

## 基于异类近邻的支持向量机加速算法

陈景年<sup>a</sup>, 胡顺祥<sup>b</sup>, 徐 力<sup>c</sup>

(a. 山东财经大学 信息与计算科学系, 济南 250014; b. 鲁南技师学院 商务管理系, 山东 临沂 276021; c. 济南市公路管理局 信息科, 济南 250013)

**摘 要:** 支持向量机的训练时间随样本增多而明显增加。为了在保持训练效果的同时提高训练速度, 给出精简训练数据集的一种算法。对每个样本, 通过选择异类近邻来构成训练集, 利用异类近邻来选择边界样本。实验结果表明, 与 FCNN 算法和 NPPS 算法相比, 该算法在保持甚至增进支持向量机分类效果的同时, 能大幅提高训练效率。  
**关键词:** 支持向量机; 样本选择; k 近邻; 异类; 分类

**中文引用格式:** 陈景年, 胡顺祥, 徐 力, 等. 基于异类近邻的支持向量机加速算法[J]. 计算机工程, 2018, 44(5): 19-24.

**英文引用格式:** CHEN Jingnian, HU Shunxiang, XU Li. Speeding Up Algorithm for Support Vector Machine Based on Alien Neighbor[J]. Computer Engineering, 2018, 44(5): 19-24.

## Speeding Up Algorithm for Support Vector Machine Based on Alien Neighbor

CHEN Jingnian<sup>a</sup>, HU Shunxiang<sup>b</sup>, XU Li<sup>c</sup>

(a. Department of Information and Computing Science, Shandong University of Finance and Economics, Jinan 250014, China;

b. Department of Commerce Management, Lunan Technician Institute, Linyi, Shandong 276021, China;

c. Information Department, Jinan Road Management Bureau, Jinan 250013, China)

**【Abstract】** The training time of Support Vector Machine (SVM) becomes much longer with the increase of training instances. To speeding up the training process without losing the effect, a method for reducing training dataset is proposed. A new training set is composed with nearest neighbors of each instance selected from different classes. A boundary sample is selected by using an alien nearest neighbor. Experimental results show that compared with FCNN algorithm and NPPS algorithm, the proposed algorithm can enormously raising the training efficiency while keeping or even improving the classification accuracy of SVM.

**【Key words】** Support Vector Machine (SVM); instance selection; k nearest neighbor; alien; classification

**DOI:** 10.19678/j.issn.1000-3428.0046453

### 0 概述

支持向量机 (Support Vector Machine, SVM) 是基于统计学习理论的一种机器学习方法, 在结构风险最小化基础上, 可有效地避免传统方法中局部极小、过拟合、维数灾难等常见问题, 并且在小样本集上依然能取得好的学习效果<sup>[1]</sup>。因此, 在文本分类、网页过滤、目标跟踪等领域得到了广泛应用<sup>[2-4]</sup>。

标准的 SVM 训练过程归结为一个凸二次规划问题。随着训练样本的增多, 其训练空间和时间会急剧增加。所以, SVM 难以适用于大规模数据集。如何提高 SVM 的学习效率以适应大规模数据处理的需求成为近年来一个研究热点<sup>[5-8]</sup>。

本文利用异类近邻来选择边界样本, 这一方法不仅适用于复杂边界的情形, 而且对多类分类问题

也非常有效。

### 1 研究背景

文献[1]指出, 在 SVM 的训练过程中, 只有支持向量对构建分类模型起作用, 并且支持向量往往位于类边界附近。因此, 边界附近的样本对 SVM 的分类结果起决定作用。所以在构建 SVM 分类模型时, 不必应用所有训练样本, 而只需选取其中最有可能成为支持向量的边界样本参与训练过程。这样, 在保持构建的分类模型效果的同时, 因训练样本的减少而可以大幅度提高训练效率。

基于上述思路, 研究人员给出了一些有效的方法以选取潜在的支持向量来构成精简的训练集, 提高了训练效率。

文献[9]利用一个样本分别到 2 类样本中每个

**基金项目:** 国家自然科学基金“融合事件关系推理和情感博弈的网络不实信息演化机理研究”(61502151)。

**作者简介:** 陈景年 (1970—), 男, 教授、博士, 主研方向为大规模数据分析、机器学习; 胡顺祥, 副教授; 徐 力, 高级工程师。

**收稿日期:** 2017-03-21

**修回日期:** 2017-06-06

**E-mail:** jnchen06@163.com

类的中心的距离之比来判断该样本是否为边界样本。该方法在保持 SVM 分类效果的情况下,可明显提高 SVM 的训练速度。文献[10]首先用一个小样本子集训练得到一个初始的 SVM 分类超平面,然后将原样本集中离此平面较远的样本删除。通过该方法可以删除一些冗余样本,在保持 SVM 泛化能力的同时,提高了其训练速度。

文献[11]利用每个样本的  $k$  近邻在训练集中的序号及相应的距离对样本打分,并按打分高低选择样本。文献[12]根据样本的  $k$  近邻中异类样本数来选择样本。

近来,文献[13]利用每个样本的最近邻信息得到训练集的一致子集,然后利用此子集训练 SVM。该选择算法的效率较高,但有时样本选择效果不够理想。文献[14]以每类样本的聚类中心作为参照来选择其他类中的边界样本。

上述研究工作在保持分类效果和提高训练效率方面都有一定作用,但在数据集的类别边界非常复杂和类别较多的情况下,选择效果往往不理想。

## 2 基于异类近邻的边界样本选择

为使提出的算法更具普适性,考虑一般的多类分类问题。由于 SVM 分类模型是针对二类分类问题建模的,因此在应用 SVM 模型进行分类时,通常将一个多类分类问题转化为多个二类分类问题来解决。常用的转化方式有 2 种:一对一方式和一对多方式。在一对一方式中,多类中的每 2 个类之间都要构建一个 SVM 模型。这样,不仅需要训练大量的模型,而且使后续的分类过程变慢。在一对多方式中,每个类轮流做一次正类。在一个类作为正类时,其余类合起来作为负类。由于一对多方式需训练的 SVM 模型较少,而且后续的分类过程效率高,因此,在下面的样本选择算法中,将采用一对多方式。

假定训练集  $A$  是由  $L$  类样本所组成, $A$  中的样本总数为  $N$ 。 $A$  中的每一类样本都轮流做一次正类样本,与此同时,其余样本作为负类样本。下面考虑第  $c$  ( $1 \leq c \leq L$ ) 类样本为当前的正类样本时的边界样本选择问题,记选择的结果为  $S_c$ 。

选择过程包括正类样本选择和负类样本选择 2 个方面。选择结果  $S_c$  初始化为空集  $\phi$ 。首先进行的是负类样本的选择。对于第  $c$  类中的每个样本  $x$ ,计算  $x$  在第  $l$  ( $1 \leq l \leq L, l \neq c$ ) 类中的  $k$  个最近邻  $x_{lj}$  ( $1 \leq j \leq k$ )。判断  $x_{lj}$  是否属于  $S_c$ 。如果  $x_{lj} \in S_c$ ,则将  $x_{lj}$  选入  $S_c$ 。这里,计算  $k$  近邻的操作是在每一小类样本中进行的,不同于传统的做法在整个数据集中进行。因此,上述  $k$  近邻的搜索过程因搜索范围的缩小而较传统做法具有更高的效率。

正类样本的选择过程与负类样本选择类似。由于在多类分类问题中采用了一对多的转化模式,这可

能会引起正类与负类之间样本数目严重不均衡的问题。因此,在正类样本的选择过程中,增加了一项调节正类和负类之间均衡性的操作,即在正类样本明显偏少的情况下(比如少于选择的负类样本数的一半),则正类样本将全部保留。整个边界样本选择过程如图 1 所示。其中, $L$  为训练集  $A$  中的样本类数, $N$  为  $A$  中的样本总数, $N_c$  为第  $c$  类样本数, $N_s$  为选择的负例数, $S_c$  为第  $c$  类作为当前正类时样本选择的结果。

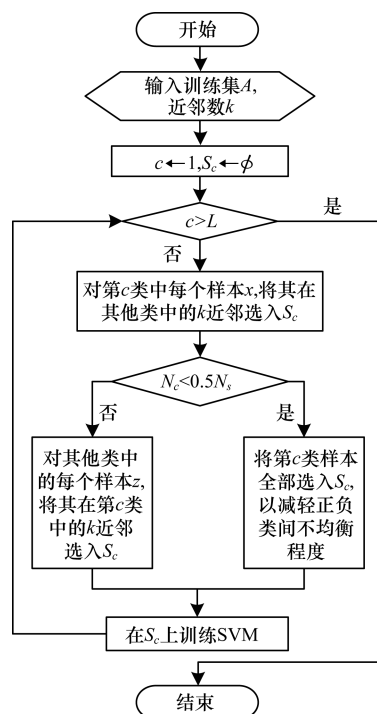


图 1 基于异类近邻的样本选择算法流程

下面是所给的样本选择和模型训练的算法描述。

**算法** 基于异类近邻的样本选择算法

**输入** 训练集  $A$  (含有  $L$  个类, 共  $N$  个样本), 近邻数  $k$

**输出** 对每个类别  $c$  ( $1 \leq c \leq L$ ), 输出以  $c$  类样本为正类, 其余样本为负类时的选择结果  $S_c$ 。

1) 初始化:  $c = 1; S_c \leftarrow \phi$ 。

2) 以  $c$  类样本为正例, 其余样本为负例。

3) 负例 (即负类样本) 的选择过程:

(1)  $l = 1$ ;

(2) 如果  $l \neq c$ , 则:

① 对于第  $c$  类中的每个样本  $x$ , 计算  $x$  在第  $l$  类中的  $k$  个最近邻  $x_{lj}$  ( $1 \leq j \leq k$ )。

② 如果  $x_{lj} \in S_c$ , 则  $S_c \leftarrow S_c \cup \{x_{lj}\}$ 。

(3)  $l \leftarrow l + 1$ ;

(4) 如果  $l \leq L$ , 则转 (2), 否则转 4)。

4) 正例 (即正类样本) 的选择过程:

(1) 比较  $c$  类样本数  $N_c$  和选择的负例数  $N_s$  的大小关系。

(2) 如果  $N_c < 0.5N_s$ , 则所有  $c$  类样本被选入  $S_c$ , 转 5)。

(3) 对每个负例  $y$ :

① 计算  $y$  在第  $c$  类中的  $k$  个最近邻  $x_{cj}$  ( $1 \leq j \leq k$ )。

② 如果  $x_{cj} \in S_c$ , 则  $S_c \leftarrow S_c \cup \{x_{cj}\}$ 。

- 5) 在得到的  $S_c$  上训练 SVM 模型。
- 6)  $c \leftarrow c + 1$ 。
- 7) 如果  $c \leq L$ , 转 2)。
- 8) 结束。

### 3 算法说明

#### 3.1 与相关算法的不同之处

尽管上述基于异类近邻的边界样本选择算法利用了样本的  $k$  近邻来进行样本选择,但与以往基于  $k$  近邻的样本选择方法相比,本文所给算法有如下不同之处:

1) 以往基于  $k$  近邻的算法往往是在整个训练集中搜索一个样本的  $k$  近邻,而上述算法是在除当前正类之外的每个小类中搜索一个正类样本的  $k$  近邻,简称异类  $k$  近邻。因此,搜索的效率会明显提高。

2) 与以往基于  $k$  近邻的算法相比,在上述算法中,异类  $k$  近邻的使用使得选择算法能够适用于各种复杂的类别边界,而不像以往算法往往受边界条件的限制。

3) 在正类样本的选择过程中,能够根据正类与负类样本数的差距来消除正、负类之间的不平衡问题,这也是上述算法的一个独到之处。

#### 3.2 算法的复杂度分析

所给算法包含了负例选择和正例选择 2 个过程,并且正例选择过程的复杂度不超过(许多情况下远低于)负例选择过程的复杂度。所以,这里只对负例选择过程的复杂度进行分析。

假设训练数据集  $A$  含有  $L$  个类,共  $N$  个样本,从而平均每个类中的样本数为  $N/L$ 。对于第  $c$  类中的每个样本  $x$ ,从其余的一小类样本中选择  $k$  个近邻的复杂度为  $\frac{N}{L} \log \frac{N}{L}$ 。因此,从其余  $L-1$  类的每类中都选

择  $k$  个近邻的复杂度为  $(L-1) \frac{N}{L} \log \frac{N}{L} < M \log \frac{N}{L}$ 。所以,整个负例选择过程乃至整个样本选择过程的复杂度不超过  $\frac{N^2}{L} \log \frac{N}{L}$ 。这一复杂度与 NPPS<sup>[12]</sup> 算法的

复杂度相近,都显著低于 SVM 模型的训练复杂度,且在类别数较多的情况下所给算法更具优势。

#### 3.3 算法的参数设置方法

在给出的样本选择算法中,一个样本的近邻数  $k$  对算法的效果和效率具有重要的作用。很显然,  $k$  越小,选择的样本越少,训练效率会越高。同时,也越容易损失一些支持向量,易引起训练效果的下降。反之,  $k$  越大,越不容易漏掉支持向量,但也往往会选入一些冗余的样本,不仅使选择的样本增多,训练效率降低,同时,较多的冗余样本也会降低训练效果。因此,  $k$  的值不宜过大或过小,而应在一定的范围内

寻找合适的  $k$  值。实验发现,绝大多数情况下,在 2~10 的范围内即可找到一个理想的  $k$  值。因此,通过实验不难确定合适的  $k$  值。

### 4 实验结果与分析

为验证本文所给算法的有效性,将它与性能显著的 2 个边界样本选择算法 NPPS<sup>[12]</sup> 以及 FCNN<sup>[13]</sup> 进行多方位的比较,具体包括如下 3 个方面:1) 样本选择的效果,即利用选择的样本得到的分类模型的分类准确率;2) 选择的样本占总训练样本的比例;3) 利用选择的样本训练分类模型所需的时间。

#### 4.1 实验数据集

实验共采用了 12 个数据集,其中 6 个为小规模数据集,另 6 个为中等规模或大规模数据集。它们大都来自于 UCI 的机器学习数据库<sup>[15]</sup>。另外,为了验证所给算法对多类分类,特别是大类别数据集分类的效果,实验中还采用了一个包含 3 755 类的手写中文字符数据集 HCL2000<sup>[16]</sup>。其中每一类都包含 1 000 个样本,被分成了 700 个训练样本和 300 个测试样本。实验中选取了其中的前 100 个类中的样本,并通过提取 8-方向梯度特征得到 512 个属性。表 1 列出了每个实验数据集中包含的样本数、属性数(也称为特征数)以及类数。其中,第 1 个~第 6 个数据集为小规模数据集,其余 6 个为中等或大规模数据集。

表 1 实验数据集参数

序号	数据集	样本数	属性数	类数
1	Dermatology	358	34	6
2	Glass	214	9	6
3	Iris	150	4	3
4	Spambase	4 601	57	2
5	Vowel	990	10	11
6	WDBC	569	30	2
7	HCL2000	100 000	512	100
8	Isolet	7 797	617	26
9	Letter	20 000	16	26
10	Optdigits	5 620	64	10
11	Pendigit	10 992	16	10
12	USPS	9 298	256	10

#### 4.2 实验参数

在所有 SVM 模型的训练过程中,采用了下式定义的高斯核函数。

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \quad (1)$$

除了大类别集的数据集 HCL2000<sup>[16]</sup> 外,在每个数据集上都采用 5 重交叉验证的方法进行实验。由于原始的数据集 HCL2000 规模很大,且已被分成了训练集和测试集,这里没再对其重新划分。

在每个数据集上,对式(1)中高斯核函数的参数 $\sigma^2$ 、SVM模型的误差限 $C$ 、本文所给算法的异类近邻数 $k$ 以及NPPS算法<sup>[12]</sup>中的近邻数 $k_n$ 都通过实验

的方法进行优化。表2列出了在每个数据集上的具体参数取值。由于FCNN<sup>[13]</sup>中采用的是最近邻算法,因此不需要对其进行参数设置。

表2 在每个数据集上的参数取值

数据集	$\sigma^2$	$C$	$k$	$k_n$	数据集	$\sigma^2$	$C$	$k$	$k_n$
Dermatology	1	20	5	8	HCL2000	1	10	3	50
Glass	0.1	0.1	5	6	Isolet	4	10	7	9
Iris	1	5	5	15	Letter	0.15	40	9	15
Spambase	1	15	7	8	Optdigits	0.5	5	5	50
Vowel	0.2	20	3	9	Pendigit	1	15	15	55
WDBC	0.1	8	10	75	USPS	1	7	4	50

### 4.3 实验结果

表3列出了在6个小规模数据集上的分类准确率、样本选择比例及训练时间。表4则给出了在其余6个中等或大规模数据集上的实验结果。为了直

观地对所给算法与FCNN以及NPPS进行比较,图2~图7用折线图描绘了它们分别在2组数据集上的分类准确率、样本选择比例及样本选择后的训练时间与未进行样本选择的训练时间之比。

表3 在小规模数据集上的实验结果

算法	比较项	Dermat	Glass	Iris	Spam	Vowel	WDBC
无选择	分类准确率/%	100	69.57	100	74.70	99.52	93.91
	样本选择比例/%	0	0	0	0	0	0
	训练时间/s	0.20	0.17	0.03	44.06	5.98	0.17
FCNN 算法	分类准确率/%	92.11	56.52	90.91	71.34	95.69	79.13
	样本选择比例/%	0.33	0.45	0.15	0.37	0.25	0.15
	训练时间/s	0.11	0.03	0.03	7.36	0.48	0.03
本文算法	分类准确率/%	97.37	69.57	100	73.51	99.52	86.96
	样本选择比例/%	0.70	0.59	0.31	0.69	0.47	0.30
	训练时间/s	0.17	0.11	0.01	27.63	1.89	0.06
NPPS 算法	分类准确率/%	94.74	58.70	63.64	74.05	96.65	89.57
	样本选择比例/%	0.55	0.63	0.32	0.64	0.78	0.60
	训练时间/s	0.14	0.08	0.01	26.53	3.59	0.17

表4 在中等或大规模数据集上的实验结果

算法	比较项	HCL2000	Isolet	Letter	Optdigit	Pendgit	USPS
无选择	分类准确率/%	99.29	96.34	97.97	98.94	98.80	95.81
	样本选择比例/%	0	0	0	0	0	0
	训练时间/s	25 173.94	290.97	2 139.5	65.03	36.22	197.78
FCNN 算法	分类准确率/%	99.02	95.77	95.67	97.61	96.80	94.67
	样本选择比例/%	0.07	0.30	0.18	0.09	0.05	0.12
	训练时间/s	7 616.83	76.80	228.06	4.23	2.06	14.53
本文算法	分类准确率/%	99.31	96.15	97.58	99.00	98.86	95.71
	样本选择比例/%	0.09	0.45	0.43	0.36	0.32	0.32
	训练时间/s	7 005.59	144.31	830.59	27.08	22.30	64.44
NPPS 算法	分类准确率/%	98.28	94.74	86.42	97.83	98.34	95.57
	样本选择比例/%	0.17	0.55	0.47	0.41	0.29	0.47
	训练时间/s	3 378.56	116.86	713.03	25.22	12.36	78.49

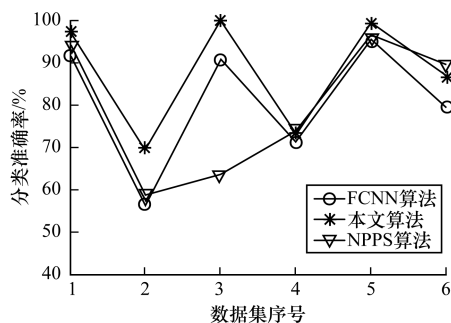


图2 小规模数据集上的分类精度

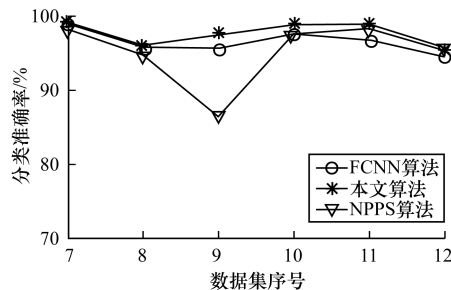


图3 中大规模数据集上的分类精度

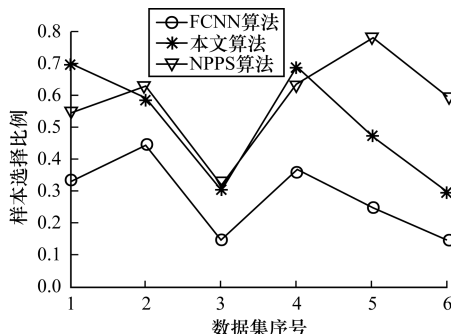


图4 小规模数据集上的样本选择比例

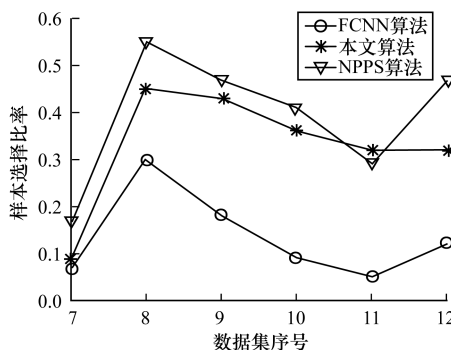


图5 中大规模数据集上的样本选择比例

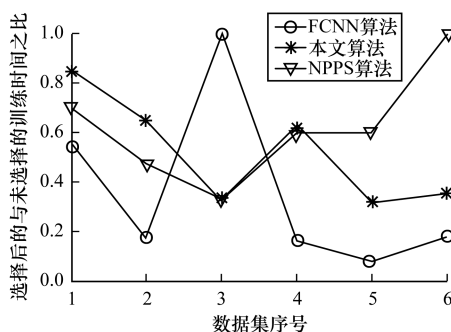


图6 小规模数据集上的训练时间比

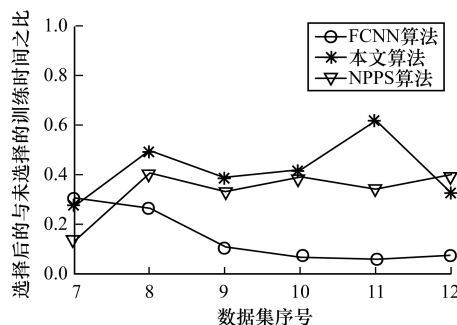


图7 中大规模数据集上的训练时间比

由图2和图3可以直观地发现,在大部分小规模数据集和所有中等规模与大规模数据集上,本文提出的样本选择算法在提高了SVM的训练效率的同时,使分类取得了最好的效果。从表3和表4可以发现,在实验的每个数据集上,本文算法的分类效果没有明显的降低。甚至在HCL2000、Optdigit以及Pendigit等数据集上分类准确率比未进行样本选择的情况下还有所提高。相比之下,FCNN算法与NPPS算法都没有很好地保持分类效果。例如在Dermat、Glass、Iris、WDBC等数据集上,FCNN使得分类精度明显下降。NPPS在Glass、Iris、Letter等数据集上引起了分类效果的明显降低。上述实验结果表明,本文给出的选择算法在选择效果上有着明显的优势。

由图4和图5可以发现,本文所给算法选择的样本的比例总体上略低于NPPS算法的选择比例而高于FCNN算法。在比较的样本选择算法中,FCNN算法选择的样本比例最低,在降低训练数据规模上具有一定优势。但由上文对选择效果的比较来看,FCNN算法容易引起分类效果明显降低。这说明选择的样本并不是越少越好。

从图6和图7综合来看,在提高训练效率方面本文所给算法和NPPS算法相当,而FCNN算法在这方面具有一定优势。这同样是由于FCNN算法选择的样本比例较低缘故。

## 5 结束语

在众多的分类算法中,支持向量机(SVM)因其较高的分类精度和坚实的理论基础而倍受关注,在诸多的应用领域中取得了显著的分类效果。然而,SVM对于样本数目来说,具有立方级的训练复杂度。因此,如何提高SVM的训练效率一直是机器学习研究中的热点问题。

为了在保持SVM效果的同时提高其效率,本文给出了精简训练集的一种算法,即利用异类近邻来选择边界样本。这一算法不仅适用于复杂边界的情形,也可以有效地用于多类分类问题,而且能在很大程度上减轻不均衡数据对分类模型的影响。在多个实验数据集上的实验结果表明,该算法在使训练效

率大幅提高的同时,能更好地保持甚至改善分类效果。

为了使所给算法能更有效地用于大规模数据集,特别是大数据处理,还需要对算法在效率上做进一步改进,这是下一步要做的一项研究内容。

### 参考文献

- [1] VAPNIK V. The Nature of statistical learning theory[M]. New York, USA: Springer, 1995.
- [2] JOACHIMS T. Text categorization with support vector machine; learning with many relevant features [C]//Proceedings of the 10th European Conference on Machine Learning. New York, USA: ACM Press, 1998: 137-142.
- [3] 王宪亮,吴志刚,杨金超,等. 基于 SVM 一对一分类的语种识别方法[J]. 清华大学学报(自然科学版), 2013, 53(6): 808-812.
- [4] 孙俊涛,张顺利,张利. 基于联合支持向量机的目标跟踪算法[J]. 计算机工程, 2017, 43(3): 266-270.
- [5] DONG Jianxiong, KRZYSAK A, SUEN C Y. Fast SVM training algorithm with decomposition on very large data sets [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(4): 603-618.
- [6] LI Boyan, WANG Qiangwei, HU Jinglu. Fast SVM training using edge detection on very large datasets[J]. IEEE Transactions on Electrical and Electronic Engineering, 2013, 8(3): 229-237.
- [7] JUNG H G. Support vector number reduction; survey and experimental evaluations novel[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 5(2): 463-476.
- [8] 包文颖,胡清华,王长忠. 基于多粒度数据压缩的支持向量机[J]. 南京大学学报(自然科学版), 2013, 49(5): 637-643.
- [9] 焦李成,张莉,周伟达. 支撑矢量预选取的中心距离比值法[J]. 电子学报, 2001, 29(3): 383-386.
- [10] 李红莲,王春花,袁保宗,等. 针对大规模训练集的支持向量机的学习策略[J]. 计算机学报, 2004, 27(5): 715-719.
- [11] PANDA N, CHANG E Y, WU Gang. Concept boundary detection for speeding up SVM [C]//Proceedings of International Conference on Machine Learning. New York, USA: ACM Press, 2006: 681-688.
- [12] SHIN H, CHO S. Neighborhood property based pattern selection for support vector machines [J]. Neural Computation, 2007, 19(3): 816-855.
- [13] ANGIULLI F, ASTORINO A. Scaling up support vector machines using nearest neighbor condensation[J]. IEEE Transactions on Neural Networks, 2010, 21(2): 351-357.
- [14] CHEN Jingnian, ZHANG Caiming, XUE Xiaoping, et al. Fast instance selection for speeding up support vector machines [J]. Knowledge-based Systems, 2013, 45(6): 1-7.
- [15] HETTICH S, BLAKE C L, MERZ C J. UCI repository of machine learning databases[EB/OL]. [2017-03-21]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [16] ZHANG Honggang, GUO Jun, CHEN Guang, et al. HCL2000—a large-scale handwritten Chinese character database for handwritten character recognition [C]//Proceedings of the 10th International Conference on Document Analysis and Recognition. Berlin, Germany: Springer, 2009: 286-289.

编辑 顾逸斐

(上接第 18 页)

- [5] 文琪,彭宏. 小波变换的离群时序数据挖掘分析[J]. 电子科技大学学报, 2005, 34(4): 556-558.
- [6] 侯澍旻. 时序数据挖掘及其在故障诊断中的应用研究[D]. 武汉: 武汉科技大学, 2006.
- [7] 侯澍旻,李友荣,刘光临. 一种基于 KS 检验的时间序列非线性检验方法[J]. 电子与信息学报, 2007, 29(4): 808-810.
- [8] WANG Yao, WU Chunguo, JI Zhaohua, et al. Non-parametric change-point method for differential gene expression detection[EB/OL]. [2017-04-14]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3104986/>.
- [9] 王小宜,卢正鼎,凌贺飞. 一个基于小波的时序数据异常探测的新算法[J]. 计算机工程与科学, 2005, 27(6): 83-85.
- [10] LIN H D. Automated visual inspection of ripple defects using wavelet characteristic based multivariate statistical approach [J]. Image and Vision Computing, 2007, 25(1): 1785-1801.
- [11] 刘丹红,张世英. 基于小波神经网络的非线性误差校正模型及其预测[J]. 控制与决策, 2006, 21(10): 1114-1118.
- [12] LIN J, KEOGH E, LONARDI S, et al. A symbolic representation of time series, with implications for streaming algorithms[C]//Proceedings of ACM Sigmod Workshop on Research Issues in Data Mining & Knowledge Discovery. New York, USA: ACM Press, 2003: 2-13.
- [13] 钟清流,蔡自兴. 基于统计特征的时序数据符号化算法[J]. 计算机学报, 2008, 31(10): 1857-1864.
- [14] SHARIFZADEH M, AZMOODEH F, SHAHABI C. Change detection in time series data using wavelet footprints[C]//Proceedings of International Symposium on Spatial and Temporal Databases. Berlin, Germany: Springer, 2005: 127-144.
- [15] BRODSKY B E, DARKHOVSKY B S. Nonparametric methods in change point problem[M]. Berlin, Germany: Springer, 1993.
- [16] ALARCON-AQUINO V, BARRIA J A. Anomaly detection in communication networks using wavelets[J]. IEEE Proceedings Communications, 2002, 148(6): 355-362.

编辑 顾逸斐