

文章编号:1007-2780(2018)02-0165-09

基于量子粒子群优化广义回归神经网络 的语音转换方法

王 民,赵 渊*,刘 利,许 娟

(西安建筑科技大学 信息与控制工程学院,陕西 西安 710055)

摘要:针对粒子群算法优化神经网络进行语音转换时容易产生收敛速度慢、早熟的问题,本文采用一种新的量子粒子群算法优化广义回归神经网络的语音转换模型。该量子粒子群通过改变量子比特相位进而改变位置矢量,并利用量子非门进行变异操作。因此首先利用量子粒子群对网络进行优化得到最佳的光滑因子参数,从而建立频谱映射规则。接着,利用频谱参数和基频参数的相关性,对韵律特征基频也进行转换。然后,联立转换后的频谱参数和基频参数,利用 STRAIGHT 模型合成目标语音。最后,采用主观和客观测评方式进行评价。实验结果表明,与传统粒子群算法优化广义回归神经网络相比,该方法转换后的语音自然度和相似度得到提升,谱失真率下降 2.1%。本文方法具有比径向基神经网络、广义回归神经网络、粒子群算法优化广义回归神经网络更好的语音转换性能。

关 键 词:语音转换;量子粒子群;广义回归神经网络;量子比特;光滑因子。

中图分类号:TN912.3 **文献标识码:**A **doi:**10.3788/YJYXS20183302.0165

Voice conversion based on quantum particle swarm optimization of generalized regression neural network

WANG Min, ZHAO Yuan*, LIU Li, XU Juan

(School of Information and Control Engineering,
Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: In this paper, a new quantum particle swarm optimization algorithm is used to optimize the voice conversion model of generalized regression neural network in order to solve the problem of slow convergence and premature phenomenon in particle swarm optimization. The quantum particle swarm optimization algorithm changes the position vector by changing the quantum bit phase and uses the quantum non-gate to perform the mutation operation. Therefore, we first use the quantum particle swarm to optimize the network to get the best smooth factor parameters, so as to establish spectrum

收稿日期:2017-09-07; **修订日期:**2017-10-27.

基金项目:住房和城乡建设部科学技术项目计划(No.2016-R2-045);陕西省教育厅专项基金(No.2013JK1081);陕西省科学技术研究发展计划项目(No.CXY1122(2));陕西省自然科学基金青年基金(No.2013JQ8003)
Ministry of Housing and Urban-Rural Development Science and Technology Project plan(No.2016-R2-045);Shaanxi Provincial Department of Education special fund(No.2013JK1081);Shaanxi Province Science and Technology Research and Development Project(No.CXY1122(2));Shaanxi Provincial Natural Science Foundation Youth Fund(No.2013JQ8003)

* 通信联系人, E-mail:1098234739@qq.com

mapping rules. After that, we use the correlation between the spectral parameters and the fundamental frequency parameters to convert the prosodic characteristic fundamental frequency. Then, the STRAIGHT model is used to synthesize the target voice in conjunction with the converted spectral parameters and the fundamental frequency parameters. Finally, we use the subjective and objective evaluation methods to evaluate. The experimental results show that the natural and similarity of the proposed method for the transformed voice are improved and the spectral distortion rate is reduced by 2.1% compared with the traditional particle swarm optimization algorithm. The proposed method has better voice conversion performance than radial basis function neural network, generalized regression neural network and generalized regression neural network optimized by particle swarm optimization.

Key words: voice conversion; quantum particle swarm optimization; generalized regression neural network; quantum bite; smooth factor.

1 引言

语音转换 (Voice Conversion, VC) 是在保证语音含义不改变的前提下, 将一个人的语音个性特征转换为另一个人的语音个性特征, 使源说话人 (source speaker) 的声音听着更像目标说话人 (target speaker) 的声音^[1]。语音转换在诸多方面有着广泛的应用前景, 如智能翻译、多媒体配音、加密通讯等。

目前语音转换较为典型的方法有: 频谱弯折 FW (Frequency Warping)^[2]、隐马尔可夫模型 HMM (Hidden Markov Model)^[3]、高斯混合模型 (Gaussian Mixture Model, GMM)^[4] 以及人工神经网络 ANN (Artificial Neural Networks)^[5]。在人工神经网络中, 广义回归神经网络 (Generalized Regression Neural Network, GRNN)^[6] 具有较强的非线性映射能力和学习速度, 有且只有一个自由参数即光滑因子 σ , 具有较大计算优势, 但选择不当对语音转换模型的准确度会造成很大影响。相关文献提出采用粒子群 (Particle Swarm Optimization, PSO) 算法优化 GRNN 进行语音转换^[7], 取得一定效果, 但在实际应用中容易产生早熟现象^[8]。

针对以上问题, 本文提出采用一种新的量子粒子群算法 (Quantum Particle Swarm Optimization, QPSO)^[9] 优化 GRNN 形成语音转换模型。该量子粒子群利用量子比特对微粒位置进行编码, 通过改变相位进而改变位置矢量, 并利用量子非门进行变异操作。因此在量子粒子群中粒子能

以一定概率自由出现在搜索空间的任意位置, 有效克服“早熟收敛”现象。QPSO 算法全局寻优能力要远远强于标准的粒子群优化 (PSO) 算法, 因此不但能提高 GRNN 的泛化性能和鲁棒性, 同时提高网络学习能力及收敛速度。此外, 本文对韵律特征基频参数也进行了转换, 得到更多的个人特征信息, 因此转换后的语音与目标语音更加接近。

2 广义回归神经网络

广义回归神经网络 (Generalized Regression Neural Network, GRNN) 是径向基神经网络的一种, GRNN 具有比 RBF 更强大的非线性映射能力和学习速度, 建模的样本数量少, 预测效果很好。当训练样本确定时, 相应的网络结构和神经元之间的连接权值也确定, 网络训练目的只是确定其唯一参数光滑因子 σ 。

GRNN 结构如图 1 所示, 输入层、模式层、求和层与输出层构建成了前馈神经网络结构。

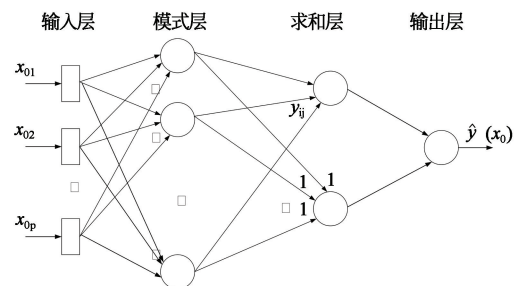


图 1 广义回归神经网络结构

Fig.1 Structure diagram of GRNN

输入层直接将输入信号传递给模式层。设输入向量 $\mathbf{X}$ 为 p 维 $\mathbf{X}=[x_1, x_2, \dots, x_p]^T$, 输出向量 $\mathbf{Y}$ 为 q 维, 样本数目为 n 。输入层中的神经元

数目等于学习样本中输入向量的维数 p , 模式层神经元数目等于学习样本数目 n , 每个神经元对应各自不同的样本, 神经元 i 的传递函数为

$$p_i = \exp\left[-\frac{(\mathbf{X}-\mathbf{X}_i)^T(\mathbf{X}-\mathbf{X}_i)}{2\sigma_i^2}\right], (i=1, 2, \dots, n), \quad (1)$$

其中: $\mathbf{X}$ 是网络输入变量; $\mathbf{X}_i$ 是第 i 个神经元对应的学习样本。

模式层又称隐回归层, 每个节点和不相同的学习样本对应。神经元数量等于训练样本数量 n , 激活函数为径向基函数 $R_i(x) = \exp\left(-\frac{\|\mathbf{X}-\mathbf{X}_i\|^2}{2\sigma_i^2}\right)$,

其中, $\mathbf{X}$ 为网络输入变量, $\mathbf{X}_i$ 为神经元 i 对应的学习样本, σ_i 为平滑参数。

求和层利用两种类型的神经元来求和。利用其中一类神经元对所有模式层神经元的输出值进行连接权值为 1 的算术求和。该类求和方式为:

$$S_D = \sum_{i=1}^n p_i = \sum_{i=1}^n \exp\left[-\frac{(\mathbf{X}-\mathbf{X}_i)^T(\mathbf{X}-\mathbf{X}_i)}{2\sigma_i^2}\right]. \quad (2)$$

另一种求和方式是将所有模式层神经元的输出值

进行加权求和, 表示为:

$$S_N = \sum_{i=1}^n y_i p_i = \sum_{i=1}^n Y_i \exp\left[-\frac{(\mathbf{X}-\mathbf{X}_i)^T(\mathbf{X}-\mathbf{X}_i)}{2\sigma_i^2}\right]. \quad (3)$$

S_N 利用模式层中输出样本 $\mathbf{Y}$ 中的各个元素 y_i 作为连接权值, 对模式层对应的神经元进行加权求和。

输出层中的神经元将求和层中所求的两种和的值相除, 得出估算结果

$$\bar{Y} = \frac{S_N}{S_D}. \quad (4)$$

3 量子粒子群优化广义回归神经网络

3.1 PSO 算法基本原理

经典 PSO 算法^[10]中设目标搜寻的空间维度为 M 维, 粒子种群规模为 N , 则第 i 个粒子的位置在 M 维空间中表示为 $\mathbf{X}_i=[x_{i1}, x_{i2}, \dots, x_{iM}]$, 速度为 $\mathbf{V}_i=(v_{i1}, v_{i2}, \dots, v_{iM})$, 第 i 个微粒“飞行”演变中的最佳位置为 $\mathbf{P}_i=[p_{i1}, p_{i2}, \dots, p_{iM}]$, 整个粒子种群中的最佳位置 $\mathbf{P}_g=(p_{g1}, p_{g2}, \dots, p_{gM})$ 。则每个微粒 i 在 M 维空间中速度和位置的变更方式分别是

$$\begin{aligned} v_{im}(t+1) &= \omega \cdot v_{im}(t) + c_1 \cdot r_1 (p_{im}(t) - x_{im}(t)) \\ &+ c_2 \cdot r_2 \cdot (p_{gm}(t) - x_{im}(t)), \end{aligned} \quad (5)$$

$$x_{im}(t+1) = x_{im}(t) + v_{im}(t+1), \quad (6)$$

其中: $i=1, 2, \dots, N$; $d=1, 2, \dots, M$; ω 是惯性因子; c_1, c_2 是加速系数; r_1, r_2 是 $[0, 1]$ 之间的随机数; p_{im} 为粒子个体最佳解, p_{gm} 为粒子全局最佳解。

3.2 QPSO 优化算法

基本 PSO 算法在实际应用中, 容易产生收敛速度较慢、早熟现象导致全局寻优能力较差^[11-12]。针对以上问题本文介绍一种新的量子粒子群优化算法(Quantum Particle Swarm Optimization, QPSO), QPSO 算法将量子运算融入 PSO 中^[13], 采用量子比特对微粒目前的位置进行编码, 利用量子旋转门通过改变微粒的量子比特相位代替 PSO 中的速度 $\mathbf{V}_i$ 矢量进行搜寻, 利用量子非门对微粒进行变异操作避免了局部收敛导致的早熟现象。

量子比特是量子信息的基本单位, 量子比特又称为量子位, 一个量子位的状态可表征为 $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$, α 和 β 是量子位的概率幅, 并且满足条件 $|\alpha|^2 + |\beta|^2 = 1$, 令 $\alpha = \cos(\theta)$, $\beta = \sin(\theta)$, 量子位也可以表示为 $[\cos(\theta)\sin(\theta)]^T$, θ 是量子

位中正余弦函数的相位。

如果把 M 维空间问题的最优解看成是 M 维空解中的点或者矢量,则求解最优解的过程可以表示 $\max f(x_1, x_2, \dots, x_M)$ 其中 $\alpha_i \leq X_i \leq \beta_i; i = 1, 2, \dots, M; M$ 为开始优化变量的数目; $[\alpha_i, \beta_i]$ 是变量 X_i 取值范围; f 为目标函数,其值可以作为粒子适应度。

QPSO 算法的实现主要分为初始种群设置,解空间转换,当前粒子状态的更新,变异操作 4 个步骤。

3.2.1 初始种群设置

在 QPSO 算法中,将量子比特的概率幅当作微粒当前位置矢量的编码。由于种群初始化时,编码存在一定的随机性,因此采用如下编码方式:

$$P_i = \left[\begin{array}{c} \cos(\varphi_{i1}) \\ \sin(\varphi_{i1}) \end{array} \middle| \begin{array}{c} \cos(\varphi_{i2}) \\ \sin(\varphi_{i2}) \end{array} \middle| \dots \middle| \begin{array}{c} \cos(\varphi_{iM}) \\ \sin(\varphi_{iM}) \end{array} \right], \quad (7)$$

其中: $\varphi_{ij} = 2\pi \times \text{rand}$; rand 是 $0 \sim 1$ 之间的随机数; $i = 1, 2, \dots, N; j = 1, 2, \dots, M; N$ 表示种群规模, M 为空间的维数。从式(7)可知,种群中所有微粒的位置均由对应搜索空间中的两个位置组成,两个位置对应的量子态 $|0\rangle$ 和 $|1\rangle$ 的概率幅为:

$$P_{ic} = (\cos(\varphi_{i1}), \cos(\varphi_{i2}), \dots, \cos(\varphi_{iM})), \quad (8)$$

$$P_{is} = (\sin(\varphi_{i1}), \sin(\varphi_{i2}), \dots, \sin(\varphi_{iM})), \quad (9)$$

其中 P_{ic} 为余弦位置, P_{is} 为正弦位置。

3.2.2 解空间转换

为了更加便捷地比较粒子在当前位置的优劣,因此将粒子两个位置由量子单位空间 $I = [-1, 1]^n$ 映射到优化问题的实际解空间上。量子比特的每一个概率幅都有解空间的一个优化矢量与之对应。若粒子 P_j 的第 i 个量子比特是 $[a_i^j, b_i^j]^T$, 则对应解空间的矢量是:

$$X_{ic}^j = \frac{1}{2} [\beta_i (1 + a_i^j) + \alpha_i (1 - a_i^j)], \quad (10)$$

$$X_{is}^j = \frac{1}{2} [\beta_i (1 + b_i^j) + \alpha_i (1 - b_i^j)]. \quad (11)$$

由公式(10)和(11)可知,每一个粒子都与优化问题的两个解相对应,量子态 $|0\rangle$ 的概率幅 a_i^j 对应 X_{ic}^j , 量子态 $|1\rangle$ 的概率幅 b_i^j 对应 X_{is}^j , 且 $i = 1, 2, \dots, N, j = 1, 2, \dots, M$ 。

3.2.3 粒子位置状态的更新

QPSO 算法利用量子旋转门转角的变更新代替基本 PSO 算法中速度的变化来实现位置的移动,故而将位置的移动转变为微粒的量子比特概率幅的变更,设粒子 P_i 目前搜寻到的自身最优位置为:

$$P_{ik} = (\cos(\varphi_{ik1}), \cos(\varphi_{ik2}), \dots, \cos(\varphi_{ikM})), \quad (12)$$

种群当前状态搜索到的全局最佳位置为:

$$P_g = (\cos(\varphi_{g1}), \cos(\varphi_{g2}), \dots, \cos(\varphi_{gM})). \quad (13)$$

如上所设,则粒子状态的更新可以通过以下方式实现:

(1)微粒 P_i 的量子比特概率幅函数相位增加量的变更方式为:

$$\Delta\varphi_{ij}(t+1) = \omega\Delta\varphi_{ij}(t) + c_1 r_1 (\Delta\varphi_k) + c_2 r_2 (\Delta\varphi_g), \quad (14)$$

公式(14)中:

$$\Delta\varphi_{ij} = \begin{cases} 2\pi + \varphi_{ikj} - \varphi_{ij} & (\varphi_{ikj} - \varphi_{ij} < -\pi) \\ \varphi_{ikj} - \varphi_{ij} & (-\pi \leq \varphi_{ikj} - \varphi_{ij} \leq \pi) \\ \varphi_{ikj} - \varphi_{ij} - 2\pi & (\varphi_{ikj} - \varphi_{ij} > \pi) \end{cases}$$

$$\Delta\varphi_g = \begin{cases} 2\pi + \varphi_{gi} - \varphi_{ij} & (\varphi_{gi} - \varphi_{ij} < -\pi) \\ \varphi_{gi} - \varphi_{ij} & (-\pi \leq \varphi_{gi} - \varphi_{ij} \leq \pi) \\ \varphi_{gi} - \varphi_{ij} - 2\pi & (\varphi_{gi} - \varphi_{ij} > \pi) \end{cases}$$

(2)量子比特概率幅利用量子旋转门更新为:

$$\begin{bmatrix} \cos(\varphi_{ij}(t+1)) \\ \sin(\varphi_{ij}(t+1)) \end{bmatrix} = \begin{bmatrix} \cos(\Delta\varphi_{ij}(t+1)) - \sin(\Delta\varphi_{ij}(t+1)) \\ \sin(\Delta\varphi_{ij}(t+1)) \cos(\Delta\varphi_{ij}(t+1)) \end{bmatrix} \cdot \begin{bmatrix} \cos(\varphi_{ij}(t)) \\ \sin(\varphi_{ij}(t)) \end{bmatrix} = \begin{bmatrix} \cos(\varphi_{ij}(t) + \Delta\varphi_{ij}(t+1)) \\ \sin(\varphi_{ij}(t) + \Delta\varphi_{ij}(t+1)) \end{bmatrix}, \quad (15)$$

其中 $i = 1, 2, \dots, N, j = 1, 2, \dots, M$ 。

则粒子 P_i 经过更新后,两个新的位置分

别为:

$$\tilde{P}_{ic} = ((\cos(\varphi_{i1}(t)) + \Delta\varphi_i), \dots,$$

$$\cos(\varphi_{iD}(t) + \Delta\varphi_{iD}(t+1))), \quad (16)$$

$$\begin{aligned} \tilde{P}_{is} = & ((\sin(\varphi_{i1}(t)) + \Delta\varphi_i), \dots, \\ & \sin(\varphi_{iD}(t) + \Delta\varphi_{iD}(t+1))). \end{aligned} \quad (17)$$

由公式(16)、(17)可知,量子旋转门可以通过更新量子比特相位,使正弦和余弦位置同时移动。因而在种群规模大小稳定时,使用量子比特编码能够极大的扩展对搜索空间的遍历性,提高获得全局最优解的概率。

3.2.4 变异操作

在 QPSO 中引进变异算子,依据变异概率对粒子进行量子非门变异,使粒子所表征的两个位置都得到变异,扩展了空间遍历性,增加了种群多样性,从而避免早熟现象。设定变异概率为 p ,将种群中每一个粒子都赋予一个 $(0, 1)$ 的随机数 rand_i ,如果该粒子 $\text{rand}_i < p$,则随机性地选择此粒子中 $[n/2]$ 个量子比特,利用量子非门对其实施变异操作,表述为:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(\varphi_{ij}) \\ \sin(\varphi_{ij}) \end{bmatrix} = \begin{bmatrix} \sin(\varphi_{ij}) \\ \cos(\varphi_{ij}) \end{bmatrix}, \quad (18)$$

其中: $i=1, 2, \dots, N, j=1, 2, \dots, M$ 。

3.3 QPSO 优化 GRNN

GRNN 只要学习样本确立,则相应的网络结构和神经元之间的连接权值也随之确定。网络训练过程是确定其唯一参数光滑因子 σ 的最优解。本文将 GRNN 的光滑因子参数 σ 看作 QPSO 的粒子。

QPSO 训练 GRNN 算法的实现步骤如下:

Step1: 初始化粒子群,设定种群大小、空间维度、加速系数、惯性因子、迭代次数等算法参数,按公式(7)生成粒子初始位置编码,组成初始种群。

Step2: 将种群中每个微粒的位置映射为 GRNN 中的光滑因子,建立 GRNN 网络。

Step3: 依据公式(10)、(11)进行解空间转换,将训练样本输出和实际值之间的均方差作为适应度函数,计算每个微粒的适应度,得到初始的个体历史最佳位置和全局最佳位置。

Step4: 利用公式(14)、(15)对所有粒子进行更新,重新评估粒子适应度。若粒子位置更新后适应度优于个体历史最佳位置时的适应度,则用此位置代替历史最优位置。若变更后微粒的个体的历史最佳位置优于当前变更后的全局最佳位置,则将此个体最佳位置作为全局最佳位置进行更换。

Step5: 依据变异概率对全局粒子进行如公式(18)的变异操作,变异后计算个体最优位置和全局最优位置,如果粒子适应度相比原值更优,则进行替换。

Step6: 判断是否达到收敛条件,若达到则输出最优光滑因子解,确定 GRNN 的最佳参数,转向 Step7,否则转向 Step4。

Step7: 得到最佳光滑因子解,算法结束。

4 基于 QPSO-GRNN 的语音转换方案

STRAIGHT^[14]模型是一种高精度的语音分析合成模型,运用 STRAIGHT 模型对源语音进行分析得到其频谱参数和基频信息。由于线性谱频率(Line Spectral Frequency, LSF)^[15]可以很好地表征声道和共振峰特性,所以将频谱包络转变为 16 阶 LSF 作为频谱特征参数进行转换。同时因为基音频率 F_0 是声音的韵律特征,考虑它与频谱特征有着极大的相关性^[16],所以本文通过 QPSO-GRNN 对基频也进行了转换,其转换原理如图 2 所示。

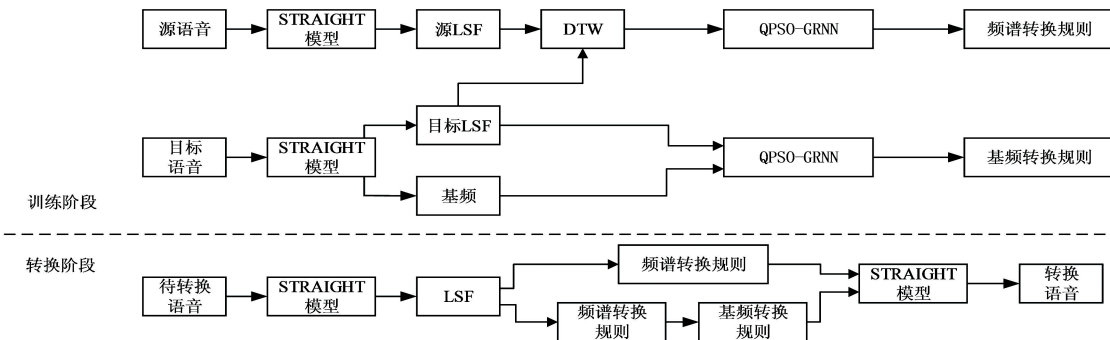


图 2 语音实现转换框图

Fig.2 Voice conversion realization diagram

训练阶段:

(1)运用 STRAIGHT 模型分别对源和目标说话人的训练语音进行分析,提取出频谱特征参数和基频参数,然后将频谱参数转换为 LSF 参数作为语音转换的特征参数。

(2)采用动态时间规整(Dynamic Time Warping,DTW)算法将源说话人和目标说话人语音的 LSF 参数对齐,而后使用 QPSO-GRNN 模型实施训练,得到频谱特征参数的映射规则。

(3)分别将目标语音的 LSF 参数和基频参数作为 QPSO-GRNN 模型的输入和输出进行训练,建立频谱特征参数和基频的映射规则。

转换阶段:

(1)用 STRAIGHT 模型对需转换的测试语音进行分析,得到频谱参数,然后将频谱参数转换为 LSF 参数,然后用已经训练好的频谱转换规则对其转换,再将转换后的 LSF 参数逆转换成频谱参数。

(2)利用基音频率映射规则将待转换声音的 LSF 参数转换为基频参数。

(3)将转换后得到的频谱参数和基频参数通过 STRAIGHT 模型进行语音合成,得到基于 QPSO-GRNN 模型转换后的语音。

5 实验

5.1 实验设计

实验语音来自卡内基梅隆大学 CMU (Carnegie Mellon University) 的 Arctic 声音库,从声音库中挑选文本内容相同的两男两女的语音各 100 段,进行同性和异性间的语音转换,其中 80 段语音作为训练句集,20 段语音作为测试句集,语音采样频率为 16 kHz。

本文设置算法参数为:粒子数目 $N=30$;空间维数 $M=24$;惯性因子 $\omega=0.5$;个体学习因子 $c_1=2$;全局学习因子 $c_2=2$;变异概率 $p=0.04$;迭代次数为 500 次。

5.2 主观评价

主观评价采用了 MOS 和 ABX 两种方法。

MOS 主要是对语音音质的评价,本文邀请 20 个专业语音测评人对转换的语音音质在 1~5 分之间评分。

1 分~5 分对应的 5 个级别为:差、比较差、一般、好和很好。

表 1 MOS 测评结果

Tab.1 Result of MOS test

转换类型	F1-M1	M1-F1	M1-M2	F1-F2
RBF	2.5	2.6	2.7	2.9
GRNN	2.7	2.8	3.0	3.1
PSO-GRNN	3.3	3.2	3.5	3.5
QPSO-GRNN	3.6	3.7	3.9	3.8

从表 1 可以看出,相同性别之间由于音色差异比较小,所以语音转换后的自然度比异性间的更好。相比其他转换方法,本文采用的 QPSO-GRNN 模型的转换效果明显优于 RBF 神经网络和 GRNN 模型以及 PSO-GRNN 模型。

ABX 是另一种主观测评方式,主要是对转换语音与目标说话者声音相似程度的评价。其分值代表转换语音和原目标语音的相似水平,分数值为 0~1 之间。本实验中邀请 20 位专业测评人对转换得到的 10 句语音进行测评。测试结果如表 2 所示。

表 2 ABX 测评结果

Tab.2 Result of ABX test (%)

转换类型	F1-M1	M1-F1	M1-M2	F1-F2
RBF	32.1	41.8	33.4	37.8
GRNN	33.4	43.1	35.2	41.5
PSO-GRNN	44.6	54.6	46.4	51.2
QPSO-GRNN	49.6	59.2	51.5	53.3

从表 2 中可知,本文提出的基于 QPSO-GRNN 模型的转换效果均优于其他 3 种算法,而且在男变女声的转变中效果更好。另外,从表格数据可知相同性间语音转换的相似度比女声变男声的相似度更高。

5.3 客观评价

将转换后的语音特征参数利用 STRAIGHT 模型合成转换语音,并同源语音、目标语音进行对比。其波形幅度归一化仿真如图 3 所示。

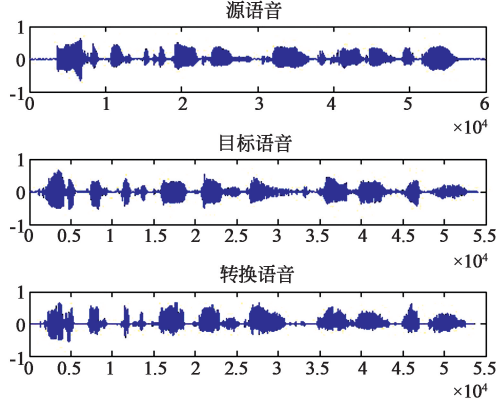


图 3 语音转换效果对比图

Fig.3 Comparison of voice conversion effect

仅依据语音波形难以精确对比转换的性能优劣,因此本文采用语音的频谱失真度作为客观评价指标:

$$SD = \frac{1}{n} \sum_{i=1}^n \frac{\|y_i - y_i'\|}{\|y_i - x_i\|}, \quad (19)$$

式中: n 为语音的帧数, x_i , y_i 和 y_i' 分别为源 LSF 矢量、目标 LSF 矢量和转换后的 LSF 矢量。图 4 和图 5 分别为男女之间语音互相转换的频谱失真图。

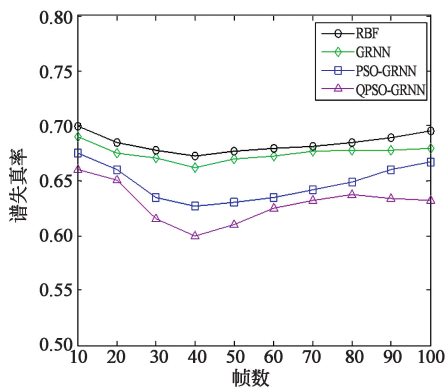


图 4 语音的频谱失真图(女声—男声)

Fig.4 Spectral distortion (F-M)

通过实验数据计算得,实现语音转换时 QPSO-GRNN 模型的平均频谱失真率比 PSO-GRNN 模型下降了 2.1%。通过图 4 和图 5 可知,采用 GRNN 模型的转换方式比采用 RBF 神

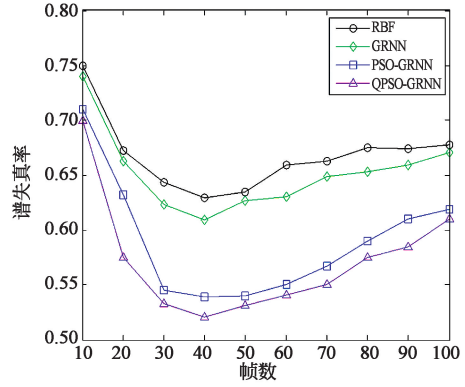


图 5 语音的频谱失真图(男声—女声)

Fig.5 Spectral distortion (M-F)

经网络的转换方式产生的频谱失真略小,因而 GRNN 模型的转换性能略优于 RBF 神经网络。然而,与采用 RBF 神经网络和采用 GRNN 模型的转换方式相比,本文采用 STRAIGHT 算法和 QPSO-GRNN 模型相结合的语音转换方式在同时考虑基频的情况下能够有效的降低失真度,使转换后语音的音质显著提高。此外相比基于 PSO-GRNN 模型的转换方法,本文的方法性能更佳,转换后的语音失真更少,尤其是男声转换女声时效果更明显。

6 结 论

本文采用新的 QPSO 算法对 GRNN 进行优化从而建立语音转换模型,在 QPSO 中通过改变量子比特相位值代替 PSO 中的速度 V_i 向量来搜寻解空间中的最优解,同时运用量子非门进行变异操作,能够避免早熟收敛。因此将 QPSO 算法训练 GRNN 的最优解作为光滑因子参数 σ ,从而找到最佳的频谱映射规则。同时利用韵律特征基频和频谱特征的相关性对基频也进行了转换,最后利用 STRAIGHT 模型对转换后的频谱参数和基频参数进行语音合成。实验结果表明与 PSO-GRNN 相比本文提出的基于 QPSO-GRNN 的语音转换方法性能更好,转换后目标语音的自然度和相似度提升,平均失真率下降 2.1%,尤其是男声转换为女声时,效果更加明显。另外,与 RBF 神经网络和 GRNN 对比,该方法具有更大的性能优势。

参 考 文 献:

- [1] GHORBANDOOST M, SAYADIYAN A, AHANGAR M, *et al.* Voice conversion based on feature combination with limited training data [J]. *Speech Communication*, 2015, 67: 113-128.
- [2] GODOYE, ROSEC O, CHONAVEL T. Voice conversion using dynamic frequency warping with amplitude scaling, for parallel or nonparallel corpora [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(4): 1313-1323.
- [3] LIL F, ZHAO Y, JIANG D M, *et al.* Hybrid deep neural network-hidden markov model (DNN-HMM) based speech emotion recognition [C]//*Proceedings of 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. Geneva, Switzerland: IEEE, 2013: 312-317.
- [4] NAKAMURA K, TODA T, SARUWATARI H, *et al.* Speaking-aid systems using GMM-based voice conversion for electrolaryngeal speech [J]. *Speech Communication*, 2012, 54(1): 134-146.
- [5] BHARTI K S, KOOLAGUDI S G, RAO K S, *et al.* Voice conversion using linear prediction coefficients and artificial neural network [C]//*Proceedings of the Cube International Information Technology Conference*. Pune, India: ACM, 2012: 240-245.
- [6] NIRMAL J, ZAVERI M, PATNAIK S, *et al.* Voice conversion using general regression neural network [J]. *Applied Soft Computing*, 2014, 24: 1-12.
- [7] 杨秀峰.基于神经网络的语音转换算法研究[D].西安:西安建筑科技大学,2017.
YANG X F. The research of voice conversion based on neural network [D]. Xi'an: Xi'an University of Architecture and Technology, 2017. (in Chinese)
- [8] 王洪涛,李丹.基于改进粒子群算法的图像边缘检测研究[J].液晶与显示,2014,29(5):800-804.
WANG H T, LI D. Image edge detection research based on improved particle swarm optimization algorithm [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(5): 800-804. (in Chinese).
- [9] 李士勇,李盼池.求解连续空间优化问题的量子粒子群算法[J].量子电子学报,2007,24(5):569-574.
LI S Y, LI P C. Quantum particle swarms algorithm for continuous space optimization [J]. *Chinese Journal of Quantum Electronics*, 2007, 24(5): 569-574. (in Chinese)
- [10] KENNEDYJ, EBERHART R. Particle swarm optimization [C]//*Proceedings of 1995 IEEE International Conference on Neural Networks*. Perth, WA, Australia: IEEE, 1995: 1942-1948.
- [11] 张玲华,姚绍芹,解伟超.基于自适应粒子群优化径向基函数神经网络的语音转换[J].数据采集与处理,2015,30(2):336-343.
ZHANG L H, YAO S Q, XIE W C. Voice conversion based on adaptive particle swarm optimization radial basis function neural network [J]. *Journal of Data Acquisition and Processing*, 2015, 30(2): 336-343. (in Chinese)
- [12] 张国梁,贾松敏,张祥银,等.采用自适应变异粒子群优化 SVM 的行为识别[J].光学精密工程,2017,25(6):1669-1678.
ZHANG G L, JIA S M, ZHANG X Y, *et al.* Action recognition based on adaptive mutation particle swarm optimization for SVM [J]. *Optics and Precision Engineering*, 2017, 25(6): 1669-1678. (in Chinese)
- [13] 张志成,林君,石要武,等.用加权子空间拟合和量子粒子群算法联合估计多普勒频率和波达方向[J].光学精密工程,2013,21(9):2445-2451.
ZHANG Z C, LIN J, SHI Y W, *et al.* Joint estimation of Dopplers and DOAs by WSF-QPSO method [J]. *Optics and Precision Engineering*, 2013, 21(9): 2445-2451. (in Chinese)
- [14] KOIKE H, NOSE T, SHINOZAKI T, *et al.* Improvement of quality of voice conversion based on spectral differential filter using STRAIGHT-based Mel-cepstral coefficients [J]. *The Journal of the Acoustical Society of America*, 2016, 140(4): 2963.
- [15] 解伟超.语音转换中声道谱参数和基频变换算法的研究[D].南京:南京邮电大学,2013.

- XIE W C. The research on vocal tract spectrum and pitch frequency transformation in voice conversion [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2013. (in Chinese)
- [16] SHAO X, MILNER B. Pitch prediction from MFCC vectors for speech reconstruction [C]//*Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Montreal, Canada: IEEE, 2004: 97-100.

作者简介:王民(1959—),男,江苏常州人,教授,硕士生导师,西安建筑科技大学通信与信息工程系主任,电子信息工程教研室主任,长期从事智能信息处理方面的研究工作。E-mail: wangmin1329@163.com

赵渊(1991—),男,陕西榆林人,硕士研究生,主要研究智能信息处理相关技术。E-mail: 1098234739@qq.com