



# Research on Automatic Speech Analysis for Teacher's Q&A in Classroom

Wenquan Chang<sup>1</sup>, Dongxing Li<sup>1</sup>, Zuying Luo<sup>1, 2, \*</sup>

<sup>1</sup>College of Information Science and Technology, Beijing Normal University, Beijing, China

<sup>2</sup>Beijing Key Laboratory of Digital Preservation and Virtual Reality for Cultural Heritage, Beijing, China

## Email address:

changcwq@mail.bnu.edu.cn (Wenquan Chang), lidx@bnu.edu.cn (Dongxing Li), luozy@bnu.edu.cn (Zuying Luo)

\*Corresponding author

## To cite this article:

Wenquan Chang, Dongxing Li, Zuying Luo. Research on Automatic Speech Analysis for Teacher's Q&A in Classroom. *Science Innovation*. Vol. 5, No. 3, 2017, pp. 144-150. doi: 10.11648/j.si.20170503.14

**Received:** March 28, 2017; **Accepted:** April 17, 2017; **Published:** April 20, 2017

**Abstract:** Question-and-answer (Q&A) is the important content of classroom evaluation. In order to meet the needs of large-scale class evaluation, this work uses voice analysis technology to implement exploratory research on automatic Q&A analyse, mainly on speaker recognition based on MFCC Gaussian mixture model and closed/open question identification based on the logistic regression method. The Q&A auto-analysis system developed in this work can perform automatic analyses for several classroom evaluation indexes including Q&A times, speaking durations of the teacher and his students, and the number of open questions and closed questions. For the real classroom teaching video down-loaded from "CCTV network Chinese public class", both the speaker recognition and closed/open questions identification of this work have obtained satisfied recognition accuracy with recognition rates above 93%.

**Keywords:** Classroom Evaluation, Interaction Among a Teacher and Students, Logistic Regression Model, Speaker Recognition

---

## 面向课堂师生问答的自动语音分析研究

常文泉<sup>1</sup>, 李东兴<sup>1</sup>, 骆祖莹<sup>1, 2\*</sup>

<sup>1</sup>信息科学与技术学院, 北京师范大学, 北京, 中国

<sup>2</sup>虚拟现实应用教育部工程研究中心, 北京, 中国

## 邮箱

changcwq@mail.bnu.edu.cn (常文泉), lidx@bnu.edu.cn (李东兴), luozy@bnu.edu.cn (骆祖莹)

**摘要:** 提问是课堂教学与评价的重要内容。为了满足大规模课堂评价的需求, 本文应用语音分析技术对课堂提问自动分析进行探索性研究, 主要对基于MFCC混合高斯模型的说话人识别和基于逻辑回归方法的封闭/开放性问题识别等内容进行了研究。本文构建的师生问答自动分析系统可以对师生问答次数、时长, 以及开放性问题与封闭性问题次数等课堂评价指标进行自动分析。对于“央视网中国公开课”给出的真实课堂教学视频, 本文系统的说话人识别和封闭/开放性问题识别都获得了满意的识别效果, 识别率达到93%以上。

**关键词:** 课堂评价, 师生交互, 说话人识别, 逻辑回归

---

## 1. 引言

课堂教学一般包括三个阶段：计划（设计）、实施（实践）、评价（反思），彼此之间是一个循环往复，周而复始的过程[1]。其中课堂评价（反思）对课堂教学效果进行过程性分析，主要针对师生交互、学生参与度、教师教学技能等诸多指标进行评价[1]。因此课堂评价是课堂教学一个必要的环节，是基础教育质量监测与提升的关键[2]。考虑到现有课堂教学过程中的量化评价指标较多，需要投入大量人力、物力才能取得有效的课堂评价数据[2]，其工作量之大，很难进行大范围常态化的课堂评价。其中，师生交互行为是师生双方进行知识交互、互相交流思想和情感、传递信息并相互影响的过程[3]，师生互动分析也就成为教学评价的主要手段。由于缺乏计算机自动评价手段，同时受到人工成本影响，目前师生互动分析研究规模相对较小，只能对少量课程进行全人工分析。因此，将信息技术应用于全自动的师生互动分析与评价是开展大规模常态化课堂评价的关键。

国内外很多学者对课堂教学师生互动量化评价进行了大量研究，并取得不错成就。其中最具有代表性的是弗兰德斯互动编码（Flanders Interaction Analysis Coding, FIAC）[4]，该编码技术由受过专业培训的编码人员看视频和听录音进行每3秒一次的手工编码，具有分类科学、编码客观的优点，所以该类手工编码技术被广泛地应用于教学科研中，一般对120分钟以内、最多3-4节课的师生互动进行量化分析[4]。考虑到该类手工编码技术分类不多的缺点，国内提出增加分类数量、细化课堂观察指标的编码系统，其中具有代表性的是方海光提出的14类iFIAS[5]和顾小清提出的18类ITIAS[6]。但上述手工编码技术都存在一个明显缺点，受限于高昂的人工费用，这类手工编码技术难以进行大规模的师生互动。

为了克服现有手工编码技术的缺点，本文基于语音端点检测、说话人识别和逻辑回归分类技术，从理论和技术角度对课堂师生互动进行自动评价研究，以降低师生互动评价的工作量、扩大师生互动评价的应用范围。为此，本文先设计了一套课堂语音采集、分割、识别系统，并自建一个“网易公开课—大学生视频公开课”语音库，对该系统识别率进行验证。通过自动标注和手工标注交叉验证的方式获取封闭性问题和开放性问题数据，使用逻辑回归技术训练该数据以得到最终问题分类模型。最后将两者结合构建了一套用户课堂评价的师生问答自动分析系统，该系统输出结果包括：教师说话总时长、学生说话总时长、教师提出开放性问题数量、封闭性问题数量。

实验结果表明：（1）在美国LDC建立的TIMIT英语语音库和自建的“网易公开课”语音库上，本文说话人识别系统均识别率均达到96%以上；（2）本文选取“央视网中国公开课”进行教师说话时长、学生说话时长、沉默时长统计，两课例识别相对误差分别为5.28%和12.76%；（3）在“央视网中国公开课”中选取六堂课进行提问类型分类模型训练，该模型对封闭性问题和开放性问题的平均识别率可达到93%以上。

## 2. 师生问答自动分析的系统架构

为了能够全自动地对师生互动进行评价，本文基于现有的语音处理技术，构建了师生问答自动分析系统。如图1所示，本文系统主要包括如下3个部分：（1）语音信号处理部分：语音端点检测和说话人识别；（2）分类模型训练部分：特征提取、封闭/开放性问题分类模型设计及训练；（3）结果输出部分：教师说话时长、学生说话时长、沉默时长、开放性问题数量和封闭性问题数量。

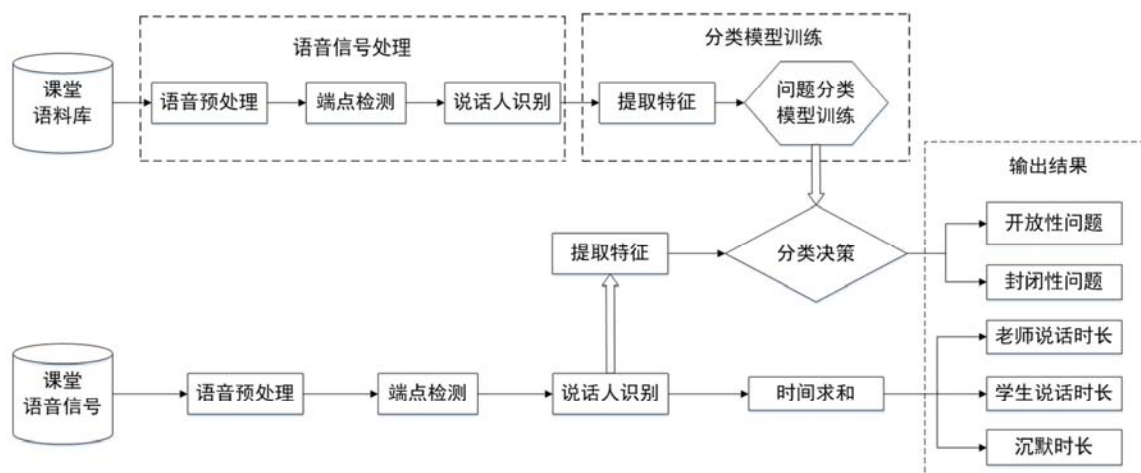


图1 师生问答自动分析的系统架构。

## 3. 基于课堂语音信号处理的说话人识别

本文师生问答自动分析系统的基础是说话人识别，只有确定话是谁说的、才能进行后续的师生说话时长统计和封闭/开放性问题的分类研究。为此，本文先采用了双门

限法端点检测方法进行课堂语音端点检测，再基于切割好的语音片段，采用Mel频率倒谱系数（Mel-Frequency Cepstrum Coefficient, MFCC）获取语音特征，使用高斯混合模型（Gaussian Mixture Model, GMM）进行说话人识别[10, 11]。

### 3.1. 课堂语音的预处理

语音信号本质上是一个非稳态、时变的信号,但语音由声门激励脉冲通过声道而形成,所以“短时间”范围内可以考虑成稳态、时不变的信号,进而可对其进行时、频域的短时分析。在分析之前,我们需要对语音信号进行一定的预处理,为后续的特征提取等提供保障。预处理过程一般包括预加重、加窗分帧等[7]。

受人们生理特点(声道、口、鼻等)的影响,信号频率在大于0.8KHz时,其平均功率谱会按照6dB/倍频程跌落,这将对信号频谱和声学特征分析造成一定影响[8]。因此,采集到的语音信号要进行预加重处理,提高高频成分,以减少高频带与低频带之间的差距,补偿信号受声带发声系统所带来的高频抑制现象。同时,语音信号在10ms至30ms之间会呈现短时平稳的特点,因此将语音信号分割成一个语音片段进行分析,并称语音片段为一帧,该过程称为分帧。为了减小分帧后截断效应的影响,需要在分帧后进行加窗处理,常用的窗函数包括矩形窗(Rectangular)、汉宁窗(Hanning)以及海明窗(Hamming)[8]。

### 3.2. 课堂语音的端点检测

所谓的端点检测是指从一段语音信号中确定出语音的起始点和结束点位置[7]。常见的语音端点检测技术有:双门限法、相关法、方差法等[8]。根据采集到的课堂语音信号中所包含的噪声种类和大小不同,我们需要采用不同的端点检测技术,这里我们只考虑采集的课堂语音信号信噪比较高即噪音微弱的情况,所以本文采用双门限法进行端点检测。双门限法端点检测技术主要是应用语音信号的时域特征参数,包括短时平均能量和短时平均过零率,其原理是韵母所包含能量较大元音,通过短时平均能量可以将其确认,而声母包含频率较高的辅音,通过短时平均过零率可以将其确认[8]。

确认了声母和韵母也就等于找到了完整的汉字字节。双门限法端点检测先根据短时能量确定一个较高门限,高于该门限阈值肯定是语音信号,由该高门限出发向下确定一个较低门限,短时能量包络与较低门限的交点初步判定为语音的起止点位置。以短时平均过零率为准再设定一个门限值,从上面初步断定的语音起止点向两侧搜索,找到平均过零率低于该门限值的端点,这便是语音段最终的起止点[8]。一段完整的课堂语音信号,经过端点检测处理后,会分割成大小不等的语音片段。至此,我们能将“说话声”和“沉默”有效的区分出来。

### 3.3. 说话人识别

所谓说话人识别是指把待测试语音和所建立的语音库里所有人进行比较,通过算法判断测试语音属于语音库中哪个人[9]。说话人识别技术分为与文本相关和与文本无关两种方式,“与文本有关”即指说话人按规定的文本发音或按提示发音,“与文本无关”指无论说话人说什么话都可进行说话人识别[9,10]。由于本文实验数据均来自于课堂教学场景,所以采用“与文本无关”说话人识别技术。说话人识别系统一般包含三部分:语音特征提取、

模型训练和识别。考虑到运算时间和识别率等方面因素,本文先选取MFCC方法获取语音特征,再使用高斯混合模型进行说话人识别[10,11]。

#### 3.3.1. MFCC语音特征提取方法

MFCC包含语音频率结构的时间变化信息,反应了声道运动的动态特征即发音方式、发音习惯等。由于该特征能很好地描述了人耳的听觉感知特性,取得了较其他特征参数更好的识别效果[9],是目前常用的特征参数之一。为了达到模拟人耳听觉感知特性的目的,在MFCC提取过程中,需要利用Mel频率尺度对实际频率进行弯折,Mel频率 $f_{mel}$ 与实际频率 $f$ 的转换公式如下:

$$f_{mel} = 2595 \lg(1 + f / 700) \quad (1)$$

MFCC特征参数提取框图如图2所示:

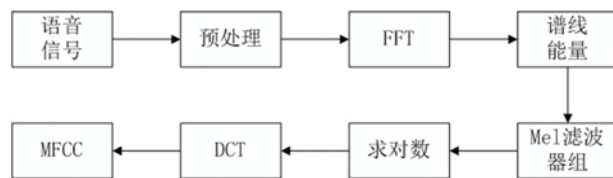


图2 MFCC特征参数提取原理框图。

MFCC倒谱频率的计算过程:(1)预处理,即将信号进行预加重、分帧、加窗函数处理。(2)快速傅里叶变换,即对每一帧信号进行FFT变换,从时域数据转变为频域数据。(3)求出谱线能量。(4)计算通过Mel滤波器的能量,即每帧谱线能量谱通过Mel滤波器组,并计算在该Mel滤波器中的能量。(5)计算DCT倒谱,即将Mel滤波器的能量取对数后计算DCT,得到MFCC系数。

#### 3.3.2. GMM说话人识别方法

GMM本质上是一种多维概率密度函数,用多个高斯分量的线性组合来表述说话人样本在声学特征空间中的分布,在目前的说话人识别研究中取得了很好的效果[12,13]。GMM的主要思想是对不同说话人的特征矢量的概率密度函数进行描述,可以对不同说话人的特点进行刻画,使系统具有更高的鲁棒性。

M阶GMM的概率密度函数如下:

$$P(X / \lambda) = \sum_{i=1}^M \omega_i b_i(X) \quad (2)$$

其中 $X$ 是一个D维语音特征向量; $M$ 表示高斯混合模型中分量的个数; $\omega_i (i = 1, 2, \dots, M)$ 是混合权重值; $b_i(X)$ 是子分布。每个子分布是D维的联合高斯概率分布,可表示为:

$$b_i(X) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) \right\} \quad (3)$$

其中 $\mu_i$ 为均值向量, $\Sigma_i$ 为协方差矩阵,混合权重满足:

$$\sum_{i=1}^M \omega_i = 1 \quad (4)$$

完整的高斯混合模型是由参数均值向量、各子分布混合权重和协方差矩阵组成，可表示为：

$$\lambda = \{\omega_i, \mu_i, \Sigma_i\}, i = 1, 2, \dots, M \quad (5)$$

对于一组长度为T的训练矢量序列  $X = (X_1, X_2, \dots, X_T)$ ，GMM的似然函数可表示为：

$$P(X / \lambda) = \prod_{i=1}^T P(X_i / \lambda) \quad (6)$$

由于上式是参数  $\lambda$  的非线性函数，常采用EM (Expectation Maximization) 算法估计参数  $\lambda$ 。说话人识别时运用贝叶斯定理，在系统中众多说话人的模型中，通过计算得到的似然概率最大的模型所对应的说话人即为识别结果：

$$i^* = \arg \max_i P(X / \lambda_i) \quad (7)$$

#### 4. 基于逻辑回归的封闭/开放性问题分类

通过说话人识别，就可以判断课堂上任意时刻是老师说话、学生说话、沉默中的哪一类。为了能进一步对师生互动进行量化分析，本文根据封闭性问题和开放性问题的特点，运用逻辑回归方法构建了一个问题分类模型，并实现了封闭/开放性问题分类。

##### 4.1. 开放性问题 and 封闭性问题

国内外众多研究表明，在中小学课堂，主要从“提问、要求、评价”来考察教师课堂教学的语言行为，其中“提问”语言行为要远远超过其他两类行为，所以“提问”行为是课堂师生交互的主要言语行为[14, 17]。按照提问类型又可以将提问分成两类：封闭性问题和开放性问题。封闭性问题容易生成一个确定的回答，回答限定在一个或少数确定性答案内。与封闭性问题相对的开放性问题，没有唯一确定的答案，很难从已经学过的知识中直接再现，需要经过逻辑推理和再创而形成答案[15, 17]。

一般来说，封闭性问题和开放问题所对应提问人数和学生回答时间长短会存在一定区别，封闭性问题等待时间相对短，开放性问题等待时间一般较长。而且从教师提问完毕到学生回答要等待3秒，如果是一个开放性问题，需要给学生全面思考和权衡的时间，15秒的等待时间也是在允许范围内[15]。所以本文根据“提问”语言行为中的提问、候答、应答、理答时间长度作为问题分类模型参数。

##### 4.2. 逻辑回归模型构建

逻辑回归 (Logistic Regression, LR) 模型是根据现有数据对分类边界线建立回归公式，以此进行分类[16]。师生交互的逻辑回归方程如下：

$$y = \omega^T \mathbf{x} = [\omega_0 \quad \omega_1 \quad \dots \quad \omega_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \quad (8)$$

其中  $y_i \in \{0, 1\}$ ，‘0’表示封闭性问题，‘1’表示开放性问题；

若  $x_i = 1$ ，则  $\omega_0$  表示回归常数；

$\omega_n$  表示第n个变量的回归系数；

$\mathbf{x}$  表示  $n \times 1$  维的特征向量，其中各参数均指时间长度。 $\mathbf{x}$  的实际物理含义如表1所示。

表1 逻辑回归模型的参数列表。

参数	实际物理含义
参数1	教师首次提问时长
参数2	教师提问候答时长
参数3	学生A回答问题时长
参数4	教师对学生A的理答时长
参数5	学生B回答问题时长
参数6	教师对学生B的理答时长
参数7	学生C回答问题时长
参数8	教师对学生C的理答时长
参数9	学生D回答问题时长
参数10	教师对学生D的理答时长

如表1所示，上述参数是从课堂师生互动过程中师生对话时长而获得，在实际课堂场景中，教师就一个问题可能提问多名同学。本文根据“央视网中国公开课”中教师对一个问题提问学生人数进行统计，发现教师一个问题提问超过4名同学情况非常少，因此得出上述模型参数。

##### 4.3. 逻辑回归模型构建

在进行参数估计中，通过将损失函数最小化的方式得到模型中的参数。使损失函数最小化的方法一般采用梯度下降法 (Gradient Descent) [16]。梯度下降的算法思路是沿着函数梯度下降的方向，直到找到全局最小值为止。令梯度记作  $\nabla$ ，则函数  $f(x, y)$  的梯度表示如下：

$$\nabla f(x, y) = \begin{Bmatrix} \frac{\partial f(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} \end{Bmatrix} \quad (9)$$

#### 5. 实验结果

##### 5.1. 说话人识别

本文采用两个语音库分别对本文使用的算法进行了验证，语音库1使用美国LDC建立的TIMIT英语语音库，选取TIMIT语音库中的168人语音数据，每个人有10句语音，采样率为8000Hz，采样精度为16bit。语音库2为本文自建的语音库，所有语音均来自于“网易公开课”视频，使用Cool Edit Pro2.1软件截取公开课中的语音，采样频率为8KHz，双声道，采样精度为16bit。自建语音库共包含200人，男130，女70人。



在实验中，均采用分帧帧长为256个采样点，帧移为100个采样点，使用hamming窗，每帧取20维的特征参数。分别选取M=4、8、16、32、64个高斯分量进行说话人识别实验。对于TIMIT语音库的168个人的语音数据，每个说话人的前7句作为训练样本，后3句作为测试样本，所有实验数据列在表2中。如表2所示，随着M的增加，识别时间也线性增加，但识别率却呈现出先增加后减小的单峰现象，即M=8时，识别率最高，达到99.40%。

表2 TIMIT语音库实验结果对比。

GMM维数	M=4	M=8	M=16	M=32	M=64
识别率	99.01%	99.40%	98.21	97.02%	81.35
时间(S)	71.69	127.12	246.23	471.46	1194.80

实验二选取自建语音库的200人的语音数据，每人录制六段共两分半语音，其中一段一分钟语音用于训练模型，五段各十五秒语音用于测试，所有实验数据列在表3中。如表3所示，随着M的增加，识别时间同样线性增加，识别率也呈现出先增加后减小的单峰现象，即M=8时，识别率最高，达到99.40%。

表3 自建语音库实验结果对比。

GMM维数	M=4	M=8	M=16	M=32	M=64
识别率	99.01%	99.40%	98.21	97.02%	81.35
时间(S)	71.69	127.12	246.23	471.46	1194.80

综合表2和表3的实验结果，可以看出GMM模型的高斯分量个数M选择8比较合适，可以在保证识别率的情况下，减小算法的时间复杂度。

5.2. 基于说话人识别的课堂师生说话时长统计

在师生互动量化分析中，教师说话时长、学生说话时长、沉默时长是三项比较重要的指标，通过端点检测和说话人识别等技术可以得到一堂课中上述三项指标。为了使实验更加贴近实际课堂，本文选取“央视网中国公开课”中两课例进行实验。

在实验中，先测出说话人说话的总时长 $T_{test}$ ，再结合实际说话总时长 $T_{true}$ ，就可以计算出说话时长相对误差，计算公式如下所示：

$$\delta = \frac{|T_{test} - T_{true}|}{T_{true}} \times 100\% \tag{10}$$

总说话人识别相对误差计算公式如下所示：

$$\delta_{\text{总}} = \frac{\sum_{i=1}^3 |T_{\text{test}_i} - T_{\text{true}_i}|}{\sum_{i=1}^3 T_{\text{true}_i}} \times 100\% \tag{11}$$

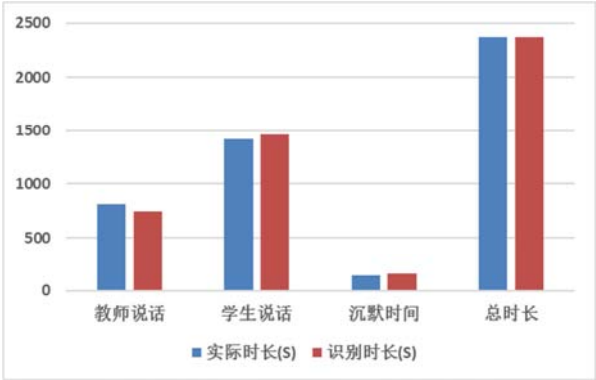


图3 课例A说话人识别结果。

如图3所示，对于授课时间为2372秒的课例A，可以测得：(1)教师实际说话时长为812秒，识别时长为750秒，相对误差为7.64%；(2)学生实际说话时长为1419秒，识别时长为1466秒，相对误差为3.31%；(3)课堂沉默真实时长为141秒，识别时长为156秒，相对误差为10.64%。(4)总相对误差为5.28%。

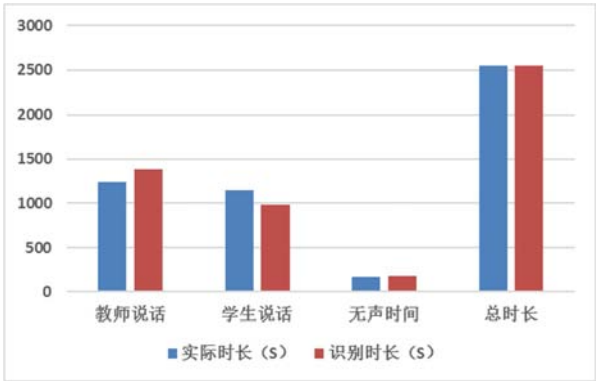


图4 课例B说话人识别结果。

如图4所示，对于授课时间为2372秒的课例B，可以测得：(1)教师实际说话时长为1239秒，识别时长为1388秒，相对误差为11.86%；(2)学生实际说话时长为1150秒，识别时长为986秒，相对误差为14.26%；(3)课堂沉默真实时长为165秒，识别时长为180秒，相对误差为9.09%。(4)总相对误差为12.76%。

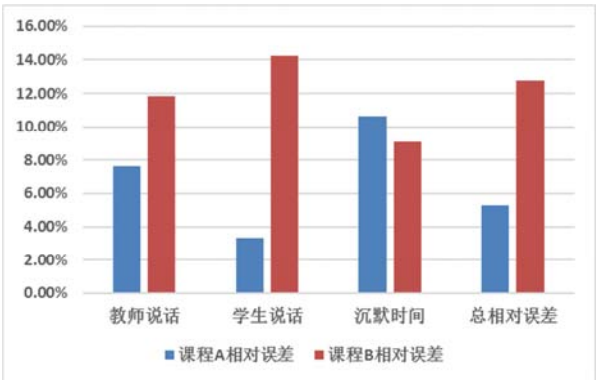


图5 课例B说话人识别结果。

如图5表示,在课程B中相对误差要比课程A的误差更大一些,由于在实际场景中采集的语音信号受周围环境和教室回声的干扰较大,所以在教师和学生身份识别上存在一定误差,但相对误差都限制在15%以内。

### 5.3. 封闭/开放性问题的逻辑回归分类

为了使数据具有说服力,本文从“央视网中国公开课”选取6位不同教师所授的6段教学视频,采用自动标注和手动标注交叉验证的方式获取师生对话数据,进行封闭/开放性问题的分类系统训练和测试。六位教师在提问类型统计图如下所示。

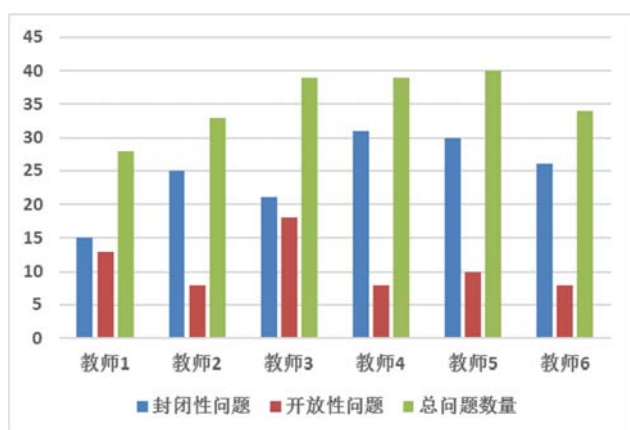


图6 封闭/开放性问题统计。

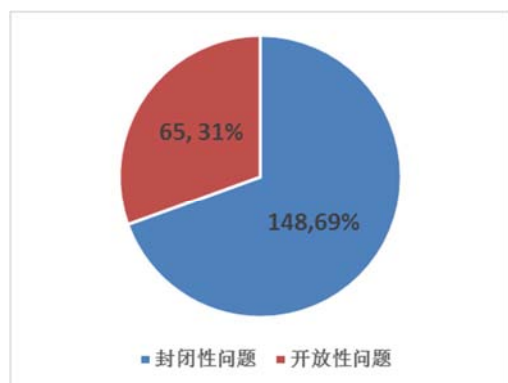


图7 封闭/开放性问题比例。

在图6、图7课例的实验数据中可以看出,一堂课的时间内、不同教师提出问题总数量、封闭性问题和开放性问题所占的比例都存在一定差异。但整体趋势是封闭性问题数量大于开放性问题数量。本文将封闭性问题和开放性问题进行汇总,共得到开放性问题数据65条,封闭性问题数据148条,随机从两类问题中各抽取50条数据(共100条数据)进行封闭/开放性问题分类模型训练。对剩余数据中抽取封闭/开发性问题各15条进行测试,实验结果如表4所示。

表4 封闭/开放性问题识别结果。

	开发性问题	封闭性问题	封闭/封闭性问题
识别率	98.31	88.67	93.49

## 6. 结论和展望

本文基于语音端点检测和说话人识别技术对课堂语音进行分析处理,并使用逻辑回归方法设计一个问题分类模型,最终构建了一套用于课堂评价的师生问答自动分析系统。系统输出数据为:教师说话总时长、学生说话总时长、教师提出开放性问题数量、封闭性问题数量。实验结果表明本文构建的师生问答自动分析系统具有满意的精度,能够从语音处理的角度对师生互动进行全自动评价,可用于日常的课堂评价与课堂管理。

为了将本文系统应用于实际课堂的师生互动分析,未来将针对教室回音在内的诸多噪音、使用多麦克去噪技术进一步增强本文系统的抗干扰能力。

## 致谢

本文得到了国家自然科学基金面上项目(61274033, 61271198)联合资助。

## 参考文献

- [1] 钟启泉. 课堂评价的挑战[J]. 全球教育展望, 2012, (01):10-16.
- [2] 崔允漷. 论课堂观察LICC范式:一种专业的听评课[J]. 教育研究, 2012, (05):79-83.
- [3] 郭绍青, 张绒, 马彦龙. “有效教学”课堂录像分析方法与工具研究[J]. 电化教育研究, 2013, (01):68-72.
- [4] 时丽莉. “弗兰德互动分析系统”在课堂教学中的应用[J]. 首都师范大学学报(社会科学版), 2004, (S2):163-165.
- [5] 方海光, 高辰柱, 陈佳. 改进型弗兰德互动分析系统及其应用[J]. 中国电化教育, 2012, (10):109-113.
- [6] 顾小清, 王伟. 支持教师专业发展的课堂分析技术新探索[J]. 中国电化教育, 2004, (07):18-21.
- [7] Lu Z, Liu B, Shen L. Speech endpoint detection in strong noisy environment based on the Hilbert-Huang Transform [C]// International Conference on Mechatronics and Automation. IEEE, 2009: 4322-4326.
- [8] 宋知用. MATLAB在语音信号分析与合成中的应用[M]. 北京航空航天大学出版社, 2013:116-154.
- [9] 熊华乔, 郑建彬等. 基于说话人模型聚类的说话人识别[J]. 计算机工程与应用. 2014, 50(02):133-137.
- [10] Zufeng Weng, Lin Li, Donghui Guo. Speaker Recognition Using Weighted Dynamic MFCC Based on GMM [C]. Anti-Counterfeiting Security and Identification in Communication (ASID), 2010(07):18-20.
- [11] 胡政权, 曾毓敏, 宗原, 等. 说话人识别中MFCC参数提取的改进[J]. 计算机工程与应用. 2014, 50(07):217-220.

- [12] Xinxing Ling, Ling Zhan, Hong Zhao, Ping Zhou. Speaker Recognition System Using the Improved GMM-based Clustering Algorithm [C]. Intelligent Computing and Integrated Systems (ICISS), 2010(10):22-24.
- [13] 赵力. 语音信号处理[M]. 机械工业出版社, 2016:228-230。
- [14] 叶立军, 周芳丽. 基于录像分析背景下的教师提问方式研究[J]. 教育理论与实践, 2012, (05):52-54。
- [15] 黄伟. 提问与对话:有效教学的入口与路径[M]. 浙江大学出版社, 2016:73-92。
- [16] PeterHarrington, 李锐等. 机器学习实战[M]. 人民邮电出版社, 2013:73-88。
- [17] Thomson R P. Strategies for effective teaching. [J]. Nln Publ, 1975(16-1538):12-24.