

融合维度构建和数据增强的评教文本匹配算法

殷秀秀, 檀健, 朱金秋, 张诗韵

(南京邮电大学理学院, 江苏南京 210023)

摘要: 高校评教文本存在着评教维度多、文本内容长的特点, 造成了评教信息难以挖掘的问题, 鉴于此, 本文设计了一种融合维度构建与数据增强的无监督评教文本匹配算法。首先, 采用TextRank方法提取评教文本中的关键词, 并根据关键词进行维度归纳与递进, 从而构建评教指标体系。接着, 对评教文本进行短文本拆解, 利用基于注意力机制的预训练模型挖掘短文本与维度间的匹配特征。最后, 在各个预训练模型的基础上, 采取SimCSE策略进行数据增强, 通过对比实验数据, 得到短文本的最佳维度匹配结果。实验结果表明, 使用该策略后的模型在准确率 R_{Acc} 和F1指标上均优于原预训练模型, 其中SimCSE-WoBERT模型匹配效果最好, R_{Acc} 达72.50%, F1达84.06%, 这表明将SimCSE模型引入评教文本匹配领域能取得较好的应用效果。本文算法能够实现评教内容与评教维度的自动化匹配, 从而更精准地挖掘高校评教人员关于各个评教维度的细粒度信息, 便于分析评教人员在听课中重点关注的教学环节, 进而为评教文本细粒度情感挖掘提供理论依据。

关键词: 高校评教; 评教体系; 数据增强; 文本匹配; 数据挖掘

中图分类号: TP391

文献标识码: A

doi: 10.62756/jnuc.issn.1673-3193.2023.07.0012

引用格式: 殷秀秀, 檀健, 朱金秋, 等. 融合维度构建和数据增强的评教文本匹配算法[J]. 中北大学学报(自然科学版), 2025, 46(1): 10-18.

YIN Xiuxiu, TAN Jian, ZHU Jinjiu, et al. Text matching algorithm based on dimension construction and data augmentation for teaching evaluation[J]. Journal of North University of China (Natural Science Edition), 2025, 46(1): 10-18.

Text Matching Algorithm Based on Dimension Construction and Data Augmentation for Teaching Evaluation

YIN Xiuxiu, TAN Jian, ZHU Jinjiu, ZHANG Shiyun

(School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Teaching evaluation text in colleges and universities has the characteristics of multiple evaluation dimensions and long text content, which makes it difficult to mine evaluation information. Based on this, this paper designed an unsupervised teaching evaluation text matching algorithm that integrated dimension construction and data enhancement. Firstly, TextRank method was used to extract keywords from the evaluation text, and the evaluation index system was constructed by dimensional induction and recursion based on the keywords. Secondly, the short text was disassembled, and the pre-training model based on the attention mechanism was used to mine the matching features between the short text and the dimensions. Finally, based on each pre-trained model, the SimCSE strategy was adopted for data enhancement, and by compared the experimental data, the best dimension matching result of the short text was obtained. The experimental results show that the models after using this strategy are better than

收稿日期: 2023-07-16

作者简介: 殷秀秀(1998-), 女, 硕士生, 主要从事应用统计、机器学习方向的研究。

通信作者: 檀健(1989-), 男, 副教授, 博士, 主要从事分析数学、机器学习中的数学理论的研究。E-mail: tj@njupt.edu.cn.

the original training model on the accuracy R_{Acc} and $F1$ indicators. Among them, the Simcse-Wobert model has the best matching effect, R_{Acc} is 72.50%, and $F1$ reaches 84.06%, which indicates that the SimCSE model is introduced into evaluation text matching fields can achieve good application effects. This algorithm can realize automatic matching of teaching evaluation content and teaching evaluation dimensions, thereby can more accurately mine the fine-grained information of university evaluation personnel on each evaluation dimension, which is convenient for analyzing the focus points on the teaching links evaluation of evaluation personnel, and can provide theoretical basis for fine-grained emotion mining of teaching evaluation texts.

Key words: college evaluation; evaluation system; data augmentation; text matching; data mining

0 引言

在高校的教务系统中,记录着大量的教学质量评教文本,这类非结构文本包含的信息多、价值高,但此类文本难以量化,对其研究时需采用一定的技术,并需花费大量的精力。随着自然语言处理技术的发展,使得挖掘和利用评教文本中蕴藏的有价值的信息成为可能。教育文本挖掘是学习分析领域中的一个重要方向^[1],对教学过程中所产生的文本进行分析有助于更好地把握教师的授课态度、内容、方式、成效,可以为精准的教学设计提供证据支持。

目前对教育评教文本的分析主要是利用高校评教系统中的文本进行评教情感分析,旨在评判学生或专家等评教人员对教师授课质量的情感倾向,从而为评估教师授课水平提供参考。该类分析的主要优点是方法简单,获得的情感倾向结果比较直观,容易被受众理解,但由于评教文本往往较长,一条评语中包含多个评教维度的信息,难以获取长文本中教师的授课态度、内容、方式等维度的细粒度信息,进而无法获取评教人员对教师各个维度的细粒度情感,所以需要长文本拆解成短文本,同时将每条短文本与评教维度进行人工匹配非常耗时。为了解决上述问题,需要实现短文本与评教维度的自动匹配,故本文从构建教学质量评教指标体系出发,建立评教内容与评教维度的文本匹配模型来分析评教人员听课中重点关注的教学环节,从而为评教细粒度情感挖掘提供理论依据。

教学质量评教指标体系和评教标准设置的公正性、科学性直接关系到评教的准确性和有效性。建立健全规范的评教指标体系,不断改进评教方法,有利于促进教师创新课堂教学,改善师生关系,提高教学质量^[2]。传统的评教指标以教师为中心,关注的重点是教师自身的知识结构和能力

素养,突出的重点是教师在教学活动中的角色和地位,而忽视了学生的学习主体的作用^[3]。传统的评教指标对师生关系以及学生收获等重视不够,对学生“学”的情况考察不足^[4]。这种情况下,师生之间是传授者与接受者的单向关系,缺少学生的教学反馈。现有的评教指标则考虑学生的主体地位,将学生对教师授课效果的反馈情况纳入评教体系考察的范围内。郭丽君^[5]建立了基于教学态度、教学内容、教学方法与手段、教学效果、教学状态的高校教师教学评教指标体系,并采用模糊综合评价法来验证其指标的有效性。Darling-Hammond^[6]认为现有的评估方法和指标存在局限性和误解,需要采用综合性的评估方法,建立学生有效的反馈和改进机制,加强师生互动和合作,真正提高教师的教学效果和教学质量。刘娟等^[7]在学生评教、同行评教、督导评教中都会不同程度地重视学生的接受程度和满意程度,将“学习收获”“学习获得感”“学习兴趣激发”等作为重要的指标项目。

基于上述研究可以发现,评教指标体系的构建需要综合考虑多个因素,包括教师自身素质、教学方法、教学内容、学生收获等多个方面。因此,本文以理论课的专家评教文本为例,在参考现有指标体系的基础上,通过对评教文本进行关键词提取与归纳,构建了一个拥有7个核心一级维度和对应的24个二级维度的专家评教指标体系。

文本匹配是自然语言处理领域的重要方向,主要研究文本与文本之间的相似度关系。近年来,深度学习技术的快速发展掀起了文本匹配的新篇章。深度学习文本匹配的方法主要分为两种:基于交互的文本匹配方法和基于表示的文本匹配方法。基于交互的文本匹配模型在表示阶段对两个文本进行不同粒度的特征交互,再加权融合得到最终的匹配分数,如ARC-II^[8]、Match-

SRNN^[9]等,但此类模型忽略了句型、句间关系等全局性信息;基于表示的文本匹配模型以经典的双塔模型为代表,将需要匹配的两个文本分别编码成高级表示,通过计算这些表示之间的相似度或距离来进行文本匹配,经典的模型如DSSM^[10]、MV-LSTM^[11]等,但是,此类模型容易失去语义焦点,难以把握词的上下文的重要性。在表示型结构中,基于注意力机制^[12]和Transformer架构^[13]的预训练模型BERT^[14]表现优异,该模型可以对输入序列进行自适应加权,以捕捉输入序列中的相关信息。BERT引起学术界关注后,许多基于BERT的改进预训练模型相继被提出。与BERT的静态掩码策略不同,RoBERTa模型引入了一种动态掩码策略,从而能更好地学习上下文信息。另外,Sentence-BERT是一种基于BERT的模型,专门用于处理句子级别的语义相似度任务,通过对BERT进行微调,该模型能够在无监督的情况下学习句子级别的语义表示。

这些基于BERT家族模型表示的方法比传统的交互式方法更加高效和准确,因为这些方法充分利用了文本序列中的所有信息,如语义和上下文信息,从而能够更好地理解文本之间的关系,使得文本匹配效果有了明显提升。马新宇等^[15]在10个通用短文本匹配数据集上,验证微调BERT模型的迁移与泛化能力。王辉等^[16]提出了BERT-Attention-MultiBiGRU模型,将新冠疫情医疗短文本中两个问句交互的信息合并到多层连接的BiGRU中用于相似度匹配,实现了医疗短文本相同语义问句的自动精确识别。赵伟等^[17]提出了基于BERT预训练模型及多视角循环神经网络的文本匹配模型,用于电力文本匹配研究。

本文研究的评教文本匹配是面向短文本内容的匹配,研究的核心任务是通过语句的表示及其相似程度的度量将短评教文本和设定好的评教维度进行无监督匹配。本文对评教文本进行无监督短文本匹配研究,主要是由于长文本研究缺乏对每个评语中各维度语义信息的挖掘利用,同时,现有的无监督短文本匹配主要基于词频或者基于交互的方式,缺乏对短文本结构和语义信息的表示能力。另外,评教短文本数量多,难以大量人工标注,而基于深度神经网络的模型在无标签数据中的学习很难。在文本匹配过程中,BERT编码生成的句向量受到句子中的词频影响,导致高频词主导句向量,使得任意两个句向量之间的相似度都非常高,造成高频词编

码的句向量紧密,低频词编码的句向量稀疏的现象。这导致向量表达呈各向异性,各维度的特征表示不一致,难以直接比较句子级别的特征向量。为了解决以上问题,本文从无标签的评教短文本数据出发,以预训练BERT等模型为基础,引入SimCSE对比学习的思想,将相似的评教样本拉近,将不相似的评教样本推远,采用dropout机制,构造正例样本来增强数据,可有效解决向量表达存在各向异性以及向量分布不均匀的问题,能更充分地学习到良好的句向量表示,从而有效解决文本表示退化的问题,进而达到增强评教文本和评教维度匹配效果的目的。

1 高校专家评教文本匹配问题描述

评教文本无监督匹配是对输入的短文本,从评教指标体系中选择语义相似度最高的维度进行匹配。为了更好地理解匹配的过程,本节给出了如下的模型问题的定义。

评教文本匹配问题可建模为评教短文本的维度预测问题。训练一个三元组 $\langle S, D, label \rangle$,其中, S 表示评语短文本 sentence, D 表示评教维度 dimensionality, $C^q = \{S_1^q, S_2^q, \dots, S_p^q\}$ 表示第 q 个评教文本的短文本集合(由预处理数据后的短句组成), $D = \{d_m^n\}$ 表示 m 个一级维度下的 n 个二级维度内容集合。 $\max Sim(S_i^q, d_m^n)$ 表示第 q 个文本评语中的短文本 i 与各二级维度内容匹配的最大相似度值, $label = index_{\max Sim(S_i^q, d_m^n)}$ 表示最大相似度值所对应的一级维度索引值为该短文本的标签值。匹配模型 $y = f(S_i^q, D_m, label)$ 将预测第 q 个文本的短文本 i 是否属于第 m 个一级维度下的内容。短文本的一级真实维度标签为 $label'$,如果 $label = label'$,则认为匹配成功,否则匹配失败。

2 高校专家评教文本匹配问题建模

本研究的目标是完成评教短文本与所构评教维度内容的匹配。为了有效利用短文本的语义信息,提高与维度内容匹配的精确度,文本引入了一种基于无监督SimCSE对比学习的评教文本匹配方法。这种无监督学习方法不需要利用标记的数据来进行训练,而是通过将对比学习嵌入到预训练的BERT模型中,从未标记的文本中学习语义相似性,该方法不仅可以解决数据缺乏标签的问题,还可以对短文本进行数据增强。

2.1 基于对比学习的评教文本数据增强

基于预训练模型的无监督 SimCSE 模型框架如图 1 所示,实线箭头指对同一个句子应用不同的 dropout masks 产生的两个相近的表示,虚线箭头指不相近的表示。SimCSE 模型是一种简单的对比句向量表示的框架。对比学习的思想就是拉近相似样本的距离,推远不相似样本的距离。

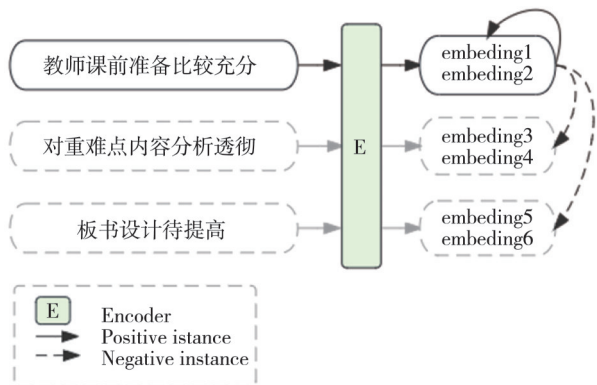


图 1 SimCSE模型架构

Fig.1 SimCSE model architecture

SimCSE 采用了两个主要的模块:文本嵌入模块和相似度计算模块。文本嵌入模块中引入对比学习的思想,使用预训练语言模型(如 BERT、RoBERTa 等)对文本进行编码,得到文本的向量表示。 N 个句子经过含 dropout 的 Encoder 后得到句向量 $h_1^{(0)}, h_2^{(0)}, \dots, h_N^{(0)}$ 。在句向量 $h_i^{(0)}$ 中,其他的句向量 $h_{j, j \neq i}^{(0)}$ 都是 $h_i^{(0)}$ 的负例。正例的构建则是采

用不同的 dropout 机制,让这 N 个句子重新经过一遍 Encoder,得到句向量 $h_1^{(1)}, h_2^{(1)}, \dots, h_N^{(1)}$,则 $(h_i^{(0)}, h_i^{(1)})$ 为一对正例样本。相似度计算模块使用余弦相似度计算文本之间的相似度,然后使用基于 infoNCE loss 的对比学习损失(Contrastive Learning Loss)来训练模型^[18]。对比损失的目标是让相似的文本向量更加靠近,不相似的文本向量更加远离。在一个 batch 中,样本 i 的训练目标为

$$-\sum_{i=1}^N \sum_{\alpha=0,1} \log \frac{e^{\cos(h_i^{(\alpha)}, h_i^{(1-\alpha)})/\tau}}{\sum_j e^{\cos(h_i^{(\alpha)}, h_j^{(1-\alpha)})/\tau} + \sum_{j=1, j \neq i}^N e^{\cos(h_i^{(\alpha)}, h_j^{(\alpha)})/\tau}}, \tag{1}$$

式中: τ 为温度超参,用于调节对困难样本的关注程度; N 为一个 batch 的大小; \cos 函数表示余弦相似度的计算; $1 - \alpha$ 表示 α 通过不同 dropout 得到的增强样本。

训练目标中的分母不仅对原样本和所有增强样本进行了累加,也将一个 batch 中不同的原样本进行了累加。

2.2 基于 BERT 等的预训练文本匹配模型

不同于传统的文本单向表示模型, BERT 使用双向 Transformer 结构对语言模型进行训练,是一个典型的基于预训练的文本表示模型,可以获得文本语言更深层次表示信息。Transformer 模型结构如图 2 所示。

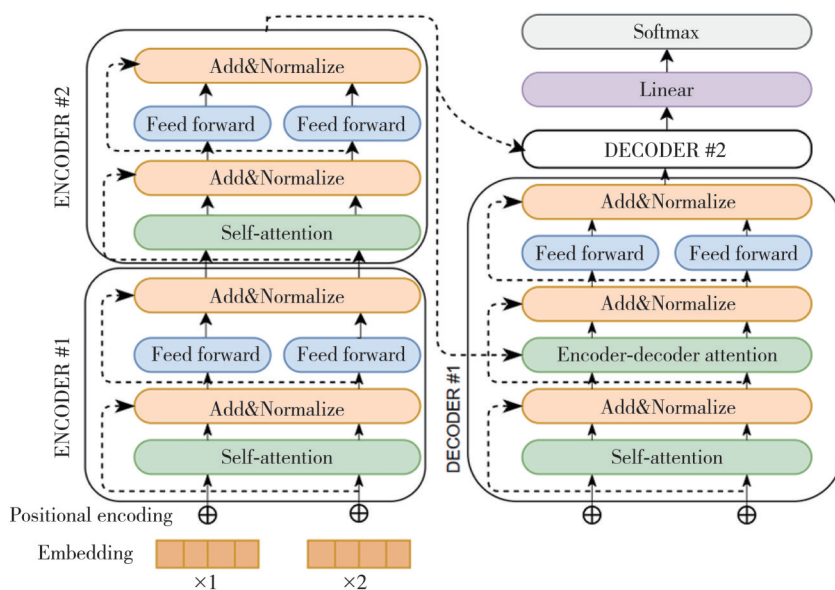


图 2 Transformer模型结构

Fig.2 Transformer model structure

Transformer 是一个 Encoder-Decoder 架构。图 2 中,左侧为编码器(ENCODER),由多个编码器层叠加而成,实线圈 ENCODER #1 为其中一个编码层,右侧为解码器(DECODER),由多个解码器层叠加而成,实线圈 DECODER #1 为其中一个解码层。编码器层包括 2 个子层,分别是自注意力层(Self-Attention)和全连接前馈神经网络层(Feed Forward)。解码器层包括 3 个子层,分别是自注意力层(Self-Attention)、第 2 个自注意力层(Encoder-Decoder Attention)和全连接前馈神经网络层。编码器与解码器中每个子层的输出会通过残差连接和层归一化(Add&Normalize)进行连接,以加速训练和提高模型性能。

x_1, x_2 作为词嵌入向量输入的底部编码器,编码器的输入会先进入自注意力层,其输出和位置编码的向量经过残差连接与层归一化后输入前馈神经网络层,该输入与前向传播后的输出再进行残差连接与层归一化,其结果作为解码器中第 2 个注意力层的一项输入。在解码器第 2 个注意力层中,会通过交叉注意力(Cross-Attention)将解码器当前位置的输出与编码器的输出进行交互。解码器后面一层是输出线性层(Linear),将解码器层

的输出映射为词汇表大小的向量,然后通过 Softmax 函数将该向量转换为概率分布,得到最终的输出概率。本文中的 BERT 是由 12 层双向 Transformer 编码器组成的模型,每个编码器的结构都是相同的,可以使用不同的权重参数。这样的双向编码器更能充分考虑词语的上下文信息,从而提取更加准确的词向量表示。这也是 BERT 模型在获取文本语义信息和上下文信息时,其性能优于传统卷积神经网络和循环神经网络的原因。

本文在获取评教文本的向量化表示时,使用 jieba 分词工具。在使用 WoBERT 模型时,采用精确模式和非 HMM 模式将句子切开,可以使文本不存在冗余,返回一个列表的精确结果。以“引导学生课上积极讨论”为例,通过分词工具得到的结果为“引导/学生/课/上/积极/讨论”,接着将分词结果作为模型的输入。每个词语在 BERT 模型中经过 3 种编码嵌入表示:词嵌入向量(Token Embeddings)、文本嵌入向量(Segment Embeddings)、位置嵌入向量(Position Embeddings),词语的最终表示由 3 种嵌入向量相加 2 次。BERT 输入如图 3 所示。标记[CLS]作为每一个句子的开始,[SEP]作为句子的结束或者句子之间的分割符。

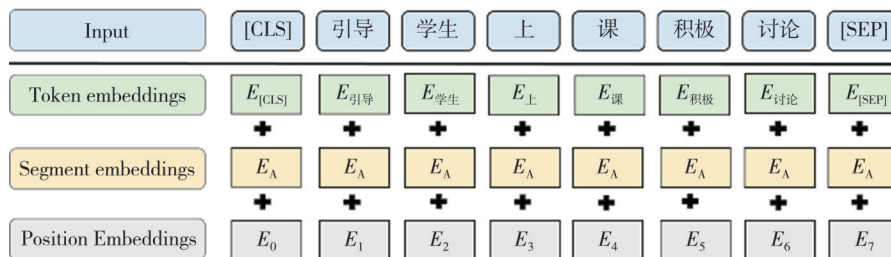


图 3 BERT 输入示例

Fig. 3 BERT input example

WoBERT 模型是以词为单位的中文 BERT 模型,针对中文语言的特点,将词汇按照语义相关性进行分组,从而可以更好地捕捉词汇之间的关联关系。同时,以词为单位,序列变短,处理速度更快。在 Masked Language Modeling 任务中,与 BERT 相比, WoBERT 不仅考虑了当前 token 的上下文信息,还考虑了该 token 的词性和位置等信息,从而更好地模拟了中文语言的特点。

3 实验过程与结果

3.1 数据准备

实验数据集采用某高校评教系统的专家评语,选取 5 296 条理论课专家评语样本,见表 1 中拆分前

示例。在预处理阶段去除评语中的空数据和无意义字符(如“”,“!!!”,“啊啊啊”等),数据预处理后的文本合并为一个总的评语文档,词性选择名词、动词、动名词以及名词化词组。采用 TextRank 关键词提取方式得到评教文本共计 5 799 个词,词汇得分最高为 0.389,得分的计算公式详见 3.2.1 节式(2),所有词汇得分分布如表 2 所示。在选取主要关键词数量恰当的情况下,结合词汇重要性得分分布,本文选取得分大于 0.04 的 77 个词汇为主要关键词。

从预处理后的 5 296 条评语样本中随机选取 3 000 条用于文本拆分。考虑本文评教数据中句子长度过长的特点,如表 1 拆分前示例,一条评语中往往包含多个维度的评教信息,为了挖掘每条

评语的细粒度评教信息, 需要把评语拆分成短文本。同时, 为了使得各短文本评教维度的界限明确, 本文首先按逗号分隔出较短的语法单元, 再按句号分隔, 进一步分化不同的语句场景。拆分完后, 去除短文本中与评教维度无关的数据(如“本节课为……”, “本次课主要内容是……”)后得到 30 791 条短文本, 见表 1 中拆分后示例。

表 1 专家评教文本拆分示例

Tab. 1 Example of expert evaluation text splitting

拆分前	拆分后
本次课主要讲授“换元积分法”。教师课前准备认真, 讲课精神饱满, 内容熟练, 对换元积分法的思路、方法和技巧的讲解条理清晰, 注重启发, 课件清晰, 多媒体课件的运用与板书的配合恰当, 学生听课注意力集中, 课堂气氛和谐。	① 教师课前准备认真 ② 讲课精神饱满 ③ 内容熟练 ④ 对换元积分法的思路、方法和技巧的讲解条理清晰 ⑤ 注重启发 ⑥ 课件清晰 ⑦ 多媒体课件的运用与板书的配合恰当 ⑧ 学生听课注意力集中 ⑨ 课堂气氛和谐

表 2 词汇重要性得分分布

Tab. 2 Score distribution of vocabulary importance

得分段	词汇数量
≥ 0.2	8
[0.1, 0.2)	22
[0.08, 0.1)	12
[0.06, 0.08)	12
[0.04, 0.06)	23
[0.02, 0.04)	76
[0, 0.02)	5 646

评语句子的长度和短文本的长度以字符数来计量, 文本匹配实验中选取的 3 000 条评语样本的最大长度为 349, 平均长度为 111, 图 4 为 3 000 条评语样本按长度段划分的数量分布, 局部图为句子长度大于 200 的数量分布, 图中句子长度 320~340 的数量为 1。

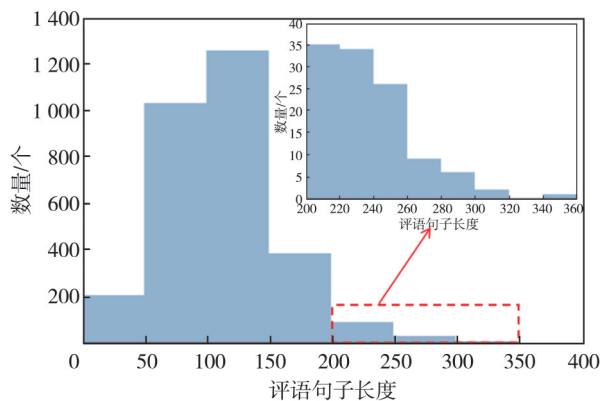


图 4 句子长度分布

Fig. 4 Sentence length distribution

30 791 条短文本最大长度为 71, 平均长度为 9, 图 5 为短文本按长度段划分的数量分布, 图中短文本长度 55~60 的数量和长度 65~70 的数量均为 0。

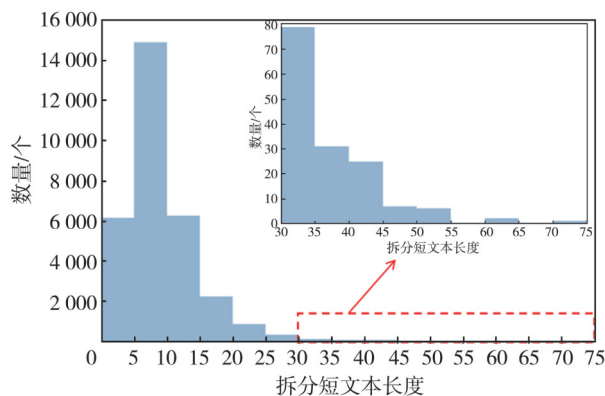


图 5 短文本长度分布

Fig. 5 Distribution of short text length

通过对评教内容的分析, 本文实验将以二级维度评语为参考, 对每条拆分后的短文本描述创建基于一级维度的标签。

3.2 评教维度的构建

3.2.1 TextRank 关键词提取

TextRank 算法提取关键词的基本思想是将文本中的每个词作为无向图中的节点, 利用文本内部的词语间的共现语义信息对节点进行排序, 取得分由高到低的若干个单词作为关键词。其具体步骤描述如下:

1) 把给定的文本 T 进行分词, 即 $T = \{t_1, t_2, \dots, t_n\}$;

2) 去除 T 中的停用词, 保留名词、动词、动名词、名词化词组等, 即 $V = \{v_1, v_2, \dots, v_m\}$;

3) 构建候选关键词 $G = (V, E)$, 其中 V 为节点集, 由 2) 生成的候选关键词组成, $E \subseteq V \times V$, 对关键词图中的任意一个节点 $v_{i,j} \in E, V_{i,j+1} \in E$, 有 $\langle v_{i,j}, v_{i,j+1} \rangle \in E$ 。令 $In(v_i) = \{v_j | \langle v_j, v_i \rangle \in E\}$, $Out(v_j) = \{v_i | \langle v_i, v_j \rangle \in E\}$, 其中, $In(v_i)$ 代表指向节点 v_i 的节点的集合, $Out(v_j)$ 代表节点 v_j 指向的节点的集合。用 ω_{ji} 表示节点 v_j 到节点 v_i 所在边的权重, 代表两个节点之间的边连接的重要程度, 则节点 v_i 的分数

$$WS(v_i) = (1 - d) + d \sum_{v_j \in In(v_i)} \frac{\omega_{ji}}{\sum_{v_k \in out(v_j)} \omega_{jk}} WS(v_j), \quad (2)$$

式中: d 为调节系数, 通常设为 0.85。

根据式(2)迭代传播各节点的权重,直至收敛。对节点权重进行倒序排序,得到最重要的 N 个单词,作为候选关键词。

由上可以看出,步骤1)对给定的语料 T 进行分词是TextRank算法提取关键词的基础。该算法提取的关键词若形成相邻词组,则组合成多词关键词,使得关键词提取更加合理。

3.2.2 评教维度指标结果

通过TextRank算法提取得分排名前77的主

要关键词,将关键词拓展为二级维度评语,用于后续语句匹配。然后根据关键词依次向更高一级的维度递进归纳,创建评教维度指标体系,过程如表3和表4所示。

表3 关键词拓展语句示例

Tab.3 Examples of keyword expansion statements

二级维度评语	关键词
教师精神饱满,有敬业精神	精神饱满、敬业精神、责任心
教师课前准备充分	课前、准备充分
信息化教学,运用多媒体资源	多媒体、信息化、视频

表4 评教维度体系

Tab.4 Teaching evaluation dimension system

一级	二级	关键词	Label
教学态度	教学精神	精神饱满、敬业精神、责任心	1
	教姿教态	端正、严谨、态度端正	1
	课前准备	课前、准备充分	1
教学内容	教学语言	表述、语速、语言表达、感染力、清晰	2
	知识素养	教学内容、熟练、讲授、透彻	2
	内容逻辑	内容、重难点、重点难点、层次、逻辑性、针对性、条理性	2
	教学信息	信息量	2
	结合应用	理论、实际、相结合、实践、案例、应用	2
教学方式	教学方法	方法灵活、灵活应用、分组	3
	引导启发	引导、启发性、兴趣	3
	能力培养	注重、学生、能力、分析	3
	组织安排	组织、时间安排、合理、完整性	3
	驾驭课堂	课堂、驾驭能力、经验丰富	3
教学资源	教学资源	教学资料、齐全	4
	工具设计	课件、板书设计	4
	资源应用	多媒体、信息化、视频	4
课堂气氛	课堂氛围	活跃、课堂秩序	5
	参与互动	积极主动、参与度、互动性、分组讨论	5
	学生态度	听课认真、记笔记、思考问题	5
	精神面貌	打瞌睡、集中、注意力	5
师生关系	对待学生	关注、理解、掌握情况、知识点	6
	学习指导	指导、纠正、课堂练习	6
学生收获	知识内化	学习效果	7
	教学效果	教学效果、课堂教学	7

注: Label为一级维度的标签; 标签号码为一级维度索引值。

3.3 实验设置

为了验证SimCSE模型在评教文本匹配上的有效性,本文基于高校专家评教文本数据集,采用SimCSE方法,对比分析了SimCSE模型与开源的BERT模型及其相关变体模型RoBERTa、WoBERT、RoFormer、SimBERT等当前主流的预训练文本匹配模型在评教文本匹配任务中的性能。

3.4 实验环境与参数设置

3.4.1 环境配置

实验利用深度学习框架Tensorflow和Keras

来实现模型网络结构,并在同一环境下进行训练、验证和测试。实验的硬件平台配置为32G内存和RTX 3090(24GB)GPU,操作系统为Linux,使用的Python版本为3.8.10, Tensorflow版本为1.14, keras版本为2.3.1。

3.4.2 参数设置

模型的复杂程度通常与Transformer编码器的层数设置有关,在效果与计算速度均衡的情况下,选择base版本,即编码器层数为12,隐藏单元数为768,注意力头数为12。模型训练过程中,设置batch size为64, dropout为0.1,采用Adam优化器,设置学习率为 1×10^{-5} ,采用first-last-avg

池化方法,即把 encoder 的第一层与最后一层的所有向量取平均。为了避免正例对长度相同导致语义性更相似的问题,根据短文本集的长度统计,设置每个句子的最大长度为 38。模型在获取数据时会截断超过文本长度的句子,或为长度不足的句子填充“0”,确保所有句子对长度相同。

3.5 评价方法

模型采用预测准确率(Accuracy, R_{Acc})与综合评价指标($F1$ -Measure)作为评价指标,对实验结果进行定量分析。 R_{Acc} 是分类任务常用的一种评价指标, $F1$ 值是对模型查准率(Precision, R_{Pre})和召回率(Recall, R_{Rec})的一种加权调和平均。

$$R_{Acc} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{FP} + N_{TN}}, \quad (3)$$

$$F1 = 2 \times \frac{R_{Pre} \times R_{Rec}}{R_{Pre} + R_{Rec}}, \quad (4)$$

式中: N_{TP} 表示成功匹配上的评教维度的样本数量; N_{FN} 表示将匹配正确维度的样本判断为匹配错误的样本数量; N_{FP} 表示将匹配错误维度的样本判断为匹配正确的数量; N_{TN} 表示将匹配错误维度的样本判断为匹配错误的数量; R_{Pre} 表示模型预测为某维度且预测正确的样本占有所有预测为该维度的样本的比例; R_{Rec} 表示模型预测为某维度且预测正确的样本占有所有真实为该维度的样本的比例。

$$R_{Pre} = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (5)$$

$$R_{Rec} = \frac{N_{TP}}{N_{TP} + N_{FN}}. \quad (6)$$

3.6 实验结果分析

本文设计了文本匹配预训练模型BERT及其相关变体模型,在此基础上分别采用SimCSE策略,进行10种模型的对比实验。实验结果如表5所示。

为了验证模型在不同数据集上的泛化性能,选取表5中结果最佳的模型SimCSE-WoBERT,从总的理论课文本中再随机选取两次3000条文本数据进行验证,结果如表6所示。

表5列出了采用SimCSE策略的模型与其原模型在评教文本数据集上的实验结果对比。本实验输入的匹配数据为评教内容文本与评教维度文本,评教内容文本为长文本通过相关策略拆分后的短文本,评教维度文本为基于维度关键词拓展后的中短文本。本文将SimCSE策略引入到评教文本匹配的预训练

模型中,在其数据集上实验的 R_{Acc} 和 $F1$ 指标均超过了各自的原模型。SimCSE模型不仅考虑了短文本的语义信息和结构信息,还可以在无标签数据的情况下,生成质量较好的句子向量,从而实现增强样本的目的。在相同的参数设置下,经过样本增强的SimCSE-WoBERT模型达到了最好的效果, R_{Acc} 提升至72.50%, $F1$ 提升至84.06%。表6的结果验证了SimCSE-WoBERT模型在不同理论课样本数据集上的泛化性能,充分证明了增强样本的SimCSE对比学习方法在预训练匹配模型上能够更加有效地学习评教文本数据的表示,从而能够提高短文本匹配的效果。

表5 模型对比实验结果

Tab. 5 Comparison experimental results of model

模型	R_{Acc}	$F1$
BERT	0.287 4	0.446 4
RoBERTa	0.254 1	0.405 2
WoBERT	0.176 6	0.300 2
RoFormer	0.291 0	0.450 8
SimBERT	0.381 4	0.552 3
SimCSE-BERT	0.657 8	0.793 5
SimCSE-RoBERTa	0.677 9	0.807 8
SimCSE-RoFormer	0.688 8	0.815 7
SimCSE-SimBERT	0.518 8	0.683 1
SimCSE-WoBERT	0.725 0	0.840 6

表6 不同样本数据集上SimCSE-WoBERT模型的实验结果

Tab. 6 Experimental results of SimCSE-WoBERT model in different sample data sets

实验	R_{Acc}	$F1$
Test1	0.725 0	0.840 6
Test2	0.715 5	0.835 6
Test3	0.718 2	0.837 1

注:Test1为表5中的实验结果;Test2、Test3为补充的两次验证结果。

4 结束语

本文介绍了评教文本维度匹配的必要性,并分析了教育评教维度和文本匹配的现状和挑战。在主流的评教维度的基础上采用关键词提取归纳的方式构建了本文的专家评教维度体系,然后针对匹配任务需求与现有模型的优缺点,将SimCSE对比学习模型引入到预训练模型中,实现了教育评教文本自动化的无监督文本匹配。在文本匹配模型中,对拆分后的评教短文本和评教维度分别进行编码表示,通过BERT模型中的注意力机制,增强了评教文本的语义特征;引入SimCSE解决了BERT模型在匹

配过程中的各向异性和均衡性的问题,实现了语义增强。将各BERT及其相关变体模型与使用SimCSE策略后模型的匹配效果进行对比实验,结果表明,采取SimCSE策略后的模型效果均有提升,其中,SimCSE-WoBERT在 R_{acc} 和F1上的表现最好,有效提升了评教文本匹配的效果,有利于后续计算分析专家的评教情感倾向和调整教师授课过程中的侧重点。尽管本文提出的评教文本无监督匹配方法提升了评教维度匹配的效果,但还是难以达到有监督的匹配效果,如果条件和人工成本允许,采用高质量的标注能够进一步提高匹配的精确度。

参考文献:

- [1] 王萌,符雅茹,牟智佳. 国际教育文本挖掘研究热点与前沿透视[J]. 开放学习研究, 2021, 26(3): 17-27.
WANG Meng, FU Yaru, MOU Zhijia. The research hotspot and frontier perspective of educational text mining [J]. Journal of Open Learning, 2021, 26(3): 17-27. (in Chinese)
- [2] 侯建,陈建成,宋洪峰,等. 教学督导和学生评教对高校教师教学质量的影响研究——以北京林业大学为例[J]. 中国林业教育, 2020, 38(6): 28-34.
- [3] 吕鹏. 以学生为中心的学评教指标体系构建[J]. 现代教育管理, 2014(3): 42-45.
LÜ Peng. Study on construction of student-centered course assessment indicator system [J]. Modern Education Management, 2014 (3): 42-45. (in Chinese)
- [4] 马力,王晓君,柳兴国. 高校教师教学质量评价指标体系与方法[J]. 统计与决策, 2005(8): 68-69.
- [5] 郭丽君. 教师发展视野下的高校教学评价制度[M]. 北京: 中国社会科学出版社, 2021.
- [6] DARLING-HAMMOND L. Getting teacher evaluation right: What really matters for effectiveness and improvement [M]. New York: Teachers College Press, 2015.
- [7] 刘娟,项林川,柯昌剑,等. 高校教学评价的实践与发展研究——以7所部属高校为例[J]. 高教学刊, 2021(4): 11-15.
- [8] WAN S X, LAN Y Y, GUO J F, et al. A deep architecture for semantic matching with multiple positional sentence representations [DB/OL]. (2015-11-26) [2023-07-12]. <http://arxiv.org/abs/1511.08277v1>.
- [9] WAN S X, LAN Y Y, XU J, et al. Match-srnn: Modeling the recursive matching structure with spatial RNN [DB/OL]. (2016-04-15) [2023-07-12]. <https://arxiv.org/abs/1604.04378>.
- [10] JADHAV S S, THEPADE S D. Fake news identification and classification using DSSM and improved recurrent neural network classifier [J]. Applied Artificial Intelligence, 2019, 33(12): 1058-1068.
- [11] WANG S H, JIANG J. Learning natural language inference with LSTM[DB/OL]. (2015-12-30)[2023-07-12]. <https://arxiv.org/abs/1512.08849>.
- [12] PARIKH A P, TÄCKSTRÖM O, DAS D, et al. A decomposable attention model for natural language inference [DB/OL]. (2016-06-06) [2023-07-12]. <https://arxiv.org/abs/1606.01933v1>.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems December, 2017: 6000-6010.
- [14] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [DB/OL]. (2018-10-11) [2023-07-12]. <https://arxiv.org/abs/1810.04805>.
- [15] 马新宇,范意兴,郭嘉丰,等. 关于短文本匹配的泛化性和迁移性的研究分析[J]. 计算机研究与发展, 2022, 59(1): 118-126.
MA Xinyu, FAN Yixing, GUO Jiafeng, et al. An empirical investigation of generalization and transfer in short text matching [J]. Journal of Computer Research and Development, 2022, 59 (1): 118-126. (in Chinese)
- [16] 王辉,章建聪,李志文. 基于BERT-Attention-MultiBiGRU模型的医疗短文本相似度匹配研究 [C]//第十六届全国信号和智能信息处理与应用学术会议, 2022: 8.
- [17] 赵伟,王文娟,甘玉芳. 基于预训练模型和多视角循环神经网络的电力文本匹配模型[J]. 重庆邮电大学学报(自然科学版), 2023, 35(3): 545-553.
ZHAO Wei, WANG Wenjuan, GAN Yufang. Electric power text matching model based on pre-training model and multi-view recurrent neural network [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(3): 545-553. (in Chinese)
- [18] Gao T, Yao X, Chen D. Simcse: Simple contrastive learning of sentence embeddings[DB/OL]. (2021-04-18) [2023-07-12]. <https://arxiv.org/abs/2104.08821v2>.