

一种基于脉冲耦合神经网络的语音情感识别新方法

梁 泽¹, 马义德¹, 张恩溯¹, 朱望飞^{1,2}, 汤书森¹

(1. 兰州大学 信息科学与工程学院, 电路与系统研究所, 兰州 730000; 2. 中国华阴兵器实验中心, 陕西 渭南 714200)

(liangz05@lzu.cn)

摘 要:针对汉语语音情感识别问题,提出了一种基于脉冲耦合神经网络(PCNN)的识别方法。该方法将语音转化为语谱图后输入到 PCNN,得到输出图像的神经元点火序列及其熵序列作为语音情感的特征,利用其特征实现语音情感识别。实验结果表明,该方法可以有效地识别“高兴”与“平常”这两种不同的情感。该方法将 PCNN 引入到语音情感识别的应用研究中,开拓了语音和图像信号结合处理的新领域,同时对于 PCNN 的理论研究和实际应用具有重要的现实意义。

关键词:脉冲耦合神经网络;语音情感识别;语谱图;神经元点火序列;熵序列

中图分类号: TP18; TP391 **文献标志码:** A

Innovative method for emotion recognition of speech based on pulse coupled neural network

LIANG Ze¹, MA Yi-de¹, ZHANG En-su¹, ZHU Wang-fei^{1,2}, TANG Shu-sen¹

(School of Information Science and Engineering, Lanzhou University, Lanzhou Gansu 730000, China;

2. Chinese Huayin Weapon Test Center, Weinan Shaanxi 714200, China)

Abstract: An innovative method for classifying emotional states of neutral or happy of Chinese phonetics by using Pulse Coupled Neural Network (PCNN) was proposed. The entropy series and neurons firing series of the image obtained by feeding the spectrogram into PCNN were used as the characteristics of emotion speech for emotion recognition. The experimental result show that this method can distinguish speaker's emotions (normal & happy) in speech effectively. The application of PCNN in emotion recognition of speech expands the combination of two important parts of signal processing-speech and image processing and it is significant for theoretical research and application of PCNN.

Key words: Pulse Coupled Neural Network (PCNN); emotion recognition of speech; spectrogram; neurons firing series; entropy series

0 引言

语音是由发音器官运动作用产生的,每个人讲话时使用的发声器官,比如舌、牙齿、喉、肺、鼻腔和声道在尺寸和形态方面的差异很大,所以不同人对同一个字的发音会有不同^[1]⁸¹;对同一个人来说,其讲话时所带有的不同情感也会导致不同的发音,致使相同的语言内容会有较大的语义差异,从而不同的情感导致相同语言内容的语义多样化。情感可以简单地划分四类:“平常”、“高兴”、“悲伤”和“生气”,其中“平常”与“高兴”是人类接触最多的两类。

目前的语音合成与识别技术都只是根据发音的简单规则做出文字层面的处理,很难根据语音本身的意义或者讲话人的不同情绪状态做出适当的调整,也就是基本上没有办法做到语义层面的处理。如何能准确地分辨出语音属于那类情感是语音技术从“文字”到“语义”的关键,而国内在这个领域的研究几乎是空白。

脉冲耦合神经网络(Pulse Coupled Neural Network, PCNN)是源于对小型哺乳动物视神经研究而提出的一种新型神经网络模型^[2],其也被称为第三代人工神经网络,它是在

生物视觉皮层模型的启发下产生由若干个神经元互连而构成的反馈型网络^[3]。与传统的神经网络不同的是:它是由脉冲耦合神经元构成的二维单层神经元阵列模型,同时也是一种以迭代算法为主的神经网络,更重要的是 PCNN 是一种自监督自学习的网络,不需要提前进行训练,因此非常适合于实时图像处理环境。

本文提出了一种基于 PCNN 神经元点火序列和熵序列的辨别汉语语音中平常情感与高兴情感的新方法:先将一维的语音信号通过语谱图(Spectrogram)^[4]算法,转化为二维的图像信号;再将得到的图像送入 PCNN 进行迭代处理,得到由 PCNN 输出描述图像信息的特征时间序列——点火时间序列以及熵值序列;之后利用点火时间序列的距离阈值进行情感差别判断;最后通过计算熵值序列的积分值来实现情感辨别。实验结果表明,该方法在平常情感与高兴情感的识别方面能取得良好的效果。

1 基于 PCNN 的语音情感辨别算法

1.1 语谱图算法

语谱图是语音信号进行时频分析后得到二维图像,是一

收稿日期:2007-09-26;修回日期:2007-12-21。

作者简介:梁泽(1983-),男,甘肃张掖人,硕士研究生,主要研究方向:数字信号处理、人工智能; 马义德(1963-),男,甘肃临夏人,教授,博士生导师,主要研究方向:计算机应用系统、生物医学图像处理、DSP与信号处理、人工智能; 张恩溯(1982-),男,甘肃兰州人,硕士研究生,主要研究方向:数字图像处理、混合信号系统; 朱望飞(1979-),男,湖南岳阳人,工程师,主要研究方向:数字图像处理、运动目标检测; 汤书森(1970-),男,陕西山阳人,工程师,主要研究方向:嵌入式系统、数字图像处理。

种能反映语音信号动态频谱特性的时频图,被称为“可视语言”。图像的每个像素的灰度值大小表示相应时刻和相应频率的信号能量密度^{[1]82}。从语谱图可以得到一些频域分析参数(如共振峰、基音周期等)随语音发生过程(时间)的变化情况;同时还可以得到能量随语音发生过程(时间)的变化情况。

为了获得语谱图,首先将语音信号分割为具有相互重叠的若干段,对每一段进行加窗处理,然后进行短时傅立叶变换,得到这个信号的短时频谱估计。语谱图的纵轴包含时间函数的谱估计。本文作者选择长度为256,交叠部分为窗长1/2的窗函数来生成语谱图。窗函数使用Hanning窗(汉宁窗),见式(1):

$$w(n) = \begin{cases} 0.5 \left[1 - \cos\left(\frac{2n\pi}{N-1}\right) \right], & 0 \leq n \leq N-1 \\ 0, & n < 0 \text{ 或 } n > N \end{cases} \quad (1)$$

信号 $S(t)$ 的短时傅立叶变换(STFT)定义如式(2),其中 $w(t)$ 为某种窗函数,在任意时刻 n 的离散时间STFT表达式如式(3),离散STFT可以通过频率抽样式(4)获得:

$$S(t, f) = \int_{-\infty}^{+\infty} s(\tau) w(t - \tau) e^{-j2\pi f\tau} d\tau \quad (2)$$

$$S(n, k) = \sum_{m=-\infty}^{+\infty} s(m) w(n - m) e^{-j2\pi km} \quad (3)$$

$$S(n, k) = S(t, f) |_{f=k/N} \quad (4)$$

其中, N 是窗函数中数据点的总数,也是频率抽样因子。把式(4)带入式(3),获得离散STFT,如式(5)所示:

$$S(n, k) = \sum_{m=-\infty}^{+\infty} s(m) w(n - m) e^{-j2\pi km/N} \quad (5)$$

最终得到的语谱图是用时间 n 作横坐标,频率 ω 作纵坐标,将能量密度谱函数的值 $S(n, k)$ 表示为灰度级的二维图像。

1.2 PCNN 模型算法

PCNN 的基础是 Eckhorn 的连接域模型^[5]。一种传统的 PCNN 神经元模型如图1所示。接收部分接收到来自其他神经元与外部的刺激输入后,将其通过两条通道传输:一条称为馈送输入通道 F ,另一条称为链接输入通道 L 。 F 通道的脉冲响应函数随时间的变化比 L 通道慢,对应着外界输入;而 L 是对相邻神经元输出脉冲的局部抽样,链接接收域的半径通常很小。链接输入 L 加上一个正的偏移量,然后和馈送输入相乘形成内部活动项 U 。脉冲发生器由一个阶跃函数和一个衰减的内部阈值 θ 组成^{[1]82}。

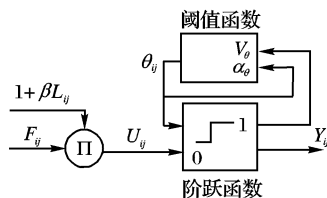


图1 传统的PCNN神经元模型

如果把二维图像矩阵 $M \times N$ 理解为 $M \times N$ 个PCNN神经元模型,每个神经元按照式(6)~(10)进行迭代循环运算,实现神经网络的各种功能。

$$F_{ij}[n] = \exp(-\alpha_F) F_{ij}[n-1] + V_F \sum m_{ijkl} Y_{kl}[n-1] + S_{ij} \quad (6)$$

$$L_{ij}[n] = \exp(-\alpha_L) L_{ij}[n-1] + V_L \sum w_{ijkl} Y_{kl}[n-1] \quad (7)$$

$$U_{ij}[n] = F_{ij}[n] (1 + \beta L_{ij}[n]) \quad (8)$$

$$Y_{ij}[n] = \begin{cases} 1, & \text{if } U_{ij}[n] > \theta_{ij}[n] \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$\theta_{ij}[n] = \exp(-\alpha_\theta) \theta_{ij}[n-1] + V_\theta Y_{ij}[n-1] \quad (10)$$

式(6)中 S_{ij} 是输入刺激信号,这里取图像像素构成的矩阵中第 i, j 像素的灰度值。 F_{ij} 是第 i, j 神经元的馈送输入, L_{ij} 是链接输入, U_{ij} 代表神经元的内部活动项, β 为耦合系数,利用链接输入对馈送输入进行调制,它是PCNN神经元之间进行三通信的关键。 θ_{ij} 为动态门限阈值,它与内部活动项相比较得到神经元的输出。 Y_{ij} 是PCNN神经元的脉冲输出值,我们定义神经元的输出由0变1为神经元的点火(即输出脉冲)。

由于PCNN本身对图像细节有较好的保留特性^[6],这些特性是其他的传统神经网络所不具备的,利用这种良好的细节保留特性,能最大限度地保存语音信号中所包含的特征。将一幅二维语谱图像输入PCNN模型,将产生一系列的二值图像输出,这些输出可以转化成特定的具有某种不变性和某种唯一性的一维时间序列,这些序列是非常好的目标识别或分类的特征向量。

1.3 语谱图特征序列提取

语谱图是一种纹理图像,它的个性特征主要表现在像素间的空间几何特征上。同一说话人的同一句话在不同的情感下有不同语谱图,分别对应不同的若干特征时间序列,这些时间序列之间不但存在差异而且具有某种规律性。

利用语谱图特征时间序列实现说话人不同情感的辨别,将它作为识别或分类的特征向量,输入到模式识别单元,进行识别。在本文中,我们使用图像中点火的神经元的总数(PCNN-ICON)^[7] $g(n)$ 以及熵序列^{[1]82} $H_n(p)$ 作为特征时间序列:

$$g(n) = \sum_{ij} Y_{ij}(n) \quad (11)$$

$$H_n(p) = -P_{n1} \times \lg P_{n1} - P_{n0} \times \lg P_{n0} \quad (12)$$

其中 $Y_{ij}[n]$ 是在 n 时刻点火神经元 N_{ij} 的输出。 $Y_{ij}[n]$ 不仅包含了图像的灰度分布信息,更重要的是它还包含了图像中相邻像素之间的相对位置信息,即图像的空间几何信息,这恰恰是纹理图像的个性特征所在。点火时刻信号 $g(n)$ 统计了 n 时刻PCNN发出脉冲的神经元总数,即每次迭代过程中整幅图像中点火的神经元的总数。

$H_n(p)$ 为PCNN第 n 次点火时输出二值图像 $Y_{ij}[n]$ 的熵。熵值是图像统计特性的一种表现形式,反映了图像包含信息量的大小。具体地也就是PCNN在每次循环迭代运算时,计算其分割输出的二值图像 $Y[n]$ 的熵值 $H_n(p)$ 。其中 P_1, P_0 分别表示 $Y_{ij}[n]$ 中像素值为1和0的概率。

PCNN的输出 $Y_{ij}[n]$ 是脉冲序列,其输出的二值图像的熵值和神经元点火序列就是随迭代次数变化而变化的。不同的图像经过一定的迭代之后,其激活的神经元的个数和分布将会不同,从而使PCNN的输出 $Y_{ij}[n]$ 出现变化,其对应相同迭代次数时,某迭代时间段内 $H(p)$ 和 $g(n)$ 的分布曲线不同。

1.4 情感辨别

由于特征序列 $g(n)$ 的数值非常大,不利于判断,所以首先利用公式(13)对特征图像 $g(n)$ 进行对数处理:

$$\hat{g}(n) = \lg g(n) \quad (13)$$

$\hat{g}(n)$ 是经过对数化了的点火时刻信号,既包含了纹理图像的个性特征,又使数量相差巨大的 $g(n)$ 值都缩减在容易进行比较的范围之内。PCNN特性保证了同样的图像内容具有的时间、尺度等许多不变的特性,利用PCNN的图像分割能力,将

输出对数化点火时间序列 $\hat{g}(n)$ 序列作为特征向量, 就可以根据不同图像之间特征向量的欧式距离 MSE 的数值大小进行辨别, 根据对欧式距离 MSE 大小设定的门限阈值 T 判断不同语句之间是否存在着情感差别, 欧式距离定义^{[1]83} 如式(14):

$$MSE = [\sum (G_0(n) - G_1(n))^2]^{1/2} \quad (14)$$

其中 $G_0(n)$ 、 $G_1(n)$ 分别表示待分类图像的特征向量对数化点火时间序列 $\hat{g}(n)$ 的值。门限阈值 T 根据式(15) 设定, 其中 δ 是经验参数, 参数的选择依赖于不同的实验样本, 取值为 1% 左右; F_n 表示 PCNN 模型的点火次数, 取值范围 30 ~ 50。

$$T = \delta F_n \quad (15)$$

$$\hat{H}_n(p) = \int H_n(p) = \sum_n H_n(p) \quad (16)$$

如果根据欧式距离 MSE 算出的结果大于阈值 T , 判断为二者之间存在情感差别。再利用 PCNN 模型输出熵序列作为特征向量, 对待分类图像进行情感判断。我们引入 $\hat{H}_n(p)$ 作为特征图像熵序列 $H_n(p)$ 的统计积分值来进行情感辨别。

2 实验仿真与结果分析

以往的情感语音研究^[8,9] 往往采用带有语义的语句作为语音辨别或者合成实验的数据, 但我们在实验中发现, 文字内容对说话人的情感具有暗示作用, 不利于通过实验客观反映算法效果, 因此我们采用不包含客观情感的阿拉伯数字串(数字 0~9 的不规则组合)的汉语发音作为本实验的数据, 很大程度上排除了文字内容对说话人的影响, 客观的反映了情感之间的联系。语音录制在安静的实验室中进行, 头戴式话筒录制, 采样频率 22 kHz, 16 位量化, 位速 352 kbps。

由于说话人对情感的理解不同, 表演的效果不完全一致, 所以在采集到录音语料后, 再由其他几名受试者进行听辨, 对判别不一致的语句进行滤除, 最后共采用有效语句 120 个。

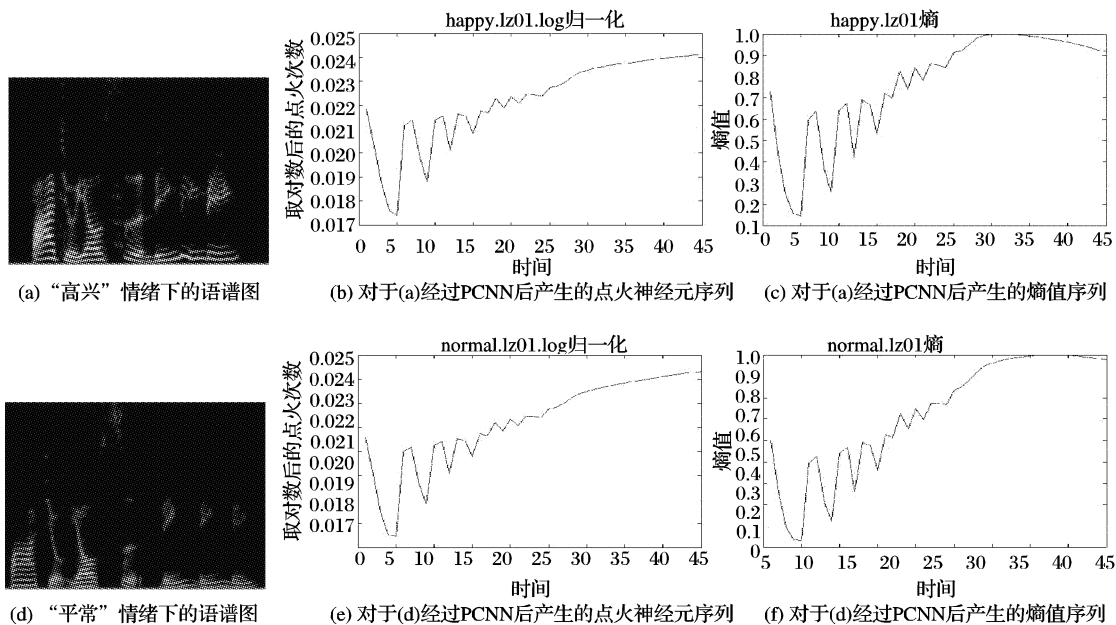


图3 男性语者

其中, 图 3(a) 和图 4(a) 的横轴为时间, 纵轴为频率, 灰度代表强弱。从图 3、4 可以看出, 由于实验语句为相同的内容, 所以点火神经元序列以及熵值序列表现出了类似的走势, 当在不同情感下发音时, 幅度以及序列位置方面存在着比较明显的差异, 只通过 PCNN 产生的图标也可以直观地区分开不同的情感。

将语音转化为分辨率为 830×465 的语谱图, 在 Matlab 7.0.4 平台下进行实验仿真, 算法见图 2。

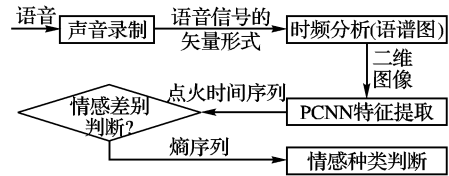


图2 算法框图

2.1 PCNN 模型的参数设置

为了保证 PCNN 输出时间序列的唯一性, 对语谱图像的处理必须在同一参数模型下进行。设定 PCNN 模型的点火次数 F_n 为 45, Y 、 L 、 U 初值设为零, S 为归一化以后的灰度值, 属于 $[0, 1]$, 连接域半径取 $r = 1.5$, 内部连接矩阵 $W = M$, 是 3×3 的方阵, 其中每一个元素数值为中心像素到周围每个像素的欧几里德距离的平方的倒数, 如式(17):

$$M_{xy} = [\sum (P_{ij} - P_{xy})^2]^{-1} \quad (17)$$

M_{xy} 为所求数值, P_{ij} 为中心像素点的坐标, P_{xy} 为周围像素点坐标。其余参数见表 1。

表1 PCNN 的参数及取值

参数	取值	参数	取值
α_L	0.1	V_L	0.2
α_E	0.5	V_E	20.0
α_F	0.1	β	0.3
V_F	0.5		

2.2 实验结果与比较分析

以汉语发音阿拉伯数字串“8914000”为例进行实验。图 3、4 分别是一名男性和一名女性说话人在“高兴”和“平常”两种情绪下发“8914000”的图像。

情感种类;如果得到的欧式距离不大于阈值,就停止判断。判断情况如表 2 所示。

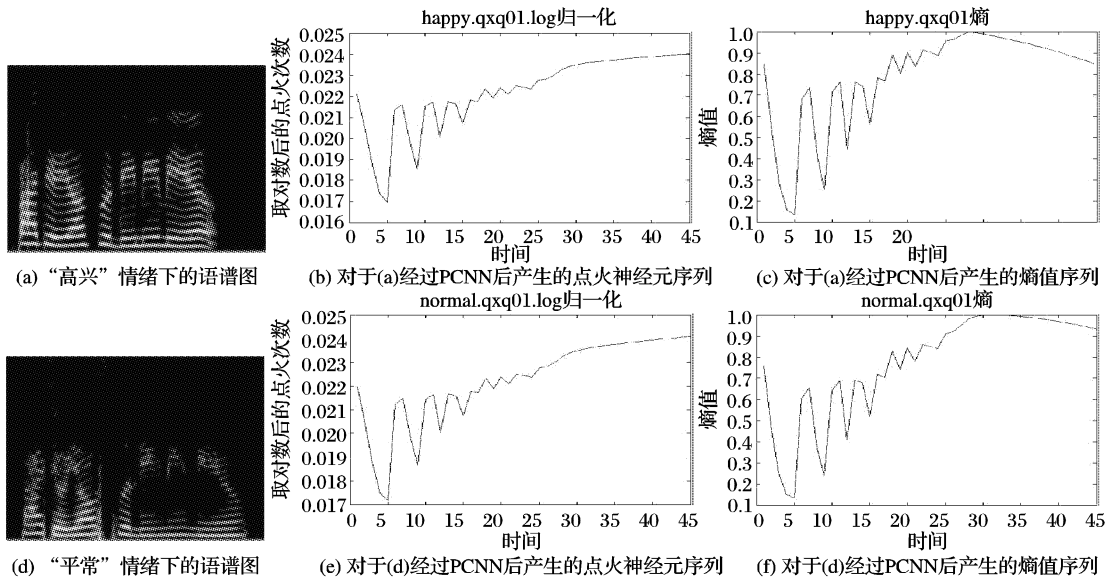


图 4 女性语者

表 2 库中 20 个说话人“8914000”的情感差别测定

语者	欧式距离	阈值判断	语者	欧式距离	阈值判断
A	0.4507	$> T$	K	0.4471	$< T$
B	1.4779	$> T$	L	1.9750	$> T$
C	1.2641	$> T$	M	0.5553	$> T$
D	0.9285	$> T$	N	2.1635	$> T$
E	0.9929	$> T$	O	2.0773	$> T$
F	0.7794	$> T$	P	0.5597	$> T$
G	1.0492	$> T$	Q	1.2650	$> T$
H	3.7494	$> T$	R	2.7405	$> T$
I	1.9202	$> T$	S	0.6589	$> T$
J	2.6510	$> T$	T	1.3798	$> T$

第二步判断,对大于阈值的语谱图熵值序列进行统计积分。我们通过大量实验发现,正常情况下熵值序列在“高兴”

时的统计积分值 $\hat{H}_n(p)^h$ 要大于“平常”时的统计积分值 $\hat{H}_n(p)^n$ 。利用统计积分值 $\hat{H}_n(p)^h$ 与 $\hat{H}_n(p)^n$ 之间数值的不同情况判断语句所属情感,见表 3。

表 3 中, $H > N$ 是指“高兴”时的熵序列统计积分值大于“平常”时的熵序列统计积分值; $H < N$ 是指“高兴”时的熵序列统计积分值小于“平常”时的熵序列统计积分值;在本文的表述中, \bigcirc 代表判断正确, \times 代表判断错误。语者 K 没有进行熵值比较,是因为第一步判断结束之后,算法根据其欧式距离 0.4471 小于阈值 0.45 (表 2 中语者 K 的数据) 而判断两句话没有情感差别,就对语者 K 停止了继续判断。根据第一步判断的结果,其他语者的欧式距离均大于阈值 (表 2 中除去语者 K 以外的语者),因此都进行了第二步判断。语者 D 与语者 P 的熵值比较出现了 $H < N$ 的判断结果,不符合实验设定的 $H > N$ 的标准,判断出现错误。

表 3 库中 20 个说话人“8914000”的情感类别判断

语者	$\hat{H}_n(p)^h$	$\hat{H}_n(p)^n$	熵值比较	情感判断	语者	$\hat{H}_n(p)^h$	$\hat{H}_n(p)^n$	熵值比较	情感判断
A	34.778	34.212	$H > N$	\bigcirc	K	—	—	—	—
B	34.674	32.689	$H > N$	\bigcirc	L	30.826	26.868	$H > N$	\bigcirc
C	31.708	29.345	$H > N$	\bigcirc	M	34.985	34.186	$H > N$	\bigcirc
D	27.273	29.031	$H < N$	\times	N	34.393	31.181	$H > N$	\bigcirc
E	35.377	34.785	$H > N$	\bigcirc	O	34.792	31.647	$H > N$	\bigcirc
F	31.426	30.001	$H > N$	\bigcirc	P	34.999	35.604	$H < N$	\times
G	35.065	33.825	$H > N$	\bigcirc	Q	29.710	27.453	$H > N$	\bigcirc
H	35.145	29.849	$H > N$	\bigcirc	R	32.725	27.708	$H > N$	\bigcirc
I	31.330	27.570	$H > N$	\bigcirc	S	33.396	32.298	$H > N$	\bigcirc
J	34.372	29.865	$H > N$	\bigcirc	T	35.048	34.991	$H > N$	\bigcirc

3 结语

本文算法对语音库中的同一说话人“平常”与“高兴”情感的正确识别率平均可达到 90% 以上。基于以上实验仿真结果,说明可以利用 PCNN 提取语音信号特征,完成说话人“高兴”与“平常”情感的识别,这对于 PCNN 的理论研究和实际应用开发具有非常重要的现实意义。另外,本文将 PCNN 引入到情感语音识别的应用研究中,开拓了信号处理中两个

极其重要的部分——语音信号处理和图像信号处理结合的新领域,实验结果说明该算法已经在这方面获得了不错的效果。

尽管本文的方法对区别两种不同的情感有比较好的性能,但在某些方面依然有一些工作要做:

1) 式(15)中的经验参数 δ 决定了欧式距离阈值 T 的大小,目前的设定是依靠实验经验来完成的。在后继的研究中我们需要寻找一种更加合理的方式来设定。

(下转第 718 页)

是垃圾短信的关键词库等特征未能得到更新而导致过滤性能降低的问题,将朴素贝叶斯分类、文本分类和支持向量机等方法引入垃圾短信过滤中,利用离线子系统的训练和学习分析结果反馈到在线子系统的快速过滤模块中,不但满足了短信的准确性和自适应性强的要求,而且根据短信的内容及时地更新在线过滤子系统的特征和内容,使得系统具有更好的自适应性和较高的智能化水平。

未来工作主要有:进一步的优化本文算法;考虑引入控制理论中的相关方法,以实现更好的自适应调节功能;将本文方法应用于下一代垃圾彩铃的过滤应用中。

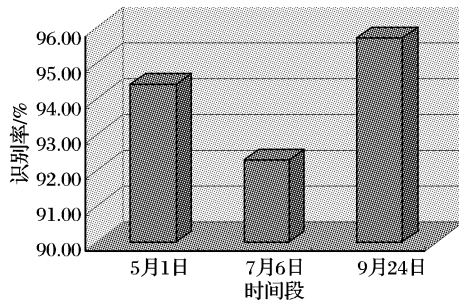


图4 过滤系统在不同时间的过滤效果比较

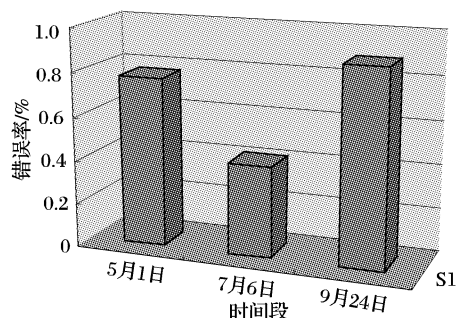


图5 过滤系统的错误率比较

参考文献:

- [1] 王丽萍. 我国垃圾短信的法律分析及防范策略[J]. 湖南税务高等专科学校学报, 2006, 19(3): 42-44.
- [2] 孙晋文, 肖建国. 基于SVM文本分类中的关键词学习研究[J].

计算机科学, 2006, 33(1): 182-184.

- [3] 易阳峰. 垃圾短信的监控与原理实现[J]. 中兴通讯技术, 2005, 11(6): 49-54.
- [4] Di NUNZIO G M. A bidimensional view of documents for text categorisation[C]// Proceedings of the 26th European Conference on Information Retrieval (ECIR 2004), Lecture Notes in Computer Science (LNCS) 2997. Germany: Springer Heidelberg, 2004: 112-126.
- [5] FORMAN G. An extensive empirical study of feature selection metrics for text classification[J]. Journal of Machine Learning Research 3, 2003(3): 1289-1305.
- [6] EYHERAMENDY S, LEWIS D D, MADIGAN D. On the naive bayes model for text categorization[C]// Proceedings of the 9-th International Workshop Artificial Intelligence and Statistics. Florida, US: IEEE Press, 2003.
- [7] ANDROUTSOPOULOS I, KOUTSIAS J, CHANDRINOS K V, et al. An evaluation of naive bayesian anti-spam filtering[C/OL]. [2007-09-15]. http://www.ics.forth.gr/~potamias/mlnia/paper_2.pdf.
- [8] ANDROUTSOPOULOS I, PALIOURAS G, KARKALETSIS V, et al. Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach[C]// Proceedings of the workshop: Machine Learning and Textual Information Access. Washington, DC: IEEE Press, 2000: 1-13.
- [9] SEBASTIANI F. Machine learning in automated text categorization [J]. ACM Computing Surveys, 2002, 34(1): 1-47.
- [10] CHANG C C, LIN C J. LIBSVM-A library for support vector machines (version 2.33) [EB/OL]. [2007-09-02]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>.
- [11] ANDROUTSOPOULOS I, KOUTSIAS J, CHANDRINOS K V. An evaluation of naive bayesian anti-spam filtering[C]// Proceedings of 2000 Workshop on Machine Learning in the New Information Age. Barcelona: IEEE Press, 2005: 9-17.
- [12] SEBASTIANI F. Machine learning in automated text categorization [J]. ACM Computing Surveys, 2002, 34(1): 1-47.
- [13] 孙晋文, 肖建国. 基于SVM的中文文本分类反馈学习技术的研究[J]. 决策与控制, 2004, 19(8): 927-930.

(上接第 713 页)

2) 我们通过大量实验发现 $\hat{H}_n(p)^h$ (“高兴”时的熵序列统计积分值)与 $\hat{H}_n(p)^n$ (“平常”时的熵序列统计积分值)之间有非常好的规律性。在算法中我们利用了这一规律,并取得了良好的效果。如何更好的诠释 $\hat{H}_n(p)^h$ 与 $\hat{H}_n(p)^n$ 二者之间的规律,提高算法的正确识别率以及算法的鲁棒性是后继的工作之一。

3) 如何将识别算法应用到其他不同的情感当中,以进行更多情感的区别判断。

参考文献:

- [1] 马义德, 袁敏, 齐春亮, 等. 基于 PCNN 的语谱图特征提取在说话人识别中的应用[J]. 计算机工程与应用, 2005, 41(2): 81-84.
- [2] ECKHORN R, REITBOECK H J, ARNDT M, et al. A neural network for feature linking via synchronous activity: results from cat visual cortex and from simulations[M]// Cambridge: Cambridge University Press, 1989: 255-272.
- [3] 刘勃, 马义德, 钱志柏. 一种基于交叉熵的改进型 PCNN 图像自动分割新方法[J]. 中国图象图形学报, 2005, 10(5): 579-584.
- [4] DENG L, KHEIRALLAH I. Dynamic formant tracking of noisy

speech using temporal analysis on outputs from a nonlinear cochlear model[J]. IEEE Transactions On Biomedical Engineering, 1993, 40(5): 456-467.

- [5] ECKHORN R, REITBOECK H J, ARNDT M, et al. Feature linking via synchronization among distributed assemblies: simulation of results from cat cortex [J]. Neural Computing, 1990, 2(3): 293-307.
- [6] SKOURIKHINE A N. A pulse couple neural network for image smoothing and segmentation[C]// International Symposium on Computational Intelligence. Kosice: IEEE Press, 2000.
- [7] JOHNSON J L. Pulse-coupled neural networks[C]// Adaptive Computing: Mathematical and Physical Methods for Complex Environments, Proceedings of SPIE. 1996, 2824(11). Proceedings of Adaptive Computing: Mathematics, Electronics and Optics. Orlando FL.
- [8] 张颖, 罗森林. 情感建模与情感识别[J]. 计算机工程与应用, 2003, 33: 98-102.
- [9] 周洁, 赵力, 邹采荣. 情感语音合成的研究[J]. 电声技术, 2005, 10: 57-59.