

基于模糊 K 近邻的语音情感识别纪正飏¹, 王吉林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: 传统的 K 近邻算法存在误判风险, 针对其不足提出了一种基于模糊 K 近邻的语音情感识别算法, 通过引入模糊隶属度的概念, 求出不同的特征参数对于不同情感识别的贡献度, 并将其与欧式距离加权应用于语音情感识别中, 实验验证了算法的有效性。

关键词: 语音情感识别; 模糊类别隶属度; 模糊 K 近邻

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2015)03-0059-04

Speech Emotion Recognition Based on FKNN

JI Zheng-biao¹, WANG Ji-lin¹, ZHAO Li²

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

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

Abstract: To solve the limitation that the traditional K -nearest neighbor algorithm may divide a sample to a wrong class, the concept of fuzzy class membership function is used to improve the K -nearest neighbor algorithm. The experiment in speech emotion recognition is implemented and the results show the effectiveness of the improved algorithm.

Key words: speech emotion recognition; fuzzy class membership function; FKNN(fuzzy K -Nearest Neighbor)

1 引言

K 近邻算法是对近邻法的扩展, 将最近邻算法中距离最近的一个样本扩展成为距离最近的 K 个样本, 把出现最多的样本类别作为待测样本的类别, 但 K 近邻法也存在缺点, 例如需要考虑全部样本, 存在决策风险等。为了解决传统 K 近邻法的局限性, 本文基于模糊集理论对 K 近邻算法进行改进, 并将改进后的模糊 K 近邻算法应用于情感语音识别。

2 K 近邻分类算法以及模糊集理论2.1 K 近邻分类算法

K 近邻(K -Nearest Neighbor, KNN)算法是一种基于实例的统计学习方法^[1], 其基本思想是在多维空间 R^n 中, 对于一个待分类样本 x , 根据距离找到与未知样本最近邻的 k 个训练样本点, 然后根据

这 k 个近邻样本点的类别决定待分类样本的类别。

对于 c 类问题, 设类 $\omega_i (i = 1, 2, \dots, c)$ 有已知类别的 N_i 个训练样本 $x_j^{(i)} (j = 1, 2, \dots, c)$, 对于一个待分类样本 x , 分别计算它与 $N = \sum_{i=1}^c N_i$ 个训练样本 $x_j^{(i)}$ 的距离, 一般采用欧式距离计算。考察其中 K 个最近邻样本, 采用多数表决制, 其中哪一类的样本最多, 就将 x 判定为哪一类。

K 近邻算法的具体实现包括两个阶段。(1) 训练阶段: 对训练样本进行离散化, 读取和存储;(2) 分类阶段: 首先对于待分类的样本, 计算它与其他的训练样本的欧式距离, 选择其中距离最近的 K 个样本, 并查看这 K 个近邻的分类标记, 根据多数表决原则确定待测样本的类别。重复以上步骤直至所有待分类的样本分类结束。

收稿日期: 2014-05-27; 修回日期: 2014-07-15

基金项目: 国家自然科学基金项目(61273266)

2.2 模糊集理论

为了解决 K 近邻算法存在的不足,将模糊集理论引入 K 近邻分类器^[2]. 模糊集理论立足于实际生活中知识表达、学习、归纳等方面的不精确和不完整的缺陷,可解决单靠经典逻辑不能解决的问题.

对于给定论域为 U ,其任意模糊子集为 A ,任意 $x \in U$,都存在一个数 $u_A(x) \in [0,1]$, $u_A(x)$ 称为 x 对 A 的隶属程度. 映射 $u_A(x): U \rightarrow [0,1], x \rightarrow u_A(x)$ 称为模糊子集 A 的隶属函数,模糊子集完全由其隶属函数所划定.

在应用模糊集理论的模式识别与分类中,某一样本以一定的隶属程度属于某一类,而不是明确的被划分为属于某一类或者不属于某一类. 用样本对各类别的隶属函数来描述样本的隶属情况更符合实际. 假设集合 S 中有 N 个样本,分别属于 C 个类别,某一样本 x_j 对于类别 i 的隶属程度为 u_{ij} , u_{ij} 应满足条件为 $u_{ij} \in [0,1]$, $\sum_{i=1}^C u_{ij} = 1$ 及 $0 < \sum_{j=1}^N u_{ij} < N$.

3 模糊 K 近邻分类算法

模糊 K 近邻算法对 K 近邻算法的改进在于引入了类别隶属度的概念,而不是像 K 近邻法那样简单的将样本唯一的划分为某一类^[3]. 待测样本隶属于不同类别的隶属度定义如式 1.

$$u_i(x) = \frac{\sum_{j=1}^k u_i[x^{(j)}] [\|x - x^{(j)}\|^{-2/(m-1)}]}{\sum_{j=1}^k [\|x - x^{(j)}\|^{-2/(m-1)}]} \quad (1)$$

在 n 维空间 R^n 中,对于 c 类问题,训练样本集合为 $X = \{X_1, X_2, \dots, X_N\}$, N 为样本数. 对于待测样本 x ,模糊 K 近邻的具体算法步骤如下:

(1) 确定待测样本的 K 个近邻的 K 值;

(2) 确定待测样本与所有训练样本的距离,本文选择采用欧式距离:

$$d = d(x, X_i) = \sqrt{\|x - X_i\|^2} \quad (2)$$

(3) 对这 N 个距离进行排序,

$$d(1) \leq d(2) \leq d(3) \leq \dots \leq d(k) \leq d(k+1) \leq \dots \leq d(N) \quad (3)$$

式中, $d(1), \dots, d(k)$ 就是待测样本与 K 个最近邻的距离.

(4) 根据式(1) 计算待测样本的类别隶属度,其中, m 为模糊权重调节因子. 对于 $n = 1, 2, \dots, c$; 如果 $u_i(x) = \max\{u_n(x)\}$, 则判定 x 属于第 i 类. 重复上

述算法直至所有待测样本都处理完毕.

近年来,模糊 K 近邻算法也被广泛应用于各个涉及到模式识别的领域. 例如在葛根类药材的模式识别中,采用模糊 K 近邻分类方法,识别的正确率达到了 100%^[4].

4 基于 FKNN 的语音情感识别

本文将模糊 K 近邻算法应用于语音情感识别研究其对识别效果的改善情况,主要工作包括两个方面即语音情感特征的提取与基于模糊 K 近邻算法的语音情感识别.

4.1 语音情感特征的提取

本文选取的语音情感特征为短时能量,短时幅度和短时过零率和基音频率四个参数. 短时能量、短时幅度和短时过零率由函数 TimePara() 来进行提取. 其流程如图 1 所示.

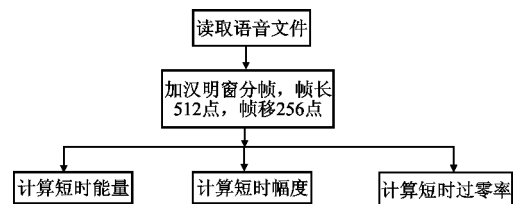


图 1 提取短时能量、短时幅度、短时过零率的流程图

基音频率由函数 FunFre() 进行提取. 该函数采用的是频谱压缩法,选择的窗为汉宁窗. 在对短时能量、短时幅度、短时过零率和基音频率分别进行提取之后,将提取到的特征参数组成特征向量,作为模糊 K 近邻分类算法的输入.

4.2 基于模糊 K 近邻的语音情感识别

在训练样本阶段,不同于传统的 K 近邻分类方法,模糊 K 近邻需要计算训练样本对于不同类别的隶属程度,主要是通过不同情感特征参数对于不同情感类别识别的贡献度来实现的. 研究证明,在语音情感识别中,不同的情感特征参数对不同的情感起到的作用是不同的,例如,对于短时能量这个参数来说,它可以很好的区分中性和悲伤这两种情感,中性与悲伤两种情感的基音频率比较相近,很难通过基音频率来进行区分,但是它们的能量具有较大差异,中性的能量明显大于悲伤的能量,因此在区分中性与悲伤情感时,能量的贡献度比基音频率大. 再者可以通过基音频率来区分愤怒和高兴这两种情感,愤怒和高兴的能量都比较大,但是愤怒的基音频率明显高于高兴,因此可以计算不同情感下不同情感特征的贡献度来对情感进行更准确的区分^[5].

要区分不同的情感特征参数的贡献度,首先要统计情感特征参数对于每种情感的离散度,离散度越高,则说明该情感特征参数在区分此类情感时的不确定性越高,即对于该种情感识别的贡献度越小.相反,如果离散度越低,则说明该情感特征参数对于此类情感识别的稳定性较好,贡献度越大.

结合上述思想,情感特征参数的贡献度具体算法步骤如下:

(1) 对于C类情感的识别,首先对训练样本集X统计出C种不同情感状态下同一个特征参数的平均值,记为 M_{ij} ($i=1,2,\dots,C, j=1,2,\dots,N, N$ 为情感特征参数的个数),再分别对每一种情感状态下的每一句语音样本的每个特征参数 M_{ijn} (n 为该情感状态下的样本, $n=1$ 表示为第一句,以此类推)进行归一化,归一化公式如式(4).

$$A_{ijn} = \frac{M_{ijn}}{\sum_{i=1}^c M_{ij}} \quad (4)$$

(2) 然后计算出某种特定情感下该特征参数的离散度:

$$\theta_{ij} = \sqrt{\sum_{k=1}^n A_{ijk}^2} \quad (5)$$

(3) 求出每种情感下每个特征参数的离散度后,根据离散度的大小来计算每种情感下每个特征参数的贡献度大小.特征参数 θ_{ij} 的贡献度 u_{ij} :

$$u_{ij} = \frac{\theta_{ij}}{\omega_i}, \omega_i = \sum_{l=1}^j \theta_{il} \quad (6)$$

并在对待分类样本使用模糊K近邻进行判别时将情感特征参数的贡献度与欧式距离加权.

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} d(x, X_j)}{\sum_{j=1}^k d(x, X_j)} \quad (7)$$

将情感特征参数对不同情感识别的贡献度作为一个加权值,在计算时加以考虑,不仅保持了K近邻算法简单易行的优点,也突出了每一情感特征参数所起作用的差异性,以及不同情感特征参数的内在联系,提高了K近邻算法的分类精确度,改善了语音情感识别的效果.

5 基于FKNN的语音情感识别实验

5.1 实验环境与语音情感数据库

实验仿真实验平台为PC 2.6G/1G, Windows2003XP操作系统/MATLAB7.0,并结合语音工具箱Voicebox.实验选用的语音库为自行录制的

汉语情感语音库,采样频率为16 kHz,16bit量化.语音库由十名录音人员(5男5女)在五种不同的情感状态(高兴(Happiness)、愤怒(Anger)、中性(Neuter)、悲伤(Sadness)、害怕(Fear))下朗读的20句不同文本的汉语语音组成,经有效性验证后总共包含624条有效情感语音.实验选取其中的四种情感(高兴、愤怒、中性、悲伤)进行识别.实验过程选取的情感特征为短时能量、短时幅度、短时过零率和基音频率.为了降低个体在情感表达上的差异对识别结果造成的影响,对提取到的语音情感特征进行归一化处理,并将归一化后的情感特征作为实验中的训练样本和测试样本.

5.2 情感特征的提取

在训练阶段对于四种情感,各选取了个句子作为样本进行特征参数提取即对高兴、愤怒、中性、悲伤提取短时平均振幅能量、短时平均能量、短时过零率和基音频率^[6].仿真实验中可以看出中性和悲伤的基音频率比较接近,但是短时平均能量有着明显的差异.而愤怒和高兴的短时平均能量比较接近,但是愤怒的基音频率更高,这说明使用特征参数的贡献度与欧式距离进行加权的模糊K近邻算法来进行语音情感识别是合理的.对四种情感的特征参数进行提取以后,按4.2中的方法求取各个特征参数对于不同情感状态识别的贡献度,如表1所示.

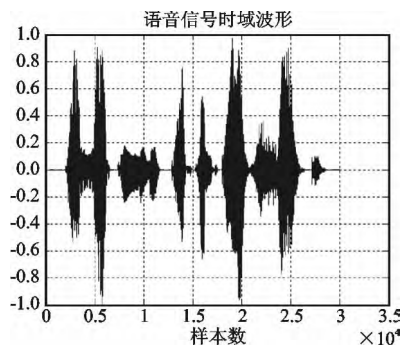
表1 各特征参数对于情感识别的贡献度

贡献度 情感状态	短时 能量	短时平均 振幅	短时 过零率	基音 频率
愤怒	0.558 7	0.732 4	0.963 6	0.639 6
悲伤	0.606 8	0.746 0	0.973 9	0.714 4
高兴	0.586 3	0.754 8	0.969 7	0.659 4
中性	0.623 7	0.756 3	0.953 4	0.705 9

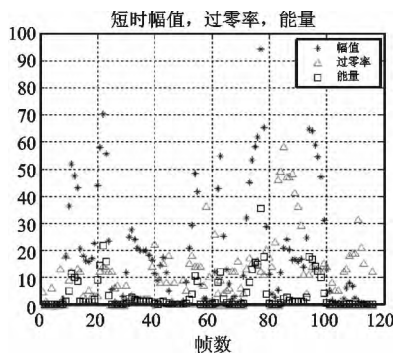
在识别阶段,从四种已知的情感语句中分别随机选择30句总共120句语音进行识别并统计识别结果,记录数值和做成图表形式进行分析比较.对其中任意一段语音信号进行特征参数提取,结果如图2所示.

5.3 实验结果与分析

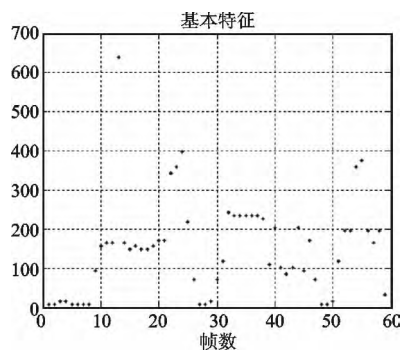
实验采用了传统的K近邻算法与模糊K近邻算法^[6]进行识别,并通过实验结果将两种方法进行了对比,分析各自的特点.在采用传统的K近邻与模糊K近邻算法时,都需要考虑k的选取,本文进行了两组实验,第一组实验取 $k=7$,第二组取 $k=13$.识别结果如表2、表3所示.



(a) 语音信号的时域波形



(b) 提取短时幅值, 短时过零率和短时能量



(c) 对语音信号进行基音频率的提取

图 2 对语音信号进行特征参数提取

表 2 $k=7$ 时两种算法的识别结果

识别方法	愤怒/%	高兴/%	中性/%	悲伤/%	平均识别率/%
KNN	82.53	81.28	76.47	78.14	79.61
FKNN	84.27	82.59	77.51	80.47	81.21

表 3 $k=13$ 时两种算法的识别结果

识别方法	愤怒/%	高兴/%	中性/%	悲伤/%	平均识别率/%
KNN	83.29	82.36	78.53	79.62	80.95
FKNN	85.32	83.59	80.14	81.47	82.63

从实验结果中可以看出,在 $k=13$ 时的平均识别率比 $k=7$ 时有所提高,四种情感的识别率也分别有不同程度的提高,这主要因为考虑了更多近邻,减

小了误判的风险,但是同时也增加了运算量.同时可以看出,传统的 K 近邻算法与模糊 K 近邻算法的识别率都比较高.对于四种情感识别中,对于愤怒的识别效果最好,主要是因为愤怒的情感特征相对于其他三种情感更为明显,在表达愤怒情感.

说话人的语速往往比较快,音调也比较高,对于中性和悲伤的识别率相对较低,主要是在悲伤和中性状态下,某些生理特征比较相近,容易混淆误分.从两种算法的对比也可以看出,模糊 K 近邻法与传统 K 近邻算法相比,在计算欧式距离时充分考虑了各个参数之间的权重比例,使得识别效果更加突出,识别率有了一定的提高.

6 结束语

针对传统 K 近邻算法不能突出不同的情感特征参数起不同作用的不足,提出了一种基于模糊 K 近邻的语音情感识别算法,通过汉语情感语音语音库进行语音情感识别实验,结果表明相对于传统的 K 近邻算法,模糊 K 近邻由于考虑了各个参数之间的权重比例,提高了识别率,可以获得更好的识别效果.

参考文献:

- [1] Davis S B, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. IEEE Transactions on Acoustic, Speech and Signal Processing, 1980, AS-SP-28(4):357-366.
- [2] Wang Hui. Nearest neighbors by neighborhood counting [J]. Pattern analysis and machine intelligence, IEEE Transaction on, 2006, 28(6):942-953.
- [3] Gau Wenlung, Buehrer Daniel J. Vague sets[J]. IEEE Transactions on systems, Man, and Cybernetics, 1993, 23(2):610-614.
- [4] 杨铭. 基于模糊 K 最近邻规则的葛根类药材的模式识别[J]. 中国生物医学工程学报, 2007, 21(5):839-840.
- [5] 朱海, 赵佳宝. 模糊 K 近邻法及其在肤色识别中的应用[J]. 计算机光盘软件与应用, 2011(23):165-166.
- [6] 王吉林, 夏菽兰, 赵力. 基于模糊 K 近邻的模糊支持向量机的语音情感识别[J]. 微电子学与计算机, 2014, 31(4): 31-35.

作者简介:

纪正飏 男, (1981-), 硕士, 讲师. 研究方向为信号与信息处理. E-mail: wjlycjs@126.com.

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.