

教育领域反馈文本情感分析方法及应用研究

欧阳元新, 王乐天, 李 想, 蒲菊华, 熊 璋

(北京航空航天大学 计算机学院, 北京 100191)

摘 要: 对教育领域反馈文本情感分析的作用进行分析, 并对相关研究现状进行介绍, 以中国大学MOOC评论文本数据集为基础, 介绍基于裁切语言模型与注意力机制的情感分析方法及在北京航空航天大学计算机导论与伦理学国家精品课线下教学课堂中的应用情况, 验证线上教育领域情感分析技术迁移至线下教育领域的可行性, 为情感分析技术在线下教育领域的应用提供实践参考。

关键词: 情感分析; 反馈文本; 自然语言处理

DOI:10.16512/j.cnki.jsjy.2020.06.019

0 引 言

2019年2月, 中共中央、国务院印发了《中国教育现代化2035》, 强调“因材施教”“知行合一”, 其十大战略任务之一“加快信息化教育时代变革”提出利用现代技术加快推动人才培养模式改革, 实现规模化教育与个性化培养的有机结合。教学的最终目的始终是为了促进学习者的学习。不同学习者的学习需求千差万别, 如何应对学生的个体差异, 真正做到“因材施教”的个性化教学(学习), 一直是摆在每个教育工作者面前的课题。在面对范围更大、层次更复杂的学习者时, 如何获得其对课程的直观反馈, 进而动态调整课程教学组织显得尤为重要。

自然语言是人类表达信息的主要途径之一, 以自然语言表述的文本信息存在于教学活动的各个环节之中。观点挖掘和情感分析是分析用户观点、反馈、评估、态度和个人情感的研究领域。过去的15年里, 作为情感计算和自然语言处理的子任务, 关于主观性和情感分析的研究已经取得了蓬勃的发展^[1]。情感分析在通常情况下不会单独使用, 其结果会作为更高层次应用的一项特征输入。情感分析技术的作用是检测文本表达出

的情感状态。这些状态序列可以通过模式识别的方式提取出情感变化特征, 作为预测用户情感状态的依据。通过对课程反馈文本信息进行情感分析, 可得到学生对当前课程有效的情感反馈, 并以此为依据, 实时调整教学方案、优化教学方法, 实现更为精准化和个性化的教学。

1 教育领域反馈文本情感分析

情感指的是一种态度、想法或感性的判断, 用来描述观点中蕴含的褒义或贬义的情感倾向。情感和观点都是人主观意愿的一种表达, 但二者之间存在明显的区别, 观点偏重于人对于某一个事物形成的具体看法, 情感更侧重于人内在的某种情感^[2]。

教育数据挖掘(Educational Data Mining)是一个重要的研究领域, 通过观察学生的表现, 了解学生的学习情况来改善教育环境。但是仅仅通过学期末获得的学生成绩等反馈数据, 不能给已经参加完该课程学习任务的学生带来帮助。为了使正在学习中的学生同样受益, 需要实时进行数据处理与分析, 并快速给出反馈, 帮助教师理解学生的学习行为和所遇到的不同问题。

基金项目: 国家自然科学基金信息和教育交叉领域面上项目(61977002); 北京航空航天大学教学改革项目。

第一作者简介: 欧阳元新, 女, 副教授, 研究方向为自然语言处理、个性化兴趣挖掘与推荐服务等, oyyx@buaa.edu.cn。

教育领域反馈文本情感分析是一个致力于从反馈文本中提取情绪和观点的任务。情感可以是消极的或者积极的,不同的情感对应于不同的意见和建议。无论是在线下还是线上课堂,教师都可以通过对反馈文本的情感分析对课堂中的情绪状态获得快速的宏观了解。这些情绪信息可以辅助教师定位到课程安排、知识体系、教学方法等方面上的问题,进而改善教学质量、提高学生学习效率。将情感分析技术应用于学生课程反馈自动化分析,在缓解教师工作压力的同时,还可有效提高教学质量。

近年来,机器学习(尤其是深度学习)方法逐渐成熟。此类方法通过对文档进行监督训练可以得到能够有效预测文档情感极性的神经网络;使用端到端训练方式可以快速利用大量的文本数据而不依赖人工分析;预训练深度语言模型的兴起使得所有自然语言处理任务都有了强大的语言表示模型基础,便于构建具体的应用模型并快速投入到实际应用中。

Piryan等人对情感分析技术的综述文献中有统计表明,机器学习方法在当前情感分析研究中占主导地位,约67.2%的文献采用了机器学习方法进行研究,其余的才是基于规则与情感词典的方法。然而教育领域的情况正相反,后者仍然广泛地应用于教育领域研究中,并且占据了主导地位^[3]。考虑到深度学习方法有数据依赖强、模型规模大、可解释性差等先天缺点,结合传统方法与深度学习方法可以使两者更好地互补、发挥优势。

2 MOOC评论文本情感分析方法

大规模开放式在线课程(MOOC)是线上教育的产物,是教育领域中的一个相对较新的发展模式。虽然与传统教室相比,在线课程具有各种独特的优势和改变教育系统未来的潜力。但是,MOOC教育模式仍然存在明显短板^[4],从教学的角度来看,大多数MOOC使用视频实现从教师到学生的内容传递,缺少教师与学生之间的直接互动,导致教师缺少视觉或听觉上的提示来区分沮丧与热情的学生。

大多数MOOC提供课程论坛作为交流和学习工具,发表与回复课程评论是学生与授课教师或其他学生互动的主要途径。其中,课程评论是最典型的反馈文本形式。MOOC的产生带动了大量课程和课程评论的产生,这些评论都是学生抒发自身情感和表达观点的载体。Tucker等人发现,学生在论坛发言表现的情感倾向与其在MOOC平台的学习表现有一定程度的正相关性^[5]。由于可以获得大量的课程评论数据,针对MOOC的数据挖掘及情感分析技术应用相对更为成熟,传统的朴素贝叶斯、最大熵和支持向量机等技术都已被证明可以很好地与在线情感数据配合使用,也获得了不错的效果。MOOC平台使用情感分析技术,能够基于学生用户对课程的情感倾向判别,快速且准确地从海量评论文本中筛选出价值较高的反馈信息,进而实现用户退课预警、个性化课程推荐等。教师则可以根据学生的反馈动态调整教学安排,以满足学生的个性化学习需求。

2.1 实验数据集的构建

从中国大学MOOC收集11个课程大类、1768门课程的评论数据并进行一定的人工标注,构建了一个量化的中文教育领域情感极性数据集,在此基础上展开学生反馈文本情感分析研究。被评论课程所属的具体领域见表1。由于语料所处的大领域与细分领域均会对情感分类中运用的自然语言处理工具产生影响,本文在数据集中融合了多种领域的文本,尝试在跨细分领域的数据集上验证模型的情感分类效果。

数据集中的每一个评论会有一个用户给出的1~5分的打分,其中5分样本占据了总评论数据的84%,3分及以下评分样本的数量仅占据总评论数据的3.8%。为了避免这种不平衡数据导致的模型偏见,从完整数据集构建相对平衡的子集用于模型的构建与训练,具体方法如下:从用户打分为5分的样本中提取了15000条评论作为正样本(积极情感)数据,并对用户打分为1~3分的6731条评论进行人工标注,最终筛选出负样本(消极情感)评论4148条,与正样本中的15000条数据共同构成实验数据集(见表2)。

表 1 MOOC 实验数据集评论领域及评分分布

类别	评分				
	1	2	3	4	5
医药卫生	120	89	253	1 647	11 855
哲学	54	40	85	490	3 194
外语	73	56	198	1 220	7 486
工学	171	101	271	1 287	10 163
心理学	66	74	159	1 008	7 786
历史与文学	199	167	477	2 543	17 182
法学	300	182	450	2 394	21 150
理学	161	100	287	1 480	11 840
经济管理	264	213	560	3 372	22 623
艺术设计	89	88	250	1 383	10 019
计算机	304	221	609	3 622	26 132
合计	1 801	1 331	3 599	20 446	149 430

表 2 MOOC 评论文本统计信息

样本类别	平均长度	中位数长度	99% 分位长度	最大长度	词组数量
正样本 (积极)	18.6	12	83	500	3 833
负样本 (消极)	22.3	16	89	500	1 958
总计	18.7	13	84	500	3 893

从表 2 中可以看出,评论负样本的平均句长、中位数样本句长与 99% 分位样本句长都相比正样本长一些,说明负样本中潜在的语义信息更丰富,也更有可能是对于课程改进有价值的评价与观点。由于 MOOC 平台的限制,评论的最大长度均为 500 个字符,因此样本最大长度均为 500。图 1 所示为数据集中不同分位的文本平均长度折线图。

2.2 基于裁切语言模型与注意力机制的情感分析方法

BERT (Bidirectional Encoder Representation from Transformers) 是由 Google Brain 提出的一种预训练深层语言模型,训练自 BooksCorpus 与

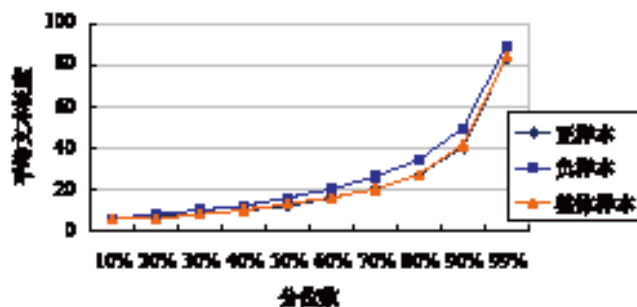


图 1 数据集中不同分位的文本平均长度

Wikipedia 语料,共计约 320 亿词的文本。其架构为多层编码器堆叠而成的栈式结构,每一个编码器都由自注意力层、全连接层与残差连接组成。BERT 模型性能提高的代价是愈发复杂的模型结构和陡增的参数数量,由此进一步导致了训练、预测时间成本的增长。此外,深层语言模型生成的词向量可能无法进一步和下游网络进行良好的协同工作。

与其他领域公开数据集不同,MOOC 评论文本以短文本为主。此类短文本分类是一种典型的分类特征抽取任务,更适合使用简单模型进行特征抽取。因此本文提出并训练得到一种基于注意力池化机制的裁切 BERT 与卷积神经网络情感分析模型^[6],尝试将浅层 BERT 词向量与卷积神经网络相结合,并在卷积操作后、池化操作前引入自注意力模块,使得该模型结构在 MOOC 评论的情感分类任务上可以达到基础 BERT 模型的效果,但模型大小与性能消耗可以大大降低,整体模型工作原理如图 2 所示。实验结果表明,本方法在中国大学 MOOC 评论文本数据集上,情感二分类(积极、消极)准确率可达 92.8%。

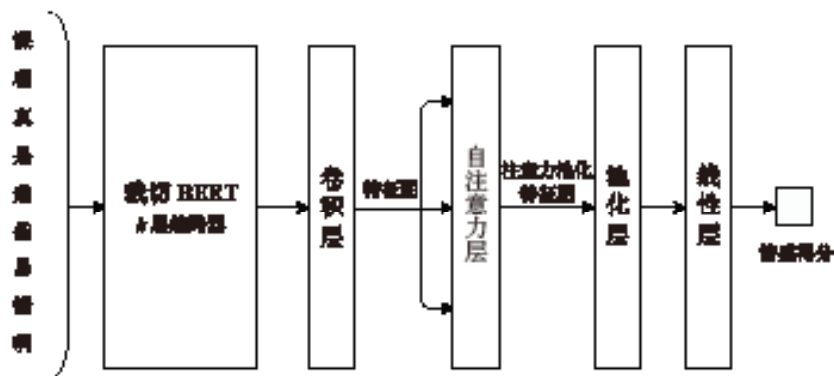


图 2 基于注意力池化机制的裁切 BERT 与卷积神经网络情感分析模型

3 从线上至线下的迁移应用

在线下课程教学中,同样可以通过情感分析技术实现学生情感自动化分析,及时发现学生情感波动,适时调整课程的教学模式与方法,以进一步提高教学质量。计算机导论与伦理学是北京航空航天大学开设的面向计算机专业的新生专业先导课程,于2008年获评国家级精品课。本文通过调查问卷的方式收集了2 078条来自选课学生的课堂反馈数据,并对正负样本(积极/消极情感)进行了人工标注作为线下测试数据集(统计信息见表3),尝试将本文提出的MOOC评论文本情感分析方法应用到计算机导论与伦理学线下教育应用中。

表3 线下课程课堂反馈文本统计信息

样本类别	平均长度	中位数长度	99%分位长度	最大长度	词组数量
正样本(积极)	31	24	127	477	1 027
负样本(消极)	31	24	130	229	1 194
总计	31	24	129	477	1 339

从文本长度上看,课堂反馈文本同MOOC评论同属于短文本。与MOOC评论相比,由于课堂反馈问卷更为正式,因此课堂反馈中的文本长度相对较长。不同于MOOC数据集中负样本平均长度较长,本课程评论数据集文本平均长度为31,中位数文本长度均为24,99%分位长度为129,正负样本间的文本长度没有体现出明显的差异,这表明正负样本没有过大的信息量差异,但正样本中的一个极长的反馈文本,使得正样本最大长度远大于负样本最大长度。此外,线上数据存在大量的无意义单字,而课堂反馈数据不存在该情况。

将通过中国大学MOOC评论数据集训练得到的情感分析模型应用于对线下课堂反馈文本的情感分析中,具体实验结果见表4。

表4 裁切复合模型在课堂反馈数据集上实验结果

数据类别	精确率	召回率	F1
正样本	0.905	0.876	0.890
负样本	0.727	0.781	0.753
整体	0.848		

由实验结果可以看出,该模型在课堂反馈数据集上精确率达到84.8%,表明通过MOOC评论数据训练得到的情感分析模型对于课堂反馈文本也具有较好的情感分类能力,但由于该数据集中同样存在正负样本不平衡的现象,导致模型存在偏见,使模型对正负样本的区分体现出一定的差别。因此,模型在对正负样本的区分上仍然体现出了与在MOOC评论数据集上相似的现象,即对正样本分类性能(90.5%)较对负样本(72.7%)分类性能更优。不同于MOOC数据负样本含有较正样本更多的信息量,线下数据中的正负样本间没有明显的信息量差异,这也会对分类预测产生一定的影响。

在线下课堂教学中,受制于一对多的教授方式,教师无法及时了解每名同学的情绪状态。以开展教学改革研究的计算机导论与伦理学课程为例,每年的选课学生在300~400人之间,让教师仅仅通过课堂上的互动,很难照顾到所有的选课学生。应用自动化分析手段后,教师可以通过对学生反馈文本的情感分析,快速获得学生的情绪状态,实现对课程安排、知识体系、教学方法等方面的针对性调整,进而改善教学质量、提高学生学习效率。另一方面,将相关方法集成到课程互动平台中,亦可根据分类预测的结果,实现对学生评论的个性化反馈和学习内容推荐。

4 结 语

使用MOOC评论文本作为训练集,对线下课程课堂反馈文本情感分析进行尝试,虽然存在模型偏见现象,模型仍能将在线上MOOC环境学习到的分类特征很好地应用至线下课堂反馈文本上,

这表明线上 MOOC 环境和线下课堂环境的语义信息、语言情感特征是相似的,语言模型及其连接的神经网络均可以有效地在线上、线下环境应用

之间迁移。本文的主要工作目前集中于对反馈文本的语句级情感分析,下一步将继续开展属性级情感分析(即观点挖掘)方面的研究和应用工作。

参考文献:

- [1] 涂慧明. 文本观点挖掘和情感分析的研究[J]. 电脑知识与技术, 2016, 12(5): 235-237.
- [2] 周立柱, 贺宇凯, 王建勇. 情感分析研究综述[J]. 计算机应用, 2008(11): 2725-2728.
- [3] Piriyani R, Madhavi D, Singh V K. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015[J]. Information Processing & Management, 2017, 53(1): 122-150.
- [4] Guo P. MOOC and SPOC, which one is better[J]. Eurasia Journal of Mathematics, Science and Technology Education, 2017, 13(8): 5961-5967.
- [5] Tucker C, Pursel B K, Divinsky A. Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes[J]. The ASEE Computers in Education (CoED) Journal, 2014, 5(4): 84.
- [6] Li X, Ouyang Y, Rong W, et al. A shallow BERT-CNN model for sentiment analysis on MOOCs comments [C]. 2019 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). Washington D C: IEEE, 2019, 1126-1131.

(编辑: 宋文婷)

(上接第79页)

的方法。采用该方法不仅可实现知识点和目标的精准对应和完整覆盖,使内容快速适应目标的变化,

而且可以辅助教师,方便、快速地完成课程教学内容、重难点和课时的调整和设计,提高工作效率。

参考文献:

- [1] 教育部. 教育部办公厅关于公布首批“新工科”研究与实践项目的通知[EB/OL]. (2018-03-21). http://www.moe.gov.cn/srcsite/A08/s7056/201803/t20180329_331767.html.
- [2] 新工科, 到底是什么? [EB/OL]. (2019-08-27). <https://zhuanlan.zhihu.com/p/60563100>.
- [3] 石兆英, 洪亲, 蔡坚勇, 等. 面向新工科教育的IT类课程教学改革探讨[J]. 计算机时代, 2018, 318(12): 107-109, 112.
- [4] 韩波, 李亦涵, 李兵, 等. 新工科建设背景下的《Web程序设计》课程实践教学方法探讨[J]. 软件导刊(教育技术), 2019, 18(1): 65-72.
- [5] 谢涛, 薄钧戈, 房琛琛. 新工科背景下Web编程技术课程的实验教学改革研究[J]. 计算机教育, 2019, 289(1): 67-71, 75.
- [6] 徐儒. 新工科背景下WEB类课程“四三式”教学体系研究[J]. 实验技术与管理, 2019, 36(6): 249-253.
- [7] 李冬霞, 张德会. 基于知识点量化分析的微课程设计策略研究[J]. 软件导刊, 2014(7): 160-162.
- [8] 周冬梅, 周阳, 鲁若愚. 美国创业课程内容构成热点及趋势研究: 基于美国3所高校的课程文本量化分析[J]. 实验技术与管理, 2017, 34(11): 283-288.
- [9] 吴倩, 关湃, 王传龙. 以项目方案为核心的商业展示设计课程DRP实践教学模式探究[J]. 北华航天工业学院学报, 2019, 29(2): 32-34.
- [10] 王海青, 钱玲, 谷亚林. 基于关联度分析的信号处理课群研究[J]. 南京理工大学学报(社会科学版), 2019, 32(6): 88-92.

(见习编辑: 郭安琪)