

基于量子逻辑线路神经网络的说话人识别方法

潘平, 罗辉, 王洋

贵州大学计算机科学与技术学院, 贵阳 550025

摘要 由于说话人的语音信号具有时变性、随机性,其特征参数也呈现出高维及相邻帧变化较大等特点。从量子信息处理理论出发,将一帧语音信号视为一个量子态,在传统神经网络的基础上,利用量子逻辑线路构造神经网络,实现说话人语音信号的有效聚类,探索一种基于量子逻辑线路神经网络的说话人识别模型与方法。利用模型固有的大量全局吸引子,可有效降低语音信号处理的时间及复杂度。通过在经典计算机上模拟仿真,并与 BP 神经网络说话人识别模型进行对比,表明该方法能够加快说话人识别模型的收敛速率,对参数变化具有更好的鲁棒性,且其系统识别率比 BP 神经网络方法平均提高了 3.34%。

关键词 量子逻辑线路;神经网络;说话人识别

中图分类号 TN912.3

文献标志码 A

doi 10.3981/j.issn.1000-7857.2013.33.001

Speaker Recognition Method Based on Quantum Logic Circuit Neural Networks

PAN Ping, LUO Hui, WANG Yang

College of Computer Science & Information, Guizhou University, Guiyang 550025, China

Abstract Under the same spatial and temporal conditions, the quantum computing is superior to the traditional computing. Because a speaker's speech signal features the time-varying property and the randomness, its characteristic parameters also show high-dimensional characters and large changes in adjacent frames. This paper, based on the quantum information processing theory, takes a frame of the speech signal as a quantum state, and uses quantum logic gate circuits to construct the neural network according to the traditional neural network, and obtains an efficient clustering of the speaker's speech signal. A speaker recognition model is built and a method based on the quantum logic circuit neural network is proposed. This model has a large number of global attractors, and the method can use them to effectively reduce the complexity of the speech signal processing. Through simulations on a classical computer and a comparison with the BP neural network speaker recognition model, it is shown that this method not only can accelerate the convergence rate of the model but also has a better robustness with respect to the parameter changes. The system's recognition rate with the method proposed in this paper is 3.34% in average higher than that with the BP neural network method.

Keywords quantum logic circuits; neural network; speaker recognition

0 引言

说话人识别是根据语音波形中反映说话人生理和行为特征的语音参数,自动鉴别说话人身份的技术^[1]。说话人识别作为一种仿生物技术,已经广泛应用于自动控制、司法取证等许多领域。与一般的说话人识别模型相比,神经网络是一

种更接近于人认知过程的计算模型,它吸取了生物神经网络的部分优点,致力于按照生物神经网络系统类似的方式处理真实世界的各种信息^[2]。

20 世纪 90 年代以来的一些研究成果表明,人脑信息处理的过程可能与量子现象有关,大脑中可能存在量子力学效

收稿日期:2013-04-15;修回日期:2013-10-25

基金项目:贵州省科学技术基金项目(黔科合 J 字[2012]2132);贵阳市科技计划项目(筑科合同[2011101]1-2 号);贵州省国际科技合作计划基金资助项目([2009]700109,[2009]700125)

作者简介:潘平,副教授,研究方向为语音信号处理、信息安全等,电子信箱:panping_17@163.com;罗辉(共同第一作者),硕士研究生,研究方向为语音信号处理,电子信箱:luohui0216@163.com

应,量子系统具有和生物神经网络相似的动力学特征^[3]。在相同时空维数下,量子计算优于传统计算。将神经网络与量子理论结合起来,能更好地模拟人脑信息处理的过程,提高神经网络的逼近能力和信息处理效率。本文在传统神经网络的基础上,研究一种基于量子逻辑线路神经网络(QNN)的说话人识别模型及方法。

1 量子逻辑线路的神经网络模型

在量子计算中,通过对量子位状态进行一系列的酉变换,可以实现某些逻辑变换功能。当一组量子逻辑门构成的逻辑线路能够以任意精度逼近任意酉运算时,称这组量子逻辑门对量子计算通用。文献[4]证明,单比特量子逻辑门和受控非门可以实现 n 量子比特上的任意酉运算,也就是说任何量子逻辑线路都可以分解为单比特量子逻辑门和受控非门的乘积形式。

1.1 一位相移门

一位相移门^[5,6]定义为

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (1)$$

其中, θ 为相移角度。一位相移门线路如图 1 所示,作用是实现量子位的相位移动。



图 1 一位相移门

Fig. 1 One-bit phase-shift door

1.2 多位受控非门

多位受控非门 $C^n(U)$ 定义为^[7,8]

$$C^n(U)(|x_1 x_2 \dots x_n\rangle |\varphi\rangle) = |x_1 x_2 \dots x_n\rangle U^{x_1 x_2 \dots x_n} |\varphi\rangle \quad (2)$$

其中, U 为单量子比特非门; φ 为量子态。若前 n 个量子比特均为 1, 则非门 U 作用到最后一位量子比特使其翻转; 否则, 状态不变, 线路如图 2 所示。

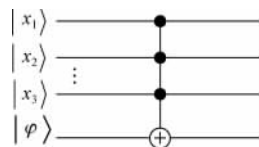


图 2 多位受控非门

Fig. 2 Multibit controlled "not" gate

1.3 量子逻辑线路神经网络模型及算法

量子逻辑线路神经元由一位相移门和多位受控非门组成^[9,10], 网络模型如图 3 所示, 其中 $|x_1\rangle, |x_2\rangle, \dots, |x_n\rangle$ 为网络输入; $|h_1\rangle, |h_2\rangle, \dots, |h_m\rangle$ 为隐层输出; $|y_1\rangle, |y_2\rangle, \dots, |y_m\rangle$ 为网络输出。

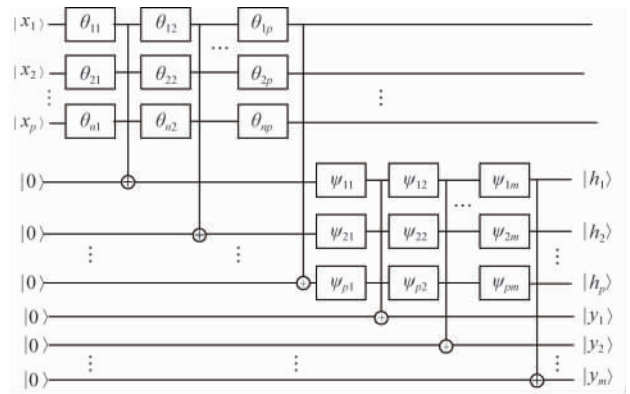


图 3 量子逻辑线路神经网络模型

Fig. 3 Quantum logic circuit neural network model

令输入量子态 $|x_i\rangle = \cos\theta_i |0\rangle + \sin\theta_i |1\rangle$, 以各层中量子位状态 $|1\rangle$ 的概率幅作为该层的实际输出, 则网络各层的实际输出为

$$h_j = \sin(\psi_j) = \prod_{i=1}^n \sin(\theta_i + \theta_{ij}) \quad (3)$$

$$y_k = \prod_{j=1}^p \sin(\psi_j + \psi_{jk}) =$$

$$\prod_{j=1}^p \sin\left(\arcsin\left(\prod_{i=1}^n \sin(\theta_i + \theta_{ij})\right) + \psi_{jk}\right) \quad (4)$$

其中, $i=1, 2, \dots, n; j=1, 2, \dots, p; k=1, 2, \dots, m$ 。

在本模型中, 以隐层和输出层一位相移门的旋转角度 θ_j 和 ψ_{jk} 为网络可调参数, 设归一化后的期望输出为 $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_m$, 则定义误差函数为

$$E = \frac{1}{2} \sum_{k=1}^m (\tilde{y}_k - y_k)^2 \quad (5)$$

根据梯度下降法, 各层旋转角度的梯度计算式为

$$-\frac{\partial E}{\partial \theta_j} = \frac{\sum_{k=1}^m (\tilde{y}_k - y_k) y_k \cot(\psi_j + \psi_{jk}) h_j \cot(\theta_j + \theta_{ij})}{\sqrt{1 - h_j^2}} \quad (6)$$

$$-\frac{\partial E}{\partial \psi_{j_1}} = \sum_{k=1}^m (\tilde{y}_k - y_k) y_k \cot(\psi_j + \psi_{jk}) \quad (7)$$

各层旋转角度的更新式为

$$\theta_j(t+1) = \theta_j(t) - \eta \frac{\partial E}{\partial \theta_j} \quad (8)$$

$$\psi_{jk}(t+1) = \psi_{jk}(t) - \eta \frac{\partial E}{\partial \psi_{jk}} \quad (9)$$

其中, t 为迭代步数; η 为学习速率。

2 基于 QNN 的说话人识别模型的构建与分析

基于量子逻辑线路神经网络的说话人识别模型, 由预处理、特征提取、特征参数量子化、模型训练和判决等组成, 如图 4 所示。

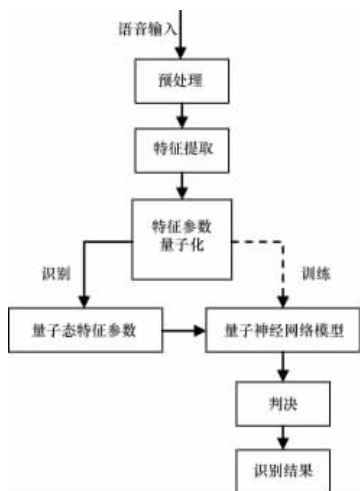


图 4 量子逻辑线路神经网络的说话人识别模型

Fig. 4 Quantum logic circuit neural network speaker recognition model

2.1 网络结构的选择

根据输出节点不同,神经网络可分为多输出型和单输出型 2 种。在多输出型方式中,由于网络要同时适应所有的说话人,势必需要更多的隐层节点,学习过程往往收敛较慢。因此,本文设计一种单输出型量子逻辑线路神经网络,网络的输入层节点对应每一帧语音样本特征,输出层节点只有 1 个。为每个说话人单独建立 1 个这样的网络模型,分别进行训练,将属于该说话人样本的期望输出设为 1。在识别阶段,将测试说话人样本输入到每个网络,让每个网络模型只判断待识别说话人是否属于该模型代表的类,根据网络输出进行判决。

2.2 QNN 说话人识别模型的训练

基于量子逻辑线路神经网络说话人识别模型的训练流程如下。

(1) 说话人特征参数量子化。对于归一化后的实值语音样本 $\bar{X}=(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$,量子态转换公式为

$$|x_i\rangle = \cos\left(\frac{2\pi}{1+e^{-x_i}}\right)|0\rangle + \sin\left(\frac{2\pi}{1+e^{-x_i}}\right)|1\rangle = \left[\cos\left(\frac{2\pi}{1+e^{-x_i}}\right) \sin\left(\frac{2\pi}{1+e^{-x_i}}\right) \right]^T \quad (10)$$

(2) 网络参数初始化。设置网络期望输出 $\tilde{y}=1$,最大迭代步数为 t_{max} ,迭代步数初始化为 $t=1$,误差精度为 ε ,学习速率为 η 。初始化隐层和输出层一位相移门的旋转角度为

$$\theta_j = 2\pi \cdot r \quad (11)$$

$$\psi_{jk} = 2\pi \cdot r \quad (12)$$

其中, r 为 $[0, 1]$ 之间的随机数。

(3) 按照式(3)、式(4)计算隐层的实际输出 $y^{(l)}$ 和输出层

的实际输出 $y^{(l)}$,其中

$$\theta_j = 2\pi / (1 + e^{-x_j}) \quad (13)$$

(4) 计算网络输出误差 $\varepsilon^{(l)}$,即

$$\varepsilon^{(l)} = \frac{1}{2} (\tilde{y} - y^{(l)})^2 \quad (14)$$

(5) 如果 $\varepsilon^{(l)} < \varepsilon$,转入流程(7),否则按照式(6)~式(9)对参数 θ_j 和 ψ_{jk} 进行更新,其中 $m=1$ 。

(6) 如果 $t < t_{max}$,转入流程(3);否则,转入流程(7)。

(7) 迭代终止,存储最新的旋转角度 θ_j, ψ_{jk} 和网络的实际输出 $y^{(l)}$ 。

2.3 QNN 说话人识别模型的识别过程

基于量子逻辑线路神经网络的说话人识别模型的识别流程如下。

(1) 将待测说话人特征样本经式(10)量子化后,制备相应的量子态 $\{|x_1^i\rangle, |x_2^i\rangle, \dots, |x_n^i\rangle\}$,其中 i 表示第 i 帧特征样本, n 表示特征维数。

(2) 将量子态特征参数依次输入到训练好的量子神经网络中,按式(14)计算输出 y_i 与存储 $y^{(l)}$ 的误差 ε_i ,若小于某一阈值,则对这一帧特征样本予以确认,否则拒绝。

(3) 计算确认帧所占的百分比,作为识别结果。

2.4 QNN 说话人识别模型的分析

由式(4)可以看出,两个迭代序列 $\{\theta_j(t)\}$ 和 $\{\psi_{jk}(t)\}$ 的全局最优解并不唯一,对于任意 $\alpha \in \mathbf{R}$,在 $[\alpha, \alpha + 2\pi]$ 内存在迭代序列 $\{\theta_j(t)\}$ 和 $\{\psi_{jk}(t)\}$ 的所有全局最优解。量子逻辑线路神经网络模型在全体实数域内均匀分布着众多全局吸引子,从而可显著提高收敛速率。因此该模型能很好地解决训练过程中由于语音样本的非线性和特征维数较高等带来的收敛速率慢等问题。

3 仿真实验

为了验证 QNN 说话人识别模型的有效性,仿真实验选择 10 个说话人作为实验对象,每人各记录 6 段语音,其中 5 段用于训练,1 段用于测试。语音的采样频率均为 8kHz,利用 MFCC 进行说话人特征提取,帧长 20ms,帧移 10ms,取前 12 阶分量作为网络的输入特征参数,识别误差取 0.01。

量子逻辑线路神经网络和 BP 神经网络结构均取 12-20-1 型,限定 $t=10000$, η 从 0.1,0.2, ..., 1.0 中取值。在相同学习速率下,对两种网络的性能进行 10 组对比仿真,其中 2 组仿真 ($\eta=0.5, \eta=0.8$) 对比如图 5 所示。可以看出,在学习速率相同的情况下,两种网络模型均能收敛,当 $\eta=0.5$ 时,QNN 模型的收敛步数为 200,BP 网络模型的收敛步数为 650;当 $\eta=0.8$ 时,QNN 模型的收敛步数为 100,普通 BP 网络模型的收敛步数为 350。10 组对比仿真的结果表明 QNN 模型不仅比普通 BP 网络有更快的收敛速率,而且当参数变化时具有较强的鲁棒性。

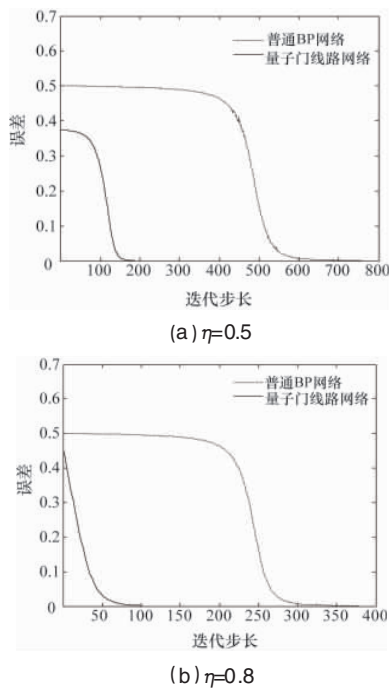


图5 两种网络模型的收敛对比

Fig. 5 Comparison of two network's convergence

利用某一说话人语音样本对网络进行训练,再选取部分说话人语音作为测试样本,QNN模型识别结果如图6所示。图6(a)为说话人1的QNN模型输出,图6(b)为测试说话人的识别误差曲线。仿真结果表明,QNN网络在训练过程中输出结果收敛于1,对于选取的3个不同说话人的10帧测试样本,均能进行正确的识别,且识别率较高。

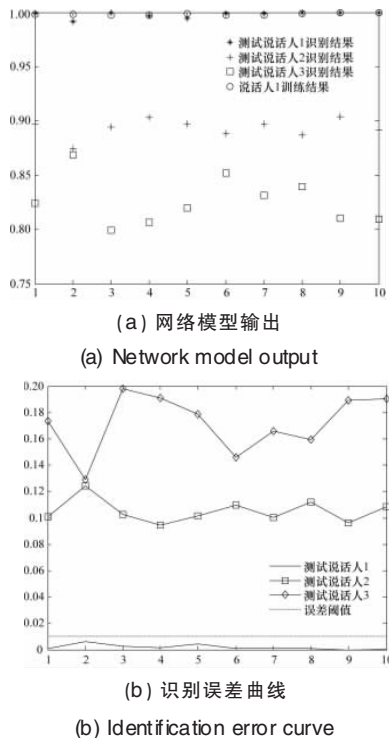


图6 量子逻辑线路神经网络的识别结果

Fig. 6 Recognition results of quantum logic circuit neural network

对于不同的隐层神经元个数,分别用QNN模型和BP网络模型对每个说话人进行100次仿真,2种模型的识别率如表1所示。从表1的识别结果可以看出,量子逻辑线路神经网络模型明显优于BP神经网络模型,同时,识别率与网络的隐层神经元个数密切相关,隐层神经元个数增多,可相应提高系统的识别率。

表1 识别率比较

Table 1 Comparison of recognition rate

隐层神经元数	识别率/%	
	BP模型	QNN模型
20	91.0	94.0
25	92.6	95.8
30	93.2	97.0
35	94.0	98.0
40	96.0	98.7

4 结论

将量子逻辑线路神经网络用于说话人识别领域,在训练过程中利用量子计算理论实现量子门旋转角度的调整,由于该模型具有周期性的多重吸引子,所以可以提高网络的收敛速率,从而降低语音信号处理的复杂度,提高识别率。在经典计算机上的仿真结果表明,量子逻辑线路神经网络在收敛速度和鲁棒性方面优于BP神经网络网络,在识别率方面,基于量子逻辑线路神经网络的说话人识别模型明显优于BP网络模型。通过理论分析和实验结果可以判断本方法能够用于说话人识别,这不仅丰富了现有的说话人识别模型,而且对说话人识别模型在未来的量子计算机上实现具有一定意义。

参考文献 (References)

- [1] 吴朝晖,杨莹春.说话人识别模型与方法[M].北京:清华大学出版社,2009. Wu Zhaohui, Yang Yingchun. Speaker recognition models and methods [M]. Beijing: Tsinghua University Press, 2009.
- [2] 赵力.语音信号处理[M].北京:机械工业出版社,2012. Zhao Li. Speech signal processing[M]. Beijing: Machinery Industry Press, 2012.
- [3] Stephen W, Kuffler J G, Nicholls A. From neuron to brain[M]. New Haven, Conn, USA: Sinauer Associates Inc, 2012.
- [4] 李士勇,李盼池.量子计算与量子优化算法[M].哈尔滨:哈尔滨工业大学出版社,2009. Li Shiyong, Li Panchi. Quantum computation and quantum optimization algorithms[M]. Harbin: Harbin Institute of Technology Press, 2009.
- [5] 李盼池,宋考平,杨二龙.基于受控旋转门的量子神经网络模型算法及应用[J].控制与决策,2011,26(6):898-901,906. Li Panchi, Song Kaoping, Yang Erlong. Control and Decision, 2011, 26(6): 898-901, 906.
- [6] 郭光灿.量子信息技术[J].重庆邮电大学学报:自然科学版,2010,22(5):521-525. Guo Guangcan. Journal of Chongqing University of Posts and Telecommunications: Natural Sciences version. 2010, 22(5): 521-525.
- [7] Kak S. Quantum neural computing [J]. Advances in Imaging and Electron Physics, 1995, 94: 259-313.
- [8] Nielsen M A, Chuang I L. Quantum computation and quantum information [M]. Higher Education Press, 2003.
- [9] 李盼池,宋考平,杨二龙.基于量子门线路的量子神经网络模型及算法[J].控制与决策,2012,27(1):143-146. Li Panchi, Song Kaoping, Yang Erlong. Control and Decision, 2012, 27(1): 143-146.
- [10] 侯旋.量子门线路神经网络及其改进学习算法研究[J/OL].计算机工程与应用,2012. [2013-03-20]. http://www.cnki.net/kcms/detail/11.2127.TP.20120820.1518.002.html. Hou Xuan. Computer Engineering and Applications, 2012. [2013-03-20]. http://www.cnki.net/kcms/detail/11.2127.TP.20120820.1518.002.html.

(责任编辑 韩星明)