

# 基于自然语言处理的主观题评分小程序设计

黄宸琰 李华楠 戚欣悦

江西财经大学

**摘要:** 随着人工智能和大数据时代的到来,以人工智能和大数据技术为基础的计算机智能评阅技术的研究逐渐深入。本文结合自然语言处理技术,从关键词提取、文本相似度等维度设计评分算法,研究开发主观题自动分析批阅小程序,为教师和学生提供自动批改、个性推荐、题库管理等功能,以期达到高效公正的批阅作业,减轻教师工作负担,缓解教师压力的目的。

**关键词:** 微信小程序; 自然语言处理; 在线批阅

【DOI】10.12252/j.issn.2096-6288.2022.02.067

## 一、研究背景

近年来,越来越多的国内和国际权威考试采用网络考试的形式,利用光学字符识别技术对客观题进行智能评阅,在主观题评阅方面,也有少量考试如TOFEL采用智能评阅技术。虽然市面上存在大量的评阅系统,但大多只能满足客观题的评价,普遍不能准确高效地评阅主观题,造成了大量人力投入。相对于人工评阅,智能评阅可以减少人员投入,节约考试成本,同时更加具有客观性。本项目以实现自动评阅不同特点的主观题、纠正中文语言使用问题为主要目标,开发微信小程序,有针对性地帮助不同学习者提高知识水平,减轻教师的工作负担。

## 二、主观题评阅技术算法

### (一) 中文分词算法

中文分词技术是自然语言处理中的重要课题,根据实现原理和特点,可分为基于词典的分词算法和基于统计的机器学习算法。由于后者需要大量的人工标注和较长的处理时间,本项目使用基于词典的分词算法。依托常用的结巴分词工具,对文本字符串进行切割,找出符合词库的词语,并提前将可能用到的学科相关词汇加入语料库,可以在一定程度上避免未收录词问题。

### (二) TF-IDF文本关键词提取算法

TF-IDF算法是一种基于统计学的文本关键词提取算法,能够表示不同关键词对文件的重要性和相关性。它的主要思想是,对文件 $d$ 中的某个关键词 $t$ ,如果 $t$ 的TF值越高,同时IDF值越低,则关键词 $t$ 越能代表文本的主题。通常包括两个算法:TF算法和IDF算法。

TF算法用于统计关键词 $t$ 在文件 $d$ 中出现的频率,算法思路可概括为:如果一个关键词在文件中反复出现,则可以认为该关键词对文件有很强的代表性,且频率越

高,这种代表性就越强。计算方法表示为关键词 $t$ 的出现次数除以文件 $d$ 中所有关键词次数之和。但在实际应用中,单用TF算法存在长文本以及“作弊文本”等问题,因此需结合IDF算法使用。

IDF算法能够反映是一个词语的普遍重要性。若某个关键词 $t$ 在文件 $d$ 中出现较多而在其他文档中出现较少,则该关键词与文件 $d$ 的相关性较高。对某一特定的关键词, IDF的计算可以由总文档数目除以包含该词语的文档数目,再将得到的商除以10为底的对数得到。

TF-IDF算法就是TF算法和IDF算法的同时使用,公式为:

$$tf \times idf(i, j) = tf_{ij} \times idf_i = \frac{n_{ij}}{\sum_k n_{kj}} \times \log\left(\frac{|D|}{1 + |D_i|}\right) \quad (2.1)$$

其中 $n_{ij}$ 表示词 $i$ 在某行文本 $j$ 中的出现次数,分母部分为某行所含有的关键词出现次数之和,  $|D|$ 表示语料库中的文档总数,  $|D_i|$ 表示语料库中含有关键词 $t$ 的文件数目。分母加1是采用了拉普拉斯平滑,同时也避免了由于关键词 $t$ 没出现而导致分母为零的情况,增强算法的健壮性。

由于该算法既用TF考虑了文本中最多的关键词,又用IDF过滤掉常见的词语,因此能很好地提取出重要的关键词。

### (三) 文本相似度计算

#### 1. 语义相似度计算

对语义相似度的计算是从词向量的角度,通过word2vec模型文本转化为词向量,再计算余弦相似度。Word2vec是google在2013年发布的词向量训练工具,包含CBOW和Skip-Gram两种模型。其中CBOW模型通过计算当前词的上下文得到当前词出现的概率, Skip-gram模

型通过当前词预测其上下文。

本项目采用word2vec作为训练工具，用Skip-gram模型对语料库进行模型训练，将得到的模型用于学生答案和参考答案的关键词向量化。通过Word2vec训练的模型得到文档向量，通过余弦相似度的方法来计算语义相似度，即用两个向量之间的距离来表示两个词语之间的相似程度，计算公式如下：

$$S(\bar{x}, \bar{y}) = \cos(\bar{x}, \bar{y}) = \frac{\sum_{k=1}^t x_k \times y_k}{\sqrt{\sum_{k=1}^t x_k^2} \times \sqrt{\sum_{k=1}^t y_k^2}} \quad (2.2)$$

其中 $t$ 表示向量维度。

## 2. 词语相似度计算

### (1) LDA模型

词语相似度的计算采用潜在狄利克雷分配模型。

LDA认为文章由若干隐含主题组成，且文档中的每一个词都是由其中的一个主题生成，并且词与词之间没有顺序和先后关系，也忽略语法结构的影响。因此形成了文章-主题-词语三层结构，如图1所示：

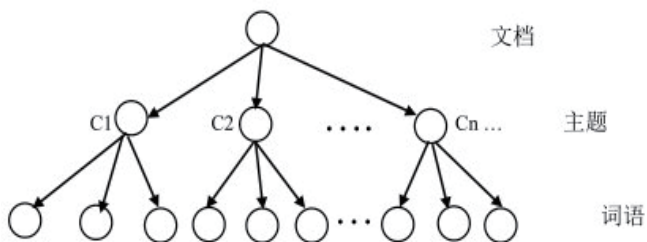


图1 LDA主题模型示意图

同时LDA模型也是一种概率模型，它认为每篇文章中单词的出现并不是偶然，而是某个主题以一定概率选择是否出现在文章中，并且主题以一定的概率选择某个词语。因此出现在同一主题或同一文章内的词语可以认为有较强的相关性。相比基于外部词典的计算方法，LDA模型更具有鲁棒性。

### (2) 基于LDA模型的词语相似度计算

使用LDA模型对语料库进行训练，可以得到文档-主题矩阵和主题-词语矩阵。为了得到词语和主题的映射关系，我们采用贝叶斯公式计算关键词 $w_i$ 出现在主题 $t_j$ 的概率：

$$p(w_i | t_j) = \frac{p(t_j) \times p(w_i | t_j)}{p(w_i)} \quad (2.3)$$

其中 $t_j$ 代表文档中的主题， $p(t_j)$ 表示语料库中主题 $t_j$ 出现的概率。

由此得出每个词语 $w_i$ 可能属于不同主题 $t_j$ 的概率，把它看作词语的特征向量，计算不同词语间的余弦相似度。

### (3) 语句相似度计算

对句子相似度的计算采用Jaccard距离计算，Jaccard相似系数是一种反映有限样本集之间的相似性与差异性的距离指标，Jaccard系数越大，则文本越相似。计算方式是用两个集合的交集除以两个集合的并集，公式如下：

$$Jaccard(u) = \frac{R(u) \cap T(u)}{R(u) \cup T(u)} \quad (2.4)$$

对学生给出答案文本 $T$ ，用TF-IDF算法将文本抽象成关键词集合，将教师参考答案 $R$ 中的关键词与LDA模型训练的同义词结合后，进行向量化，得到语句相似度。

引入Jaccard相似度算法考虑到了主观题批阅时“踩点得分”给分原则，符合实际阅卷时的一般规律。同时关键词与同义词的集合提高了评分准确性，避免了由于用词不同而造成的误判等情况。

### (四) 综合得分计算

本项目共使用了三种相似度计算方法，由用户提前为这三种方法设置权重，得到加权相似度，再用总分相乘得到最后的得分，计算公式如下：

$$Score = total \times (\cos Sim \times w_1 + wordSim \times w_2 + jacSim \times w_3) \quad (2.5)$$

其中 $total$ 表示题目总分， $cosSim$ 表示语义相似度， $wordSim$ 表示词语相似度， $jacSim$ 表示句子相似度， $w_i$ 表示各个相似度的权重。

## 三、微信小程序功能介绍

### (一) 微信小程序设计简介

主观题自动评阅系统借助微信小程序平台，无须下载安装。目标群体不仅定位于各年龄段学习者身上，也着力于为老师减轻负担，辅助教师教学。小程序包括在线测试、自动批改、个性化题库推荐和用户社区等功能模块，融合多样化功能，高效快速的完成一站式学习和自动化智慧批改。

### (二) 前台功能模块

自动评阅小程序主要实现的功能有：在线测试、自主训练、学习乐园和辅助功能，各种功能相辅相成，相互独立，满足不同用户的个性化需求，下文将依次对功能设计进行具体介绍。

### 1. 在线测试

在线测试板块包含设置试题、作答提交、自动评阅和反馈结果四个步骤，老师可以用此功能发布作业，反馈学习情况；学习者可以通过本功能自主上传试题完成自测，定向练习。此外本功能通过算法保障评阅的准确性，极大程度上降低误判、错判的概率，为学习者和老师构建一个省时省力的平台。

#### 1) 设置试题

学生可选择题库中的题目类型进行练习，也可由教师进行试题设置操作。根据系统提示，输入题目及参考答案，并对不同的题目进行不同的分值设置以及词语相似度、语句相似度和语义相似度的权重设置。设置完毕后，即可发布试题，供学生实时作答。

#### 2) 作答提交

学习者接受到试题后，在答题区输入文本进行答案上传，作答完毕后可检查答题卡，确保答题完成；同时未完成的题目可跳过，完成作答后成功提交即可等待评阅。

#### 3) 自动评阅

作答者提交后，系统会对文本进行处理，与参考答案进行对比，给出相应答案和评分细则，以便学习者对照修订。

#### 4) 反馈结果

系统生成并保存答题结果及统计分析报告，收录进错题集，实时反馈给教师端和学生端，既能帮助教师了解学生答题情况，又能帮助学生全面、科学地审视自己的能力，精准地找出自己的优势与不足。

### 2. 自主训练

学习者可以通过微信小程序在试题资源库中选择适合自己的试题类型进行更有针对性的自主训练或考试模拟，一些技能考试也可以在小程序上导入题库进行学习，给师生提供便捷途径。

### 3. 学习乐园

即学习社群，个人版用户注册登录后可以在学习贴吧上发表自己的学习经历、经验分享，共享学习资源等，还可以进行对战答题，提高学习趣味性。

### 4. 辅助功能

若用户想要查询自己的课程，观看记录等，同时想要直观的学习记录，可以到个人中心搜索学习周报或者学习记录，为之后的学习提供参考。

### 结语

本文提出了一种基于自然语言处理的主观题评阅小程序的设计，详细介绍了主观题自动评阅的主要技术和算法实现思路，并且详细描述了小程序在线测试、自主训练等不同模块的设计，同时也介绍了模块中具体页面的实现，帮助学习者和老师在平时学习教学过程中更加公平高效的完成阅卷，减轻压力，同时提高监督效力。从目标群体的需求出发，使用户在一个平台内可进行一站式阅卷和学习，满足了用户的需求，助力国内相关领域的发展。

### 参考文献

[1] 吕亚伟, 李芳, 戴龙龙. 基于LDA的中文词语相似度计算[J]. 北京化工大学学报(自然科学版), 2016, 43(05): 79-83.

[2] 李晓, 解辉, 李立杰. 基于Word2vec的句子语义相似度计算研究[J]. 计算机科学, 2017, 44(09): 256-260.

[3] 石凤贵. 基于自然语言处理的Word2Vec词向量应用[J]. 黑河学院学报, 2020, 11(07): 173-177.

[4] 彭云. 提取商品特征和情感词的语义约束LDA模型研究[D]. 江西财经大学, 2016.

[5] 孙昌年, 郑诚, 夏青松. 基于LDA的中文文本相似度计算[J]. 计算机技术与发展, 2013, 23(01): 217-220.

[6] 王振振, 何明, 杜永萍. 基于LDA主题模型的文本相似度计算[J]. 计算机科学, 2013, 40(12): 229-232.

作者简介: 黄宸琰(2021.7-), 女, 汉族, 江西新余人, 江西财经大学本科在读, 研究方向: 自然语言处理

李华楠(2001.6-), 女, 汉族, 黑龙江铁力人, 江西财经大学本科在读, 研究方向: 信息与计算科学。

戚欣悦(2002.1-), 女, 汉族, 安徽滁州人, 江西财经大学本科在读, 研究方向: 国际经济与贸易。

基金项目: 2022“江西财经大学国家级大学生创新创业训练计划”项目资助 项目编号: 202210421110。