

# 基于语谱图的江西境内赣方言自动分区研究

作者 1 作者 2 作者 3

(一级单位名称 二级单位名称, 省 市 邮编)

**摘要:** 汉语方言分区研究是语言学的重要组成部分。鉴于传统基于词汇和语法的人工方言分区方法具有一定的主观性, 该文研究了如何有效利用语音本身特征进行方言的自动分区。该文首先构建了江西省 11 个省辖市, 91 个下辖县级行政区的时长约 1500 分钟的 1223 条语音语料库。然后在传统的 MFCC 语音特征提取基础上, 提出了基于 CNN 的自编码降维语谱图的深度学习特征提取模型, 对降维后的语音特征分别采用 k 均值算法聚类、高斯混合聚类和层次聚类对方言自动分区。实验结果表明新型语谱图特征的聚类性能量度内部指标 DBI 指数以及 DI 指数显著优于传统 MFCC 特征, 维度为 16 时语谱图和 MFCC 下的拼接特征聚类效果和传统人工方言分区较为接近。

**关键词:** 语谱图; 聚类; 赣方言; 方言分区

**中图分类号:** TP391

**文献标识码:** A

## Automatic Partition of Gan Dialect in Jiangxi Province Based on Spectrogram

Author, author, and author

(Unit ,city, province zip code, China)

**Abstract:** Chinese dialect partition is a significant part of linguistics. Due to the subjectivity of traditional artificial dialect partition methods based on vocabulary and grammar, this paper studies how to effectively use the features of speech itself to auto-partition the dialect. This paper first constructs 1,223 speech corpora of about 1,500 minutes in the 11 municipalities and 91 county-level administrative regions under the jurisdiction of Jiangxi Province. Then based on the traditional MFCC (Mel-Frequency Cepstral Coefficient) speech feature extraction, a deep learning feature extraction model based on CNN self-coding dimension reduction spectrogram has been put forward. After dimension reduction, the speech feature of the dialects will be auto-partitioned through k-means clustering, Gaussian mixture clustering and hierarchical clustering respectively. The results revealed that the internal indicators of cluster performance metrics, DBI and DI index, of the new language spectrogram features significantly outperform traditional MFCC features. With the dimension of 16, the clustering effect of feature stitching under the spectrogram and MFCC is close to that of the traditional artificial dialect partition.

**Key words:** spectrogram; clustering; Gan Dialect; dialect partition

### 0 引言

汉语方言的变化与人类历史的变迁、社会活动有着密切的关系。某一地域的方言与其历史方言的关系纷繁复杂, 有的是在其历史方言的基础上传承演变而来, 有的是由于战争动乱、人类迁徙等原因消亡, 有

的语音已经发生明显变化却依旧保留历史特征<sup>[1]</sup>。其中, 方言的分区是文化交互的内在关系的实质体现。对于方言分区的原则、依据和条件, 现代语言学家各持一词。主流的分区方式主要以语言因素为重要依据, 在对方言语音、方言词汇以及方言语法进行充分的调查研究基础上, 通过古今语音比较的方式, 辅以社会

---

历史背景方面的资料，并结合地理类型和行政区域等其他因素，根据各地方言中表现出来的语言特征对方



言片区进行划分。计算机自动识别汉语方言对发掘方言文化的内涵关系有着重要的进步意义，而方言的自动分区对提高方言识别精度有着重要的作用。

赣方言（赣语）是中国汉语七大方言之一，为汉族江右民系使用的主要语言。使用范围主要在江西省境内，分布在赣江的中下游、抚河流域、鄱阳湖流域以及周边、湘东和闽西北、皖西南、鄂东南和湘西南等地区，使用人口约在 5500 万左右。目前，现代语言学家对江西省境内赣方言（以下简称赣方言）分区的主流方案都是采用人工划分方式<sup>[2, 3, 4, 5, 6]</sup>，主要采用方言词汇和语法特点进行人工分区。在汉语方言与计量研究上，先后有学者发表了一些颇有影响的文章和专著<sup>[7, 8, 9, 10, 11, 12]</sup>，这些文章和专著都从理论和实践两方面对计量研究在汉语方言关系研究中的地位、作用和意义作了探讨。近年来，部分学者开始在汉语方言的分区上尝试采用计量分析方法，通过聚类分析对现有方言的语音特征进行方言的分类或方言分区。而在如何利用计算机自动提取方言的语音特征，并对其进行分类分析鲜有文献著作。

基于此，本文首先构建了江西省 11 个省辖市，91 个下辖县级行政区的时长约 1500 分钟的 1223 条语音语料库。然后分别提取方言语音中梅尔频率倒谱系数 (Mel-Frequency Cepstral Coefficient, 以下简称 MFCC) 特征以及语谱图 2 种不同的语音特征。针对语音特征提取的维度过大问题，在 MFCC 特征上采取了 PCA (Principle Component Analysis) 降维处理，在语谱图特征上采用基于卷积神经网络 (Convolutional Neural Networks, 以下简称 CNN) 系统的自编码降维处理。最后，对降维后的语音特征分别采用 k 均值算法聚类、高斯混合聚类和层次聚类对方言自动分区。实验结果表明新型语谱图特征的聚类性能量度内部指标 DBI 以及 DI 指数显著优于传统 MFCC 特征，维度为 16 时语谱图和 MFCC 下拼接特征的聚类效果和传统人

工方言分区较为接近。

本文后续内容安排如下：第 1 节介绍赣方言分区的相关工作；第 2 节介绍本文采集的赣方言语音语料库；第 3 节详细阐述了本文使用的 2 种语音特征及聚类算法；第 4 节描述了在不同算法下评价指标的实验结果和分析；第 5 节是结论及后续工作展望部分。

## 1 相关工作

本节对方言的人工分区、计量分区方法以及语音的特征提取进行简要介绍。

### 1.1 人工分区方法

针对江西省境内赣方言（以下简称赣方言）的分区，颜森先生将其划为五个片区，分别是昌靖片、宜萍片、吉莲片、鹰弋片和抚广片，分区标准既采取了语音标准，也采取了词汇标准<sup>[2]</sup>。李如龙先生划为三个片区，分别为赣东区、赣中区和赣北区，分区标准主要以词汇为标准，根据相同词汇的接近总数的多少来分区<sup>[3]</sup>。刘纶鑫先生的分区是将赣方言划为五个片区，分别是波阳片、宜春片、临川片、都昌片和奉新片，分区标准主要采取了综合标准，将语音、词汇和语法特点综合考虑<sup>[4]</sup>。孙宜志、陈昌仪和徐阳春先生划为南北两大区，共七个小片区，北区包括都昌片、乐平片和奉新片，南区包括崇仁片、铅山片、泰和片和分宜片，分区标准主要采取了语音标准，也考虑了自然地理和行政区划的关系<sup>[5]</sup>。谢留文先生在前任学者基础上划为九个片区，分别是昌靖片、宜浏片、吉茶片、抚广片、鹰弋片、大通片、耒资片、洞绥片和怀岳片，分区标准主要采取了语音标准划分<sup>[6]</sup>。可以看出，这些研究运用了传统的语言学方法，通过田野调查收集语料，根据语音特征对赣方言进行分区。

### 1.2 计量分区方法

从 20 世纪 70 年代初开始，郑锦全、陆致极、陈海伦、王士元等语言学家将计量方法应用于汉语方言研究，讨论了方言关系的材料和计量单位问题，区分

了方言的亲疏关系和亲缘关系，并在方言上进行了不同计量方法的实践和研究<sup>[7, 8, 9, 10, 11, 12]</sup>。现代语言学家对赣方言分区的主流方案主要是根据方言词汇和语法特点，结合行政地理特征进行人工分区。近几年，已有部分学者开始尝试采用计量分析方法对汉语方言分区进行研究，例如项梦冰先生的浙南方言分区<sup>[13]</sup>、王荣波先生的江淮官话洪巢片分区<sup>[14]</sup>等，主要采取聚类分析的方法对现有方言的语音特征进行方言的分类或方言分区。而在如何利用计算机自动提取方言的语音特征，并对其进行聚类分析鲜有文献著作。

### 1.3 语音特征提取

特征提取是从说话人语音信号中获得能够描述语音信号特征参数的过程，是语音识别过程中至关重要的一步。现有的特征提取方法包括线性预测编码系数（Linear Predictive Coding, LPC）<sup>[15]</sup>、线性预测倒谱系数（Linear Predictive Cepstral Coefficient, LPCC）<sup>[16]</sup>以及梅尔频率倒谱系数（Mel Frequency Cepstral Coefficients, MFCC）<sup>[17]</sup>等。

90年代初，潘凌云<sup>[18]</sup>等就提出了使用了语谱图进行语音实验，利用语谱图密度变化的形变函数，以及自适应阈值技术来定位每个音素段的边缘，文章所得结果与语音学家分割的结果进行比较，得到的识别率高于93%。近几年，语谱图特征的应用也是较为广泛，如文献<sup>[19]</sup>提出将语谱图特征应用于语音情感识别。文献<sup>[20]</sup>将语谱图输入到有生物视觉依据的人工神经网络——脉冲耦合神经网络，得到输出图像的时间序列及其熵序列作为说话人语音的特征，利用它的不变性实现说话人识别等。

## 2 赣方言语音语料库

本节主要介绍赣方言语音语料库的采集工作。

### 2.1 语料库设计

如何选取录音文本语料，是语料库建库工作的关键。为了保证语料库的质量，体现方言语料的特点，

在语料库构建之前，本文按照以下原则选取了语料库的文本语料：(1) 语料库中的单字、词尽量涵盖声韵现象，以便更好地反映该方言语音的音系特征；(2) 语料库中的词汇依据汉语调查常用表为基础，选取了具备客赣方言特色的口语语料，以便更加符合语音识别面对的真实情形；(3) 语料库中的句子在内容和语义上尽量保证完整，能够尽可能地反映一个句子的韵律信息；(4) 要求发音人在自然状态下说方言，从而反映语音特征。本文依据此原则，参照国际上语音语料库的设计标准，结合汉语方言之间的差异性，选取了江西省11个省辖市，91个下辖县级行政区进行录音采样（如图1所示）。



图1 方言点采样

在确定方言点之后，本文采用自然语音的朗读文本，提示是汉语文字材料。说话人选择的是生活或居住在方言采集点十年以上的高校新生，包含学生姓名、性别、出生年月、出生地、现居住地、方言区生活时间和录音时长该方言语音语料库将语音中的性别、年龄、地域等信息的进行综合处理，适用于语音识别和方言特征识别等研究。例如姓名：某某某，性别：女，

出生年份：2000 年，民族：汉族，出生地：九江市星子县南康镇迎春桥热，现居住地：九江市星子县南康镇黄泥岭，在方言区居住年份：18 年，方言所在地经纬度：东经 116.0517, 北纬 29.46204, 录音 1 时长：37s, 录音 2 时长：34s。

方言语料采集参与人数共 740 人，其中男性 186 人，占总比率 25.1%，女性 554 人，占 74.7%。17 至 20 岁学生人数为 537 人，占 72.6%。最后，共采集录制 1223 条语音，其中用于特征提取的有效录音 936 条，时长约 1500 分钟。方言点南昌地区录音人数 78 人，九江地区 64 人，上饶地区 62 人，抚州地区 38 人，宜春地区 111 人，吉安地区 134 人，赣州地区 158 人，景德镇地区 29 人，萍乡地区 31 人，新余地区 16 人，鹰潭地区 29 人。地域分布基本符合方言分区的均匀采样原则。

### 3 赣方言语音特征提取及自动分区

本节主要描述了语谱图的特征提取及基于 CNN 的自编码器降维、MFCC 特征的提取和 PCA 降维以及所采用的聚类算法。

#### 3.1 语谱图及基于 CNN 的自编码器降维描述

作者首先提取每一条语音文件的音频参数，例如

声道数 (nchannels:1)，量化位数 (sampwidth:2)，采样频率 (framerate:16000)，采样点数 (nframes:不同长度语音采样点数不同，大概范围 300000~910000)。将这些得到的语音参数 (字符串类型) 转化为整型参数并且进行归一化处理，可以得到语音的帧长和帧叠点数等参数。最后将这些参数作为输入得到对应语音的语谱图。

为了对语谱图特征进行降维，本文构造了基于卷积神经网络 (Convolutional Neural Networks, CNN 网络) 的自编码降维系统。自编码器 (Autoencoder) 是一种利用反向传播算法使得输出值等于输入值的神经网络，它先将输入压缩成潜在空间表征，然后通过这种表征来重构输出。作者使用自编码器提取语谱图的瓶颈特征，输入是原始语谱图，输出是生成的语谱图。自编码器由编码器和解码器组成，编码器将语谱图压缩成瓶颈特征，解码器将瓶颈特征还原成语谱图。

本文所使用的 CNN 网络结构包括输入层 (Input layer)、卷积层 (Conv layer)、编码层 (Encoder layer)、解码层 (Decoder layer)、最大池化层 (Max pool)、和输出层 (Output)。

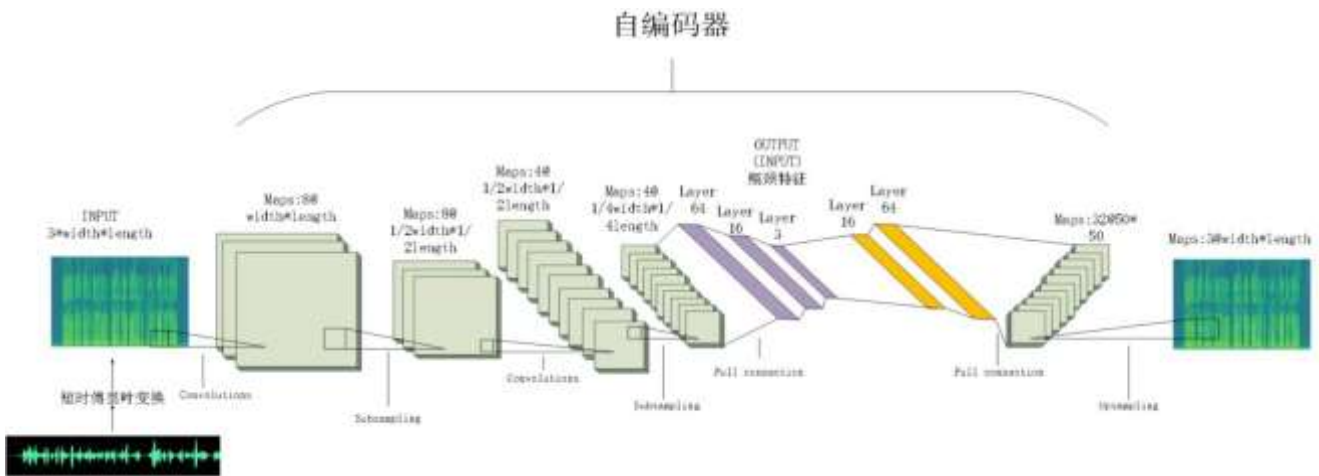


图 2 语谱图自编码器

如图 2 所示，模型结构图中输入层输入的是由语音文件产生对应的语谱图，输出层最后输出每一个语谱图对应的编码层和解码层计算后的特征向量。编码

器由两层卷积层、两层最大池化层和 3 层全连接层组成，两层卷积层的卷积核大小为 3\*3 和 2\*2，步长为 (1, 1), (2, 2)。其中第一层卷积核数量为 8，第二层卷

积核数量为 4。两层最大池化层的卷积核大小均为  $2 \times 2$ ，步长均为 2。卷积层不改变图像的大小，最大池化层将图像长宽减半。三层全连接层分别将数据降维到 64 维、16 维和 3 维，其中 3 维是瓶颈特征的大小。解码器由三层全连接层和一层反卷积层组成，三层全连接层分别将 3 维的瓶颈特征升维到 16、64 以及解码器第二层池化后的维度（记此特征为 Linear\_3），再经过一层反卷积将 Linear\_3 还原至语谱图。作者使用 L1 计算输入语谱图与生成语谱图之间的损失。通过 Adam 算法更新模型参数，同时学习率设置为 0.001。

### 3.2 MFCC 特征及 PCA 降维描述

本文将每一个语音信号首先分帧处理，将信号帧化为 15 ms/帧，帧移为 10ms，对每帧进行快速离散傅里叶变换(Fast Fourier Transformation, 以下简称 FFT)，从时域数据转变为频域数据能量分布来观察。对 FFT 的数据计算谱线的能量，得到向量特征，在梅尔域内能量谱经三角带通滤波器后得到 26 个对数滤波器组能量。最后，采用 26 个对数滤波输出经过离散余弦变换 (Discrete Cosine Transform, 以下简称 DCT)，最终得到每帧语音的 13 维的 MFCC 特征向量。其中，1s 的语音按照帧移为 10ms 来计算，可以切分出 100 帧，每帧的特征向量是 13 维，即长度为 1s 的语音得到的特征向量长度为 1300 维。如此高维的特征向量对于聚类来说计算量巨大。因此，本文使用 PCA 方法对 MFCC 的高维特征进行降维处理（如图 3 所示）。主成分分析 PCA 也称主分量分析，它是一种将原有的多个变量通过线性变换转化为少数几个新的综合变量的统计分析方法。这些新变量互不相关，即能有效地表示原变量的信息，也在降维之后依然能够最大化保持数据的内在信息。MFCC 特征在经过 PCA 处理后，为后续进行聚类分析工作既减少了运算量又降低了数据

的存储量，同时还对语音的特征参数进行了最优化。

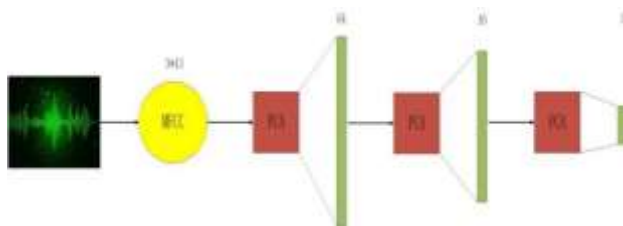


图 3 MFCC 特征提取及 PCA 处理过程

### 3.3 聚类分析

聚类是将没有分类的标签数据集分为若干个簇的过程，是一种无监督的分类方法。聚类分析的过程则是将聚类对象的数据集进行特征的选择或变换，再通过聚类算法得出结果进行评价。如上文所述，本文将采集到的赣方言语音语料进行预处理，提取每条语音的 MFCC 特征和语谱图特征，作为聚类分析的特征选择，并对特征进行降维处理。由于特征的选择并不会改变其原有属性，所以结果只是一个原始属性的优化特征子集，保留了原属性的物理意义。聚类簇的选择依靠聚类结束准则函数得到，所以，这种准则函数一般由人为设定的终止条件实现。本文在传统语言学家对赣方言分类的基准上，人工将聚类簇定为 3、5、7、9 类，并分别采用传统的 k-means 聚类，语言特征常用的层次聚类和语音识别常用的 GMM 聚类方法进行比较。

## 4 实验结果及分析

本节描述性能评价指标，不同语音特征下的聚类结果及对比分析。

### 4.1 性能评价指标

本文使用聚类性能度量内部指标来评价不同聚类之间的效果。常用的内部指标有 DB 指数 (Davies-Bouldin Index, 简称 DBI) 和 Dunn 指数 (Dunn Index, 简称 DI)。

(1) DB 指数 (Davies-Bouldin Index, 简称 DBI)

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{avg(C_i) + avg(C_j)}{d_{ccn}(\mu_i, \mu_j)} \right) \quad \text{公式 (1)}$$

(2) Dunn 指数 (Dunn Index, 简称 DI)

$$DI = \min_{1 \leq i \leq k} \left\{ \min_{j \neq i} \left( \frac{d_{\min}(C_i, C_j)}{\max_{1 \leq l \leq k} diam(C_l)} \right) \right\} \quad \text{公式 (2)}$$

其中, DBI 的值越小说明类间距离越小, 同时类内距离越大, 而 DI 则相反。

## 4.2 实验结果分析

本文依据上述工作, 在赣方言地区的 76 个下辖县、区中随机抽取一条语音进行实验。表 1 及表 2 列出语谱图和 MFCC 特征下的三种聚类方法的评价指标的四种结果 (字体加粗数据为更优数据)。实验结果表明, 3 分类上, MFCC 特征的 DBI 数据要优于语谱图特征, 5 分类上, MFCC 特征的层次聚类效果优于语谱图特征。总体而言, 语谱图特征下的聚类效果要优于 MFCC 特征的聚类效果。

表 1 语谱图特征在不同聚类下比较

		3 分类	5 分类	7 分类	9 分类
K-mean s	DBI	0.1257	<b>0.1113</b>	<b>0.1105</b>	<b>0.1179</b>
	DI	<b>9.8283</b>	<b>1.1794</b>	<b>1.7865</b>	<b>2.1773</b>
GMM	DBI	0.1257	<b>0.2474</b>	<b>0.1200</b>	<b>0.1179</b>
	DI	<b>9.8283</b>	<b>8.6775</b>	<b>1.7161</b>	<b>2.1773</b>
层次	DBI	0.1257	0.6688	<b>0.1105</b>	<b>0.1179</b>
	DI	<b>9.8283</b>	0.4878	<b>1.7865</b>	<b>2.2398</b>

表 2 MFCC 特征在不同聚类下比较

		3 分类	5 分类	7 分类	9 分类
K-mean s	DBI	<b>0.0659</b>	0.2049	0.3042	0.5248
	DI	1.3840	0.7935	0.8104	0.7251
GMM	DBI	<b>0.0659</b>	0.5340	0.5337	1.4261
	DI	1.3840	0.5010	0.4946	0.5280

层次	DBI	<b>0.0659</b>	<b>0.3068</b>	0.5212	0.4525
	DI	1.3840	<b>0.7668</b>	0.6135	0.7664

一条方言语音不仅反映了方言的音位系统、声韵调系统、音节系统, 还包含不同地域的有连续音变的多音词的变调、变声、变韵以及和音的规则。通过实验, MFCC 下能得到语音特征中频率的出现, 却无法得知该频率出现的时间点。而语谱图特征中则蕴含了大量的与语音的语句特性有关的信息, 它综合了频谱图和时域波形的特点, 明显地显示出语音频谱随时间的变化情况, 对方言而言是一种很好的区分性特征。

## 4.3 实验验证

为验证语音特征下方言自动分区的效果, 本文将方言自动分区的数据和语言学家的方言分区进行对比。实验结果表明, 在 16 维的语谱图和 MFCC 拼接特征下, 语谱图特征权重为 1 的拼接效果最为接近。

本文选取颜森老师的方言分区为例。5 分类而言, 如表 3 所示, 文章所采用的语音拼接特征聚类效果与颜森老师方言分区有部分重叠 (字体加粗为重合部分)。

表 3 语谱图权重为 1 的拼接特征下  
5 个片区的接近数据

		方言片				
5 大片区		昌靖片	宜萍片	吉莲片	抚广片	鹰弋片
方言小片	南昌	宜春	<b>峡江</b>	<b>临川</b>	鹰潭	
	新建	宜丰	吉安	崇仁	<b>贵溪</b>	
	安义	<b>上高</b>	<b>吉水</b>	<b>南城</b>	<b>万年</b>	
	永修	新干	泰和	黎川	乐平	
	<b>修水</b>	新余	安福	资溪	景德镇	
	<b>德安</b>	<b>分宜</b>	莲花	金溪	余干	
	<b>都昌</b>	萍乡	<b>永新</b>	<b>东乡</b>	<b>波阳</b>	



	湖口	丰城	井冈山	进贤	彭泽
	高安	万载	万安	南丰	横峰
	奉新		遂川	广昌	弋阳
	武宁				铅山

颜森老师将南昌市等十四个市县划入昌靖片，并总结出该片区共同的 2 个音系特点和若干特例。而本文将两种语音特征进行拼接，不同权重得出不同的聚类效果，如语谱图特征权重为 0.9 的特征拼接下，新建、安义、德安、都昌以及湖口语音特征聚在一类；语谱图特征权重为 1 的特征拼接下修水、武宁、都昌和德安语音特征聚在一类。从江西地势上而言，武宁、修水、奉新等地有九岭山脉穿过，都昌、德安、永修等地绕鄱阳湖水域，南昌、新建、安义等地属于省会城市区域范围。语言学家在进行方言调查的同时，会将音系特征、地理因素和行政区域或多或少的考虑在划分范围。而计算机的自动分区更关注语音本身特征的划分。说明在语音特征提取下的方言聚类可以为传统语言学上具有争议的片区划分提供一定的参考。

本文又选取萍乡市为例，将语音特征提取的聚类和传统的语言学家的方言分区的聚散作可行性分析比较。传统语言学家在萍乡话的分区上就有不同的意见。对萍乡话的集中分区的代表有：颜森老师将萍乡话分在宜萍片。对萍乡话内部离散的分区的代表有：陈昌仪老师将萍乡话分在宜春片，将莲花话分在吉安片；孙宜志老师将萍乡话分别分在北区的奉新片和南区的泰和片宜春片区；谢留文老师将萍乡市内的萍乡话、上栗话和芦溪话分在宜浏片区，莲花话分在吉茶片区。萍乡属江西省下辖设区的地级市，下辖莲花、上栗、芦溪三县和安源、湘东两区。在本文所提取的语音特征中，作者根据语音特征值为坐标，计算机既能在语

谱图特征下将萍乡地区的三县两区语音聚合（如图 4 蓝色标示所示），也能在 MFCC 特征下将三个县两区分散（如图 4 红色标示所示）。这也为具有争议的方言归属提供了一定的参考价值。

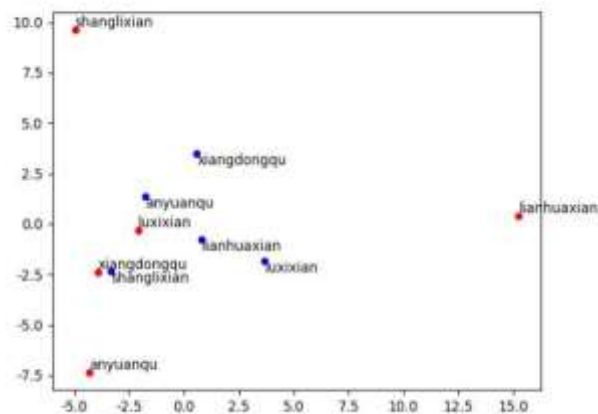


图 4 萍乡方言在 MFCC 特征（红色）和语谱图特征（蓝色）下的聚类效果

## 5 总结

采用计算机对汉语方言进行自动分区对语言学研究有着重要的印证作用。在言语工程领域，方言识别能够为带口音的语音识别、说话人识别等方面的研究打下良好的基础。在信息查询和检索服务领域，方言的语音识别可以作为一个前端处理，预先区分用户的方言类别，以便于接受不同方言的语音服务。

本文构建了江西省范围内的赣方言语音语料库，在传统的语音特征提取基础上，设计了基于语谱图的深度学习模型提取特征。最后，通过聚类性能量度内部指标评价了不同聚类方法上的方言分区的可行性。实验结果表明了在语谱图特征的有效性，维度为 16 时语谱图特征下的聚类效果和传统人工方言分区较为接近。在后续的工作中，作者将研究其他类型的学习模型对赣方言语音特征提取的作用，提高方言自动分区精度。

## 参考文献

- [1] 李国强, 傅伯言. 赣文化通志[M]. 江西教育出版社, 2004.
- [2] 颜森. 江西方言的分区(稿)[J]. 方言, 1986(01):19-38.
- [3] 李如龙. 客赣方言调查报告[M]. 厦门: 厦门大学出版社, 1992.
- [4] 刘纶鑫. 客赣方言比较研究[M]. 北京: 中国社会科学出版社, 1999.
- [5] 孙宜志, 陈昌仪, 徐阳春. 江西境内赣方言区述评及再分区[J]. 南昌大学学报(人文社会科学版), 2001(02):110-117.
- [6] 谢留文. 赣语的分区(稿)[J]. 方言, 2006(03):264-271.
- [7] 陆致极. 汉语方言间亲疏关系的计量描写[J]. 中国社会科学, 1987(01):155-166.
- [8] 郑锦全. 汉语方言沟通度的计算[J]. 中国语文, 1994(01):35-43.
- [9] 沈榕秋. 谈汉语方言的定量研究[J]. 语文研究, 1994(02):45-52.
- [10] 陈海伦. 论方言相关度、相似度、沟通度指标问题[J]. 中国语文, 1996(05):361-368.
- [11] 陈海伦. 方言间韵母系统相似度测度研究[J]. 中国语文, 2000(02):139-145+190.
- [12] 陈鹏飞. 计量方法在汉语方言关系研究中的运用[J]. 天津师范大学学报(社会科学版), 2006(02):66-69+80.
- [13] 项梦冰. 从语言特征分布图到方言分区图——以沂南方言为例[J]. 现代语言学, 2018, 6(1): 61-76.
- [14] 王荣波, 贾桂云. 类聚分析在方言分区上的应用——以江淮官话洪巢片为例[J]. 国际汉语学报, 2012, 08(01):47-56.
- [15] 鲁晓倩. 基于VP树和GMM的说话人识别研究[D]. 中国科学技术大学, 2014.
- [16] Skowronski M D, Harris J G. Increased MFCC filter bandwidth for noise-robust phoneme recognition[C]// IEEE International Conference on Acoustics. IEEE, 2002.
- [17] HuFengsong, ZhangXuan. Speaker recognition method based on Mel frequency cepstrum coefficient and inverted Mel frequency cepstrum Coefficient [J]. Journal of Computer Applications, 2012, 32(9): 2542-2544 (inChinese).
- [18] 潘凌云, 孙达传, 吴美朝. 语音识别中基于语谱图的语音音素分割方法[J]. 杭州大学学报(自然科学版), 1995(01):42-46.
- [19] Badshah A M, Ahmad J, Rahim N, et al. Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network[C]// International Conference on Platform Technology & Service. 2017.
- [20] 马义德, 袁敏, 齐春亮, 刘悦, 刘映杰. 基于PCNN的语谱图特征提取在说话人识别中的应用[J]. 计算机工程与应用, 2005(20):81-84.
- [21] 项梦冰. 方言地理、方言分区和谱系分类[J]. 龙岩学院学报, 2012, 30(04):21-30.
- [22] 刘新中, 杨蔚, 崔淑慧. 关于汉语方言分区的语言条件[J]. 语言研究, 2003(03):113-118.

- [23] 孙宜志.江西赣方言语音的特点[J].南昌大学学报,2001,32 ( 04):122-127.
- [24] 刘镇发.百年来汉语方言分区平议[J].学术研究,2004(04):125-130.
- [25] 熊正辉.汉语方言的分区[J].方言,2005 ( 02):97-108.
- [26] 王临惠.汉语方言分区的语音标准刍议[J].南开语言学刊, 2005(01):18-22+226.
- [27] 李小凡.汉语方言分区方法再认识[J].方言,2005(04):356-363.
- [28] 谢留文.赣语的分区(稿)[J].方言,2006 ( 03):264-271.
- [29] 侯兴泉.汉语方言区片划分语音标准的层级和主次[J].暨南学报(哲学社会科学版),2013,35(09):147-154+164.
- [30] 李蓝.方言分区的历史与方法平议[J].语言战略研究,2018,3(02):38-49.
- [31] 李爱军.863语音识别语音语料库 RASC863—四大方言普通话语音库[A].中国中文信息学会,2003:4.
- [32] 殷志刚.语音语料库的建设和作用[J].中国社会科学院院报,2006 ( 03).
- [33] 陈小莹,陈晨,华侃,于洪志.语音语料库的设计研究[J].科技信息,2008(36):5-6
- [34] 杨鸿武,梁青青,郭威彤,李锦珑,陈龙.一个面向言语工程的兰州方言语料库[J].西北师范大学学报(自然科学版),2009,45(06):54-59.
- [35] 邹法欣.语音语料库的设计与实现[D].广西师范大学,2012.
- [36] 邓海龙.赣南客家方言语音语料库及其检索平台的设计与实现[J].文学语言研究,2016(81):16-17.
- [37] 高原,顾明亮,孙平,王侠,张长水.多用途汉语方言语音数据库的设计[J].计算机工程与应用,2012,48(05):118-120.