

基于融合评价指标 BERT-RGCN 的油田评价区块调整措施推荐方法

王梅¹, 朱晓丽¹, 孙洪国², 王海艳², 濮御³

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 大庆油田勘探开发研究院 开发规划研究室, 黑龙江 大庆 163712; 3. 东北石油大学 提高油气采收率教育部重点实验室, 黑龙江 大庆 163318)

摘要:为解决油田领域区块调整措施推荐过程中存在的样本数据稀疏和语义特征复杂等问题,提出基于融合评价指标(EI)的变换器双向编码(BERT)与关系图卷积神经网络(RGCN)的油田评价区块调整措施推荐方法(EI-BERT-RGCN方法)。根据评价指标、评价区块及措施之间的交互信息构建异构图,利用BERT模型生成评价指标、评价区块及措施术语词向量,共同作为输入词向量,将融合评价指标信息的异构图和输入词向量放入RGCN模型训练,学习评价区块的有效表征;在某油田评价区块提供的数据集上进行实验对比。结果表明:EI-BERT-RGCN方法能够捕捉文本中隐含的复杂语义并缓解数据稀疏问题,能更好理解未观察到的评价指标与调整措施之间的潜在关系,提升节点的表示质量。EI-BERT-RGCN模型在精确率、召回率、 F_1 分数及ROC曲线下面积等评价指标上优于其他基准模型,在保持较高精确率的同时,展现更好的泛化能力和鲁棒性。该结果为油田评价区块调整措施推荐提供参考。

关键词:异构图;变换器双向编码(BERT);预训练模型;关系图卷积神经网络(RGCN);推荐算法;措施推荐;油田评价区块

中图分类号:TP391.1 文献标识码:A 文章编号:2095-4107(2025)05-0110-11

0 引言

油田产业是国家经济的重要支柱,随油田开发进入中后期阶段,地层原油储量持续减少,采油难度日益加剧^[1-2]。在社会对石油能源需求持续增长的背景下,需要实时分析油藏单元的开发情况并采取有针对性的提升措施^[3]。以每个评价区块的监测数据为依据,对储量资源、开发技术、资产运行和油田开发效益等评价指标组合分析,为评价区块提供压裂、补孔和转注等措施,对挖掘油井潜力非常关键^[4-7]。

目前,制定常规的措施推荐方案主要依赖于专家经验和传统的推荐方法。为解决储层非均质性强、难以有效动用等问题,刘丹阳^[8]分析松辽盆地东南隆起区A油田历年转向压裂效果,考虑地层物性参数、地应力状态、岩石力学性质和经济效益评价指标,采取压裂措施,提升A油田低产区块的开发效果;为提高PQ油藏开发效果,黄易泽等^[9]考虑地层系数、孔隙度等评价指标,分析补孔潜力层位。油藏内部影响因素繁杂多变,制定措施方案需要大量的时间和成本,利用人工智能技术为评价区块高效推荐开发措施具有重要的现实意义^[10-11]。传统的推荐方法包括基于内容的推荐^[12]、基于协同过滤的推荐^[13]及混合推荐^[14]等,主要关注评价区块与措施的显式交互信息。为提升推荐结果的个性化程度,研究者将深度学习引入推荐任务,将深度神经网络^[15]、循环神经网络^[16]、卷积神经网络^[17]和图神经网络^[18]作为基础模型的改进推荐算法,用于挖掘记录非线性的潜在交互信息,从而提升推荐系统的准确性。随自然语言处理技术的飞速发展,文本信息可以作为重要的输入特征,提高推荐系统的性能。预训练词向量表示方法ELMo利用双向LSTM提取特征^[19],但受限于计算复杂性及特征融合方式。基于Transformer的解码器,生成式预训练变换模型GPT提高长距离文本信息的捕捉能力并加速训练^[20],但单向结构限制上下文信息的全面利用。变换器双向编码BERT引入位置编码和双向上下文结构,获取上下文相关的信息,获得更丰富的特

收稿日期:2024-12-24;编辑:蔡田田

基金项目:国家自然科学基金项目(52274037);黑龙江省科技创新基地项目(JD24A009);黑龙江省自然科学基金项目(LH2024F005);黑龙江省博士后科研启动金资助项目(LBH-Q20080)

作者简介:王梅(1975—),女,博士,教授,主要从事机器学习、核方法和模型选择等方面的研究。

通信作者:孙洪国,E-mail:sunhongguo@petrochina.com.cn

征表示^[21]。

近年来,人工智能技术在油气行业的应用日益广泛,围绕措施推荐领域展开相关探索^[22]。王晓丽等^[23]提出基于领域本体的油气勘探开发知识获取模式,利用改进的索引树词典构造和分词方法提取已登录命名实体,根据规则特征和聚类统计分析处理未登录命名实体。基于最近邻算法模式匹配模型,高丽洁等^[24]实现措施匹配。李权等^[25]应用知识图谱的语义分析及路径推理算法,推荐特定情况下的优化措施。在处理油田开发问题时,涉及大量描述地质特征、开发特征等信息的文本数据,包含丰富的专业术语和复杂的上下文关联。BERT 可以精确区分相似表述但本质不同的语义特征,将文本信息转化为富含语义的向量表示,为处理油田开发问题提供高质量的特征输入。此外,随油藏含水率持续上升,区块、储层、措施等实体类型间的关系日益复杂,RGCN^[26]可以捕捉各实体类型间的隐式依赖关系。

有效融合文本深层语义及多维度的评价指标,可以从样本数据获取精确的措施推荐模型。笔者提出基于融合评价指标的 BERT-RGCN 措施推荐方法(EI-BERT-RGCN 方法),根据历史评价信息构建评价信息异构图;利用 BERT 预训练模型和多头注意力机制,获取每个文本的稠密分布式嵌入表示向量,学习上下文依赖的深层语义特征;将高阶语义向量嵌入 RGCN,在缓解数据稀疏的场景下,融合评价指标信息并捕捉评价区块与措施间的深层偏好交互,为油田评价区块调整措施推荐提供参考。

1 EI-BERT-RGCN 模型框架

EI-BERT-RGCN 模型的总体框架主要分为异构图构建、文本节点特征初始化及措施推荐算法模型构建 3 部分。首先,根据以知识图谱形式存储的结构化历史推荐数据,构建“评价指标—评价区块—措施”异构图;然后,利用 BERT 模型提取评价指标、评价区块和措施文本的语义特征,融合上下文信息,作为 RGCN 模型的节点特征嵌入;最后,利用 RGCN 捕捉异构图中隐含的异常评价指标对措施的间接影响,根据预测结果制定推荐措施方案。EI-BERT-RGCN 模型实现框架见图 1。

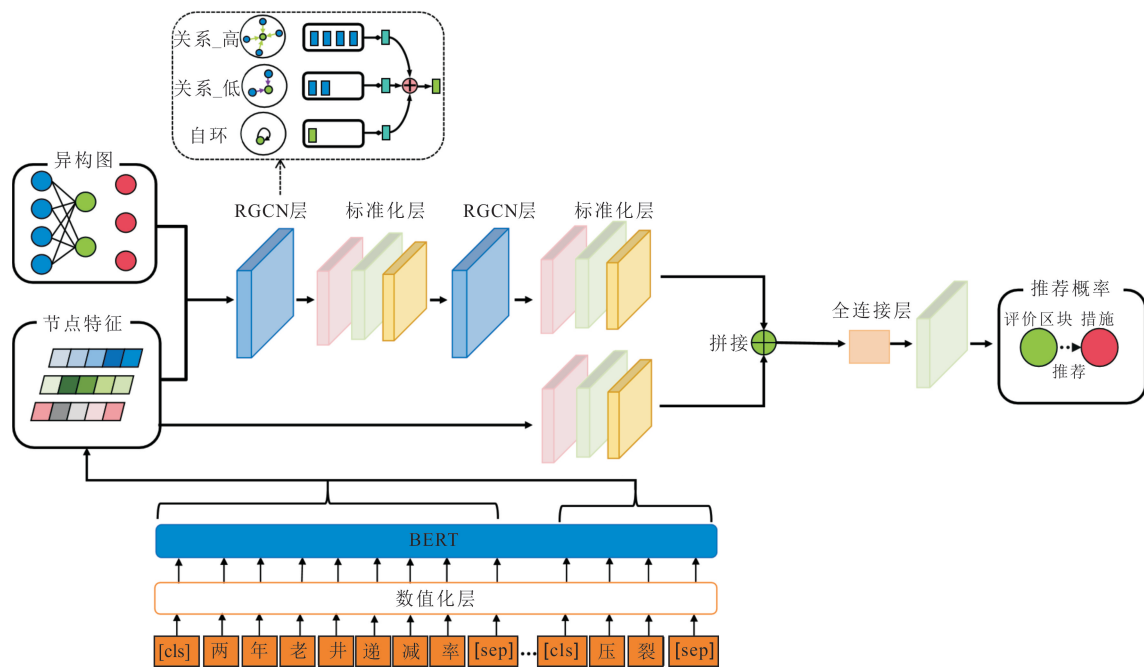


图 1 EI-BERT-RGCN 模型实现框架
Fig. 1 EI-BERT-RGCN model implementation framework

1.1 数据预处理

对 EI-BERT-RGCN 模型中出现的数据进行处理,设评价指标 $Z = \{z_1, z_2, \dots, z_m\}$ 、评价区块 $Q = \{q_1, q_2, \dots, q_n\}$ 、措施 $C = \{c_1, c_2, \dots, c_s\}$, $|Z| = m, |Q| = n, |C| = s$ 。由专家给出某油田不同含水阶段下每种评价指标的绝对划分空间,确定当前开发阶段每个评价指标的合理取值范围,取值边界为阈值,分为最大阈

值边界和最小阈值边界。选取某油田 15 种评价指标,构建开发评价指标体系(见表 1)。

收集历史评价指标、评价区块、评价日期和措施数据,作为历史评价数据进行数据预处理。首先,统一历史评价数据中评价日期的格式,删除重复的历史评价数据。其次,对缺失数据进行处理,运用随机森林回归模型补充,将无缺失数据的评价区块作为训练集,将其他的评价区块作为测试集,先填充缺失数据少的评价指标列,再用补充的评价指标列辅助补充缺失数据多的评价指标列。然后,采用箱线图法识别异常数据,将评价指标从小到大排列并划分为 4 个部分,3 个分割点 Γ_{SEG1} 、 Γ_{SEG2} 、 Γ_{SEG3} 为四分位数, Γ_{IQR} 为四分位数间距。箱线图的下界为 $\Gamma_{SEG1} - (1.5\Gamma_{IQR})$, 上界为 $\Gamma_{SEG3} + (1.5\Gamma_{IQR})$, 将超过箱线图上下界的数据视为异常数据,用对应的边界值补充。推荐措施为转注、气驱、微生物驱、化学驱、加密、转注、补孔、压裂、酸化及单砂体注水 10 种类型。最后,构建“评价区块—措施”交互矩阵 O' 记录历史交互标签,当评价区块 Z_c 与措施 C_c 产生交互行为时, o'_{cc} 为 1; 否则,为 0。

表 1 评价指标体系

Table 1 Evaluation index system

类别	评价指标
储量资源类	采出程度
	地层能量保持程度
开发技术类	水驱控制程度
	水驱动用程度
	渗透率
	分注率
	含水率
	油水井数比
资产运行类	采油速度
	油水井综合生产时率
	长关井比例
油田开发效益类	吨油操作成本
	两年老井递减率
	含水上升率
	注采比

1.2 评价信息异构图构建

评价指标—评价区块知识图谱(EI-EBKG)由评价指标和评价区块两种类型的节点构成,多个评价指标影响一个评价区块。在措施推荐算法中,根据制定的措施推荐规则判定当前评价区块的开发阶段,对每个开发阶段下的每个评价指标制定不同评价指标的合理取值范围;根据当前阶段对应的合理取值范围,确定每个评价指标与评价区块之间的关系。对于一个评价区块,某一个评价指标不在阈值范围内时,若评价指标高于最大阈值边界,则评价指标与评价区块之间的关系为高;若评价指标低于最小阈值边界,则评价指标与评价区块之间的关系为低。评价区块—措施知识图谱(EB-MKG)由评价区块和措施两种类型的节点构成,对一个评价区块推荐多种措施。当对一个评价区块采取某种措施时,在评价区块与措施之间设定“推荐”标签。根据不同关系将 EI-EBKG 转换三元组的形式,提取 EB-MKG 中包含的所有措施,构建“评价指标—评价区块—措施”异构图(见图 2)。

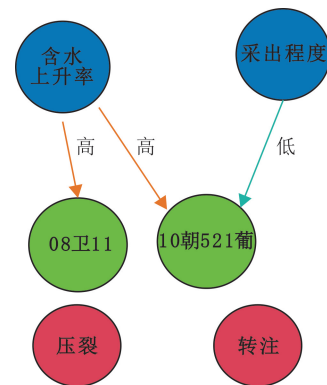


图 2 评价信息异构图
Fig. 2 Heterogeneous graphs for evaluation information

1.3 文本节点特征初始化

“评价指标—评价区块—措施”异构图有 15 个评价指标文本、780 个评价区块文本和 10 个措施文本。首先,使用 BERT tokenizer 分词器分别将 805 条节点文本信息转换为模型可以接受的输入格式;然后,按照转化的输入格式传递给 BERT 模型,得到 768 维的嵌入表示,如其中一个节点文本的嵌入向量为 $\{v_1, v_2, \dots, v_{768}\}$, 作为下一步推荐模型的输入。

1.4 措施推荐算法模型

评价信息异构图有两种类型的边关系,采用 RGCN 层处理输入特征。标准化层包括层归一化(LN)、ReLU 激活函数及 Dropout 正则化,将输入特征与卷积特征进行标准化,提升模型训练效率并增强模型泛化能力。采用全连接层及 Sigmoid 激活函数,输出当前评价区块对措施是否推荐的预测概率。一个评价区块对措施的预测概率为 $\{p_{rec1}, p_{rec2}\}$, 其中, p_{rec1} 为推荐概率, p_{rec2} 为不推荐概率。

2 措施推荐方法

基于 EI-EBKG 提取评价区块与评价指标之间所有的交互关系, 以及基于 EB-MKG 提取其中包含的所有措施, 构建融合评价指标信息的“评价指标—评价区块—措施”异构图。其中, 评价区块节点与异常评价指标节点之间存在高和低的关系连接, 在一个统一的异构图中组织所有节点及其交互关系, 以显示的方式捕捉所有的成对关系。

2.1 BERT 模型

BERT 模型是一种基于 Transformer 的自然语言处理模型, 结构见图 3。在大规模语料库上完成预训练后, BERT 通过微调获得文本语义特征。Transformer 由 Encoder 和 Decoder 两部分组成, BERT 只包含 Encoder 网络。

Trm 模型为 Transformer 编码器, 主要包括 Multi-Head Attention、Feed Forward 和 Add & Norm 层。Multi-Head Attention 层关注来自不同位置和不同表示子空间的信息; Feed Forward 层主要提供非线性变换; Add & Norm 层中的残差连接操作用于缓解梯度消失问题, 保持输入信息, 层归一化用于加速模型收敛并提高训练的稳定性。

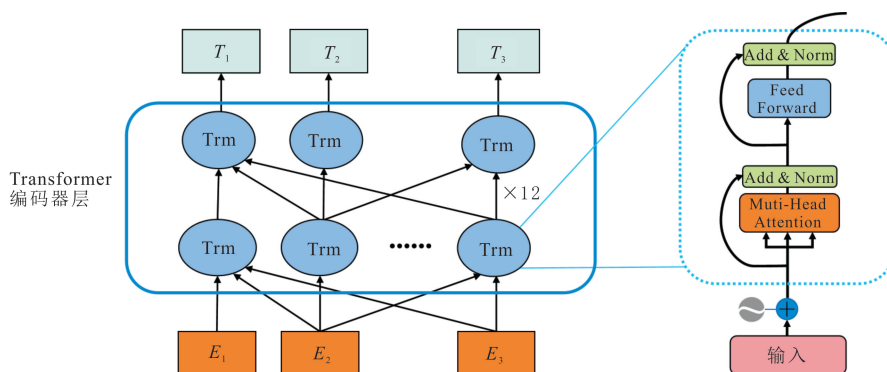


图 3 BERT 模型结构
Fig. 3 BERT model structure

应用 BERT 模型的编码器, 获取评价指标文本、评价区块文本和措施文本的词向量。通过使用词嵌入、位置嵌入和分段嵌入相加得到最终的输入嵌入 B_{BERT} 。评价指标文本矩阵 \mathbf{O}_Z 、评价区块文本矩阵 \mathbf{O}_Q 和措施文本矩阵 \mathbf{O}_C 分别为

$$\mathbf{O}_Z = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_i, \dots, \mathbf{Z}_m) = B_{\text{BERT}}(z_1, z_2, \dots, z_m), \quad (1)$$

$$\mathbf{O}_Q = (\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_j, \dots, \mathbf{Q}_n) = B_{\text{BERT}}(q_1, q_2, \dots, q_n), \quad (2)$$

$$\mathbf{O}_C = (\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_k, \dots, \mathbf{C}_s) = B_{\text{BERT}}(c_1, c_2, \dots, c_s), \quad (3)$$

式(1-3)中: $\mathbf{O}_Z \in \mathbb{R}^{m \times d}$, d 为向量维度大小, $d=768$; $\mathbf{O}_Q \in \mathbb{R}^{n \times d}$; $\mathbf{O}_C \in \mathbb{R}^{s \times d}$; $\mathbf{Z}_i, \mathbf{Q}_j, \mathbf{C}_k$ 分别为评价指标文本、评价区块文本和措施文本的第 i, j, k 个词向量。

将文本嵌入表示矩阵 $\mathbf{O}_t \in \mathbb{R}^{(m+n+s) \times d}$ 输入 Trm 模型的多头注意力层, 增强模型的表达能力。每个“头”专注于输入序列的不同方面, 对不同“头”的输出进行拼接, 多头注意力层的计算过程为

$$\boldsymbol{\psi}_{\text{head}, f} = \text{Softmax} \left(\frac{(\mathbf{W}_f^Q \mathbf{O}_t) (\mathbf{W}_f^K \mathbf{O}_t) (\mathbf{W}_f^V \mathbf{O}_t)}{\sqrt{d}} \right), \quad (4)$$

$$\mathbf{M}_{\text{multihead}} = \text{Concat}(\boldsymbol{\psi}_{\text{head}, 1}, \boldsymbol{\psi}_{\text{head}, 2}, \dots, \boldsymbol{\psi}_{\text{head}, f}) \mathbf{W}_0, \quad (5)$$

式(4-5)中: $\boldsymbol{\psi}_{\text{head}, f}$ 为第 f 个头的注意力; $\text{Softmax}(\cdot)$ 为激活函数; $\mathbf{W}_f^Q, \mathbf{W}_f^K, \mathbf{W}_f^V$ 为第 f 个头的自注意力权重矩阵; $\text{Concat}(\cdot)$ 为连接操作; $\mathbf{M}_{\text{multihead}}$ 为多头注意力; \mathbf{W}_0 为线性变换矩阵。

2.2 RGCN 模型

考虑节点间有多种不同的边关系类型, 采用 RGCN 模型, 对不同边关系类型的邻居节点设置不同的权重参数。在聚合过程中, 当前节点先对同一边关系类型的邻居节点聚合, 然后根据不同边关系类型的聚

合结果进行总体聚合,最终获得各个节点的潜在向量表示^[27]。

RGCN 使用不同的权重矩阵对不同的边关系类型进行加权求和,与节点 u 存在边关系类型为 e 的直接邻居节点 v 的特征向量 $\mathbf{t}_{v \rightarrow u, e}^{(l)}$ 为

$$\mathbf{t}_{v \rightarrow u, e}^{(l)} = \frac{1}{\sqrt{N(u)} \sqrt{N(v)}} \mathbf{h}_v^{(l)} \mathbf{W}_e^{(l)}, \quad (6)$$

式中: $\sqrt{N(u)} \sqrt{N(v)}$ 为归一化常数; $\mathbf{h}_v^{(l)}$ 为节点 v 在第 l 层的特征向量,采用 BERT 初始化节点的第一层特征向量为 $\mathbf{h}_v^{(0)}$; $\mathbf{W}_e^{(l)}$ 为第 l 层的参数矩阵。

节点 u 获取多个邻居节点的特征信息后,聚合边关系类型相同的邻居节点特征信息并更新自身节点表示。根据不同边关系类型的多个聚合向量,更新节点在第 $l+1$ 层的特征向量 $\mathbf{h}_u^{(l+1)}$ 为

$$\mathbf{h}_u^{(l+1)} = \text{ReLU} \left(\frac{1}{|E(u)|} \sum_{e \in N_e(u)} \mathbf{t}_{v \rightarrow u, e}^{(l)} \right), \quad (7)$$

式中: $\text{ReLU}(\cdot)$ 为激活函数; $E(u)$ 为节点 u 的边关系类型集合; $N_e(u)$ 为节点 u 的边关系类型为 e 的直接邻居节点的集合。

将式(6-7)用于措施推荐任务两种边关系节点信息的计算时,RGCN 网络参数迅速增多,易产生过拟合问题,引入基函数分解和块对角分解两种方法,减少模型计算参数的数量。基函数能够共享不同边关系类型的权重,分解式为

$$\mathbf{W}_r^{(l)} = \sum_{b=1}^B a_{rb}^{(l)} \mathbf{V}_b^{(l)}, \quad (8)$$

式中: $\mathbf{W}_r^{(l)}$ 为基函数, $\mathbf{W}_r^{(l)}$ 的基变换可以看作是参数 $a_{rb}^{(l)}$ 和 $\mathbf{V}_b^{(l)}$ 的线性组合, $a_{rb}^{(l)}$ 依赖于关系类型 r , $\mathbf{V}_b^{(l)}$ 为关系类型 r 在基底 b 上的基变换,是所有关系共享的参数矩阵。

块对角分解可以进行稀疏约束,对稀疏矩阵求和得到每个关系类型的权重矩阵为

$$\mathbf{W}_r^{(l)} = \sum_{b=1}^B \hat{\mathbf{Q}}_{br}^{(l)}, \quad (9)$$

式中: $\hat{\mathbf{Q}}_{br}^{(l)}$ 为低维稀疏矩阵。求和得到的对角矩阵作为不同关系类型下的稀疏约束。

为建立更加全面的决策边界,提高推荐任务的鲁棒性和准确性。将已经定义的评价区块与措施间的所有连接关系作为正例,同时替换三元组中的尾节点,将构建不存在的连接关系作为负例。在措施推荐任务中出现样本分布不均匀的情况,采用平衡因子 λ 增加少样本类别的权重,提高模型对少样本类别的识别性能,计算公式为

$$\lambda = \frac{\mu_{\text{neg}}}{\mu_{\text{pos}} + \epsilon}, \quad (10)$$

式中: μ_{pos} 、 μ_{neg} 分别为正、负样本的数量; ϵ 为一个趋于 0 的参数,防止在极端不平衡数据集下发生异常。

对异构图上的节点特征进行传递与更新,达到稳定状态的表示信息能够充分建立节点与其高阶邻居节点间的复杂关系,为推荐任务提供丰富的向量表征。在消息传递过程中,评价区块节点根据与之关联的评价指标节点的信息更新嵌入表示(见图 4)。评价区块的节点表示融合不同关系的评价指标信息。在消息传递和聚合操作中,评价指标节点利用评价区块节点中已聚合的信息更新自身嵌入表示,从而获得评价指标感知的评价区块节点表示。采用 RGCN 编码器,经过消息传递、聚合操作,获得各节点的潜在向量表示。评价指标的潜在向量表示为 \mathbf{h}_e ,评价区块的潜在向量表示为 \mathbf{h}_q ,措施的潜在向量表示为 \mathbf{h}_c 。

将 EI-EBKG 交互信息以三元组 $(E_{\text{head}}, R, E_{\text{tail}})$ 的形式表示。其中, E 为实体集合, R 为实体间的关系集合。与 EB-MKG 的所有措施共同建立“评价指标—评价区块—措施”异构图,使用 RGCN 将不同类型的节点映射到同一向量空间,每个评价区块的向量表示融合评价指标信息。为实现措施推荐任务,设计分类模型,实现评价区块对每种措施的推荐概率预测。

2.3 实现过程

激活函数 $\sigma(x)$ 的定义为

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (11)$$

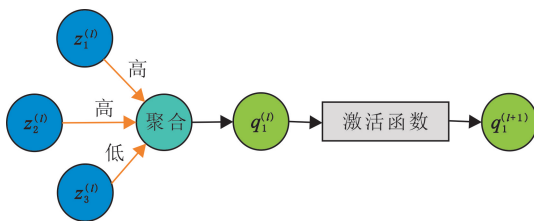


图 4 节点消息聚合

Fig. 4 Node message aggregation

对应的代价敏感加权交叉熵损失函数 $L^{[28]}$ 为

$$L = -\frac{1}{U} \sum_{u=1}^U (\lambda y^{(i)} \log \sigma(x^{(i)}) + (1 - y^{(i)}) \log(1 - \sigma(x^{(i)}))) + \gamma \|\theta\|^2, \quad (12)$$

式中: U 为样本数量; $x^{(i)}$ 为第 i 个训练样本的得分; $y^{(i)}$ 为第 i 个样本对应的标签, $y^{(i)} \in \{0, 1\}$; γ 为模型参数; θ 为正则化参数。

通过全连接层将提取的特征转换到进行最终决策的维度 2, 每一维度代表每个评价区块节点对每种措施推荐概率 (p_{rec1}) 与不推荐概率 (p_{rec2})。

3 实验与分析

3.1 数据集选取

选取某油田一年的 780 个评价区块的数据, 包括 15 种评价指标和 10 种措施。EI-EBKG 中有关系 $R_1 = \{r_1: \text{高}, r_2: \text{低}\}$, EB-MKG 中有关系 $R_2 = \{r_3: \text{推荐}\}$ 。其中, 评价指标与评价区块的交互记录有 10 060 条, 评价区块与措施之间的交互记录有 7 343 条。将 780 个评价区块与措施之间的 4 197 条交互记录作为正样本, 未发生交互记录的 3 146 条样本作为负样本。为解决样本不均衡问题, 设置平衡因子平衡正、负样本的权重。将评价区块与措施交互记录数据集的 80% 数据划分为训练集与验证集, 20% 数据划分为测试集。实验数据集统计见表 2。

表 2 实验数据集统计

Table 2 Statistics of experimental dataset

数据集	评价区块/个	评价指标/种	措施/种	关系	训练三元组记录/条	测试三元组记录/条
EI-EBKG	780	15	—	2	10 060	10 060
EB-MKG	780	—	10	1	5 874	1 469

3.2 参数设置

实验环境为 CPU: 12th Gen Intel(R) Core(TM) i5-12400F; 内存: 16 GB; GPU: NVIDIA GeForce GT 1030。使用 Pytorch 框架实现 EI-BERT-RGCN 模型。BERT 模型隐藏单元数即词向量维度为 768。RGCN 模型的实体嵌入维度为 768, 隐藏层维度为 256, Dropout 比率为 0.1, 使用 Adam 进行模型参数优化, 优化器的初始学习率设置为 0.01, 每经过 20 轮迭代训练, 学习率降低 50%。

3.3 指标评估

为对比算法性能, 选用的评估指标包括精确率 (P_{Pre})、召回率 (R_{Recall})、 F_1 分数、ROC 曲线下面积 (A_{AUC}), 计算公式分别为

$$P_{Pre} = \frac{T_P}{T_P + F_P}, \quad (13)$$

$$R_{Recall} = \frac{T_P}{T_P + F_N}, \quad (14)$$

$$F_1 = \frac{2P_{\text{Pre}}R_{\text{Recall}}}{P_{\text{Pre}} + R_{\text{Recall}}} = \frac{2T_P}{2T_P + F_P + F_N}, \quad (15)$$

$$A_{\text{AUC}} = \frac{\sum g_{\text{pos}} > g_{\text{neg}}}{N_{\text{pos}}N_{\text{neg}}}, \quad (16)$$

式(13-16)中: T_P 为真阳性, 表示评价区块被正确推荐措施的实例; F_P 为假阳性, 表示评价区块被错误推荐措施的实例; F_N 为假阴性, 表示错误预测评价区块不需要推荐当前措施的实例; g_{pos} 、 g_{neg} 分别为模型对正样本和负样本的预测值; N_{pos} 、 N_{neg} 分别为正样本和负样本的数量。

为衡量推荐算法的增益, 引入相对增益 R_{Rela} ^[29]。对于一个随机高斯分类器, $A_{\text{AUC}} = 0.5$, R_{Rela} 为

$$R_{\text{Rela}} = \left(\frac{A_{\text{AUC, OurModel}} - 0.5}{A_{\text{AUC, OtherModel}} - 0.5} - 1 \right) \times 100\%, \quad (17)$$

式中: $A_{\text{AUC, OurModel}}$ 为文中推荐算法的 ROC 曲线下面积; $A_{\text{AUC, OtherModel}}$ 为其他推荐算法的 ROC 曲线下面积。

选用参数量 (P_{Params}) 及浮点运算次数 (F_{FLOPs}) 两种评估指标评估模型复杂度。 P_{Params} 为模型中可训练的参数总数, 包括权重和偏置等; F_{FLOPs} 为衡量模型的计算复杂度, F_{FLOPs} 越小, 模型的推理速度越快。

3.4 实验结果分析

3.4.1 方法对比

选取主流模型框架作为基准模型, 在 EI-BERT-RGCN 模型构建的数据集上进行对比。选取的基准模型包括: MF-GCN 模型^[29], 将具有不同边缘类型的子图分别输入图卷积层, 学习每种节点的特征表示; ERNIE-RCNN 模型^[30], 将外部知识嵌入语义表示, 结合局部依赖和全局语义建模; ERNIE-TEXTCNN 模型^[31], 捕捉文本中的上下文特征和局部关键特征; TM-BERT 模型^[32], 使用文本挖掘和 BERT 模型进行特征表示; BERTGCN 模型^[33], 使用 BERT 模型捕捉上下文信息, 通过 GCN 邻域聚合技术学习实体节点表示。

6 种方法的推荐效果见表 3。由表 3 可知, MF-GCN 模型考虑异构图中不同边类型的结构信息, 融合高、低关系子图和节点属性, 模型较为简单, 但学习的上下文节点稀疏, 整体性能较差。ERNIE-RCNN 模型可以学习复杂语义, 但主要用于对长文本依赖建模。对于短文本的节点类型, ERNIE-TEXTCNN 模型局部语义捕捉能力较强, 综合性能整体比 ERNIE-RCNN 模型的高。TM-BERT 模型基于 BERT 模型考虑实体节点的深层上下文语义, 但未有效学习评价指标与措施之间的高阶隐式关系。BERTGCN 模型结合文本深层语义信息及图拓扑信息, 推荐性能较好, 但没有考虑两种关系的不同权重。EI-BERT-RGCN 模型的 P_{Pre} 为 94.60%, R_{Recall} 为 94.62%, F_1 为 94.61%, A_{AUC} 为 98.23%, 相较于其他基准模型的最好结果, 分别提高 0.21%、0.20%、0.23%、0.20%。EI-BERT-RGCN 模型优于其他基准模型, 主要原因为 BERT 模型能够更快适应新的异构网络场景, 更好理解实体节点在整个文本序列中的位置和关系, 提取更丰富的词向量表征。此外, RGCN 模型考虑不同关系对节点的影响, 有效挖掘异常评价指标组合对推荐措施的影响, 从而提升推荐性能。

表 3 6 种方法推荐效果

Table 3 Recommended effect of six methods

模型	$P_{\text{Pre}}/\%$	$R_{\text{Recall}}/\%$	$F_1/\%$	$A_{\text{AUC}}/\%$	$R_{\text{Rela}}/\%$	$P_{\text{Params}}/10^6$	$F_{\text{FLOPs}}/10^9$
MF-GCN	93.34	93.30	93.32	97.25	2.07	0.27	0.11
ERNIE-RCNN	93.71	93.73	93.72	96.67	3.34	103.06	1.71
ERNIE-TEXTCNN	93.61	94.37	93.99	97.72	1.06	103.19	1.72
TM-BERT	94.25	94.42	94.33	97.56	1.41	101.68	1.70
BERTGCN	94.39	94.37	94.38	98.03	0.42	101.94	2.13
EI-BERT-RGCN	94.60	94.62	94.61	98.23	—	102.73	2.55

3.4.2 消融实验对比

为验证 BERT、RGCN 和 EI-EBKG 对模型的影响, 对 EI-BERT-RGCN 模型进行消融实验。其中,

BERT-RGCN 模型不考虑评价指标信息, EI-BERT 模型未融合图结构信息, EI-RGCN 模型不使用 BERT 提供的包含上下文信息的词嵌入, 采用独热编码作为 RGCN 模型的嵌入。不同消融实验推荐效果见表 4。由表 4 可知, 在需要精准推荐措施的场景下, EI-BERT-RGCN 模型对措施推荐任务有增益作用。不包含语义增强模块的 EI-RGCN 模型计算复杂度低, 但推荐效果受限于图结构信息。未融合评价指标与评价区块之间关系的 EI-BERT 模型计算复杂度低于其他需要预训练的模型的, 但未考虑图结构信息, 性能较低。BERT-RGCN 模型在表示学习时只依赖措施记录, 节点嵌入表示同质化, 推荐效果明显低于 EI-BERT-RGCN 模型的, 表明在进行措施推荐的同时还要考虑评价指标信息。EI-BERT-RGCN 模型能够捕捉文本中隐含的复杂语义并缓解数据稀疏问题, 在计算复杂度没有显著增加的情况下, 能更好理解未观察到的评价指标与措施之间的潜在关系, 提升节点的代表质量, 推荐效果最佳。

表 4 不同消融实验推荐效果

Table 4 Recommended effect of different ablation experimentals

模型	$P_{Pre}/\%$	$R_{Recall}/\%$	$F_1/\%$	$A_{AUC}/\%$	$R_{Rela}/\%$	$P_{Params}/10^6$	$P_{FLOPs}/10^9$
EI-RGCN	93.51	93.12	93.31	97.18	2.23	1.05	0.85
EI-BERT	94.32	94.36	94.34	97.79	0.92	101.68	1.70
BERT-RGCN	91.79	91.49	91.64	96.80	3.06	101.94	2.13
EI-BERT-RGCN	94.60	94.60	94.61	98.23	—	102.73	2.55

3.4.3 敏感性分析

为分析重要的超参数对 EI-BERT-RGCN 模型的影响, 分析 BERT 配置参数、RGCN 层数 R_L 、RGCN 隐藏层维度 R_H 及平衡因子 λ , 采用 F_1 、 A_{AUC} 作为性能评估的标准指标。选用 BERT-base 和 BERT-large 两种标准的中文预训练语言模型^[21] 及 BERT-wwm^[34]。设置 B_L 为编码器层数, B_A 为编码器的注意力头数, 每个注意力头的隐藏单元数为 64 个, B_H 为编码器的全连接网络包含的隐藏单元维度, 表示从模型中得到的词向量维度。BERT-base 和 BERT-wwm 配置参数: $B_L=12$, $B_A=12$, $B_H=768$ 。BERT-large 配置参数: $B_L=24$, $B_A=16$, $B_H=1024$ 。

不同 BERT 模型配置推荐效果见图 5。选用 BERT-base 配置能够捕捉上下文语义信息, 但对文本字粒度的掩蔽方式破坏词语的完整性, 推荐性能指标略低。BERT-wwm 配置采用全词遮蔽, 学习整体的评价信息语义表示, 在模型参数保持一致的情况下, 两项指标最高, 能够高效区分推荐与未推荐措施。选用 BERT-large 配置模型的指标低于前二者的且参数更多, 原因是训练数据规模较小且导致模型过拟合。

RGCN 层数 R_L 是 EI-BERT-RGCN 模型重要的超参数, 设置 R_L 为 1、2、3、4, R_L 对模型性能的影响见图 6。由图 6 可知, 当 R_L 增大到 2 时, 模型能够在计算资源和训练时间更优的情况下学习丰富的高阶嵌入; 当 R_L 继续增大时, 模型对节点关键特征的捕捉能力下降, 措施推荐性能下降。

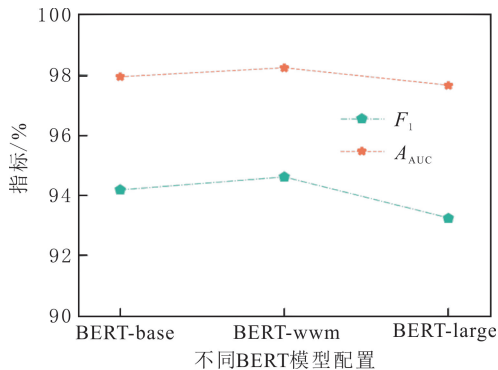
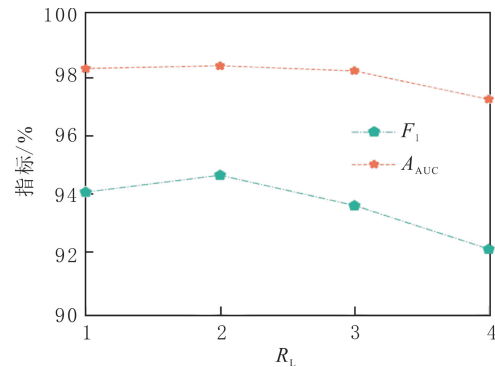


图 5 不同 BERT 模型配置推荐效果

Fig. 5 Recommended effect of different BERT configurations

图 6 R_L 对模型性能的影响Fig. 6 Impact of R_L on model performance

设置 RGCN 隐藏层维度 R_H 为 128、256、384、512, R_H 对模型性能的影响见图 7。由图 7 可知, R_H 为 256 时, 两个指标值最高, 可以挖掘高阶评价区块重要特征, 性能达到最佳; 当 R_H 继续增大时, 模型参数增多导致过拟合, 泛化能力下降。

为评估正、负样本平衡对实验结果的影响, 引入平衡因子 λ 优化模型。 R_{ROC} 曲线和 A_{AUC} 评估 λ 对模型性能的影响见图 8。由图 8 可知, 原始模型的 A_{AUC} 为 96.88%; 加入平衡因子 λ 为正样本, 设置较高的样本权重并重新训练模型, A_{AUC} 提升到 98.23%。这表明使用平衡因子使模型在识别少数样本类和分类整体样本的性能上增强, 具有更强的鲁棒性, 能够很好区分正、负样本, 模型在对评价区块推荐措施的过程中更可靠、精准。

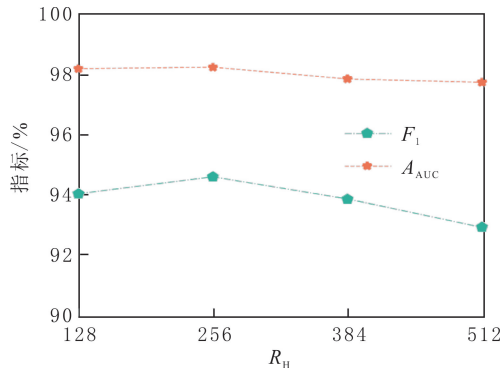


图 7 R_H 对模型性能的影响

Fig. 7 Impact of R_H on model performance

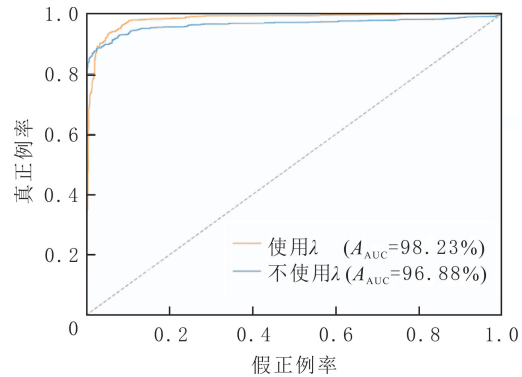


图 8 λ 对模型性能的影响

Fig. 8 Impact of λ on model performance

3.4.4 一致性分析

为验证 EI-BERT-RGCN 模型推荐结果的合理性, 进行各评价区块的推荐结果与油藏机理一致性分析(见表 6)。由表 6 可知, 对于高含水区块, 模型优先推荐单砂体注水, 辅以压裂、补孔等措施, 强化油水波及并改善注采连通性; 对于中含水区块, 模型倾向于推荐单砂体注水、压裂、补孔、转注及加密等措施, 改善流体流动通道, 优化采油效率; 对于渗透率、采出程度和采油速度低的区块, 模型推荐转注方式调整注采策略, 以激活低产区, 提高剩余油动用程度。这与水驱开发中“注采平衡、提高驱替效率”的油藏机理相符合, 与油藏物理规律保持较好一致性。整体上, 模型在不同评价指标的条件下体现较强的针对性与科学性。

表 6 模型推荐结果与油藏机理一致性分析

Table 6 Analysis of the consistency of model recommendation results with reservoir mechanism

评价区块	渗透率/ $10^{-3} \mu\text{m}^2$	含水率/ %	采出程度/ %	采油速度/ %	主要推荐措施	机理一致性
区块 1	66.20	81.74	14.96	1.80	单砂体注水、压裂、补孔	高含水阶段, 采油速度高, 主要推荐单砂体注水, 以强化波及效率, 辅以压裂、补孔措施改善注采连通性
区块 2	2.10	63.20	8.87	0.31	单砂体注水、压裂、转注	含水中等偏高, 采油速度低, 主要推荐单砂体注水, 以强化波及效率, 辅以压裂、转注, 以提高驱动能量
区块 3	40.20	87.43	14.02	0.43	单砂体注水、压裂	高含水阶段, 采出程度低, 主要推荐单砂体注水, 以强化波及效率, 辅以压裂改善注采连通性
区块 4	208.70	44.69	8.88	0.18	单砂体注水、压裂、补孔、转注、加密	渗透率高, 含水中等, 采油速度低, 主要推荐单砂体注水, 以增强波及, 辅以压裂、补孔和转注改善注采连通性, 采取加密措施优化采油效率
区块 5	2.50	63.09	4.12	0.02	转注方式	渗透率、采出程度和采油速度低, 主要推荐转注方式调整注采策略, 以激活低产区, 提高剩余油动用程度

4 结论

(1)基于BERT学习每个节点文本信息的稠密分布式嵌入向量,应用RGCN融合评价指标信息,提出基于融合评价指标的EI-BERT-RGCN措施推荐方法,处理具有高、低多种边关系类型图并完成措施推荐。在措施推荐任务中,EI-BERT-RGCN方法能有效捕捉节点文本丰富上下文信息,挖掘异常评价指标与措施间的潜在关系,缓解措施推荐任务中样本数据稀疏和语义特征复杂的情况。

(2)在某油田提供的真实数据集上进行对比实验,相较于其他基准模型,EI-BERT-RGCN模型在保持较高精确率的同时,展现更好的泛化能力和鲁棒性,适用于评价区块调整措施推荐任务。

(3)对于EI-BERT-RGCN模型,可以通过模型压缩、量化等方法,降低模型计算复杂度,提升推理速度;考虑时间特征,学习推荐措施随评价区块中评价指标变化的动态调整规律,为评价区块调整措施推荐提供参考。

参考文献(References):

- [1] 周志军,张国青,崔春雪,等. 页岩储层孔隙结构表征及物性下限确定方法及应用[J]. 特种油气藏,2024,31(4):96-102.
ZHOU Zhijun, ZHANG Guoqing, CUI Chunxue, et al. Methods and applications for characterizing pore structure and determining physical property lower limit in shale reservoirs[J]. Special Oil & Gas Reservoirs, 2024,31(4):96-102.
- [2] 王双龙,岳宝林,祝晓林,等. 渤海锦州南变质岩潜山油田储层分类及其剩余油分布规律[J]. 东北石油大学学报,2024,48(4):89-99.
WANG Shuanglong, YUE Baolin, ZHU Xiaolin, et al. Reservoir classification and remaining oil distribution patterns of Jinzhou south metamorphic buried mountain oilfield in Bohai Sea[J]. Journal of Northeast Petroleum University, 2024,48(4):89-99.
- [3] 王洪亮,李宁,李欣,等. 基于数据驱动的致密油藏产能预测方法[J]. 特种油气藏,2025,32(2):82-88.
WANG Hongliang, LI Ning, LI Xin, et al. Productivity prediction method of tight reservoir based on data-driven approach[J]. Special Oil & Gas Reservoirs, 2025,32(2):82-88.
- [4] 张凤远,邹林君,崔维,等. 基于压裂液返排数据的页岩油气藏裂缝参数反演方法[J]. 东北石油大学学报,2022,46(1):76-87.
ZHANG Fengyuan, ZOU Linjun, CUI Wei, et al. Inverse method based on fracturing fluid flowback data for estimating fracture properties in shale reservoirs[J]. Journal of Northeast Petroleum University, 2022,46(1):76-87.
- [5] 彭神奇. 大牛地气田补孔措施效益后评价及效益界限产量研究[J]. 中外能源,2020,25(6):53-56.
PENG Shenqi. Benefit evaluation and benefit limit production research of reperforating measures in Daniudi Gas Field[J]. Sino-Global Energy, 2020,25(6):53-56.
- [6] 苑登御,侯吉瑞,宋兆杰,等. 塔河油田缝洞型碳酸盐岩油藏注水方式优选及注气提高采收率实验[J]. 东北石油大学学报,2015,39(6):102-110.
YUAN Dengyu, HOU Jirui, SONG Zhaojie, et al. Optimization of water injection methods and N₂ flooding for EOR in Tahe fractured-vuggy carbonate reservoir[J]. Journal of Northeast Petroleum University, 2015,39(6):102-110.
- [7] 欧阳雨薇,龙明,张吉磊,等. 渤海河流相Q油田提液挖潜规律分析[J]. 石油化工高等学校学报,2023,36(1):32-39.
OUYANG Yuwei, LONG Ming, ZHANG Jilei, et al. Law of liquid production in increase in fluvial facies reservoir of Bohai Q Oilfield[J]. Journal of Petrochemical Universities, 2023,36(1):32-39.
- [8] 刘丹阳. A油田压裂井效果分析与研究[J]. 石化技术,2015,22(8):174.
LIU Danyang. Fracturing wells in A Oilfield[J]. Petrochemical Industry Technology, 2015,22(8):174.
- [9] 黄易泽,刘国栋,李洲,等. 安塞油田致密砂岩油藏补孔压裂措施效果分析[J]. 石油地质与工程,2024,38(2):88-92.
HUANG Yize, LIU Guodong, LI Zhou, et al. Analysis of perforation fracturing measures effect in tight sandstone reservoir of Ansai Oilfield[J]. Petroleum Geology and Engineering, 2024,38(2):88-92.
- [10] 高雅田,李春生,富宇. 基于知识解析的油田动态开发辅助决策系统研究[J]. 科学技术与工程,2010,10(11):2767-2771.
GAO Yatian, LI Chunsheng, FU Yu. Research on assistant decision system for oilfield dynamic development based on analysis of knowledge[J]. Science Technology and Engineering, 2010,10(11):2767-2771.
- [11] 孟军晓. 先进技术在石油开采率提升中的应用分析[J]. 中国石油和化工标准与质量,2025,45(17):146-148.
MENG Junxiao. Application analysis of advanced technologies in improving oil recovery rate[J]. China Petroleum and Chemical Standard and Quality, 2025,45(17):146-148.
- [12] MA S, ZHANG C, LIU X. A review of citation recommendation: from textual content to enriched context[J]. Scientometrics, 2020,122(3):1445-1472.
- [13] LU Y, HE Y, CAI Y, et al. Time-aware neural collaborative filtering with multi-dimensional features on academic paper recommen-

- dation[C]//International Conference on Computer Supported Cooperative Work in Design. Dalian, China; IEEE, 2021;1052-1057.
- [14] JEONG C, JANG S, PARK E, et al. A context-aware citation recommendation model with BERT and graph convolutional networks[J]. *Scientometrics*, 2020,124(1):1907-1922.
- [15] COVINGTON P, ADAMS J, SARGIN E. Deep neural networks for YouTube recommendations[C]//ACM Conference on Recommender Systems. Boston, MA, USA; ACM, 2016;191-198.
- [16] CHENG H T, KOC L, HARMSSEN J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, MA, USA; ACM, 2016;7-10.
- [17] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. *Nature*, 1986,323(6088):533-536.
- [18] LYU Z, WU Y, LAI J, et al. Knowledge enhanced graph neural networks for explainable recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022,35(5):4954-4968.
- [19] SARZYNSKA-WAWER J, WAWER A, PAWLAK A, et al. Detecting formal thought disorder by deep contextualized word representations[J]. *Psychiatry Research*, 2021,304:114135.
- [20] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. *Open AI Blog*, 2019,1(8):9.
- [21] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota; ACM, 2019;4171-4186.
- [22] 郭军辉,郑宪宝,王治国,等. 大庆长垣油田水驱开发技术智能化实践与展望[J]. *大庆石油地质与开发*, 2024,43(3):203-213.
GUO Junhui, ZHENG Xianbao, WANG Zhiguo, et al. Intelligent practice and prospects of water flooding development technology in Daqing Placanticline Oilfield[J]. *Petroleum Geology & Oilfield Development in Daqing*, 2024,43(3):203-213.
- [23] 王晓丽,邓达康,孟祥龙,等. 基于领域本体的油气勘探开发知识获取模式及实现[J]. *东北石油大学学报*, 2016,40(4):74-79.
WANG Xiaoli, DENG Dakang, MENG Xianglong, et al. Oil and gas exploration and production knowledge processing mode and implementation based on domain ontology[J]. *Journal of Northeast Petroleum University*, 2016,40(4):74-79.
- [24] 高丽洁. 基于模式匹配与最近邻算法的油井维护措施推荐[D]. 北京:中国石油大学(北京), 2018:28-34.
GAO Lijie. Measures recommendation of oil well maintenance based on pattern matching and nearest neighbor algorithm[D]. Beijing: China University of Petroleum(Beijing), 2018:28-34.
- [25] 李权,李金蔓,张艳英,等. 知识图谱在油井措施推荐方面的应用[J]. *石化技术*, 2022,29(3):124-125.
LI Quan, LI Jinman, ZHANG Yanying, et al. Application of knowledge atlas in oil well measure recommendation[J]. *Petrochemical Industry Technology*, 2022,29(3):124-125.
- [26] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//European Semantic Web Conference. Cham; Springer International Publishing, 2018;593-607.
- [27] 成全,余德昕. 融合患者体征与用药数据的图神经网络药物推荐方法研究[J]. *数据分析与知识发现*, 2022,6(9):113-124.
CHENG Quan, SHE Dexin. Drug recommendation based on graph neural network with patient signs and medication data[J]. *Data Analysis and Knowledge Discovery*, 2022,6(9):113-124.
- [28] HUANG C, LI Y, LOY C C, et al. Learning deep representation for imbalanced classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas; IEEE, 2016;5375-5384.
- [29] LU J, CHEN Z, DENG X L. A graph convolutional neural network model based on fused multi-subgraph as input and fused feature information as output[J]. *Engineering Applications of Artificial Intelligence*, 2025,139:109542.
- [30] BAO D, QIN D, HONG L, et al. Multi-channel text classification model based on ERNIE[C]//Proceedings of the 2022 11th International Conference on Computing and Pattern Recognition. Beijing, China; ACM, 2022;321-327.
- [31] WANG M, XU J. Research on Chinese short text classification based on ERNIE-TEXTCNN model[C]//Third International Conference on Advanced Algorithms and Neural Networks(AANN 2023). Qingdao, China; SPIE, 2023,12791:315-319.
- [32] MACEDO J B, MOURA M D C, AICHELE D, et al. Identification of risk features using text mining and BERT-based models: application to an oil refinery[J]. *Process Safety and Environmental Protection*, 2022,158:382-399.
- [33] LIU Q, XIAO K, QIAN Z. A hybrid re-fusion model for text classification[J]. *Scientific Reports*, 2025,15(1):9333.
- [34] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for Chinese BERT[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019,29:3504-3514.