USA: Morgan Kaufmann, 2000:359-366.
- [13] KOHAVI R, JOHN G H. Wrappers for Feature Subset Selection[J]. Artificial Intelligence, 1997, 97(1-2):273-324.
- [14] DAS S. Filter, Wrappers and a Boosting-based Hybrid for Feature Selection[C]. // Proceedings of the 18th International Conference of Machine Learning, San Francisco, USA: Morgan Kaufmann, 2001:74-81.
- [15] MILLER D J, UYAR H S. A Mixture of Experts Classifier with Learning based on Both Labelled and Unlabelled Data [C]. // M. M, J. M, I (eds.). Advances in Neural Information Processing Systems, Cambridge, Massachusetts: MIT Press, 1997:571-577.
- [16] YAO J, MAO Q, GOODISON S, et al. Feature Selection for Unsupervised Learning Through Local Learning[J]. Pattern Recognition Letters, 2015(53):100-107.
- [17] PENG X G, MING Z, WANGH T. Text Learning and Hierarchical Feature Selection in Webpage Classification[C]. // AD-MA, Berlin, Heidelberg: Springer-Verlag, 2008:452-459.
- [18] VAPNIK V N, LERNER A. Pattern Recognition Using Generalized Portrait Method[J]. Automation and Remote Control, 1963(24):28-36.
- [19] VAPNIK V N. The Nature of Statistical Learning[M]. Berlin: Springer, 1995:1-50.
- [20] MIRANDA J, MONTOYA R, WEBER R. Linear Penalization Support Vector Machines for Feature Selection[C] // PReMI 2005. 2005:188-192.
- [21] 代琨, 于宏毅, 马学刚, 等. 基于支持向量机的特征选择算法综述[J]. 信息工程大学学报, 2014, 15(1):85-91.
- [22] WESTON J, MUKHERJEE S, CHAPELLE O, et al. Feature Selection for SVMs[J]. Advances in Neural Information Processing Systems, 2000(13):668-674.
- [23] GUYON I, WESTON J, BARNHILL S, et al. Gene Selection for Cancer Classification Using Support Vector Machines[J]. Machine Learning, 2002, 46(1-3):389-422.
- [24] RAKOTOMAMONJY A. Variable Selection Using SVM-based Criteria[J]. Journal of Machine Learning Research, 2003 (3):1357-1370.
- [25] MUNDRA P A, RAJAPAKSE J C. SVM-RFE with MRMR Filter for Gene Selection[J]. IEEE Transactions on NanoBio-science, 2010, 9(1):31-37.
- [26] 王俭臣, 单甘霖, 张岐龙, 等. 基于改进 SVM-RFE 的特征选择方法研究[J]. 微计算机应用, 2011, 32(2):70-74.
- [27] ZHOU X, TUCK D P. MSVM-RFE: Extensions of SVM-RFE for Multi-class Gene Selection on DNA Microarray Data[J]. Bioinformatics, 2007, 23(9):1106-1114.
- [28] ZHOU Q F, HONG W C, SHAO G F, et al. A New SVM-RFE Approach Towards Ranking Problem[C] // Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009:270-273.
- [29] MALDONADO S, WEBER R. A Wrapper Method for Feature Selection Using Support Vector Machines[J]. Information Sciences, 2009, 179(13):2208-2217.
- [30] MANGASARIAN O L, WILD E W. Feature Selection for Nonlinear Kernel Support Vector Machines[C] // Proceedings of the Seventh IEEE International Conference on Data Mining-workshops. 2007:231-236.
- [31] WANG L F, SHEN X T. Multi-category Support Vector Machines, Feature Selection and Solution Path[J]. Statistica

Sinica, 2006(16):617-633.

- [32] WANG L, ZHU J, ZOU H. The Doubly Regularized Support Vector Machine[J]. Statistica Sinica, 2006, 16(2):589-615.
- [33] LIU Y, ZHENG Y F. FS_SFS: A Novel Feature Selection Method for Support Vector Machines[J]. Pattern Recognition, 2006, 39(7):1333-1345.
- [34] LEE M C. Using Support Vector Machine with a Hybrid Feature Selection Method to the Stock Trend Prediction[J]. Expert Systems with Applications, 2009, 36(8):10896-10904.
- [35] PENG Y H, WU Z Q, JIANG J M. A Novel Feature Selection Approach for Biomedical Data Classification[J]. Journal of Biomedical Informatics, 2010(43):15-23.
- [36] ALADEEMY M, TUTUN S, KHASAWNEH M T. A New Hybrid Approach for Feature Selection and Support Vector Machine Model Selection Based on Self-adaptive Cohort Intelligence[J]. Expert Systems with Applications, 2017(88):118-131.
- [37] SHAO Y H, LI C N, LIU M Z, et al. Sparse, L_Q-norm Least Squares Support Vector Machine with Feature Selection[J]. Pattern Recognition, 2018(78):167-181.
- [38] ALJARAHI I, AL-ZOUBI A M, FARIS H, et al. Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm[J]. Cognitive Computation, 2018(2):1-18.
- [39] LABBÉ, MARTINE, MARTÍNEZ-MERINO, et al. Mixed Integer Linear Programming for Feature Selection in Support Vector Machine[J]. Discrete Applied Mathematics, 2019(261):276-304.
- [40] LIN K C, HUNG J C, WEI J T. Feature Selection with Modified Lion's Algorithms and Support Vector Machine for High-dimensional Data[J]. Applied Soft Computing, 2018(68):669-676.

A Research Review of Feature Selection Algorithm

LIANG Wuqi, WANG Ronghua, LIU Keli, LI Bin

(School of Information and Engineering, Anhui Radio and TV University, Hefei 230022, China)

Abstract: Feature selection is the key step of the preprocessing module in pattern recognition, which is used to reduce the dimension of feature space and improve the classification performance of the classifier. First, the paper analyzes the main process of feature selection, and discusses the classification method of feature selection from different perspectives. Then the paper analyzes the function of feature selection based on SVM, and summarizes the feature selection algorithm based on SVM. Finally, the paper analyzes and compares the advantages and disadvantages of three kinds of feature selection based on SVM, and also points out the research hotspots and application development directions in theoretical research and practical application.

Keywords: pattern recognition; text categorization; feature selection; support vector machine; Wrapper method

[责任编辑 李潜生]