

文章编号: 1001-0920(2013)04-0506-05

## 基于双重支持向量机的网络故障诊断

温祥西, 孟相如, 马志强

(空军工程大学信息与导航学院, 西安 710077)

**摘要:** 为了提高网络故障诊断模型的建立和诊断速度, 提出一种双重支持向量机 (Dual-SVM) 算法. 通过两次 SVM 训练确定诊断模型: 第1次 SVM 训练通过两类样本的类中心和样本点在类中心连线方向上的投影分布情况直接确定近似最优分类面; 第2次 SVM 训练根据第1次的训练结果选择边界样本点并计算其模糊隶属度, 通过模糊 SVM 获得诊断模型. 在 DARPA 数据集上的实验表明, Dual-SVM 相较 SVM 能够获得更快的训练速度和更加简单的故障诊断模型.

**关键词:** 网络故障诊断; 支持向量机; 样本选取; 双重训练

**中图分类号:** TP393

**文献标志码:** A

## Network fault diagnosis based on Dual-SVM

WEN Xiang-xi, MENG Xiang-ru, MA Zhi-qiang

(Information and Navigation Institute, Air Force Engineering University, Xi'an 710077, China. Correspondent: WEN Xiang-xi, E-mail: wxxajy@163.com)

**Abstract:** The dual support vector machine (Dual-SVM) is proposed to promote the speed of establishing model and diagnosing. The diagnosis model is established by two SVM training processes. In the first process, the approximate classifying hyperplane is directly obtained by the two classes centers and the distribution of the samples on the connecting direction of centers. In the second fuzzy SVM process, the boundary samples are selected, fuzzy memberships are calculated, and the diagnosis model is established. The experiments on DARPA data-sets show that the Dual-SVM can get higher training speed and more simplified model compared to SVM.

**Key words:** network fault diagnosis; SVM; training set selection; dual training

### 0 引言

网络在日常生活以及军事中的地位越来越重要, 保证网络通信的正常和可靠已成为当前经济、国防和 社会的迫切需求. 近年来, 人们对网络故障诊断展开了大量的研究<sup>[1-3]</sup>. 网络故障诊断的本质是模式识别问题<sup>[1]</sup>. 支持向量机 (SVM) 作为一种先进的模式识别方法, 已初步应用于网络故障诊断<sup>[2-3]</sup>. 随着网络规模的扩大, 网管系统实时采集到的数据量也不断增大, SVM 训练时间随着样本的增大而急剧增加, 诊断模型更加复杂, 从而影响诊断的时效性. 可见 SVM 方法在网络故障诊断中应用的主要问题是训练和测试效率较低.

当前提高 SVM 训练效率的方法主要有 3 种: 1) 快速求解二次规划问题, 典型方法有 Chunking 算法、

Decomposition 算法、序贯最小优化 (SMO) 算法、双支持向量机 (TWSVM)<sup>[4]</sup> 以及采用 PSO 等智能搜索算法求解<sup>[5]</sup> 等; 2) 避免求解二次规划问题, 典型的有临近支持向量机 (PSVM)<sup>[6]</sup> 和最小二乘支持向量机 (LSSVM)<sup>[7]</sup>; 3) 减少训练样本个数, 典型的有 RSVM<sup>[8]</sup> 以及文献 [9-10] 提出的采用预选支持向量的方法等. 本文选择最后一种思路, 采用预选边界样本作为训练样本集的方式来提高 SVM 的训练效率. 这种方法还有一个优点是能够对所得模型进行稀疏化, 简化诊断模型. 现有预选样本的方法都是通过阈值比较来确定的, 例如文献 [10] 依据  $k$  近邻中两类样本个数的比值进行样本选取, 比值的设定决定了所选样本集的精度. 但是这些文献并没有给出阈值的选取方法, 影响了其在现实中的应用.

收稿日期: 2011-12-19; 修回日期: 2012-03-22.

基金项目: 国家自然科学基金项目 (61003252); 全军军事学研究生课题 (2011JY002-524); 空军工程大学创新基金项目 (201105).

作者简介: 温祥西 (1984-), 男, 博士生, 从事网络故障预测与智能管理的研究; 孟相如 (1963-), 男, 教授, 博士生导师, 从事宽带通信网络技术的研究.

为了提高SVM算法的训练和测试速度, 满足网络故障诊断中应用的要求, 本文提出一种双重支持向量机 (Dual-SVM) 算法, 通过两次 SVM 训练确定最终的诊断模型. 初次训练为第2次训练确定训练样本; 第2次训练简化诊断模型. 通过在 DARPA 数据集上的实验验证了本文方法在网络故障诊断中的高效性.

## 1 Dual-SVM

### 1.1 初次快速SVM训练

Dual-SVM 方法旨在减少 SVM 的训练时间和测试时间. 初次 SVM 主要是寻找边界点, 为第2次训练确定训练样本, 因此它必须具有很快的速度. 本文提出一种非常简单的 SVM 训练方法, 不需要求解二次规划问题, 也不需要复杂的矩阵运算, 直接确定近似最优分类面  $H_f$ .

首先引入样本类中心的概念. 在  $N$  维空间中, 假设正负样本为独立分布的, 其类别中心分别记为  $M_+$  和  $M_-$ . 对于线性可分问题, 若已知样本向量组  $\{x_1, x_2, \dots, x_l\}$ , 其中正类样本数为  $l_+$ , 负类样本数为  $l_-$ , 则它们各自类中心为

$$M_+ = \frac{1}{l_+} \sum_{i_+=1}^{l_+} x_{i_+}, \quad M_- = \frac{1}{l_-} \sum_{i_-=1}^{l_-} x_{i_-}. \quad (1)$$

**定理 1** 若使样本点到各自超平面  $H_1$  和  $H_2$  距离  $\xi_i$  的平方和最小, 则  $H_1$ 、 $H_2$  必分别过两类的类中心.

**证明** 设超平面的法向量为  $w$ , 则有

$$H_1 : wx + c = 0, \quad H_2 : wx + d = 0. \quad (2)$$

其中:  $w$  为任意方向, 可以认为已知;  $c$  和  $d$  为变量, 代表同一法向的一族直线. 由几何知识知, 对于正类样本, 有

$$\xi_i^2 = (wx_i + c)^2 / \|w\|;$$

对于负类样本, 有

$$\xi_i^2 = (wx_i + d)^2 / \|w\|.$$

则原问题的数学化表示为

$$\min \sum_{i=1}^l \xi_i^2 = \sum_{i_+=1}^{l_+} \xi_{i_+}^2 + \sum_{i_-=1}^{l_-} \xi_{i_-}^2 = \sum_{i_+=1}^{l_+} \frac{(wx_{i_+} + c)^2}{\|w\|} + \sum_{i_-=1}^{l_-} \frac{(wx_{i_-} + d)^2}{\|w\|} = f(c, d), \quad (3)$$

其中  $x_i$  为已知样本. 若要得到式(3)的最小值, 则需对  $c$  和  $d$  求偏导, 即

$$\begin{cases} \frac{\partial f(c, d)}{\partial c} = \frac{2}{\|w\|} \sum_{i_+=1}^{l_+} (wx_{i_+} + c) = 0, \\ \frac{\partial f(c, d)}{\partial d} = \frac{2}{\|w\|} \sum_{i_-=1}^{l_-} (wx_{i_-} + d) = 0. \end{cases} \quad (4)$$

因  $w$  为已知, 故  $2/\|w\|$  为常数. 于是式(4)等价于

$$\begin{cases} \sum_{i_+=1}^{l_+} (wx_{i_+} + c) = \sum_{i_+=1}^{l_+} wx_{i_+} + l_+c = 0, \\ \sum_{i_-=1}^{l_-} (wx_{i_-} + d) = \sum_{i_-=1}^{l_-} wx_{i_-} + l_-d = 0, \end{cases} \quad (5)$$

$$c = -w \frac{1}{l_+} \sum_{i_+=1}^{l_+} x_{i_+}, \quad d = -w \frac{1}{l_-} \sum_{i_-=1}^{l_-} x_{i_-}. \quad (6)$$

将  $c$  和  $d$  分别代入  $H_1$ 、 $H_2$  中, 有

$$\begin{cases} H_1 : wx + c = w \left( x - \frac{1}{l_+} \sum_{i_+=1}^{l_+} x_{i_+} \right) = 0, \\ H_2 : wx + d = w \left( x - \frac{1}{l_-} \sum_{i_-=1}^{l_-} x_{i_-} \right) = 0. \end{cases} \quad (7)$$

因此, 当  $x = \frac{1}{l_+} \sum_{i_+=1}^{l_+} x_{i_+}$  时,  $w \left( x - \frac{1}{l_+} \sum_{i_+=1}^{l_+} x_{i_+} \right) \equiv 0$ ; 当  $x = \frac{1}{l_-} \sum_{i_-=1}^{l_-} x_{i_-}$  时,  $w \left( x - \frac{1}{l_-} \sum_{i_-=1}^{l_-} x_{i_-} \right) \equiv 0$ . 即  $H_1$ 、 $H_2$  必分别过  $M_+$  和  $M_-$ .  $\square$

对于非线性可分问题, 将样本  $\{x_1, x_2, \dots, x_l\}$  映射到高维空间  $\{\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)\}$ , 其中正类样本数为  $l_+$ , 负类样本数为  $l_-$ , 则它们各自在高维空间中的中心为

$$M_+ = \frac{1}{l_+} \sum_{i_+=1}^{l_+} \varphi(x_{i_+}), \quad M_- = \frac{1}{l_-} \sum_{i_-=1}^{l_-} \varphi(x_{i_-}). \quad (8)$$

类似于定理1, 可以证明在高维空间中使得  $\sum_{i=1}^l \xi_i^2$  最小的两超平面同样必过类中心.

增加结构风险最小化约束, 即两超平面的距离最大化. 则由几何公理知, 过两个点的平行面之间的最大距离即是该两点的距离, 而且超平面的法向量为两点的连线方向, 即

$$w = \frac{1}{l_+} \sum_{i_+=1}^{l_+} \varphi(x_{i_+}) - \frac{1}{l_-} \sum_{i_-=1}^{l_-} \varphi(x_{i_-}).$$

最优分类面的法向量确定之后, 还需确定参数  $b$  才能最终确定最优分类面的位置. 在 SVM 算法中, 最优分类面是在  $H_1$  和  $H_2$  中间位置, 这里有个默认的前提: 这两类样本是同分布的. 但是, 现实中的两类样本往往分布情况不同, 这会造成最优分类面的偏移, 有学者对此做过相关研究<sup>[11]</sup>. 另外, 在本文方法中, 约束条件与其他 SVM 算法中的有所不同, 因此  $b$  的确定不能简单地采用平分  $H_1$  和  $H_2$  的方式, 而要根据样本的分布情况确定. 在分类问题中, 分类信息主要集中于分类方向上, 为此, 本文提出一种依据各类样本点到各自超平面  $H_1$ 、 $H_2$  平均距离的比值来确定最优分类面的方法. 原空间可以视为向高维空间中的一种映

射, 只是此时的映射函数为  $\varphi(x) = x$ , 在下面的讨论中统一用高维空间中映射的形式表示.

对于正类样本, 由上一小节知

$$H_1 : w \left( \varphi(x) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_i) \right) = 0,$$

$x_{i_+}$  到  $H_1$  的距离  $d_{i_+}$  为

$$d_{i_+} = \frac{\left| w \left( \varphi(x_{i_+}) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_i) \right) \right|}{\|w\|}; \quad (9)$$

对于负类样本,  $x_{i_-}$  到  $H_2$  的距离  $d_{i_-}$  为

$$d_{i_-} = \frac{\left| w \left( \varphi(x_{i_-}) - \frac{1}{l_-} \sum_{i=1}^{l_-} \varphi(x_i) \right) \right|}{\|w\|}. \quad (10)$$

于是正类样本点到  $H_1$  的平均距离  $\bar{d}_+$  以及负类样本点到  $H_2$  的平均距离  $\bar{d}_-$  分别为

$$\bar{d}_+ = \frac{\sum_{i=1}^{l_+} \left| w \left( \varphi(x_{i_+}) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_i) \right) \right|}{l_+ \|w\|}, \quad (11)$$

$$\bar{d}_- = \frac{\sum_{i=1}^{l_-} \left| w \left( \varphi(x_{i_-}) - \frac{1}{l_-} \sum_{i=1}^{l_-} \varphi(x_i) \right) \right|}{l_- \|w\|}. \quad (12)$$

对类中心连线进行比例分割, 分割的比例为

$$\frac{\bar{d}_+}{\bar{d}_-} = \frac{\sum_{i=1}^{l_+} \left| w \left( \varphi(x_{i_+}) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_{i_+}) \right) \right| l_-}{\sum_{i=1}^{l_-} \left| w \left( \varphi(x_{i_-}) - \frac{1}{l_-} \sum_{i=1}^{l_-} \varphi(x_{i_-}) \right) \right| l_+}. \quad (13)$$

计算两类中心之间的距离为

$$d = \left[ \left( \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_{i_+}) - \frac{1}{l_-} \sum_{i=1}^{l_-} \varphi(x_{i_-}) \right)^T \times \left( \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_{i_+}) - \frac{1}{l_-} \sum_{i=1}^{l_-} \varphi(x_{i_-}) \right) \right]^{\frac{1}{2}} = \|w\| = \sqrt{w^T w}, \quad (14)$$

其中  $k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$  为核函数. 则最终得到的分类面为

$$f(x) = w \left( \varphi(x) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_{i_+}) \right) + \frac{\|w\| \bar{d}_+}{\bar{d}_- + \bar{d}_+} d = w \left( \varphi(x) - \frac{1}{l_+} \sum_{i=1}^{l_+} \varphi(x_{i_+}) \right) + \frac{d^2 \bar{d}_+}{\bar{d}_- + \bar{d}_+}. \quad (15)$$

这样, 利用快速 SVM 算法直接确定了  $H_f$ , 相较传统的 SVM 算法, 本算法的计算复杂度大大降低. 此次训练的目的是为了第 2 次 SVM 训练选择样本, 边界样本点集即是离最优分类面最近的点的集合. 因此, 本文选择边界样本的方法是选择离  $H_f$  最近的  $n$  个样

本点, 选择过程不涉及阈值选择. 考虑  $H_f$  近似地等同于 SVM 的分类结果, 此时选择的边界样本相较其他方法更加准确.

## 1.2 二次 SVM 训练

第 2 次 SVM 训练是以边界样本点为样本集进行训练的, 目的是用一个更加简单的模型表示分类函数, 这在本质上是对 SVM 训练所得模型的稀疏化过程. 样本的减少使得各个样本点在分类面确定中的作用增大, 与 RSVM 类似, 需要在训练时采用较大的惩罚因子. 但是这样将会增大噪声点对最优分类面的影响, 本文采用模糊 SVM 算法对不同的样本点赋予不同的模糊隶属度以提高模型的泛化性.

样本的模糊隶属度通过其到  $H_f$  的距离来确定. 这里考虑所选择的样本点集在初次训练后的分类情况: 若离  $H_f$  近且分类正确, 则令其模糊隶属度为 1; 若样本点在  $H_f$  附近但是又被错分, 则根据其到最优分类面的距离给出模糊隶属度. 具体表示如下:

$$\mu_i(d_i) = \begin{cases} 1, & y_i y'_i = 1; \\ \exp\left(-\frac{d_i^2}{2d'^2}\right), & y_i y'_i = -1. \end{cases} \quad (16)$$

其中:  $y'$  为初次分类测试得到的结果;  $d_i$  为第  $i$  个样本点到  $H_f$  的距离;  $d'$  为设定值, 正类样本取  $0.05\bar{d}_+$ , 负类样本取  $0.05\bar{d}_-$ . 若样本点错分且离最优分类面大于设定值, 则该样本点对最终分类面的确定贡献很小. 此次 SVM 训练过程为普通的求解二次规划问题, 获得的解具有稀疏性.

## 1.3 算法验证

为了验证本文方法的有效性, 采用 Benchmarks 数据集进行实验<sup>[12]</sup>. 实验环境为: Pentium(R)Dual-Core 2.7 G CPU, 2 G 内存, Windows XP 系统, Matlab 7.4.0, 选用 RBF 核函数.

**例 1** 线性可分实验. 采用 Iris 数据集中样本 1 类和 2 类的前两维作为训练样本, 得到初次的分类面  $H_f$ 、边界点以及第 2 次训练的分类面, 如图 1 所示.

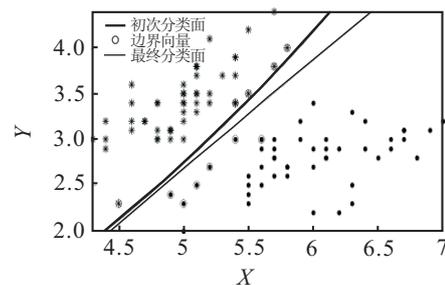


图 1 Iris 数据集分类结果

观察  $H_f$ , 它可以将 Iris 两类样本完全分开, 位置上略偏向于“\*”类样本, 这是由于“\*”类样本在法向上的离散度较小. 所选择的边界向量很好地描述了分

类边界的情况. 最后通过 FSVM 训练获得的最优分类面(所有隶属度均为 1), 同样可以将两类样本完全分开.

**例 2** 非线性可分实验. 选择 Banana 数据集, 同样经过两次 SVM 训练, 得到  $H_f$ 、边界点以及第 2 次训练所得分类面, 如图 2 所示.

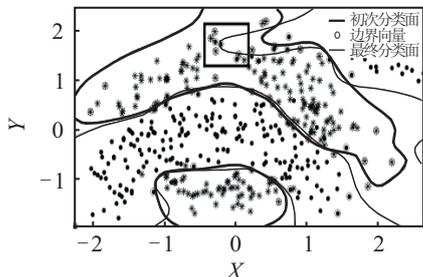


图 2 Banana 数据集分类结果

从分类结果可以看出, 在非线性的情况下,  $H_f$  同样可以较好地将两类样本区分, 所选的边界向量也准确地描述了 Banana 数据集的边界情况. 需要注意的是: 粗框内的样本点虽然在二维空间中较初次训练获得的最优分类面较远, 但在高维空间中距离  $H_f$  是较近的.

为了说明本文方法的有效性, 对 Benchmarks 中的其他数据集分别进行分类实验. 分别采用 SVM、LSSVM 以及 Dual-SVM 方法, 得到的分类结果如表 1 所示, 其中  $Tr$  和  $Te$  分别表示训练样本个数和测试样本个数.

表 1 Benchmarks 实验精度比较

数据集	Tr+Te	分类精度/%		
		SVM	LSSVM	Dual-SVM
Diabetis	468+300	75.3	73.0	76.3
Flare-solar	666+800	65.8	66.8	67.1
Image	1 300+1 010	97.9	96.6	97.3
Splice	1 000+2 175	89.1	88.2	88.7
Thyoid	1 400+750	96.0	96.0	97.3
German	2 100+3 000	79.7	79.3	79.1

从表 1 可以看出, 本文提出的方法具有很好的分类效果, 获得了与 SVM、LSSVM 相当的分精度, 在 Diabetis、Flare-solar 和 Thyoid 三个数据集上的实验结果甚至高于 SVM 和 LSSVM 方法. 这主要是因为本文方法在选择训练样本集时将一些偏离较大的野点进行了排除, 并对训练集中样本分配了不同的模糊隶属度, 从而提高了模型的泛化性.

另外, 为了说明本文方法训练的高效性和所得模型的简化性, 对 Dual-SVM 算法和 SVM 的训练时间以及测试时间进行比较. 类似于文献 [9], 本文通过训练时间比以及测试时间比这两个参数来说明本文方法在训练和测试速度上的优势, 并通过对支持向量

(SV) 个数的比较说明得到模型的简化性. 其中, 训练时间比是指 Dual-SVM 训练时间与 SVM 训练时间的比值; 测试时间比是指样本在 Dual-SVM 测试时间与在 SVM 测试时间的比值.

实验仍采用 Benchmarks 数据集, 训练和测试样本与表 1 相同, 获得的实验结果如表 2 所示.

表 2 Benchmarks 实验速度比较

数据集	SV 个数		训练时间比	测试时间比
	SVM	Dual-SVM		
Diabetis	243	124	0.105	0.51
Flare-solar	451	319	0.023	0.73
Image	336	242	0.015	0.68
Splice	800	213	0.032	0.24
Thyoid	328	94	0.013	0.28
German	689	258	0.006	0.32

从实验结果可以看出: Dual-SVM 训练得到的 SV 个数少于采用 SVM 训练所得的 SV 个数, 这表明 Dual-SVM 训练得到的分类模型较 SVM 更加简单, 相应的测试时间也更短; 从训练时间比可以看到, 虽然经过两个训练过程, 但其速度远远快于 SVM, 因此 Dual-SVM 的训练速度更快. 总之, Dual-SVM 与 SVM 和 LSSVM 等传统方法相比较, 能够获得更快的训练和测试速度, 且得到的分类模型性能与这些方法相当.

## 2 基于 Dual-SVM 的网络故障诊断

### 2.1 诊断框架

本文提出的 Dual-SVM 在不失诊断精度的前提下提高了 SVM 的训练效率, 同时得到的诊断模型简单, 诊断速度快. 基于该方法的网络故障诊断框架如图 3 所示.

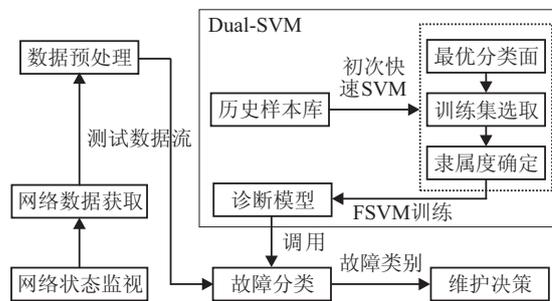


图 3 基于 Dual-SVM 的网络故障诊断框架

诊断模型的建立过程中首先采用快速 SVM 从大量的训练样本集中选择边界样本, 然后确定所选样本的隶属度, 最后通过 FSVM 得到诊断模型. 在诊断过程中, 实时地采集网络数据, 并经过数值化、归一化等处理后送入故障诊断模型, 进而得到诊断结果.

### 2.2 仿真与实验

本文选择美国国防部高级研究计划署 (DARPA) 数据集作为实验数据集, 以受攻击下的网络状态模拟

网络故障<sup>[13]</sup>. 通过对历史样本的训练获得故障诊断模型, 并用测试样本集检验所得诊断模型的诊断精度和速度. 由于本文仅研究故障的识别过程, 这里的诊断过程和测试过程是等价的. 实验所选用的训练样本和测试样本的结构如表 3 所示.

表 3 实验数据结构

数据集	训练样本/条	测试样本/条
Normal	5 000	8 000
Probe	2 700	4 500
Normal	5 000	8 000
R2L	1 000	1 600
Normal	5 000	8 000
DOS	3 000	4 800

分别用 SVM 和 Dual-SVM 对这 3 组样本进行训练和诊断, 比较两种方法获得模型的故障诊断精度、训练时间以及测试时间. 实验结果如表 4 所示.

表 4 实验结果

数据集	诊断精度/%		训练时间比	测试时间比
	SVM	Dual-SVM		
Normal Probe	99.92	100	0.023	0.294
Normal R2L	100	100	0.018	0.385
Normal DOS	100	100	0.012	0.313

从表 4 的实验结果可以看出: Dual-SVM 在训练时间上远远低于 SVM 算法, 在诊断精度上也没有任何下降; 而测试时间较 SVM 算法提高了约 2 倍, 这是由于 Dual-SVM 相较 SVM 获得的支持向量个数少, 诊断模型更加简单.

### 3 结 论

本文提出了 Dual-SVM 方法以解决 SVM 方法在网络故障诊断领域所存在的训练、诊断效率问题. 通过预选训练样本的方法提高了训练速度. 与其他样本预选方法不同, 本文通过快速 SVM 训练获得最优分类面以确定边界样本集, 避免了其他方法中样本选择阈值确定的问题. 根据所选样本到该分类面的距离确定样本点的模糊隶属度. 最终通过 FSVM 得到了诊断模型. 实验结果表明, 该方法有效提高了诊断模型建立的速度, 同时也简化了诊断模型, 获得了更快的诊断速度. 本文的研究为 SVM 在网络故障诊断中应用提供了一条可行的途径.

### 参考文献(References)

[1] Jesper Grønbaek, Hans-Peter Schwefel, Andrea Ceccarelli, et al. Improving robustness of network fault diagnosis to

uncertainty in observations[C]. IEEE Int Symposium on Network Computing and Applications. Cambridge, 2010: 229-232.

- [2] Guo Jiang-wei, Wu Xiao-ping, Ye Qing. Network fault diagnosis based on rough set-support vector machine[C]. ICCASM 2010. Taiyuan, 2010, 14: 314-315.
- [3] Zhang Li, Meng Xiang-ru, Zhou Hua. Network fault diagnosis using hierarchical SVMs based on kernel method[C]. WKDD2009. Moscow: IEEE Press, 2009: 753-756.
- [4] Jayadeva, Khemchandai R, Chandra S. Twin support vector machines for pattern classification[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910.
- [5] Tang Xian-lun, Zhuang Ling, Cai Jun, et al. Multi-fault classification based on support vector machine trained by chaos particle swarm optimization[J]. Knowledge-Based Systems, 2010, 23(5): 486-490.
- [6] Fung G, Mangasarian O L. Proximal support vector machine classifiers[C]. KDD-2001. New York: Association for Computing Machinery, 2001: 77-86.
- [7] Suykens J A K, Vandewalt J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [8] Lee Yuh-Jye, Olvi L Mangasarian. RSVM: Reduced support vector machines[C]. Proc of the 1st SIAM Int Conf on Data Mining. Chicago, 2001: 5-7.
- [9] Li Yu-hua. Selecting training points for one-class support vector machines[J]. Pattern Recognition Letters, 2011, 32(11): 1517-1522.
- [10] 韩德强, 韩崇昭, 杨艺. 基于  $k$ -最近邻的支持向量预选方法[J]. 控制与决策, 2009, 24(4): 494-498. (Han D Q, Han C Z, Yang Y. Approach for pre-extracting support vectors based on  $k$ -NN[J]. Control and Decision, 2009, 24(4): 494-498.)
- [11] 周皓, 李少洪. SVM 最优分类面相对位置的修正[J]. 北京航空航天大学学报, 2009, 35(11): 1302-1305. (Zhou H, Li S H. Relative position modification of SVM's optimal hyperplane[J]. J of Beijing University of Aeronautics and Astronautics, 2009, 35(11): 1302-1305.)
- [12] University of tuebingen benchmarks[DB/OL]. (2011-09-08). <http://ftp.tuebingen.mpg.de/pub/fml/raetschlab/benchmarks/>.
- [13] University of California Irvine. UCI KDD archive[DB/OL]. (2010-11-21). <http://kdd.ics.uci.edu/>.