

# TensorFlow 在大学英语语言训练中的应用

孙雅琳, 文福安

(北京邮电大学 网络教育学院, 北京 100876)

**摘 要:**北京邮电大学人文学院每学期都利用通用计算机化考试平台(大学英语语言技能训练系统)来进行英语测试,其中客观题系统能够对照答案直接给出分数,但是主观题只能依靠英语教师来逐个批改。本研究是利用 tensorflow 深度学习平台对英语口语表达题进行智能批改,取音素后验概率、语速 ROS、关键词覆盖率、文本覆盖率、用词变化程度等维度作为特征,学生考试成绩为目标,用 tensorflow 深度学习平台进行模型训练,实现学生成绩的预测。

**关键词:** tensorflow; 深度学习; 大学英语; 神经网络

**中图分类号:** TP391.42 **文献标识码:** A **DOI:** 10.3969/j.issn.1003-6970.2017.08.027

**本文著录格式:** 孙雅琳, 文福安. TensorFlow 在大学英语语言训练中的应用 [J]. 软件, 2017, 38 (8): 142-144

## The Application of TensorFlow in College English Language Training

SUN Ya-lin, WEN Fu-an

(Beijing University of Posts and Telecommunications, School Of Network Education, Beijing 100876, China)

**【Abstract】:** Beijing University of Posts and Telecommunications College of Humanities each year using the general computerized examination platform (College English language skills training system) for English testing, in which the objective system can directly give the answer to the score. But the subjective questions can only rely on English teachers to change one by one. In this study, we use the tensorflow, deep learning platform, to intelligently correct the oral expression of English and take the dimension of phoneme posteriori probability, speed rate ROS, keyword coverage, text coverage, degree of change of words, and so on. With tensorflow deep learning platform for model training to achieve student achievement prediction.

**【Key words】:** Tensorflow; Deep learning; College english; Neural network

## 0 引言

国内的批改系统主要有以下几个方面:电路类虚拟实验自动批改与智能指导系统、智能导师系统、英语作文智能评阅系统、英语口语智能评阅系统。对于正确答案有一定规则的问题可以利用智能导师系统进行评阅,像用英语口语围绕主题阐述一定观点这种开放性题目,现有的系统不能评阅这种题型。所以,在语言训练中,智能批改仍有一定局限性<sup>[1-6]</sup>。

国外研究现状与国内类似,对于口语口语表达题目等答案固定或者具有一定规则的题目评分有比较成熟的机器学习模型,而对于答案规则不明显的题型目前还不能利用深度学习模型来进行智能批改。在 2000 年本领域的一篇经典论文就列举了人工

老师打分情况,他提出的观点是使用“平均”带来打分稳定性和准确度提高。多个句子的平均,可以更好的对整体水平进行打分,句子越多,打分越准确<sup>[7]</sup>。瑞士 IDIAP 在 2016 年也提出了更新和复杂的深度学习技术来进行评分。

## 1 人工智能

近几年,人工智能是计算机领域中的一个热门话题,它的应用非常广泛,可以利用机器学习算法解决一些传统计算机无法解决的复杂问题。而深度学习可以看做是机器学习的另一次升华,深度学习可以解决传统机器学习算法无法解决的问题,大大地提高了人工智能的发展速度。

作者简介:孙雅琳(1992-),女,硕士研究生,主要研究方向:教育软件工程;文福安(1962-),男,教授,主要研究方向:计算机科学与应用。  
通讯联系人:文福安(1962-),教授,主要研究方向:虚拟仿真技术。

一些深度学习平台有 tensorflow、Caffe、Deeplearning4j、CNTK、MXNet、Theano 等。Tensorflow 作为是最主流和应用最广泛的深度学习平台之一。

**Caffe:** 一款集表达、速度和模块化思想于一身的深度学习框架。其图像处理优势明显, 只用一个单一的 NVIDIA K40 的 GPU 就可以每天处理 60M 图像数据。**Deeplearning4j:** Deeplearning4j 是 java 科学计算引擎驱动的矩阵操作。**CNTK:** 是一个描述神经网络的深度学习工具。**MXNet:** MXNet 能够运行在 CPU/ GPU、集群、服务器、手机上。支持分布式计算, 也能够 AWS, GCE, Azure 和 Yarn 集群上运行。**Theano:** 一个可以定义、优化、评估多维数组的 python 库。**Torch:** 有强大的 N 维数组, 能够索引、切片、换位, 快速高效的 GPU 支持。**H2O:** H2O 可以与最流行的开源产品 Apache Hadoop 和 Spark 结合来给客户灵活地解决具有挑战性的数据问题的机器学习平台, 可以进行大规模可

扩展的数据分析。**OpenNN:** opennn 是一个用 C++ 语言编写的实现神经网络的开源类库, 机器学习研究的主要领域。用于实现监督学习的任意数量的非线性处理单元。它的存储速度和处理速度都很高。

**OpenCyc:** OpenCyc 是 Cyc 的开源版本, Cyc 是世界上最庞大的、最完备的通用知识库与常识推理引擎。**OpenCyc** 是对人工智能与语义万维网科学研究的贡献, 在人工智能、语义万维网及其应用领域具有很广泛的应用价值。

## 2 TensorFlow 平台

早在 2011 年, 谷歌内部就开发了一个深度学习工具——DistBelief, 这个工具只在谷歌内部使用。之后, 由 Jeff Dean 带领的谷歌大脑团队将 DistBelief 框架进行改进, 终于在 2011 年发布了一款基于 Apache 2.0 的开源通用计算框架。它是一个可以支持大规模数据计算的机器学习框架, 工作流程如下<sup>[8]</sup>:



图 1 tensorflow 的工作流程  
Fig1. Tensorflow workflow

目前 Tensorflow 的应用有如下几个方面: 图像识别、语言识别、自然语言处理与人机博弈等。像最近的人机博弈大战, 例如李世石和 DeepMind alphago 的围棋之战、腾讯 Master 围棋大战、百度大脑和人类大脑的对战都无疑是机器略胜一筹。其实早在 1997 年, IBM 的 deep blue 就打败了国际象棋冠军 Garry Kasparov。这二十年里, 机器学习技术的发展是巨大的, 不仅能掌握国际象棋的规则, 还能通过大量的数据运用深度学习方法训练出解决无规则复杂问题的程序。这对于人工智能的发展来说意义重大, 因为现实生活中, 很多问题都是无规则的。比如, 自动驾驶、人机对话等。让机器通过数据学习像人一样处理和解决问题的能力, 这是飞跃性的进步。目前人工智能还不能够理解语义含义, 不能理解人类语言中的情绪, 但是我相信在不久的将来, 人工智能通过大量的数据和深度学习框架一定能够解决这些问题。

## 3 评分系统

### 3.1 评分系统工作原理

评分系统主要分为识别端和评分端:

(1) 识别端: 只有通过识别, 计算机才能“听懂”发音人的发音内容。首先对于当前考生, 基于考生口语表达题的数据, 通过科大讯飞语音识别接口, 生成口语表达题语音所对应的词图来表示识别结果<sup>[9]</sup>。

(2) 评分端: 首先基于词图提取多维与专家人工评分标准相关的机器评分特征, 其过程可以利用范文本身及聘请专家手工标注的本篇范文在被口语表达中应该提及的一些关键词列表。然后, 将各维评分特征输入训练好的评分模型, 得到的口语表达题自动评分结果<sup>[10]</sup>。

### 3.2 评分特征选取

通过对分类、回归、聚类等深度学习算法的研究, 提取出评分系统评分特征, 包括音素后验概率、语速、吸收模型占有声段比例、关键词覆盖率、文本覆盖率等。利用现有的通用计算机化考试平台收集到的学生考试数据以及评分特征, 对智能批改模型进行多次训练, 最终利用机器来批阅批改英语口语表达题。通过机器评分和教师评分的相关度来验证机器评分的准确性。与人工评分相关度越高, 机器评分准确度越高<sup>[11-13]</sup>。

### (1) 音素后验概率

PPP(Phone Posterior Probability)发音的标准程度是评价一个英语口语水平的重要方面, 提取了音素后验概率这样一维评分特征。采用的声学模型是基于标准发音数据训练得到的标准发音模型。

### (2) 语速 ROS(Rate of Speech)

语速是评判说英语流利程度的一个比较好的特征, 语速越快而且中间没有停顿犹豫说明英语说得很流利。相反, 语速很慢说明思考时间长而且不流利。

### (3) 关键词覆盖率(key cover rate, KCR)

口语表达题要求考生的口语表达在语义上覆盖范文的主要内容, 但是目前语义分析的技术水平还很难达到实用水平, 因此, 需要手工标注对于一篇范文在被表达中应该提及的一些关键性的词的集合(关键词未必一定包含在原始范文中), 以这样一个关键词集合的覆盖度来代替范文内容在语义上的覆盖度, 这虽然是一种很大程度的近似, 但是一种目前比较可行的办法。

### (4) 文本覆盖率(text cover rate, TCR)

经分析发现, 人工选出的关键词并不能表征完整程度, 因为许多考生会用自己的语言绕过关键词进行描述。因此, 定义文本覆盖率作为表征考生口语表达完整与否的度量。

## 4 总结

目前国内对语言训练的指导和批改大多还停留在智能导师系统的阶段, 通过深度学习模型来实现主观题目的评价少之又少, 本文在训练通用计算机化考试平台(大学英语语言技能训练系统)数据的基础上, 阐述了各个模型在语言训练指导和批改中的应用, 通过各模型在语言训练指导和批改中的应用, 将理论落实到具体实践中, 实现学生成绩智能评分。根据教师评分规则, 通过深度学习技术对

语进行评分, 实现批改的自动化, 是语言训练指导和批改的重大突破。

## 参考文献

- [1] 蒋艳, 马武林. 中国英语写作教学智能导师系统: 成就与挑战——以句酷批改网为例[J]. 电化教育研究, 2013, 07: 76-81.
- [2] 梁贺君. 支持智能批改的作业在线系统设计与实现[D]. 复旦大学, 2008.
- [3] 刘雷. 英语作文智能批改中语法检查的研究与实现[D]. 北京邮电大学, 2013.
- [4] 王振宇. 电路类虚拟实验自动批改与智能指导系统的设计及实现[D]. 北京邮电大学, 2011.
- [5] 张理志. 网络环境下英语作文在线智能批改与师生角色定位——以句酷批改网为例[A]. 外语教育与翻译发展创新研究(第五卷)[C]., 2016: 3.
- [6] 贾硕. 智能批改系统的研究与设计[D]. 北京邮电大学, 2008.
- [7] Franco H, Neumeyer L. Automatic scoring of pronunciation quality for language instruction[J]. Journal of the Acoustical Society of America, 1996, 100(4): 2763.
- [8] Martín Abadi, Paul Barham, Jianmin Chen. TensorFlow's High-level Module for Distributed Machine Learning[J]
- [9] 刘庆升, 魏思, 胡郁, 等. 基于语言学知识的发音质量评价算法改进 [J]. 中文信息学报, 2007, 21 (6): 92 - 96.
- [10] 严可. 面向大规模英语口语机考的复述题自动评分技术[J]. 清华大学学报 (自然科学版) 2009, Vol. 49,
- [11] 王明涛. 英语作文智能评阅系统中高分作文判别模型的研究 [D]. 北京邮电大学, 2012.
- [12] 葛诗利, 陈潇潇. 国外自动作文评分技术研究[J]. 外语电化教学, 2007, (05): 25-29. [2017-08-03].
- [13] 陈贻东. 计算机辅助英语作文批阅系统的研究与设计[D]. 中国科学技术大学, 2016.