南京農業大學
NANJING AGRICULTURAL UNIVERSITY

基于主题分类模型的在线学术社区答案质量评价研究

汇报人：林立涛

南京农业大学信息管理学院
南京农业大学人文与社会计算研究中心

目录

CONTENT

一 研究背景、目的、意义

1.研究背景、目的、意义

◆背景

- 学术社区中的非学科知识对于促进科研活动的有序开展具有积极意义。
- “问题-答案”对形式的用户生成内容（User Generated Content，UGC）是在线学术社区的重要组成。然而，在线学术社区中的用户生成答案的质量参差不齐，难以为用户提供高效的决策支持。

◆目的与意义

- 从每个帖子下的用户生成答案中筛选出优质答案，促进在线学术社区用户生成内容中的问答类型UGC的高效利用。

二 研究方法

2. 1.基于主题分类模型的答案质量评价方法的提出依据

◆网络社区知识抽取

- 问题视角：通过判别网络社区中的优质提问，总结归纳高质量问题的句法结构特征，进而实现对高质量问题及其关联内容的获取[1]。
- 答案视角：可以进一步分为面向单条答案的知识抽取和面向多条答案的知识抽取，即（1）基于多答案的文本摘要[2]；（2）单条答案排序与优选[3,4]。

◆用户生成内容质量评价

- 评价角度：从结构化特征、文本特征、用户社交属性、答案主题、用户权威度、文本属性、答题者影响力属性、答题者专业性、答案表层属性、社会情感特征、答案时效特征以及融合多特征的答案质量评价方法[3,5,6]。
- 评价方法：逻辑回归、SVM、随机森林、 BERT-CNN、 GA-BP神经网络模型等

1.基于主题分类模型的问答相关性计算方法的提出依据

◆过往研究的不足

- 用户社交属性、用户权威度、文本属性、答题者影响力属性、答题者专业性、社会情感特征等特征数据不易轻易获取。
- 没有充分利用问答帖中的问题描述文本。
- 机器学习能够有效对非文本特征建模，但是对语义特征的建模能力不足。

◆理论基础

- 高质量的答案与对应的问题应该属于同一个话题。
- 基于问答主题相似度的答案质量评价是有效的。
- 深度预训练语言模型能够实现高质量语义表征。

2.基于问答主题相似度的答案质量评价方法描述

◆核心特点

主题分类模型构建及应用过程

- （1）获得主题分类体系
- （2）利用有监督的问题描述文本构建主题分类模型
- （3）利用构建的主题分类模型计算问题和答案的主题相似度

2.基于问答主题相似度的答案质量评价方法描述

◆ 基于问答主题相似度的答案排序方法示意图

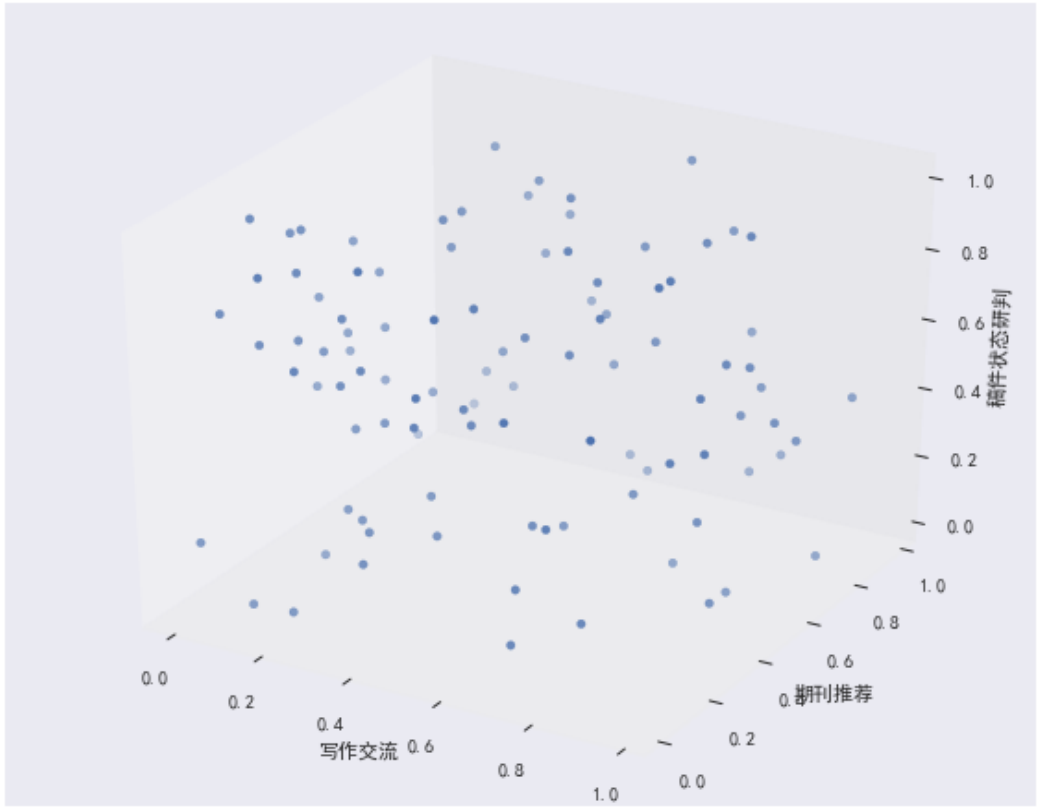


图5-1 三维向量空间中的文本主题向量

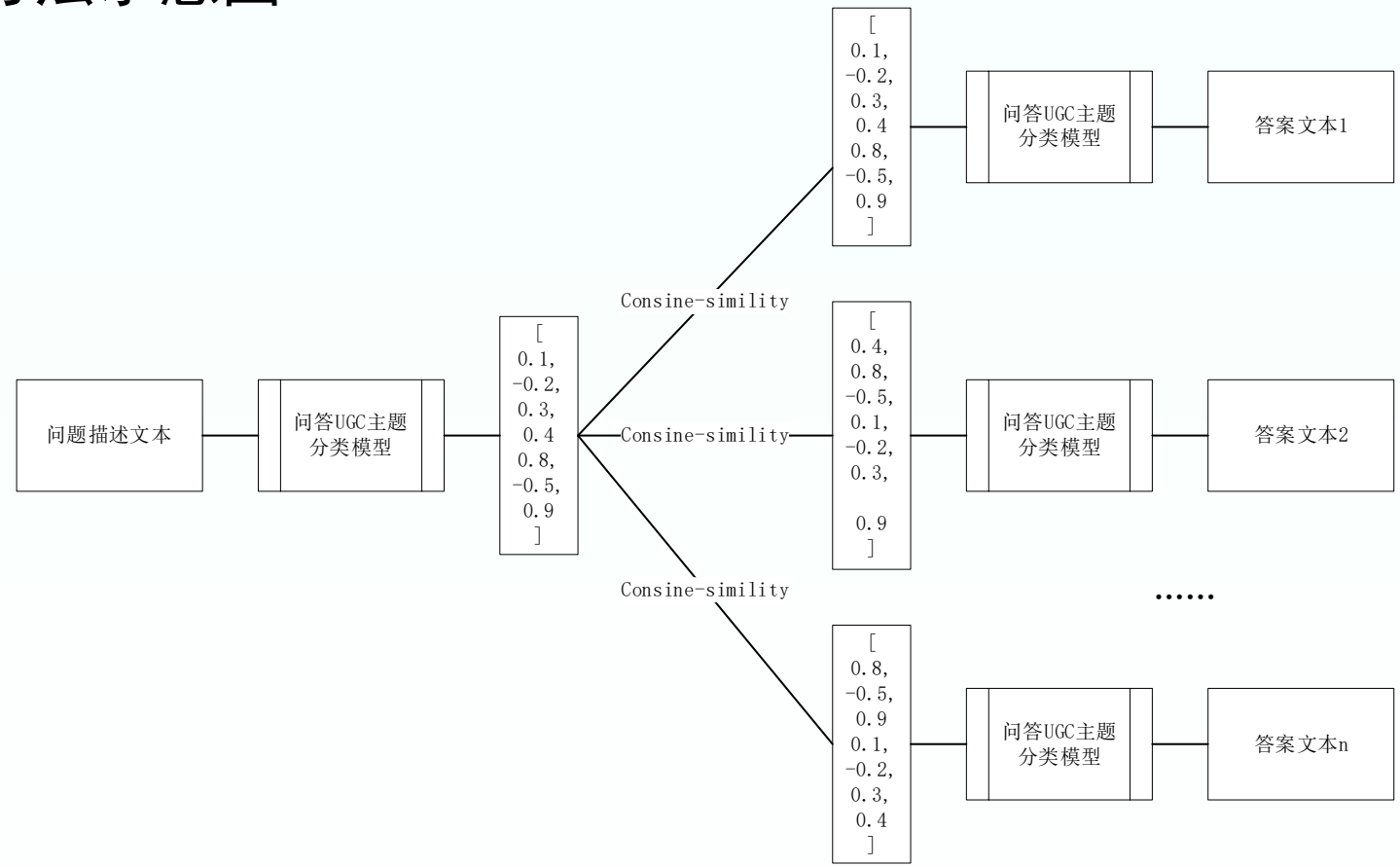
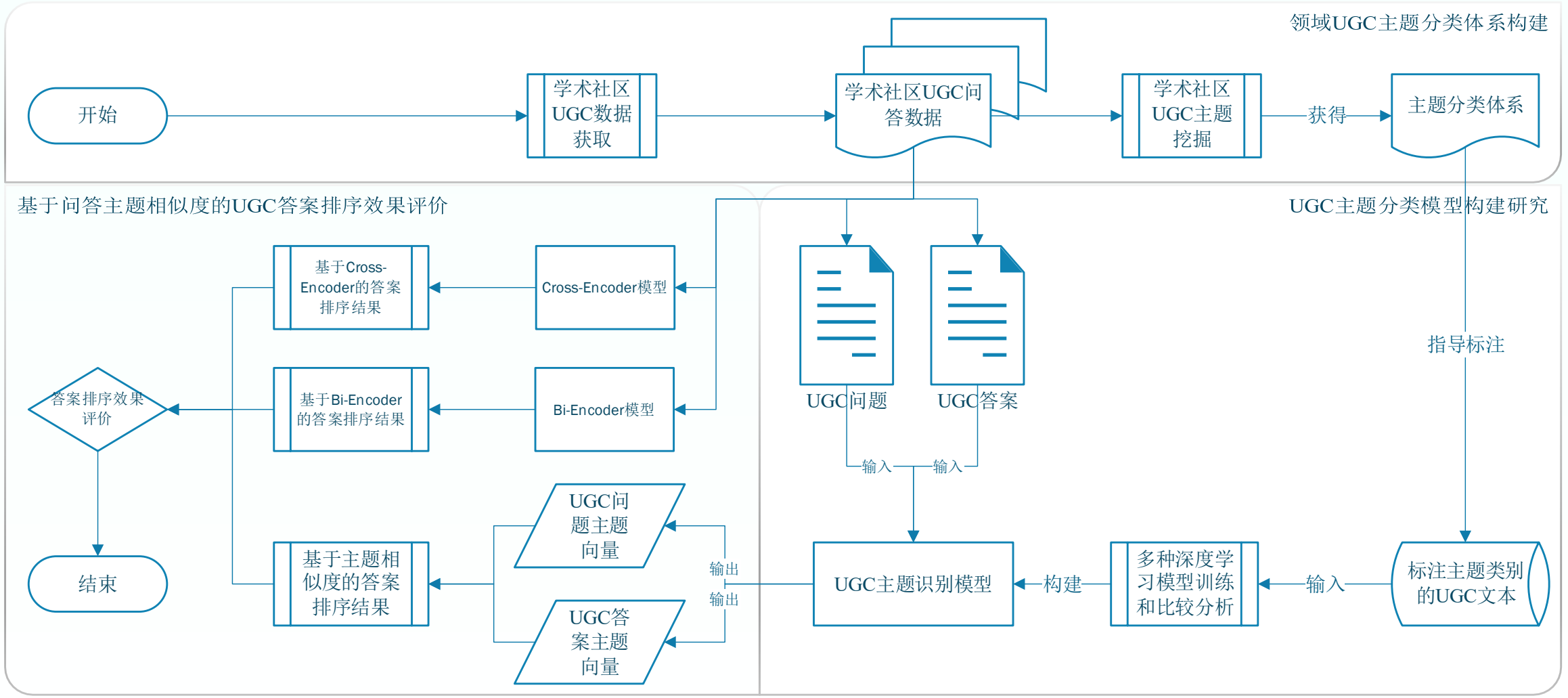


图5-1 基于UGC主题相似度的答案排序策略示意图

三 实证研究

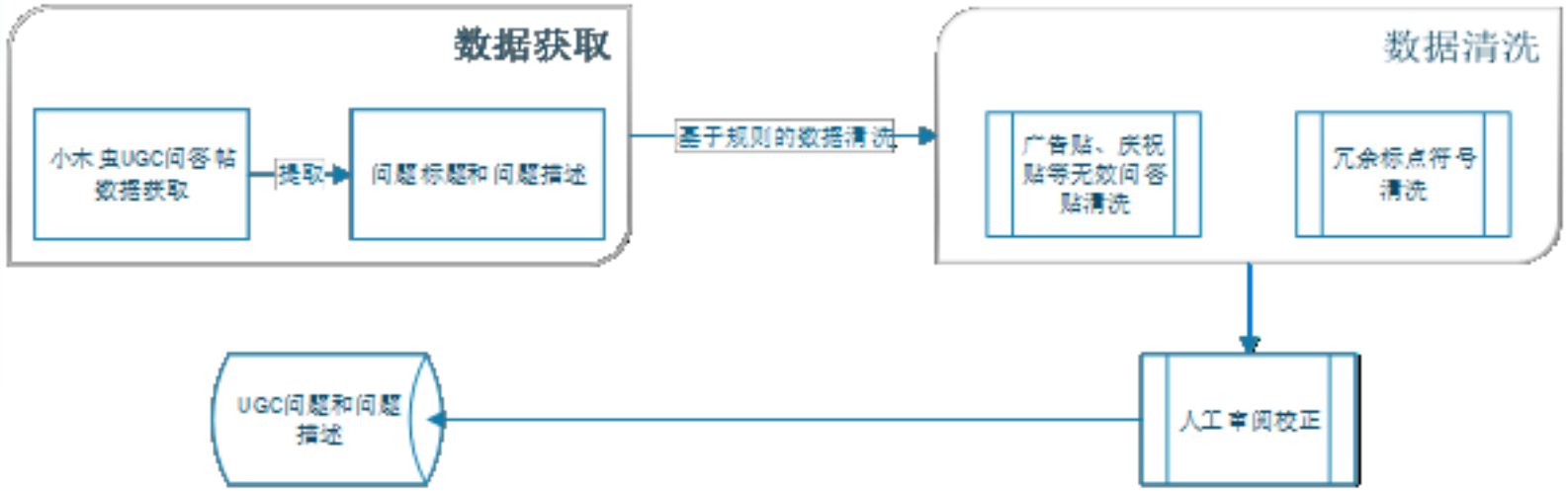
1.整体技术路线


领域UGC主题分类体系构建
开始
学术社区UGC数据获取
学术社区UGC问答数据
学术社区UGC主题挖掘
获得
主题分类体系
基于问答主题相似度的UGC答案排序效果评价
UGC主题分类模型构建研究
基于Cross-Encoder的答案排序结果
Cross-Encoder模型
基于Bi-Encoder的答案排序结果
Bi-Encoder模型
答案排序效果评价
结束
基于主题相似度的答案排序结果
UGC问题主题向量
UGC答案主题向量
输出
输出
UGC问题
UGC答案
输入
输入
指导标注
UGC主题识别模型
构建
多种深度学习模型训练和比较分析
输入
标注主题类别的UGC文本

2.在线学术社区问答UGC主题类别挖掘

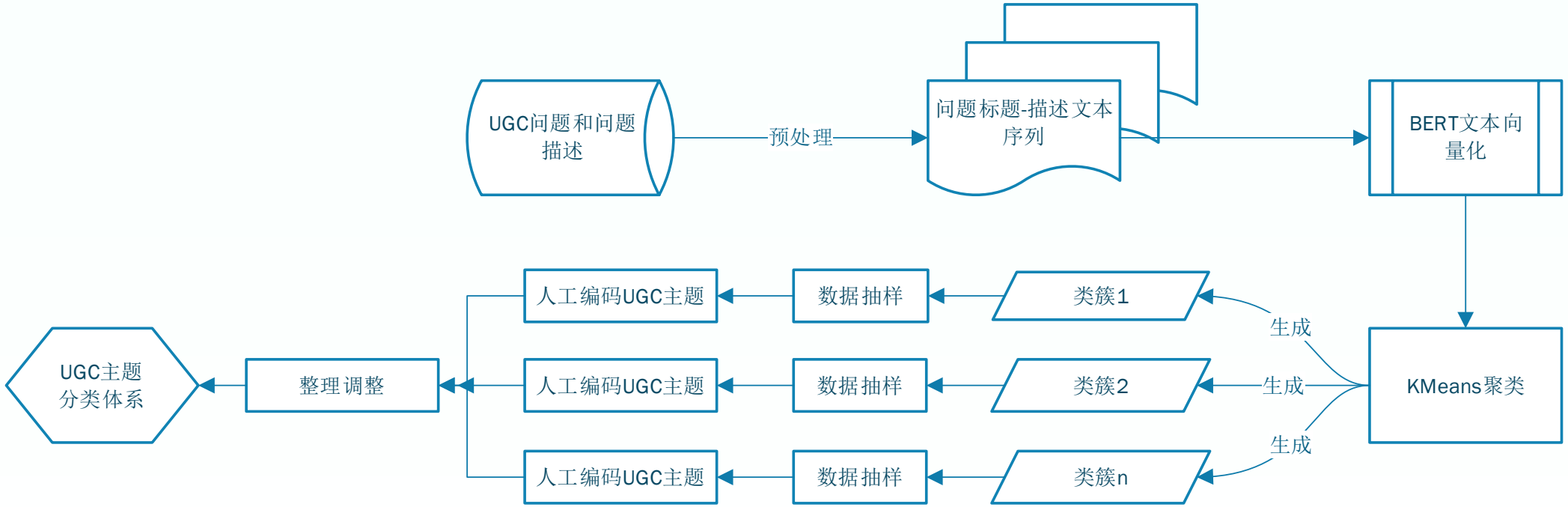
◆数据获取和预处理

- 数据源：“小木虫”社区“论文投稿”板块的用户发帖数据（含问题、问题描述、用户生成答案）
- 数据获取和清洗：Python程序



2.在线学术社区问答UGC主题类别挖掘

◆基于问题文本的UGC主题类别挖掘过程



2.在线学术社区问答UGC主题类别挖掘

◆基于问题描述文本的UGC主题类别挖掘结果

图3-6 学术社区UGC主题编码结果

3.在线学术社区问答UGC主题分类模型构建

◆基础模型选取：TextCNN、BiRNN-Attention、BiLSTM-Attention、BiGRU-Attention

◆模型训练策略优化策略：将损失函数由CrossEntropy Loss替换为PolyCrossEntropyLoss[7]

◆实验结果：Chinese-roberta-wwm-ext-poly模型最优，加权F1值为0.779

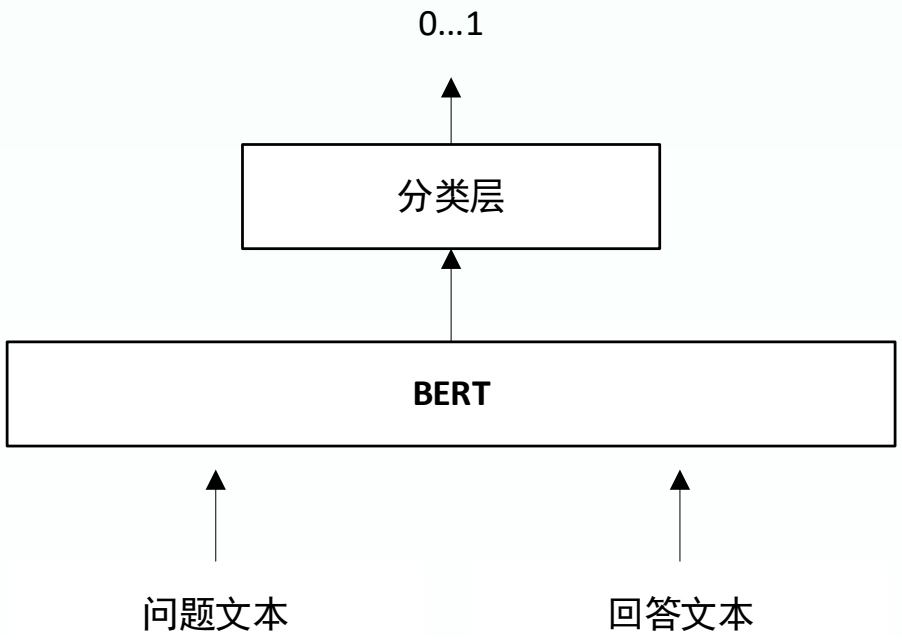
表4-6 基于深度预训练模型（PolyCrossEntropy）问答UGC主题分类结果

Model	Accuracy	Pweighted	Rweighted	Fweighted
bert-base-Chinese-poly	0.745	0.746	0.745	0.744
bert-base-multilingual-cased-poly	0.735	0.734	0.735	0.734
Chinese-bert-wwm-ext-poly	0.745	0.748	0.745	0.745
Chinese-roberta-wwm-ext-poly	**0.780**	**0.782**	**0.780**	**0.779**
Avg	0.751(+0.001)	0.753(+0.002)	0.751(+0.001)	0.751(+0.002)

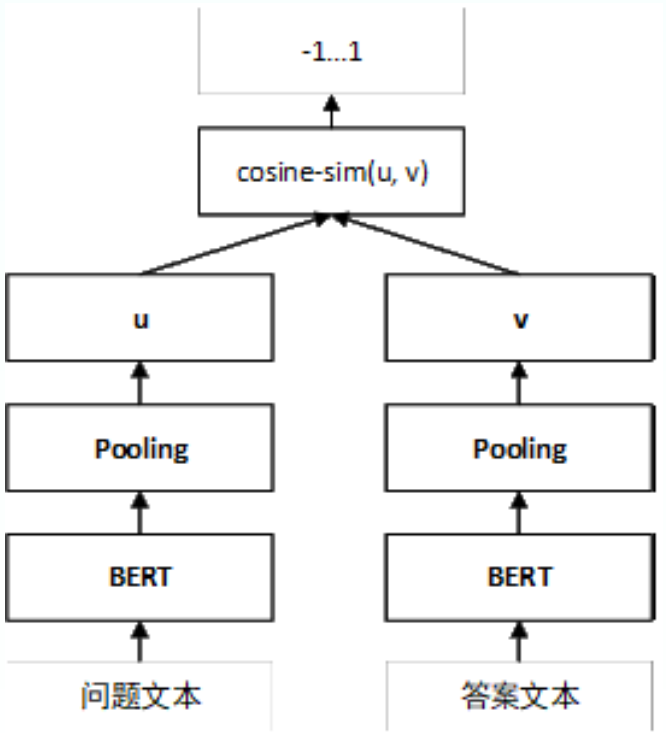
4.基于问答主题相似度的答案质量评价方法有效性验证

◆对比方法介绍

- Cross-Encoder：问题和答案联合建模，输出相关性得分[8]。
- Bi-Encoder：孪生神经网络，输出文本语义相似度内积值[9]。



基于Cross-Encoder架构的问答相关性计算过程



基于Bi-Encoder架构的问答相关性计算过程

4.基于问答主题相似度的答案质量评价方法有效性验证

◆评价指标

- 信息检索评价指标：NDGC、NEER、Q-Measure

◆参考排序结果获取

- “小木虫”学术社区论文投稿板块“求助完结”栏目下的所有提问及每条提问下的全部答案。对于任意答案，将其自带的“点赞数”作为评价答案相关性的依据。最终，构建了包含2715个真实用户问题和14492个用户答案的参考数据集。

注：“求助完结栏目”下的问答帖处于低活跃状态，其中用户回答的点赞数已基本不会变动，能够重返反映真实场景下的答案优劣。因此，该数据适合作为参考答案排序结果。

4.基于问答主题相似度的答案质量评价方法有效性验证

表5-4 三种问答相关性计算方法Top3结果

计算方法	模型名	NDCG@Top-3	NEER@ Top-3	Q-Measure@Top-3
问答主题相似度	Chinese-roberta-wwm-ext	**0.897**	**0.896**	**0.945**
Cross-Encoder	msmarco-distilbert-base-tas-b	0.891	0.888	0.941
	msmarco-roberta-base-ance-firstp	0.895	0.891	0.943
	msmarco-distilbert-base-dot-prod-v3	**0.896**	**0.892**	**0.943**
Bi-Encoder	multi-qa-MiniLM-L6-cos-v1	0.881	0.877	0.936
	multi-qa-distilbert-cos-v1	0.880	0.876	0.935
	multi-qa-mpnet-base-cos-v1	**0.943**	**0.901**	**0.969**

5.研究贡献与展望

◆研究贡献

- （1）所提出的答案质量评价方法更多依靠用户生成的问题文本及答案文本的内在特征，减少了对发帖时间、答题者特征、社会情感特征等外部数据的依赖。
- （2）提出了一种新的基于主题语义相似度的答案质量评价方法。
- （2）探索了面向文本分类任务的预训练模型微调训练策略优化方案。

◆研究展望

- 本研究的主题类别体系的构建过程自动化程度不高，不仅有人工参与，还参照了前人的研究成果。
- 后续研究将探索自动化程度更高的无监督主题分类体系构建方法。

参考文献

[1] 张扬．基于深度学习的社区问答关键技术研究［D］．北京：北京理工大学，2018.

[2] 王宝勋．面向网络社区问答对的语义挖掘研究［D］．哈尔滨：哈尔滨工业大学，2013.

[3] 王伟，冀宇强，王洪伟，等．中文问答社区答案质量的评价研究:以知乎为例［J］．图书情报工作，2017，61（22）：36-44.

[4] 沈旺，李世钰，刘嘉宇，等．问答社区回答质量评价体系优化方法研究［J］．数据分析与知识发现，2021，5（2）：83-93.

[5] 袁健，刘瑜．基于混合式的社区问答答案质量评价模型［J］．计算机应用研究，2017，34（6）：1708-1712.

[6] 郭顺利，张向先，陶兴，等．社会化问答社区用户生成答案质量自动化评价研究——以“知乎”为例［J］．图书情报工作，2019，63（11）：118-130.

[7] Leng Z, Tan M, Liu C, et al. PolyLoss: A Polynomial Expansion Perspective of Classification Loss Functions[A]. arXiv, 2022.

[8] Mahurkar S, Patil R. LRG at SemEval-2020 Task 7: Assessing the Ability of BERT and Derivative Models to Perform Short-Edits Based Humor Grading[C]//Proceedings of the Fourteenth Workshop on Semantic Evaluation, SemEval@COLING 2020. New York, NY, USA: International Committee for Computational Linguistics, 2020: 858-864.

[9] Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings Using Siamese BERTNetworks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 3982-3992.

感谢！

恳请各位专家学者批评指正！

南京农业大学信息管理学院 林立涛
litaolin@njau.edu.cn