

基于模糊核超球的快速分类算法在股票预测中的应用^①

龙真真¹, 张正文²

¹(北京大学 经济学院, 北京 100871)

²(国家统计局, 北京 100826)

摘要: 基于 SVM(支持向量机)的 SVDD(支持向量数据描述)分类算法存在计算复杂、分类准确率较低的缺陷, 针对股票数据非线性、高噪声的特点, 在传统的 SVDD 分类算法基础上, 模糊核超球快速分类算法(FCABFKH)通过合并法寻找超球集, 并依据最大隶属度原则构建分类器, 排除了离群点和超球集的重叠问题, 同时避免了复杂的二次规划, 具有分类速度快, 分类结果准确率高的特点. 采用中国沪市上市公司数据验证该方法的有效性, 实验结果表明, 运用 FCABFKH 算法得到的组合回报率超过了市场基准.

关键词: 支持向量机; 支持向量数据描述; 分类算法; 股票预测

Application of Fast Classification Algorithm Based on Fuzzy Kernel Hypersphere to Stock Forecasting

LONG Zhen-Zhen¹, ZHANG Zheng-Wen²

¹(School of Economics, Peking University, Beijing 10871, China)

²(National Bureau of Statistics, Beijing 10826, China)

Abstract: SVDD classification algorithm based on SVM has defects, such as high calculation complexity property and low accuracy. According to nonlinear and high-noise characteristics of stock data, inspired from the idea of traditional SVDD classification algorithm, the proposed algorithm (FCABFKH) adopts mergence method to find hypersphere sets and maximum membership degree law to construct classifier. By this means, the algorithm can rule out off-group points and hypersphere sets overlap problem. Furthermore, it can avoid complex quadratic programming. Consequently, FCABFKH provides faster rate and higher accuracy. Using the data of listed companies of China A stocks market, experiments are done to test the validity of the method mentioned above. The result indicates that portfolio's return rate using classification method of FCABFKH is higher than the market benchmark.

Key words: support vector machine; support vector data description; classification algorithm; stock forecasting

1 引言

股票市场具有高收益与高风险并存的特性, 关于股市分析与预测领域的研究一直为人们所关注. E. F. Fama 和 K. R. French^[1,2]在实证研究中发现股票的未收益在一定程度上是可预测的. 考虑到股票数据的非线性和高噪声特点, 基于 SVM(支持向量机)的分类算法近期得到了股票预测研究的关注^[3-5,13], 但由于此算法设计初衷是解决为二值分类问题, 直接应用于股票预测领域存在运算复杂度高、精确度低的缺点.

SVDD(支持向量数据描述)也是一种核化形式的分类方法^[6], 基本思想是寻找一个最小包围超球, 使其包含尽可能多的正常数据和尽可能少的异常数据, 从而实现分类应用, 相对于经典的 SVM 分类算法, 该算法具有复杂程度和对训练样本数据分布要求相对较低等优点^[7], 但仍然存在只考虑了正负两类分类, 难以妥善处理少数离群点等缺陷, 因此, 文献[8-12]提出了一些改进算法试图解决上述问题, 主要思想是为多类训练样本构造多个超球实现多类分类, 但由于没有

^① 收稿时间:2013-08-18;收到修改稿时间:2013-09-17

完全排除超球重叠和二次规划计算量大等问题, 导致分类结果准确率和算法泛化能力下降, 同时由于针对超球重叠的调整方法复杂, 增加了计算复杂度。

本文在 SVDD 算法基本思想基础上提出了模糊核超球快速分类算法(Fast Classification Algorithm based on Fuzzy Kernel Hypersphere, FCABFKH), 该算法通过合并法寻找超球集, 并依据最大隶属度原则构建分类器, 排除了离群点和超球集的重叠问题, 同时避免了复杂的二次规划, 具有分类速度快, 分类结果准确率高特点, 实验证明, 该算法为进行股票

预测从而获得超额收益的股票组合提供了一条可行的途径。

2 数据预处理

财务指标是反映和提供公司基本财务信息的重要渠道及分析判断公司财务状况的基本方法, 也是股票预测的基础。如表 1 所示, 根据 wind 终端的划分, 财务指标一般分为资产回报、财力杠杆、盈利能力、成长性、投资性、短期流动性、风险指标 7 个大类, 这 7 大类又可细分为 33 项分指标。

表 1 财务指标

资产回报	财务杠杆	盈利能力	成长性
税前利润/总资产	负债/资产净值	税前利润/销售	销售增长率
税前利润/总资本	总负债/总资本	税后利润/销售	税前收入增长率
净收入/总资本	总负债/股东净资产	净收入/销售	净经常性盈利增长率
现金流/总资产	总资产/股东总资产	现金流/销售	营业利率增长率
现金流/总资本	总资产/总市值	税后利润/股本	总资产增长率
		现金流/总市值	
投资性	短期流动性	风险指标	
每股净有形资产	流动资产/流动负债	税前利润/流动负债	
股息率	流动负债/总资产	税后利润/流动负债	
收益率	流动负债/资产净值	现金流/流动负债	
所有者权益/总市值	长期负债/总债务		

为消除每个分指标的量纲和数量级差别, 同时考虑到分类器收敛的需要, 需要对数据进行处理^[2,13]。具体步骤为: ①对原始数据进行预处理, 其中大于 95%分位的数用 95%分位的数来代替, 小于 5%分位的数用 5%分位的数来代替; ②对预处理后的数据进行均值化处理; ③计算均值化后的 p 个指标的协方差矩阵, 均值化的具体方法为用各指标的原始数据除以其均值, 即

$$z_{ij} = \frac{x_{ij}}{x_j}$$

此方法没有改变指标间的相关系数, 因此在

消除指标量纲与数量级的影响的同时, 能够保留原始数据的特征信息; ④计算相关矩阵 R 的特征根和特征向量; ⑤计算各主成分的方差贡献率及累计方差贡献率; ⑥选择主成分的个数; ⑦由主成分计算综合评分值, 并以此对被评价对象进行排序和比较。先按累计方差贡献率不低于某阈值的原则确定前 t 个主成分, 然后以所选每个主成分的方差贡献率为权数, 将其线性加权求和, 求得综合评价价值指标 F 。

3 分类器

设 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 是未知分布的训练集, 这里 $x_i \in R^n$, $y_i \in \{1, 2, \dots, p\}$ 是相应类别标签。对于样本空间的训练样本 x_i 通过非线性变换映射到高维特征空间表示为 \tilde{x}_i , \tilde{x}_i 称为 x_i 的像, x_i 称为 \tilde{x}_i 的原像。

3.1 Mercer 核

Mercer 核是 SVDD 算法的重要支撑, 如式(1)所示, 它将出现使得可以将高维 Hilbert 空间中两个点的内积计算, 转化为样本空间中的两个模式的简单核函数的求值。

$$K(x_i, x_j) = \tilde{x}_i \bullet \tilde{x}_j \tag{1}$$

三种常见的核函数为:

线性核函数: $K(x_i, x_j) = \langle x_i, x_j \rangle$

多项式核函数: $K(x_i, x_j) = (\langle x_i \bullet x_j \rangle + 1)^p$

径向基核函数: $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$

通过 Mercer 核的应用, 可以将样本 x_i 和 x_j 之间

在特征空间的欧氏距离表示为式(2):

$$\|\tilde{x}_i - \tilde{x}_j\| = \sqrt{K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j)} \quad (2)$$

由此推导出样本和关于样本集的核超球球心之间的距离表示为式(3):

$$d_{kh} = \|\tilde{x}_k - \tilde{w}_h\| = \left\| \tilde{x}_k - \frac{1}{N_h} \sum_{i \in y_i=h} \tilde{x}_i \right\|$$

$$= \sqrt{K(x_k, x_k) - \sum_{i \in y_i=h} \frac{2}{N_h} K(x_k, x_i) + \sum_{i \in y_i=h} \sum_{q \in y_q=h} \left(\frac{1}{N_h}\right)^2 K(x_i, x_q)} \quad (3)$$

值得注意的是, 当选用线性核函数时, 核超球就转化为了普通的超球, 普通超球其实就是核超球的特例.

3.2 核超球隶属函数

超球隶属函数 $b_j(\tilde{x}_i, \tilde{w}_j, \tilde{r}_j)$ 用来度量在特征空间中输入样本的像 \tilde{x}_i 与半径 \tilde{r}_j 和球心 \tilde{w}_j 构成的核超球之间的相似关系, 当 \tilde{x}_i 远离核超球时, 核超球隶属函数接近 0; 当 \tilde{x}_i 在核超球内时, 核超球隶属函数为 1. 由此定义核超球隶属函数如下:

$$b_j(\tilde{x}_i, \tilde{w}_j, \tilde{r}_j) = \begin{cases} \exp(-\|\tilde{x}_i - \tilde{w}_j\| - \tilde{r}_j)^{\gamma_h}, & \text{if } \|\tilde{x}_i - \tilde{w}_j\| - \tilde{r}_j > 0 \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

式中灵敏参数 γ_h 反映输入样本在远离核超球 h 时, 它们之间的隶属度降低速率.

3.3 分类器结构

分类器拓扑结构由训练集自适应形成, 如图 1 所示, 包括 F_B 层和 F_C 层. 输入层 F_A 有 n 个信息处理单元. F_B 层节点代表超球集, F_A 和 F_B 层联结权是核超球的球心和半径, F_B 层有 m 个训练集学习后得到的核超球处理单元, 其传递函数是由(4)式定义的核超球隶属函数. 分类器的输出层有 p 个输出类别, F_B 和 F_C 层连接权是二值矢量 $\{0,1\}^p$, 计算公式为:

$$u_{jk} = \begin{cases} 1, & \text{if 球}_j \text{属于类别}k \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

式中, c_k 是输出层的第 k 个节点输出值. 输出层节点的传递函数是同一类核超球模糊集的并集, 定义为:

$$c_k = \max_{j=1}^m b_j u_{jk} \quad (k=1, \dots, p) \quad (6)$$

4 分类学习算法

离群点是样本集中一些远离其他数据点的数据,

由于离群点在整个样本集中只占很小的一部分, 所以经典的 SVDD 算法从运算效率的角度允许部分离群点位于超球之外, 从而导致分类结果的精确度下降. 此外, 现有的 SVDD 改进算法在超球集的构建过程中难以根除重叠问题. 如图 2 所示, 经过学习, 在特征空间内形成了两类核超球 H_1 和 H_2 , 两球在特征空间中相交. 设 \tilde{x}_k 为 H_2 的支撑向量, 由于分类函数选用距离作为判定分类的依据, 显然 \tilde{x}_k 更加靠近的 H_1 球心, 因此 \tilde{x}_k 将被划入错误的类别, 导致分类精度的下降. 为解决这个问题, 文献[11,12]给出了改进算法, 但调整过程十分复杂, 一方面使得算法时间效率大大降低, 另一方面也无法根除消除覆盖问题.

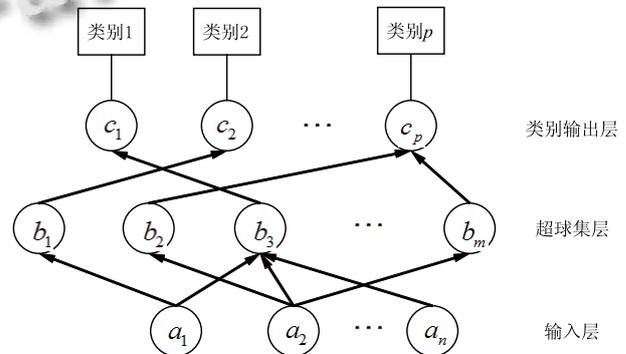


图 1 三层模糊核超球分类器

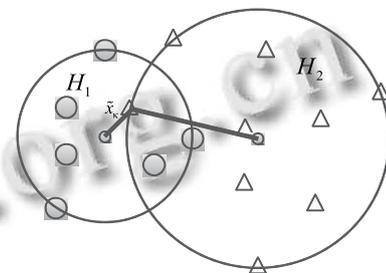


图 2 超球重叠导致的分类错误

由此看来, 构建一个对训练集完全覆盖的且没有重叠问题的超球集是提高分类结果精确度的重要手段, 本文将此类超球集定义为纯净核超球完全集.

定义 2. 纯净核超球完全集: 设关于训练集 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ 的核超球集为 $\Omega = \{(\Theta_1, y_1), (\Theta_2, y_2), \dots, (\Theta_p, y_p)\}$, 覆盖 Θ_i 为类别标签为 i 的核超球子集覆盖范围的并, y_i 为子集的分类标签. 对于任意 $x_i \in S$ 必然存在于一个覆盖 $\Theta_i \in \Omega$ 范围内, 且不被任意 $\Theta_j (j \neq i)$ 所覆盖, 则 Ω 被称为关于 S 的纯净核超球完全集.

由以上定义可知, 如果找到某种算法能够找到纯净核超球完全集, 则可以在保证对所有离群点的覆盖的前提下避免类似图 2 的分类错误. 不同于经典 SVDD 算法引入松弛变量和以往的 SVDD 改进算法在产生错误划分后再进一步调整的做法, FCABFKH 算法依靠样本合并的方法给出了寻找纯净核超球完全集的办法.

FCABFKH 算法主要分为两步, 第一步先将每一个训练样本视作一个核超球, 设置所有训练样本的标志为 1, 然后随机挑选不同类别的训练样本集中的一个点作为起始点, 不断合并同类样本, 任意训练样本被合并后, 设置该样本标志为 0, 同时对比新加入训练样本与未合并超球球心在特征空间中的距离是否比未合并超球半径大, 如确定, 则更新超球半径. 合并过程中, 如果发现合并后的核超球包含了异类训练样本, 则该次合并终止退回, 重新挑选该类训练样本集中任意标志为 1 的点作为新球的起始点, 不断合并同类训练样本集中其它标志为 1 的样本, 直到训练样本集中所有点的标志都为 0. 最后再设置所有只含有一个样本的核超球的球心为该样本, 半径为该样本到任意异类核超球球心距离与该球半径相减的最小值. 经过第一步的计算, 可以得到完全包含所有训练样本且不包含异类样本的核超球集. 仍然选取图 2 的训练样本作为范例, 通过第一步计算, 可以形成如图 3 所示的纯净核超球完全集.

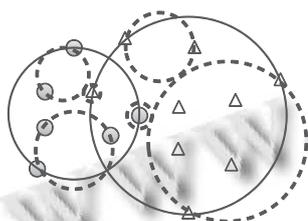


图 3 运用新方法生成的核超球集对比

算法的第二步则按照式(3)分别计算检验样本关于某类别的隶属度, 最后依据最大隶属度原则决定检验样本的类别标签.

下面将详细介绍计算方法:

第一步: F_A 到 F_B 学习算法

输入: 样本集 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, $x_i \in R^n$, $y_i \in \{1, 2, \dots, p\}$

输出: 核超球集

$W = \{(\tilde{w}_{i1}, \tilde{r}_{i1}), (\tilde{w}_{i2}, \tilde{r}_{i2}), \dots, (\tilde{w}_{im}, \tilde{r}_{im}), \dots, (\tilde{w}_{pm}, \tilde{r}_{pm})\}$, 其中下标 n_i 为划分为 i 类的核超球个数, $m = n_1 + n_2 + \dots + n_p$.

算法描述:

Step1 初始化 W , $n_i = 1 (i = 1, 2, \dots, p)$, 随机选取各类训练样本中任意一个样本作为该类核超球球心

Step2 设置所有训练集样本标志为 1;

Step3

for ($j = 1; j \leq p; j++$) //学习得到由 p 个核超球组成的核超球集

{while (有类别标签为 j 的样本标志为 1)

{for ($k = 1; k \leq N_j; k++$) // N_j 为训练集中类别标签为 j 的样本个数

{随机挑选 j 类样本标志为 0 的样本;

if(该样本与未合并超球球心在特征空间中的距离比未合并超球半径大){更新超球半径; 检验更新半径后的超球中是否存在异类样本, 如存在则回退至原超球, n_i++ ;} //end if

else {保留原超球半径, 设置样本标志为 1;}

} //end for

} //end while

} //end for

第二步: F_B 到 F_C 分类算法

输入: 核超球集 W

输出: $Y = \{y_1, y_2, \dots, y_g\}$ // g 为检验样本数量

算法描述:

for ($i = 1; i \leq g; i++$)

{ for ($j = 1; j \leq p; j++$)

{for ($k = 1; k \leq n_j; k++$)

{由式(4)计算 b_i ;} // end for

由式(6)计算得到 c_{ij} ;

} //end for

依据最大隶属度原则得到检验样本的类别标签;

} //end for

第一步算法的复杂度为 $O(p \cdot N_i \cdot (l - N_i))$, 在 $p \rightarrow l$ 时, 复杂度趋向于 $O(l^2)$; 第二步复杂度为 $O(g \cdot p \cdot n_j)$, 在 $g \rightarrow l$ 且 $p \rightarrow l$ 时, 复杂度趋向于 $O(l^2)$, 因此整个算法的复杂度为 $O(l^2)$. 而基于经典的 SVDD 改进的分类算法复杂度为 $O(l^3)$, FCABFKH 无疑大大提高了分类效率.

5 实验及其结果

实验数据来自于 Wind 终端, 类别标签分别为“好”、“差”、“中”, 年收益率定义: 当年最后一个交易日收盘价-当年第一个交易日开盘价+每股分红及派送)/第一个交易日开盘价, 划分标准为年收益率占总量的前 20%、后 20%及剩余的股票. 取径向

基核函数作为实验核函数, 采用经预处理后($t=7$)的上证 200 家上市公司 2006 年年报数据作为训练集, 学习得到超球集, 对 2007 年、2008 年情况进行预测分类, 分类结果由这两年年报数据验证准确率. 平均准确率为上述 2 年 3 个类别准确率的算术平均值.

表 2 实验结果

结果	核参数 σ	经典 SVDD 改进 ^[10]				子超球 SVM 法 ^[11]				FCABFKH			
		“好”	“中”	“差”	平均	“好”	“中”	“差”	平均	“好”	“中”	“差”	平均
划分 准确 率	4.43	28.5%	32.9%	45.7%	35.7%	29.1%	33.5%	46.9%	36.5%	35.3%	48.9%	51.4%	45.2%
	2.83	29.7%	31.3%	49.1%	36.7%	28.3%	33.1%	50.2%	37.2%	35.1%	49.1%	54.4%	46.2%
	1.53	29.9%	31.5%	52.3%	37.9%	31.9%	35.2%	59.8%	42.3%	34.9%	49.5%	52.1%	45.5%
	1.00	31.5%	32.9%	51.7%	38.7%	35.5%	36.2%	51.9%	41.2%	35.3%	49.7%	50.9%	45.3%
	0.32	28.1%	31.2%	40.6%	33.3%	31.3%	35.6%	48.9%	38.6%	35.3%	49.8%	50.5%	45.2%
平均运算时间		1.78s				1.47s				0.59s			

如表 2 所示, 一方面, 子超球 SVM 法由于采用了一定的改进手段修正误差使得平均准确率高于经典 SVDD 改进算法, 而 FCABFKH 的分类准确率又高于其他两种算法; 另一方面, 经典 SVDD 改进及子超球 SVM 法分类准确度随 σ 的变化波动较大, 原因在于 σ 取值不同导致形成的超球集重叠程度不同, 导致分类结果准确率产生波动, 而 FCABFKH 不产生超球重叠, 因此对 σ 的变化不敏感. 同时, 由运算时间可以看出由于避免使用二次规划, FCABFKH 的运行效率有较大提高.

下面运用分类结果对 2009 年至 2011 年股市进行投资预测. 设 k 为平均收益率, x_i 为第 i 支股票的年收益率, n 为股票个数, 平均收益率计算公式为:

$$k = \frac{\sum_{i=1}^n x_i}{n}$$

挑选前两年预测类别标签都为“好”的 33 支股票作为投资组合, 计算得到当年的投资组合收益率, 例如, 利用 2006 年数据对 2007 年和 2008 年股票投资组合进行预测, 挑选这两年预测结果都为“好”的 33 支股票作为投资组合, 计算得到 2009 年预测投资组合收益率. 上证指数收益率定义: (当年最后一个交易日收盘价-当年第一个交易日开盘价)/当年第一个交易日开盘价. 由表 3 可以看出, 由 FCABFKH 分类预测所得投资收益率要高于市场基准.

表 3 预测投资组合收益率对比

时 间	FCABFKH 分类预测投资组合收益率	上证指数收益率
2009 年	84.3%	77.2%
2010 年	-11.3%	-14.6%
2011 年	-20.8%	-22.2%

6 结论

本文尝试用基于模糊核超球的快速分类算法 (FCABFKH) 来对股票收益率进行预测, 并与经典 SVDD 改进及子超球 SVM 算法的实验结果进行对比, 最后得出 FCABFKH 算法优于前两者. 同时该研究结果也表明, 本文提出的算法可以用于选股投资, 利用实验数据计算所得的股票投资组合回报率超过了市场基准.

参考文献

- 1 Fama EF, French KR. Dividend yields and expected stock returns. *Journal of Financial Economics*, 1988, 22: 3-26.
- 2 Fama EF, French KR. The cross-section of expected stock returns. *Journal of Finance*, 1992, 47: 427-465.
- 3 Ding Z. Application of support vector machine regression in stock price forecasting. *Business, Economics, Financial Sciences, and Management*. Springer Berlin Heidelberg, 2012. 359-365.
- 4 数据挖掘在股票价格组合预测中的应用. 计算机仿真,

(下转第 148 页)

KKT 的算法,虽然识别率较非增量算法有一定提高,但其没考虑样本集中非 SV 集对分类信息的影响,因此其分类精度提高的不快;而本文算法,在增量到 4、5 步的时候,分类器的性能已远远优于其他算法,并能在接下来的学习中一直保持较好的性能.

表 1 各算法识别率的比较表

算法 学习数	非增量算法	基于 KKT 算法	本文算法
1	67.74%	67.74%	67.74%
2	70.97%	70.97%	74.19%
3	64.52%	64.52%	74.19%
4	48.39%	54.84%	70.97%
5	48.39%	54.84%	100%
6	61.29%	70.97%	96.77%
7	54.84%	64.52%	90.32%
8	35.48%	48.39%	100%
9	41.94%	58.06%	96.77%
10	51.61%	67.74%	100%

5 结语

本文提出了一种新的基于 KKT 条件的错误驱动增量学习算法.考虑了在增量学习过程中有可能成为 SV 的非支持向量,以及新增样本中满足 KKT 条件但被错分类的样本,并将新增样本做改进的 KKT 条件和错误驱动策略的筛选.实验结果显示,该算法在优化分类器效果,提高分类器性能方面上有良好的作用.

参考文献

1 王作为.具有认知能力的智能机器人行为学习方法研究[学

位论文].哈尔滨:哈尔滨工业大学,2010.

- Coppock H, Freund J. All or None versus Incremental Learning of Errorless Shock Escapes by the Rat. *Science*, 1962, 135: 318–319.
- Epifanio B, Takio K, Masakatsu H, Hiroaki I. Adapting SVM Image Classifiers to Changes in Imaging Conditions Using Incremental SVM: An Application to Car Detection, *Computer Vision-AACV2009*.
- Werner B. Incremental Regularization to Compensate Biased Teachers in Incremental Learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010: 1–8.
- Syed NA, Sung K. Handling Concept Drifts in Incremental Learning with Support Vector Machines. *Proc. of the 5th ACM SIGKDD international conference*, 1999: 316–321.
- Christopher P, Johns H. SVM Incremental Learning Adaptation and Optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003: 2685–2691.
- Zhao Y, Wang K, Zhong P. Incremental support vector machine based on border samples. *Computer Engineering and Design*, 2010, 31: 161–164.
- Luo J, Pronobis A, Caputo B. Incremental Learning for Place Recognition in Dynamic Environments. *IEEE International Conference on Intelligent Robots and Systems*. 2007. 721–729.
- 王珏,石纯一.机器学习研究.广西师范大学学报,2003,21: 1–15.
- 李祥纳,艾青,秦玉平.支持向量机增量学习算法综述,渤海大学学报,2007,9(4):187–189.

(上接第 201 页)

- 2012,7(29),375–378.
- 马耀兰.基于支持向量机方法的股票预测模型.鸡西大学学报,2013,4(13):124–125.
- 胡文军,王士同等.适合大样本快速训练的最大夹角间隔核心集向量机.电子学报,2011,39(5):1178–1184.
- 朱孝开,杨德贵.基于推广能力测度的多类 SVDD 模式识别方法.电子学报,2009,37(3):464–469.
- 林雄,冯海.基于 SVDD 多类分类新方法的研究.信息技术, 2008,7:100–103.
- 谢志强,高丽,杨静.基于球结构的完全二叉树 SVM 多类分类算法.计算机应用研究,2008,25(11):3268–3271.
- Ban T, Abe S. Implementing Multi-class Classifiers by One-class Classification Methods. *International Joint Conference on Neural Networks Vancouver*. BC, Canada. Sheraton Vancouver Wall Centre Hotel. 2006. 3272–332.
- 吴强,贾传炎,张爱锋,刘爽.球结构支持向量机的改进算法及仿真研究.系统仿真学报,2008,20(2):345–348.
- 刘小平.一种新的不平衡数据 v-NSVDD 多分类算法.南京大学学报(自然科学),2013,2(49):150–157.
- 汪东.基于支持向量机的选时和选股研究[学位论文].上海:上海交通大学,2007.