

基于多任务学习的同行评审论文接收预测

蔡涛, 檀健, 杨珂, 秦一天

南京邮电大学理学院, 江苏 南京

收稿日期: 2024年4月21日; 录用日期: 2024年5月21日; 发布日期: 2024年5月28日

摘要

同行评审论文接收预测是一项具有重要意义任务, 其有效提升了同行评审的效率和质量。以往的同行评审论文接收预测方法大多以单任务的形式研究, 并未充分利用论文评分等其它辅助信息, 同时也未有效提取同行评审文本的语义特征。针对上述问题, 文中提出了一种多任务同行评审文本分析模型 BCLJ(BERT-CNN-LSTM-Joint Model, BCLJ)。首先, 使用BERT作为词向量获得文本的矩阵表示; 然后, 引入卷积神经网络(Convolutional Neural Network, CNN)和长短期记忆网络(Long Short-Term Memory Network, LSTM)进行语义特征的提取, 并运用注意力机制增强对文本信息的理解能力; 最后, 利用不同的全连接层进行多任务学习, 获得论文接受预测和评分预测两种输出, 通过评分预测任务来优化主分类任务。实验结果表明, 多任务模型在论文接收预测任务和评分预测任务中表现出色高于其他的基线模型, 在论文接收预测任务中准确率达到0.7117, F1值达到了0.7101, 在论文评分预测任务中MSE, RMSE和MAE分别为1.3690, 1.1700和0.9324。

关键词

同行评审, BERT, 多任务学习, 长短期记忆神经网络, 卷积神经网络, 注意力机制

Prediction of Peer-Reviewed Paper Acceptance Based on Multi-Task Learning

Tao Cai, Jian Tan, Ke Yang, Yitian Qin

School of Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Apr. 21st, 2024; accepted: May. 21st, 2024; published: May. 28th, 2024

Abstract

Peer-reviewed paper acceptance prediction is a task of great significance, which effectively improves the efficiency and quality of peer review. Most of the previous peer-reviewed paper accep-

tance prediction methods are in the form of single-task research, which do not make full use of other auxiliary information such as paper ratings, and do not effectively extract the semantic features of peer-reviewed text. To address the above problems, a multi-task peer-review text analysis model BCLJ (BERT-CNN-LSTM-Joint Model, BCLJ) is proposed in the paper. First, BERT is used as the word vector to obtain the matrix representation of the text; then, Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) are introduced for semantic feature extraction, and the attention mechanism is applied to enhance the comprehension of the textual information; finally, multi-task learning is performed by utilizing different fully-connected layers, and two outputs of paper acceptance prediction and scoring prediction are obtained to optimize the main classification task through the rating prediction task to optimize the main classification task. The experimental results show that the multi-task model performs better than other baseline models in the paper acceptance prediction task and the rating prediction task, with an accuracy of 0.7117 and an F1 value of 0.7101 in the paper acceptance prediction task, and MSE, RMSE, and MAE of 1.3690, 1.1700 and 0.9324 in the paper rating prediction task, respectively.

Keywords

Peer Review, BERT, Multi-Task Learning, LSTM, CNN, Attention Mechanisms

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着科技的发展,科技论文投稿数量不断增加,确保论文科学性、可靠性和创新性至关重要。同行评审作为推动科技领域发展的基本学术过程[1],其透明度和可信度仍有待改进。在线评审平台的兴起使同行评审逐渐向更加公开透明的方向发展。自动评审分类技术的应用可为编辑减轻工作负担,提供决策支持,并节省时间和精力。在文献中,确定论文最终是否被接受通常被视为文本分类任务,而对专家评分的预测被视为多分类或回归任务[2]。本文将专家评分预测作为回归任务。目前主流的自动评审分类方法包括机器学习和深度学习。例如,Keith [3]等人使用支持向量机等机器学习算法并结合 TF-IDF 词向量模型来预测国际会议的研究论文评审结果。Ribeiro [4]等人使用了人工智能会议和机器人会议的 2313 篇评论文本作为数据集,基于回归和分类任务的七种机器学习方法以及 VADER 应用进行预测论文的录用和推荐得分。实验证明,基于传统的机器学习模型进行自动评审分类时难以有效地捕捉和表示大量的复杂语义和上下文信息,因此预测的效果较差。随着深度学习的快速发展,同行评审分类方法通过深度模型进行分类。Li [5]等人结合了摘要-评审匹配机制和交叉评审匹配机制,从同行评审文本中学习深度特征来进行引文计数预测。Leng [6]等人提出了一种基于 CNN 的 DeepReviewer 模型,用于自动论文审阅。该模型通过学习文章的语义、语法和创新特征,并将这些特征通过一个注意力层整合,得出论文的最终评分。Deng [7]等人设计了一种用于论文评审评级预测和推荐的分层双向自注意力网络框架 (HabNet),作为学术论文评审过程的有效决策工具。该框架利用论文评论的层次结构,包括句子、审稿内和审稿间级别的编码器,通过双向自注意力机制从不同级别的评论中提取有用的语义信息。Chen [8]等人提出了一种利用科技论文语义信息和开放同行评审中专家评分的自动评审分类方法,发现融合语义信息和评分信息的评审分类模型效果优于仅依靠评分均值进行评审判断。Lin [9]等人提出了一种整合定性评价的论文质量评价模型,利用论文标题、摘要、Twitter 评论和同行评议作为输入,以论文审稿平均

分为标准，运用情感分析模型探索开放获取文本的情感极性，并据此对论文进行评价。Li [10]等人提出了一个多任务学习模型用于自动预测学术论文同行评审的分数，并采用附加信息以提升预测性能。通过引入多任务共享结构编码方法，实现了自动选择合适的共享网络结构和辅助资源。Zhu [11]等人构建了一个多任务学习模型 BLBC，通过在 BERT-LCF 模型的基础上增加 BiLSTM-CRF 模块，从而使其具备了同时完成属性词抽取和细粒度情感分析任务的能力。这些研究虽然在一定程度上解决了同行评审自动分类任务和专家评分预测任务，但是尚未有人将这两个任务同时结合。然而，评分的高低与是否接收之间存在一定关联，这导致在论文接收预测方面存在一定的局限性。不仅如此，若要同时解决同行评审自动分类任务和专家评分预测任务，就需要构建两个模型，这会极大地增加模型占用空间和复杂度[12]，从而造成模型间的误差。

为应对上述问题，本文提出了一种同行评审文本分析模型 BCLJ。该模型利用多任务学习框架，把论文接收预测作为主任务，同时将专家对论文的评分预测作为辅助任务，利用 BERT 作为预训练词向量，同时利用共享的 CNN-LSTM 层，实现对论文接收预测和评分预测的双重任务。

本文的主要贡献如下：(1)本文采用多任务学习的方式对论文接收与评分进行预测，相较于单任务学习，多任务学习使用共享参数提升了两个任务的预测效果。(2)本文首次在同行评审中将论文接收预测和评分预测结合考虑，利用专家对论文的评分作为辅助信息来影响论文是否被接收的决策。(3)本文首次在同行评审论文接收和评分预测中引入多任务学习方法，为该领域研究提供了新的思路和方法。

2. 模型构建

本文提出的 BCLJ 模型任务模型主要由三个组件组成：输入层、共享模块和特定任务层，模型的结构如图 1 所示。该模型的任务为给定一个同行评审句子序列 $S = \{w_1, w_2, \dots, w_n\}$ ，其中 w_i 为句子中的单词， n 为句子的长度。模型的输出结果包括论文是否被接受以及对论文评分的预测。

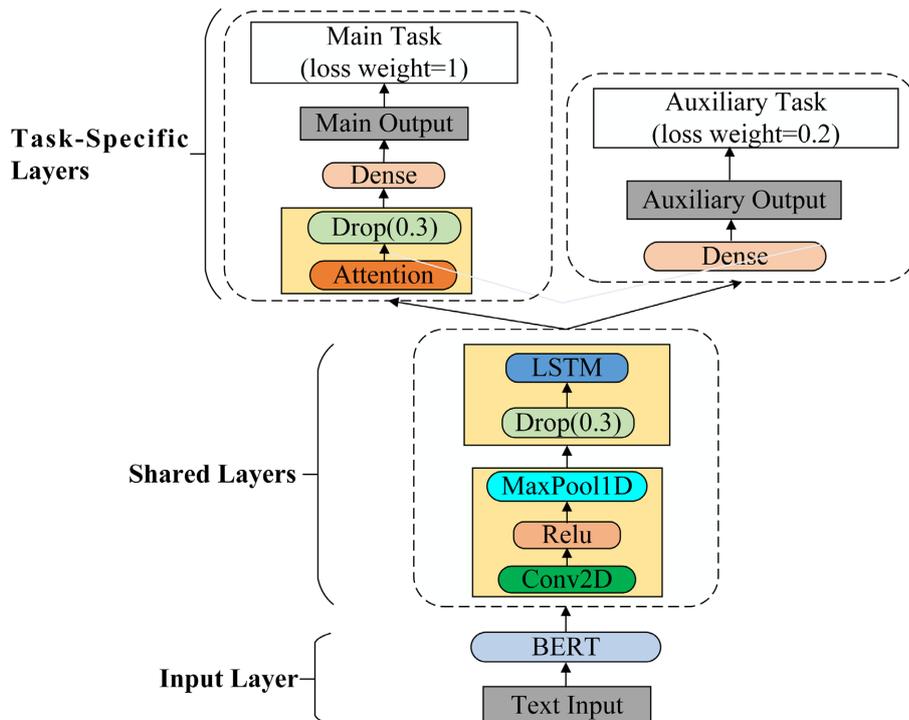


Figure 1. BERT-CNN-LSTM-Joint Model

图 1. BERT-CNN-LSTM-Joint Model 预测模型

2.1. 输入层

输入层是深度学习模型的初始部分，其任务是接收经过预处理的文本数据，将处理后的文本转化为向量形式，并使其准备好供模型进一步处理。本文输入层采用基于 BERT 的词嵌入方法，使用预训练模型 BERT-BASE 来进行文本表示。BERT (Bidirectional Encoder Representations from Transformers) 是一种在大规模语料上进行预训练的语言模型，它能够捕捉文本中丰富的语义信息[13]。BERT-BASE 模型的输入由词向量，块向量和位置向量组成，如图 2 所示。多任务学习模型的输入序列为 $Input$ 。

$$Input = ([CLS], w_1, \dots, w_m, [SEP]) \quad (1)$$

输入序列通过 BERT 编码后得到了文本词向量为 O_b 。

$$O_b = BERT(Input) \quad (2)$$

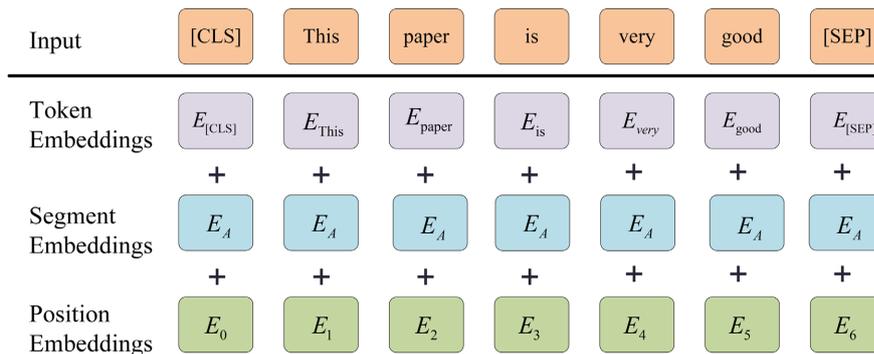


Figure 2. BERT-BASE input features

图 2. BERT-BASE 输入特征

2.2. 共享层

共享层能够提取不同任务之间通用的特征表示。通过共享层，模型可以学习到适用于多个任务的抽象特征表示，从而提高模型的泛化能力和效率。本文将 CNN-LSTM 作为共享层，用来对输入词嵌入进行转换和提炼，以便让模型更好地理解文本的语义特征。

卷积神经网络由卷积层和池化层构成。卷积层通过卷积核在输入数据上进行滤波操作，提取不同的文本特征。进行卷积操作的计算公式如下所示：

$$O_b \otimes w = ReLU \left(\sum_{k=1}^M \sum_{j=1}^{W_f} \sum_{i=1}^{H_i} (O_b^k(i, j) w^k(i, j)) \right) \quad (3)$$

其中， O_b 表示输入文本序列的词嵌入矩阵， \otimes 表示卷积操作， w 表示卷积核的权重参数， M , W_f 和 H_i 分别表示卷积核的通道数，宽度和高度。

池化层的主要目的是通过降维和抽样来减少数据量并突出显著的特征。最大池化和平均池化是两种常见的池化操作类型。本文采用最大池化操作来提取特征，其计算公式如下：

$$MaxPool(O_b \otimes w) = \max_{i,j} (O_b \otimes w[i, j]) \quad (4)$$

在深度学习中，出现过拟合是一种常见的现象。Hinton [14] 等人提出了一种 Dropout 的正则化技术，用于减少神经网络的过拟合。Dropout 层通过在神经网络的训练过程中随机地将某些神经元的输出置零，以一定概率抑制神经元的连接从而减少神经元之间的依赖关系和提高模型的泛化能力。本文将 Dropout

层中抑制神经元连接的概率设置为 0.3。输入文本序列的词嵌入矩阵经过卷积和池化操作之后，进行 Dropout 处理，得到 X_D 。

$$X_D = Dropout(MaxPool(O_b \otimes w)) \quad (5)$$

LSTM 是一种用于对序列数据进行建模的神经网络结构。相比于普通的 RNN，LSTM 能够更好地捕获长序列中的依赖关系，同时有效地避免梯度消失的问题。LSTM 通过引入了输入门，遗忘门和输出门来控制信息的输入、遗忘和输出，使得模型能够选择性地保留或遗忘输入数据中的信息。LSTM 网络具体的计算公式如下所示：

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (6)$$

$$c_t = (1 - i_t) \cdot c_{t-1} + i_t \cdot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (7)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (8)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (9)$$

其中： i_t 是输入门的输出， c_t 是 t 时刻细胞状态， o_t 表示 t 时刻的输出门， h_{t-1} 表示上一个细胞的输入， x_t 表示 t 时刻的输入， σ 表示 sigmoid 函数， W 表示权重矩阵。在利用卷积神经网络来捕获文本中的局部特征之后接着使用 LSTM 来进一步编码序列中的语义和上下文信息。将 X_D 经过 LSTM 特征提取，编码后为 X_L 。

$$X_L = LSTM(X_D) \quad (10)$$

2.2. 多任务输出层

多任务输出层负责针对每个任务执行特定的预测。多任务学习是指一个模型同时处理多个任务或目标的机器学习方法，并且在一定程度上实现了共享参数[12]。在本文中，BCLJ 模型包括两个多任务输出。第一个任务是预测论文是否被接收，而第二个任务是预测论文的评分。

对于论文接收预测任务，本文使用自注意力机制来使模型关注重要的文本信息。自注意力机制(Self-Attention)是一种用于序列数据建模的注意力机制，能减少对于外部信息的依赖从而更好的捕捉特征之间的内部关系。自注意力机制的计算公式如下所示：

$$S_A(Y) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

$$Q = W_q Y, K = W_k Y, V = W_v Y \quad (12)$$

其中 Q, K 以及 V 是通过线性变换得到的， W_q, W_k 和 W_v 是可以学习的权重参数， $Softmax$ 是用来权重归一化的函数。自注意力机制的框架如图 3 所示。将经过共享层的 X_L 进行自注意力机制编码后得到 X_L^{SA} 。

$$X_L^{SA} = selfAttention(X_L) \quad (13)$$

在进行自注意力机制后再一次进行 Dropout 操作来减少模型的过拟合风险，提高模型的泛化能力。 X_L^{SA} 在进行 Dropout 操作之后得到 X_{LD}^{SA} 。

$$X_{LD}^{SA} = Dropout(X_L^{SA}) \quad (14)$$

本文提出的多任务模型有两个输出头，二者分别使用 Sigmoid 函数和全连接层来预测论文被接收的概率分布 \hat{Y}_p 和论文评分的数值 \hat{S} 。

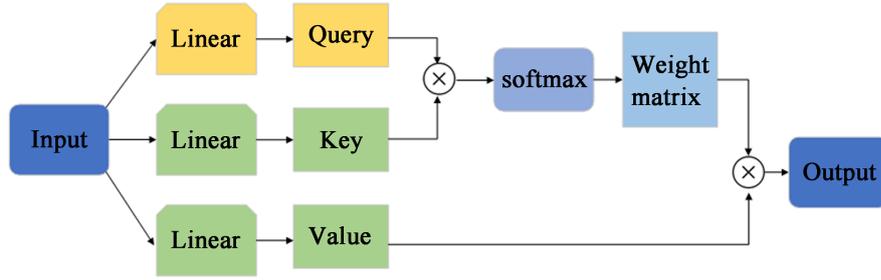


Figure 3. Self-attention mechanisms
图 3. 自注意力机制

$$\widehat{Y}_p = \text{Sigmoid}\left(\text{Dense}\left(X_{LD}^{SA}\right)\right) \quad (15)$$

在多任务学习中，模型的训练损失是由多个子任务的损失函数加权得到的。本文使用交叉熵损失函数作为论文接收预测任务的损失函数，并使用均方误差损失作为论文评分预测任务的损失函数。设定论文接收预测任务的损失权重为 1，评分预测任务的损失权重为 λ 。因此，模型的联合损失为 $Loss$ 。

$$Loss = 1 \cdot Loss_p + \lambda \cdot Loss_s \quad (16)$$

$$Loss_p = -\left(\widehat{Y}_p \cdot \log(p) + (1 - \widehat{Y}_p) \cdot \log(1 - p)\right) \quad (17)$$

$$Loss_s = (\widehat{S} - S)^2 \quad (18)$$

其中 $Loss_p$ 为论文接收预测任务的损失， $Loss_s$ 为分数预测任务的损失， λ 为联合损失的调整系数。

3. 实验过程及结果

3.1. 数据准备

本文实验使用的数据是从 OpenReview 平台上爬虫获取的。本文收集了 2023 年提交给 ICLR 会议的所有评论，最终得到 3851 篇文章的数据和 15,400 篇专家评审数据。对于这 15,400 篇评审数据，每个专家都为每篇论文提供了一个整体的推荐分数，其范围在 1 到 10 分之间，10 分代表最高评价。为了综合考量专家们对于某篇论文的评价，本文引入了一个新的变量 S 来度量该论文的得分。这个得分是各专家对同一篇论文推荐分数的平均值，计算公式如下所示：

$$S = \frac{1}{n} \sum_{i=1}^n score_i \quad (19)$$

其中 n 表示由 n 个专家对于同一篇论文进行评价， $score_i$ 表示第 i 个专家对于某篇论文的推荐得分。通过绘制直方图展示了综合得分与论文最终决策的数据分布情况，如图 4 所示。图中显示具有较高评分的文章更有可能被接收，这进一步表明论文评分预测任务对于预测论文接收具有潜在的辅助作用。

本文将专家对于每篇论文的总体评价进行综合，并进行预处理后，按 8:2 的比例划分为训练集和测试集。最终的分布情况如表 1 所示。

3.2. 对比实验

为了验证本文提出的多任务学习模型在 ICLR 同行评审数据集上的性能，本文进行了一系列实验。这些实验分为三组，两组单任务对比实验，一组多任务对比实验。

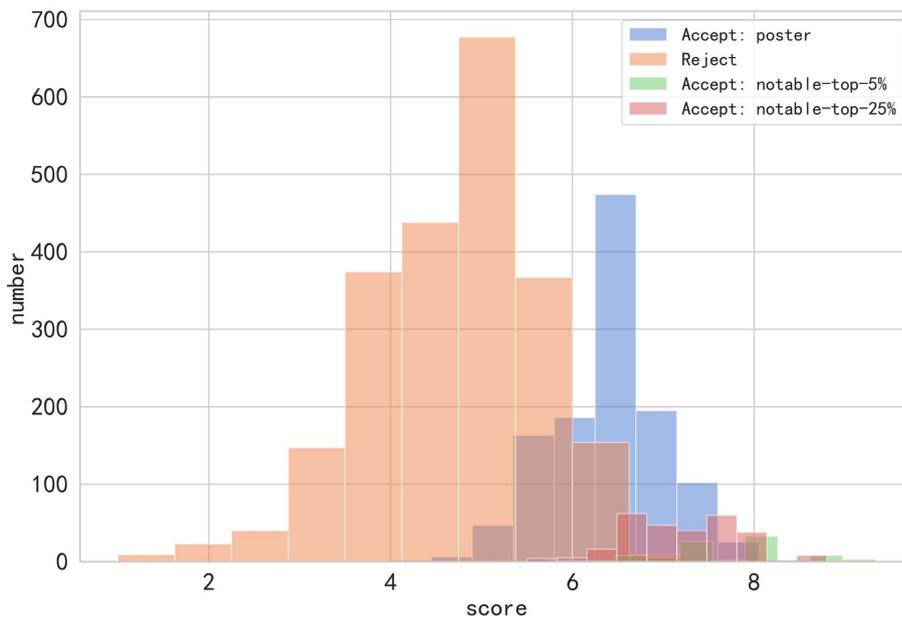


Figure 4. Histogram of scores and paper decisions

图 4. 综合得分与论文决策直方图

Table 1. Data distribution of experimental samples

表 1. 实验样本数据分布

实验文本的分布	训练集数量	测试集数量
接收	1302	310
未接收	1778	460

针对实验一进行的论文接受预测任务，本文设计了以下实验。

(1) SVM/XGBOOST: 以 TF-IDF 词向量作为输入，再使用 SVM/XGBOOST 进行论文接受预测分类，如 Ribeiro [4] 的研究所示。

(2) TextCNN (GloVe or BERT): Kim [15] 等人提出了一种卷积神经网络的文本分类模型。本文以 GloVe 或者 BERT 词向量作为输入，利用 TextCNN 进行论文接受预测分类。

(3) LSTM (GloVe or BERT): 以 GloVe 或者 BERT 词向量作为输入，利用长短期记忆网络捕获全局句子信息，将编码后的信息输入到全连接层进行分类，如 Fernandes [1] 的研究所示。

(4) BiLSTM-Attention (GloVe or BERT): Zhou [16] 等通过结合双向长短期记忆网络和注意力机制来进行分类。以 GloVe 或者 BERT 词向量作为输入，利用 BiLSTM-Att 进行论文接受预测分类。

针对实验二进行的论文评分预测任务，本文设计了以下实验。

(1) SVM/XGBOOST: 以 TF-IDF 词向量作为输入，再使用 SVM/XGBOOST 进行论文评分预测，如 Ribeiro [4] 的研究所示。

(2) TextCNN/BiLSTM/CNN-LSTM: 以 GloVe 或者 BERT 词向量作为输入，再分别使用 TextCNN, BiLSTM 和 CNN-LSTM 模型进行论文评分预测，如 Yang [17] 的研究所示。

针对实验三进行的基于多任务学习的论文接收预测和评分预测任务，本文设计了以下实验。

(1) BCLJ: 输入序列以 BERT 编码，共享 CNN-LSTM 进行语义特征提取，分别用 sigmoid 函数和全连接层完成接收预测和评分预测。

(2) GCLJ: 输入序列以 GloVe 编码, 共享 CNN-LSTM 进行语义特征提取, 分别用 sigmoid 函数和全连接层完成接收预测和评分预测。

3.3. 实验模型参数设置

为了构建多任务学习模型, 本文采用了 PyTorch 深度学习框架, 并设置了一些重要参数, 如表 2 所示。其中 BERT 和 GloVe 的词嵌入维度大小 Embedding size 分别为 768 与 100。Learning rate 为模型的初始学习率并设置为 0.0001。为了提高模型的鲁棒性并避免过拟合, 选择 Adam 作为本文实验的优化器。为了防止模型过拟合, 将 Dropout 的概率设置为 0.4。将训练周期设置为 20。卷积层中滤波器的数量设置为 100, 每个卷积核的尺寸为 [3, 4, 5]。

Table 2. Experimental parameter settings

表 2. 实验参数设置

参数	设定值
BERT	Bert-base-uncased
BERT Embedding size	768
GloVe	glove.6B.100d
GloVe Embedding size	100
Learning rate	10^{-4}
优化器	Adam
Dropout	0.4
Epochs	20
Batch size	8
Num_filters	100
Filter-size	[3, 4, 5]
λ	0.5

在多任务学习中, 辅助任务的权重用于平衡各个任务对模型整体训练的影响, 旨在找到一个合适的辅助任务的权重能够使得各个评价指标尽可能的达到最优。图 5 显示了各个不同的权重下各个评价指标的比较结果。在权重为 0.2 和 0.5 时, 论文接收预测任务的准确率和 F1 值都相对较高。然而, 在论文评分预测任务中, 权重为 0.2 时的预测效果优于权重为 0.5 时的表现。因此, 综合考虑各方面因素, 选择了权重为 0.2。

3.4. 实验结果分析

表 3 统计了各个单任务模型在论文接收预测任务中的预测效果。在论文接收预测模型中, 基于 BERT 词向量的 BiLSTM-Attention 模型相较于基于 GloVe 词向量的 BiLSTM-Attention 模型, 准确率提高了 1.77%。这表明基于 BERT 的词向量模型在预测效果上优于基于 GloVe 和 TF-IDF 词向量的模型。在传统的机器学习方法中, 如 SVM 和 XGBOOST 算法, 它们在该任务上虽然劣于 BiLSTM-Attention 模型, 但与 CNN 和 LSTM 模型的表现类似。值得注意的是, 基本的 CNN 模型和 LSTM 模型在无论是使用 BERT 还是 GloVe 词向量时, 均不及 BiLSTM-Attention 模型, 这表明注意力机制和双向长短期记忆网络能够更有效地学习文本特征。

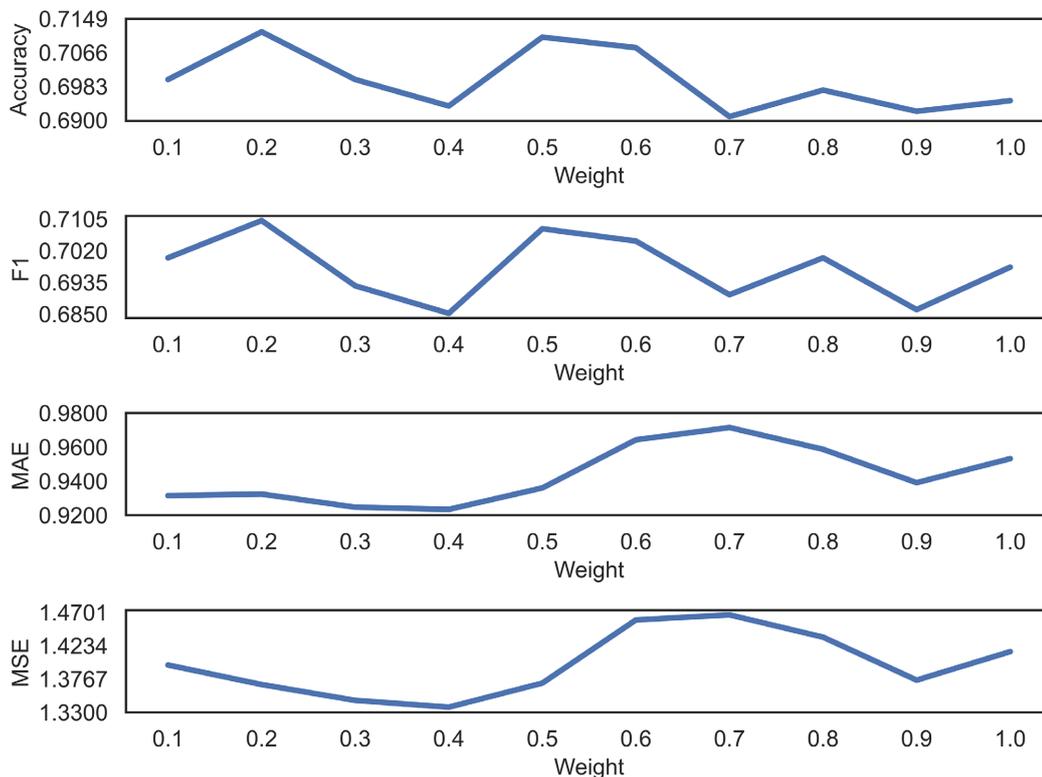


Figure 5. Performance comparison of different weights
图 5. 不同权重的性能比较

Table 3. Paper receiving predictive single-task modeling experimental results
表 3. 论文接收预测单任务模型实验结果

	Models	Accuracy	F1 value
论文接收预测任务	SVM (TF-IDF)	0.6493	0.6457
	XGBOOST (TF-IDF)	0.6415	0.6349
	TextCNN (GloVe)	0.6351	0.6241
	TextCNN (BERT)	0.6481	0.6172
	LSTM (GloVe)	0.6338	0.6221
	LSTM (BERT)	0.6487	0.6409
	BiLSTM-Att (GloVe)	0.6537	0.6733
	BiLSTM-Att (BERT)	0.6714	0.6747

表 4 统计了各个单任务模型在论文分数预测任务中的预测效果。基于 GloVe 词向量的 BiLSTM 模型的预测效果最差。在基于 BERT 类模型中, CNN-LSTM 模型的预测效果最好, 表明该模型更加充分地捕捉了文本的局部特征和全局特征, 并且对序列信息进行了充分建模。在 TF-IDF 词向量结合机器学习方法中, SVM 和 XGBoost 的预测效果也相当可观, 展现了对稀疏数据的良好处理能力, 有效捕捉文本特征。整体来看, 基于 BERT 词向量的模型优于基于 GloVe 词向量的模型。BiLSTM 模型无论基于 BERT 还是 GloVe 词向量, 效果均不如 CNN-LSTM, 可能是因为未能捕捉文本的局部特征, 并未充分利用词向量的语义特征。

Table 4. Experimental results of a single-task model for paper scoring prediction
表 4. 文评分预测单任务模型实验结果

	Model	MSE	RMSE	MAE
论文评分预测任务	SVM (TF-IDF)	1.3312	1.1218	0.9384
	XGBOOST (TF-IDF)	1.4179	1.0573	0.9416
	TextCNN (GloVe)	1.3328	1.1545	0.9507
	TextCNN (BERT)	1.4107	1.2805	0.9430
	BiLSTM (GloVe)	1.5472	1.1465	0.9864
	BiLSTM (BERT)	1.3990	1.3642	0.9509
	CNN-LSTM (GloVe)	1.3889	1.1785	0.9415
	CNN-LSTM (BERT)	1.2477	1.1088	0.9308

表 5 中对比实验为本文提出的多任务论文接收预测和论文评分预测联合模型 BCLJ 和使用 GloVe 词向量的多任务模型的对比。正如结果所示，多任务模型能同时完成论文接收预测和论文评分预测任务。本文提出的多任务模型在论文接收预测任务中击败了所有的基线模型。具体来说，BCLJ 模型在论文接收预测任务中相较于单任务的 BiLSTM-Attention (BERT)模型，准确率提高了 4.03%，F1 值提高了 3.54%。与单任务模型 LSTM 和 TextCNN 相比，在准确率和 F1 值上的提升超过了 6%。针对论文评分预测任务，BCLJ 模型相较于基于 BERT 的 BiLSTM 单任务模型，MAE、MSE 和 RMSE 分别降低了 0.0185、0.03 和 0.1942。在论文评分预测任务中，BCLJ 模型表现与基于 BERT 或 GloVe 的 CNN-LSTM 模型相当。然而，通过论文评分预测任务对论文接收预测任务的辅助，BCLJ 模型在论文接收预测任务上效果显著提升，这表明论文评分与论文接收具有一致性。进一步注意到与经典的机器学习分类器相比，BCLJ 模型对论文接收预测任务有着显著的提升并且对论文评分预测任务也有一定的改善。在进行多任务学习的模型时，使用基于 BERT 词向量的多任务模型表现优于使用基于 GloVe 词向量的模型。

Table 5. Experimental results of a single-task model for paper scoring prediction
表 5. 文评分预测单任务模型实验结果

多任务模型	论文接收预测任务		论文评分预测任务		
	Accuracy	F1 value	MSE	RMSE	MAE
GCLJ (GloVe)	0.6987	0.6927	1.4024	1.1842	0.9424
BCLJ (BERT)	0.7117	0.7101	1.3690	1.1700	0.9324

4. 结论

为了解决论文接收预测的难题，本文提出了一种同行评审文本分析模型 BCLJ，引入了多任务学习的思想。首先，通过 BERT 词嵌入获得词向量矩阵；其次利用 CNN 和 LSTM 提取语义特征，并引入注意力机制增强对文本信息的理解能力；最后，利用不同的全连接层进行多任务学习，获得论文接受预测和评分预测两种输出。该模型将论文评分预测任务作为辅助任务，将论文接收预测任务作为主要任务，显著提高了论文接收预测任务的准确率。此外，通过论文评分预测任务来增加论文接收预测任务的准确率，从侧面说明了论文评分与论文接收存在一致性。

参考文献

- [1] Fernandes, G.L. and Vaz-de-Melo, P.O.S. (2022) Between Acceptance and Rejection: Challenges for an Automatic

- Peer Review Process. *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, Cologne, 20-24 June 2022, 1-12. <https://doi.org/10.1145/3529372.3530935>
- [2] Pang, B. and Lee, L. (2005) Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. Ann Arbor, 25-30 June 2005, 115-124. <https://doi.org/10.3115/1219840.1219855>
- [3] Keith, B., Fuentes, E. and Meneses, C. (2017) A Hybrid Approach for Sentiment Analysis Applied to Paper. *Proceedings of ACM SIGKDD Conference*, Halifax, August 2017, 10.
- [4] Ribeiro, A.C., Sizo, A., Lopes Cardoso, H., et al. (2021) Acceptance Decision Prediction in Peer-Review through Sentiment Analysis. Springer International Publishing, Cham, 766-777. https://doi.org/10.1007/978-3-030-86230-5_60
- [5] Li, S., Zhao, W.X., Yin, E.J., et al. (2019) A Neural Citation Count Prediction Model Based on Peer Review Text. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, November 2019, 4914-4924. <https://doi.org/10.18653/v1/D19-1497>
- [6] Leng, Y., Yu, L., Xiong, J. (2019) Deepreviewer: Collaborative Grammar and Innovation Neural Network for Automatic Paper Review. *2019 International Conference on Multimodal Interaction*, Suzhou, October 2019, 395-403. <https://doi.org/10.1145/3340555.3353766>
- [7] Deng, Z., Peng, H., Xia, C., et al. (2020) Hierarchical Bi-Directional Self-Attention Networks for Paper Review Rating Recommendation. *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, November 2020, 6302-6314. <https://doi.org/10.18653/v1/2020.coling-main.555>
- [8] 陈红玉, 胡文俊, 路永和. 开放同行评审中自动评审分类方法研究[J]. 现代情报, 2024, 44(5): 95-106.
- [9] 林原, 王凯巧, 丁莹. 学术论文的定性评价定量化研究[J]. 情报理论与实践, 2021, 44(8): 28-34.
- [10] Li, J., Sato, A., Shimura, K., et al. (2020) Multi-Task Peer-Review Score Prediction. *Proceedings of the First Workshop on Scholarly Document Processing*, Stroudsburg, November 2020, 121-126. <https://doi.org/10.18653/v1/2020.sdp-1.14>
- [11] 朱金秋, 檀健, 韩斌彬, 等. 基于多任务学习的同行评审细粒度情感分析模型[J]. 中北大学学报(自然科学版), 2024, 45(1): 105-113.
- [12] Vandenhende, S., Georgoulis, S., Van Gansbeke, W., et al. (2021) Multi-Task Learning for Dense Prediction Tasks: A Survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **44**, 3614-3633. <https://doi.org/10.1109/TPAMI.2021.3054719>
- [13] Kenton, J.D. and Toutanova, L.K. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, Minnesota, June 2019, 4171-4186.
- [14] Hinton, G.E., Srivastava, N., Krizhevsky, A., et al. (2012) Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. <https://doi.org/10.48550/arXiv.1207.0580>
- [15] Kim, Y. (2014) Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, conference date, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [16] Zhou, P., Shi, W., Tian, J., et al. (2016) Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, conference city, conference date, 207-212. <https://doi.org/10.18653/v1/P16-2034>
- [17] Yang, P., Sun, X., Li, W., et al. (2018) Automatic Academic Paper Rating Based on Modularized Hierarchical Convolutional Neural Network. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, May 2018, 496-502. <https://doi.org/10.18653/v1/P18-2079>