doi:10.16104/j.issn.1673-1891.2020.04.014

基于深度学习的智能语音问答系统研究

董 钰[1],郭军华[2]
(淄博师范高等专科学校,山东 淄博 255130)

摘要:随着人工智能行业的不断发展,智能语音问答技术逐步得到国内外学者的广泛关注和研究,但是语音识别方面仍然存在两个技术瓶颈,第一是语音识别系统,第二是根据识别的语音进行问题的回答。基于此,开展了基于深度学习的智能语音问答系统研究。首先介绍了基于隐马尔科夫模型的语音识别系统,然后研究了基于梅尔频率的语音信号特征提取技术,并建立了声学和语言模型,最后研究了基于GRU算法的问答匹配模型,并基于以上模型开发了智能语音问答系统。经实际实验验证分析,文章所提出的算法在语音识别和问答的准确度方面都相比传统算法具有很高的精确度,本算法具有较大的实用价值。

关键词:深度学习;智能语音;问答
中图分类号:TN912.34;TP181 **文献标志码**:A **文章编号**:1673-1891(2020)04-0058-04

Research on Question-Answering System with Intelligent Voice Based on Deep Learning

DONG Yu[1], GUO Junhua[2]
(Zibo Normal College, Zibo, Shandong 255130, China)

Abstract: With the continuous development of artificial intelligence industry, the technology of question-answering intelligent voice has been widely concerned and studied by scholars at home and abroad. However, there are two technical bottlenecks in speech recognition. The first lies in the speech recognition system, and the second lies in answering to the recognized voice. In response, this paper develops a question answering system with intelligent voice based on in-depth learning. This paper introduces the speech recognition system based on Hidden Markov Model at first , then studies the extraction technology of speech signal features based on Meier frequency, establishes some acoustic and linguistic models, and finally studies the question-answer matching model based on GRU algorithm. A question answering system has been developed as a result of above study. The experimental results show that the proposed algorithm, with considerable practical value, has higher accuracy in speech recognition and question answering than the traditional algorithm.

Keywords: deep learning; intelligent voice; question and answer

0 引言

近年来,随着计算机信息技术和人工智能技术的不断发展,智能机器人等智能语音问答系统应用逐渐增多。由于智能语音系统应用范围较广,并且可以很大程度上实现减员、降本增效的目的,因此被国内外学者广泛研究。智能语音问答系统中需要解决的关键技术主要有两点,其一是语音识别系统,其二是根据识别的语音进行问题的回答。只有以上两个关键技术得到解决,才能构建出一套比较完善的智能语音问答系统。归根结底,智能语音问答系统的效率和精度,需要强大的人工智能技术支撑[1-4]。例如人工神经网络已与上世纪80年代逐步被应用至语音识别领域,并且对其采用BP方法进行训练,其语音识别的精度得到很大程度提升。但是随着语言的不断丰富和研究的不断完善,传统的神经网络已经无法满足实际需求,其各种弊端也逐步凸显[5-7]。因此,以深度神经网络为代表的其他声学模型得到广泛的研究和应用,极大的提升了语音识别效果。以苹果等公司为例,其在语音识别研究上投入了巨大的人力和物力,并且成绩不菲。文章首先研究了基于隐马尔科夫模型的智能语音识别系统,然后研究了基于卷积神经网络算法的问答系统。通过实验数据证明,文章所做研究在语音识别

收稿日期:2020-03-17
基金项目:山东省社会科学规划研究项目(17CXWJ05)。
作者简介:董钰(1981—),男,山东淄博人,讲师,硕士,研究方向:计算机网络管理。

的精确度和问答回复的准确度方面相较其他算法具相对较高的精度。本文最大的创新点在于,改造了传统的声学模型,引入隐含层的非线性激活函数取代了以往DNN-HMM模型中的sigmod为基础的非线性激活函数,并将基于梅尔频率的语音信号特征提取技术、声学模型、语言模型、基于GRU算法的匹配模型和系统开发有机结合,形成了一套完备的智能语音问答系统。

1 基于隐马尔科夫模型的语音识别系统

隐马尔科夫模型的基本形式可以采用式(1)表示。

$$\gamma=\{N,M,\pi,A,B\} \tag{1}$$

式1中N代表检测到的语音状态个数;M代表可被检测的具体符号数量;A为矩阵,矩阵中的每个元素代表某时间间隔内状态转化发生的基本概率。每个元素的计算公式如式(2)。

$$a_{ij}=P[q_{t+1}=S_j|q_t=S_j] \tag{2}$$

式(2)中 $i/j\in[1,N]$, $\sum_{j=1}^{N}a_{ij}=1$,且矩阵中每个元素的值为非负。B表示可以被检测到的状态分布概率。在具体分析过程中可以根据被观测变量的不同将模型分为连续和离散两类。图1代表隐马尔科夫链的基本结构,其下一节点出现状态转移的几率和此节点所在的位置关系较大,不会受到其历史位置的关联影响[8]。

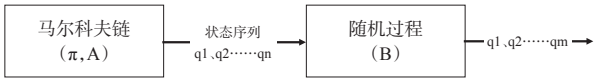

图1　隐马尔科夫链

采用此模型需要经历三个基本流程,分别为评估流程、译码流程、训练流程三个。评估流程主要实现基于前向算法的模型评估,译码流程主要实现基于Viterbi算法的最佳路径寻找,训练流程主要是根据已知序列,调整设置参数,使得输出最大化,一般使用Baum-Welch算法来实现。

2 基于深度学习技术的语音识别系统

2.1 基于梅尔频率的语音信号特征提取技术

特定的参数对语音识别系统的精度有很大影响,所以在建立语音识别系统时应该选择与之匹配的听觉感知参数。一般来说,使用频率较高的参数主要有梅尔频谱系数和线性预测倒普系数。由于线性预测倒普系数相比于梅尔频谱系数在语音识别质量上容易受到噪声的污染和影响,对输出信号的精确度达不到问答系统的要求,因此文章选择了梅尔频谱系数作为语音信号特征提取的主要方法[9-10](图2)。

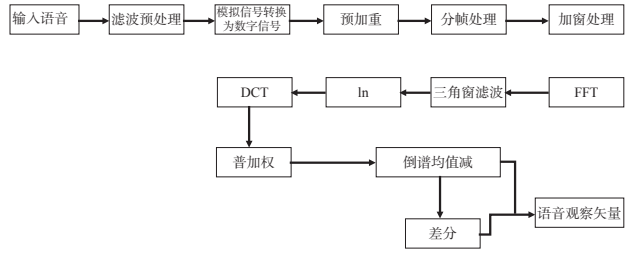

图2　基于梅尔频谱系数进行语音特征识别的一般过程

2.2 建立声学模型

常用的GMM-HMM模型,其优势明显,主要为系统构成简单,但是当算法面对大规模的语音数据处理时,其语音降噪工作主要体现在特征处理上,因此处理效率并不高,程序的鲁棒性比较差。与之相比,DNN-HMM算法其基本结构是使用DNN代表原来的GMM,此深度学习算法可以对海量的语音数据进行有效的数学建模,并通过底层的网络将噪声去掉,然而在高层网络中可以把语音特征中比较具有区分性的特征识别出来,并将其保存。此种方式,很大程度上增强了系统程序的鲁棒性,在准确率等方面相比于GMM-HMM算法提升了近两成。

本文在传统DNN-HMM算法的基础上,又对其进行改造,引入隐含层的非线性激活函数取代了以往DNN-HMM模型中的sigmod为基础的非线性激活函数,采用了ReLUs算法实现非线性激活。其声学模型如图3所示。

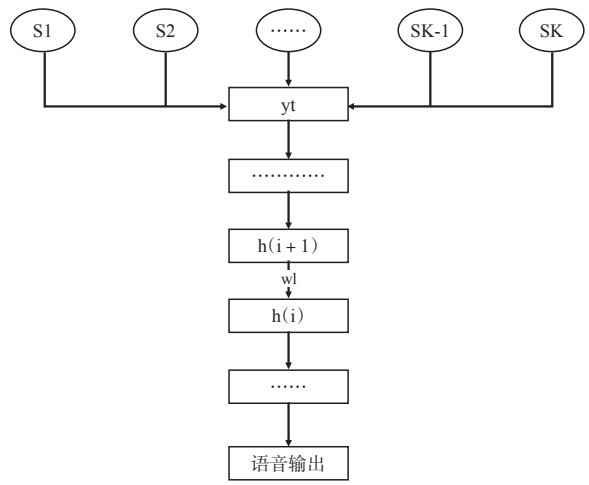

图3　基于DNN–HMM算法的声学模型图

从数学建模层面来讲,其模型可以利用式(3)~(5)表示。

$$h^0=x(1) \tag{3}$$

$$h^l=f(W^l h^{l-1}+b^l) \qquad 1\leqslant l\leqslant L \tag{4}$$

$$y=\mathrm{Softmax}(W^{L+1}h^L+b^{L+1}) \tag{5}$$

在以上式中,x代表输入语音序列的声学特征,W^l和b^l分别代表在某一层中的连接权重和偏量。函数f代表隐含层的非线性激活函数,此函数取代了以往DNN-HMM模型中的sigmod为基础的非线性激活函数,而是采用了ReLUs算法来进行,其表达式如(6)所示。

$$\text{ReLUs:}f(x)=\max(x,0) \tag{6}$$

Softmax()函数表示某个单元对应的后检验发生概率。此函数表征输出和对应标注之间的差异度,是模型性能优劣的典型描述特征。实验选择了式(7)作为目标优化函数,此函数可以描述输入和输出之间的匹配度,其匹配度越高则说明模型的性能越强。

$$F_{CE}(W)=-\sum_{r=1}^{N}\sum_{t=1}^{T}\log y_n(S_n) \tag{7}$$

在式(7)中,t代表语音中第r个句子在状态s下对应的实际$y_n(S_n)$。

2.3 建立语言模型

作为自然语言处理范畴的基础性问题,语言问题是指对客观的语音信息进行抽象,并对其进行数学建模,表达字和词之间的基本含义。此模型可以动态的检测是否存在语音上的边界,从而区分不同字词的界限,模型可以充分反映词句之间的联系和基本语义。在实际语义过程中一般使用最多的为三音子模型。此模型综合考虑了前后音各个声音元素的关系,充分使用上下文的基本信息,从而对语音识别的效率有了很大程度的提升,文章首先采用基于决策树的方法对三音子模型进行聚类分析,被聚类的各个模型均可以进行独自的训练,其训练数据和算法参数是均可以被共享的,在对算法进行上下分裂的过程中,在输入一定的语音学知识库后,一方面可以输出从未使用过的三音子模型,从而丰富知识库,另一方面运算效率可以得到成倍提升。

2.4 基于深度学习的语音识别系统建立

基于以上研究,文章建立了基于深度学习的语音识别系统。在特征提取模型、声学模型和语音模型建立完成后,加之解码器和语音数据库作为知识输入,可以输出语音信息对应的文本信息,其基本处理流程如图4所示。详细步骤如下:

1)对原始声音进行加重、加窗和相关帧处理;

2)对帧数据,基于梅尔频率倒谱系数进行特征识别处理,可以得到语音序列对应的特征矩阵,其中r代表序列的帧总数;

3)基于卷积神经网络算法实现前向计算,最终获得特征识别的输出矩阵;

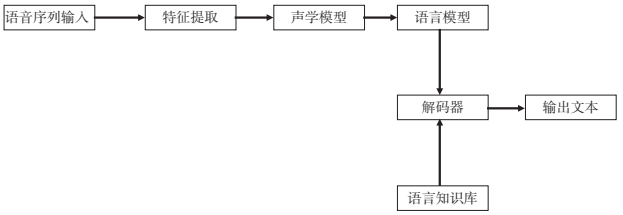

图4 基于深度学习的语音识别系统的基本流程

4)基于上述矩阵,计算其中的最大输出因素概率值,以此形成马尔科夫链,其中每个元素表示每一帧声音的基本元素;

5)对4)步形成的模型进行分析,确定概率最高的声音路径,从而得到最终的文本信息,并对文本信息进行输出。

2.5 基于卷积神经网络的问答匹配模型

如今,卷积神经网络被广泛应用于长度变化的文本序列识别中,由于其可以根据句子的语句序列特征识别语句内容,其处理模式是采用记忆单元的模式。现常用来解决循环卷积神经网络问答问题的卷积神经网络算法主要有两类,一类是LSTM算法,另一类是GRU算法。两类算法各有优劣,但是相对于LSTM算法来讲,GRU算法在收敛速度和计算效率等方面都具有优势,这是由于其内部隐藏单元内缺少一个控制门的原因。鉴于此文章使用GRU算法来实现问答匹配。基于GRU算法的匹配模型详细如下。

在时刻t内,经基于深度学习的语音识别系统得到的句子输入如式(8)所示。

$$X=\{x_1,x_2,\cdots,x_T\} \tag{8}$$

其上一时刻隐藏层输出假设为ht-1,此时GRU内部单元详细情况如下。

遗忘门如式(9),更新门如式(10),时刻t的内部状态如式(11),时刻t模型的输出输入式(12)。

$$f_t=\sigma(W_{xf}x_t+U_{hf}h_{t-1}) \tag{9}$$

$$z_t=\sigma(W_{xz}x_t+U_{hz}h_{t-1}) \tag{10}$$

$$h'_t=\tanh(W_{xz}x_t+U_{hh}(f_t\odot h_{t-1})) \tag{11}$$

$$h_t=(1-z_t\odot h_{t-1}+z_t\odot h_t) \tag{12}$$

上式中函数σ称作激活函数,$\odot$称作各元素叉乘,U和W分别代表参数矩阵。鉴于算法在某一固定时刻不能对之前和之后的语义信息进行学习,因此文章采用将输入的语音序列从正反两个方向分别输入的方式进行,将两个方向在同一时刻的输出层作为一个新的输出向量,从而便可构造出一个满足双向运算且包含更丰富信息的模型,此模型对于输入语句的特征表示更加丰富,模型的基本情况如图5所示。

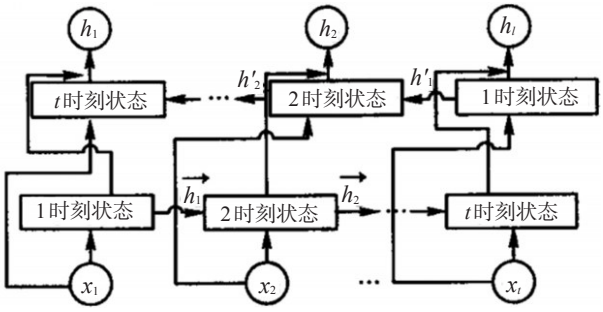

图5　双向运算的GRU算法的匹配模型

为了获取输入的语音序列更深层次的音素,将此模型的输出序列作为输入元素输出到另一卷积层中,卷积层中将语音序列串接为一个新的向量,并且通过特定函数将其映射为全新向量。此环节中卷积窗口的大小设置为k,从而得到矩阵S,卷积核数目假设为n,卷积操作的具体模型如式(13)所示。

$$G=f(W_{gn}S+b) \tag{13}$$

上式中,为了加速模型的收敛采用了激活函数f,W_{gn}和b分别代表权重和偏差,其遵循均匀分布原则。与传统的卷积算法不同,本算法在每一次卷积操作中都可以产生一个新的n-gram特征。同时文章为了防止过拟合现象产生,在算法执行过程中采用了dropout方法,整个算法模型如图6所示。

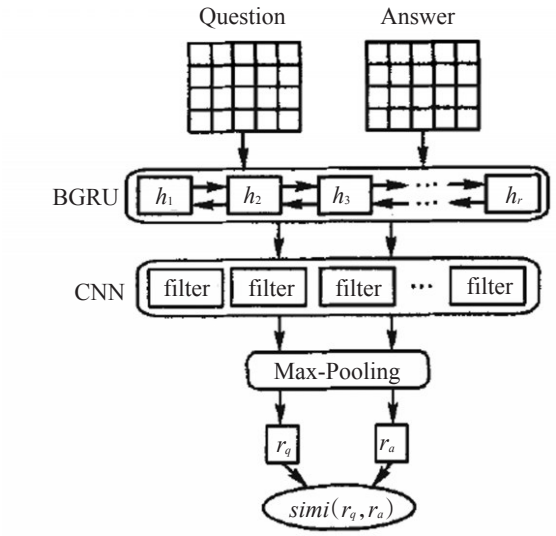

图6　基于卷积神经网络的问答匹配模型

3 实验及结论分析

实验在实验室环境下进行,语音类型主要有连续单词、孤立单词、噪声干扰和方言等,以上述类型语音作为输入,并采用16 kHz的语音采样频率,每一帧长度为30 ms,声音为单声道。为了验证文章所提出的算法的效率,文章将传统DNN-HMM和本文提出的DNN-HMM算法对照实验,对各类语音输入类型进行了识别,并用声音识别的准确率作为评判标准。其实验结果如表1所示。

表1　对比实验结果分析表　　%

语音类型	传统DNN-HMM	本文DNN-HMM
连续语音类型	77	88
单个单词	84	98
噪声干扰	65	75
方言	66	77

从表1可知,文章所提出算法在各类语音类型的识别准确率方面都比传统DNN-HMM算法高出近10个百分点。识别的准确率有了很大程度的提升,为语音问答的准确性打下了很好的输入基础。

在问答方面,同样选择了准确率作为评判标准,实验数据来源于现今的NPLPCC2016 DBQA Task数据库。采用文章所提算法,并与传统的模型相比较,最终实验结果如表2所示。

表2　对比实验结果分析表

算法	常用模型	准确率/%
基于人工构造特征算法	TF	45.44
	Edit	21.05
双向长短期记忆算法	Word overlap	51.55
	All features	81.80
本文算法	AR-BGRU	78.81
	AP-BGRU	85.82

从表2可知,在输入相同的情况下,语音问答的准确率,本文算法相比于其他算法具有很大的优势,更加验证了文章所提出算法的准确性和可行性。

针对智能语音问答业务中存在的关键问题,文章提出了基于隐马尔科夫模型和卷积神经网络的智能语音问答系统,首先利用隐马尔科夫模型解决了语音识别问题,然后建立了基于卷积神经网络的语音问答系统,经过实验验证所提算法在语音识别准确度和语音问答的准确度方面,相比其他传统算法都有很高的优势。

参考文献:

[1] 张建明,詹智财,成科扬,等.深度学习的研究与发展[J].江苏大学学报(自然科学版),2015,36(2):191-200.

[2] 李佳,刘振宇. SVM与BP神经网络在石煤提钒行业清洁生产评价中的对比研究[J].中南民族大学学报(自然科学版).2018(4):18-21.

(下转第81页)

自信心是大学生健康成长与能力发展的原动力,缺乏自信的大学生,往往会导致丧失发展的内在动力,而有自信心的大学生,在学习过程中必然有更多的主动性,良好的自信心是大学生自主学习能力的保障[11]。自信心是决定学生能否产生自主学习行为动机和促使自主学习行为发生的重要因素,自信的学生,能够乐观处理和应对学习活动中面临的各种机遇、威胁和挑战,进而更好地获得自身的发展。在设计的深度混合教学中,鼓励学生发挥自身的主观能动性参与到教学互动中来,借助各种资源实现学习目标,从而为自主学习能力的培养提供保障。

4.4 教学实施中的问题与对策

以结构主义教学论为基础的混合教学实施中,存在以下3方面问题,一是部分学生对混合式教学模式的认识不够,认为课外线上学习占用了自己的课余时间,甚至有的学生不知道如何进行线上学习,在线学习主动性和积极性不高,这就要求教师课前分析学生现状,课程设计要避免过多占用课余时间,帮助学生熟悉线上学习的策略。第二是教师对于线上教学改革的热情不高、混合深度不够,面对当前互联网+教育形势,教育信息化发展已是大势所趋,这对于高校教师的专业化也提出更高要求,从课前准备、课中学习、课后反馈等多个维度混合。第三是线上线下、课内课外内容衔接不紧密,虽然国内混合式教学模式各不相同,但不是简单的线下内容搬到线上,要求教师重新设计课程结构,真正体现"教与学"的深度混合。

参考文献:

[1] CHIGERWE M,BOUDREAUX K A,ILKIW J E.Self-directed learning in veterinary medicine:are the students ready [J]. International Journal of Medical Education,2017(8):229-230.

[2] 刘玉萍,梁瑞芳,苏旭.大学生自主学习影响因素分析及其提升对策[J].青海师范大学学报(自然科学版),2019,35(2):85-88.

[3] 楚东杰.结构主义教学论探析[J].教育教学论坛,2016(19):179-180.

[4] 蔡静.结构主义学习理论下的课堂教学设计——以《学前教育学》课程为例[J].教育现代化,2017,4(40):262-264.

[5] 张伟,李化树.基于结构主义教学理论的高中生英语自主学习能力的培养策略[J].文史博览(理论),2012(8):81-83.

[6] MA Xingming, LUO Yanping, ZHANG Lifeng, et al. A trial and perceptions assessment of APP-based flipped classroom teaching model for medical students in learning immunology in China [J].Education Sciences,2018,8(2):45.

[7] PAONAN C. The development of a measurement tool to assess Chinese engineering students' self-directed learning abilities [J]. Global Journal of Engineering Education,2012,14(2):196-199.

[8] LI C, HE J, YUAN C, Chen B, *et al*. The effects of blended learning on knowledge, skills, and satisfaction in nursing students: a meta-analysis [J]. Nurse Educ Today, 2019,82:51-57.

[9] SÁIZ-MANZANARES M C, ESCOLAR-LLAMAZARES M, GONZ ÁLEZ Á A. Effectiveness of blended learning in nursing education [J]. International Journal of Environmental Research and Public Health,2020,17(5):E1589.

[10] 李浩光.浅谈数学教师在大学生自学能力的培养中的定位作用[J].科教文汇(中旬刊),2018(4):70-71.

[11] 迟宝策.刍议大学生的自信心的培养方式[J].辽宁师专学报(社会科学版),2019(3):93-95.

(上接第61页)

[3] 毛博,徐恪,金跃辉,等.Deep Home:一种基于深度学习的智能家居管控模型[J].计算机学报.2018,41(12):2689-2701.

[4] 王毅,谢娟,成颖.结合LSTM和CNN混合架构的深度神经网络语言模型[J].情报学报, 2018,37(2):194-205.

[5] 李志义, 黄子风, 许晓绵.基于表示学习的跨模态检索模型与特征抽取研究综述[J].情报学报,2018,37(4):86-99.

[6] 吴飞,韩亚洪,廖彬兵,等.多媒体技术研究:2017——记忆驱动的媒体学习与创意[J].中国图象图形学报, 2018,23(11): 1617-1634.

[7] 柯登峰, 俞栋, 贾珈.语音图文信息处理中的深度学习方法进展专刊序言[J].自动化学报, 2016, 42(6):805-806.

[8] 王斌,范冬林.深度学习在遥感影像分类与识别中的研究进展综述[J].测绘通报, 2019(2) :99-102+136.

[9] 王万良,张兆娟,高楠,等.基于人工智能技术的大数据分析方法研究进展[J].计算机集成制造系统, 2019,25(3):529-547.

[10] 刘彪,黄蓉蓉,林和.基于卷积神经网络的盲文音乐识别研究[J].智能系统学报,2019, 14(1):186-193.