

基于 BERTGAT-Contrastive 的语义匹配模型

袁靖舒, 李 武, 赵兴雨, 袁 满

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

摘 要: 针对仅使用 BERT 最后一层信息进行预测导致丢失一些文本的词法、句法和语义信息的问题, 提出了一种基于 BERT 与图注意力网络 (GAT) 的 BERTGAT 模型。首先, 通过利用 BERT 多个中间层的隐藏状态矩阵和注意力矩阵分别作为对应数量 GAT 的节点特征矩阵和邻接矩阵, 并采用动态权重策略对不同的 GAT 层进行加权, 再应用激活函数判断句子间的相似性。其次, 为了使 BERTGAT 模型能够更好地学习到句子对之间的语言表征, 在 BERTGAT 的基础上引入了对比学习方法, 提出了 BERTGAT-Contrastive 模型, 增强了模型对文本之间语义相似性的识别能力。最后, 通过在 LCQMC 和 BQ 数据集上进行实验, 结果表明: 本文提出的模型与对比学习方法相比效果更显著, 准确率和 F_1 值均有明显提升。

关键词: 深度学习; 语义匹配; BERT; 图注意力网络; 对比学习

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1671-5497(2025)07-2383-10

DOI: 10.13229/j.cnki.jdxbgxb.20240913

Semantic matching model based on BERTGAT-Contrastive

YUAN Jing-shu, LI Wu, ZHAO Xing-yu, YUAN Man

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: Using only the information of the last layer of BERT for prediction can result in the loss of some lexical, syntactic and semantic information of the text. Regarding this problem, this paper firstly puts forward the BERTGAT model based on BERT and graph attention networks (GAT). By making use of the hidden state matrices and attention matrices of multiple intermediate layers of BERT as the node feature matrices and adjacency matrices of the corresponding number of GAT respectively, and adopting a dynamic weighting strategy to weight different GAT layers, and then applying the activation function to determine the similarity between sentences. Secondly, to enable the BERTGAT model to better acquire the language representations between sentence pairs, the contrastive learning approach is introduced on the basis of the BERTGAT model to propose the BERTGAT-Contrastive model, which enhancing the model's ability to recognize semantic similarities among texts. Finally, through experiments conducted on the LCQMC and BQ datasets, the results indicate that the proposed model and its contrastive learning method are highly effective, with notable improvements in accuracy and F_1 score.

Key words: deep learning; semantic matching; bert; graph attention network; contrastive learning

收稿日期: 2024-08-19.

基金项目: 东北石油大学人才引进科研启动经费项目 (2023KQ17).

作者简介: 袁靖舒 (1992-), 男, 讲师, 博士. 研究方向: 自然语言处理, 深度学习, 知识组织.

E-mail: yuanjingshu@nepu.edu.cn

0 引言

语义匹配(Semantic matching)是自然语言处理的一个重要任务,其目标是确定两个文本是否表达相似的含义,常用于问答系统^[1]、信息检索^[2]以及推荐系统^[3]等。传统的语义匹配方法主要通过基于统计和规则模型评估两个文本片段之间的语义关系,主要有基于 TF-IDF(Term frequency-inverse document frequency)^[4]、基于词袋模型^[5]以及基于 N-gram^[6]的方法等。然而,这些传统方法存在一些局限性,如无法处理复杂语境^[7]、忽略了词序关系、泛化能力较差等。

随着深度学习的兴起,近年来语义匹配研究更多地集中在利用神经网络等深度学习技术,以提高模型在更复杂语义任务上的性能。Huang 等^[8]提出了 DSSM(Deep Structured Semantic Model,深度结构化语义模型),通过 one-hot 方法和 Word Hashing 方法将中英文文本映射到一个向量空间中,然后使用多层全连接网络进行编码,最后利用余弦相似度和 Softmax 函数计算两文本相似度。Shen 等^[9]将 DSSM 模型中的全连接层替换成卷积神经网络(CNN)模型,提出了 CNN-DSSM 模型,从而提升了语义匹配性能。Wang 等^[10]提出了一种基于 Word2vec 模型的短文本匹配算法。翁兆琦等^[11]采用 Word2vec 模型计算句子之间的语义相似度,并融合 LTP 平台获取的句法结构信息,有效规避了中文复杂句式和句法结构对语义匹配任务的影响。这些研究表明:深度学习技术在语义匹配领域具有显著的潜力和优势。

在最近的研究中,多角度信息交互^[11]、实体上下文特征融合以及知识增强等策略^[12,13]被广泛用于提高语义匹配的准确性。例如,一些研究通过融合汉字的形音义多元知识和标签嵌入加深模型对中文语义的理解^[14]。此外,多头注意力机制和孪生网络^[15,16]也被引入语义匹配中,这些技术可以更好地捕捉文本中的长距离依赖关系和复杂模式。在多任务语义匹配中自监督学习^[17]可以从未标注的数据中学习有用的表示,而基于多维特征表示的文本匹配方法^[18,19]也为提高匹配效果提供了新思路。为了进一步提高匹配效果,许多研究者还尝试引入外部知识、利用先验知识指导模型的注意力机制等方法^[20]。随着时间的推移,预训练语言模型 BERT 的出现进一步提高了深

度学习在语义匹配中的性能。一些研究对 BERT 进行微调即可取得很好效果^[21,22]。关于 BERT 的研究重点更多是在利用 BERT 和其他模型或方法进行语义匹配任务,如宋爽等^[23]提出了一种基于 BERT 与图像自注意力机制的文本匹配模型,陈岳林等^[24]提出了一种基于 BERT 与密集复合网络的长文本语义匹配模型,王钦晨等^[25]提出了一种基于 BERT 字句向量与差异注意力的短文本语义匹配策略。

通过上述调研可知,在语义匹配任务中,出现了多种对现有算法进行改进或多种算法相结合的方法,但基于 BERT 模型的语义匹配研究依旧属于热点。多数 BERT 相关的研究是在其基础上融合其他算法,并主要依赖于其最后一层的输出信息完成预测或语义匹配等任务。尽管取得了不错的效果,但却忽略了 BERT 在多层架构中逐步积累到的丰富多样化词法、句法和语义信息。

针对这一问题,本文提出了一种 BERT 与图注意力网络(GAT)相结合的 BERTGAT 模型。该模型充分运用了 BERT 在不同层次上分别捕获的文本信息,包括从表层特征逐步深入到句法特征,再到更为丰富的语义特征。首先,利用 BERT 多个中间层的隐藏状态矩阵和注意力矩阵分别作为对应数量 GAT 的节点特征矩阵和邻接矩阵,并采用动态权重策略对不同的 GAT 层进行加权,再应用激活函数判断句子间的相似性。其次,本文在 BERTGAT 的基础上引入了对比学习方法设计了 BERTGAT-Contrastive 模型,通过比较正样本和负样本之间的相似性和差异性,引导模型学习更准确的表示,使正样本在表示空间中更为接近,使负样本在表示空间中更为分散,从而增强模型对文本之间语义相似性的识别能力。最后,分别在两个公开领域问答数据集上验证了两个模型的有效性。实验结果表明:本文提出的模型与对比学习方法相比效果更显著,具有较高的准确率和 F_1 值。

1 BERTGAT 语义匹配模型设计

本文提出了一种能够利用 BERT 在预训练过程中学到的多层语义特征并通过 GAT 有效整合信息的 BERTGAT 模型(见图 1)。首先获取 BERT 中多个层的隐藏状态矩阵和注意力矩阵;然后将隐藏状态矩阵作为 GAT 的节点特征矩阵,同时计算注意力矩阵的平均值作为 GAT

的邻接矩阵,采用动态权重策略对不同的 GAT 层进行加权;最终,对 GAT 和 BERT 的输出进行独立分类,得到两个独立的损失值,模型的总体损失由这两个部分的损失组成。

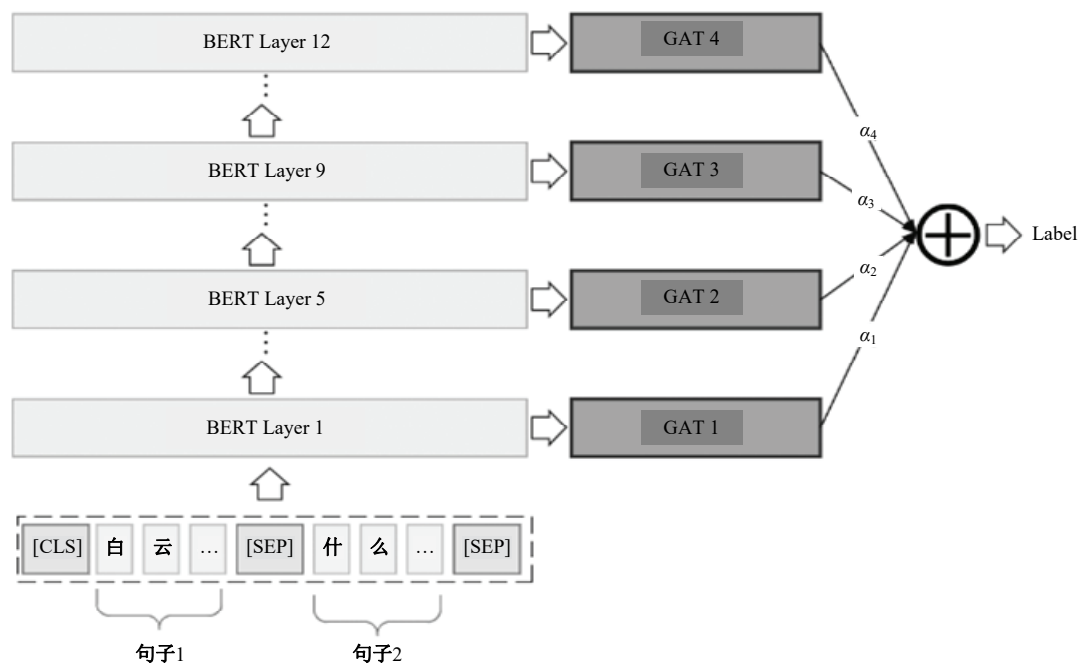


图1 BERTGAT模型框架图
Fig. 1 BERTGAT model frame diagram

1.1 BERT层

BERT 是一种基于 Transformer 架构的预训练语言模型,其中包含了多个编码器层。在 BERT 预训练过程中,模型目标是通过融合所有层的信息理解语言的上下文和关系。每个层都生成了一个表示,而最终的预训练表示是通过将所有层的输出组合在一起得到的。这种多层设计使 BERT 能够捕捉从底层到高层不同层次的语言信息,包括词法、句法和语义特征。分析文献[26]发现,BERT 在不同层上捕获的信息是不同的,并对 BERT 不同层级学习到的表示特征进行了深入探究,总结发现词法信息特征在底层网络,句法信息特征在中间层网络,语义信息特征在高层网络。

假设语义匹配任务中的句子对为 (s_1, s_2) ,其

中 s_1 和 s_2 分别对应语义匹配任务中的句子1和句子2,则 BERT 的输入为 $([CLS], s_1, [SEP], s_2, [SEP])$,如图2所示。

对于下游任务,通常会在 BERT 的最后一层之后添加一个额外的分类层,以便输出与任务相关的预测。BERT 将输入文本特征编码为隐藏状态和自注意力权重。其中,隐藏状态作为文本的中间表示形式,包含了输入文本的深层语义信息。而自注意力权重则是用于确定每个位置对生成隐藏状态的贡献程度。将不同层的输出组合起来,或者对中间层的输出进行进一步处理(例如加权组合或降维),这两种方式都有助于获取更充分的信息。

隐藏状态矩阵定义为:
 $H=[H_1, H_2, \dots, H_n]$,其中 H_i 是 BERT 中第 i 层

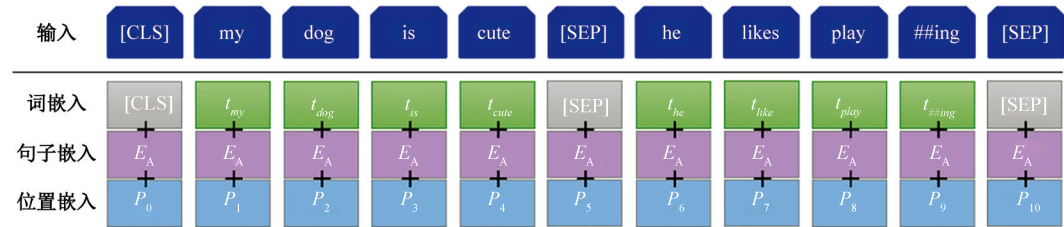


图2 BERT模型的输入表示
Fig. 2 Input representation of BERT model

的隐藏状态矩阵。注意力矩阵定义为: $B = [B_1, B_2, \dots, B_n]$, 其中 B_i 是 BERT 中第 i 层的注意力状态矩阵。

本文受 Jawahar 等^[26]研究成果的启发, 选择第 1、5、9 和 12 层的隐藏状态矩阵和注意力矩阵作为下一层的输入。

1.2 GAT 层

在 BERT 等预训练模型的基础上添加额外结构, 如 CNN、图卷积神经网络(GCN)等, 可以更适应于特定任务并提高算法性能。本文从全面提升语义匹配的性能角度出发, 利用 BERT 多个中间层的隐藏状态矩阵和注意力矩阵分别作为对应数量 GAT 的节点特征矩阵和邻接矩阵, 可以更有效地利用 BERT 模型在多层结构中积累的信息, 从而在语义匹配任务中提高匹配的准确性。

首先对 BERT 的第 i 个中间层注意力矩阵 B_i 进行加权平均, 得到一个与输入序列长度相同的平均注意力矩阵 W_i 。 W_i 用于表示不同位置之间的关系强度, 作为第 m 个 GAT 中的邻接矩阵 A_m :

$$A_m = W_i = \frac{1}{n} \sum_{j=1}^n B_{i,j} \quad (1)$$

式中: n 为输入序列的长度。BERT 中的每个 token 视为图中的节点。

第 m 个 GAT 中的节点特征矩阵 X_m 来源于 BERT 模型中第 i 层的隐藏状态矩阵 H_i 。每一层的隐藏状态都被视为节点在不同层次的特征表示。

假设有 L 个 GAT 层, 每个层 i 对应的权重为 w_i , 第 i 个 GAT 层的归一化权重为 α_i , 计算公式如下:

$$\alpha_i = \frac{\exp(w_i)}{\sum_{j=0}^{L-1} \exp(w_j)} \quad (2)$$

将不同 GAT 层的输出按照归一化后的权重进行线性加权, 得到最终的序列输出 O , 即

$$O = \sum_{i=0}^{L-1} \alpha_i \cdot Z_i \quad (3)$$

式中: Z_i 为第 i 个 GAT 层的输出。

1.3 损失函数

损失函数用于衡量模型预测结果与实际标签之间的差距。旨在提供一个量化的指标以评估模型的预测性能。通过计算损失, 可以了解模型在匹配任务中的表现。

BERTGAT 模型的损失函数由两部分组成: BERT 模型的损失和 GAT 的损失。BERT 模型的损失函数通常采用交叉熵损失, 用于二分类或多分类问题, 衡量模型输出的概率分布与真实标签的一致性。GAT 部分的损失函数关注于图注意力网络的预测与实际标签之间的差异, 这有助于优化 GAT 层的参数, 使其更好地捕捉文本中的语义信息。

BERTGAT 模型的分类层计算公式如下:

$$\text{logits} = x \cdot W^T + b \quad (4)$$

式中: logits 为模型的预测输出; x 为经过图注意力层和加权平均后的特征表示; W 为分类层的权重矩阵; b 为分类层的偏置参数。

BERT 模型的损失函数 L_{bert} 为:

$$L_{\text{bert}} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \text{labels}_{i,j} \cdot \log(\text{Softmax}(\text{logits}_{i,j})) \quad (5)$$

式中: N 为样本数量; C 为类别数量; $\text{labels}_{i,j}$ 为样本 i 是否属于类别 j 的二元标签; $\text{labels}_{i,j}$ 为模型对样本 i 属于类别 j 的预测得分; $\text{Softmax}(\text{logits}_{i,j})$ 为将 $\text{logits}_{i,j}$ 应用 Softmax 函数得到的概率分布。

GAT 部分的损失函数 L_{gat} 为:

$$L_{\text{gat}} = -\frac{1}{N} \sum_{i=1}^N (\text{labels}_i \cdot \log(\text{Softmax}(\text{logits}_{i,\text{true}})) + (1 - \text{labels}_i) \cdot \log(1 - \text{Softmax}(\text{logits}_{i,\text{false}}))) \quad (6)$$

式中: $\text{logits}_{i,\text{true}}$ 为模型对样本 i 中真实类别的预测值; $\text{logits}_{i,\text{false}}$ 为模型对样本 i 中其他类别的预测值。

最终的总体损失 L_{ce} 通过权重线性组合, 计算公式如下:

$$L_{\text{ce}} = (1 - \alpha) \cdot L_{\text{bert}} + \alpha \cdot L_{\text{gat}} \quad (7)$$

为了平衡 BERT 和 GAT 在语义匹配中的贡献, 本文将权重 α 设置为 0.6。

本节提出了 BERTGAT 模型, 模型融合了 BERT 的深层语义表示和 GAT 的局部特征捕获能力, 阐述了如何有效利用 BERT 模型的多个中间层输出, 并在此基础上构建 GAT 层。此外, 还分析了如何通过加权平均注意力机制形成邻接矩阵, 并使用 BERT 的隐藏状态矩阵作为图节点的特征表示。最终, 通过引入动态权重调整策略和特定的损失函数优化模型性能。

2 BERTGAT-Contrastive语义匹配模型设计

为了进一步提升BERTGAT模型在学习句子对之间语言特征方面的能力,本节在其基础上引入了对比学习方法并提出了BERTGAT-Contrastive模型,指导模型更精确地捕捉句子间的语义表示,从而增强模型识别文本语义相似性的能力。对比学习的设计主要包括两个部分:正负样本构建和损失函数。

2.1 正负样本构建

构建正负样本是对比学习的第一步,也是关键一步。在语义匹配任务中处理句子对时,通常会将句子对视为一个整体。这是因为任务的目标是评估或学习句子之间的语义关系,而不仅是单独处理每个句子。经过上述方法的处理,模型可以更好地捕捉句子之间的关联性和语义匹配程度。

由于BERT模型中dropout机制的存在,将同一个句子进行两次前向传播,会得到不同的特征向量,因此,SimCSE方法^[27]将句子本身作为正样本用于数据扩增。基于此思想,本节将句子对中的两个文本交换顺序,标签不变,用于构建对比学习的正例,这种方法并不会引入噪声。

正负样本构建方法如下:

(1)将batch中的句子对交换顺序作为正样本。

(2)选择batch中与句子对A标签不同的句子对B,从A和B中各选一句生成新的句子对,将其作为A的负样本。

例如,假定同一个batch中存在句子对 $h_1(s_1, s_2)$ 、 $h_2(s_3, s_4)$ 、 $h_3(s_5, s_6)$,它们的语义匹配任务标签分别为0、1、1,则句子对 $h_1(s_1, s_2)$ 对应的正样本为 $h_1^+(s_2, s_1)$,负样本为 $h_{11}^-(s_1, s_3)$ 、 $h_{12}^-(s_1, s_4)$ 、 $h_{13}^-(s_1, s_5)$ 、 $h_{14}^-(s_1, s_6)$ 、 $h_{15}^-(s_3, s_2)$ 、 $h_{16}^-(s_4, s_2)$ 、 $h_{17}^-(s_5, s_2)$ 、 $h_{18}^-(s_6, s_2)$;句子对 $h_2(s_3, s_4)$ 的正样本为 $h_2^+(s_4, s_3)$,负样本为 $h_{21}^-(s_3, s_2)$ 、 $h_{22}^-(s_2, s_4)$;句子对 $h_3(s_5, s_6)$ 的正样本为 $h_3^+(s_6, s_5)$,负样本为 $h_{31}^-(s_5, s_2)$ 、 $h_{32}^-(s_1, s_6)$ 。

2.2 损失函数

通过Kullback Leibler(KL)散度衡量正负样本之间的相似性,可以更有效地引导模型调整正样本在表示空间中的距离,使它们更为紧密。同

时又会增加负样本在空间中的距离,使它们更为分散。

对从文本 h_i 中得到的正例句子 h_i^+ 和负例句子 h_i^- ,KL散度的计算公式如下:

$$D_{KL}(p \parallel q) = \sum_{i=1}^n p(x) \log \frac{p(x)}{q(x)} \quad (8)$$

式中: $p(x)$ 和 $q(x)$ 分别为 h_i^+ 和 h_i^- 的概率分布。

由于KL散度是非对称的,即

$$D_{KL}(P \parallel Q) \neq D_{KL}(Q \parallel P) \quad (9)$$

通过计算两个方向上的KL散度并取平均得到最终的KL损失:

$$KL = \frac{D_{KL}(p \parallel q) + D_{KL}(q \parallel p)}{2} \quad (10)$$

最终的BERTGAT模型的整体损失 L 为式:

$$L = L_{ce} + \lambda \cdot KL \quad (11)$$

式中: L_{ce} 为BERTGAT模型的损失函数; λ 为用于平衡原始损失和KL散度损失的权重参数。

本节提出了BERTGAT-Contrastive模型。模型中整合了对比学习理论,并采用正负样本构建机制,结合KL散度度量,实现了对句子对间语义相似性的精确表征,优化了语义匹配的性能。

3 实验过程与结果分析

3.1 实验数据与实验参数

为验证本文所提两个模型的有效性,本文选择在LCQMC(Large-scale Chinese question matching corpus)^[28]以及BQ(Bank question corpus)^[29]两个数据集上进行评估。其中,LCQMC是从百度知道的用户问题中抽取的中文语义匹配数据集,涵盖了多个领域,专注于解决中文领域缺乏大规模问题匹配数据的问题。BQ是银行金融领域的问题匹配数据集,其中的问题对是从线上银行系统日志中抽取的,是目前最大的垂直领域问题匹配数据集。

LCQMC和BQ数据集参数与示例如表1和表2所示,标签0代表两句话语义不匹配,1代表两句话语义匹配。

BERT-base是BERT模型框架中的一个具体实现,它代表了具有一定参数量的BERT模

表1 数据集参数

数据集	训练集大小	验证集大小	测试集大小
LCQMC	238 766	8 802	12 500
BQ	100 000	10 000	10 000

表 2 数据集样本示例

Table 2 Sample data set examples

数据集	句子 1	句子 2	标签
LCQMC	这种图片是用什么软件制作的?	这种图片制作是用什么软件呢?	1
	这腰带是什么牌子的	护腰带什么牌子的好	0
	什么牌子的空调最好!	什么牌子的空调扇最好	0
	说身份错误无法借款次	身份信息输入错误过多无法借钱是怎么回事	1
BQ	多久才有贷款	凌晨以后的申请何时到账	1
	为什么我申请后没有收到验证电话	什么时候全线开通微粒贷	0

型。BERT-base 因其较少的参数而在对计算资源要求上更为适中,同时又保持了BERT模型的核心优势,如通过预训练获得的深层双向语义理解能力,使其成为多种自然语言处理任务中的首选模型。本文选择了参数更少的BERT-base模型,并设置了一些关键超参数,具体的超参数配置如表3所示。

表 3 实验超参数设置

Table 3 Experimental hyperparameter Settings

超参数	超参数设置
batch_size	32
Dropout	0.2
learning-rate	$1.6e-5$
epoch	10

3.2 实验设计与评价标准

实验环境为 Windows 系统,显卡型号为 NVIDIA 4090, 24 GB 显存,开发环境为 Python 3.8, PyTorch 版本为 1.11.0。评价指标分别为准确率(Accuracy, ACC)和 F_1 值(F_1 score, F_1),计算

公式如下:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (15)$$

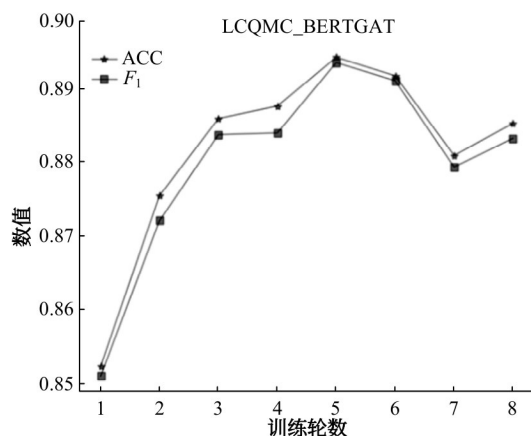
式中:TP(True positive)表示正例预测为正例的个数;TP(False positive)表示负例预测为正例的个数;TN(True negative)表示正例预测为负例的个数;FN(False negative)表示负例预测为负例的个数; P (Precision)为精准率; R (Recall)为召回率。

3.3 实验结果

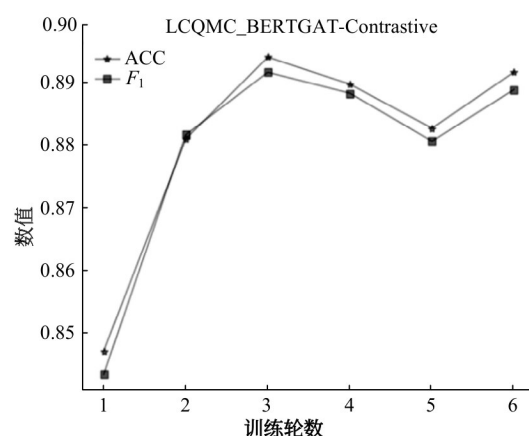
早停法通常用于防止训练中模型过度表达而导致泛化性能差的问题。如果迭代次数太少,算法容易欠拟合,而迭代次数太多,算法则容易过拟合。早停法通过确定迭代次数解决这一问题,不需要对特定值进行手动设置。本文分别在LCQMC和BQ数据集上进行测试,设置早停参数为3,训练结果如图3和图4所示。

通过图3和图4可以发现,在LCQMC数据集上,BERTGAT的最佳ACC为0.88598,最佳 F_1 分数为0.88381;BERTGAT-Contrastive的最佳ACC为0.89878,最佳 F_1 分数为0.89840。在BQ数据集上,BERTGAT的最佳ACC为0.85923,最佳 F_1 分数为0.85591;BERTGAT-Contrastive的最佳ACC为0.86581,最佳 F_1 分数为0.85855。

为了进一步对比本文所提模型在语义匹配任



(a) BERTGAT模型的训练结果



(b) BERTGAT-Contrastive模型的训练结果

图 3 模型在LCQMC数据集上的训练结果

Fig. 3 Training results of the model on LCQMC data set

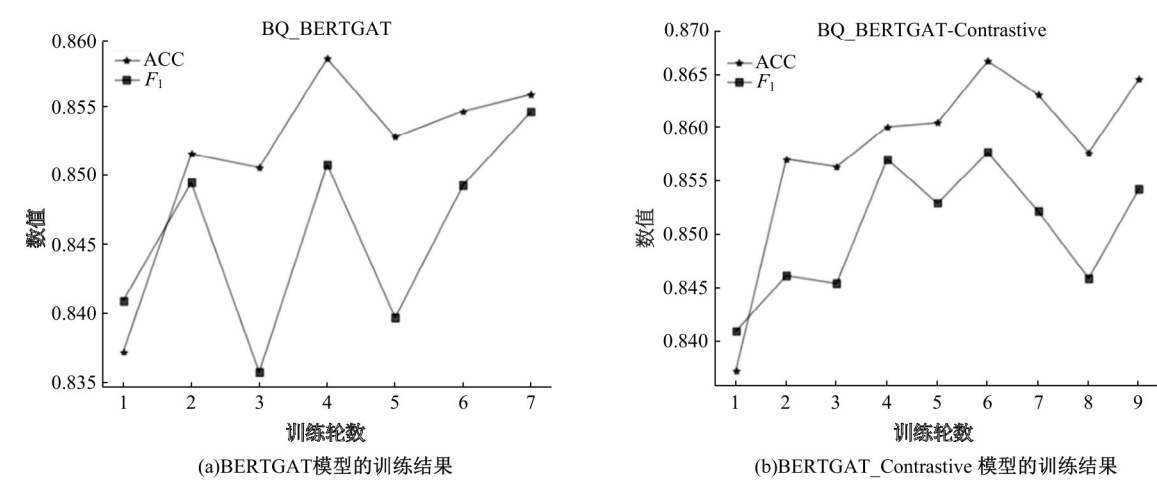


图 4 模型在 BQ 数据集上的训练结果

Fig. 4 Training results of the model on BQ data set

务中的显著效果,选择 Lattice-CNN^[30]、ERNIE、BERT 与多头差异注意力机制结合模型^[25]、Bert+FastText+BiLSTM^[31]、GMN-BERT^[32]和 ICE^[33]共 6 种模型进行对比实验。表 4 展示了本文所提模型和对比模型在 LCQMC 及 BQ 数据集上进行语义匹配任务的实验结果。

表 4 不同语义匹配模型在 LCQMC 和 BQ 数据集上的实验对比效果

Table 4 Experimental comparison effects of different semantic matching models on LCQMC and BQ data sets

模型	LCQMC		BQ	
	ACC	F ₁	ACC	F ₁
Lattice-CNN	82.1	82.4	78.2	78.3
ERNIE	87.00	88.06	84.67	84.20
BERT+多头差异注意力机制	88.55	88.99	85.63	85.36
Bert+FastText+BiLSTM	87.34	87.28	85.42	85.4
GMN-BERT	87.3	88.0	85.6	85.5
ICE	86.73	87.33	84.71	84.77
BERTGAT (本文)	89.81	89.52	86.36	85.81
BERTGAT-Contrastive (本文)	90.84	90.07	87.63	86.78

通过对比实验分析发现,采用词格方法的 Lattice-CNN 由于结构限制,在语义匹配任务上的性能不及本文所提模型。ERNIE 改进了两种 mask 策略,尽管在实验数据集上的语义匹配性能获得一定提升,但不及 BERT 与其他模型相结合的方法。利用 BERT 对句子对进行向量化表示,并引入多头差异注意力机制的方法考虑了文本之间的差异,有效地提取文本语义的差异信息,但没有充分利用 BERT 的多层文本特征,效果不及本文所提模型。利用 BERT 对输入的句子对进行

编码后使用 FastText 进行训练,获得文本词向量,最后加入 BiLSTM 进一步提取上下文之间的语义信息,但该模型并没有充分利用这些信息,因此,模型效果不及本文模型。GMN-BERT 模型通过充分利用词信息,有效提升了语义匹配任务性能,但没有充分利用句子的语义信息。相比之下,本文所提模型可以充分利用文本中的多层次文本特征,因此,对语义匹配性能的提升具有显著效果。

表 5 消融实验结果显示,BERTGAT 模型在 LCQMC 和 BQ 数据集上的性能都明显优于 BERT 模型。相对于基础的 BERT 模型,BERTGAT 模型在 LCQMC 和 BQ 数据集上的准确率分别提高了 4.11、1.86 个百分点,F₁ 值分别提高了 2.72%、1.81%。更进一步地,引入对比学习的 BERTGAT-Contrastive 模型相较于 BERTGAT 模型表现更为出色,其在 LCQMC 和 BQ 数据集上的准确率分别提升了 1.03、1.27 个百分点,F₁ 值分别提升了 0.55、0.97 个百分点。

表 5 消融实验结果

Table 5 Ablation experiment results

模型	LCQMC		BQ	
	ACC	F ₁	ACC	F ₁
BERTGAT-Contrastive	90.84	90.07	87.63	86.78
-Contrastive	89.81	89.52	85.92	85.59
-(GAT-Contrastive)	85.7	86.8	84.5	84.0

结合表 4 和表 5 结果可知,BERT 在 1、5、9 和 12 层通过利用 BERT 多个中间层获取了文本中更多的词法、句法和语义信息。在此基础上结合 GAT 模型捕获文本中的局部信息和全局信息,取

得了更好的效果。BERTGAT-Contrastive 引入对比学习进一步增强了模型性能,它通过将相似样本映射到相邻特征空间并同时将不相似样本分离,有效提高了语义匹配任务的性能。

4 结束语

本文提出了一种 BERT 与 GAT 结合的 BERTGAT 模型。该模型采用一种多层次特征整合策略,通过提取 BERT 模型在多个中间层级的语义编码,将其作为 GAT 层的节点特征输入,通过加权平均方法生成邻接矩阵,为 GAT 层提供了丰富的结构化关系信息,并引入动态权重分配机制和定制化的损失函数,增强了模型对中文语义的理解和匹配能力,解决了只使用 BERT 最后一层信息进行预测所导致的问题。此外,为了进一步提升模型的语义匹配性能,在 BERTGAT 模型的基础上引入了对比学习方法 BERTGAT-Contrastive。实验结果显示,本文提出的两个模型在各项评估指标上均取得了显著的提升。

参考文献:

- [1] 徐若卿. 融合知识图谱和语义匹配的医疗问答系统[J]. 现代电子技术, 2024, 47(8): 49-54.
Xu Ruo-qing. Medical question answering system integrating knowledge graph and semantic matching[J]. Modern Electronic Technique, 2019, 47(8): 49-54.
- [2] 黄翔. 基于中心句语义相似度的信息检索方法研究[D]. 武汉: 华中师范大学计算机学院, 2021.
Huang Xiang. Research on information retrieval method based on semantic similarity of topic sentences[D]. Wuhan: School of Computer Science, Central China Normal University, 2021.
- [3] 于敬, 石京京, 刘文海. 基于文本语义匹配的物品相关推荐算法[J]. 电子技术与软件工程, 2022(7): 206-211.
Yu Jing, Shi Jing-jing, Liu Wen-hai. Article related recommendation algorithm based on text semantic matching[J]. Electronic Technology and Software Engineering, 2022(7): 206-211.
- [4] 李佳歆. 基于序列模型的文本语义匹配方法研究[D]. 重庆: 重庆邮电大学计算机科学与技术学院, 2018.
Li Jia-xin. Research on text semantic matching method based on sequence model[D]. Chongqing: School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, 2018.
- [5] 王娜娜. 基于词袋模型的图像分类方法研究[D]. 兰州: 兰州理工大学计算机与通信学院, 2018.
Wang Na-na. Research on image classification based on bag of words model[D]. Lanzhou: Computer and Communication, Lanzhou University of Technology, 2018.
- [6] 张莹, 亚森·艾则孜, 吴顺祥. 利用 N-gram 和语义分析的维吾尔语文本相似性检测方法[J]. 计算机应用研究, 2019, 36(9): 2722-2725, 2729.
Zhang Ying, Yasen Ai-ze-zi, Wu Shun-xiang. A uighur text similarity detection method using N-gram and semantic analysis[J]. Applied Research of Computers, 2019, 36(9): 2722-2725, 2729.
- [7] 杨名. 面向复杂语境的文档级关系抽取方法研究[D]. 大连: 大连海事大学信息科学技术学院, 2023.
Yang Ming. Research on document-level relation extraction for complex context[D]. Dalian: School of Information Science and Technology, Dalian Maritime University, 2023.
- [8] Huang P S, He X, Gao J, et al. Learning deep structured semantic models for web search using click-through data[C]//Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, Francisco, USA, 2013: 2333-2338.
- [9] Shen Y, He X, Gao J, et al. A latent semantic model with convolutional-pooling structure for information retrieval[C]//Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 2014: 101-110.
- [10] Wang J, Luo L, Wang D. Research on Chinese short text classification based on Word2Vec[J]. Computer Systems & Applications, 2018, 27(5): 209-215.
- [11] 翁兆琦, 张琳. 基于多角度信息交互的文本语义匹配模型[J]. 计算机工程, 2021, 47(10): 97-102.
Weng Zhao-qi, Zhang Lin. Text semantic matching model based on multi-angle information interaction[J]. Computer Engineering, 2019, 47(10): 97-102.
- [12] 徐文峰, 杨艳, 张春风. 融合实体上下文特征的深度文本语义匹配模型[J]. 武汉大学学报: 理学版, 2020, 66(5): 483-494.
Xu Wen-feng, Yang Yan, Zhang Chun-feng. Deep text semantic matching model combining entity context features[J]. Journal of Wuhan University (Science Edition), 2019, 66(5): 483-494.

- [13] 张贞港, 余传明. 基于知识增强的文本语义匹配模型研究[J]. 情报学报, 2024, 43(4): 416-429.
Zhang Zhen-gang, Yu Chuan-ming. Research on text semantic matching model based on knowledge enhancement[J]. Journal of Information Technology, 2024, 43(4): 416-429.
- [14] 赵云肖, 李茹, 李欣杰, 等. 基于汉字形音义多元知识和标签嵌入的文本语义匹配模型[J]. 中文信息学报, 2024, 38(3): 42-55.
Zhao Yun-xiao, Li Ru, Li Xin-jie, et al. Text semantic matching model based on multiple knowledge of Chinese character phonology and tag embedding[J]. Journal of Chinese Information Processing, 2019, 38(3): 42-55.
- [15] 臧洁, 周万林, 王妍. 融合多头注意力机制和孪生网络的语义匹配方法[J]. 计算机科学, 2023, 50(12): 294-301.
Zang Jie, Zhou Wan-lin, Wang Yan. Semantic matching method combining multi-head attention mechanism and twin network[J]. Computer Science, 2023, 50(12): 294-301.
- [16] 于碧辉, 王加存. 孪生网络中文语义匹配方法的研究[J]. 小型微型计算机系统, 2021, 42(2): 231-234.
Yu Bi-hui, Wang Jia-cun. Research on Chinese semantic matching in twin networks[J]. Small and Micro Computer Systems, 2019, 42(2): 231-234.
- [17] 陈源, 丘心颖. 结合自监督学习的多任务文本语义匹配方法[J]. 北京大学学报: 自然科学版, 2022, 58(1): 83-90.
Chen Yuan, Qiu Xin-ying. Multi-task text semantic matching method combined with self-supervised learning[J]. Journal of Peking University (Natural Science Edition), 2012, 58(1): 83-90.
- [18] 王明, 李特, 黄定江. 基于多维特征表示的文本语义匹配[J]. 华东师范大学学报: 自然科学版, 2022(5): 126-135.
Wang Ming, Li Te, Huang Ding-jiang. Text semantic matching based on multidimensional feature representation[J]. Journal of East China Normal University (Natural Science Edition), 2022(5): 126-135.
- [19] 赵小虎, 赵成龙. 基于多特征语义匹配的知识库问答系统[J]. 计算机应用, 2020, 40(7): 1873-1878.
Zhao Xiao-hu, Zhao Cheng-long. Knowledge base question answering system based on multi-feature Semantic matching[J]. Journal of Computer Applications, 2020, 40(7): 1873-1878.
- [20] Xia T, Wang Y, Tian Y, et al. Using prior knowledge to guide bert's attention in semantic textual matching tasks[DB/OL]. [2024-07-25]. <https://doi.org/10.1145/3442381.3449988>
- [21] Wang Z Q, Li P, Zhang H, et al. Multi-passage BERT: a globally normalized BERT model for open-domain question answering[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP), Hong Kong, China, 2019: 4892-4903.
- [22] Alberti C, Lee K, Collins M, et al. A bert base-line for the natural questions[J/OL]. [2024-07-25]. arXiv preprint arXiv:1901.08634
- [23] 宋爽, 陆鑫达. 基于BERT与图像自注意力机制的文本匹配模型[J]. 计算机与现代化, 2021(11): 12-16, 21.
Song Shuang, Lu Xin-da. Text matching model based on BERT and image Self-attention mechanism[J]. Computer and Modernization, 2021(11): 12-16, 21.
- [24] 陈岳林, 高铸成, 蔡晓东. 基于BERT与密集复合网络的长文本语义匹配模型[J]. 吉林大学学报: 工学版, 2024, 54(1): 232-239.
Chen Yue-lin, Gao Zhu-cheng, Cai Xiao-dong. Long text semantic matching model based on BERT and dense composite network[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(1): 232-239.
- [25] 王钦晨, 段利国, 王君山, 等. 基于BERT字句向量与差异注意力的短文本语义匹配策略[J]. 计算机工程与科学, 2024, 46(7): 1321-1330.
Wang Qin-chen, Duan Li-guo, Wang Jun-shan, et al. Short text semantic matching strategy based on BERT sentence vector and differential attention[J]. Computer Engineering and Science, 2019, 46(7): 1321-1330.
- [26] Jawahar G, Sagot B, Seddah D. What does BERT learn about the structure of language?[C]//ACL57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019: 729-739.
- [27] Gao T, Yao X, Chen D. Simcse: simple contrastive learning of sentence embeddings[J/OL]. [2024-07-26]. arXiv preprint arXiv: 2104.08821
- [28] Liu X, Chen Q, Deng C, et al. Lcqm: a large-scale chinese question matching corpus[C]//Proceedings of

- the 27th International Conference on Computational Linguistics, 2018: 1952-1962.
- [29] Chen J, Chen Q, Liu X, et al. The bq corpus: a large-scale domain-specific chinese corpus for sentence semantic equivalence identification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018: 4946-4951.
- [30] Lai Y, Feng Y, Yu X, et al. Lattice CNNs for matching based chinese question answering[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019: 6634-6641.
- [31] 刘东旭, 段利国, 崔娟娟, 等. 融合义原相似度矩阵与字词向量双通道的短文本语义匹配策略[J]. 计算机科学, 2024(12): 250-258.
- [32] Liu Dong-xu, Duan Li-guo, Cui Juan-juan, et al. A short text semantic matching strategy combining semantic similarity matrix and word vector[J]. Computer Science, 2024(12): 250-258.
- [33] Chen L, Zhao Y, Lyu B, et al. Neural graph matching networks for Chinese short text matching[DB/OL]. [2024-07-26]. <https://aclanthology.org/2020.acl-main.547>
- [33] Zhang X, Li Y, Lu W, et al. Intra-correlation encoding for chinese sentence intention matching[DB/OL]. [2024-07-27]. <https://aclanthology.org/2020.coling-main.455>