

序贯散列近邻法及其在光谱识别中的应用

李乡儒

(华南师范大学 数学科学学院, 广东 广州 510631)

摘要:基于近邻的方法是海量光谱数据获取、自动处理和挖掘中的一类重要方法,在应用中它们的主要问题是效率较低,为此文中提出了基于序贯计算的散列近邻法(SHNN)。在SHNN中,首先使用PCA方法对光谱数据进行正交变换,使数据按照各成分的散列能力进行组织;然后在PCA空间中快速查找待识别光谱的近邻数据,在此过程中通过散列思想快速约减搜索空间,并用序贯计算法高效地排除非近邻光谱数据,提高计算效率。文中主要贡献是,提出了SHNN算法,并研究了该算法在恒星光谱、正常星系光谱和类星体光谱识别中的应用。SDSS光谱实验研究表明,SHNN效率提高约96%以上,速度提高26.45倍以上。由于近邻法的广泛适用性,文中研究结果不仅对恒星光谱、正常星系光谱和类星体光谱的识别研究有重要的参考意义和一定的实用价值,亦对恒星大气参数的估计和基于光谱的红移测量有一定的参考意义。

关键词:光谱数据挖掘;光谱识别;高效计算

中图分类号:TN911.7

文献标识码:A

文章编号:1673-629X(2012)06-0182-05

Sequential Computation-based Hash Nearest Neighbor Algorithm and Its Application in Spectrum Classification

LI Xiang-ru

(School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China)

Abstract: The nearest neighbor (NN) method is one of the most typical methods in spectral retrieval, automatic processing and data mining. The main problem in NN is the low efficiency. Therefore, focus on the efficient implementation problem and introduce a novel and efficient algorithm SHNN (sequential computation-based hash nearest neighbor algorithm). In algorithm SHNN, firstly, decompose and recognize the spectrum flux components based on their hashing power; Secondly, the nearest neighbor is computed in PCA space based on sequential computation idea. In the second procedure, the putative nearest spectra can be reduced based on hash idea, and the un-nearest spectra can be rejected as early as possible. The contributions of this work are: 1) a novel algorithm SHNN is introduced, which improve the efficiency of the most popular spectramining method nearest neighbor significantly; 2) Its application in star spectrum, normal galaxy spectrum and Qso spectrum classification is investigated. Evaluated the efficiency of the proposed algorithms experimentally on the SDSS (Sloan Digital Sky Survey) released spectra. The experimental results show that the proposed SHNN algorithm improves the efficiency of nearest neighbor method more than 96%. The nearest neighbor is one of the most popular and typical methods in spectra mining. Therefore, this work is useful in a wide scenario of automatic spectra analysis, for example, spectra classification, spectra parameter estimation, redshift estimation based on spectra, etc.

Key words: spectral data mining; spectrum classification; efficient computation

1 引光谱数据挖掘的原型法

基于近邻的方法是海量光谱数据的获取、自动处理中的一类重要方法,关于他的应用已经得到了广泛研究^[1]。例如,类星体测光红移估计^[2],测光红移概率密度函数估计^[3],恒星光谱分类^[4],基于光学、X射线

和红外波段的AGN天体识别^[5],类星体与正常星系光谱分类^[6],恒星、星系、类星体光谱分类^[7],恒星大气参数的估计^[8]和恒星族参数估计^[9]等。在应用中该方法的主要问题是效率较低。为此,文中结合恒星、正常星系、类星体光谱的分类研究了最近邻方法的高效计算,并提出了基于序贯计算的散列近邻法(sequential computation-based hash nearest neighbor algorithm, SHNN)。该研究既有一定的应用价值,亦有重要的参考意义。

2 快速计算解决方案

假设有 n 个天体的观测光谱及其类型数据 $S =$

收稿日期:2011-10-31;修回日期:2012-02-03

基金项目:国家自然科学基金(61075033);模式识别国家重点实验室开放课题(201001060);广东省自然科学基金(S2011010003348);华南师范大学教学改革项目(2009jg28)

作者简介:李乡儒(1972-),男,博士,副教授,主要研究领域为计算机视觉、模式识别及其应用。

$\{(x_i, y_i), i = 1, \dots, n\}$, 其中 $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,m})^T$ 是第 i 个天体目标的观测光谱, y_i 是天体目标的某个感兴趣的属性值, 例如, 类型、红移或恒星大气参数等。对于属性的观测光谱 x_Δ , 最近邻规则是: 如果

$$d(x_{i_0}, x_\Delta) = \min\{d(x_i, x_\Delta), i = 1, \dots, n\} \quad (1)$$

则按照最近邻规则, 将光谱 x_Δ 的属性判定 y_{i_0} , 其中 $d(x_{i_0}, x_\Delta)$ 表示两个观测光谱之间的距离。

在应用中, 上述 d 一般采用欧氏距离, 这时式(1)可写为

$$\sqrt{\sum_{j=1}^m (x_{i_0,j} - x_{\Delta,j})^2} = \min\left\{\sqrt{\sum_{j=1}^m (x_{i,j} - x_{\Delta,j})^2}, i = 1, \dots, n\right\} \quad (2)$$

在理论上, 式(2)等价于

$$\sum_{j=1}^m (x_{i_0,j} - x_{\Delta,j})^2 = \min\left\{\sum_{j=1}^m (x_{i,j} - x_{\Delta,j})^2, i = 1, \dots, n\right\} \quad (3)$$

在应用中, 式(3)比式(2)少 $n+1$ 次求平方根运算, 所以文中的后续探讨中均采用式(3), 它的效率更高, 实现的伪代码如图1所示, 其目标是计算一个样本

```

初始状态建立:
d02 = 0;
For j = 1:m
    d02 = d02 + (x1,j - xΔ,j)2;
End
i0 = 1;
快速遍历搜索
For i = 1:n
    d2 = 0;
    For j = 1:m
        d2 = d2 + (xi,j - xΔ,j)2;
    End
    If d2 < d02
        d02 = d2;
        i0 = i;
    End
End
d02 中存储的是 xΔ 最近邻距离的平方
i0 中存储的是与 xΔ 距离最近光谱的序号

```

图1 基于欧氏距离的传统最近邻搜索算法(TNN)
 x_Δ 在数据集 $\{x_i, i = 1, \dots, n\}$ 中最近邻。该算法的基本操作是乘法运算, 需要进行的乘法次数是

$$m \times n \quad (4)$$

其中 m 表示每条光谱的观测流量数量, n 表示观测天体光谱的数量。

随着天文观测技术的进步, 以及 2df, SDSS 和郭

守敬望远镜(LAMOST)等大型巡天计划的实施, 天体光谱的分辨率和获取速度均得到了大幅度提高, 即式(4)中的 m 和 n 都在不断增加。 m 和 n 的急剧增加分别对应于信息处理中的高维和海量问题。他们导致的结果是, 光谱数据挖掘对算法的计算速度提出了越来越高的要求。文中将结合恒星、正常星系和类星体光谱的识别问题, 从光谱样本空间和特征计算空间的自适应约简两方面入手探讨最近邻方法的快速计算。

3 流量空间的自适应约简

流量空间, 即上述特征计算空间, 其维数是光谱的流量采样数量。最近邻算法的效率与光谱分辨率成反比, 光谱分辨率在式(4)中体现为流量观测数量 m 。本节将结合分支定界^[10]思想和主成分分析(PCA, Principal Component Analysis)^[11]技术探讨怎样尽快结束式(3)中不必要的光谱流量差异评估, 提高计算效率。通过它们实现了最近邻查找中的序贯计算。

在式(3)中, 如果已经计算得到

$$d^2(x_{i_1}, x_\Delta) = \sum_{j=1}^m (x_{i_1,j} - x_{\Delta,j})^2, 1 \leq i_1 < n$$

且某个样本 x_{i_k} 满足

$$d^2 = \sum_{j=1}^k (x_{i_k,j} - x_{\Delta,j})^2 \geq d^2(x_{i_1}, x_\Delta), 1 \leq k < m$$

则不必进行 $d^2(x_{i_k}, x_\Delta)$ 的后续计算 $\sum_{j=k+1}^m (x_{i_k,j} - x_{\Delta,j})^2$ 即可断定 x_{i_k} 不可能是 x_Δ 的最近邻, 从而节省大量的计算时间, 提高算法的效率。为了便于阐述, 将经过上述优化的最近邻方法简称为分支定界最近邻法, 相应的算法如图2所示。

对于在不同方向上散列程度不一样的数据集, 如果先在数据较分散的方向上进行计算, 则能较快地按照上述方法排除非最近邻计算。例如, 在图3中, 如果已经计算了点A与点B之间的距离, 则仅通过 x 分量的计算即可断定C、D等点不可能为点A的最近邻; 但是, 如果先计算 y 分量、再计算 x 分量, 则必须完全计算出点C、点D与点A之间的距离才能正确做出他们是否可能为最近邻的判决。同时, 研究表明, 光谱数据在高维空间中的分布一般是各向异性的, 且光谱数据的区分性往往集中在少数几个方向^[12,13], 在其他的方向上则区分性非常微弱。因此, 文中建议首先对原始光谱数据进行主成分分析变换, 将数据成份进行分解、重组, 并按照数据的散列程度将各成分进行排序; 然后, 在变换空间中使用分支定界最近邻法寻找最近邻。

4 光谱样本空间的快速约简

由式(4)可知, 最近邻算法的效率与可疑最近邻

样本数量 n 成反比。在该算法的初始状态,任何一个样本均可能是光谱 x_{Δ} 的最近邻,即 n 实际上是算法初始状态时可疑最近邻样本的数量。本节将探讨如何通过 Hash 和聚类思想快速批量缩减可疑最近邻样本的数量,以进一步提高算法的速度。

```

初始状态建立:
 $d_0^2 = 0;$ 
For  $j = 1:m$ 
     $d_0^2 = d^2 + (x_{1,j} - x_{\Delta,j})^2;$ 
End
 $i_0 = 1;$ 
快速遍历搜索
For  $i = 1:n$ 
     $d^2 = 0;$ 
    For  $j = 1:m$ 
         $d^2 = d^2 + (x_{i,j} - x_{\Delta,j})^2;$ 
        If  $d^2 \geq d_0^2$ 
             $d_0^2 = d^2$ 
            跳过  $x_i$ , 继续处理  $x_{i+1}$ 
        End
    End
     $d_0^2 = d^2;$ 
     $i_0 = i;$ 
End
 $d_0^2$  中存储的是  $x_{\Delta,j}$  最近邻距离的平方
中存储的是与  $x_{\Delta,j}$  距离最近光谱的序号
    
```

图 2 基于分支定界思想的最近邻查找算法(NNBB)

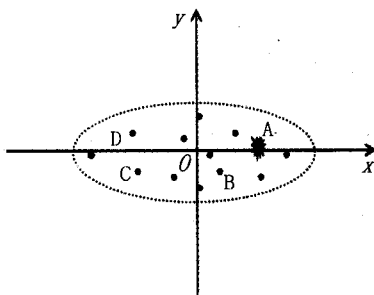


图 3 根据样本在空间中的分布特点对最近邻计算方法进行加速的示意图

对于由样本数量众多所引起的搜索算法效率下降,可通过聚类的方法对数据进行组织以提高算法的效率。例如,在图 4 中,将虚线圆中的样本作为一个聚类 S ,并用这些样本的中心 o_1 和半径 r 表征该聚类;在计算出点 A 与点 B 之间的距离后,只需要通过检查 A 与聚类 S 中心的距离满足

$$d(A, B) + r \leq d(A, o_1) \quad (5)$$

即可断定该聚类中的所有观测样本均不可能为点 A 的最近邻。

如前所述,随着天文观测技术的进步,观测光谱的分辨率越来越高。但是,在高维空间中,光谱数据一般

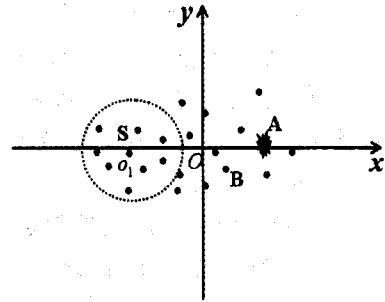


图 4 基于聚类的快速最近邻查找原理示意图

分布较为分散,这一方面会给聚类的划分造成困难,另外,各个聚类的半径亦较大,对后续快速计算不利。因此,文中工作的思路是将高维的光谱数据约简至一维空间后进行聚类 and 样本约简。该投影需要满足的条件是,

1) 离散性:样本投影到该空间后需要尽可能离散。即,在该空间中能够较好地地区分不同的光谱数据,这是样本约简的必要前提。

2) 连续性:距离较近的光谱投影后亦较近,反之,投影后距离较远的数据,原始光谱之间的距离亦较远。

该投影约简样本空间方法与 Hash 方法^[14]的思想既有类似之处,亦有很大区别。类似之处是:它们均是空间换时间的解决方案,通过改变数据的组织,实现更快的数据查找;不同之处在于:1) Hash 法是精确匹配查找,而文中方法是最相似查找;2) Hash 法对连续性无要求。上述算法如图 5 所示。

```

1) 对训练数据  $S = \{x_i, i = 1, \dots, n\}$  进行 PCA 分析,变换后的数据集记为  $S' = \{z_i, i = 1, \dots, n\}; x_{\Delta}$ ; 变换为  $z_{\Delta}$ 。
2) 对数据集  $S'$  在第一主成分方向上进行聚类,各个聚类分别记为  $S_1, \dots, S_k$ ,他们的中心和半径分别记为  $(x_{o_1}, r_1), \dots, (x_{o_k}, r_k)$ ,其中  $k$  是一个给定的正整数。
3) 按照聚类中心到样本  $z_{\Delta}$  的距离从小到大对  $S_1, \dots, S_k$  进行排序,排序后记为  $S_{R(1)}, \dots, S_{R(k)}$ 
4) 在第一主成分空间中按照与  $z_{\Delta}$  的距离由小到大,对  $S_{R(1)}$  中的样本排序,得到  $\mathcal{S}_{R(1)}$ 。
5) 运用表 2 中算法从  $\mathcal{S}_{R(1)}$  中找  $z_{\Delta}$  的最近邻,结果存至  $d_0^2$  和  $i_0$  中。
6) 进一步遍历查找
For  $i = 2:k$ 
    If  $d(z_{\Delta}, x_{o_{R(i)}}) + r_{R(i)} > d_0$ 
        以  $d_0^2$  和  $i_0$  为初始状态,运用表 2 中算法从  $\mathcal{S}_{R(i)}$  中找  $z_{\Delta}$  的最近邻,结果存至  $d_0^2$  和  $i_0$  中。
    End
 $d_0^2$  中存储的是  $x_{\Delta,j}$  最近邻距离的平方
 $i_0$  中存储的是与  $x_{\Delta,i}$  距离最近光谱的序号
    
```

图 5 基于序贯思想的最近邻计算实现算法(SHNN)

5 实验比较与探讨

下面研究最近邻方法在恒星、正常星系和类星体光谱识别中的应用效果,以及文中方法对计算效率的提高情况。由于文中算法包括样本空间的约简和流量空间的约简两个核心成分,所以从算法的整体效率、流量空间约简对效率的影响和样本空间约简对效率的影响这三方面进行评估。

5.1 恒星、正常星系和类星体光谱的识别

文中研究使用的是 Sloan 发布的一维光谱数据^[15],恒星、正常星系和类星体光谱各 4000 条,它们所在的天区编号分别是 2701–2974,0267–0276,0267–0389,波长范围截取为 3800 Å–9000 Å。由于研究表明,在光谱识别中采用对数波长–流量数据格式有较好的效果^[12,13],所以本研究采用该数据格式,并均匀采样 3791 个流量值。为了消除同一类天体观测光谱之间流量数量级的不确定性,采用 S_{median} 方法进行光谱流量标准化^[12,13]。

在基于最近邻的天体光谱识别研究中,需要将数据集集合分为两个子集:训练集和测试集,为了便于后文阐述,两个子集的样本数分别记录为 n 和 n' 。在文中研究中,从上述恒星、正常星系和类星体数据中分别随机选择 $n/3$ 条光谱作为训练数据,剩余的的光谱作为测试数据。对于不同规模的训练数据集选择 $n = 10500, 9000, 7500, 6000$,相应的正确识别率分别为 91.40%, 90.23%, 90.51%, 90.98%。如果将文中算法扩展到 k 近邻,并添加拒绝策略则能进一步提高算法的识别准确率,并使算法具有特殊天体发现能力,由于这不是文中工作的重点,所以在此不作进一步探讨。

5.2 算法效率研究

效率评估方法。在算法效率的评估中需要首先考虑的是算法运行时间的度量单位。常用的小时、分、秒等度量方法的明显缺陷是最终的结果依赖于所使用的计算机和算法程序实现的质量。文中采用基本操作作为时间的度量单位。基本操作是对算法总运行时间贡献最大的算法运算。由于使用中只需统计算法实现中需要进行的基本操作数量,他既不依赖于所使用的计算机、算法实现的质量,也不依赖于算法程序的语言和编译器,因此,在算法效率研究中,该方法比常规的时间单位更客观、通用。

算法 SHNN 的效率。如图 1,图 2,图 5 中的算法所示,在最近邻的搜索中,乘法运算重复的次数最多,在最内层循环中,所以,该问题的基本操作是乘法。假设训练数据集 S 中有 n 条光谱,且每条光谱有 m 个观测流量,则由式(4)知,在传统算法 TNN 中,每搜索一条光谱 $x = (x_1, x_2, \dots, x_m)^T$ 在 S 中的最近邻需要进行 $m \times n$ 次乘法运算,文中研究中取 $m = 3791, n = 10500$,

9000,7500,6000,相应的基本操作次数请见表 1 的 ON (TNN)行。在文中算法 SHNN,对于不同的输入样本 x ,所需要的计算量一般亦是不同的,难于进行严格的数学分析,因此,运用经验分析法,即通过大量输入,统计平均意义上查找一个最近邻所需要的基本操作乘法次数,具体结果请见表 1 的 MBON (SHNN)行。由表 1 中的结果可见,文中方法大大提高了算法的计算效率,提高幅度高达 96.73%,这对于海量光谱数据的及时处理有重要的意义。

在表 1 中, n 表示训练集中的样本数量; MBON (SHNN)和 MBON (NNBB) 分别表示是 SHNN 算法和 NNBB 算法中需要的平均基本操作次数; ON (TNN)表示 TNN 算法需要的基本操作次数, $TR<SHNN, TNN>$ 和 $TR<NNBB, TNN>$ 分别是文中算法 SHNN 和算法 NNBB 对传统方法 TNN 效率的提高百分比; $TR<SHNN, NNBB>$ 代表 SHNN 算法对 NNBB 算法的效率的提高百分比。

表 1 SHNN 算法和 NNBB 算法的效率比较

n	10500	9000	7500	6000
ON (TNN)	39805500	34119000	28432500	22746000
MBON (SHNN)	1300208	1118145	981008	828369
TR<SHNN, TNN> (%)	96.73	96.72	96.55	96.36
MBON (NNBB)	5403063	4612800	3871330	3142052
TR<NNBB, TNN> (%)	86.43	86.48	86.38	86.19
TR<SHNN, NNBB> (%)	75.93	75.76	74.66	73.64

算法中不同成分的评估。在文中算法 SHNN 中,包括样本空间的约简和流量空间的约减,核心技术环节是分支定界和 PCA 数据重整,下面将分别评价这两个环节对效率提高的贡献。研究方法仍然是经验分析法和数学分析法相结合。分析结果如表 1。表 1 实验研究的是分支定界技术对最近邻查找效率的提高,结果表明,在文中研究的情况中,计算效率均提高了 86% 以上。表 1 结果则表明,通过 PCA 数据重整,能在此基础上进一步提高 73% 以上。因此,这两个技术环节在近邻的高效计算中都是不可或缺的。

6 结束语

文中研究了高维空间中最近邻的快速计算问题。快速计算的解决方案是:首先,通过 PCA 方法对数据成分进行分解、重组,并按照各成分对判断样本是否为最近邻的贡献率排序,使算法可通过少数几个成分的计算排除大批候选样本,以节省计算复杂度,提高运算速度;然后,通过采用散列 (Hash) 思想对训练数据进行有效地组织和存储,实现对近邻候选样本的批量排除,从而进一步大幅提高最近邻计算的速度。另外,文中工作亦结合天文数据挖掘中的恒星、正常星系和类

星体光谱的识别问题,通过实验验证了所给算法的有效性,结果表明文中方法可将计算复杂度降低 96.36% 以上,速度提高 26.45 倍以上。由于最近邻的计算是模式识别、机器学习、计算机视觉相关算法中的一个基本环节,所以,文中研究不仅对天文光谱数据的挖掘有重要的应用价值,在文档和图像检索、视觉目标识别、生物特征识别和异常检测中众多应用中均有一定的参考意义。

参考文献:

- [1] Ball N M, Brunner R J. Data Mining and Machine Learning in Astronomy [J]. International Journal of Modern Physics D (IJMPD), 2010, 19(7): 1049-1106.
- [2] Ball N M, Brunner R J, Myers A D, et al. Robust Machine Learning Applied to Astronomical Datasets II: Quantifying Photometric Redshifts for Quasars Using Instance-based Learning [J]. The Astrophysical Journal, 2007, 663(2): 774-774.
- [3] Ball N M, Brunner R J, Myers A D, et al. Robust Machine Learning Applied to Astronomical Data Sets. III: Probabilistic Photometric Redshifts for Galaxies and Quasars in the SDSS and GALEX [J]. The Astrophysical Journal, 2008, 683(1): 12-12.
- [4] 张健楠, 赵永恒, 刘 蓉. 基于非参数回归与最近邻方法的恒星光谱自动分类 [J]. 光谱学与光谱分析, 2009, 29(12): 3424-3428.
- [5] Li Lili, Zhang Yanxia, Zhao Yongheng. K-nearest neighbors for automated classification of celestial objects [J]. Science in China Series G: Physics, Mechanics & Astronomy, 2008, 51(7): 916-916.
- [6] 李乡儒, 卢 瑜, 周建明, 等. 基于最近邻方法的类星体与正常星系光谱分类 [J]. 光谱学与光谱分析, 2011, 31(9): 2582-2585.
- [7] Duan Fuqing, Liu Rong, Guo Ping, et al. Automated spectral classification using template matching [J]. Research in astronomy and astrophysics, 2009, 9(3): 341-348.
- [8] Ramirez J F, Fuentes O, Gulati R K. Prediction of stellar atmospheric parameters using instance-based learning and genetic algorithms [J]. Experimental Astronomy, 2001, 12(3): 163-178.
- [9] Solorio T, Fuentes O, Terlevich R, et al. An active instance-based machine learning method for stellar population studies [J]. Monthly Notices of the Royal Astronomical Society, 2005, 363(2): 543-554.
- [10] Thakoor N, Gao J. Branch-and-bound for Model Selection and Its Computational Complexity [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(5): 655-668.
- [11] Ulfarsson M O, Solo V. Vector Sparse Variable PCA [J]. IEEE Transactions on Signal Processing, 2011, 59(5): 1949-1958.
- [12] Li Xiangru, Hu Zhanyi, Zhao Yongheng, et al. Spectral Preprocessing and Its Effect on Galaxy/Quasar Classification [J]. Chinese Astronomy and Astrophysics, 2008, 32(1): 13-22.
- [13] 李乡儒, 胡占义, 赵永恒, 等. 光谱预处理及其对星系/类星体分类结果的影响 [J]. 天文学报, 2007, 48(3): 280-288.
- [14] Dungan K E, Potter L C. Classifying Vehicles in Wide-angle Radar Using Pyramid Match Hashing [J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(3): 577-591.
- [15] Abazajian K N, Adelman-McCarthy J K, Agüeros M A. The seventh data release of the sloan digital sky survey [J]. The Astrophysical Journal Supplement, 2009, 182(2): 543-558.

(上接第 181 页)

故障特征,而且达到了理想的诊断精度。

参考文献:

- [1] 陈 洁, 芮延年. 基于模糊理论液压系统故障诊断方法的研究 [J]. 机电工程, 2003, 20(6): 67-70.
- [2] 吕卫永, 熊诗波, 林 选, 等. 基于小波包与 EMD 处理的滚动轴承故障诊断 [J]. 太原理工大学学报, 2010, 41(2): 178-182.
- [3] 王国鹏, 翟永杰, 王东风, 等. 模糊支持向量机在汽轮机故障诊断中的应用 [J]. 华北电力大学学报, 2003, 30(4): 47-50.
- [4] 李 刚, 张来斌, 段礼祥. 基于 EMD 与模糊聚类的发动机故障诊断方法 [J]. 石油机械, 2010, 38(7): 49-53.
- [5] Davide A. Improved Neural Network for SVM Learning [J]. IEEE Transactions on Neural Networks, 2002, 13(5): 1243-1244.
- [6] 施其权, 李小明. 一类新型快速模糊支持向量机 [J]. 计算机技术与发展, 2010, 20(2): 103-105.
- [7] 杨 斌, 路 游. 基于统计学理论的支持向量机的分类方法 [J]. 计算机技术与发展, 2006, 16(11): 56-58.
- [8] 陈家德, 吴小俊. 自适应模糊支持向量机算法研究 [J]. 计算机工程与应用, 2010, 46(25): 191-194.
- [9] Xu Tu, He Dake. Theory of Hyper Sphere Multiclass SVM [J]. Control Theory and Applications, 2009, 26(11): 1293-1297.
- [10] Ma Junshui. SVM Training with Duplicated Samples and Its Application in SVM-based Ensemble Methods [J]. Neurocomputing, 2004, 61(1-4): 455-459.
- [11] 张海波, 陈国军. 液压系统压力数据采集装置的研制 [J]. 液压与气动, 2003, 16(3): 8-9.
- [12] 凌海峰, 高亚鹏, 韩 军, 等. 工程车辆液压系统检测仪的关键技术及实现 [J]. 液压与气动, 2006, 5(2): 24-25.