王江涛. 基于预训练大模型的电信诈骗案件分类研究[J]. 智能计算机与应用,2025,15(12):69-73. DOI:10.20169/j.issn.2095-2163.25060902

基于预训练大模型的电信诈骗案件分类研究

王江涛
(中国人民公安大学,信息网络安全学院,北京 100038)

摘 要:电信诈骗案件的增加对社会安全和经济稳定造成严重威胁。传统的诈骗案件分类方法主要依赖于人工经验,分类效率低、准确性不高。本文提出了一种基于双向 Trnsformer 编码器表征模型(Bidirectional Encoder Representations from Transformers, BERT)与潜在狄利克雷分配模型(Latent Dirichlet Allocation, LDA)主题建模的特征融合模型,用于电信诈骗文本笔录的特征提取以及分类。该模型结合了 BERT 模型的深度语义理解能力与 LDA 主题建模的文本主题分析能力,通过特征融合技术,将两者提取的特征进行有效整合,从而更全面地捕捉电信诈骗文本笔录的关键信息。实验结果表明,该模型的分类准确率达 95.24%, *F*1 - score 为 95.04%,显著优于 GLM-4 模型;在 12 类诈骗案件中如刷单返利、冒充电商客服等,均表现出色,分类效果稳定,数据依赖性较强。融合 BERT 与 LDA 的模型能有效捕捉文本语义与主题特征,为电信诈骗案件智能化分类提供了高效解决方案,对提升警务工作效率具有重要实践价值。

关键词:BERT 模型;LDA 主题建模;案件分类;电信诈骗;GLM-4 模型
中图分类号:TP311.13;D631 **文献标志码**:A **文章编号**:2095-2163(2025)12-0069-05

Research on classifying telecommunication fraud cases based on pre-trained large models

WANG Jiangtao
(School of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China)

Abstract: The increase in telecom fraud cases poses a severe threat to social security and economic stability. Traditional methods for classifying fraud cases primarily rely on human experience and are confronted with issues such as low classification efficiency and poor accuracy. This paper proposes a feature fusion model based on the pre-trained large-scale model BERT and LDA topic modeling for feature extraction and classification of telecom fraud text transcripts. This model combines the deep semantic understanding capability of the BERT model with the text topic analysis capability of LDA topic modeling. Through feature fusion techniques, it effectively integrates the features extracted by both methods, thereby more comprehensively capturing key information in telecom fraud text transcripts. Experimental results demonstrate that the model achieves a classification accuracy of 95.24% and an *F*1 - score of 95.04%, significantly outperforming the GLM-4 model. The model performs excellently across 12 categories of fraud cases (such as click farming rebates, impersonating e-commerce customer service, etc.), with stable classification performance and strong data dependency. The experimental results indicate that the model fusing BERT and LDA can effectively capture textual semantic and thematic features, providing an efficient solution for intelligent classification of telecom fraud cases and offering significant practical value in enhancing police work efficiency.

Key words: BERT models; LDA topic modeling; classification of cases; telecom fraud; GLM-4 model

0 引 言

电信诈骗案件的增加对社会安全和经济稳定造成了严重威胁。传统的诈骗案件分类方法主要依赖于人工经验,分类效率低、准确性不高[1]。随着人工智能技术的迅猛发展,特别是预训练大模型在自然语言处理领域的显著突破,电信诈骗案件的分类逐渐向自动化和智能化方向发展[2]。

在电信诈骗案件分类中,纪杰等[3]提出基于情境分析的分类方法,结合提示学习技术对案件文本进行自动分类,其准确率和 *F*1 值较传统 BERT 模型提升 1%~2%。当前更为主流研究是以预训练语言

作者简介:王江涛(2000—),男,硕士研究生,主要研究方向:预训练大模型。Email:791508714@qq.com。

收稿日期:2025-06-09

模型为基础,结合对抗训练提升模型鲁棒性,刘伟等[4]提出的模型融合了对抗扰动生成、长短期记忆网络(Bidirectional Long Short - Term Memory, BiLSTM)上下文句法分析以及卷积神经网络(Convolutional Neural Network, CNN)局部语义提取,在电信诈骗案件数据上达到83.9%的分类准确率。Hilas[5]等聚焦电子平台欺诈类型识别,通过系统文献回顾发现信用卡欺诈占比达82.61%,并验证随机森林算法在欺诈检测中的最优性能。上述研究通过微调与深度学习等方式提高了分类的准确性,但在标注数据稀缺或数据质量较低时的模型效果容易受限,对数据中所包含特征的提取不够全面[6]。

此外,目前研究数据来源主要集中于警情数据、接处警数据、互联网电诈案例等,存在案件描述不够细致、真实性和时效性难以保证等特点[7]。因此,本文借助电信诈骗笔录文本数据,在保证案件基本特征与细节丰富性的前提下,采用BERT模型与LDA主题建模结合的特征融合模型,将两者提取的特征进行有效整合,用于电信诈骗案件分类研究。

1 电信诈骗案件文本分类预训练大模型构建

1.1 BERT模型基本架构

BERT模型的核心结构为多层Transformer编码器[8]。每一层Transformer编码器内部包含自注意力机制和前馈神经网络,这些组件共同工作,使模型能够深入理解文本数据。在输入端,文本被分割成一系列词元(token),并通过嵌入层转换为高维向量,这些向量随后被送入多层Transformer编码器中进行处理,每一层都会进一步提取和整合文本中的信息[9-10];最终,模型输出一个包含丰富上下文信息的向量表示,为后续的任务,如分类、命名实体识别等提供强有力的支持[11]。

1.2 LDA主题建模模型基本架构

在文本挖掘中,LDA主题模型是较为常用的一种识别文档主题的统计模型[12],该模型基于概率生成式假设,认为文档是由一系列主题的混合生成的,而每个主题又是由一系列单词的分布定义的[13],如图1所示。

(1)α和β是模型的超参数,K是主题的数量,N是文档的数量,N_i表示文档i中的所有单词,θ_i表示文档i的主题分布,ϕ_k表示主题k的单词分布,Z_{ij}表示文档i中第j个单词的主题,W_{ij}表示文档i的第j个单词。

(2)文档i的主题分布θ_i,服从参数为α的狄利克雷(Dirichlet)分布;

(3)主题k的单词分布ϕ_k,服从参数为β的狄利克雷(Dirichlet)分布;

(4)文档i中第j个单词的主题Z_{ij},服从参数为θ_i的多项式分布;

(5)文档i的第j个单词,服从参数为$\varphi_{Z_{ij}}$的多项式分布。

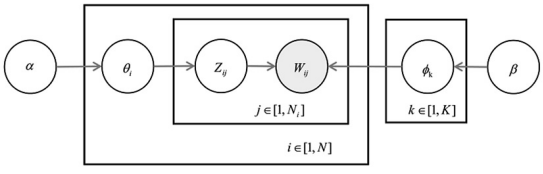

图1 LDA主题模型结构图

Fig. 1 LDA topic model structure diagram

1.3 基于LDA-BERT的电诈笔录文本分类模型构建

数据进行预处理操作,主要包括数据清洗、分词处理、去除停用词等,数据预处理阶段确保输入数据的质量和一致性,为后续的特征提取提供可靠的基础;

利用BERT模型与LDA主题建模进行语义特征提取,BERT模型通过预训练学习到的语言知识和上下文信息,能够捕捉到文本中的深层语义特征;LDA主题建模则通过分析文本中单词的共现关系,挖掘出文本的主题结构,为文本分类提供另一种维度的特征;

特征融合将BERT提取的语义嵌入表示与LDA生成的主题分布相结合,形成联合特征表示;

使用线性分类器Softmax对融合后的特征进行分类,预测案件类别。

2 实验与分析

2.1 数据集与分类指标

本文以北方T市某分局300份电信网络诈骗案件被害人笔录数据为研究对象,保留被害人年龄、性别、户籍地、工作单位、政治面貌、文化程度、职业、民族等特征信息,将其他涉及个人隐私的敏感信息作删除处理,共保留文字717 631字。

本文采用准确率、精确率、召回率、$F1$ - score等指标评估模型性能。

(1)准确率(Accuracy)是指所有正确预测的样本(包括正类和负类)占总样本数量的比例:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

其中，TP 表示模型正确预测为正类的样本数；TN 表示模型正确预测为负类的样本数；FP 表示模型错误预测为正类的样本数；FN 表示模型错误预测为负类的样本数。

(2)精确率(Precision)是指正确预测为正类的样本数量与模型预测为正类的样本总数的比例，关注预测为正类的样本的准确性：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

(3)召回率(Recall)是指正确预测为正类的样本数量与实际为正类的样本总数的比例，关注模型捕捉正类的能力：

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

(4) $F1$ - score 是精确率和召回率的调和平均数，提供了精确率和召回率之间的平衡：

$$F1 - \text{score} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

2.2　基于 LDA-BERT 模型的电诈文本笔录特征提取

LDA 模型性能评估的一个重要指标就是模型的困惑度[14]。困惑度越低，表示 LDA 模型对于文档的生成概率越高，模型对文档的主题分布拟合得越好。因此在最优主题数的确定上，通过对不同主题数量的困惑度进行分析，得出困惑度与主题数量的关系如图 2 所示，得出在主题数为 12 时，困惑度较低，从而确定了模型的主题数为 12。

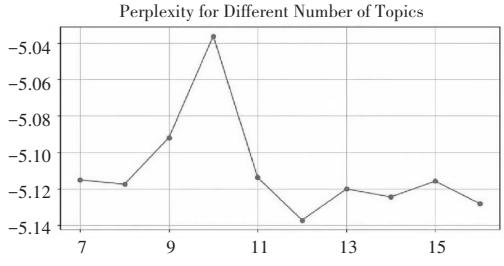

图 2　困惑度与主题数量关系

Fig. 2　Relationship between perplexity and the number of topics

对笔录文本数据进行预处理的操作后，利用 BERT 计算每个 token 的嵌入向量与整个句子嵌入向量的相似度。BERT 特征提取结果见表 1。

表 1　BERT 特征提取结果(部分)

Table 1　BERT feature extraction results (partial)

编号	文本内容(摘要)	BERT 特征提取结果(前 5 维)
1	"银行客服要求转账到安全账户"	[0.24, -0.56, 0.78, 1.23, -0.12]
2	"网购平台退款诱导提供验证码"	[0.18, -0.61,0.82, 0.97, -0.45]
3	"公检法声称涉嫌洗钱要求配合"	[0.75, -0.33, 0.45, 1.56, 0.23]
4	"中奖要求支付手续费领取奖金"	[-0.12, 0.45, 0.67, 0.89, -0.78]
5	"投资平台高收益"	[0.31, 0.28, -0.56, 1.12, 0.45]

BERT 模型的特征提取能力在该结果中得到了初步验证，可见不同电信诈骗类型的文本内容对应的 BERT 特征向量在前五维上呈现出明显的差异，这种差异性为后续的分类任务提供了有力的特征支持。

2.3　特征融合

本文使用自动编码器将 BERT 模型的语义特征向量与 LDA 的主题特征向量融合，借助 LDA-BERT 模型成功识别出电信诈骗类犯罪的 12 类核心主题，并从每个主题中筛选出题词对应概率排名前十的词语，电诈主题识别结果见表 2。

2.4　分类依据

在分类器的选取中，本文选取 Softmax 分类器输出每个类别的概率分布。

刑法中并没有直接对电信诈骗案件进行详细的分类。而在实际司法实践中，为了更有效地打击和防范电信诈骗犯罪，执法和司法机关可能会根据电信诈骗的具体手法、特点和危害程度等因素，对其进行一定的分类和归纳，有助于警方更好地识别、预防和打击电信诈骗犯罪，也有助于公众提高警惕，增强防范意识[15-17]。

公安部反诈中心根据诈骗手法和案件特点，将电信诈骗类犯罪大致归纳为 12 种类别，各类电信诈骗定义见表 3。

2.5　分类结果

为更直观地理解模型的性能，从而为模型的优化和实际部署提供依据。除此之外，为了方便后续模型评测性能效果，最终分类结果见表 4。模型的准确率、精确率、召回率、$F1$ - score 指标见表 5，各项指标指数均在 0.95 之上，证明模型的分类效果较好。

表 2 电诈主题识别结果

Table 2 Telecom fraud subject identification results

序号	主题标识	主题核心词(10 个)
1	刷单返利	兼职、高佣金、垫付、返利、轻松赚钱、任务单、扫码、群聊、保证金、信用分
2	冒充电商物流客服	退款、退货、快递丢失、质量问题、链接、账户异常、赔偿金、操作失误、安全认证、验证码
3	虚假网络投资理财	高回报、内幕消息、稳赚不赔、导师带单、数字货币、平台漏洞、提现冻结、对冲、私募、杠杆
4	贷款代办信用卡	低利息、秒放款、无抵押、包装资质、手续费、解冻费、银行关系、黑户可办、额度、征信修复
5	虚假征信	征信污点、影响贷款、账户异常、学生贷注销、银监会、清零记录、安全账户、转账解封、授权
6	虚假购物服务	低价促销、海外代购、限量款、预付定金、货到付款、海关扣留、系统卡单、砍单、走私、定金
7	冒充公检法及政府机关	涉案、通缉令、保密调查、资金清查、安全账户、开庭通知、警官证、检察院、冻结、配合调查
8	冒充领导熟人	急用钱、转账、新号码、项目款、开会不便、代付、账户变更、关系疏通、借款、伪造签名
9	网络游戏产品虚假交易	装备交易、账号代练、低价皮肤、充值、解封申诉、外挂、担保客服、私下交易、回购、CDK
10	网络婚恋交友	杀猪盘、跨国恋、投资漏洞、未来规划、生病急用钱、赌博网站、见面路费、创业、彩礼、基金
11	网黑	黑产、技术攻击、数据泄露、有偿删帖、负面压制、舆论操控、付费删记录、黑客接单、曝光
12	冒充军警购物	部队采购、指定供应商、垫资、罐头帐篷、军用物资、后勤合同、定金预付款、发票、领导批示

表 3 各类电信诈骗定义

Table 3 Definition of various types of telecom fraud

电诈类别	具体表述
刷单返利	刷单返利诈骗通过发布兼职广告,以小额返利诱骗受害者刷单,逐步加大金额后不返还本金和报酬,实施诈骗。
冒充电商物流客服	诈骗分子冒充电商平台或物流客服,以理赔退款或激活账户为由,骗取受害者银行信息或转账。
虚假网络投资理财	骗子构建虚假投资平台,承诺高额回报诱骗投资,一旦投入资金便消失或拒绝提现。
贷款代办信用卡	不法分子冒充银行或贷款公司,以提升额度或低息贷款为由,收取手续费或保证金实施诈骗。
虚假征信	骗子冒充客服,以影响征信为由诱骗贷款并转账至指定账户。
虚假购物服务	通过虚假广告诱导付款后失联,或以加缴费用为由骗取钱财。
冒充公检法及政府机关	冒充公检法人员,以涉嫌犯罪为由要求转账至“安全账户”。
冒充领导熟人	利用信任关系,以资金周转或帮忙办事为由骗取钱财。
网络游戏产品虚假交易	发布虚假游戏装备交易信息,骗取钱财或无效道具。
网络婚恋交友	通过婚恋平台建立感情关系后骗取钱财。
网黑	黑客或犯罪团伙利用网络技术从事非法活动,如窃取信息或破坏系统。
冒充军警购物	冒充军队或警方人员,以采购物资为由骗取货款或货物。

表 4 文本分类结果

Table 4 Text classification results

编号	类别	分类数量
0	刷单返利	42
1	冒充电商物流客服	39
2	虚假网络投资理财	12
3	贷款代办信用卡	27
4	虚假征信	21
5	虚假购物服务	21
6	冒充公检法及政府机关	30
7	冒充领导熟人	20
8	网络游戏产品虚假交易	18
9	网络婚恋交友	17
10	网黑	18
11	冒充军警购物	35

表 5 LDA-BERT 模型分类指标

Table 5 Classification indicators of LDA-BERT model

评估指标	数值
准确率	0.952 4
精确率	0.960 9
召回率	0.952 4
*F*1 分数	0.950 4

2.6 模型对比评测

在电信诈骗案件分类中,预训练大模型的效果评测基于多种指标,以确保其在实际应用中的有效性和可靠性。为了使模型效果更加直观,将 LDA-BERT 模型与智谱 GLM-4 模型进行对比实验。选取包含 300 条电信诈骗案件笔录的数据集,采用 70%的训练集和 30%的测试集划分方式,以 50 个 epoch 为训练

周期进行训练，分类指标对比结果见表 6。

表 6　分类指标对比

Table 6　Comparison of categorical metrics

评估指标	GLM-4	LDA-BERT
准确率	0.747 8	0.952 4
精确率	0.658 0	0.960 9
召回率	0.747 8	0.952 4
*F*1 分数	0.647 9	0.950 4

LDA-BERT 模型的性能优于 GLM-4 模型，说明 LDA-BERT 模型在此特定任务中具有更强泛化能力。在实际部署中，模型的推理速度也是重要的评测指标，LDA-BERT 模型每条推理速度为 0.08 s，快于 GLM-4 模型。综合各项评测结果，LDA-BERT 模型在电信诈骗案件分类中表现出色，尤其在分类精确度和误判率方面均更优秀，更适合该领域的应用需求。

3　结束语

本文将 BERT 预训练大模型与 LDA 主题建模的特征融合，使得模型在电信诈骗案件分类中显示出显著的效果提升与应用潜力。LDA-BERT 模型的 *F*1 - score 达到 0.95 以上，显示出良好的分类能力，也显示出其对文本理解的更高敏感度及上下文捕捉能力。相较于 GLM-4 模型，LDA-BERT 在分类精确度、召回率以及 *F*1 - score 等多个评估指标上优势显著，为实际警务工作中对电信诈骗案件进行分类提供了可行性与可操作性。此外，LDA-BERT 模型在推理速度上的优势也为其在实时应用中的部署提供了有力支持，在处理大量电信诈骗案件文本数据时，能够迅速给出准确的分类结果，这对于提高警务工作效率、及时响应诈骗案件具有重要意义。

当前预训练大模型在电信诈骗案件分类中的应用展现出一定的优势，但也存在一些不足之处。数据依赖性较强，模型性能受限于数据的质量与多样性。在训练过程中，本文使用的电信诈骗案例数据样本为 300 个，模型对未被充分学习的新型诈骗手法的分类准确率较低。

模型的可解释性也是一个明显的不足，当前使用的注意力机制虽带来了部分可解释性，但仍无法全面揭示分类决策的内部逻辑，需进一步探索可解释模型的开发。

未来可通过增强学习改善模型的适应能力，设计一个自适应性强的模型使其能够实时更新，学习新出现的诈骗模式。此外，建立更大规模、多样化的标注数据集，以丰富模型的训练数据，提升其分类能力。探讨融合多个模型的集成学习方法，借助不同模型的优点，提升整体性能。在处理推理效率方面，应考虑量化和剪枝等模型压缩技术，以降低推理时间和内存占用，使模型部署更具灵活性，可以通过结合可解释性 AI（eXplainable Artificial Intelligence）方法，如 LIME（Local Interpretable Model - Agnostic Explanations）或 SHAP（SHapley Additive exPlanations），增强模型输出的透明度，从而提升用户的信任感，确保模型在实际应用中的有效性。

参考文献

[1] 马宇畅. 电信网络诈骗犯罪侦查中合成作战机制研究[J]. 网络安全技术与应用，2024(11)：156-159.

[2] 邓翠艳，齐小刚. 基于 Transformer 及多任务学习的电信网络诈骗识别[J]. 应用科技，2024 (5)：256-262.

[3] 纪杰，孙承杰，单丽莉，等. 基于提示学习的电信网络诈骗案件分类方法[J]. 山东大学学报(理学版)，2024，59(7)：113-121.

[4] 刘伟. 电信诈骗犯罪的新趋势与防治对策[J]. 辽宁警专学报，2011，13(6)：53-56.

[5] HILAS C S. Designing an expert system for fraud detection in private telecommunications networks[J]. Expert Systems with Applications，2009，36(9)：11559-11569.

[6] 陈锡清，叶蕴芳. 电信诈骗风险防控模型的研究与应用[J]. 通信世界，2024(19)：44-45.

[7] 牛硕. 面向电信诈骗案例数据的事件联合抽取方法[D]. 北京：中国人民公安大学. 2024. DOI：10.27634/d.cnki.gzrgu.2024.000408

[8] 金辉. 大模型赋能涉网新型犯罪的打击和治理[C]//中国网络空间安全协会人工智能安全治理专委会. 2024 世界智能产业博览会人工智能安全治理主题论坛论文集. 国投智能信息股份有限公司. 2024：35-38. DOI：10.26914/c.cnkihy.2024.009838

[9] 狄婷. 电信网络诈骗案件特征及防控对策研究[D]. 太原：山西师范大学，2020. DOI：10.27287/d.cnki.gsxsu.2020.000921

[10] 尹金鑫，尹军祖. 基于改进预训练模型的裁判文书摘要生成研究[J]. 智能计算机与应用，2025，15(6)：50-57. DOI：10.20169/j.issn.2095-2163.24122604

[11] BORKETEY B. Real-time fraud detection using machine learning[J]. Journal of Data Analysis and Information Processing，2024，12(2)：189-209. DOI：10.4236/jdaip.2024.122011

[12] SUBUDHI S，PANIGRAHI S. Quarter-sphere support vector machine for fraud detection in mobile telecommunication networks[J]. Procedia Computer Science，2015，48：353-359. DOI：10.1016/j.procs.2015.04.193

[13] 林国荣. 论当前电信诈骗问题与防治对策[D]. 福州：福建师范大学，2014.

[14] 陈斌. 电信诈骗犯罪的犯罪学研究[D]. 宁波：宁波大学，2012.

[15] 蒋涵雨. 基于通信行为特征的诈骗号码识别研究[D]. 上海：华东师范大学，2022. DOI：10.27149/d.cnki.ghdsu.2022.002762

[16] 刘润程. 基于多模态特征融合的谎言行为识别模型的研究与实现[D]. 北京：北京邮电大学，2022. DOI：10.26969/d.cnki.gbydu.2022.000907

[17] 赖云龙. 论电信诈骗犯罪的防治对策[D]. 长春：吉林大学，2019. DOI：10.27162/d.cnki.gjlin.2019.000833