

# 基于遗传算法数据降维的汉语数字语音识别<sup>①</sup>

高文曦<sup>1</sup>, 孙小琪<sup>2</sup>, 镇丽华<sup>3</sup>

<sup>1</sup>(南通大学 商学院, 南通 226000)

<sup>2</sup>(南通大学 教育科学学院, 南通 226000)

<sup>3</sup>(南通大学 管理学院, 南通 226000)

**摘要:** 针对语音信号特征参数 LPCC 和 MFCC 相结合后数据维数过高, 导致识别器性能下降的问题, 提出采用遗传算法对初始特征参数进行降维, 来提高识别性能. 首先提取语音信号的 LPCC 和 MFCC, 然后采用遗传算法对其进行特征降维, 最后将得到的低维数据送入支持向量机进行识别. 仿真实验结果表明, 采用遗传算法进行特征降维与传统的 PCA 降维相比, 识别率提高了 12.2%, 和初始特征相比识别率降低了 1.23%, 但是识别时间提高了 4.5 倍.

**关键词:** 汉语数字语音识别; 遗传算法; PCA; SVM

## Feature Dimension Reduction Based on Genetic Algorithm for Mandarin Digit Recognition

GAO Wen-Xi<sup>1</sup>, SUN Xiao-Qi<sup>2</sup>, ZHEN Li-Hua<sup>3</sup>

<sup>1</sup>(Business College, Nantong University, Nantong 226000, China)

<sup>2</sup>(College of Education Science, Nantong University, Nantong 226000, China)

<sup>3</sup>(School of Management, Nantong University, Nantong 226000, China)

**Abstract:** The dimensions are higher after combining MFCC with LPCC, which affect the performance of recognition. To solve the problem, genetic algorithm parameters for initial feature dimension reduction, to improve recognition performance. Firstly, we extract MFCC and LPCC of the speech signal. Then we reduce the dimensionality based on genetic algorithm. Finally, recognition is presented based on SVM with the low dimensional data. Simulation results show that compared to traditional PCA, the genetic algorithm recognition can increase the recognition rate by 12.2%. Meanwhile, recognition rate was reduced by 1.23% compared with the initial feature, but the recognition time increased 4.5 times.

**Key words:** mandarin digit recognition; GA; PCA; SVM

汉语数字语音识别是识别“0”到“9”十个非特定人汉语数字语音<sup>[1]</sup>. 汉语数字语音识别是语音识别领域中一个具有广泛应用背景的分支, 信用卡号码、电话语音拨号、个人身份证号码、电子密码等都具有数字化特征, 因此, 数字语音识别是语音识别领域中一个具有广泛应用背景的分支<sup>[2]</sup>. 语音信号的特征提取是语音识别的重要前提和基础. 目前, 常用的语音信号特征有线性预测倒谱系数和 Mel 倒谱系数<sup>[3]</sup>. 但是语音信号的特征参数的数据量大, 会影响识别性能, 所以在分类之前, 先将数据降维后, 在输入识别器. 传统的数据降维算法, 常用的线性降维法有主成分分析

法、独立成分分析和线性判别分析等<sup>[4]</sup>. 主成分分析是最经典的线性降维法, 通过对原始变量的相关矩阵或协方差矩阵内部结构的研究, 将多个变量转换为少数几个综合变量, 从而达到数据降维的目的<sup>[5]</sup>. 但是语音识别中提取的高维特征数据大多是非线性的.

遗传算法由美国的 J.Holland 教授于 1975 年首先提出, 其主要特点是直接对结构对象进行操作, 不存在求导和函数连续性的限定, 具有内在的隐并行性和更好的全局寻优能力, 采用概率化的寻优方法, 能自动获取和指导优化的搜索空间, 自适应地调整搜索方向, 不需要确定的规则<sup>[6]</sup>. 遗传算法是模拟自然进化过

① 收稿时间:2015-04-23;收到修改稿时间:2015-06-15

程搜索最优解的方法,采用概率化的寻优方法,不需要确定的规则.与PCA等线性数据降维法相比,基于遗传算法的数据降维更能适用于语音识别过程中提取出的非线性特征数据.仿真实验结果表明:语音信号初始特征参数(LPCC和MFCC)为 $5000 \times 24$ 维,通过遗传算法降维后得到 $1500 \times 24$ 维的特征,识别率降低了1.2%,但是识别时间缩短了416.96s,速度提高为原来的4.2倍.

## 1 算法原理

### 1.1 遗传算法

遗传算法是一种借鉴生物界的进化规律演化而来的随机化搜索方法<sup>[7]</sup>.遗传算法是从代表问题可能潜在的解集的一个种群开始的,而一个种群则由经过基因编码的一定数目的个体组成.初代种群产生之后,按照适者生存和优胜劣汰的原理,逐代演化产生出越来越好的近似解,在每一代,根据问题域中个体的适应度大小选择个体,并借助于自然遗传学的遗传算子进行组合交叉和变异,产生出代表新的解集的种群.这个过程将导致种群向自然进化一样,后代种群比前代更加适应于环境,末代种群中的最优个体经过解码,可以作为问题近似最优解.遗传算法中的遗传操作是随机的,但又不是完全随机搜索.它能够有效利用历史信息来推测下一代的寻优点集,通过一代代不断进化最终收敛到最适应环境的个体.遗传算法设计五个基本要素:参数编码、初始群体设定、适应度函数设计、遗传算子设计和控制参数设计.

遗传算法的基本流程如下:

- 1) 确定编码策略,确定参数集 $X$ 并将解域转换为位串结构空间 $S$ ;
- 2) 确定适应性函数 $f(x)$ ;
- 3) 定义遗传策略和遗传算子,包括群体大小 $N$ ,选择、交叉、变异的方法,交叉概率 $P_c$ 和变异概率 $P_m$ 等控制参数;
- 4) 确定初始群 $P$ ;
- 5) 根据适应性函数 $f(x)$ 计算出当前代群体的适应值;
- 6) 根据适应值评价当前代群体是否满足停止条件(满足某一指标或达到最大迭代次数);满足则结束算法,否则转向7);
- 7) 按照遗传策略,对当前代群体进行选择、交叉和

变异,生成下一代群体并转向5)对新一代群体进行评价.

### 1.2 支持向量机

支持向量机(Support Vector Machine, SVM)是基于统计模式理论的模式分类器. SVM原理概括来讲就是利用内积函数定义的非线性变换将输入空间变换到一个高维空间,在这个空间中求最优分类面对输入样本进行分类<sup>[8]</sup>. SVM的结构类似于神经网络,具体来说SVM类似于一个三层的神经网络结构,有输入层、隐含层、输出层.假定训练样本数据 $(x_i, y_i), x_i \in R^n, y_i \in \{-1, +1\}$ 可以被一个超平面 $(w \cdot x) + b = 0$ 分开,如果距离超平面最近的向量与超平面之间的距离是最大的,则判定这个向量被最优超平面分开,即:

$$\begin{cases} (w \cdot x_i) + b \geq 1, & \text{if } y_i = 1 \\ (w \cdot x_i) + b \leq -1, & \text{if } y_i = -1 \end{cases} \quad (1)$$

式(1)中,位于 $(w \cdot x_i) + b = 1$ 和 $(w \cdot x_i) + b = -1$ 上的向量被称为支持向量机. SVM问题可用下式带约束条件的优化问题来描述:

$$\begin{cases} \min \phi(w) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \\ y_i (w \cdot x_i + b) \geq 1 + \zeta_i, & i = 1, 2, \dots, l \end{cases} \quad (2)$$

最优分类函数为:

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \right] \quad (3)$$

式(3)中, $K(x_i, x_j)$ 为核函数,常用的核函数有:多项式核函数、径向基函数、Sigmoid核函数.

## 2 算法实现与仿真实验

### 2.1 基于遗传算法数据降维的汉语数字识别算法设计

基于遗传算法特征数据降维的汉语数字语音识别算法流程图如图1所示,具体步骤如下:

- 1) 读入0-9是个汉语数字语音信号预加重、分帧和加窗等预处理,帧长取256,帧移取128,窗函数为汉明窗;
- 2) 预处理后的每帧语音信号为原始数据 $x(1), x(2), \dots, x(N)$ ,依据第二章中LPCC的计算步骤,求出每个数字的12阶LPCC;
- 3) 将每帧语音信号FFT变换到频域上,计算频谱幅度的平方,通过Mel滤波器组后取反离散余弦变换,得到12阶的MFCC;
- 4) 将提取的LPCC和MFCC相结合,得到 $5000 \times 24$ 维的初始特征(其中每个样本的特征维数为 $50 \times 24$ );

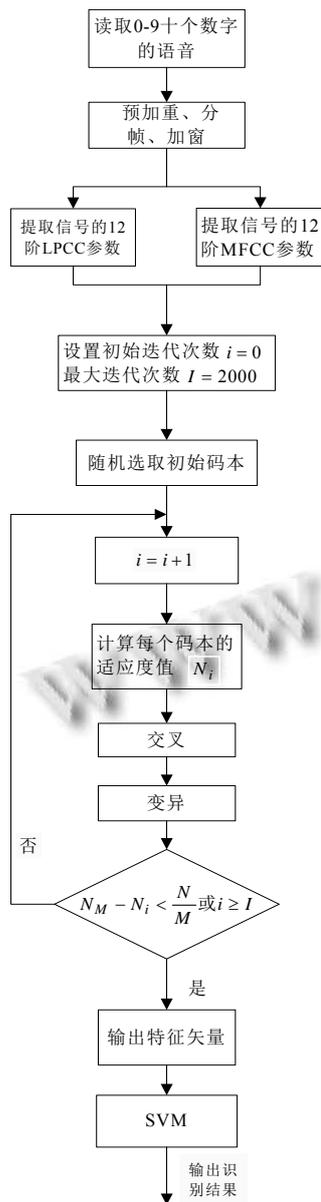


图 1 基于遗传算法特征数据降维的汉语数字语音识别算法流程图

5) 在所有向量中随机选出  $M$  个向量,  $Y = \{y_i | i=1,2,\dots,M\}$  作为初始码本(初始种群), 置迭代次数  $i = 0$ , 最大迭代次数为 2000 次;

6) 把所有向量按最小距离准则分配给  $Y$  的  $M$  个向量, 计算每个向量所包含的向量个数(适应度), 并把  $Y$  的  $M$  个码字按向量个数的多少进行排序, 得到  $y_1, y_2, \dots, y_M$ , 各码字对应的向量个数分别为  $N_1, N_2, \dots, N_M$ ;

7) 若  $N_M - N_1 < \frac{N}{M}$  或迭代次数达到最大时, 将  $Y = \{y_i | i=1,2,\dots,M\}$  作为最后的最优特征输出, 否则执行第 8)步;

8) 令  $i = i + 1$ , 对码本进行变异和交叉, 以概率  $P_m$  进行变异:

$$y_1 = y_M - \alpha(y_M - y_1) \tag{4}$$

以概率  $P_c$  进行交叉:

$$\begin{cases} y_j = y_j - \beta \left( y_{j+\frac{M}{2}} - y_j \right) \\ y_{j+\frac{M}{2}} = y_{j+\frac{M}{2}} + \gamma \left( y_{j+\frac{M}{2}} - y_j \right) \end{cases}, j=1, \dots, \frac{M}{2} \tag{5}$$

其中  $\alpha, \beta, \gamma$  为待定系数;

9) 返回第 7)步执行, 如果遗传算法结束, 则输出最终选取的特征矢量;

10) 送入 SVM 中进行分类, 输出为分类的结果和分类的正确率.

### 2.2 仿真实验与结果分析

仿真实验软件为 Matlab7.0, 实验中所用的语音库由 10 人录制, 发音为清晰标准的普通话, 实验室环境安静, 录音软件为 cooledit, 采用 22050Hz 的采样频率, 单声道, 16 位采样精度进行语音信号的录制, 每人每个数字发音 2 遍, 共 200 个样本, 前 100 个语音样本个用于训练, 后 100 个语音样本中随机抽取 90 个用于测试.

1) 采用两种降维算法进行数据降维后识别率的比较

将初始特征维数为  $5000 \times 24$  维的数据, 降低为  $1500 \times 24$  维, 同时采用传统的 PCA 降维至相同维数, 比较两种降维方法的识别率. 初始特征参数的维数为  $5000 \times 24$  维, 如表 1 所示.

表 1 语音信号特征参数 MFCC 和 LPCC 的数据

	1	2	3	...	24
1	-14.404	-4.455	-7.044	...	-0.102
2	-15.088	-4.464	-5.483	...	-0.067
3	-14.895	-3.498	-4.001	...	0.045
4	-14.862	-2.526	-3.725	...	0.056
5	-15.087	-2.532	-6.020	...	-0.086
6	-14.288	-3.257	-7.308	...	-0.127
7	-15.015	-5.208	-9.276	...	-0.068
8	-15.062	-5.528	-9.608	...	-0.073
9	-15.274	-5.021	-9.502	...	0.013
...	...	...	...	...	...
5000	-17.129	-3.578	-3.133	...	-0.233

采用遗传算法对初始参数降维后数据维数为 $1500 \times 24$ 维,由于数据量较大,选取部分数据如表2所示。

表2 采用遗传算法对特征参数降维后的低维数据

	1	2	3	...	24
1	-17.267	-6.748	1.950	...	-0.047
2	-8.358	-0.849	1.6508	...	-0.089
3	-9.334	0.662	-2.773	...	-0.041
4	-16.229	0.294	-7.129	...	-0.131
5	-17.593	-6.669	0.920	...	-0.057
6	-10.784	-0.304	-6.971	...	-0.056
7	-13.627	-2.603	-8.197	...	0.063
8	-14.148	-0.342	-7.665	...	-0.060
9	-16.399	-2.708	-3.522	...	-0.038
...	...	...	...	...	...
1500	-15.101	-4.102	-7.074	...	-0.097

采用PCA算法对初始参数降维至相同维数即 $1500 \times 24$ 维,选取部分数据如表3所示。

表3 采用PCA算法对特征参数降维后的低维数据

	1	2	3	...	24
1	-119.1	-28.877	-25.717	...	-0.787
2	0.758	-1.410	-4.386	...	0.431
3	1.26	0.933	-0.526	...	0.083
4	0.535	-0.713	4.081	...	0.063
5	-0.331	-0.669	-2.785	...	0.017
6	-0.346	4.325	0.954	...	-0.154
7	0.539	-2.601	0.483	...	0.191
8	1.454	-1.632	-3.441	...	-0.002
9	-0.007	2.125	-1.594	...	0.084
...	...	...	...	...	...
1500	0.021	-0.010	0.061	...	0.059

将这两种算法降维后的数据(表2和表3)分别送入SVM分类器,识别结果如表4所示。

表4 两种降维方法识别率的比较

降维算法	识别率
PCA	77.65%
GA	89.85%

采用遗传算法进行特征数据降维,与传统的PCA算法相比,将原始数据降低到相同维数,采用遗传算法来进行降维的识别率要高于PCA降维,由此可说明:基于遗传算法的数据降维更能适用于语音识别过程中提取出的非线性特征数据。

2)采用遗传算法识别率和识别速度的比较在100个测试语音中任意选取90个,提取语音的LPCC和MFCC,然后采用遗传算法对提取的特征参数进行降

维,采用SVM识别器,核函数选择径向基核函数,比较降维前后识别率的变化和识别的速度,实验结果如表5所示。

表5 识别率和运行时间的比较

参数	识别率(%)	运行时间(s)
LPCC+MFCC	91.08	545.41
GA+LPCC+MFCC	89.85	128.45

由表5可知,采用遗传算法对提取的特征参数(LPCC和MFCC)进行降维,识别率降低了1.23%,但是识别时间提高了416.96s。由此可见,采用遗传算法对特征参数进行降维,减少了输入分类器的数据量,有效的提高了识别速度。

### 3 结语

针对特征参数LPCC和MFCC相结合后数据维数过高,不利于后继识别器的识别,提出采用遗传算法对初始特征参数进行降维。初始特征为 $5000 \times 24$ 维,通过遗传算法降维后得到 $1500 \times 24$ 维的特征,识别率几乎没变,速度却提高为原来的5倍。PCA是一种线性的投影,不能正确的处理位于非线性流行上的数据,而遗传算法是一种具有指向性的随机搜索方法,有效的利用历史信息来推测下一代期望性能有所提高的寻优点集,最后收敛到一个最适应环境的个体上,求得问题的最优解,仿真实验表明采用遗传算法进行数据降维比传统的PCA算法更能保留语音信号的本质信息。

### 参考文献

- 马静.基于HMM模型的汉语数字语音识别算法的研究[硕士学位论文].太原:太原理工大学,2008.
- 高朝煌.非特定人汉语连续数字语音识别系统的研究与实现[硕士学位论文].西安:西安电子科技大学,2011.
- 赵力.语音信号处理.北京:机械工业出版社,2003.
- 毕达天,邱长波,张晗.数据降维技术研究现状及其进展.情报理论与实践,2013,36(2):125-128.
- 陈炜亮.基于MFCCG-PCA的语音情感识别.北京大学学报,2015,51(2):269-274.
- 李敏强.遗传算法的基本理论与应用.北京:科学出版社,2002.
- 包希日莫,高光来,张璟.基于遗传算法的声学模型拓扑结构优化.计算机工程与应用,2014,50(14):5-8.
- 高家宝.支持向量机在语音识别中的应用.软件导刊,2015,14(1):39-40.