Modeling and Simulation 建模与仿真, 2022, 11(6), 1424-1433
Published Online November 2022 in Hans. http://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2022.116134

Hans汉斯

基于深度学习的大学生幸福感测度分析

付瑞鑫，张　莉，屈启兴

对外经济贸易大学信息学院，北京

收稿日期：2022年9月9日；录用日期：2022年10月27日；发布日期：2022年11月7日

摘　要

大学生情绪容易受到各类事件影响，学业求职的激烈竞争和疫情的突如其来导致其幸福感明显降低，利用社交媒体分析大学生幸福感引起了国内外学者的关注。本文提出一种利用社交媒体数据，基于深度学习ALBERT-TextCNN模型的大学生幸福感测度分析方法。首先使用ALBERT预训练语言模型将社交媒体文本描述转化成向量表示，提取文本描述中的关键特征，然后将提取到的特征送入TextCNN模型进行分类预测，得出社交媒体文本的情感极性，最后将积极情感文本占比作为大学生用户的幸福感指数。在公开微博文本数据集上进行实验，ALBERT-TextCNN模型在情感极性分类预测上准确率、精确率、召回率和F1值均达到较高水平，同时训练时间短成本低。最终本文利用此模型确定了北京某高校466名大学生的幸福感指数情况。

关键词

幸福感测度，ALBERT模型，TextCNN模型，社交媒体

Analysis of College Students' Happiness Perception Based on Deep Learning

Ruixin Fu, Li Zhang, Qixing Qu

School of Information Technology & Management, University of International Business and Economics, Beijing

Received: Sep. 9th, 2022; accepted: Oct. 27th, 2022; published: Nov. 7th, 2022

Abstract

College students' emotions are easily affected by various events. The fierce competition for academic job hunting and the sudden outbreak of the epidemic have led to a significant decrease in their happiness. The use of social media to analyze college students' happiness has attracted the attention of scholars at home and abroad. This paper proposes an analysis method of college students' happiness perception based on deep learning ALBERT-TextCNN model using social media

文章引用: 付瑞鑫, 张莉, 屈启兴. 基于深度学习的大学生幸福感测度分析[J]. 建模与仿真, 2022, 11(6): 1424-1433.
DOI: 10.12677/mos.2022.116134

data. First, the ALBERT pre-trained language model is used to convert the social media text description into a vector representation, and the key features in the text description are extracted, and then the extracted features are sent to the TextCNN model for classification and prediction, and the emotional polarity of the social media text is obtained. Finally, the proportion of positive emotional texts is taken as the happiness index of college students. Experiments were carried out on the public microblog text dataset, and the ALBERT-TextCNN model achieved a high level of accuracy, precision, recall, and F1 value in sentiment polarity classification prediction, and at the same time, the training time was short and the cost was low. Finally, this paper uses this model to determine the happiness index of 466 college students in a university in Beijing.

Keywords

Happiness Perception, ALBERT Model, TextCNN Model, Social Media

1. 引言

大学生网民比重大，其情绪容易受到各类事件的影响，他们也热衷于在社交媒体平台发布自己对某一事件的意见、看法，加之疫情突如其来，大学生群体在学业、求职等各方面竞争更加激烈，压力更大，幸福感逐渐降低。通过对大学生发布的信息进行分析，可以在不对他们造成干扰的情况下实时、大规模地了解这类群体的情绪和幸福感。

幸福感是指人类基于自身的满足感与安全感而主观产生的一系列欣喜与愉悦的情绪[1]。以往专家学者对幸福指数测量方法主要包括基于量表的问卷调查、基于统计数据的模型构建和基于社交媒体的数据挖掘，以此得出研究群体的幸福指数水平。Di Wang [2]和毛良斌[3]等学者利用社交媒体数据建立了幸福感监测分析体系，尤其是后者验证了积极自我呈现和真实自我呈现均能显著提高主观幸福感。

在此背景下，本文将研究群体在社交媒体发布内容的积极情感占比定义为幸福感指数，并利用在短文本分类上效果佳且轻量化的 ALBERT-TextCNN 模型，针对大学生发布在社交媒体上的短文本数据，对该群体的幸福感指数进行研究分析，这既是利用社交媒体数据进行幸福感测度的一种尝试，也是将深度学习模型应用于幸福感测度的一项创新。

2. 相关研究

2.1. 幸福指数相关研究

幸福指数是一种主观感受、心理体验、愉悦心情的量化[1]，近些年来众多专家、学者相继展开了各种方面幸福指数相关的研究，采用的方法包括文献研究、问卷调查、实证分析、数据挖掘等多种方法，对大学生、居民等不同群体的幸福感测度做出了研究贡献。

祝琳[4]和付文宁[5]等学者使用幸福感量表量化幸福指数，即量表得分越高幸福感越高，前者使用美国国立统计中心制定的总体幸福感量表，验证了团体辅导可以有效地提高大学生的主观幸福感水平，后者则应用 Campbell 等人编制的幸福感指数量表得到了商洛市 140 名大学生的幸福感基本情况；郝乐[6]和 Strotmann [7]等学者均提出结合主观幸福感和客观条件的幸福感分析方法，前者用改进的距离综合评价法统计测量客观幸福指数，将客观幸福指数与主观幸福指数的加权平均值作为测量幸福感的综合指标，

后者则对印度四个村庄 2300 多人进行实证分析，并证明多维贫困指数反映的客观条件缺失与幸福感缺失存在正相关；Di Wang [2]等利用社交媒体 Twitter 数据对双语国家幸福指数进行监测分析，其提出了一种利用社交媒体通过多级过滤进行总体人口情绪监测的系统；毛良斌[3]基于社交媒体的角度，探讨了自我呈现与主观幸福感的关系，验证了社交媒体自我呈现能够体现主观幸福感的高低，积极自我呈现和真实自我呈现均能显著提高主观幸福感，消极自我呈现则显著降低主观幸福感。

2.2. 情感分析相关研究

在情感分析领域，新兴起的深度学习模型表现了良好的效果，在计算机视觉、语音识别、自然语言处理等领域都得到广泛应用，本文选择了轻量化的 ALBERT 和 TextCNN 相结合，在最大化深度学习模型效果的情况下，尽可能的降低模型成本。

Google 在 2018 年提出了 BERT 预训练模型[8]，其主要是运用了预训练 + 微调的模式，预训练阶段使用大规模的数据让模型以无监督方式学习到大量知识，在模型具备大量先验知识后，微调阶段再采用自身标注数据完成最后一步的监督学习，这大大提高了模型的效果。其模型结构如图 1 所示，初始字向量 E1，E2，…，EN 通过由多个 transformer 组成的编码器，最后输出包含丰富语义信息的字向量 T1，T2，…，TN。

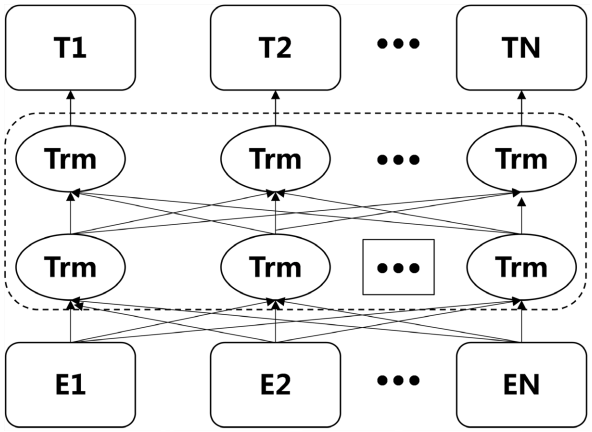

Figure 1. Schematic diagram of Bert structure
图 1. Bert 结构示意图

在深度学习中，把网络变深可以增加模型效果，但将 BERT 的网络变深会出现明显的参数爆炸问题，而 ALBERT 模型[9]可以解决这个问题。ALBERT 保留了和 BERT 相同的模式，通过使用参数因式分解和跨层参数共享两个技术减少了模型参数量，同时构建句子连贯性预测任务 SOP (Sentence-Order Prediction)使模型学习更细粒度的文本层次连贯性，提升了模型的效果。ALBERT 参数更少，对语义特征提取能力更强，更加适合于本文的短文本情感分析任务。

Kim 在 2014 年提出的 TextCNN [10]是一种文本分类模型，是卷积神经网络 CNN 的一个变体，其使用不同尺寸的卷积核对文本局部特征进行训练提取，从而得到多样性且代表性更强的特征，在短文本分类上取得了较好的效果。从结构上讲，TextCNN 分为输入层、卷积层和池化层，如图 2。

综上所述，以往学者对幸福感指数的度量主要还是依托于各种量表、调查问卷和实证分析来展开，通过提炼各项影响幸福感指数的指标进行幸福感指数的计算，这需要大量的人力和时间，并且通过调查获得的数据无法确保内容真实有效。本文主要在 Di Wang [2]和毛良斌[3]等学者的启发和研究基础下，基于数据规模大、由用户自主创作且能够真实客观地反映用户幸福感情况的社交媒体数据，使用对用户博文进行情感分析并进一步计算幸福感指数的方法对大学生幸福感测度进行分析研究。

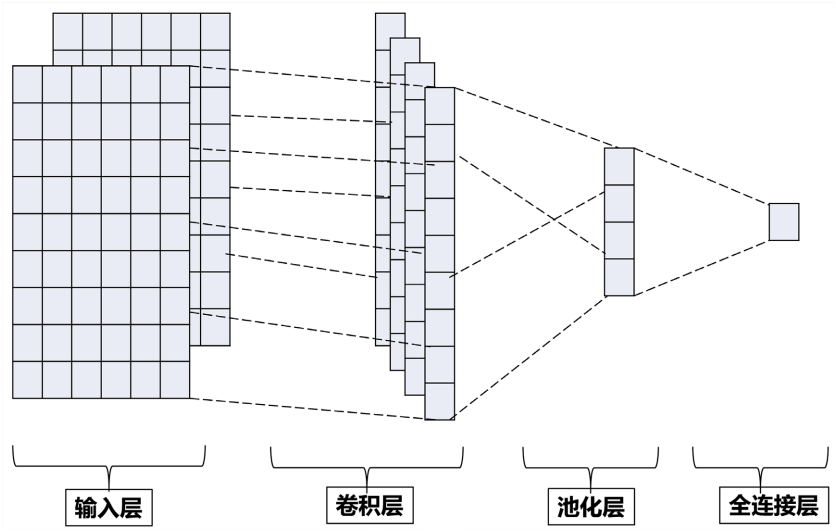

Figure 2. A schematic diagram of the structure of TextCNN
图 2. TextCNN 结构示意图

3. 幸福指数计算模型构建

3.1. 幸福指数计算

首先使用 ALBERT 预训练语言模型获取微博文本的动态特征表示，使得句子中同一个词在不同上下文语境中具有不同的词向量表达；然后利用 TextCNN 对特征进行训练，充分考虑文本中的局部特征信息和上下文语义关联，得出微博文本的情感极性；最后，利用微博文本的情感极性计算大学生用户的幸福感指数。如图 3 所示。

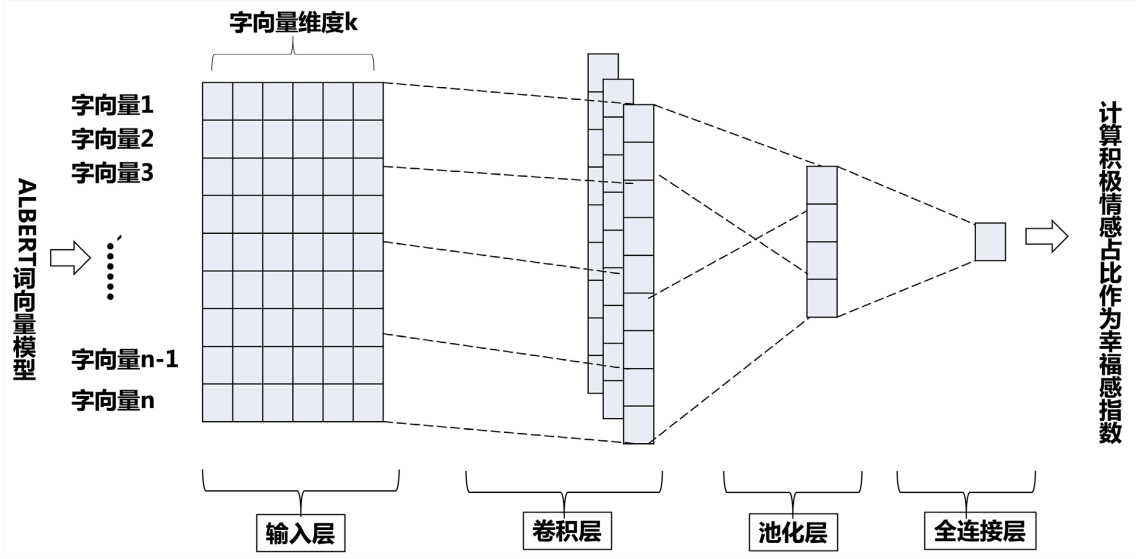

Figure 3. Schematic diagram of the structure of calculating happiness index based on ALBERT-TextCNN model
图 3. 基于 ALBERT-TextCNN 模型计算幸福感指数结构示意图

幸福感是指人类基于自身的满足感与安全感而主观产生的一系列欣喜与愉悦的情绪，而幸福指数是一种主观感受、心理体验、愉悦心情的量化[1]。本文基于社交媒体数据的情感分析计算幸福感指数，将

研究群体在社交媒体发布内容的积极情感占比定义为幸福感指数，即社交媒体发布内容越积极，欣喜与愉悦情绪越多，越能体现其自身的满足感和安全感，则幸福感指数越高。

如果一个人表现出更多积极情绪，说明其幸福感更高，所以本文将积极情感所占比重定义为幸福感指数。

即：

$$HI = \frac{NP}{TB}$$

其中，*HI* (*Happiness index*)为幸福感指数，取值范围为 0~1，*NP* (*Number of positive emotional blogs*)为用户所发布的积极微博数，*TB* (*Total number of blogs*)为用户所发布的全部微博总数。

3.2. 研究设计

本文研究框架包括 5 个步骤。

1) 数据收集与预处理。通过 GitHub 开源网站获得微博文本情感标注语料作为训练集，对评论文本进行数据清洗，过滤特殊字符如“@【:】”等，删除多余空白字符等。

2) 文本分词、清洗、标准化。利用开源 jieba 分词工具，对微博文本进行分词处理，并利用结合了哈工大停用词表、四川大学机器智能实验室停用词库及百度停用词表等在内的停用词汇总表进行停用词过滤。

3) 文本特征提取。将分词序列结果传入 ALBERT 模型进行微调训练，得到每个微博文本的特征序列。

4) 分类模型构建。构建 ALBERT-TextCNN 模型训练微博文本的情感分类模型，与其他模型在模型效果和训练时间等评估指标上进行对比，应用对比之后较好的模型。

5) 计算幸福感指数。用模型对微博上大学生发布的实时数据进行分析，并计算大学生积极情感占所有情感的比重，作为幸福感指数。

4. 实验结果和分析

4.1. 实验环境配置

本实验基于 tensorflow1.14.0 环境，具体软硬件配置如表 1 所示。

Table 1. Experimental environment configuration table
表 1. 实验环境配置表

操作系统	macOS 10.15.7
处理器	2.6 GHz 六核 Intel Core i7
内存	16 GB 2667 MHz DDR4
启动磁盘	Macintosh HD
python	3.6
tensorflow	1.14.0

模型训练采取 adam 优化器，学习速率 learning_rate 取值 0.00005，batch_size 大小为 64，sequence_length 为 200，具体参数配置如表 2 所示。

Table 2. Model parameter configuration table
表 2. 模型参数配置表

类别	参数	值
训练参数	num_train_epoch	5
	print_step	10
	batch_size	64
	batch_size_eval	128
	summary_step	10
	num_saved_per_epoch	3
	max_to_keep	100
	ALBERT	albert_small_zh_google
优化参数	optim	adam
	warmup_proportion	0.1
	use_tpu	None
	do_lower_case	TRUE
	learning_rate	5.00E−05
TextCNN 参数	num_filters	128
	filter_size	[2, 3, 4, 5, 6, 7]
	embedding_size	384
	keep_prob	0.5
其他参数	sequence_length	200
	weight_decay	1.00E−06
	seed	666666
	dropout	0.3

4.2. 实验数据

本实验采用代码托管平台“github”上的公开语料库进行模型训练。该语料库是对一些微博文本进行正负情感标注的数据集合，其建设时间较新，在数据标注期间采取多人核验保证数据质量。语料库共包含 119,988 条数据，其中正向 59,993 条，负向 59,995 条，将该语料库划分为训练集和测试集两个部分。训练集用于对情感分类模型的训练评估；测试集用于检验构建的模型能否准确得出正确分类标签。公开语料库的具体情况如表 3 所示。

Table 3. Description of public datasets
表 3. 公共数据集描述

类别	训练集	测试集
积极情感	41,996	17,997
消极情感	41,996	17,999
总数	83,992	35,996

4.3. 实验流程

为验证 ALBERT-TextCNN 文本情感分析模型的有效性，将本文模型与 BERT-TextCNN、BERT、ALBERT、TextCNN 等模型进行对比，在同一微博文本数据集上分别进行实验。BERT-TextCNN、ALBERT-TextCNN 采用 Google 发布的中文预训练模型 BERT-Base 和 ALBERT-Base 来进行文本特征表示。本研究构建的模型分别对经过文本预处理及文本向量化的测试集进行情感倾向判定，即正向情感标记为“1”，负向情感标记为“0”。

实验流程如图 4 所示。

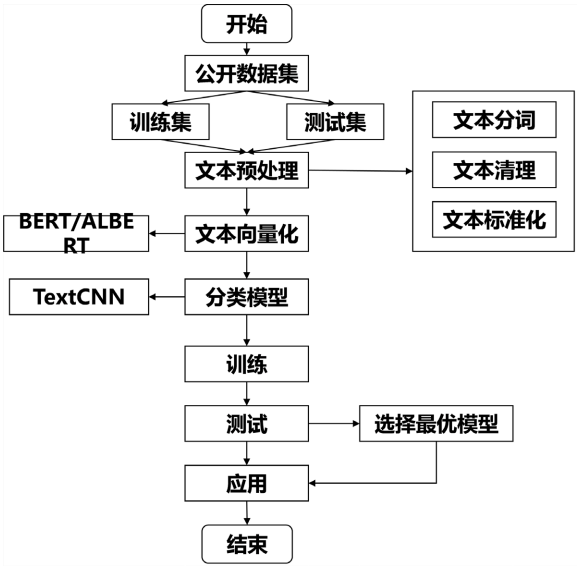

Figure 4. The analysis diagram of the experimental flow
图 4. 实验流程分析图

4.4. 实验结果与分析

采用准确率(accuracy)、精确率(precision)、召回率(recall)和 F1 值(F1-score)等评估标准对构建的情感分类模型进行评估结果如表 4 所示。

Table 4. Model evaluation metrics
表 4. 模型评价指标

评价指标	Accuracy	Recall	Precision	F1
TextCNN	0.79	0.80	0.81	0.79
ALBERT	0.88	0.84	0.89	0.88
BERT	0.91	0.91	0.90	0.90
BERT-TextCNN	0.95	0.92	0.96	0.95
ALBERT-TextCNN	0.94	0.92	0.95	0.93

从上表中可以看出，BERT-TextCNN 模型和 ALBERT-TextCNN 模型较其他模型，各项性能指标均大幅提升，说明在文本情感分析问题中，此类模型本身的特点保证了情感分类的准确性和实时性。

在 BERT-TextCNN 模型和 ALBERT-TextCNN 模型比较上，由于 ALBERT 模型采取了因式分解和跨

层共享，大幅减少了模型参数，降低了模型空间复杂度，但也使模型的性能略有损失，从表 4 中可以看到，准确度 A、精确度 P 和 F1 值均有下降。但随着模型复杂度的降低，模型的运行效率大大提高，训练时间明显减少。本文在此次实验环境下，对两个模型上的训练时间进行了测试，结果如表 5 所示。

Table 5. Model training time
表 5. 模型训练时间

模型	训练时间
BERT-TextCNN	1 小时 58 分
ALBERT-TextCNN	55 分

可以发现，ALBERT-TextCNN 模型在能够保证 A、R、P、F 四个标准的值比较高的情况下，训练效率相比 BERT-TextCNN 模型提升到约 2.15 倍。因此，本文后续研究最终选择应用成本低、效果好的 ALBERT-TextCNN 模型进行基于情感分析的大学生幸福感测度研究。

4.5. 模型应用

本研究选取了 2021 年 9 月~12 月曾定位北京某高校发博 5 次以上的 466 个学生账户，对数据脱敏后获取了此群体 2019 年 1 月 1 日到 2021 年 12 月 31 日的共 97,703 条有效微博数据，其中 2021 年 54,815 条，2020 年 25,938 条，2019 年 16,950 条。微博文本及发布时间作为本文主要的实验数据。

本研究对 2019 年 1 月 1 日到 2021 年 12 月 31 日期间 466 名学生的 97,703 条有效文本数据进行分词、去停用词等文本预处理及文本向量化，选择 ALBERT-TextCNN 构建的情感分类模型对其情感极性进行了分类，并进行幸福感指数的计算。统计结果如下。

1) 大学生总体幸福感指数分布情况

其次，本研究对每个月份学生总体幸福感指数进行了分析，结果如图 5 所示。可以发现，从年度来看，2019 年最高，2020 年次之，2021 年最低，三年都在 6~8 月出现低谷，2020 年和 2021 年更是在 12~2 月之间出现了低谷。其中 2019 年和 2020 年 9-10 月期间幸福感指数均处于上升期，但 2021 年从 9 月就表现出情绪持续下降的情况。

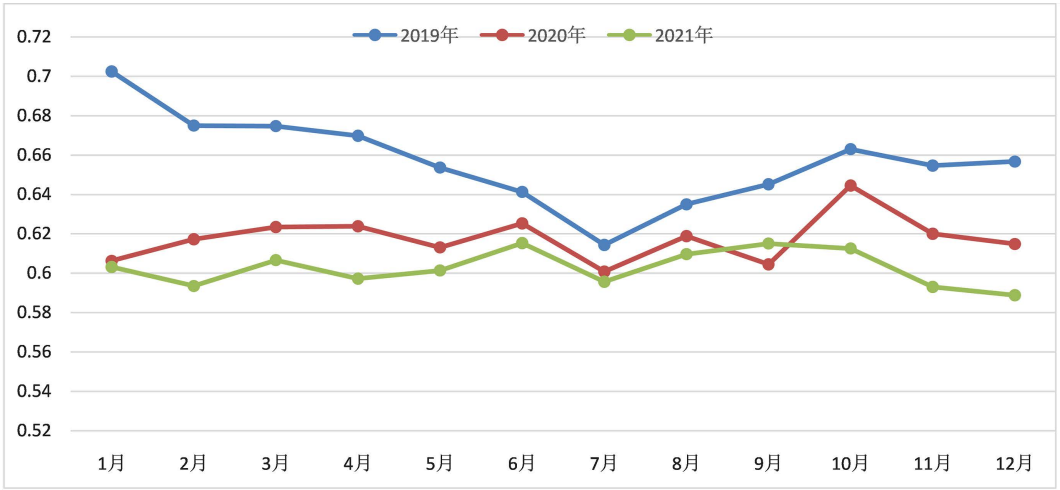

Figure 5. The distribution of the overall well-being index of college students
图 5. 大学生总体幸福感指数分布

2) 大学生个体幸福感分布情况

将 466 名大学生的情感分析情况进行汇总统计，用每个大学生所有博文的积极情感所占比例衡量大学生个体的幸福感，得到每一位同学的幸福感指数。统计发现大学生幸福指数分布情况如图 6 所示。

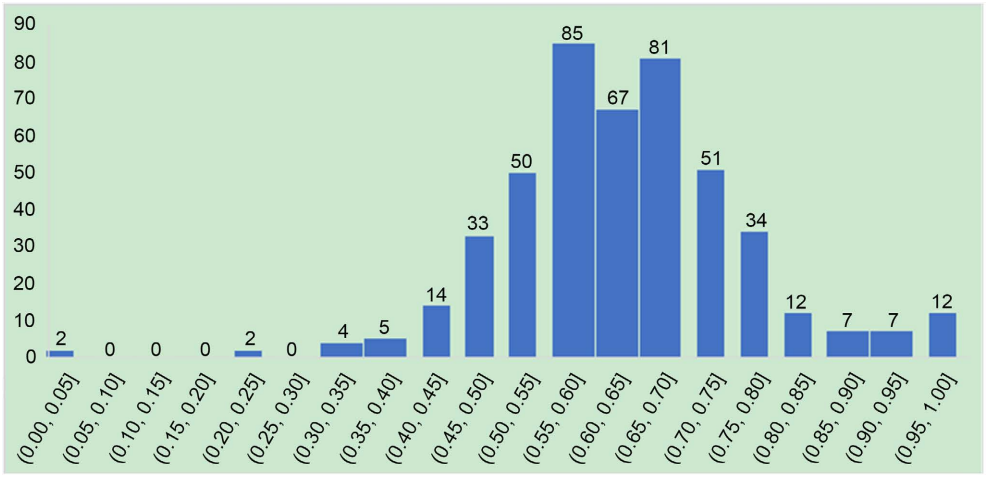

Figure 6. The distribution of individual happiness index of college students
图 6. 大学生个体幸福感指数分布

根据上图结果可以看出，466 个大学生中，有 60 个同学幸福感指数低于 0.50，有 406 名同学幸福感指数高于 0.50，大多数同学的幸福感指数集中在 0.45~0.80 之间。所有同学的幸福感指数的平均值是 0.63，说明同学们整体幸福感比较高，但是在广大人群中会有少数同学存在幸福感低的情况，这些同学需要特殊关注。

利用此模型结果进行大学生幸福感指数计算后，发现近三年大学生幸福感逐年降低，其中每年 7 月份是全年幸福感的低谷，在本文研究的 466 名学生中，406 名同学幸福感指数超过 0.5 水平，60 名学生低于 0.5 水平，说明多数同学幸福感比较高，但仍有少数同学有不幸福的表现；同时在 2021 年中，后半年比前半年不幸福的情况更加明显，学生发表不幸福博文数量增加，并表现出了对期末、实习、就业等日常生活和疫情、封校、车祸等社会或校园事件的关注。

5. 研究结论

本文提出一种利用深度学习模型进行大学生幸福感度量分析的方法，即利用 ALBERT-TextCNN 模型进行情感分类，计算积极情绪占比作为幸福感指数。在公开微博数据集上进行实验，ALBERT-TextCNN 模型进行情感分类时准确率、精确率、召回率和 F1 值均达到较高水平，同时训练时间短成本低。最终本文利用此模型确定了北京某高校 466 名大学生的幸福感指数情况。

基金项目

1) 国家社科基金项目：数据资产价值视角下“网红”影响力及其行为规范研究(编号：21BXW098)；
2) 对外经济贸易大学 2022 年度党建研究课题(项目编号：DJ20220203)。

参考文献

[1] 梁兴辉, 车娟娟. 幸福指数测量方法研究综述[J]. 山西财经大学学报, 2012, 34(S4): 4.

[2] Wang, D., Al-Rubaie, A., Hirsch, B. and Pole, G.C. (2021) National Happiness Index Monitoring Using Twitter for

Bilanguages. *Social Network Analysis and Mining*, **11**. https://doi.org/10.1007/s13278-021-00728-0

[3] 毛良斌. 社交媒体自我呈现与主观幸福感关系的元分析[J]. 现代传播(中国传媒大学学报), 2020, 42(8): 141-148.

[4] 祝琳, 杨志刚. 大学生主观幸福感的干预[J]. 中国健康心理学杂志, 2016(3): 437-440.

[5] 付文宁, 刘冰. 商洛市大学生主观幸福感及其影响因素[J]. 中国健康心理学杂志, 2015, 23(5): 743-746.

[6] 郝乐, 张启望. 幸福指数及其统计测量[J]. 统计与决策, 2020, 36(17): 38-42.

[7] Strotmann, H. and Volkert, J. (2018) Multidimensional Poverty Index and Happiness. *Journal of Happiness Studies*, **19**, 167-189. https://doi.org/10.1007/s10902-016-9807-0

[8] Devlin, J., Chang, M.W., Lee, K., *et al*. (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Computation and Language*, arXiv:1810.04805.

[9] Lan, Z., Chen, M., Goodman, S., *et al*. (2019) ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations. *Computation and Language*, arXiv:1909.11942.

[10] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), Doha, October 2014, 1746-1751. https://doi.org/10.3115/v1/D14-1181