

利用 PCNN 神经元点火位置信息的语音特征研究

张歆奕, 阮柏尧

ZHANG Xin-yi, RUAN Bai-yao

五邑大学 信息学院, 广东 江门 529020

School of Information, Wuyi University, Jiangmen, Guangdong 529020, China

E-mail: xzhang@wyu.cn

ZHANG Xin-yi, RUAN Bai-yao. Speech feature based on position of PCNN neurons fired and its application to speaker recognition. *Computer Engineering and Applications*, 2009, 45(34): 124-126.

Abstract: Spectrogram is a time-frequency representation of speech signal. This image contains rich information of the speech. The traditional speech feature extracted from spectrogram using Pulse Couple Neural Network (PCNN) is the time series of the numbers of neurons fired at each iteration. Usually the length of the series is 50 data points. This paper proposes a new approach to extract speech feature from spectrogram using PCNN, which is based on the distribution of the neurons fired. Experiments on speaker recognition show that the new speech feature yields much better results than the traditional one.

Key words: spectrogram; Pulse Couple Neural Network (PCNN); speaker recognition

摘 要: 语谱图是语音信号的时频表示, 含有丰富的信息。把语谱图输入到脉冲耦合神经网络(PCNN)可以获得语音的特征矢量。传统的语音特征采用 PCNN 50 次迭代的点火次数。提出了一种新的语音特征参数, 该参数基于 PCNN 神经元点火位置的信息。说话人识别的实验表明, 新语音特征比传统的特征更能反映话者语音信号的特点, 获得更好的识别结果。

关键词: 语谱图; 脉冲耦合神经网络; 说话人识别

DOI: 10.3778/j.issn.1002-8331.2009.34.038 文章编号: 1002-8331(2009)34-0124-03 文献标识码: A 中图分类号: TP183

1 引言

如何获得有效和鲁棒的语音特征是语音识别和说话人识别的关键问题。语音特征一般从时域分析或频域分析得到。语谱图是语音信号的时频域表示, 其横轴是帧数, 代表时间的次序; 纵轴是每帧语音信号的功率谱。由于人类听觉系统对信号的相位不敏感, 因此语谱图几乎包含了语音信号的全部信息, 而且反映了语音的整体信息。因此语谱图可应用于语音特征的提取^[1], 语音音素的分割和语音识别^[2-3]等。

脉冲耦合神经网络(PCNN)是一种较新的神经网络, 它模拟了猫的视觉神经元的活动特点, 成功地应用于图像处理特别是图像分割等领域^[4]。马义德等人把语谱图输入到 PCNN, 把 PCNN 每次迭代的点火次数作为语音的特征, 并进行了基于该特征的说话人识别实验^[5]。

2 传统的基于 PCNN 点火次数的语音特征

利用 PCNN 提取语音特征参数的传统方法是: (1) 把输入语音分帧和加汉明窗, 窗长 10~20 ms; (2) 用 FFT 计算信号的功率谱; (3) 由各帧的功率谱构成语谱图; (4) 并把语谱图输入到 PCNN, PCNN 的神经元数等于输入图像的像素个数; (5) 最后通过统计 PCNN 每次迭代的点火次数获得语音的特征参数。

$$F_{ij}[n] = I_{ij}$$

$$L_{ij}[n] = \sum w_{ijk} Y_{ijk}[n-1]$$

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

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n] > \theta_{ij}[n-1] \\ 0, & U_{ij}[n] \leq \theta_{ij}[n-1] \end{cases}$$

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

采用简化的 PCNN 模型, PCNN 的迭代方程如式(1)所示, 其中 $F = I_{ij}$ 为输入, 与像素 (i, j) 连接; L_{ij} 为相邻神经元输出的加权; U_{ij} 为内部活动项; θ_{ij} 为动态门限; Y_{ij} 为脉冲输出, β , α_{θ} 和 V_{θ} 为 PCNN 的三个参数。语谱图输入到 PCNN 后, 每迭代一次都统计点火神经元的个数, 经过 50 次迭代, 得到 50 点的时间序列。该时间序列与原始的语谱图相比, 数据量大大减少, 一定程度上反映了输入图像的空间分布信息, 因此可作为语音的特征参数, 用于语音识别或说话人识别等用途^[6]。

图 1 为两个语谱图及其对应的 PCNN 点火次数时间曲线, 其中语谱图分别为两个不同话者的同一发音。从图 1 可见, 图 1(b) 和图 1(d) 是不同的, 因此 PCNN 点火次数时间曲线包含有话者的个性信息, 可输入到后面的分类器进行说话人识别。

3 新的基于 PCNN 神经元点火位置的语音特征

上面的点火次数曲线统计了每次迭代点火神经元的个数,

作者简介: 张歆奕(1965-), 男, 副教授, 博士, 主要研究方向为信息与信号处理和电子系统 EDA; 阮柏尧(1982-), 男, 在读硕士, 主要研究方向为信息处理理论与技术。

收稿日期: 2009-06-15 修回日期: 2009-10-09

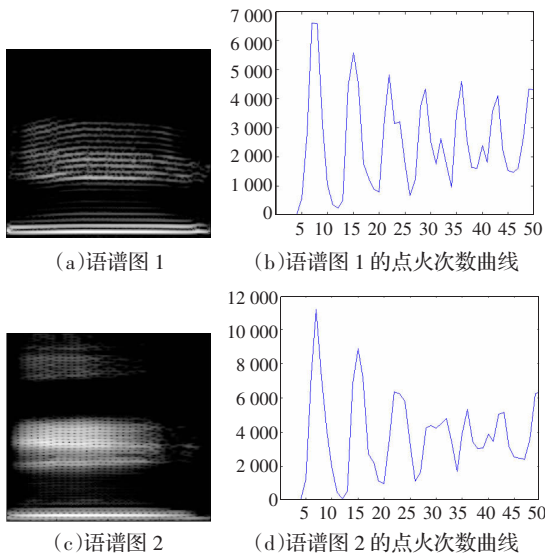


图1 语谱图及其 PCNN 点火次数曲线

虽然有数据量少和鲁棒性较好的优点^[4],但点火次数不能充分反映对应语谱图的分布特点,而神经元的点火位置及其分布应能更好地反映语谱图的分布特点,即能更好地反映语音的特点。

图2给出了语谱图及其每次迭代时的点火神经元的位置分布。图中 PCNN 的输入语谱图与图1(c)是一样的,图2(a)为第一次迭代的点火神经元的位置分布;图2(b)为第二次迭代的点火神经元的位置分布;如此类推,图2(d)为第六次迭代的点火神经元的位置分布。从图2可以看到,点火神经元的位置分布图反映了输入语谱图像素的灰度值及其位置的信息,而语谱图像素的灰度值及其位置反映了语音的时频特点。因此,点火神经元的位置分布比点火次数更能反映语谱图的分布特点,含有更丰富的话者信息。但每次迭代的点火神经元的位置分布的数据量比点火次数的数据量也大得多。采用以下方法减少数据量:(1)把每次迭代获得的点火神经元的位置分布图向频率轴和时间轴投影,如图3所示;(2)由频率轴和时间轴的投影构成一个特征矢量,如图4上半部分所示;(3)由特征矢量构成特征矩阵,如图4下半部分所示。

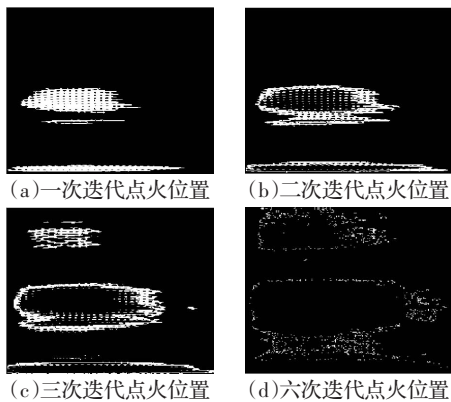


图2 语谱图及其每次迭代点火神经元位置分布图

4 说话人识别实验

4.1 实验准备

为了测试以上算法,在实验室录制了实验用语音库。该语音库包含有10人的发音,每个人录制20个发音,发音内容为

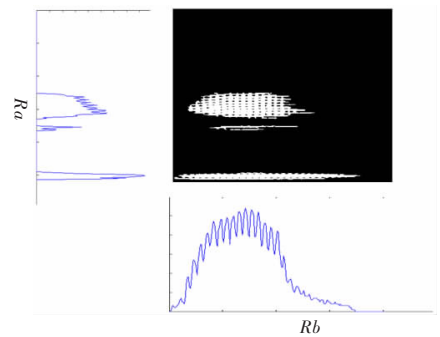


图3 把点火位置分布图投影到频率轴和时间轴

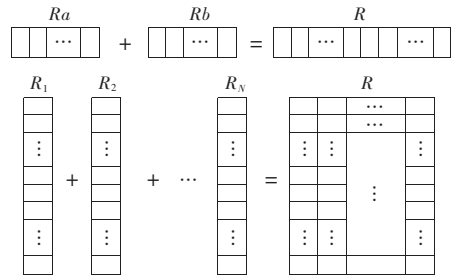


图4 频率轴和时间轴的投影构成一个特征矢量和特征矩阵

相同的普通话发音。采样频率16 000 Hz,窗长为256点,采用简化的 PCNN 模型。计算点火次数时间曲线时 PCNN 参数为 $\beta=0.8, \alpha_{\sigma}=0.1, V_{\sigma}=140$;计算新特征矩阵时 PCNN 的参数选定为 $\beta=0.8, \alpha_{\sigma}=0.3, V_{\sigma}=20$ 。

4.2 说话人确认实验

用目标话者的20个发音来训练目标话者的模型,计算这20个发音的特征矢量的均值矢量;用其他9个人的4个发音共36个发音作为闯入者的发音。特征矢量采用点火次数曲线。通过计算目标话者的发音特征矢量与均值矢量的距离和计算闯入者的发音特征矢量与均值矢量的距离,可以测试特征矢量的有效性,测试结果如图5所示。从图5可以看出,当阈值等于0.1801时,等错误率 $EER=0.1893$ 。若采用 BP 网络作为后级分类器,特征矢量仍采用点火次数曲线,识别性能提高很少^[5]。若采用新的语音特征,用前面的简单分类器和相同的条件进行实验,结果如图6所示^[5]。图6的结果比图5好。

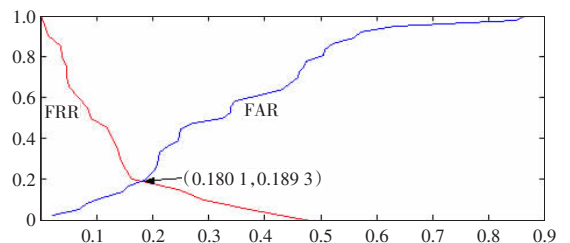


图5 采用旧特征的说话人确认结果

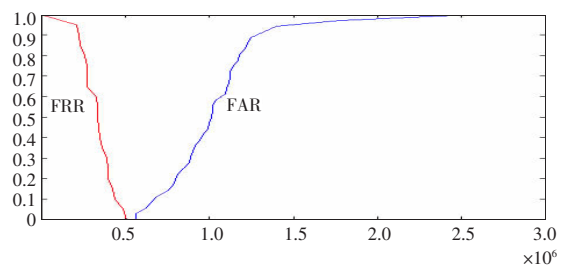


图6 采用新特征时的说话人确认结果

4.3 说话人辨认实验

在每人 20 段语音, 选 3 段有代表性的样本求平均作为模板, 求其他样本与这些模板的距离(选特征求模板的准则是类间离散度与类内离散度的比值最大), 再求剩下每人 17 段语音(共 170 段语音)的特征与各模板的距离。用最小距离判别法来辨认^[6]。采用点火次数曲线特征时, 识别率只有 51.76%, 采用新特征时, 识别率达到 100%^[5], 具体如表 1 所示。

表 1 两种特征用于说话人辨认结果对比

说话人	原特征		新特征	
	正确辨认个数	错误辨认个数	正确识别个数	错误识别个数
1	5	12	17	0
2	8	9	17	0
3	8	9	17	0
4	9	8	17	0
5	11	6	17	0
6	8	9	17	0
7	10	7	17	0
8	7	10	17	0
9	14	3	17	0
10	8	9	17	0
10人总共	88	82	170	0
识别率	51.76%		100%	

5 结论

通过分析 PCNN 各时刻的点火位置分布图的情况, 提出了

一种以 PCNN 各时刻点火位置图为基础的新特征的提取方法。与传统的 PCNN 提取语谱图的特征相比, 用新特征提取方法提取所得的新特征, 由于包含语谱图每一个点火时刻的位置, 更能反映语谱图的特点, 因而包含更完整的说话人语音信息。而且把新特征用于说话人确认和闭集的说话人辨认时, 实验效果比传统特征实验效果好, 所以, 新特征在说话人识别中更有应用价值。

参考文献:

- [1] Pinkowski B. Principal component analysis of speech spectrogram images[J]. Pattern Recognition, 1997, 30(5): 777-787.
- [2] Kingsbury B E D, Morgan N, Greenberg S. Robust speech recognition using the modulation spectrogram[J]. Speech Communication, 1998, 25: 117-132.
- [3] 张晓俊, 陶智. 改进脉冲耦合神经网络的语音识别研究[J]. 计算机工程与应用, 2007, 43(8): 51-53.
- [4] 马义德, 李廉. 脉冲耦合神经网络原理及其应用[M]. 北京: 科学出版社, 2006.
- [5] 阮柏尧. 脉冲耦合神经网络在基于语谱图的说话人识别中的应用[D]. 江门: 五邑大学, 2008.
- [6] 肖健华. 智能模式识别方法[M]. 广州: 华南理工大学出版社, 2006.
- [7] 顾晓东, 余道衡. PCNN 的原理及其应用[J]. 电路与系统学报, 2001, 6(3): 45-50.

(上接 49 页)

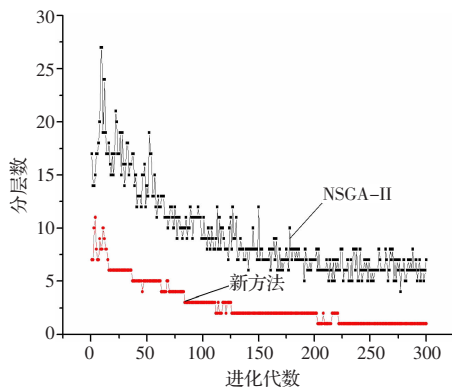


图 3 ZDT1 测试函数的分层数随进化代数的变化情况

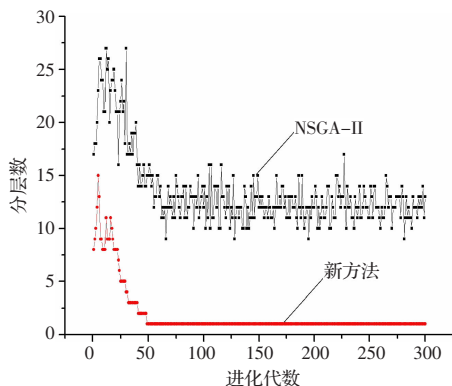


图 4 DTLZ1 测试函数的分层数随进化代数的变化情况

则此时的时间复杂度近似为 $O(M \log N)$, 远远好于 NSGA2-II 的 $O(N^2)$ 。综上所述, 与 NSGA-II 相比, 新方法在双目标优化的效率上有了很大的提高。

5 结论

多目标进化算法的一个重要研究方面是如何有效地对进化群体进行非支配分层, 以便于确定个体的适应值。在 Deb 提出的 NSGA-II 算法研究基础上, 针对算法双目标优化的运行效率, 采取了按需分层的策略, 提出了新的构造非支配前沿集的方法, 将其原来的时间复杂度由 $O(N^2)$ 降至 $O(kN + M \log N)$ 。最后以 6 个函数优化问题为例, 验证了新方法比 NSGA-II 具有更好的双目标优化效率。至于三个以上目标的优化问题, 比双目标优化更为复杂, 该文的方法可以借鉴, 但具体的解决方案尚待进一步地研究。

参考文献:

- [1] Knowles J D, Corne D W. Approximating the non-dominated front using the Pareto archived evolution strategy [J]. IEEE Transactions on Evolutionary Computation, 2000, 8(2): 149-172.
- [2] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach [J]. IEEE Transactions on Evolutionary Computation, 1999, 3(4): 257-271.
- [3] Deb K, Pratap A, Agrawal S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.
- [4] 郑金华, 蒋浩, 邝达, 等. 用擂台赛法则构造多目标 Pareto 最优解集的方法[J]. 软件学报, 2007, 18(6): 1287-1297.
- [5] Zitzler E, Deb K, Thiele L. Comparison of multiobjective evolutionary algorithm: Empirical results [J]. Evolutionary Computation, 2000, 8(2): 173-195.
- [6] Deb K, Thiele L, Laumanns M, et al. Scalable test problems for evolutionary multi-objective optimization [C]// Yao X. Proc of the 2002 Congress on Evolutionary Computation (CEC 2002). New Jersey: IEEE Service Center, 2002: 825-830.