

基于核 Fisher 判别的说话人辨认

谭萍^{1,2} 邢玉娟^{1,2} 李明¹

(兰州理工大学计算机与通信学院¹, 兰州 730050; 甘肃联合大学理工学院², 兰州 730000)

摘要 针对语音特征的自身特点,通过对 Fisher 判别技术的研究,提出采用核 Fisher 判别技术来解决说话人辨认。核 Fisher 判别技术在判别中使用所有训练样本,显著提高了系统的识别率。实验数据表明,该方法在不影响系统实时性的情况下,有效地提高了说话人辨认系统的识别率。

关键词 核 Fisher 判别 说话人辨认 支持向量机 核函数

中图分类号 TN912.34; 文献标志码 A

说话人辨认技术^[1]是指通过说话人测试语音来判断该说话人是训练集合 N 个人中的哪个人过程。在网络安全、银行证券系统、公安等领域有着广泛的应用前景。

目前说话人辨认多是采用基于支持向量机的分类器。在文献 [2] 中,侯风雷等人成功地将支持向量机应用于说话人辨认,并取得了较好的实验结果,但在支持向量机进行判别的时候仅仅使用了训练样本中的一些特殊样本即支持向量,这势必会影响系统的识别性能。核 Fisher 判别技术^[3]是基于 Fisher 线性判别提出的一种二元分类方法,其核心思想是通过核函数将样本向量映射到一高维空间,在此空间使用 Fisher 线性判别进行分类。在文献 [3] 中,使用核 Fisher 判别技术进行判别分类,指出核 Fisher 判别在识别中使用所有的训练样本,显著提高了系统的识别性能,实验结果验证了这种分类方法的性能优于其他分类技术。本文提出采用核 Fisher 判别技术来实现说话人辨认。

1 Fisher 线性判别 (FLD)

Fisher 线性判别经典的二元分类技术,它根据

2008年1月8日收到 甘肃省教育厅科研项目(0603-10)资助
第一作者简介:谭萍,女,讲师,硕士生,研究方向:智能信息处理。

类间离散度最大化类内离散度最小化的准则,确定原始向量的投影方向,使两类最大程度的分离,从而达到正确的分类。

设 $X_1 = \{x_1^1, \dots, x_{c_1}^1\}$ 和 $X_2 = \{x_1^2, \dots, x_{c_2}^2\}$ 是两类样本: $X = X_1 \cup X_2 = \{x_1, x_2, \dots, x_N\}$, $C_1 + C_2 = N$, 最佳投影方向可通过最大化下面目标函数求得:

$$J(w) = \frac{w^T S_b w}{w^T S_w w} \quad (1)$$

(1) 式中 w 为投影方向, S_b 为类间离散度矩阵, S_w 为类内离散度矩阵, 分别定义如下。

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (2)$$

$$S_w = \sum_{i=1}^2 \sum_{x \in X_i} (x - m_i)(x - m_i)^T \quad (3)$$

(2) 式、(3) 式中 m_i ($i=1, 2$) 是各类样本的均值向量,

$$m_i = \frac{1}{C_i} \sum_{j=1}^{C_i} x_j^i \quad (i=1, 2) \quad (4)$$

(4) 式中 x_j^i 表示第 i 类的第 j 个样本。最大化式 (1) 可得:

$$w = S_w^{-1} (m_1 - m_2) \quad (5)$$

在 Fisher 线性判别中,关键问题就是寻找最优投影方向;而对于语音数据来说,数据分布呈非线性,线性判别显得过于简单,并不能达到好的分类性能。因此,在本文的说话人辨认系统中提出使用基于核函数的 Fisher 判别进行分类。

2 基于核 Fisher判别 (KFD)的说话人辨认

基于核 Fisher判别 (KFD)的说话人辨认系统框图如图 1所示。

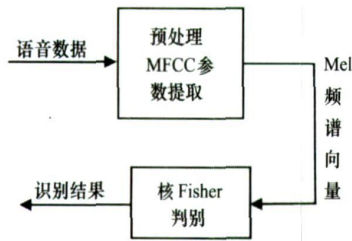


图 1 基于 PCA 和核 Fisher判别的说话人确认

在训练阶段,语音数据经过预处理得到 Mel参数向量,通过核 Fisher判别准则(类间离散度最大,类内离散度最小)找到最优的映射方向 w 。在识别阶段,待识别说话人语音中导出的特征向量通过训练中最优映射方向进行投影,最后由分类判别函数进行判决。说话人辨认是一个多类分类问题,需构造基于核 Fisher判别的多元分类器。

2.1 核 Fisher判别

核 Fisher判别 (KFD)的核心思想是,通过一个非线性映射 Φ 将原始特征空间映射到一个新的特征空间 H 。在新的特征空间 H 中使用 Fisher线性判别进行分类^[4,5]。

在新的特征空间 H 中目标函数相应变为:

$$J(w) = \frac{w^T S_b^\Phi w}{w^T S_w^\Phi w} \quad (6)$$

(6)式中 S_b^Φ 和 S_w^Φ 是相应的在空间 H 中的类间离散度矩阵和类内离散度矩阵, w 是投影方向。

$$S_b^\Phi = (m_1^\Phi - m_2^\Phi)(m_1^\Phi - m_2^\Phi)^T \quad (7)$$

$$S_w^\Phi = \sum_{i=1}^2 \sum_{x \in X} (\Phi(x) - m_i^\Phi)(\Phi(x) - m_i^\Phi)^T \quad (8)$$

$$m_i^\Phi = \frac{1}{C_i} \sum_{j=1}^{C_i} \Phi(x_j^i), \quad i = 1, 2 \quad (9)$$

由于空间 H 的维数很高甚至是无穷维,直接求解是不可能的,可通过核函数的技巧^[6],而不涉及到具体的非线性运算。

根据核函数理论,任何一个目标函数的解 w 都

可以用特征空间中元素的线性组合表示:

$$w = \sum_{i=1}^N a_i \Phi(x_i) \quad (10)$$

将(10)式和(9)式相乘并用核函数 $K(x_j, x_k^i)$ 代替相应的点积 $\langle \Phi(x_j), \Phi(x_k^i) \rangle$ 得:

$$w^T m_i^\Phi = \frac{1}{C_i} \sum_{j=1}^N \sum_{k=1}^{C_i} a_j K(x_j, x_k^i) = a^T \mu_i \quad (11)$$

(11)式中 $\mu_i = \frac{1}{C_i} \sum_{k=1}^{C_i} K(x_j, x_k^i)$ 这里,根据(7)式和(11)式,可得

$$w^T S_b^\Phi w = a^T M a \quad (12)$$

$$w^T S_w^\Phi w = a^T P a \quad (13)$$

(12)式和(13)式中 $M = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$, $P = P_1 + P_2$, $P_i = K_i K_i^T - C_i(\mu_i \mu_i^T)$; $i = 1, 2$ 。 K_i 为核函数矩阵, $(K_i)_{jk} = K(x_j, x_k^i)$ 其中 $i = 1, 2$; $j = 1, 2, \dots, N$; $k = 1, 2, \dots, N$; x_k^i 表示第 i 类第 k 个样本点。则(6)式可变为:

$$J(a) = \frac{a^T M a}{a^T P a} \quad (14)$$

根据广义 Rayleigh 熵并忽略比例因子得 $a = P^{-1}(\mu_1 - \mu_2)$, 特征空间 H 中任一向量 $\Phi(x)$ 在 Fisher 判定最优方向上的投影为^[7]:

$$w\Phi(x) = \sum_{i=1}^N a_i K(x_j, x) \quad (15)$$

选择合适的阈值可得在新特征空间 H 中的分类判别函数为:

$$f(x) = \text{sgn}[w^T \Phi(x) + b] = \text{sgn}\left[\sum_{i=1}^N a_i K(x_j, x) + b\right] \quad (16)$$

目标说话人的训练语音特征参数根据 Fisher 准则得到最优映射方向,其测试语音特征参数通过式(15)计算出投影,最后使用分类判别函数(16)式即可得出识别结果。

2.2 说话人辨认模型

在说话人辨认中,辨认模型是多类分类器,在此参照支持向量机构造多类分类器的方法,设计一个基于核 Fisher判别的多类分类器,其结构如图 2 所示。

在 N 个人的说话人辨认系统可按照图 2 进行分类。

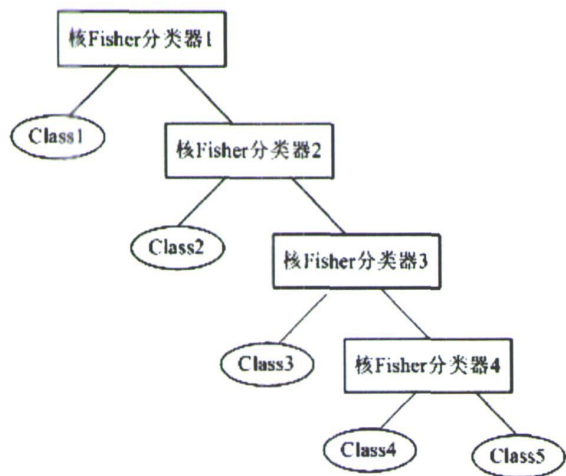


图2 基于核 Fisher判别的多类分类器

3 实验结果及分析

实验中采用自己录制的语音库,录音的人数为58人,其中男29人,女29人,采用普通的话筒进行录音,录音后的数据通过采样频率为8 kHz量化位数为16 bit单声道的A/D转换成数字信号存储。为了反映说话人的发声随时间变化,录入是间隔一段时间多次进行的,每个人都录制了12个语音段,每个语音段为30 s随机选取6个语音段用于训练寻找最优映射方向,剩余6个语音段用于测试。采用13维的MFCC及它的一阶偏导和二阶偏导构成39维输入特征向量。

实验中,核 Fisher判别函数的阈值由线性SVM估值,为了测试KFD的分类性能,首先将几种典型的核函数应用于KFD分别进行分类测试,结果显示基于径向基核函数(RBF)的Fisher判别方法分类识别率最高。然后,将传统的基于RBF的SVM方法、基于高斯混合模型法和基于RBF的核Fisher判别说话人辨认系统的识别率进行比较。

表1列出应用不同核函数的说话人辨认的识别率比较结果。

表1 不同核函数识别率比较

核函数	识别率/%
径向基核函数	88.6
三阶多项式核函数	75.2
线性核函数	65.7

在表1中很容易看出,在KFD算法中使用径向基核函数进行分类的系统的识别率,比使用三阶多项式核函数提高了13.4%,比使用线性核函数提高了22.9%。因此,在几种典型的核函数中,径向基核函数应用到KFD中进行分类的性能是最好的。表2列出基于RBF的核Fisher判别说话人辨认与几种传统的说话人辨认分类方法的比较结果。

表2 不同分类器识别率比较

核函数	识别率/%
基于径向基核函数支持向量机	88.3
基于径向基核函数核Fisher判别	89.7
基于高斯混合模型	87.6

在表2中,基于RBF核Fisher判别分类器的识别率,比基于RBF的支持向量机分类器的识别率提高了1.4%,比基于高斯混合模型分类器的识别率提高了2.1%。因此,本文提出的基于核Fisher判别技术的说话人辨认具有高的识别率。

4 结束语

通过对多元统计分析方法的研究,本文提出采用核Fisher判别技术来实现说话人辨认的方法。核Fisher判别方法在求解中使用了所有的训练样本,从而保证了其较好的分类性能。实验结果表明,核函数的选择对KFD分类性能的影响较大,基于RBF的KFD性能最好,且优于传统的说话人辨认方法。

参考文献

- 何听,刘重庆,李介谷.基于支撑向量机的文本无关的说话人识别的研究.计算机工程,2000;26(26):61-63
- 侯风雷,王炳锡.基于支持向量机的说话人辨认研究.通信学报,2002;23(6):61-67
- Mikat S, fisch G, Weston J. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop. Volume Issue Aug 1999; 41-48
- Chao Y,hsiang Tsai W,eiho Wang H,sim in. A kernelbased discrimination framework for solving hypothesis testing problems with application to speaker verification. Pattern Recognition. ICPR 2006. 18th International Conference.2006; 4: 229-232
- Yang Shu, Yan Shuicheng Zhang Chao. Bilinear analysis for kernel

selection and nonlinear feature extraction. IEEE Transactions on Neural Networks Accepted 2006

Fourth IEEE International Conference on Publication Date: Nov. 2004; 1-4

6 Park C H, Park H, Pardalos P. A comparative Study of linear and nonlinear feature extraction methods. Data Mining 2004. ICDM '04.

7 张宝昌,陈熙霖,山世光,等.基于支持向量的 Kernel判别分析.计算机学报,2006;29(12):2143-2150

Speaker Identification Based on Kernel Fisher Discriminant

TAN Ping^{1,2}, XING Yu-juan^{1,2}, LIM ing¹

(College of Computer and Communication Lanzhou University of Technology¹, Lanzhou 730050, P.R.China

School of Science and Engineering Gansu Lianhe University², Lanzhou 730000, China)

[Abstract] Aiming at inherent characteristic of phonetic data, a new method based on kernel fisher discriminant was proposed to implement speaker identification system. Kernel fisher discriminant uses all training samples in the process of classification, so it can improve the performance of system remarkably. The experiment results validate the system by comparing with the traditional speaker identification method.

[Key words] kernel fisher discriminant speaker identification SVM kernel function

(上接第 2205 页)

孔隙水压力消散的效果较明显。而且孔隙水压力的消散的速率与夯击能也有关,随夯击能的增大而增大。

⑥在模型试验中,通过静载和模拟强夯的对比可知,在强夯作用下排出水的体积较大,排水速度较快。而且排水的速度也随夯击能的增大而增大。

参 考 文 献

- 1 杨熙章.土工试验与原理.上海:同济大学出版社,1993:187-196
- 2 叶正强,张森安.性土的渗透规律性研究.东南大学学报:自然科学版;1999.(5);
- 3 高大钊.土力学与基础工程.北京:中国建筑工业出版社,1998:105-115
- 4 钱家欢,殷宗泽.土工原理与计算.北京:中国水利水电出版社,

- 1996; 107-155
- 5 黄文熙.土的工程性质.北京:水利电力出版社,1983:230-245
- 6 杨青.强夯法处理湿陷性黄土地基的试验研究土工基础.1995,2
- 7 朱晓羲,叶兴永.降水强夯处理软弱粉土地基试验研究.地基处理,1998;35-38
- 8 徐金明,等.强夯法加固软土地基的现场对比试验研究.工程勘察,1996;19-22
- 9 Mochizuki N A, et al. Field compaction test at a fill of weathered granite compaction method. Thomas Telford Ground Improvement Geosynthetics. London: 1997; 83-88
- 10 Nakaoka T, et al. Evaluation of results of improvement of ground by dynamic compaction method.50 and Foundations 1992; 40(5):35-40

Study on the Pore Pressure Dissipation Test and the Model Test for Clays

CHEN Chen, Zhang Guang-chao, GU Liang

(Central South University, Shanyang 410083, P.R. China)

[Abstract] In the laboratory, the static triaxial test is improved, study on rules of pore pressure dissipation. And the model test is designed, the rules of the pore pressure dissipation and permeability coefficient of the different pressure are learned.

[Key words] heavy tamping model test permeability coefficient pore pressure