

基于时点分割的核Fisher判别分析-顺序回归机 多类分类建模方法

张洪祥^{1,2}, 毛志忠^{1,3}

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110004; 2. 辽东学院 信息技术学院, 辽宁 丹东 118001;

3. 东北大学 流程工业综合自动化国家重点实验室, 辽宁 沈阳 110004)

摘要: 针对多维时间序列的多类分类问题, 本文提出基于时点分割思想的核Fisher判别分析-顺序回归机(KFDA-ORM)多类分类建模方法. 该方法利用核Fisher判别分析(KFDA)与顺序回归机(ORM)的互补性得到分类决策函数; 对分类样本的多维时间序列进行时点分割处理, 使用决策函数得到各时点的分类级别; 通过指数平滑分析得到采样周期内样本的最终分类结果. 通过实验验证, 该方法对多维时间序列的分类具有较好效果, 是一种有效的多类分类方法.

关键词: 多维时间序列; 核Fisher判别分析; 顺序回归机; 指数平滑法

中图分类号: TP181; TP391 **文献标识码:** A

The kernel Fisher discriminant analysis-order regression machine multi-class classification modeling method based on time point partition

ZHANG Hong-xiang^{1,2}, Mao Zhi-zhong^{1,3}

(1. School of Information Science and Engineering, Northeastern University, Shenyang Liaoning 110004, China;

2. College of Information Technology, Eastern Liaoning University, Dandong Liaoning 118001, China;

3. State Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Shenyang Liaoning 110004, China)

Abstract: To tackle the multi-class classification problem of multidimensional time series, we present a kernel Fisher discriminant analysis (KFDA) and order regression machine (ORM) multi-class classification model based on the time point partition. The classification decision function of this method is obtained from the complementarities of KFDA and ORM. The time point segmentation is used to deal with the multidimensional time series of the samples. The classification level at each time point is determined by the decision function. The final classification results of the samples in the sampling period are obtained by using the exponential smoothing. The experimental results show that the algorithm has desirable results in the classification of multi-dimensional time series, and provides an effective multi-class classification model.

Key words: multidimensional time series; KFDA; ORM; exponential smoothing model

1 引言(Introduction)

多类分类技术作为数据挖掘技术中的一个重要研究方向, 已经取得了较大进展, 并得到了广泛地应用. 特别是多类分类混杂方法不但增强模式识别的鲁棒性和泛化能力, 同时也提高了分类器的正确率及稳定性. 近年来, 众多学者在多类分类问题上取得了较多的研究成果^[1], 如: 2004年, 陈才扣等提出一种融合主成分分析(principal component analysis, PCA)和核Fisher判别分析(kernel Fisher discriminant analysis, KFDA)的多类分类方法, 将特征抽取的速度分别提高了3倍和7倍, 而分类准确率没有降低^[2]. 2005年, 孔锐等探讨了核Fisher判别分析算法, 提出了一种基于KFDA的高性能多类分类算法, 该算法在保证分类精度的条件下, 提高了分类器的训练和

分类的速度^[3]. 2006年, 赵晖等提出了基于线性支持向量机度量类间相似程度的方法, 该方法有效地提高了多类分类器的分类精度和速度^[4]. 2007年, 张晓平提出了一种新的二叉树生成算法, 该算法克服了现有算法的一些缺陷, 应用上具有较好的推广性能和抗干扰能力^[5]. 2009年, 张宏达等提出一种线性判别分析(linear discriminant analysis, LDA)与支持向量机(support vector machine, SVM)混合的多类分类算法, 针对有向无环图支持向量机需训练大量支持向量机和误差积累的问题^[6]. 2010年, 郭显娥等总结基于支持向量机的主要分类方法, 并对各自的优缺点及性能做了比较^[7]. 2010年, 何菲等运用PCA-SVM方法对TE(tennessee eastman)过程中的故障数据进行多分类故障分类, 突破了2类或3类故障识别

的局限, 达到较高的多分类准确率^[8]. 2011年, 刁智华等用平衡决策树结构, 得到了一种新的决策树支持向量机多分类算法, 大大减少系统的训练时间^[9]. 在这些算法中, 支持向量机及其衍生算法成为了多类分类技术的热点^[10-11].

随着社会经济的迅速发展, 大量具有时间特性的数据, 进入到数据挖掘中, 并逐渐成为研究的热点. 时间序列数据是指按时间顺序取得的一组观测数据, 通常是按时间顺序排列. 而时间序列的多维特性会使数据挖掘的过程复杂化, 这是由于一方面多维时序存在着存储的多维性与本质的低维性两个维度, 数据的多维特性会给数据处理过程带来计算效率低下、维数灾难等问题. 传统的多类分类方法都要对多维数据进行预处理, 常用的降维技术如文献中的PCA, ICA, FDA^[12]等都是基于线性提取, 提取结果也是线性特征; 然而现实生活中, 数据大多都是非线性, 特别是多维时间序列, 这样就无法保证对数据提取仍保持其几何结构和关系, 较难直接获得多维空间数据对应低维空间的映射表示; 另一方面传统多类分类方法大多是基于相似性度量的聚类方法^[13-14], 这些方法的距离测度针对的是确定性向量空间, 而时间序列的长度通常是变化的不确定的, 特别是在对小样本情况下, 采样时间点较少的多维时间序列分类, 传统方法难以寻找其规律性和周期性.

针对上述问题, 本文提出了基于时点分割思想的核Fisher判别分析-顺序回归机(KFDA-ORM)多维时间序列多类分类建模方法. 该方法通过KFDA对多维数据提取最佳鉴别特征, 使数据在高维特征空间的映射变为线性可分, 呈现有序性, 便于顺序回归机进行分类建立决策函数; 对样本的多维时间序列按照时点分割思想进行分割处理, 利用决策函数得到样本在各时点上的分类结果, 最后通过指数平滑确定采样周期内样本的最终类别.

2 多维时间序列(Multidimensional time series)

时间序列(time series)是一组变量在连续的时间上获得的一系列采样值. 长度为 n 的 d 维时间序列:

$$x_{i,j}, i = 1, 2, \dots, d, j = 1, 2, \dots, n,$$

其中 $x_{i,j}$ 表示索引时间点 j 上第 i 维变量的值. 如果仅对单个变量进行采样, 即 $d = 1$, 那么所获得的时间序列称为一维时间序列.

如果采样对象与一个时刻对应的是多个实数变量, 其相互之间的关系也不再局限于变量与时间之间的相关性, 变量与变量之间也会具有一定的联系, 此时所获得的时间序列则称为多维时间序列. 一般地, 一个长度为 n 的 $d \times n$ 维时间序列可以用一个 X 矩

阵表示: $X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & & \vdots \\ x_{d,1} & \cdots & x_{d,n} \end{bmatrix}$, 其中 $x_{i,j}$ 为第 i 维变量

在时间点 j 上的采样值^[15].

3 模型的建立(Model establish)

3.1 原始数据标准化(Original data standardization)

考虑到实际数据的属性度量单位不同, 需要对数据进行标准化处理, 将属性值映射到某个数值区间. 数据标准化模型如下:

$$x'_{i,j} = \frac{x_{i,j} - \bar{x}_i}{s_i}, \quad (1)$$

其中: $x_{i,j}$ 为第 j 个样本第 i 项变量的原始数据,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{i,j}, s_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_{i,j} - \bar{x}_i)^2}. \quad (2)$$

3.2 KFDA模型(KFDA model)

假设有 n 个 d 维实数样本, 可分为 l 个类别 $\{X_1, X_2, \dots, X_l\}$, 每一类有 n_i 个样本, 其中

$$X_i = [x_1^i \ x_2^i \ \cdots \ x_{n_i}^i].$$

KFDA是通过一个非线性映射 φ , 将数据映射到高维特征空间, 即 $\varphi: \mathbb{R}^d \rightarrow F, x \rightarrow \varphi(x)$, 再进行线性Fisher判决分析.

KFDA通过最大化其特征空间中的Fisher准则来得出最优鉴别向量

$$W_{\text{opt}} = \arg \max_{w^\varphi} \frac{|(W^\varphi)^\text{T} S_B^\varphi W^\varphi|}{|(W^\varphi)^\text{T} S_W^\varphi W^\varphi|} = [w_1^\varphi \ w_2^\varphi \ \cdots \ w_c^\varphi], \quad (3)$$

其中 S_B^φ 和 S_W^φ 分别为特征空间的类间散布矩阵和类内散布矩阵,

$$S_B^\varphi = \sum_{i=1}^l n_i (\mu_i^\varphi - \mu_0^\varphi) (\mu_i^\varphi - \mu_0^\varphi)^\text{T}, \quad (4)$$

$$S_W^\varphi = \sum_{i=1}^l \sum_{k=1}^{n_i} (\varphi(x_k) - \mu_0^\varphi) (\varphi(x_k) - \mu_0^\varphi)^\text{T}, \quad (5)$$

其中: $\mu_i^\varphi = \frac{1}{n_i} \sum_{k=1}^{n_i} \varphi(x_k)$, $\mu_0^\varphi = \frac{1}{n} \sum_{j=1}^n \varphi(x_j)$. 式(3)

中 $[w_1^\varphi \ w_2^\varphi \ \cdots \ w_c^\varphi] (c \leq l - 1)$ 是对应 $S_W^{\varphi^{-1}} S_B^\varphi$ 的最大特征值向量的解^[16], 通过求解下式可以求解:

$$S_B^\varphi w_i^\varphi = \lambda S_W^\varphi w_i^\varphi. \quad (6)$$

再生核理论认为: 鉴别向量 W 位于特征空间中所有训练样本组成的子空间中^[17], 可表示为全体训练样本的线性组合, 用Mercer核 $k(x_i, x_j)$ 代替 $\varphi^\text{T}(x_i)\varphi(x_j)$:

$$w_i^\varphi = \sum_{i=1}^n a_i \varphi(x_i), \quad (7)$$

$$w_i^\text{T} \varphi(x) = \sum_{i=1}^l \sum_{j=1}^{n_i} a_i k(x_i, x_j). \quad (8)$$

将式(7)-(8)代入式(3)可得

$$(w_i^\varphi)^\text{T} S_B^\varphi w_i^\varphi = a_i^\text{T} M a_i, (w_i^\varphi)^\text{T} S_W^\varphi w_i^\varphi = a_i^\text{T} N a_i, \quad (9)$$

$$a_{\text{opt}} = \arg \max_a \frac{|a^T M a|}{|a^T N a|}. \quad (10)$$

求解最佳鉴别向量就转化为上式取得最大值的向量 a 的问题. 其中

$$M = \sum_{i=1}^l (m_i - m_0)(m_i - m_0)^T, \quad (11)$$

式中: $m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} k(x_j, x_i)$, $m_0 = \frac{1}{n} \sum_{j=1}^n k(x_j, x_i)$.

$$N = \sum_{j=1}^l \kappa_j (I - l_{N_j})(\kappa_j)^T, \quad (12)$$

式中: $\kappa_j = [k(x_1, x_j) \cdots k(x_l, x_j)]^T$, I 是单位矩阵, l_{N_j} 是一个 $n_j \times n_j$ 所有元素都是 $1/n_j$ 的矩阵. 则最优子空间为

$$W_{\text{opt}} = [\sum_{i=1}^n a_1^i \varphi(x_i) \quad \sum_{i=1}^n a_2^i \varphi(x_i) \quad \cdots \quad \sum_{i=1}^n a_c^i \varphi(x_i)]. \quad (13)$$

获取一个模式 X 在特征空间 F 中的映射, 可将 $\varphi(x)$ 映射在该空间, 通过下式进行计算:

$$\varphi(X)W_{\text{opt}} = \sum_{i=1}^l \sum_{j=1}^{n_i} a_j^i k(x_i, x_j), \quad (14)$$

其中 $\sum_{i=1}^n a_j^i k(x_i, x_j) = [\sum_{i=1}^n a_1^i k(x_i, x) \quad \sum_{i=1}^n a_2^i k(x_i, x) \quad \cdots \quad \sum_{i=1}^n a_c^i k(x_i, x)]$.

通过KFDA对数据进行的特征提取, 会使多维数据在特征空间变得相对有序, 有利于对样本进行下一步的分类.

3.3 顺序回归机的数学提法(Mathematical formulation of ORM)

设输入样本可分为 l 类, 且在空间 \mathbb{R}^d 的位置是从 1 到 l 有顺序、有确定的相邻关系, 即第 i 类和第 $i - 1$, $i + 1$ 类相邻, 而第 $i - 1$ 类和 $i + 1$ 类并不相邻. 顺序回归问题可表述为^[18], 即: 给定样本集

$$T = \{x_j^i, y_j^i\}_{i=1, \dots, l}^{j=1, \dots, n_i},$$

其中: $x_j^i \in \mathbb{R}^d$ 为样本输入, $X_i \in \{1, 2, \dots, l\}$ 表示相应的类别, n_i 是第 i 类样本的个数.

样本在空间 \mathbb{R}^d 上寻找的 $l - 1$ 个互相平行的超平面:

$$(w \cdot x) = b_r, \quad r = 1, \dots, l - 1, \quad (15)$$

其中: $w \in \mathbb{R}^d$, $b_1 \leq b_2 \leq \dots \leq b_{l-1}$, 且 $b_1 > -\infty$, $b_{l-1} < +\infty$. 对于给定的样本 x_j^i , 采用如下决策函数:

$$f(x) = \min_{r \in \{1, \dots, l\}} \{r : (w \cdot x) - b_r < 0\}. \quad (16)$$

确定出使 $f(x)$ 取最小值的 r 值, 则可以将 x_j^i 归为第 r 类.

3.4 顺序回归机(ORM)

与建立两类分类模型先考虑线性可分情形类似, 这里先用 $l - 1$ 个超平面完全正确分开的样本, 作为

最大间隔法的推广, 最大化相邻最近的两类之间的间隔, 可求解下列问题^[19-20]:

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2, \\ \text{s.t.} \quad & (w \cdot x_j^i) - b_i \leq -1, \quad i = 1, \dots, l, \quad j = 1, \dots, n_i, \\ & (w \cdot x_j^i) - b_{i-1} \geq 1, \quad i = 1, \dots, l, \quad j = 1, \dots, n_i, \end{aligned} \quad (17)$$

其中: $b = (b_1, \dots, b_{l-1})^T$, $b_1 > -\infty$, $b_{l-1} < +\infty$.

对于一般的样本集, 需要引入松弛变量 $\xi^{(*)}$ 和惩罚参数 $V > 0$, 上述问题变为凸二次规划:

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 + V \sum_{i=1}^l \sum_{j=1}^{n_i} (\xi_j^i + \xi_j^{*i}), \\ \text{s.t.} \quad & (w \cdot x_j^i) - b_i \leq -1, \quad i = 1, \dots, l, \quad j = 1, \dots, n_i, \\ & (w \cdot x_j^i) - b_{i-1} \geq 1, \quad i = 1, \dots, l, \quad j = 1, \dots, n_i, \\ & \xi_j^i \geq 0, \quad \xi_j^{*i} \geq 0, \end{aligned} \quad (18)$$

其中: $b = (b_1, \dots, b_{l-1})^T$, $b_1 > -\infty$, $b_{l-1} < +\infty$, $\xi^{(*)} = (\xi_1^1, \dots, \xi_{l_1}^1, \dots, \xi_1^l, \dots, \xi_{l_l}^l; \xi_1^{*1}, \dots, \xi_{l_1}^{*1}, \dots, \xi_1^{*l}, \dots, \xi_{l_l}^{*l})^T$.

式(18)的求解可转化为其对偶问题解的求解

$$\begin{aligned} \min_{a^{(*)}} \quad & \frac{1}{2} \sum_{j, i} \sum_{j', i'} (a_j^{*i} - a_j^i)(a_{j'}^{*i'} - a_{j'}^i)(x_j^i \cdot x_{j'}^{i'}) - \\ & \sum_{j=1}^l \sum_{i=1}^{l^j} (a_j^i + a_j^{*i}), \\ \text{s.t.} \quad & \sum_{i=1}^{l^j} a_j^i = \sum_{i=1}^{l^{j+1}} a_j^{*i+1}, \\ & 0 \leq a_j^i, a_j^{*i} \leq V, a_i^{*1} = 0, \\ & i = 1, \dots, l, \quad j = 1, 2, \dots, n_i, \\ & \bar{a}^{(*)} = (\bar{a}_1^1, \dots, \bar{a}_{n_1}^1, \dots, \bar{a}_1^l, \dots, \bar{a}_{n_l}^l, \bar{a}_1^{*1}, \\ & \quad \dots, \bar{a}_{n_1}^{*1}, \dots, \bar{a}_1^{*l}, \dots, \bar{a}_{n_l}^{*l})^T. \end{aligned} \quad (19)$$

设 $\bar{a}^{(*)}$ 是对偶问题(19)的解, 若对 $i = 1, \dots, l - 1$ 都存在 $\bar{a}^{(*)}$ 的分量 $\bar{a}_i^j \in (0, V)$ 或者 $\bar{a}_i^{*j+1} \in (0, V)$, 则支持向量顺序回归机的原始问题(17)对 (w, b) 的解是存在唯一的, 在此情况下, 它对 (w, b) 的解 $(\bar{w}, \bar{b}) = (\bar{w}, \bar{b}_1, \dots, \bar{b}_{l-1})$ 可按下列方式计算: 令

$$\bar{w} = \sum_{i'=1}^l \sum_{j'=1}^{n_i} (\bar{a}_{j'}^{*i'} + \bar{a}_{j'}^{i'}) x_j^i, \quad (20)$$

其中

$$a^{(*)} = (a_1^1, \dots, a_{n_1}^1, \dots, a_1^l, \dots, a_{n_l}^l, a_1^{*1}, \dots, a_{n_1}^{*1}, \dots, a_1^{*l}, \dots, a_{n_l}^{*l})^T.$$

对 $i = 1, \dots, l - 1$, 若存在 $\bar{a}_i^j \in (0, V)$, 则令

$$\bar{b}_j = 1 + \sum_{i'=1}^l \sum_{j'=1}^{n_i} (\bar{a}_j^{*i} - \bar{a}_{j'}^{i'}) (x_j^i \cdot x_{j'}^{i'}); \quad (21)$$

对 $i = 1, \dots, l - 1$, 若存在 $\bar{a}_i^{*j+1} \in (0, V)$, 则令

$$\bar{b}_j = \sum_{i'=1}^l \sum_{j'=1}^{n_i} (\bar{a}_j^{*i} - \bar{a}_{j'}^{i'}) (x_j^i \cdot x_{j'}^{i'}) - 1. \quad (22)$$

根据 (w, b) 构造决策函数

$$f(x) = \min_{i \in \{1, \dots, l\}} \{r : (\bar{w} \cdot x_j^i) - \bar{b}_j < 0\}. \quad (23)$$

通过决策函数 $f(x)$ 就可以对分类样本进行多类分类的处理.

3.5 指数平滑算法(Exponential smoothing)

指数平滑是指所有过去数据的加权平均. 所有的权重由 a 唯一决定, 而且随着时间的往前推移加权指数递减, 所以称为指数平滑. 本文指数平滑计算公式如下:

$$\bar{A} = aA_t + a(1-a)A_{t-1} + a(1-a)^2A_{t-2} + \dots \quad (24)$$

在指数平滑中, 对 a 的选择较大则最近的数据赋予的权重大; 反之, a 较小, 过去的数据仍对未来有影响. 本文采用该方法对分类样本在采样周期内的分类情况进行加权累加, 使分类结果更加客观、真实.

3.6 算法步骤(Steps of the algorithm)

Step 1 原始数据规范化. 原始数据量纲不同, 对所有数据进行规范化处理.

Step 2 对训练数据进行KFDA鉴别分析. 利用式(13)求出 $W_{opt} = [w_1^\varphi \ w_2^\varphi \ \dots \ w_c^\varphi]$, 使数据在高维空间中构成一个子空间. 再通过式(14)求得 $T = \varphi(X)W_{opt}$, 从而获取训练集在特征空间 F 中的映射, 便于顺序回归机的分类处理.

Step 3 用顺序回归机对KFDA得到的有序特征输入进行多类分类. 选取适当的惩罚参数 V , 构造并求解凸二次规划问题得解 $a^{(*)}$. 通过式(20)求得 w , 将其代入式(21)-(22)求得 b_j . 将 w 和 b_j 代入式(23)得到决策函数 $f(x)$.

Step 4 对分类样本按照采样时间点进行分割处理, 分别通过决策函数 $f(x)$ 对每个时点上的样本进行分类, 得出样本在不同时点的分类等级.

Step 5 最后对样本在各个采样时间点上的分类结果进行综合分析, 本文对受评样本在时点上采用指数平滑算法, 计算出综合的分类等级.

4 实验(Experiments)

本文选取深、沪两市某板块15家上市公司作为信用评价对象, 进行3等级分类, 以2001年到2010年年报为时间采样点, 指标变量依据财务常用信用评价指标体系选取(每股收益、每股净资产、每股公积金、每股未分配利润、每股经营现金流、净资产收益率、资产负债率、存货周转率、净利润增长率、净利润率). 构成变量个数为10, 采样时间点数10, 样本个数为 10×15 的多维时间序列矩阵数据库.

4.1 实验过程(Experimental process)

Step 1 样本集原始数据进行规范化处理, 由于数据结构的无序性特征明显, 不能直接进行分类.

Step 2 使用核KFDA函数对训练集进行特征提取可以看到数据在特征空间变得有序线性可分(见图1), 对训练集执行 $l = 3$ 时的顺序回归机分类算法, 求得的决策函数.

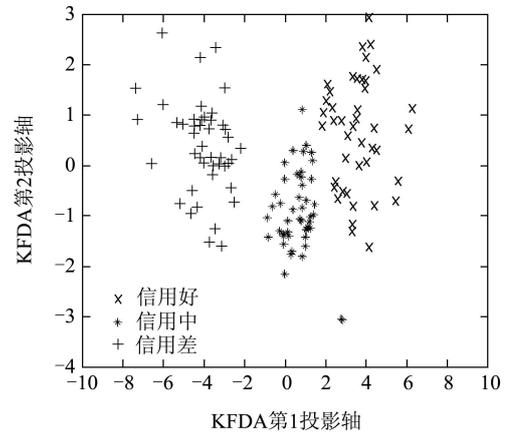


图 1 KFDA前两个轴上的投影特征提取

Fig. 1 The characteristics of the data extraction on the first two axis of KFDA

Step 3 对分类样本数据按照采样时间点进行分割处理成10个数据库, 分别通过决策函数 $f(x)$ 对每个时点上的样本进行分类分析, 得出样本在不同时点的信用分类等级(见图2). 对样本在采样时间点的信用级别进行仿真, 便可得到受评样本在采样周期的信用状态趋势以及其波动情况(见图3-4).

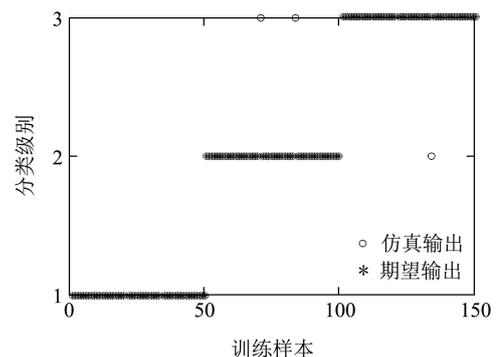


图 2 ORM分类结果

Fig. 2 ORM classification results

Step 4 采用指数平滑算法, 计算出样本在各时点上的最终类别.

4.2 仿真结果分析(Simulation results analysis)

从图3-4可以观察到受评样本在采样时间点上信用级别情况. 图中可以看到受评样本 x_2, x_6, x_{15} 的波动幅度较大, 其中: x_2 在02时点的级别为信用差, 03时点的级别为信用好, 08时点的级别为信用差, 其他时点一直保持在信用中, 因此指数平滑后的 x_2 稳态信用级别为信用中; x_6 在03, 05, 07时点的级别为信用中, 而其他时点一直保持在信用好, 因此指数平滑后 x_6 的稳态信用级别为信用好; x_{15} 在04, 06, 09时点的级别为信用好, 其他时点一直保持在信用

中,因此指数平滑后 x_{15} 的稳态信用等级为信用中.对分类样本各时点指数平滑后,可以得到最终的信用评价分类结果,其中:信用好包括 $x_1, x_5, x_6, x_{10}, x_{11}$;信用中包括 $x_2, x_7, x_{12}, x_{14}, x_{15}$;信用差包括 $x_3, x_4, x_8, x_9, x_{13}$.

为了验证本文方法的有效性,本文分别与文献

[2,6-8]中多类分类方法进行了仿真比较.通过比较(如表1所示),可以看到该方法的在信用评价上的分类准确率有所提高,通过对比受评样本的经营业绩以及行业综合地位,证明该方法的信用评价结果较为真实,具有很好的稳定性,不会受到孤立时点由于数据突变导致的评价失真的问题.

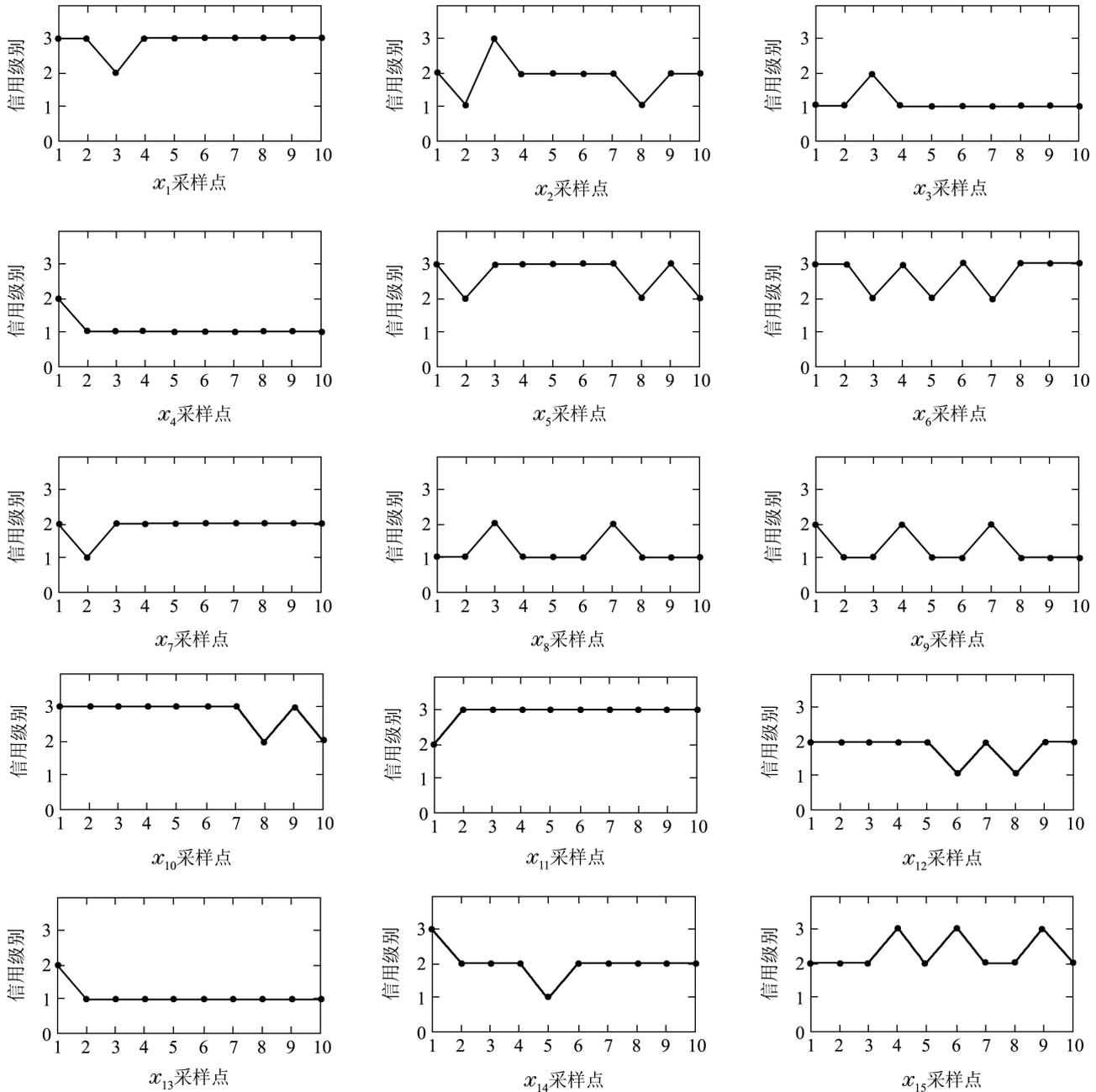


图3 x_1-x_{15} 采样时间点趋势

Fig. 3 x_1-x_{15} sampling point trends

表1 多类分类方法的比较

Table 1 Comparison of the commonly multi-class classification methods

| 模型 | SVM | PCA+KFDA | LDA+SVM | PCA+SVM | KFDA+ORM |
|-------|-----|----------|---------|---------|----------|
| 准确率/% | 88 | 91 | 92 | 96 | 98 |

5 结论(Conclusions)

本文提出针对多维时间序列的KFDA-ORM多类分类建模. 该方法的优点在于: 通过KFDA分析可以有效地对多维时间序列进行特征选取, 使数据在特征空间变得有序, 有利于顺序回归机的分类; 其次采用分割时点的方法, 对多维时间序列进行时点分割, 再利用指数平滑方法将各时点分类结果进行加权累加, 得到较为稳定可信的分类结果. 通过实例验证, 该方法对多维时间序列的分类具有较好效果, 是一种有效的多类分类方法, 适用于小样本情况下, 采样时间点较少的多维时间序列多类分类问题.

参考文献(References):

- [1] 郝红卫, 王志彬, 殷绪成, 等. 分类器的动态选择与循环集成方法 [J]. 自动化学报, 2011, 37(11): 1290 – 1295.
(HAO Hongwei, WANG Zhibin, YIN Xucheng, et al. Dynamic selection and circulating combination for multiple classifier systems [J]. *Acta Automatica Sinica*, 2011, 37(11): 1290 – 1295.)
- [2] 陈才扣, 杨静宇, 杨健. 一种融合PCA和KFDA的人脸识别方法 [J]. 控制与决策, 2004, 19(10): 1147 – 1154.
(CHEN Caikou, YANG Jingyu, YANG Jian. Fusion of PCA and KFDA for face recognition [J]. *Control and Decision*, 2004, 19(10): 1147 – 1154.)
- [3] 孔锐, 张冰. 基于核Fisher判决分析的高性能多类分类算法 [J]. 计算机应用, 2005, 25(06): 1327 – 1329.
(KONG Rui, ZHANG Bing. New perfect performance multiclass classification algorithm based on KFDA [J]. *Computer Applications*, 2005, 25(06): 1327 – 1329.)
- [4] 赵晖, 荣莉莉, 李晓. 一种设计层次支持向量机多类分类器的新方法 [J]. 计算机应用研究, 2006, 23(6): 34 – 37.
(ZHAO Hui, RONG Lili, LI Xiao. New method of design hierarchical support vector machine multi-class classifier [J]. *Application Research of Computers*, 2006, 23(6): 34 – 37.)
- [5] 张晓平, 杨洁明. 一种新的支持向量机多类分类二叉树算法 [J]. 机械工程与自动化, 2007, (3): 1 – 3.
(ZHANG Xiaoping, YANG Jieming. A new algorithm of generating binary tree on multiclass classification methods for support vector machines [J]. *Mechanical Engineering & Automation*, 2007, (3): 1 – 3.)
- [6] 张宏达, 王晓丹, 徐海龙. 一种LDA与SVM混合的多类分类方法 [J]. 控制与决策, 2009, 24(11): 1723 – 1728.
(ZHANG Hongda, WANG Xiaodan, XU Hailong. Hybrid multi-class classification approach based on LDA and SVM [J]. *Control and Decision*, 2009, 24(11): 1723 – 1728.)
- [7] 郭显娥, 武伟, 刘春贵, 等. 多类SVM分类算法的研究 [J]. 山西大同大学学报(自然科学版), 2010, 26(3): 6 – 14.
(GUO Xiane, WU Wei, LIU Chungui, et al. Research of multi-class support vector machines classification algorithm [J]. *Journal of Shanxi Datong University (Natural Science)*, 2010, 26(3): 6 – 14.)
- [8] 何菲, 杜文莉, 钱锋. PCA-SVM的多故障分类方法在TE过程中的应用 [J]. 计算机与应用化学, 2010, 27(10): 1321 – 1324.
(HE Fei, DU Wenli, QIAN Feng. Multifaults classification method based on PCA-SVM and its application to the TE process [J]. *Computers and Applied Chemistry*, 2010, 27(10): 1321 – 1324.)
- [9] 刁智华, 赵春江. 一种新的基于平衡决策树的SVM多类分类算法 [J]. 控制与决策, 2011, 26(1): 149 – 156.
(DIAO Zhihua, ZHAO Chunjiang. A new SVM multi-class classification algorithm based on balance decision tree [J]. *Control and Decision*, 2011, 26(1): 149 – 156.)
- [10] 李仁兵, 李艾华, 白向峰, 等. 支持向量机的进化多核设计 [J]. 控制理论与应用, 2011, 28(6): 793 – 798.
(LI Renbing, LI Aihua, BAI Xiangfeng, et al. Evolutionary multiple kernels design for support vector machines [J]. *Control and Decision*, 2011, 28(6): 793 – 798.)
- [11] 任斌, 程良伦. 多项式光滑的支持向量回归机 [J]. 控制理论与应用, 2011, 28(2): 261 – 265.
(REN Bin, CHENG Lianglun. Polynomial smoothing support vector regression [J]. *Control Theory & Applications*, 2011, 28(2): 261 – 265.)
- [12] 程健, 陈光响, 龚平华, 等. 非线性多维时间序列模式分类的新方法 [J]. 计算机工程与应用, 2011, 47(32): 128 – 131.
(CHENG Jian, CHEN Guangyuan, GONG Pinghua, et al. Novel method for patterns classification of nonlinear multidimensional time series [J]. *Computer Engineering and Applications*, 2011, 47(32): 128 – 131.)
- [13] 贾澎涛, 何华灿, 刘丽, 等. 时间序列数据挖掘综述 [J]. 计算机应用研究, 2007, 24(11): 15 – 18.
(JIA Pengtao, HE Huacan, LIU Li, et al. Overview of time series data mining [J]. *Application Research of Computers*, 2007, 24(11): 15 – 18.)
- [14] 李春生, 王耀南. 聚类中心初始化的新方法 [J]. 控制理论与应用, 2010, 27(10): 1435 – 1440.
(LI Chunsheng, WANG Yaonan. New initialization method for cluster center [J]. *Control Theory & Applications*, 2010, 27(10): 1435 – 1440.)
- [15] 欧阳为民, 蔡庆生. 数据库中的时态数据挖掘研究 [J]. 计算机科学, 1998, 25(4): 60 – 63.
(OUYANG Weimin, CAI Qingsheng. Research on temporal data mining in databases [J]. *Computer Science*, 1998, 25(4): 60 – 63.)
- [16] 徐勇, 杨强, 杨静宇. 基于核的快速特征抽取及识别方法 [J]. 解放军理工大学学报(自然科学版), 2005, 6(2): 128 – 131.
(XU Yong, YANG Qiang, YANG Jingyu. Fast kernel-based feature extraction and recognition method [J]. *Journal of PLA University of Science and Technology(Natural Science)*, 2005, 6(2): 128 – 131.)
- [17] JIAN Y, ZHONG J, YANG J Y, et al. Essentel of kerne Fishier discriminant: KPCA plus LDA [J]. *Pattern Recognition*, 2004, 37(10): 2097 – 2100.
- [18] 杨志霞, 邓乃扬. 线性支持向量顺序回归机的原始问题的解集分析 [J]. 运筹学学报, 2007, 11(3): 105 – 112.
(YANG Zhixia, DENG Naiyang. The analysis to the solution set of the primal problem in linear support vector regression [J]. *OR Transactions*, 2007, 11(3): 105 – 112.)
- [19] FRANK E, HALL M A. A simple arproach to ordinal classification [C] // *Proceeding of the European Conference on Machine Learning*. Hamilton, New Zealand: the University of Waikato, 2001, 145 – 165.
- [20] XIA F, ZHANG W S, WANG J. An effective tree-based algorithm for ordinal regres-sion [J]. *IEEE Intelligent Information Bulletin*, 2006, 7(1): 22 – 26.

作者简介:

张洪祥 (1978-), 男, 博士研究生, 目前研究方向为复杂系统建模与优化、经济控制, E-mail: zhx780711@yahoo.com.cn;

毛志忠 (1961-), 男, 教授, 博士生导师, 目前研究方向为复杂控制建模与控制等, E-mail: maozhizhong@ise.neu.edu.