

基于预训练模型的专家匹配任务分类重构

邓家亮^{1,2}, 祝家东², 邵长城³, 陈国梁⁴, 陈平华¹

(1. 广东工业大学 计算机学院, 广东 广州 510006; 2. 佛山职业技术学院 人工智能学院, 广东 佛山 528137;
3. 广东工业大学 管理学院, 广东 广州 510006; 4. 广东省科技创新监测研究中心, 广东 广州 510033)

摘要: 在要求“小同行”评审的科研项目中, 传统的专家匹配方法通常依赖项目描述与专家简历之间的双文本语义对齐, 但在实际应用中面临专家简历数据稀缺、跨学科术语语义漂移等问题。为此, 提出一种基于预训练模型的匹配任务分类重构模型: 基于论文文本—专家标签数据对, 训练一个从论文文本到专家标签的神经网络, 将项目—专家匹配任务转化为“文本到标签”的分类问题。模型仅需项目描述作为输入, 直接输出专家身份标签, 从而绕过简历依赖并降低术语对齐难度。为解决中文专家标注数据稀缺问题, 引入情感分类作为代理任务, 系统评估 BERT-base-Chinese、CNN、LSTM、RNN 与 SVM/TF-IDF 五类模型的性能差异。实验表明, BERT-base-Chinese 在自建跨学科专家数据集上的准确率和 F1_weighted 值均达到 97% 以上, 明显优于其他模型, 其全词掩码技术与分层微调策略能有效解决术语语义漂移问题。

关键词: 专家匹配; 预训练模型; 文本分类; 跨域术语消歧

DOI: 10.11907/rjdk.251554

扫描二维码阅读全文: 

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2026)004-0123-09

Classification-based Reframing of Expert Matching Tasks Powered by Pretrained Models

DENG Jialiang^{1,2}, ZHU Jiadong², SHAO Changcheng³, CHEN Guoliang⁴, CHEN Pinghua¹

(1. School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China; 2. School of Artificial Intelligence, Foshan Polytechnic, Foshan 528137, China; 3. School of Management, Guangdong University of Technology, Guangzhou 510006, China; 4. Guangdong Provincial Center for Science and Technology Innovation Monitoring, Guangzhou 510033, China)

Abstract: In research project evaluation scenarios requiring "small-circle peer review," traditional expert retrieval/matching methods rely on semantic alignment between project descriptions and expert resumes (dual-text), facing practical challenges such as scarcity of expert resume data and semantic drift of cross-disciplinary terms. To address these issues, this paper proposes a classification-based reframing model for matching tasks using pretrained models. Based on paper text-expert label data pairs, the model trains a neural network that maps paper text to expert labels, transforming the project-expert matching task into a "text-to-label" classification problem. The model requires only the project description as input and directly outputs expert identity labels, thereby bypassing the dependency on resumes and reducing the difficulty of term alignment. To mitigate the scarcity of annotated Chinese expert data, sentiment classification is introduced as a proxy task to systematically evaluate the performance differences among five models: BERT-base-Chinese, CNN, LSTM, RNN, and SVM/TF-IDF. Experiments show that BERT-base-Chinese achieves an accuracy rate above 97% and an F1_weighted-score above 97% on a self-constructed cross-disciplinary dataset, significantly outperforming other models. Its Whole Word Masking (WWM) technique and layered fine-tuning strategy effectively resolve term semantic drift.

Key Words: expert matching; pretrained models; text classification; cross-domain term disambiguation

收稿日期: 2025-10-09

基金项目: 广东省软科学研究计划项目(2025A1010010002); 佛山市密码工程与可信系统重点实验室项目(FS2025016); 广东省普通高校重点科研项目(2024KCXTD090); 佛山职业技术学院校级科研项目(KY2025Z12)

作者简介: 邓家亮(1995-), 男, 广东工业大学计算机学院硕士研究生, 佛山职业技术学院助教, 研究方向为自然语言处理、深度学习; 祝家东(1989-), 男, 硕士, 佛山职业技术学院副教授, 研究方向为深度学习、大模型; 邵长城(1992-), 男, 广东工业大学管理学院博士研究生, 研究方向为推荐系统、深度学习、大模型; 陈国梁(1982-), 男, 广东省科技创新监测研究中心工程师, 研究方向为推荐系统、政务信息化; 陈平华(1967-), 男, 硕士, 广东工业大学计算机学院教授、硕士生导师, 研究方向为大数据、推荐系统、人工智能。本文通讯作者: 陈国梁。

0 引言

在国家和省市自然科学基金等要求“小同行”评审的项目评审中,科学的专家抽取评审机制是公平公正评审的重要保障。传统的评审专家抽取机制依赖“项目描述”和“专家简历”,存在两大难以规避的缺陷:①受隐私保护与更新滞后的影响,专家简历数据可获取性低且时效性差;②同一术语(如“神经网络”“深度学习”等)在不同学科语境中语义漂移明显。随着交叉学科项目日益增多,中文学术场景中高质量跨领域对齐语料严重稀缺,导致跨域术语对齐成本陡增,匹配精度随之下降^[1]。

针对以上问题,本文提出如下创新方案:①任务重构:仅以科研项目描述为唯一输入,直接输出对应专家身份标签,将匹配任务转化为“单文本”到“单标签”映射的分类问题,可以绕过对专家简历的依赖,从而解决数据不可获取性问题,规避对专家简历的依赖^[2];②实验验证:研究引入情感分类作为代理任务,初步筛选和验证不同模型架构在中文语义表征和分类任务上的基础能力。情感分类任务要求模型具备关键词捕捉、语境理解等能力,这些底层能力与专家匹配任务中要求模型识别研究领域关键词、理解其在本学科语境下含义的能力是相通的。因此,在专家标注数据稀缺的前提下,在公开情感数据集上表现优异的模型,其强大的特征提取和语义编码能力有望迁移至专家匹配任务中,为中文专家匹配提供新路径。与现有研究相比,本文的创新性主要体现在以下3个方面:

(1)任务重构的创新。摒弃了传统依赖“项目—简历”双文本语义匹配的方法,首次将专家匹配任务重构为“项目描述”到“专家标签”的单文本分类问题。这一重构从根本上摆脱了对难以获取的专家简历数据的依赖,将复杂的跨文本语义对齐问题转化为单文本内部的语义理解与分类问题,显著降低了任务复杂度。

(2)代理任务的引入与迁移机制。针对中文专家标注数据稀缺的痛点,引入情感分类这一公开资源丰富的任务作为代理任务进行模型验证与筛选。情感分类与专家匹配同属语义理解范畴,模型在情感分类中学习到的语义特征提取、上下文理解等能力可有效迁移至专家匹配任务。此方法为小样本场景下的模型性能预评估提供了可行路径。

(3)跨域术语消歧的有效性。系统验证了预训练模型(特别是采用全词掩码技术的BERT-base-Chinese)在解决跨学科术语语义漂移问题上的卓越性能。通过分层微调等策略,模型能够精准捕捉术语在特定学科语境下的真实含义,实现了优于传统方法的匹配精度。

1 相关工作

1.1 专家匹配研究现状

现有的专家匹配研究与实践主要遵循“基于规则与知

识库”和“基于双文本语义匹配”两种方法:①早期系统依赖专家自行填报的研究领域关键词(如“机器学习”“生物医学”),通过关键词词频^[3]、热点主题^[4]来匹配实现项目与专家的粗粒度对齐。该方法严重依赖关键词体系的完备性与专家填报的准确性,无法处理一词多义(语义漂移)和新兴交叉领域术语,自动化程度低;②随着NLP技术的发展,研究者尝试挖掘项目描述与专家简历(如已发表论文、专利摘要)的深层语义关联。这类方法通常采用传统的文本匹配技术,或采用深度神经网络、预训练模型计算两个文本的语义相似度^[5-6]。

尽管第二种方法取得了一定进展,但其根本瓶颈在于对“项目—简历”双文本数据的重度依赖,这带来了数据可获取性差和跨领域对齐难度大双重挑战。因此,探索一种不依赖专家简历数据的新匹配方法,已成为该领域一个亟待突破的问题。

1.2 文本匹配技术综述

在自然语言处理(NLP)领域,传统文本匹配任务一般有两种技术路线^[7]:①把文本在向量空间映射(Word2vec)后,通过连续度量评估两段文本语义贴近程度的“相似度计算”^[8];②对匹配问题进行特征工程处理,然后交给分类器,将其转化为“匹配/不匹配”的二元相关性分类。两种路线均共享底层的文本表征技术,但任务目标本质不同,他们最终回答了不同的问题,相似度计算输出连续相关性,而相关性分类输出离散决策。传统文本匹配技术路线,如图1所示。文本匹配技术路线不同的核心在于输出形式。

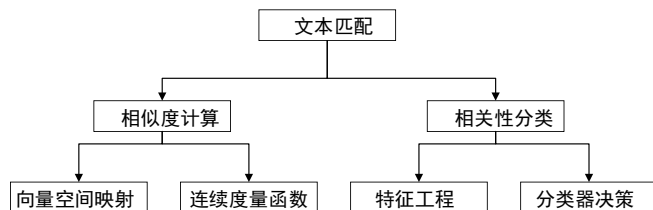


Fig. 1 Traditional text matching technical approaches

图1 传统文本匹配技术路线

1.2.1 相似度计算

相似度计算是一种广泛应用于文本匹配的方法,其核心在于通过数学公式或模型来衡量两段文本之间的相似程度。常见的步骤如下:首先,对文本进行分词、去除停用词等操作,以提取有意义的特征^[9];其次,针对预处理后的文本,通过词袋模型(Bag-of-Words, BOW)、TF-IDF、词嵌入(Word Embedding)等方法,将其转换为数值向量(即文本特征)^[10];然后,根据转换后的向量,使用余弦相似度(Cosine Similarity)^[11]、Jaccard相似度^[12]、编辑距离等方法计算文本之间的相似度;最终,根据计算出的相似度值,对结果进行排序,选取最相似的文本作为匹配结果。

TF-IDF(Term Frequency-Inverse Document Frequency)是一种用于衡量一个词 w 在文档 d_i 中的重要性的统计方法。它结合了词频(Term Frequency, TF)和逆文档频率(In-

verse Document Frequency, IDF), 如公式(1)所示。

$$TF-IDF(w, d_i) = tf_{w, d_i} \cdot \log \frac{N}{df_w} \quad (1)$$

其中, N 是文档总数, df_w 是包含词 w 的文档数量。

余弦相似度(Cosine Similarity)是衡量两个向量之间夹角余弦值的指标,用于评估两个文本向量之间的相似性。其值范围为 $[-1, 1]$,值越接近1表示两个向量越相似,如公式(2)所示。

$$\text{CosineSimilarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (2)$$

其中, A 和 B 分别表示两个文本向量, $\cdot$ 表示点积, $\|A\|$ 和 $\|B\|$ 分别表示向量范式。

Jaccard 相似度是一种基于集合论的相似度度量方法,用于衡量两个集合之间的相似性。其值范围为 $[0, 1]$,值越大表示两个集合越相似,如公式(3)所示。

$$\text{Jaccard Similarity} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

1.2.2 相关性分类

相关性分类是一种将文本匹配问题转化为分类问题的技术路线。在这种方法中,文本匹配任务被建模为一个二分类问题,即判断两个文本是否匹配(1)或不匹配(0),其优势在于可以利用现有的分类算法(如逻辑回归、支持向量机、神经网络等)来解决匹配问题^[13]。相关性分类的方法一般包括以下步骤:首先,从文本中提取特征,如词频、TF-IDF、词嵌入(Word Embedding)等;其次,对特征使用分类算法对提取的特征进行训练,以学习如何区分匹配和不匹配的文本对;最终,使用新的文本对进行预测,并通过准确率、召回率、F1值等指标评估模型的性能。

准确率表示模型预测正确的样本比例,是评估整体性能的基础指标,如公式(4)所示。在专家匹配任务中,其扩展为多分类场景下正确预测样本的总占比,反映模型全局识别专家身份的可靠性。高准确率表明模型能精准映射项目描述至对应专家ID,减少误判风险。然而,该指标在类别分布不均衡时可能掩盖少数专家类别的识别偏差,如公式(4)所示。

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

其中, TP 是真正例(True Positive),即模型预测为正例,且实际为正例(项目描述匹配专家A,模型正确输出专家A); TN 是真负例(True Negative),即模型预测为负例,且实际为负例(项目描述不匹配专家B,模型正确排除专家B); FP 是假正例(False Positive),即模型误报,模型预测为正例,但实际为负例(项目描述不匹配专家C,模型错误输出专家C); FN 是假负例(False Negative),即模型漏报,模型预测为负例,但实际为正例(项目描述匹配专家D,模型错误排除专家D)。

召回率衡量模型识别正类样本的完备性,定义为实际正类样本中被正确预测的比例,如公式(5)所示。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

在专家匹配中,召回率评估模型发现所有相关专家的能力。高召回率可最小化漏报风险(FN),避免在交叉学科项目中遗漏本应匹配的专家,对保障评审公平性与资源利用率至关重要。鉴于专家标签可能存在样本不均衡,宏平均召回率(Recall_macro)更能反映模型对少数专家类别的鲁棒性。

F1值作为精确率与召回率的调和均值,是综合评估分类性能的核心指标,尤其适用于类别不平衡或需平衡误报和漏报代价的任务,如公式(6)所示。

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

在专家匹配任务中,F1值量化模型兼顾匹配精准性与覆盖完备性的能力,具体表现为减少误报(FP)。例如,避免将项目错误分配至无关专家,同时减少漏报(FN),确保专家不被遗漏。该指标在跨学科研究场景中具有重要意义,尤其是在涉及人工智能与管理学术语交叉的研究项目中。它可直接验证预训练模型通过单文本内部语义解析来解决术语歧义问题的有效性,例如“深度学习”在人工智能^[14]和教育心理学^[15]语境中的语义漂移情况。

1.3 预训练模型

随着数据规模的指数级增长,深度学习技术实现重大突破。然而,模型性能高度依赖大规模高质量标注数据,其收集与处理过程面临显著挑战^[16]。预训练模型(Pre-trained Models)通过在大规模无标注语料上学习通用语义表征,有效降低了下游任务对标注数据依赖的数据稀缺瓶颈问题,建立了通用语义空间到垂直领域的知识迁移通道。通过上下文编码解决跨域术语歧义,成为提升训练效率的核心技术路径。2017年提出的Transformer架构对序列建模方法进行了革新^[17]。其核心创新在于:①实现长距离语义关联建模的自注意力机制;②显著提升训练效率的并行计算架构;③在非序列结构中保留文本顺序信息的位置编码技术。这三者协同突破递归神经网络(RNN)的序列计算限制。

作为Transformer的里程碑式应用,原始BERT(Bidirectional Encoder Representations from Transformers)通过掩码语言模型(Masked Language Model, MLM)随机遮蔽15%输入词元并预测原值,学习深层语境表征^[18]。此外,采用下一句预测(Next Sentence Prediction, NSP)判断句子连贯性,增强语义关系理解。

针对中文场景专项优化的“BERT-base-Chinese”,继承了原始BERT的12层Transformer编码器架构,并针对中文特性引入了全词掩码(Whole Word Masking, WWM)技术,通过遮蔽完整词语来增强对术语边界的感知能力。此外,在训练策略方面也有所调整,BERT模型原始NSP任务被移除,仅保留MLM任务进行优化。基于百科/新闻/论坛等100GB中文语料训练,在CLUE中文基准测试中文本分

类平均F1达89.7%,证实其对复杂语义场景的适应性,展现出对中文复杂语义场景的深度理解能力,包括跨领域术语消歧、长距离依赖解析以及专业语境建模。

2 模型重构

2.1 任务定义

本文的核心任务分为两个阶段:情感分类(代理任务)和专家匹配(目标任务)。

情感分类任务定义为 $x \rightarrow y \in \{正, 负\}$,其中,输入 x 为单条文本(如用户评论),输出 y 为情感极性标签(正面或负面)。

专家匹配任务重构为 $x \rightarrow y \in \{E_1, \dots, E_k\}$,输入 x 为科研项目描述文本,输出 y 为离散的专家身份标签(对应专家ID)。这种重构避免了传统匹配中对专家简历的依赖,将跨域对齐难题转化为单文本内部语义解析,显著提升数据可行性和计算效率。

2.2 传统专家匹配模型

传统的专家匹配模型被定义为一个双文本匹配问题。给定一个项目描述 d 和一个专家简历 r ,模型的目标是学习一个匹配函数 f ,输出一个匹配分数 $s=f(d, r)$,用以衡量两者之间的相关性。其架构通常如图2所示。

模型接收两个输入,通过编码器(如BERT)获取各自的表征,再通过一个交互层(如计算余弦相似度、拼接后过分类器)得到最终的匹配决策。该方法的弊端在于严重依赖大规模、高质量的配对数据(项目描述 d ,专家简历 r)进行训练或计算,而这在现实中难以获取。

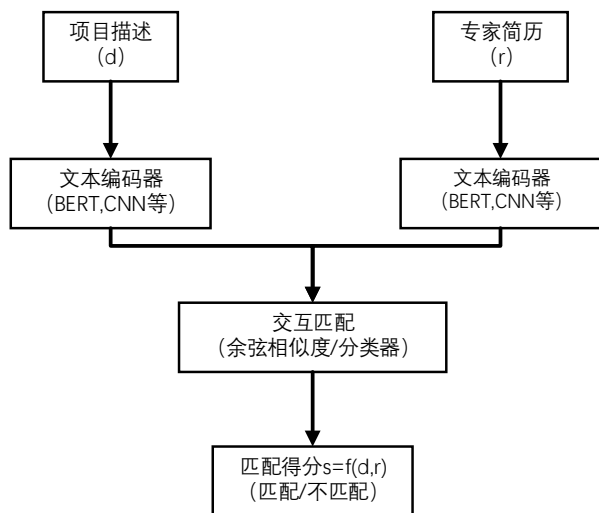


Fig. 2 Schematic diagram of the traditional expert matching model

图2 传统专家匹配模型原理图

2.3 重构后的专家匹配模型

本文彻底摒弃了双文本匹配的方法,将专家匹配重构为一个单文本多分类任务。

2.3.1 模型原理

重构后的模型如图3所示,仅以项目描述 d 为输入,通

过一个分类模型,直接输出其属于第 k 个专家 E_k 的概率 $P(y=E_k|d)$ 。其中, $k \in [1, k]$, K 为专家库中的专家总数。

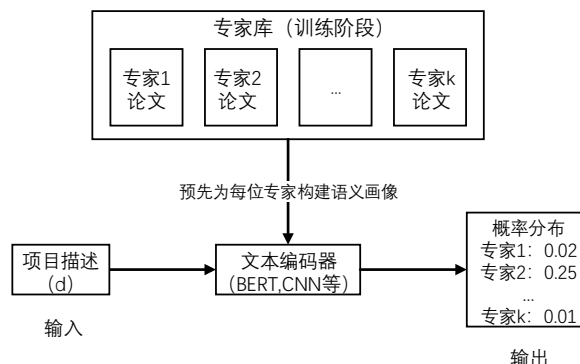


Fig. 3 Schematic diagram of the reframed expert matching model

图3 重构专家匹配模型原理图

该模型为单输入分类结构,关键在于其训练数据源自公开的专家论文,而非保密的内部简历。模型通过学习论文内容与专家ID的映射关系,规避了对“项目—简历”配对数据的直接依赖,解决了数据稀缺的瓶颈。在训练阶段,专家知识(论文)仅用于构建带标签的训练集;在推理阶段,模型仅需项目描述即可直接输出专家ID概率分布。

2.3.2 模型训练

模型的训练目标是最小化预测分布与真实标签之间的交叉熵损失。对于一条训练数据 (d_i, y_i) ,其中, y_i 是表示专家ID的one-hot向量,损失函数定义如公式(7)所示。

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{i,k} \cdot \log(P(y_i = e_k | d_i)) \quad (7)$$

2.3.3 模型推理

在推理阶段,给定一个新的项目描述 d_{new} ,训练好的模型会输出一个在所有 k 个专家上的概率分布,最终选择概率最大的专家作为匹配结果,如公式(8)所示。

$$\hat{y} = \arg \max_k P(y = e_k | d_{new}) \quad (8)$$

3 实验

3.1 基线模型

为系统验证不同模型在单文本分类任务中的适应性,本文选取5类文本分类模型进行对比实验,涵盖传统机器学习(SVM/TF-IDF组合)、基础神经网络(CNN^[19]、LSTM^[20]、RNN)及预训练模型(BERT-base-Chinese),各模型在统一实验环境NVIDIA GeForce GTX 1050Ti GPU上运行,架构细节如下:

(1)SVM/TF-IDF。作为传统机器学习基线,验证基于统计特征的分类方法在专家匹配任务中的有效性。TF-IDF负责文本向量化,SVM负责构建分类决策边界,共同验证简单特征在跨域任务中的局限性。在本文中采用特征工程TfidfVectorizer(max_features=10 000),保留最高频10 000个词项并包含1-3元语法特征;分类器使用线性核

SVM(kernel='linear'),正则化系数 $C=0.8$ 平衡过拟合风险;通过 One-vs-Rest 策略实现多分类扩展,支持概率输出(probability=True)用于置信度评估;特征维度为 10 000 维,直接反映术语频率分布特征。

(2)卷积神经网络(CNN)。本模型可以验证局部语义模式捕捉能力在专家匹配中的有效性。通过卷积核提取 n -gram 级术语特征,适用于中文文本的短距离结构识别。

本文架构使用 BERT 词表初始化嵌入层,维度统一为 128;卷积操作保留位置信息(padding=1 保持序列长度不变);全局最大池化压缩特征维度,降低参数量;输出层直接映射到情感任务 2 类/专家匹配任务对应类别数(6, 15, 25 类)的对应分类空间。本文所采用的卷积神经网络通过多层卷积核提取 n -gram 级局部术语特征,结合全局最大池化捕捉核心语义,其具体架构如图 4 所示。

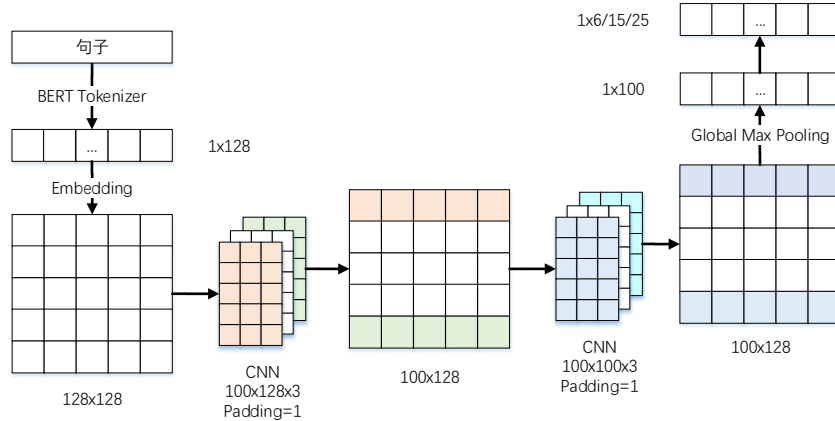


Fig. 4 Convolutional neural network

图4 卷积神经网络

(3)长短期记忆网络(LSTM)。本文模型用于验证序列建模能力对科研项目描述的理解效果。通过门控机制建模长距离依赖关系,解决梯度消失问题。采用单向 LSTM 结构(非双向)架构,隐藏单元 256 个;仅提取最终时间步隐

藏状态作为序列表征;Adam 优化器(lr=0.001)配合交叉熵损失函数;训练 10 个 epoch,批次大小 64 适配 GPU 显存。本文利用单向 LSTM 的门控机制对科研项目描述进行时序建模,以捕捉长距离的语义依赖关系,其结构如图 5 所示。

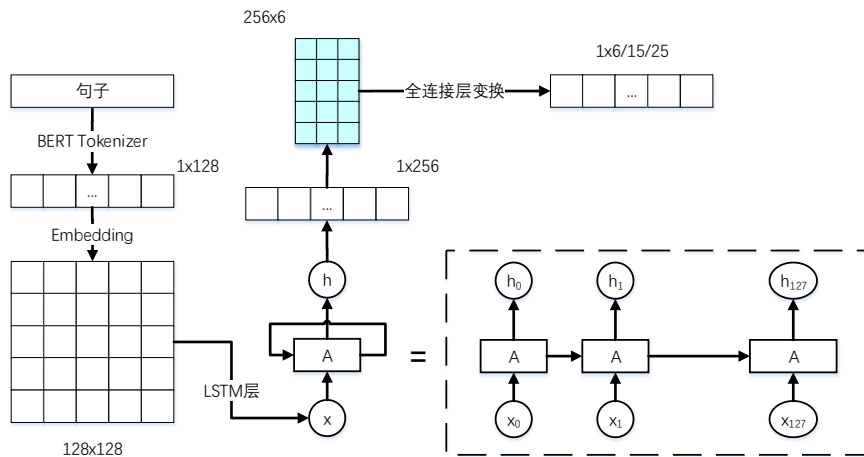


Fig. 5 Long short-term memory network

图5 长短期记忆网络

(4)循环神经网络(RNN)。本文模型在研究中作为时序模型基线,显式验证梯度消失问题对专家匹配任务的影响。与 LSTM 同参数规模对比,凸显门控机制在长文本建模中的必要性。采用基础 RNN 单元(非 GRU/LSTM 变体);Tanh 激活函数,隐藏层 256 单元;与 LSTM 相同的嵌入维度和隐藏层规模;作为对照组暴露长序列建模缺陷。本文构建了基础循环神经网络作为对比基线,旨在显式验证梯度消失现象对长文本解析的影响,其逻辑结构如图 6 所示。

(5)BERT-base-Chinese。基于 Transformer 的预训练

模型,专为中文优化,旨在验证预训练语言模型在跨域术语消歧中的优势。通过迁移大规模通用语义知识,解决专业术语在不同学科的语义漂移问题。架构包含 12 层编码器,采用全词掩码(WWM)技术适配中文分词特性;移除 NSP 任务,专注 MLM 表示的迁移学习;微调 3 个 epoch,学习率 $2e-5$,批次大小 16;采用梯度累积(steps=4)解决 GTX 1050Ti 显存限制;进行分层微调策略,顶层参数高学习率($5e-5$),底层参数低学习率($1e-5$)^[15]。本文旨在验证任务重构本身的有效性,由于 BERT-base-Chinese 能代表预训

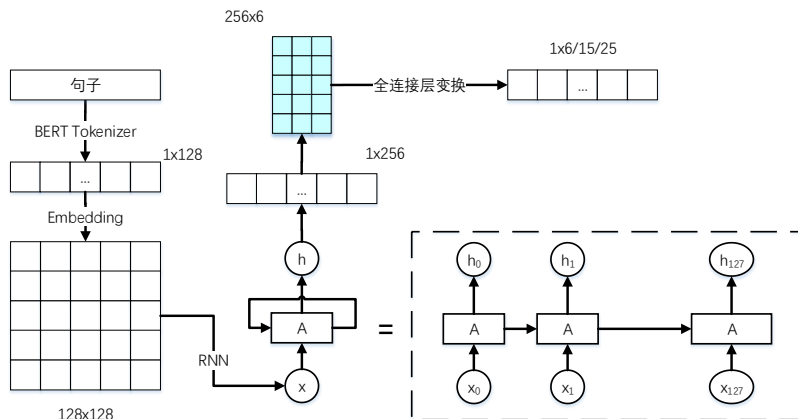


Fig. 6 Recurrent neural network

图6 循环神经网络

练模型的先进水平,因此将其作为核心预训练模型。本文核心采用的 BERT-base-Chinese 模型基于 12 层 Transform-

er 架构,通过全词掩码技术实现深层语境表征,其整体架构如图 7 所示。

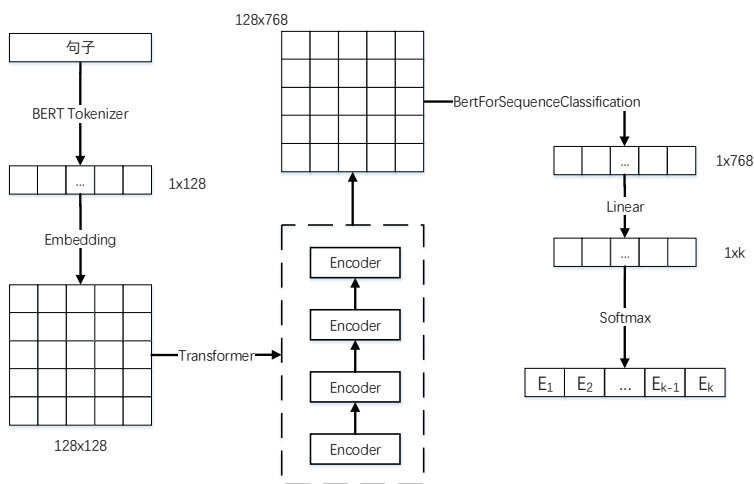


Fig. 7 BERT-base-Chinese architecture

图7 BERT-base-Chinese 架构

3.2 数据集

为系统验证不同模型在专家匹配任务中的性能表现,本文构建了两类数据集进行对比:情感分类基准数据集(用于代理任务验证)和专家匹配核心数据集(用于目标任务验证)。通过代理任务与目标任务的递进验证,系统评估模型性能。其中,代理任务用于验证语义基础能力,初步确定实验中 BERT-base-Chinese 的优势;目标任务则用于验证初步的猜想,从而系统评估模型性能。

3.2.1 情感分类基准数据集

(1)ChnSentiCorp_hdt_all(酒店评论)。包含 7 766 条真实用户评论,构成自然语言理解试金石。其中,5 322 条正向评价是聚焦服务体验亮点,2 444 条负向评价揭示了痛点表述模式。该数据集用于验证模型在开放域中文场景的语义解析鲁棒性。

(2)waimai_10k(外卖评价)。某外卖平台收集的 11 987 条用户评价,生活化短文本构成细粒度情感识别挑战,包含 4 000 条正向评价与 7 987 条负向评价。其口语化表达和领域术语有效测试模型对非规范文本的适应能力。

3.2.2 专家匹配核心数据集

为系统验证模型性能随专家规模扩展的变化趋势,本文分阶段构建了 3 个不同规模的专家数据集。

(1)myself-6(6 名专家)。初始数据集,包含来自 6 个不同学科的 6 位专家的论文文本片段,用于概念验证。

(2)myself-15(15 名专家)。扩充数据集,在 myself-6 的基础上,新增 9 位专家,共包含 15 位来自不同学科的专家。

(3)myself-25(25 名专家)。大规模数据集,在 myself-15 的基础上,再新增 10 位专家,共包含 25 位来自广泛学科的专家。

所有论文文本均从 CNKI 下载,统一处理为长度不超过 500 字符的文本片段,并标注对应专家 ID。

3.3 实验设置

为了在有限硬件条件下,通过批次/轮次调节平衡计算效率与模型收敛,实验配置遵循“控制变量”原则,确保结果可比性,在 NVIDIA GTX 1050Ti 硬件约束下优化计算效率。

3.3.1 统一基础配置

实验在 NVIDIA GeForce GTX 1050 Ti GPU(4 GB 显存)硬件平台上执行,为确保实验可复现性,所有随机种子固定为 42,训练集与测试集按 8:2 比例分割。采用 Adam 优化器进行模型训练,基础学习率设置为 0.001,配合交叉熵损失函数实现梯度更新。评估阶段统一采用五项核心指标:准确率(Accuracy)、加权 F1 值(F1_weighted)、宏平均 F1 值(F1_macro)、加权召回率(Recall_weighted)和宏平均召回率(Recall_macro)。为适配显存限制,BERT 模型采用梯度累积策略(累积步数=4),所有文本输入统一截断为最大长度 128 个词元,使用 BERT 分词器进行标准化预处理。

3.3.2 分模型实现细节

(1)BERT 微调模型。基于 BERT-base-Chinese 预训练模型,继承其 12 层 Transformer 编码器架构,移除原始 NSP 任务,专注 MLM 语义表示迁移。微调阶段采用分层学习率策略:顶层参数学习率设置为 5e-5,底层参数学习率设置为 1e-5,全局学习率 2e-5,训练 3 个 epoch。批次大小动态调整为训练阶段 16/评估阶段 64,通过 TrainingArguments 配置关闭模型保存(save_strategy="no")以提升效率。输出层直接映射到分类空间(情感任务 2 类/专家匹配 6、15、25 类)。

(2)SVM/TF-IDF 模型。文本向量化采用 TfidfVectorizer 生成 10 000 维稀疏特征矩阵,包含 1—3 元语法特征以捕捉局部术语组合。支持向量机选用线性核(kernel='linear'),正则化系数 C=0.8 平衡过拟合风险,启用概率输出(probability=True)支持置信度分析。多分类场景采用 One-vs-Rest 策略扩展,决策函数基于 LibSVM 库实现,特征工程与分类器训练均在 CPU 端执行。

(3)卷积神经网络(CNN)。嵌入层使用 BERT 词表初始化,维度固定为 128。卷积架构包含两个卷积层(通道数 100,卷积核尺寸 3,填充模式为 1),通过 ReLU 激活函数与全局最大池化压缩特征维度。输出层直接映射至标签空间,优化器采用 Adam(lr=0.001),训练 10 个 epoch,批次大小 64。卷积操作保留序列位置信息(padding=1),专门设计用于捕捉“深度学习”等局部术语模式。

(4)长短期记忆网络(LSTM)。单向 LSTM 结构设计隐藏单元 256 个,仅提取最终时间步状态作为序列表征。嵌入维度 128,与 CNN 共享词表初始化策略。训练配置与 CNN 一致(10 epoch, batch_size=64, lr=0.001),通过门控机制建模长距离依赖,验证其对跨句术语关联的解析能力。

(5)循环神经网络(RNN)。基础 RNN 单元(非 GRU/LSTM 变体)作为对照组,隐藏层 256 单元,Tanh 激活函数。与 LSTM 同参数规模(嵌入维度 128,训练轮次 10),显式暴露梯度消失问题对长文本建模的影响,为 LSTM 的性能优势提供对比基准。

3.4 结果与分析

本实验以分类准确率为核心评估指标,系统分析 5 类

模型在两个公开情感数据集(ChnSentiCorp_htl_all, waimai_10k)和 3 个不同规模专家匹配数据集(myself-6, myself-15, myself-25)上的性能表现,如表 1—表 5 所示。实验结果表明,预训练模型 BERT 在不同领域和不同任务规模下均展现出显著且稳定的优势。

Table 1 Key metrics (ChnSentiCorp_htl_all)
表 1 主要指标(ChnSentiCorp_htl_all)

model	accuracy	f1_weighted	f1_macro	recall_weighted	recall_macro
BERT (fine-tuned)	0.907 276	0.906 725	0.893 131	0.907 276	0.888 719
SVM/TF-IDF	0.751 449	0.716 889	0.651 731	0.751 449	0.642 959
LSTM	0.804 250	0.803 523	0.775 352	0.804 250	0.773 395
CNN	0.873 793	0.873 114	0.854 696	0.873 793	0.851 085
RNN	0.674 823	0.543 802	0.402 922	0.674 823	0.500 000

Table 2 Key metrics (waimai_10k)
表 2 主要指标(waimai_10k)

model	accuracy	f1_weighted	f1_macro	recall_weighted	recall_macro
BERT (fine-tuned)	0.917 848	0.917 886	0.907 932	0.917 848	0.908 309
SVM/TF-IDF	0.805 254	0.788 746	0.750 045	0.805 254	0.729 611
LSTM	0.892 827	0.893 315	0.880 895	0.892 827	0.884 557
CNN	0.897 415	0.896 742	0.883 452	0.897 415	0.878 761
RNN	0.664 721	0.530 844	0.399 299	0.664 721	0.500 000

Table 3 Key metrics (myself-6)
表 3 主要指标(myself-6)

model	accuracy	f1_weighted	f1_macro	recall_weighted	recall_macro
BERT (fine-tuned)	0.978 723	0.978 709	0.976 455	0.978 723	0.976 699
SVM/TF-IDF	0.567 921	0.536 681	0.484 198	0.567 921	0.419 889
LSTM	0.710 311	0.680 519	0.540 606	0.710 311	0.524 358
CNN	0.968 903	0.968 888	0.967 908	0.968 903	0.964 438
RNN	0.358 429	0.286 114	0.181 661	0.358 429	0.210 257

Table 4 Key metrics (myself-15)
表 4 主要指标(myself-15)

model	accuracy	f1_weighted	f1_macro	recall_weighted	recall_macro
BERT (fine-tuned)	0.977 251	0.977 188	0.968 180	0.977 251	0.965 382
SVM/TF-IDF	0.228 436	0.106 353	0.070 933	0.228 436	0.091 504
LSTM	0.596 209	0.536 084	0.360 603	0.596 209	0.374 865
CNN	0.943 128	0.941 986	0.925 322	0.943 128	0.912 541
RNN	0.291 943	0.252 247	0.175 964	0.291 943	0.171 314

Table 5 Key metrics (myself-25)
表 5 主要指标(myself-25)

model	accuracy	f1_weighted	f1_macro	recall_weighted	recall_macro
BERT (fine-tuned)	0.973 514	0.973 393	0.971 427	0.973 514	0.969 381
SVM/TF-IDF	0.128 311	0.046 315	0.041 057	0.128 311	0.057 175
LSTM	0.721 601	0.711 311	0.604 788	0.721 601	0.609 894
CNN	0.924 073	0.923 873	0.911 328	0.924 073	0.900 861
RNN	0.186 580	0.170 612	0.130 027	0.186 580	0.126 830

3.4.1 情感分类任务

在情感分类任务中,BERT 模型凭借其强大的语义表征能力,在两个数据集上均取得了最高的分类准确率。在

酒店评论数据集(ChnSentiCorp_h1l_all)上,BERT的准确率达90.727 6%,在外卖评论数据集(waimai_10k)上达到91.784 8%。这表明BERT对正式评论文本和非正式口语化文本均具备优异的泛化能力。

相比之下,传统模型表现出了明显的局限性。SVM/TF-IDF模型在两个情感数据集上的准确率分别为75.144 9%和80.525 4%,其性能依赖于人工特征工程,难以捕捉深层语义。RNN模型因梯度消失问题,准确率最低(67.482 3%,66.472 1%),在处理长文本时几乎失效。CNN与LSTM模型表现居中,准确率均超过80%,证明了神经网络模型在特征自动提取方面的有效性,但其性能仍与BERT存在显著差距(约3~10个百分点)。

3.4.2 专家匹配任务

专家匹配任务的结果有力验证了本文任务重构方法的有效性,并揭示了模型性能随任务规模(专家数量)扩大的变化趋势。本文专家匹配任务按照3种规模三步走的方式进行。

(1)小规模(6类)验证可行性。由表3可知,在6位专家的匹配任务中,所有模型的准确率均达到最高。BERT模型以97.872 3%的准确率排在首位,充分证明了“单文本”到“单标签”重构路径的可行性。CNN模型以96.890 3%的准确率紧随其后,展现出作为轻量化替代方案的巨大潜力。而SVM和RNN模型准确率仅为56.792 1%和35.842 9%,难以对项目进行专家匹配。

(2)中规模(15类)凸显稳定性。由表4可知,当专家数量扩展至15人时,模型间性能差距进一步拉大。BERT模型的准确率依然保持在97.725 1%的极高水准,波动极小(较6类任务仅下降0.147 2%),展现了惊人的稳定性。CNN模型准确率下降至94.312 8%,但仍维持在可用水平。传统模型性能出现断崖式下跌,SVM准确率暴跌至22.843 6%,RNN也仅为29.194 3%,表明其难以处理类别增长带来的复杂度提升。

(3)较大规模(25类)证明泛化性与可扩展性。由表5可知,在25位专家的更具挑战性的场景下,BERT模型的准确率依然高达97.351 4%,相较于6类任务仅微降0.520 9%,充分证明了其优异的泛化能力和可扩展性。CNN模型准确率为92.407 3%,虽有所下降,但仍显著优于其他基线模型,其实用性再次得到确认。与此形成鲜明对比的是,SVM和RNN模型准确率已分别降至12.831 1%和18.658 0%,其性能已趋近随机猜测水平,完全无法满足实际应用需求。

3.5 综合讨论

3.5.1 量化迁移增益与规模扩展性

实验数据展现了预训练模型的迁移增益与规模扩展性。在情感分类代理任务中,BERT模型展现出卓越的语义解析能力。通过对近两万条涵盖酒店服务与口语化评价数据的学习,其加权F1值达到0.912,验证了其跨领域语义理解的稳定性。当该模型迁移至专家匹配目标任务后,

其性能优势得到延续,在myself-25数据集上的准确率依然高达97.351 4%。这一跨越领域的性能保持,验证了基于预训练模型的迁移学习策略在解决专家匹配问题上的高效性^[21]。更重要的是,随着专家类别从6个扩展至25个,BERT的准确率仅下降0.520 9个百分点,展现出卓越的任务规模扩展性,为后续扩展到真实大规模专家库提供了坚实的数据支撑。

3.5.2 BERT跨域优势原理

BERT在跨领域专家匹配任务中保持的高准确率源于其预训练语料、模型结构与微调策略的协同作用。首先,模型在100 GB通用中文语料上预训练积累的多学科术语知识,为领域迁移奠定基础;其次,全词掩码(WWM)技术以整词为掩码单元,强化了中文术语边界感知,有效缓解了“神经网络”“深度学习”等词语的跨学科语义漂移;最后,分层微调策略在顶层使用较高学习率(5e-5),快速适配专家标签分布,在底层保持较低学习率(1e-5)以稳固通用语义表示,从而兼顾领域适应性与知识稳定性^[22]。

3.5.3 CNN在专家任务中的潜力与局限性

CNN模型在各类专家规模任务中准确率稳定在第二位(25类任务中92.407 3%),凸显了其在特定场景下的重要价值。其优势在于:①局部语义模式捕捉:尺寸为3的卷积核能够高效识别关键局部术语组合^[23];②计算效率:全局最大池化层显著降低了模型参数量和计算复杂度;③资源友好:在显存受限的硬件环境下,CNN是BERT的高性能轻量化替代方案。然而,随着任务复杂度继续增加,CNN与BERT之间的性能差距逐渐扩大(从6类的1.0%到25类的4.9%),表明其在极端复杂场景下的表征能力仍有限。

3.5.4 传统模型失效分析

本文的实验结果凸显了传统模型在专家匹配任务上的显著局限:①RNN的梯度消失问题:基础RNN单元在处理长文本序列时,因梯度消失问题导致其在25类任务中准确率仅18.658 0%;②SVM/TF-IDF的特征稀疏性与跨域失效:依赖词频统计特征的SVM模型在跨域术语对齐上表现最差(25类任务准确率12.831 1%),其根本原因在于无法理解词语在特定学科语境下的语义差异;③统计模型对类别增长的敏感性:在专家规模扩大时,传统模型性能急剧下降,而深度学习模型表现出更好的可扩展性。

4 结语

本文提出了一种基于预训练模型的专家匹配任务分类重构方法,通过将“项目—专家”匹配转化为单文本分类问题,摆脱了对专家简历数据的依赖。实验证明,以BERT-base-Chinese为核心的重构模型在6类、15类及25类任务中均保持了97%以上的准确率,优于传统的统计和基础神经网络模型,验证了该路径在跨学科术语消歧与规

模扩展上的有效性。同时,代理任务机制的引入成功实现了语义表征能力的预筛选,为专家数据稀缺场景下的匹配优化提供了新思路。

尽管本文验证了重构方案的可行性,但受限于实验规模,未来仍需在以下方面深入研究:①进一步扩展专家库至千人级别,探索层级分类与元学习策略以应对超大规模分类挑战;②引入Qwen、ERNIE等能力更强的最新大语言模型,并尝试融合专利、项目经历等多源异构信息构建全方位专家画像;③推动框架从单标签分类向多标签、重排序方向演进,并引入注意力可视化等技术增强匹配结果的可解释性,从而更好地服务于现实科研评审需求。

参考文献:

- [1] JIANG X Y, LI D L. Research on the construction of a multi-level intelligent project review expert management system: optimization based on the management of scientific research project review experts in Guangdong [J]. Science and Technology Management Research, 2018, 38(16): 212-217.
江笑颜,李栋亮.多层次、智能化的项目评审专家管理体系构建研究——基于广东科研项目评审专家管理的优化[J].科技管理研究, 2018, 38(16): 212-217.
- [2] YU Y, FU Y, WU X P. A survey of Chinese text classification methods [J]. Journal of Cyber Security, 2019, 5(5): 1-8.
于游,付钰,吴晓平.中文文本分类方法综述[J].网络与信息安全学报, 2019, 5(5): 1-8.
- [3] KUNCORO B A, ISWANTO B H. TF-IDF method in ranking keywords of Instagram users' image captions [C]//Piscataway: Proceedings of 2015 International Conference on Information Technology Systems and Innovation (ICITSI), 2016.
- [4] ZHENG Y, MENG Z P, XU C. A short-text oriented clustering method for hot topics extraction [J]. International Journal of Software Engineering and Knowledge Engineering, 2015, 25(3): 453-471.
- [5] LIN Y, ZHANG J S, LIU Z H, et al. Application of technical review expert selection method in disruptive technology expert prediction platform [J]. China Science & Technology Resources Review, 2024, 56(2): 54-62.
林毅,张均胜,刘志辉,等.技术评审专家遴选方法在颠覆性技术专家预判平台上的应用[J].中国科技资源导刊, 2024, 56(2): 54-62.
- [6] XIE X J, LIANG Y, WANG Z S, et al. Semantic matching method for academic expertise of experts integrating multi-source information [J]. Chinese Journal of Computers, 2023, 46(5): 1045-1059.
谢小杰,梁英,王梓森,等.融合多源信息的专家学术专长语义匹配方法[J].计算机学报, 2023, 46(5): 1045-1059.
- [7] WEI X L, SUN X. Research progress and development trends of natural language processing technologies and methods: a review [J]. China-Arab Science and Technology Forum (Chinese and English), 2025(5): 84-88.
卫宣伶,孙翔.自然语言处理技术方法研究进展与发展趋势综述[J].中阿科技论坛(中英文), 2025(5): 84-88.
- [8] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [DB/OL]. https://arxiv.org/abs/1310.4546.
- [9] CHANG J Q, SHEN W. Research on Chinese word segmentation algorithm based on string matching [J]. Industrial Control Computer, 2016, 29(2): 115-116, 119.
常建秋,沈炜.基于字符串匹配的中文分词算法的研究[J].工业控制计算机, 2016, 29(2): 115-116, 119.
- [10] MALDONADO S, LÓPEZ J. Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification [J]. Applied Soft Computing, 2018, 67: 94-105.
- [11] SALTON G, WONG A, YANG C S. A vector space model for automatic indexing [J]. Communications of the ACM, 1975, 18(11): 613-620.
- [12] JACCARD P. The distribution of the flora in the alpine zone [J]. New Phytologist, 1912, 11(2): 37-50.
- [13] HE Y. Research on support vector machines in text classification [J]. Henan Science and Technology, 2019(29): 8-10.
何焱.文本分类中支持向量机研究[J].河南科技, 2019(29): 8-10.
- [14] YANG Q, WANG A Z, WU J T, et al. A survey of dichotomous image segmentation based on deep learning [J/OL]. Computer Engineering and Applications, 2025-09-14. https://link.cnki.net/urlid/11.2127.TP.20250818.1649.004.
杨茜,王安志,吴锦涛,等.基于深度学习的二分图像分割综述[J/OL].计算机工程与应用, 2025-09-14. https://link.cnki.net/urlid/11.2127.TP.20250818.1649.004.
- [15] KANG X X, NIE Y. The construction and application of dual-teacher teaching model for enhancing students' deep learning capabilities: from perspective of ICAP theory [J]. China Educational Technology, 2025(9): 126-135.
康欣欣,乜勇.面向学生深度学习能力提升的双师教学模式构建及应用研究——基于ICAP理论的视角[J].中国电化教育, 2025(9): 126-135.
- [16] GONG L N, ZHOU Y R, QIAO Y, et al. Research progress on the application of pre-trained models in the field of software engineering [J]. Journal of Software, 2025, 36(1): 1-26.
宫丽娜,周易人,乔羽,等.预训练模型在软件工程领域应用研究进展[J].软件学报, 2025, 36(1): 1-26.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceeding of the 31st International Conference on Neural Information Processing Systems, 2017: 1-11.
- [18] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bi-directional transformers for language understanding [DB/OL]. https://arxiv.org/abs/1810.04805.
- [19] HU B, LU Z, LI H, et al. Convolutional neural network architectures for matching natural language sentences [DB/OL]. https://arxiv.org/abs/1503.03244.
- [20] SEPP H, JÜRGEN S. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [21] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [22] SUN C, QIU X, XU Y, et al. How to fine-tune bert for text classification? [C]//China National Conference on Chinese Computational Linguistics, 2019: 194-206.
- [23] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014: 1746-1751.

(责任编辑:陈维捷)