

一种引入冗余控制的特征排序模型

周 星¹,刁兴春²,曹建军²

(1. 解放军理工大学 指挥信息系统学院,江苏 南京 210007;2. 南京电讯技术研究所,江苏 南京 210007)

摘 要:针对特征排序方法较少考虑特征之间的相关关系,导致选择的特征子集存在冗余的问题,提出一种引入冗余控制的特征排序模型。将特征子集判别能力最大且冗余程度最小作为模型的目标函数,以降低特征之间的冗余;使用贪心方法和非线性规划方法对模型进行求解。在9个开源数据上的实验及与特征排序方法比较表明,本模型在大部分数据上,所选择的特征子集能够获得更好的分类准确性且个数更少;使用非线性规划方法求解时,能够直接得到特征子集,有利于确定特征个数。本模型可用于特征之间存在冗余时的特征选择。

关键词:特征选择;特征排序;特征相关;非线性规划

中图分类号:TP391

文献标志码:A

A Feature Ranking Model with Redundancy Control

ZHOU Xing¹, DIAO Xingchun², CAO Jianjun²

(1. School of Command Info. System, PLA Univ. of Sci. and Technol., Nanjing 210007, China;

2. Nanjing Inst. of Telecommunications Technol., Nanjing 210007, China)

Abstract: Aimed at problems of feature redundancy caused by the fact that feature correlation was seldom considered in the feature ranking methods, a feature ranking model with redundancy control was proposed. Maximum discrimination ability and minimum redundancy of a feature subset were used as the objective functions of the very model so as to reduce the redundancy among features, and greed and non-linear programming methods were employed to solve the model. Experiments were conducted on 9 public datasets and compared with feature ranking, and the result showed that the model can obtain a better classification accuracy and less feature size on most datasets. When non-linear programming method is employed, the model can yield a feature subset, on benefit for determining the feature size. This model can be used when correlation exists among features.

Key words: feature selection; feature ranking; feature redundancy; nonlinear programming

特征选择是模式识别的关键技术之一,它移除无关或冗余的特征,降低数据的维度,提高预测的准确性,提升数据挖掘的效率^[1]。根据是否考虑特征之间的相关关系,将特征选择方法分为特征排序和特征子集选择两类方法^[2]。特征排序方法通常按照某一评价准则对特征排序,如方差^[3]按照特征的代表能力排序;Laplacian 分数^[4]按照特征的代表能力以及局部保持能力排序;Fisher 分数^[3]按照特征的判别能力排序;约束分数^[5]按照特征使得同类样本尽量近,不同类样本尽量远的能力排序;Relief^[6]

根据样本在各特征上的假设间隔排序;信息增益^[7]根据特征与类标之间的相关性排序。特征子集选择方法则考虑了特征之间的相关关系,如最大投影最小冗余^[8]选择最能够重构原特征且冗余最小的子集、最大相关最小冗余^[9]选择代表能力最强且冗余最小的特征子集。

特征排序方法不考虑特征之间的相关关系,因此计算复杂性更低。特征子集选择方法的结果为一个子集,特征个数易确定。而特征排序方法通常选择前 k 个或者前 $k\%$ 个特征作为特征选择结果,由

收稿日期:2015-10-05

基金项目:国家自然科学基金资助项目(61371196)

作者简介:周 星(1988—),男,博士生.研究方向:数据工程;数据质量. E-mail:zx0327@163.com

网络出版时间:2016-9-18 15:11:36

网络出版地址: <http://www.cnki.net/kcms/detail/51.1596.T.20160918.1511.006.html>

<http://jsuese.scu.edu.cn>

于不考虑特征之间的相关关系,选择的结果可能存在冗余。为了移除特征排序后的冗余特征,进一步提高特征排序方法的效果,一种可行的思路是将特征子集选择中的冗余控制引入特征排序方法中。

Javed 等^[2]针对高维文本分类问题,提出了一种基于两阶段马尔可夫毯的方法,先用特征排序算法对特征进行排序,选择前 k 个特征,再用马尔可夫毯滤除冗余特征,实验结果表明,相比仅使用特征排序,准确性得到了较大提高。然而,姚旭等^[10]指出,马尔可夫毯的计算复杂性较高,针对高维文本,移除冗余特征的计算复杂性将较高。Liu 等^[11]针对特征排序方法仅考虑特征与类标之间的相关关系,无法移除冗余特征的问题,提出一种基于依赖边缘的方法,同时考虑已选特征和剩余特征与类标之间的相关关系,利用依赖边缘对各特征进行排序,并采用前向贪心搜索进行选择。Liu 等提出的方法也采用马尔可夫毯,且需要为所有特征计算马尔可夫毯,计算复杂性高。Zhao 等^[1]对常见的排序方法,如 Laplacian 分数、Fisher 分数、Relief 等进行了研究,指出它们都选择能够保持样本相似度的特征,并提出了一个涵盖上述方法的统一的框架。理论分析表明,该框架克服了这些排序方法无法处理特征冗余的不足。然而,该框架的相似性和冗余度量仅考虑了特征与类别之间、特征与特征之间的线性相关关系,无法涵盖非线性相关关系。

本文将冗余控制引入特征排序方法,提出一种引入冗余控制的特征排序模型,即将选择的特征子集满足最大判别能力且冗余最小作为目标函数,使得该模型选择的特征子集既具有最大的判别能力,又具有最小的冗余,相比现有的特征排序方法,本模型选择的特征能够获得更高的分类正确率、更小的特征个数,且冗余度更小。

1 引入冗余控制的特征排序模型

1.1 特征冗余的度量

对特征判别能力的度量研究较多,比如可以用特征排序函数度量特征的判别能力,对特征冗余的研究相对较少。本文将常用的冗余度量分为基于子集的度量和基于两两比较的度量。

1.1.1 基于子集的度量

马尔可夫毯是一种典型的基于子集的度量,自从 Lei 等^[12]利用马尔可夫毯定义冗余特征以来,马尔可夫毯在度量特征的冗余程度方面得到了越来越广泛的应用,如文献^[2,9]均使用马尔可夫毯滤除

无关特征。马尔可夫毯的定义如下:

给定特征 $F_i \in F$, 令 $M_i \in F (F_i \notin M_i)$, 当且仅当 $p(F - M_i - \{F_i\}, C | F_i, M_i) = p(F - M_i - \{F_i\}, C | M_i)$ 时, M_i 是 F_i 的马尔可夫毯。其中, F_i 为第 i 个特征, F 为特征集合的全集, M_i 为特征子集, C 为类标。根据马尔可夫毯, Lei 等定义冗余特征如下:

令 S 是一个特征子集, 当且仅当 F_i 在 S 中有一个马尔可夫毯, F_i 是冗余的。

由于马尔可夫毯的求解需要为每个特征在所有的特征子集中搜索, 时间复杂度为 $O(2^k)$ (k 为特征的维数)^[10]。因此, 姚旭等提出利用特征之间的相关性近似马尔可夫毯, 该定义如下^[10]:

对特征 F_i 和 F_j , E 为一种度量准则, 当且仅当 $E(F_i; C) > E(F_j; C)$, 称 F_i 是 F_j 的一个近似马尔可夫毯。

该定义相比马尔可夫毯效率更高, 但是, 所定义的马尔可夫毯只是一个特征, 而更多情况下, 马尔可夫毯是一个特征集合。

1.1.2 基于两两比较的度量

基于两两比较的度量多根据两两特征之间的相关性定义冗余。Zhao 等^[1]为评估其提出的框架, 给出了一个评估特征子集冗余程度的度量。该度量定义如下^[1]:

$$r(F_s) = \frac{1}{m(m-1)} \sum_{F_i, F_j \in F_s, i > j} \rho_{i,j} \quad (1)$$

式中, F_s 为已选特征集合, m 为特征的个数, $\rho_{i,j}$ 为特征 F_i 和 F_j 的皮尔逊相关系数。

Wang 等^[8]提出的最大投影最小冗余方法, 采用了 Zhao 等^[1]提出的冗余度量, 在计算过程中, 利用特征向量的内积代替皮尔逊相关系数。

Chakraborty 等^[13]根据特征之间的相关性也给出了冗余的度量, 定义如下:

$$r(F_s) = \frac{1}{m(m-1)} \sum_{i=1}^m f(\beta_i) \sum_{j \neq i} f(\beta_j) dep(F_i, F_j) \quad (2)$$

式中: m 为特征的个数; dep 为特征相关性的度量, 当考虑线性依赖时, 使用皮尔逊相关系数的平方或绝对值, 当考虑非线性依赖时, 使用互信息; $f(\cdot)$ 为一个门函数, 当特征 F_i 很重要时, $f(\beta_i)$ 接近 1。

Chakraborty 等^[13]提出的冗余度的度量可以看作 Zhao 等^[1]定义的扩展, 它既考虑了线性相关的情况, 又考虑了非线性相关的情况。

1.2 引入冗余控制的特征排序模型

本文将所选特征子集判别能力最大看作一个目

标函数,特征子集冗余最小看作另一个目标函数,给出式(3)所示的多目标优化模型。

$$\begin{cases} \max d(F_s), \\ \min r(F_s) \end{cases} \quad (3)$$

式中, F_s 为选择的特征子集, $d()$ 为特征集合判别能力的度量函数, $r()$ 为特征集合冗余程度的度量函数。

该模型直观意义为要求选择的特征集合的判别能力最大,同时冗余程度最小。这2个目标函数不能同时满足,因此它是一个典型的多目标优化问题。

本文的特征排序方法使用 Relief 和信息增益作为线性和非线性特征排序方法的代表对特征进行排序,冗余度量采用式(1)的定义,使用互信息代替皮尔逊相关系数。

2 模型的求解

提出贪心(Greed)搜索和非线性规划(nonlinear programming, NLP)2种方法对模型进行求解。

2.1 贪心搜索

贪心搜索的思想是每次都从候选特征集合中选择判别能力最大,并与已选特征集合的冗余最小的特征。

式(3)是一个多目标优化问题,为简化求解,参考文献[5,8,13],将多目标转换为单目标。本文给出2种转换方法。其中:式(4)对应线性转换,即判别能力和冗余度2个目标线性组合成一个目标;式(5)对应非线性转换,即判别能力和冗余度2个目标非线性组合成一个目标。

式(4)是最常见的转换方法,它引入一个新的 α 参数,平衡判别能力和冗余度的权重;式(5)不引入任何新的参数:

$$\min_{F_s} -\alpha d(F_s) + (1 - \alpha)r(F_s) \quad (4)$$

$$\min_{F_s} \frac{r(F_s)}{d(F_s)} \quad (5)$$

以式(4)为例,第1个选择的特征为:

$$F_1 = \arg \min_i ((1 - \alpha)r_i - \alpha d_i) \quad (6)$$

式中: r_i 为第 i 个特征与其余所有特征的冗余度的均值, $r_i = \frac{1}{m-1} \sum_{j \neq i} r(f_i \cup f_j)$, m 为特征的个数; d_i 为第 i 个特征的判别能力。

第 n 个选择的特征为:

$$F_n = \arg \min_i ((1 - \alpha)r(F_i \cup F_s) - \alpha d_i) \quad (7)$$

式中, F_s 为已选特征集合, $r(F_i \cup F_s)$ 为第 i 个特征与已选特征集合的冗余度。

选择的先后顺序即是特征排序的顺序,结果是特征的有序排列。

2.2 非线性规划方法

非线性规划方法求解是将式(3)转换为一个非线性规划问题,其结果是一个特征子集。

将式(3)改写为式(8)的形式:

$$\begin{aligned} \min_{\mathbf{w}} \quad & -\alpha \mathbf{D} \mathbf{w}^T + (1 - \alpha) \mathbf{w} \mathbf{R} \mathbf{w}^T \\ \text{s. t.} \quad & 0 < \mathbf{w} < \boldsymbol{\beta} \end{aligned} \quad (8)$$

式中: $\mathbf{D} = (d_1, d_2, \dots, d_n)$ 为各特征的判别能力组成的向量;

$$\mathbf{R} = \frac{1}{m(m-1)} \begin{bmatrix} 0 & r(F_1 \cup F_2) & \dots & r(F_1 \cup F_n) \\ r(F_1 \cup F_2) & 0 & \dots & r(F_2 \cup F_n) \\ \vdots & \vdots & \ddots & \vdots \\ r(F_1 \cup F_n) & r(F_2 \cup F_n) & \dots & 0 \end{bmatrix}$$

为所有两两特征向量的冗余度构成的冗余度矩阵; $\mathbf{w} = (w_1, w_2, \dots, w_n)$ 为特征的权重向量; α 为平衡判别能力和冗余度的权重的参数; $\boldsymbol{\beta}$ 为控制特征个数的参数。

式(8)是一个典型的线性约束二次规划问题,其结果是特征的权重向量,仅选择其中权重大于0的特征,即得到选择的特征子集。

3 实验

3.1 实验设置

实验数据来自于 UCI 机器学习数据库^[16]中的分类数据。各数据的基本信息描述如表1所示。

表1 数据的基本信息

Tab.1 Basic information of data

数据	样本数	特征数	类别
Vote	435	16	2
Sonar	208	60	2
Hypothyroid	3 772	29	4
Hepatitis	155	19	2
Glass	214	9	7
Ionosphere	351	34	2
Diabetes	768	8	2
Brest cancer	286	9	2
Colic	368	27	2

Hypothyroid 数据中由于仅有一个样本在 hypopituitary 特征上取值不一样,另外所有样本在 TBG-measured 和 TBG 2个特征上的取值都一样,因此去除这3个特征后再进行特征选择。Ionosphere 数据的 a02 特征取值全为0,因此去除该特征。

实验包括 3 个方面:首先,从分类正确率和特征个数 2 个方面比较本模型与特征排序方法,说明本模型的有效性;其次,比较贪心搜索求解中,多目标到单目标的 2 种转换方法;最后,比较冗余度和判别能力的权重对于模型结果的影响。

利用 Weka 提供的特征选择模块计算各特征的 Relief 和信息增益分数。采用 Zi^[16]的代码实现计算特征之间的互信息。

由于 K 近邻的参数个数少,分类正确率受分类器本身参数的影响较小,因此,采用 K 近邻分类器验证特征选择的效果,将 K 设置为 5。在估计分类正确率时,采用 10 轮 5 重交叉验证,并取 10 轮结果的分正确率的均值作为最终的分正确率。

实验环境为 1 台 Intel i7-4790 的 CPU,4 GB 内存,Windows7 操作系统的 PC,算法基于 MATLAB 7.1 实现。采用 MATLAB 优化工具箱中的 fmincon 函数求解式(8)。

3.2 模型的有效性验证

验证模型的有效性时,使用贪心搜索和非线性规划 2 种方法对本模型进行求解。

对贪心搜索,采用式(5)将特征个数从 2 开始逐步增加,直到特征总数,将得到的最大分类正确率作为最终的分正确率,将得到最大分正确率时的特征个数作为最终的特征个数。

对非线性规划方法,将 α 设为 0.5,由于 β 的取值将影响选择出的特征的个数,进一步影响分正确率,且经过实验发现, β 的倒数与选出的特征个数直接相关,因此,将 β 的取值设置为 $1/n$,并使得 n 从 1 到总特征数逐步增加,选择得到最高分正确率时的特征个数的倒数为 β 的值。

在后续实验结果中,使用 Relief 表示仅使用 Relief 排序,使用 IG(information gain)表示仅使用信息增益排序,使用 Greed 表示使用贪心搜索求解,使用 NLP 表示使用非线性规划方法求解。

1) 只用 Relief 排序和本模型对比分正确率和所选特征个数。

Relief 排序和本模型的分正确率对比如表 2 所示,选择的特征个数对比如表 3 所示。

从表 2 可知:相比仅使用 Relief 排序,本模型使用贪心搜索求解时,在 3 个数据上分正确率得到了较大提高,在 5 个数据上分正确率略有提高,在 1 个数据上分正确率略有降低;本模型使用非线性规划方法求解时,在 5 个数据上分正确率得到了较大提高,在 2 个数据上有较大降低,在 2 个数据

上略有降低。

表 2 Relief 排序和本模型的分正确率对比

Tab. 2 Comparison of classification accuracy between Relief ranking and the proposed model

数据	Relief	Greed	NLP
Vote	0.961 6	0.966 3	0.960 5
Sonar	0.816 7	0.821 4	0.823 8
Hypothyroid	0.958 2	0.973 3	0.922 7
Hepatitis	0.806 5	0.822 6	0.867 7
Glass	0.639 5	0.676 7	0.667 4
Ionosphere	0.884 3	0.894 3	0.910 0
Diabetes	0.740 3	0.740 3	0.733 8
Brest cancer	0.966 4	0.967 9	0.972 1
Colic	0.725 7	0.716 2	0.804 1

表 3 Relief 排序和本模型选择的特征个数对比

Tab. 3 Comparison of number of features selected between Relief ranking and the proposed model

数据	Relief	Greed	NLP
Vote	2	2	2
Sonar	13	14	7
Hypothyroid	17	3	5
Hepatitis	9	10	11
Glass	5	5	6
Ionosphere	4	4	5
Diabetes	5	5	5
Breast cancer	7	4	6
Colic	6	16	5

从表 3 可知:相比仅使用 Relief 排序,本模型使用贪心搜索求解时,在 2 个数据上选择的特征个数有较大减少,在 2 个数据上略有增多,1 个数据上有较大增多;本模型使用非线性规划方法求解时,在 2 个数据上有较大减少,在 3 个数据上略有增多。

2) 只用信息增益排序和本模型对比分正确率和所选特征个数。

信息增益排序和本模型的分正确率对比如表 4 所示,选择的特征个数对比如表 5 所示。

从表 4 可知:相比仅使用信息增益排序,本模型使用贪心搜索求解时,在 3 个数据上分正确率得到了较大提高,在 1 个数据上分正确率有较大降低,其余数据上的分正确率略有提高;本模型使用非线性规划方法求解时,在 4 个数据上分正确率得到了较大提高,在 1 个数据上分正确率略有提高,但是,在 3 个数据上分正确率有较大降低,在

1 个数据上分类正确率略有降低。

表4 信息增益排序和本模型分类正确率对比

Tab.4 Comparison of classification accuracy between IG ranking and the proposed model

数据	IG	Greedy	NLP
Vote	0.969 8	0.964 0	0.957 0
Sonar	0.811 9	0.826 2	0.826 2
Hypothyroid	0.956 5	0.969 5	0.968 6
Hepatitis	0.809 7	0.845 2	0.832 3
Glass	0.711 6	0.714 0	0.674 4
Ionosphere	0.890 0	0.907 1	0.918 6
Diabetes	0.738 3	0.738 3	0.737 7
Brest cancer	0.965 0	0.965 0	0.970 7
Colic	0.727 0	0.663 5	0.647 3

表5 信息增益排序和本模型选择的特征个数对比

Tab.5 Comparison of number of features selected between IG ranking and the proposed model

数据	IG	Greedy	NLP
Vote	5	3	3
Sonar	31	7	15
Hypothyroid	12	3	3
Hepatitis	9	5	3
Glass	5	7	3
Ionosphere	4	3	2
Diabetes	7	7	3
Breast cancer	9	9	6
Colic	6	2	2

从表5可知:相比仅使用信息增益排序,本模型使用贪心搜索求解时,在4个数据上选择的特征个数有较大减少,在1个数据上略有减少;本模型使用非线性规划方法求解时,在6个数据上有较大减少,在3个数据上略有减少。

为比较引入冗余控制后的冗余度变化,以 Hypothyroid 为例,比较仅用信息增益排序和在利用贪心搜索求解本模型时冗余度随特征个数变化的情况,结果如图1所示。

从图1可知,在相同特征个数时,相比仅使用信息增益排序,本模型所选特征的冗余度有较大下降。

综合来看,相比原特征排序方法,本模型在运用贪心搜索和非线性规划方法求解时,分类正确率和特征个数上都有一定的改进。本模型在部分数据上效果不够好,这可能是参数不够好,或者这些数据的特征之间不存在冗余。

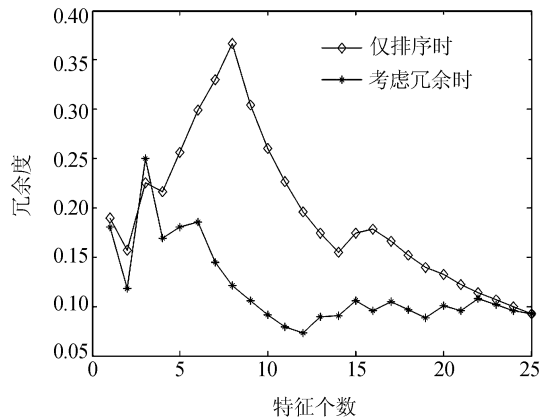


图1 信息增益排序和本模型的冗余度对比

Fig.1 Comparison of redundancy between IG ranking and the proposed model

3.3 两种多目标到单目标的转换方法对比

针对两种多目标到单目标的转换方法,以 Hypothyroid 数据集为例,仍然令 α 为 0.5, 进行对比实验。比较其在利用贪心搜索求解本模型时,两种多目标到单目标转换方法的效果。用 1~26 对原特征进行顺序编号。

线性转换方法对应的结果: {26, 3, 16, 11, 6, 2, 18, 10, 8, 22, 14, 4, 12, 25, 15, 9, 13, 5, 20, 21, 24, 7, 1, 19, 23, 17};

非线性转换方法对应的结果: {26, 10, 8, 11, 22, 3, 6, 18, 2, 16, 25, 14, 15, 21, 20, 9, 24, 4, 17, 5, 13, 19, 7, 1, 23, 12}。

分类正确率和冗余度对比分别如图2、3所示。

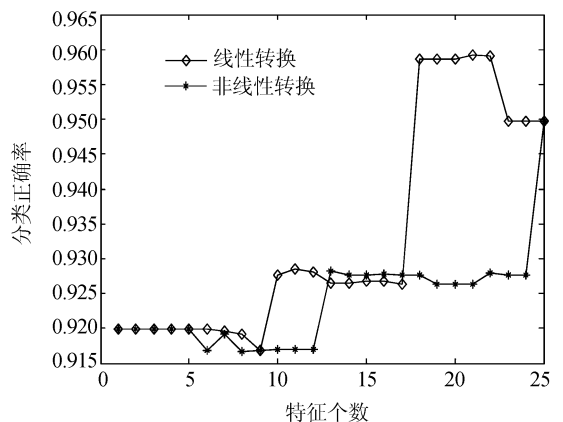


图2 线性转换和非线性转换的分类正确率对比

Fig.2 Comparison of classification accuracy between linear and non-linear transformation

从图2和3可知,线性转换和非线性转换的结果存在一些区别。在本数据上,线性转换能够获得更高的分类正确率。线性转换可以为每个数据,调整 α 的取值,进一步改变冗余度和判别能力的权重,以寻求最优结果。然而, α 参数的引入在提高灵活性

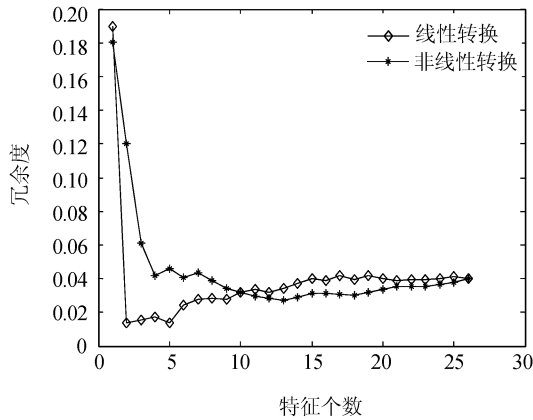


图 3 线性转换和非线性转换的冗余度对比

Fig. 3 Comparison of redundancy between linear and non-linear transformation

的同时将使得模型的复杂性增加。非线性转换不引入额外的参数,复杂性更小。

3.4 冗余度和判别能力的权重对模型结果的影响

为进一步说明冗余度和判别能力的权重对模型结果的影响,比较贪心搜索求解中,线性转换中的 α 取值对于特征排序结果的影响。 α 取值越大,则冗余度的权重越小,越接近原排序方法; α 取值越小,则冗余度的权重越大。

以 Glass 数据集为例,将原特征按 1 ~ 9 进行顺序编号,使 α 从 0.1 到 0.9 变化,分类正确率以及排序结果如表 6 所示。

表 6 线性转换中的 α 取值比较

Tab. 6 Comparison of α in linear transformation

α	正确率	特征排序结果
0.1	0.672 1	{3,8,9,6,4,5,7,2,1}
0.2	0.696 7	{3,8,9,6,4,7,2,5,1}
0.3	0.696 7	{3,8,9,6,4,7,2,5,1}
0.4	0.687 0	{3,8,9,4,6,7,2,5,1}
0.5	0.676 7	{3,8,9,4,6,7,2,1,5}
0.6	0.680 0	{3,8,4,9,7,2,6,1,5}
0.7	0.668 8	{3,8,4,7,9,2,1,6,5}
0.8	0.690 2	{3,8,4,7,2,1,6,9,5}
0.9	0.669 8	{3,8,4,7,2,1,5,6,9}

从表 6 可知,分类正确率集中在 (0.668 8, 0.696 7) 的区间,在本实验中, α 在取值为 0.2 和 0.3 时分类正确率取得最大值。实际应用中,由于各数据特征之间的冗余程度不同,需要为每个数据选择合适的 α 值。

4 结论

将冗余控制引入特征排序模型中,将选择的特

征子集既得到最高分类正确率又保持最小冗余作为目标函数,并使用贪心搜索和非线性规划方法对模型进行求解。实验表明,本模型相比原特征排序方法,在大部分数据上,可取得更高的分类正确率以及更少的特征个数,且所选特征的冗余度得到降低。下一步研究需要进一步为每个数据选择最优参数。

参考文献:

- [1] Zhao Zheng, Wang Lei, Liu Huan, et al. On similarity preserving feature selection [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(3): 619-631.
- [2] Javed K, Maruf S, Haroon A B. A two-stage markov blanket based feature selection algorithm for text classification [J]. Neurocomputing, 2015, 157(2015): 91-104.
- [3] Bishop C M. Neural networks for pattern recognition [M]. Oxford: Oxford University Press, 1995.
- [4] He Xiaofei, Cai Deng, Niyogi P. Laplacian score for feature selection [C] // Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2005.
- [5] Zhang Daoqiang, Chen Songcan, Zhou Zhihua. Constraint score: A new filter method for feature selection with pairwise constraints [J]. Pattern Recognition, 2008, 41(5): 1440-1451.
- [6] Kira K, Rendell L A. A practical approach to feature selection [C] // Proceedings of the Ninth International Workshop on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc, 1992: 249-256.
- [7] Mitchell T. Machine learning [M]. Columbus: McGraw-Hill, 1997.
- [8] Wang Shiping, Pedrycz W, Zhu Qingxin, et al. Unsupervised feature selection via maximum projection and minimum redundancy [J]. Knowledge-Based Systems, 2015, 75: 19-29.
- [9] Song Qinbao, Ni Jingjie, Wang Guangtao. A fast clustering-based feature subset selection algorithm for high dimensional data [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(1): 1-14.
- [10] Yao Xu, Wang Xiaodan, Zhang Yuxi, et al. Ensemble feature selection algorithm based on markov blanket and mutual information [J]. Systems Engineering and Electronics, 2012, 34(5): 1047-1050. [姚旭, 王晓丹, 张玉玺, 等. 基于 Markov Blanket 和互信息的集成特征选择算法 [J]. 系统工程与电子技术, 2012, 34(5): 1046-1050.]
- [11] Liu Yong, Tang Feng, Zhang Zhiyong. Feature selection based on dependency margin [J]. IEEE Transactions on Cybernetics, 2015, 45(6): 1209-1221.
- [12] Yu Lei, Liu Huan. Efficient feature selection via analysis of relevance and redundancy [J]. Journal of Machine Learning Research, 2004, 10(5): 1205-1224.
- [13] Chakraborty R, Pal N R. Feature selection using a neural framework with controlled redundancy [J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(1): 35-50.
- [14] Lichman M. UCI Machine learning repository [EB/OL]. [2015-10-01]. <http://archive.ics.uci.edu/ml>.
- [15] Zi Qiao, 互信息和标准化互信息 [EB/OL]. (2011-11-23) [2015-10-01]. <http://www.cnblogs.com/ziqiao/archive/2011/12/23/2286273.html>.

(编辑 赵婧)