

Primi Language Spectrogram Recognition Based on Learning Vector Quantization

Hua Yang, Tao Jiang*, Lihua He, Shaoxiong Chen, Wenlin Pan

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming Yunnan
Email: *2497821757@qq.com

Received: Dec. 5th, 2018; accepted: Dec. 17th, 2018; published: Dec. 26th, 2018

Abstract

In this paper, we proposed a Primi language speech recognition algorithm based on Learning Vector Quantization (LVQ). Firstly, the algorithm uses the Fourier transform to plot each Primi language pragmatics; secondly, it extracts the gray level co-occurrence moments of each spectrogram to construct feature vectors; and finally, it uses learning vector quantization to realize the classification of Primi isolated words. The classification accuracy based on the Learning Vector Quantization (LVQ) is 80.0%.

Keywords

Primi Isolated Words, Gray-Level Co-Occurrence Matrix, Learning Vector Quantization (LVQ)

基于LVQ的普米语语谱图识别

杨花, 江涛*, 和丽华, 陈绍雄, 潘文林

云南民族大学, 数学与计算机科学学院, 云南 昆明
Email: *2497821757@qq.com

收稿日期: 2018年12月5日; 录用日期: 2018年12月17日; 发布日期: 2018年12月26日

摘要

本文使用基于学习向量量化算法实现了对普米语语谱图的识别。该算法首先通过傅立叶变换绘制出每条普米语语料的语谱图; 再次, 提取每张语谱图的图像灰度共生矩构建特征向量; 最后, 运用学习向量量化实现对普米语孤立词的分类。基于学习向量量化算法的分类准确率达到80.0%。

*通讯作者。

关键词

普米语, 灰度共生矩阵, 学习向量量化(LVQ)

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

普米族是我国边疆少数民族之一, 主要分布在云南省怒江州的兰坪白族普米族自治县、丽江市古城区、永胜县、宁蒗彝族自治县[1]。怒江州兰坪县是全省普米族人口最集中中的一个县, 据2010年统计, 怒江州兰坪县普米族人口有18,643人[2]。普米语是一种无文字的少数民族语言, 目前由于普米族人数极少, 对普米语的研究也是空白的, 因此基于普米语的研究具有重要的文化价值和保护价值。

LVQ神经网络模型是一种具有自学习能力强、网络结构简单等优点的有监督式神经网络分类方法[3]。通过这些优点提出了一种基于LVQ神经网络的模型用于普米语语谱图分类的方法, 并对每一种网络的参数进行调整使得得到最优的一个网络结构。另外基于共生矩阵提取的纹理特征作为输入向量辅助分类[4], 对普米语语谱图的纹理特征进行数据处理, 充分挖掘该数据中的有效信息, 然后建立普米语语谱图的LVQ神经网络模型, 该模型有着较好的鉴别准确率。

2. 语谱图生成

2.1. 基于傅里叶变换绘制语谱图

语谱图, 亦称语音频谱图, 它是通过像素的灰度值大小反映相应时刻和相应频率的语音信号能量密度。本文采用短时傅里叶变换(STFT)把语音信号从时域转换到频域, 求取每帧信号的能量密度谱, 最终绘制出语谱图, 其流程如图1所示。

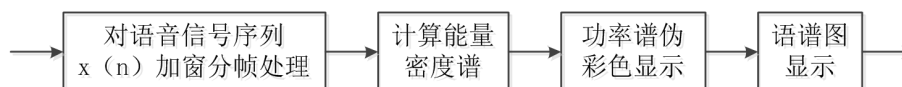


Figure 1. Flow charts for generating spectrograms

图1. 语谱图生成的流程图

2.1.1. 语谱图显示的功能函数[5]

1) 伪彩色映射函数: colormap (MAP)。其中 MAP 是所采用的伪彩色映射矩阵, 默认值为 JET, 可以通过 $MAP = colormap$ 获得当前的伪彩色映射矩阵, 它可以是一个任意行的矩阵, 但其必须有且只有三列, 并分别表示红色、绿色和蓝色的饱和度。

2) 语谱图显示函数: imagesc (t,f,L)。当数据比例映射使用整个色域时使用该函数, 其他情况下与 image() 函数作用相同。其中 t 是时间坐标, f 是频率坐标, L 则是从功率谱值经伪彩色映射后的彩色电平值。

通过上述流程, 画出了普米语汉语为“出去”的语谱图, 如图2。

在图2中横坐标表示时间, 纵坐标表示频率, 坐标点值表示语音数据能量。

2.2. 语谱图特征提取——灰度共生矩阵

灰度共生矩阵是一种研究图像纹理特征的方法, 其能反映图像灰度关于变化幅度、相邻间隔、方向

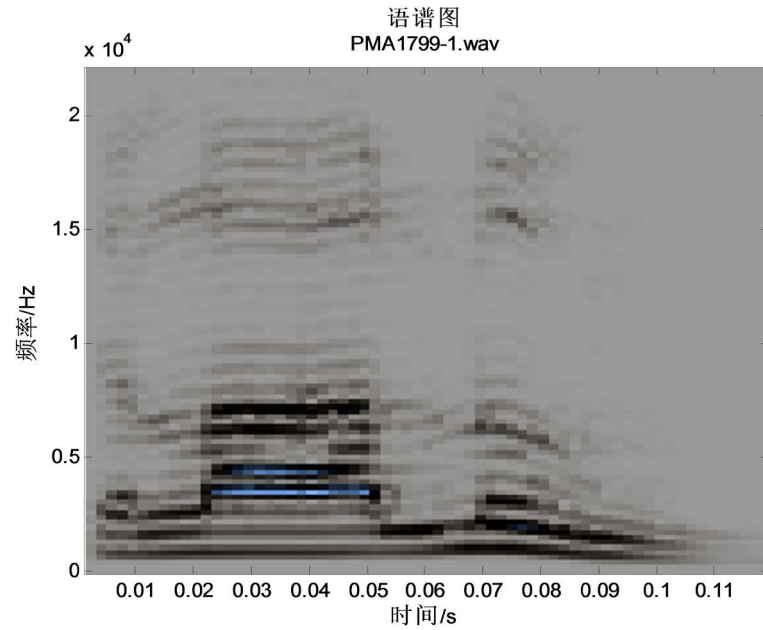


Figure 2. The spectrum of “Going Out”
图 2. “出去”的语谱图

的综合信息，也是一种对图像相邻像素的灰度相关性进行定量描述的方法。灰度共生矩阵就是从影像 (x, y) 灰度 i 的像素出发，统计与距离为 δ ，灰度为 j 的像素 $(x + \Delta x, y + \Delta y)$ 同时出现的概率 $P(i, j, \delta, \theta)$ 。用数学公式表示则为：

$$P(i, j, \delta, \theta) = \left\{ \left[(x, y), (x + \Delta x, y + \Delta y) \right] \mid f(x, y) = i, f(x + \Delta x, y + \Delta y) = j, \right. \\ \left. x = 0, 1, 2, \dots, N_x - 1; y = 0, 1, 2, \dots, N_y - 1 \right\} \quad (式 1)$$

式中 $i, j = 0, 1, 2, \dots, L - 1$ ； x, y 是影像中的像素坐标； N_x, N_y 分别为影像的行列数。

根据灰度共生矩阵，可以定义出大量反映图像在方向、间隔、变化幅度及快慢的纹理特征。

2.2.1. 算法实现步骤

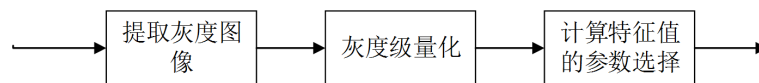


Figure 3. Flow chart of algorithm implementation
图 3. 算法实现流程图

图 3 中计算出的特征值为灰度共生矩阵。

2.3. 共生矩阵的计算

计算灰度共生矩阵时，我们取 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 四个方向作为灰度共生矩阵的方向：

$$P_{0^\circ} \cdot d(a, b) = \left\{ \left[(k, l), (m, n) \right] \in D : k - m = 0, \right. \\ \left. |l - n| = d, f(k, l) = a, f(m, n) = b \right\} \quad (式 2)$$

$$P_{45^\circ} \cdot d(a, b) = \left\{ \left[(k, l), (m, n) \right] \in D : (k - m = d, l - n = -d) \right. \\ \left. \vee (k - m = -d, l - n = d), f(k, l) = a, f(m, n) = b \right\} \quad (式 3)$$

$$P_{90^\circ} \cdot d(a,b) = \left| \left\{ \left[(k,l), (m,n) \right] \in D : |k-m|=d, \right. \right. \\ \left. \left. l-n=0, f(k,l)=a, f(m,n)=b \right\} \right| \quad (\text{式 4})$$

$$P_{135^\circ} \cdot d(a,b) = \left| \left\{ \left[(k,l), (m,n) \right] \in D : (k-m=d, l-n=-d) \right. \right. \\ \left. \left. \vee (k-m=-d, l-n=-d), f(k,l)=a, f(m,n)=b \right\} \right| \quad (\text{式 5})$$

式中 d 表示像素间隔, $(k,l), (m,n)$ 分别为原像素和偏移后的像素坐标, 其中 k, m 为纵坐标, D 为图像范围。

2.4. 灰度共生矩阵特征量

1) 对比度

$$Con = \sum_{n=0}^{L-1} n^2 \left\{ \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \hat{p}(i,j|d,\theta) \right\}^2 |i-j|=n \quad (\text{式 6})$$

2) 能量

$$Asm = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \hat{p}(i,j|d,\theta)^2 \quad (\text{式 7})$$

3) 熵

$$Ent = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \hat{p}(i,j|d,\theta) \lg \hat{p}(i,j|d,\theta) \quad (\text{式 8})$$

4) 相关性

$$Coor = \frac{1}{\sigma_x \sigma_y} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_x)(j - \mu_y) \hat{p}(i,j|d,\theta) \quad (\text{式 9})$$

其中 $\mu_x, \mu_y, \sigma_x, \sigma_y$ 分别定义为:

$$\mu_x = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} \hat{p}(i,j|d,\theta) \quad (\text{式 10})$$

$$\mu_y = \sum_{j=0}^{L-1} j \sum_{i=0}^{L-1} \hat{p}(i,j|d,\theta) \quad (\text{式 11})$$

$$\sigma_x = \sqrt{\sum_{i=0}^{L-1} (i - \mu_x)^2 \sum_{j=0}^{L-1} \hat{p}(i,j|d,\theta)} \quad (\text{式 12})$$

$$\sigma_y = \sqrt{\sum_{j=0}^{L-1} (j - \mu_y)^2 \sum_{i=0}^{L-1} \hat{p}(i,j|d,\theta)} \quad (\text{式 13})$$

其中 μ_x, μ_y 为均值, σ_x, σ_y 为标准差。

3. 分类模型

3.1. LVQ 神经网络原理

LVQ (Learning Vector Quantization)神经网络是一种有监督的聚类学习方法, LVQ神经网络由输入层、竞争层和线性输出层三层神经元组成。如下图 4 所示:

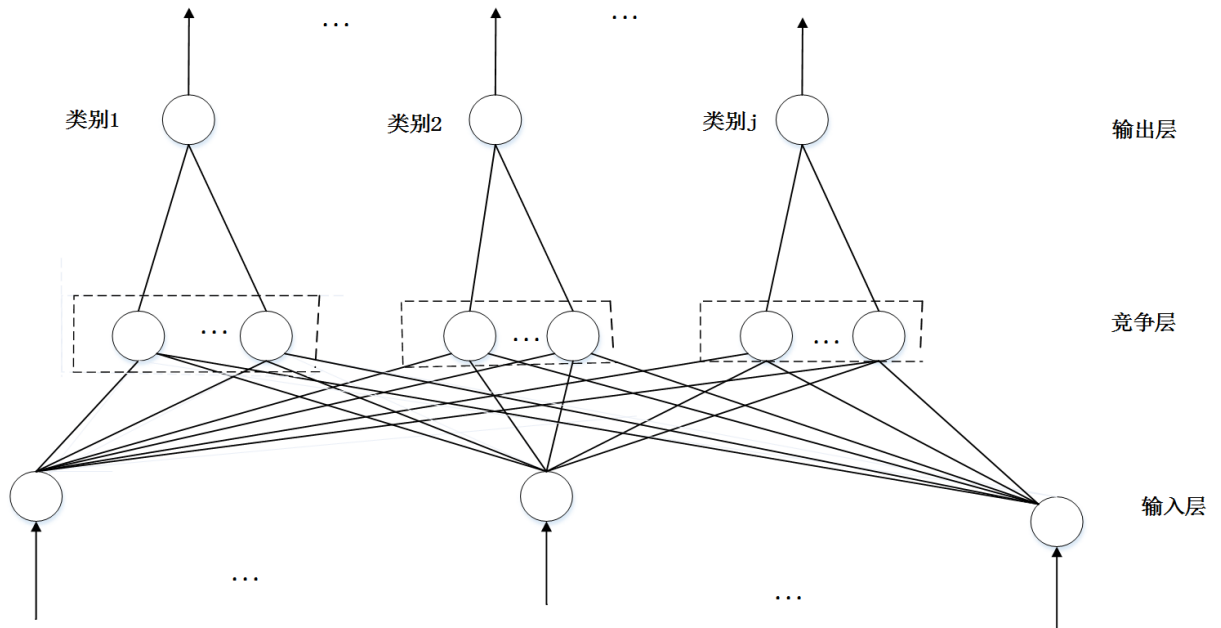


Figure 4. Structure of LVQ network
图 4. LVQ 网络的结构

3.2. 分类训练过程

分类的流程如图 5 所示:

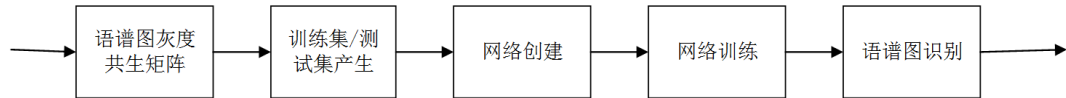


Figure 5. Flow chart of design steps
图 5. 设计步骤流程图

3.2.1. 灰度共生矩阵特征向量提取

- 步骤 1: 从语料库中随机抽取 100 个特定人发音的普米语孤立词, 每个孤立词读八遍, 共有 800 个语料;
- 步骤 2: 生成的 800 个样本语谱图, 随机抽取 600 个作为训练集, 200 个作为测试集;
- 步骤 3: 计算出训练集和测试集能量、熵、惯性矩、相关的均值、标准差、一阶矩、二阶矩、最大值、最小值作为最终纹理特征向量。

3.2.2. LVQ 分类训练步骤

步骤 1: 选定训练集和测试集后, 建立一个向量量化神经网络函数 newlvq, 其中学习速率设定为 0.01, learnlv1 为学习方法(LVQ1), 其 MATLAB 代码如下:

```

for i = 1:8
rate{i} = length(find(Tctrain == i))/300;
end
net = newlvq (minmax (Ptrain),10,cell2mat (rate),0.01,'learnlv1');
net.trainParam.epochs = 100;
net.trainParam.goal = 0.001;
net.trainParam.lr = 0.1;
    
```

步骤 2: 网络创建好后, 对网络中的迭代次数、学习速率和训练精度等参数进行设置, 并运用 `train` 函数对 LVQ 网络训练, 其 MATLAB 代码如下:

```
net = train(net, P_train, T_train);
```

步骤 3: 待 LVQ 神经网络模型训练好后, 使用 `sim` 函数将普米语语谱图的测试集数据输入到模型, 得到测试集语谱图的分类结果。其 MATLAB 代码如下:

```
T_sim = sim(net, P_test);
```

```
T_csim = vec2ind(T_sim);
```

```
result = [Tc_sim; Tc_test]
```

4. 实验结果

本实验首先从语料库中随机抽取 100 个特定人发音的普米语孤立词, 每个孤立词读八遍, 共有 800 个语料; 其次从生成的 800 张语谱图中随机抽取 600 张语谱图作为训练集, 200 张语谱图作为测试集进行实验; 最后根据灰度共生矩阵算法定义: 设相邻间隔距离 $d = 1$, 即可得到 0° 、 45° 、 90° 、 135° 四个方向上的共生矩阵, 通过计算训练集和测试集的灰度共生矩阵, 将能量、熵、对比度、相关性的均值、最大值、最小值和标准差作为最终 16 维纹理特征第 LVQ 网络进行训练, 训练结果如图 6 所示。

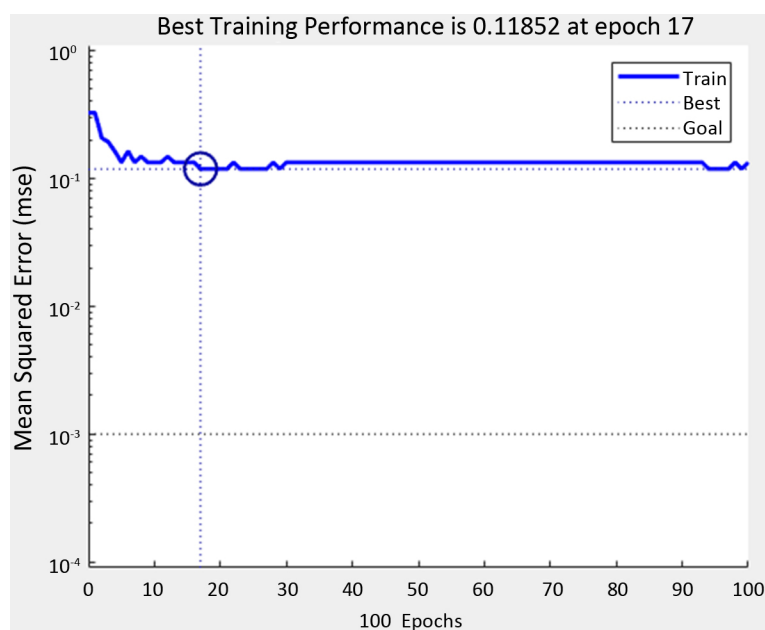


Figure 6. Error change process of network iteration 100 times

图 6. 网络迭代 100 次的误差变化过程

网络训练迭代 100 次的误差变化如图 5 所示, 从图可以看出, 随着迭代次数的增加, 当迭代次数 epoch = 17 时, 训练结果是最优的, 最终的分类准确率为 Accuracy = 80.0%。

5. 总结

本文主要是针对普米语进行分析研究的, 最后的实验结果表明, 数据归一化预处理结合 LVQ 神经网络可以实现对普米语的识别, 识别率为 80.0%, 可以证明实验结果的合理性和价值性。然而本文只基于特定人的小样本数据且仅针对 LVQ 一种方法对普米语语谱图进行识别, 因此在今后对大样本数据和不同的识别方法可以作进一步对比研究。

基金项目

国家自然科学基金项目(61363022); 云南民族大学研究生创新基金(2018YJCXS228); 云南省教育厅科学研究基金项目(2017YJS056)。

参考文献

- [1] 黄成龙. 普米语研究综述[J]. 阿坝师范高等专科学校学报, 2013(1): 17-20
- [2] 马祥林, 左显彪. 人口较少民族“民族镜”探究——以普米族、景颇族为例[J]. 经营管理者, 2017(9): 341-343.
- [3] 剑峰, 徐俊艳. 学习矢量量化的推广及其典型形式的比较[J]. 计算机工程与应用, 2006, 42(17): 82-85.
- [4] 程程, 惠晓威. 基于灰度共生矩阵的纹理特征提取[J]. 计算机系统应用, 2010, 19(6): 195-198.
- [5] 黄忠霖. 控制系统 MATLAB 计算与仿真[M]. 北京: 国防工业出版社, 2001.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org