

基于模糊 K 近邻的模糊支持向量机的语音情感识别

王吉林¹, 夏菽兰¹, 赵 力²

(1 盐城工学院 信息工程学院, 江苏 盐城 224051; 2 东南大学 信息科学与工程学院, 江苏 南京 210096)

摘 要: 基于模糊 K 近邻算法对模糊支持向量机中模糊类别隶属度的计算进行了改进, 将距离与样本之间的关系相结合进行加权弥补了 FSVM 算法的不足. 引入 CCA 算法对语音特征矢量进行降维处理, 有效减小了特征之间的冗余信息, 通过识别实验对传统的 SVM、FSVM 以及基于模糊 K 近邻的 FSVM 的算法性能进行了比较和分析.

关键词: 语音情感识别; 模糊 K 近邻; 模糊支持向量机

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-7180(2014)04-0031-05

Speech Emotion Recognition of FSVM Based on FKNN

WANG Ji-lin¹, XIA Shu-lan¹, ZHAO Li²

(1 Yancheng Institute of Technology, College of Information Engineering, Yancheng 224051, China;

2 College of Information Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: In the process of calculating the fuzzy class membership function, the distance and relationships between the samples are weighted based on fuzzy K neighbor algorithm. In the experiment, firstly, the dimensions of the speech feature vector are reduced by CCA method. Then, with the results of experiment, the performance of traditional SVM, Fuzzy SVM, Fuzzy SVM based on Fuzzy K neighbor nearest algorithm is compared.

Key words: speech emotion recognition; FKNN; FSVM

1 引言

Vapnik 和 Chervonenkis 在多年研究统计学习理论基础于 1970 年代提出支持向量机 (Support Vector Machine, SVM) 理论^[1-2]. SVM 是一种可训练的机器学习理论. SVM 在解决小样本、非线性和高维模式识别问题中有着良好的分类效果, 因而在语音情感识别、说话人识别、人脸识别等领域获得了广泛应用. 但是 SVM 算法从本质上来说是一种不适当问题的正则化理论和非线性规划计算方法, 当大量训练样本之间存在相互混叠时, 不仅造成支持向量数目急剧增加导致训练难度增大, 而且 SVM 分类性能也有明显下降. 对此构造基于模糊 K 近邻 (Fuzzy K-Nearest Neighbor, FKNN) 的模糊 SVM 算法并将其应用于语音情感识别, 通过对比实验验证此算法识别效果及其抗噪性能.

2 支持向量机

SVM 算法是统计学习理论的一种实现方式. 最基本思路就是要找到使测试样本的分类错误率达到最低的最佳超平面, 也就是要找到一个分割平面, 使得训练集中的训练样本距离该平面的距离尽可能的远, 该分割平面两侧的空白区域 (margin) 最大.

在 n 维空间 R^n 中, 对于两类问题进行分类时, 设输入空间中一组样本为 $(x_i, y_i), i = 1, \dots, n, x \in R^n, y \in \{+1, -1\}$ 是类别标号. 在线性可分的情况下, 存在多个超平面将两类样本分开, 其中可以使得两个类别中离它最近的样本与它的距离最大的那个超平面即最优超平面.

设超平面方程为

$$wx + b = 0 \quad (1)$$

它使得

收稿日期: 2013-05-29; **修回日期:** 2013-07-15

基金项目: 国家自然科学基金项目 (61273266)

$$\omega x_1 + b = 1 \quad \omega x_2 + b = -1 \quad (2)$$

进而

$$\frac{(\omega(x_1 - x_2))}{\|\omega\|} = \frac{2}{\|\omega\|} \quad (3)$$

则分类函数就是 $g(x) = \omega x + b$, 且分类函数归一化以后, 两类中的所有样本都满足 $|g(x)| \geq 1$, 距离分类超平面最近的样本的 $|g(x)| = 1$, 分类间隔即为 $\frac{2}{\|\omega\|}$, 当 $\|\omega\|$ 最小的时候, 分类间隔最大. 实际上, 寻找最优分类面的问题就简化成一个简单的优化问题.

如果样本是线性不可分的, 则可以通过非线性映射算法实现低维输入空间线性不可分样本到高维特征空间线性可分样本的映射, 再用针对线性可分情况的方法进行分析. 同样给定一组样本 (x_i, y_i) , $i = 1, \dots, n$, $x \in R^n$, $y \in \{+1, -1\}$, 假如 Φ 是低维输入空间 R^n 到高维特征空间 F 的一个映射, 核函数 k 对应高维特征 F 中向量内积运算, 从而将最优分类问题转化为一个约束条件下的二次优化问题

上面针对的是对于两类问题的分类. 如果需要对于 N 类问题进行分类, 则需要对 SVM 进行组合. 组合的策略有“one-vs-one”和“one-vs-all”. 若选用“one-vs-one”策略, 即每两类中训练一个分类器, 由于有 N 个类别, 一共需要 $N(N-1)/2$ 个 SVM.

选择使用不同的内积核函数, 相当于将输入样本投影到不同的高维特征空间之中, 对识别的效果会有影响, 常用的几种核函数有多项式形式的核函数、径向基形式的核函数以及 S 形核函数, 这三类核函数各有利弊, 而且其参数选择也很重要. 目前 SVM 技术还没有统一的核函数的选取标准. 一般在选取核函数及其参数时需要经过多次实验才能确定, 目前识别效果较好的是径向基 RBF 函数.

3 基于模糊 K 近邻的模糊支持向量机(FSVM)

3.1 FSVM 的基本原理

用传统的 SVM 方法对待测样本进行分类时, 与核判别分析一样都是将某一样本唯一隶属于某一类, 但是这与实际应用并不符合, 实际中, 不同的训练样本对于分类结果有着不同的影响. 训练样本集中的一些样本点对于分类影响很大, 但是同时也有一些点对于分类的影响作用可以忽略不计. 因此使用传统的 SVM 方法会使得分类存在决策风险. 在这样的前提下, 针对传统的 SVM 方法存在的误分

风险, 将模糊集理论引入 SVM, 提出具有模糊类别隶属度的 SVM 即模糊支持向量机(FSVM).

模糊类别隶属度是表征某一样本对于某一类别的隶属程度, 用 u_{ij} 表示, $u_{ij} \in [0, 1]$, u_{ij} 越大, 则表示对于某一类的隶属程度越高, 在进行分类时的影响更大.

假设原来样本集为

$X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n); x_i \in X, y_i \text{ 为类别标号}\}$, 加入模糊类隶属度之后成为

$X = \{(x_1, y_1, u(x_1)), (x_2, y_2, u(x_2)), \dots, (x_n, y_n, u(x_n))\}$, $u(x_i)$ 为模糊类别隶属度. 并通过映射关系 $\Phi(x)$ 将原始样本集从低维输入空间 R^n 映射到多维特征空间 F 中.

定义 δ_i 为测量错分程度的变量. $u(x_i)\delta_i$ 就是带权值的误差项, 可以用来表示重要性不同的变量的错分程度. 同传统 SVM 方法一样, 模糊 SVM 也是要找到最优超平面使得分类间隔最大. FSVM 的最优超平面问题就变成了在一个约束条件下的二次优化问题:

模糊类别隶属度可以用样本与最优超平面的距离来表示. 对于两类分类问题, 样本 x 和最优超平面 $g_{ij}(x)$ 的距离为 $d = g_{ij}(x)/\|\omega\|$, 当 $d \geq 1$ 或者 $d \leq -1$, 则认为样本 x 完全隶属于类别 i 或者类别 j , 当 $g_{ij}(x) \in (-1, 1)$ 时, 则可以认为是隶属于类别 i 或者类别 j 的隶属度.

以上是针对两类问题的 FSVM 算法, 对于多类问题进行分类时, 如果采用 one-vs-one 策略, 针对 N 类问题需要采用 $N(N-1)/2$ 个 SVM, 对样本划分模糊类别隶属度的过程如下:

(1) 对于输入样本 x , 如果 $|g_i(x)| \geq 1$, 则直接将样本 x 归入相对应的类别. 否则转入步骤(2);

(2) 如果 $g_i(x) > 0$, 且样本 x 不只隶属于某一类, 将样本 x 归入 $g_i(x)$ 最大所对应的那一类;

(3) 如果 $g_i(x) < 0$, 且样本 x 隶属于所有类, 则将 x 归入 $g_i(x)$ 绝对值最小对应的那一类.

3.2 基于 FKNN 的 FSVM

由前述的 FSVM 算法可知隶属度函数可以将对分类有重要影响的和非重要影响的样本分开, 是整个模糊算法中的重要部分, 隶属度函数必须是可以真实并且精确地反映样本的实际分布特性的. 因此, 在 FSVM 中, 类别隶属度函数的设计是一个关键问题. Huang 等人提出类别隶属度函数计算是基于样本与类中心的距离的, 将隶属度函数看成是样本与类中心距离的线性函数^[3]. 文献[4]使用 S 型函

数,将样本的隶属度函数和样本与类中心的距离看成是非线性的关系,但是实际中,基于样本与类中心之间距离来计算隶属度的方法在样本密度不一样的时候会引起误判,因此在设计隶属度函数的时候,不仅需要考虑到样本与类中心之间的距离,也要对样本之间的关系进行考虑。

在这种情况下,在计算类隶属度函数的时候采用FKNN,将样本与类中心的距离和样本之间的关系进行综合的考虑.首先通过样本与类中心的距离来计算样本的初始隶属度,然后采用模糊K近邻法对每个样本的K个近邻的隶属度进行计算排序,并与初始隶属度进行加权,这样就完成了样本之间的关系与隶属度计算的融合^[2].具体的步骤如下:

(1) 假设样本集为 $X = \{x_1, x_2, \dots, x_N\}$, 其样本均值为 x_c , 即为类中心, 将类中样本与类中心的距离的最大值定义为半径 $R = \max \|x_i - x_c\|$, 初始隶属度为

$$u_0(x_i) = 1 - \|x_i - x_c\| / (R + \delta) \quad (4)$$

为了防止 $u_0(x_i)$ 等于 0, 设 $\delta > 0$.

(2) 对于类中某一样本 x_i , 考虑其 k 个近邻 $x_1, x_2, \dots, x_k$, 定义样本的隶属度为

$$u(x_i) = \frac{\sum_{j=1}^k u_0(x_j) \left(\frac{1}{\|x_i - x_j\|} \right)^b}{\sum_{j=1}^k \left[\frac{1}{\|x_i - x_j\|} \right]^b} \quad (5)$$

从式(5)可以看出,在第二次求解隶属度的过程中,极大的削弱了样本自身的影响,当样本集密度较大时,则样本的隶属度不会受到第二次计算很大的影响,但是当样本集较为稀疏时,虽然已经考虑到样本之间的距离差别,但是样本的隶属度会受到周围K个近邻的极大影响,而样本自身所携带的隶属度信息被忽略掉,样本的隶属度偏差有可能会被过度修正.当样本空间中有噪声样本存在时,如果噪声样本和正常样本的K个近邻是相同的,则正常样本的隶属度和噪声样本有可能是相同的,图1所示为正常样本与噪声样本有两个相同近邻的情况.当 $k=2$ 时,正常样本与噪声样本的两个近邻都为 a 和 b ,很有可能两个样本的隶属度相同.

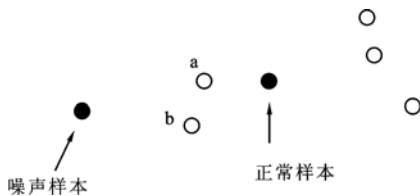


图1 正常样本与模糊样本在模糊K近邻($k=2$)时

为了避免这一情况的出现,在计算隶属度的时候一般采用初始隶属度和第二次计算的隶属度的加权作为最终的隶属度.将样本最终的隶属度定义为

$$u(x_i) = \alpha u_0(x_i) + (1 - \alpha) \frac{\sum_{j=1}^k u_0(x_j) \left(\frac{1}{\|x_i - x_j\|} \right)^b}{\sum_{j=1}^k \left(\frac{1}{\|x_i - x_j\|} \right)^b} \quad (6)$$

式中, α 为权重因子, 针对样本密度不同取不同的值, 一般经多次实验对比后确定. 从基于模糊K近邻定义的隶属度可以看出, 当近邻数目一定时, 样本与类中心的距离和隶属度成反相关关系, 距离越远, 隶属度越小, 反之越大; 样本集密度越大的时候, 则样本隶属度的偏差越小, 反之越大. 按照这种方法计算得到的模糊隶属度就用于模糊支持向量机分类中.

4 基于模糊K近邻的FSVM的语音情感识别

将构建的基于模糊K近邻算法计算模糊类别隶属度的模糊支持向量机应用到语音情感识别中. 主要的工作有两部分即基于CCA(典型相关分析, Canonical Correlation Analysis)对语音情感特征进行降维和基于模糊K近邻的FSVM的语音情感识别.

4.1 基于CCA的语音情感特征的降维

在语音情感识别中, 提取的特征参数一般包括有短时能量、短时幅度、短时过零率、基音频率、共振峰、MFCC等参数及其派生参数. 如果直接将这些参数组成特征向量, 由于其维数较高, 且包含了大量噪声信息, 使用这样的高维特征向量训练FSVM, 会导致FSVM的训练速度降低, 为此本文使用典型相关分析对情感特征矢量进行降维, 以降低后续进行FSVM算法的复杂度. CCA是考虑两组变量的线性组合, 并在组合中找到一对相关系数最大的组合, 用这个组合的相关关系来描述两组变量的相关性. 核心方法是根据判别准则, 找到投影方向, 使得在该投影方向上两组向量之间的相关性最大^[5].

假设 $f_x \in R^{n_x}$, $f_y \in R^{n_y}$ 是两组语音情感特征矢量, 采用CCA方法的目的就是找到投影方向 w_x 和投影方向 w_y , 使得在这两个投影方向上, 两组矢量之间的相关性最大. 即满足:

$$\rho(f_x, f_y, w_x, w_y) = \frac{E\{w_x^T f_x f_y^T w_y\}}{\sqrt{E\{w_x^T f_x f_x^T w_x\}} \sqrt{E\{w_y^T f_y f_y^T w_y\}}} \quad (7)$$

首先计算第一对投影方向 w_x^1 和 w_y^1 , 使得两组矢量在这两个方向上的投影之间的相关度最大, 称 $(w_x^1)^T f_x$ 和 $(w_y^1)^T f_y$ 为第一对典型相关特征, 以此类推求得 CCA 所求解的 k 组投影矢量. 令

$$w_x = [w_x^1, w_x^2, \dots, w_x^k] w_y = [w_y^1, w_y^2, \dots, w_y^k] \quad (8)$$

f_x, f_y 在这两个投影方向上的投影值为

$$\rho_x = w_x^T f_x, \rho_y = w_y^T f_y \quad (9)$$

通过对 f_x 和 f_y 进行典型相关分析, 得到包含有效判别信息的 k 对典型相关特征向量, 不仅对原特征向量进行了降维, 也消除了特征之间的冗余信息.

4.2 基于模糊 K 近邻的 FSVM 的语音情感识别

基于模糊 K 近邻的模糊支持向量机的语音情感识别流程如图 2 所示.

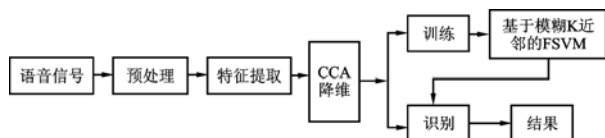


图2 基于模糊 K 近邻的 FSVM 的语音情感识别流程

对于输入的语音信号, 经过预处理中的预加重, 分帧加窗等后, 进行特征提取, 并使用 CCA 方法进行降维, 最后再使用基于模糊 K 近邻的 FSVM 进行分类, 得到最终的识别结果.

计算类别隶属度的过程中, 需要寻找样本的 K 个近邻, 首先是存储所有的样本, 对于待测样本需要计算它与其他所有训练样本的距离, 再进行排序, 因此计算量与存储量都比较大, 因此在求解样本到类中心距离的过程中, 使用二分法进行查找. 该方法的具体过程是: 首先计算全部训练样本到类中心的距离并进行排序, 再计算样本与类中心之间的距离, 并将距离在序列中进行插入, 找到其附近的 K 个近邻. 将过程反应在二维空间, 就是在同心圆中找到近邻的过程, 如图 3 所示.

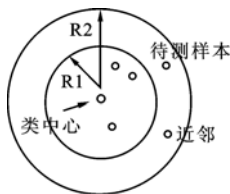


图3 二维空间中的二分法查找

在训练 FSVM 的过程中, 本文将对四种情感进行识别, 根据“one-vs-one”策略需要构造 6 个 FSVM 分类器.

5 识别实验

5.1 实验环境与语音情感数据库

本文的实验仿真平台为 PC2.6 GHz/1 GB, Windows XP2003 操作系统/Matlab 7.0, 并结合语音工具箱 Voicebox 和 SVM Toolbox1.0. 实验采用的情感语音库为自行录制的汉语情感语音库, 语音库中包含有 10 名录音人员(5 男 5 女), 每名表演人员分别用 5 种不同情感(高兴、愤怒、中性、悲伤、害怕)朗读的 20 句不同文本的汉语, 总共包含的有效语句为 624 句, 采样频率为 16 kHz, 采样精度为 16 位.

5.2 情感特征的提取

实验对四种情感(高兴、愤怒、中性、悲伤)进行识别, 每类情感取 60 句, 30 句用于训练, 30 句用于测试. 提取的情感特征参数为短时能量、短时振幅、共振峰、基音频率以及 MFCC 及其派生的特征参数, 总共 42 维作为特征向量.

5.3 实验结果

5.3.1 识别效果

实验中将传统的 SVM 方法、模糊支持向量机与基于模糊 K 近邻的模糊支持向量机算法进行了比较. 径向基核函数的分类效果比较好, 因此在实验中采用, 并设 $\sigma=1.3$. 惩罚系数 $C=200$. 计算初始隶属度时取 $\delta=0.1$, 基于模糊 K 近邻算法计算模糊类别隶属度时取 $k=13$, 初始隶属度与第二次计算的隶属度进行加权时, 权重因子取 $\alpha=0.9$. 实验结果如表 1 所示.

识别方法	愤怒	高兴	中性	悲伤	平均识别率
SVM	79.13	80.25	78.94	76.51	78.71
FSVM	81.34	82.59	80.32	79.95	81.05
FKNN+FSVM	83.75	84.68	82.31	81.47	83.05

从实验结果中可以看出, 本文提出的基于模糊 K 近邻的模糊支持向量机的方法较传统的 SVM 和模糊支持向量机的平均识别率有了提高, 比 SVM 的识别率提高了 4.34%, 比 FSVM 的识别率提高了 2%; 在对四种情感的识别中, FKNN+FSVM 较 SVM 和 FSVM 的识别率也有相应的提高, 对于愤怒情感, 比 SVM 提高了 4.62%, 比 FSVM 提高了 2.41%, 对于高兴情感, 比 SVM 提高了 4.43%, 比 FSVM 提高了 2.09%, 对于中性, 比传统的 SVM 提高了 3.37%, 比 FSVM 提高了 1.99%, 对于悲伤,

比SVM提高了4.96%,比FSVM提高了1.52%。从中也可以看出四种情感中,对于愤怒和高兴的识别率较高,主要是因为这两种情感的发音特征较明显,而中性与悲伤的有些生理特征较相似,容易混淆。

5.3.2 抗噪性能实验

为了验证基于模糊K近邻的模糊支持向量机方法的抗噪性能,本文实验中对无噪声的120句测试语句添加了高斯白噪声,选择的信噪比分别为:35 dB,30 dB,25 dB,20 dB,15 dB,10 dB,5 dB。

从实验结果中可以看出,随着噪声的增加,三种算法的识别率都有下降,但是基于模糊K近邻的模糊支持向量机下降的幅度小于其他两种方法,且平均的识别率高于其他两种方法。因此基于模糊K近邻的模糊支持向量机具有良好的抗噪性能。

6 结束语

本文基于模糊K近邻对模糊支持向量机进行了改进并应用于语音情感识别,并通过实验比较分析了支持向量机、模糊支持向量机和改进的模糊支持向量机的算法性能,证明了基于模糊K近邻的模

糊支持向量机有较好的识别效果和抗噪性能。

参考文献:

- [1] Vapnik V N. Statistical learning theory[M]. New York: Wiley,1998.
- [2] Christopher J C Burges. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge,1998,2(2):121-167.
- [3] Huang H P, Liu Y H. Fuzzy support vector machines for pattern recognition and data mining [J]. International Journal of Fuzzy Systems,2002,4(3): 826-835.
- [4] 边肇祺,张学工. 模式识别[M]. 2版. 北京:清华大学出版社,2000.
- [5] Perez M, Rubin D M, Scott L E, et al. A hybrid fuzzy-SVM classifier, applied to gene expression profiling for automated leukaemia diagnosis [C]//Proceedings of the IEEE Conference Israel, Eilat, Israel: IEEE, 2008.

作者简介:

王吉林 男,(1966—),硕士,副教授. 研究方向为电子技术与信息处理.

赵力 男,(1958—),博士,教授,博士生导师. 研究方向为信号与信息处理.

(上接第30页)

- [2] Kessler PB. Fast breakpoints: Design and implementation[C]//ACM SIGPLAN Conference on Programming Language Design and Implementation. White Plains: ACM Press, 1990:78-84.
- [3] 刘建. 一种具有快速条件断点的并行程序调试器. 软件学报[J]. 2003,14(11):1827-1833.
- [4] Brown J, Zosel M, Zwakenberg R, et al. ASCI debugging requirements. Los Alamos National Laboratory,

1998[EB/OL]. [2013-03-21]. <http://www.lanl.gov/projects/asci/PSE/ASCIdebug.html>.

- [5] Bryant R E. Graph-based algorithms for boolean function manipulation[J]. IEEE Transactions on Computers,1986,c-35(8):677-691.

作者简介:

朱浩 男,(1987—),博士研究生. 研究方向为代码优化.