

基于高效深度神经网络的社交网络隐私数据自动感知模型

郭 祯, 范晨阳, 徐 嘉, 刘 干

海南大学网络空间安全学院 海口 中国 570228

摘要 目前大多数针对在线社交网络用户的隐私感知算法难以精确识别和定位隐私泄露的具体文本位置, 并且在处理嵌套隐私方面表现较差, 难以满足用户多样化的隐私保护需求。针对嵌套隐私识别, 尽管已有较多研究成果, 然而, 由于嵌套隐私本身的复杂性, 实际应用中依然面临识别准确率低和识别速度慢等挑战。具体而言, 嵌套隐私识别不仅仅局限于简单的实体识别任务, 还需要深入理解隐私信息的层级关系和上下文。这些层级关系可能包括个人信息、社交关系以及行为模式等, 它们在不同情境下的隐私敏感度可能存在显著差异。为了解决这些问题, 本文提出了一种基于深度学习的社交网络隐私信息内容提取模型, 旨在更精准地识别和定位在线社交网络中隐私泄露的具体文本位置, 有效处理嵌套隐私问题, 提升隐私识别的准确率与速度。该模型结合了改进的 RoformerBERT 模型、BI-LSTM 模型和高效的全局指针算法(Efficient Global Pointer, EGP)构建了一个基于深度学习的快速隐私实体识别模型(Deep Learning-based Fast Privacy Entity Recognition Model, FPERM)。FPERM 模型可以自动感知社交网络中用户共享的隐私信息, 并准确定位文本中泄露的敏感信息位置, 提升训练速度的同时保持模型性能。实验结果表明, FPERM 模型高达 94.01% 的出色总体准确率, 并将提出的模型与当前最先进的方法进行了比较分析, 结果表明提出的模型具有更优越的性能以及更强的泛化能力, 能够有效保护社交网络的隐私信息。

关键词 社交网络; 深度学习; 隐私感知; 嵌套实体

中图分类号 TP183 DOI 号 10.19363/J.cnki.cn10-1380/tn.2026.05.08

Automatic Sensing Model for Private Data in Social Networks Based on Efficient Deep Neural Networks

GUO Zhen, FAN Chenyang, XU Jia, LIU Gan

School of Cyberspace Security, Hainan University, Haikou 570228, China

Abstract Most of the current privacy-aware algorithms for online social network users have difficulty in accurately identifying and locating the specific text location of privacy leakage, and perform poorly in dealing with nested privacy, making it difficult to meet the diverse privacy protection needs of users. Although there have been many research results on nested privacy recognition, due to the complexity of nested privacy, it still faces challenges such as low recognition accuracy and slow recognition speed in practical applications. Specifically, nested privacy recognition is not only limited to the simple entity recognition task, but also requires an in-depth understanding of the hierarchical relationships and contexts of privacy information. These hierarchical relationships may include personal information, social relationships, and behavioral patterns, etc., and their privacy sensitivities may differ significantly in different contexts. To solve these problems, this paper proposes a deep learning-based content extraction model for social network privacy information, which aims to more accurately identify and locate the specific text location of privacy leakage in online social networks, effectively deal with the nested privacy problem, and improve the accuracy and speed of privacy recognition. The model combines the improved RoformerBERT model, BI-LSTM model, and efficient global pointer algorithm (EGP) to construct a Deep Learning-based Fast Privacy Entity Recognition Model (FPERM). The FPERM model can automatically sense the privacy information shared by users in the social network and accurately locate the location of leaked sensitive information in the text, improving the training speed of the model while maintaining the model performance. The experimental results show that the FPERM model achieves 94.01% in overall accuracy, and the proposed model is compared and analyzed with the current state-of-the-art methods, and the results show that the proposed model has a superior performance as well as a stronger generalization ability, which can effectively protect the private information in social networks.

Key words social networks; deep learning; privacy-awareness; nested entities

通讯作者: 徐嘉, 博士研究生, Email: jiaxu@hainanu.edu.cn。

本课题得到国家自然科学基金项目(No. 62262013)资助。

收稿日期: 2024-10-13; 修改日期: 2025-02-26; 定稿日期: 2026-04-09

1 引言

在线社交网络(Online Social Network, OSN)是一种集多功能服务于一体的数字交互社区^[1]。作为人们对现实生活和活动的有效延伸,社交网络凭借其便捷、灵活的信息传播和快速高效的网络沟通,充分满足了人们表达个人诉求、分享各种信息的需求,已成为观点、信息、行为和文化传播的主要载体。例如,在美国总统大选中,社交网络发挥了至关重要的作用,候选人通过 X(原推特)和 Facebook 等平台与选民建立联系,有效动员支持者。目前,大规模在线社交平台已经形成了一个独立的生态系统(如微信、微博、Facebook、Instagram 等)。社交平台用户数量急剧增长,用户也越来越热衷于通过社交网络相互交流和分享信息。

OSN 具有信息类型多样、规模庞大和共享公开等特点,但是多数社交网络用户对个人隐私的保护意识较为淡薄,他们使用社交网站发布的文字、照片、视频、定位等各类形式复杂的内容有意或无意间包含了用户真实的隐私信息。网络攻击者可以很容易地获取大量用户数据,经过数据分析可以很容易地挖掘用户的隐私信息,可能导致一系列恶意行为^[2-3],包括但不限于跟踪、诽谤、垃圾邮件、网络钓鱼、身份盗窃、个人数据克隆、Sybil 攻击等隐私泄露事件。例如,分享旅游住宿的环境可能会泄露用户的具体位置,威胁到自身的安全。此外,分享关于二手物品的信息可能会吸引很多来自职业诈骗人的垃圾邮件和电话促销。这些信息的披露会导致无休止的安全事件。因此,如何有效管理用户在社交网络上的隐私需求,以及如何准确高效地感知和识别社交网络上发布信息的隐私泄露已成为亟待解决的问题。这不仅对于帮助用户更好地理解 and 应对隐私风险、防止隐私泄露具有重要的理论意义,同时也为社交网络服务提供者改进隐私保护策略、提升用户体验提供了实践价值。因此,深入研究用户的隐私管理行为、隐私感知能力以及隐私保护技术,对于构建安全、可信赖的社交网络环境具有非常重要的理论意义和应用前景。

目前,针对社交网络中用户的隐私和安全问题,已经提出了许多隐私保护技术。最直接的方法旨在保护社交网络^[4]中的隐私信息。提出了包括匿名化、去中心化、加密、信息安全监管、细粒度的隐私设置和访问控制等隐私保护技术,以及增强用户对隐私的意识和行为^[5]。这些技术的前提是识别出社交网络中的隐私数据,并进行相应的隐私处理。然而,这

些方法通常只能识别特定类别的隐私信息,对于间接隐私泄露保护效果较差。同时,一些研究人员基于社交网络的动态特性,对社交网络中的隐私保护进行了深入研究。他们提出利用数据监控和数据分类的方法来分析动态社会网络中的隐私信息,并探讨了用于隐私保护的集中式^[6]和分布式^[7]技术。尽管这些方法在隐私保护方面取得了一定的成效,但无法满足个体对隐私控制和追求个性化保护的特定需求。

为了更好地保护用户的隐私,研究人员开始尝试利用深度学习(Deep Learning, DL)和机器学习(Machine Learning, ML)技术进行隐私感知。一些研究将隐私信息分为几个相对宽泛的类别,利用 ML 算法来预测用户隐私感知的类别^[8]。其他研究人员将文本标记为敏感或非敏感,然后使用 DL 模型进行文本分类^[9]。然而,考虑到社交网络文本的复杂性、文本长度的变化,以及采用相对位置信息的局限性等问题,这些方法只能判定隐私泄露的大致类别,对文本中一个实体的内部或下一级包含另一个实体的嵌套实体(如:“我在北京字节传媒公司拍了一张照片。”其中北京字节传媒公司为嵌套实体,包含北京和北京字节传媒公司两个实体)识别较差且难以精确地定位到隐私泄露的具体文本位置。

在已有的相关工作中已经初步解决了隐私实体的感知问题,但是训练时间相对较长,以及对部分隐私实体的识别效果还有所提升。为了解决上述问题,本文提出了一种基于 OSN 隐私信息内容提取的隐私感知模型。该模型结合 RoformerBERT 模型、基于跨度的位置嵌入和注意力机制,并通过优化训练结构,显著提高嵌套实体和长文本的识别准确率。此外,模型采用 BI-LSTM 模型增强语义信息和上下文特征,并生成基于上下文的动态权重,从而更好地适应上下文变化。最后,通过改进的 EGP 算法对特征信息进行解码,从而精准定位隐私泄露位置。该模型能够自动感知用户在社交网络中共享的文本敏感信息,准确定位隐私泄露的具体文本位置并反馈给用户,以增强用户的隐私意识。本文的贡献可归纳如下:

(1) 建立了一个高效的 OSN 隐私感知模型 FPERM。该模型通过对文本信息进行充分的特征化和编码,实现自动感知社交网络中用户共享的敏感文本信息并准确定位敏感信息位置。

(2) 使用基于跨度的旋转位置编码(Span-based Rotary Position Embedding, SRPE)融合绝对位置和相对位置的信息,增强模型对跨距离依赖关系的感知能力,结合注意力机制有效提高嵌套实体和长文本的识别准确率。通过优化归一化层,提升训练速度的

同时保持模型的稳定性。

(3) 采用 EGP 算法在大幅减少参数的同时提高效率 and 准确率, 引入动态权重使模型更好地适应上下文变化, 解决社交网络中嵌套隐私实体和非嵌套隐私实体分布不平衡的问题。

2 相关工作

社交网络的发展带来便利的同时, 也对用户隐私保护提出了巨大挑战。因此, 如何对社交网络用户信息进行隐私保护是亟须解决的问题。本文从传统的社交网络隐私研究以及基于自然语言处理(Natural Language Processing, NLP)和 DL 的社交网络隐私研究两个方面介绍相关工作。

2.1 传统的社交网络隐私研究

在社交网络隐私感知研究方面, 研究人员考虑了不同因素对隐私的影响, Srivastava 等人^[10]对用户发布字符信息造成的隐私泄露进行了深入研究, 利用项目反应理论模型(Item Response Theory, IRT)来衡量在线社交网站用户发布的信息和文本中的隐私泄露情况。Liang 等人^[11]发现在各种社交网络上删除已发布的图片后, 在一定时间内仍然可以被访问, 而这种删除延迟可能会造成不必要的隐私泄露。随着研究的发展, 越来越多的因素被考虑到隐私度量中。Maximilien 等人^[12]测量了用户简介中多个属性信息的敏感度和可见性, 然后通过贝叶斯统计计算出统一的分数。Liu 等人^[13]提出了一种 IRT 融合属性的敏感度和可见性, 更合理地计算出隐私评分。Aghasian 等人^[14]将属性敏感度和可见性相结合, 使用统计模糊系统解决了在多个社交网络平台上计算隐私分数的问题。Pensa 等人^[15]认为单个用户的隐私泄露取决于他对周围朋友的信任, 提出了一个基于信任感知的框架来度量用户的隐私泄露。Alsakal 等人^[16]利用信息熵理论为整个社交网络中的用户引入了信息度量, 讨论了个体识别信息和不同片段信息的组合对用户信息泄露的影响。

在社交网络隐私保护研究方面, 研究人员提出了基于 k 匿名的隐私数据保护技术、数据扰动技术、基于密码学的隐私保护技术和基于差异隐私的数据保护方法等。基于 k 匿名的隐私数据保护技术主要是将 k 匿名模型应用于社交网络, 对隐私数据进行泛化和隐藏^[17]。数据扰动技术主要基于数据随机化保护的思想, 利用数据随机化对敏感信息进行加密^[18]。基于密码学的隐私保护技术提供了差异化的社交网络隐私保护^[19]和同态加密网络数据隐私保护^[20]。这些技术的前提是识别社交网络中的隐私数据, 然后进行

相应的隐私处理^[21]。这些技术大多通过隐私信息测量表来识别隐私信息。然而, 这种方法只能识别特定类别的隐私信息, 对于间接的隐私泄露无法提供有效保护。

2.2 基于 NLP 和 DL 的社交网络隐私研究

在过去的研究中, 研究人员对社交网络中个人隐私信息的保护展开了广泛研究, 其中 NLP 和 DL 技术是目前主流的研究方法。NLP 是实现社交网络隐私保护的重要技术, 通过 NLP 技术可以对文本信息进行脱敏处理, 去除或替换其中的敏感信息^[22], 同时还可以结合加密技术保护用户的隐私数据, 在传输和存储过程中进行加密处理。但是这些方法表现出高度的模糊性, 被替换或去除的信息可能导致原始信息的可读性较差。一些社交平台也使用基于 NLP 技术来分析用户的隐私数据和行为, 制定针对性的隐私保护策略, 防止用户隐私数据被滥用。虽然这些策略在时间和空间上阻止了个人共享内容在网络上的传播, 但它们对用户文本内容中的敏感信息识别和保护并不全面。DL 技术可以捕捉到词向量和文本向量之间的复杂关系, 提高模型的性能及可扩展性, 达到隐私保护和增强隐私意识的目的。许多研究人员利用各种模型和方法对隐私数据进行分析识别, Gu 等人^[23]利用图卷积神经网络(Graph Convolutional Network, GCN)提出了一种数据隐私保护模型, 通过半监督学习方法构建 GCN 分类模型, 推断出目标用户隐藏的隐私属性类别, 但预测准确率并不高, 均在 60% 的范围内。Tang 等人^[24]融合 BI-LSTM 模型和条件随机场(Conditional Random Field, CRF)识别临床医疗文本中的隐私实体, 该算法的最终 F1 值达到 91.82%。Zhang 等人^[25]通过研究超参数设置对中文命名实体识别(named entity recognition, NER)的影响, 加入随机失活机制进行辅助。通过不断改变参数设置使模型在 NER 任务中表现出更高的准确性和鲁棒性。Qiu 等人^[26]提出了一种弱监督的中文地名识别架构, 利用了训练数据集生成器和双向循环神经网络, 实现可扩展的地名识别体系结构。Mehdy 等人^[27]使用 NLP 技术对文本进行处理, 获取文本特征, 如语言标签、句法依赖和实体关系等特征。利用得到的文本特征训练卷积神经网络模型(Convolutional Neural Networks, CNN), 经过训练的模型可以识别文本是否存在隐私泄露风险。该方法通过特定技术获取文本特征, 进而训练预测模型, 并将该模型用于文本隐私感知。然而, 这些方法未充分考虑文本数据的语义特征, 在预测方面表现出较差的可解释性。用户使用这些方法时只知道存在隐私泄露的可能, 却无法

准确了解具体的隐私泄露内容。

3 方法

为了更加全面地保护用户在社交网络平台发表的文本隐私安全, 本文提出了一个社交网络用户自披露隐私感知模型。在介绍该模型之前, 本文提出的模型使用的各种参数和数学符号如表 1 所示。

表 1 计算中主要数学符号

符号	描述
$\text{Re}[\cdot]$	取实部函数
h	提取特征后的向量序列
W	表示全连接层的权重参数
a	模型或层的输入
$[h_i, h_j]$	跨度表示法表示实体从 i 开始到 j 结束
$\tilde{q}, \tilde{k}$	将绝对位置信息注入 q 和 k 中
$\vec{C}, \bar{C}$	前向和后向的细胞状态
λ	调整跨度的衰减速度

3.1 FPERM 模型

FPERM 模型是由改进的 RoformerBERT 预训练模型、BI-LSTM 模型和 EGP 算法组成的。改进的 RoformerBERT 预训练模型可以深入学习文本特征

和解决隐私实体分布不平衡的问题, 在提高训练速度的同时减少计算资源和内存^[28]。BI-LSTM 模型主要用于提取句子的序列特征。EGP 算法不仅解决了隐私实体的嵌套问题, 而且显著提升了模型的识别准确率。FPERM 模型整体图如图 1。首先, 输入的文本通过预处理、分词、嵌入操作得到相应数据格式。其次, 通过改进的 RoformerBERT 模型和 BI-LSTM 模型提取序列特征, 最后, 利用 EGP 算法通过打分对实体进行预测。其中参数 i 表示输入数据的长度, R_a 表示每个实体的共享参数。

隐私实体识别模型的具体操作步骤如下:

(1) 首先对数据集进行预处理操作, 通过分词操作将文本分割成单词或子单词。然后读取 Vocab.txt 文件(该文件包含了模型训练过程中所有单词对应索引编号的映射关系), 根据单词和对应的索引将文本中的单词转换为相应的向量表示。将词嵌入向量和位置嵌入向量通过旋转位置编码(Rotary Position Embedding, RoPE)重编码嵌入融合, 利用相对位置的偏移量生成包含跨度依赖信息的特征表达式, 从而得到符合改进的 RoformerBERT 预训练模型输入要求的文本数据表示。其中 RoPE 算法进行重新编码的目的是将绝对位置编码转换成相对位置编码, 从而增加数据的输入量。

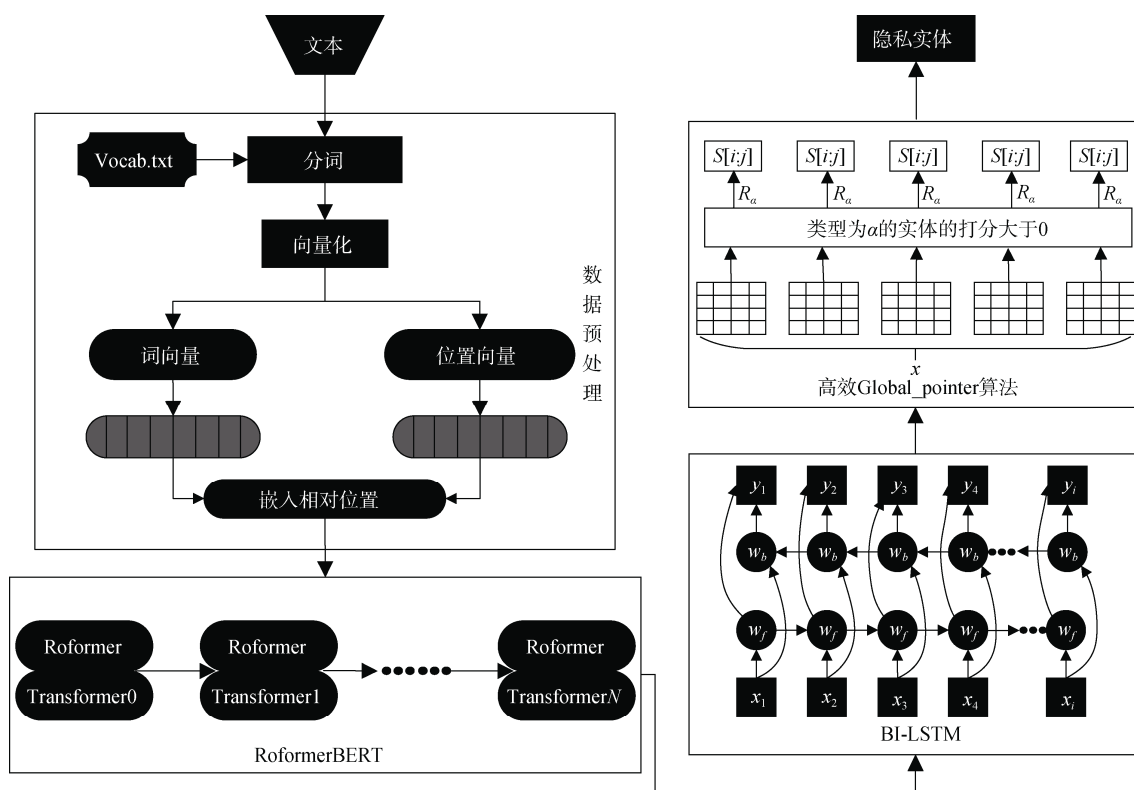


图 1 FPERM 模型结构

Figure 1 Structure of the FPERM model

(2) 利用改进的 RoformerBERT 模型对转换后的编码向量进行特征提取处理, 得到文本数据的特征表达式。通过改进的 RoformerBERT 模型特征提取和表征学习, 进行多维特征的增强和变换, 一个 $1 \times L$ 的向量会转化成一个 $348 \times L$ 矩阵, 可以更好地表示文本中的隐藏特征。其中 L 代表的是每个输入标记(Token)的向量维度。

(3) 将改进的 RoformerBERT 模型处理后的数据输入到 BI-LSTM 模型中进行序列处理, 从而捕捉数据的长期依赖关系和双向信息, 提高模型在处理序列数据时的性能。BI-LSTM 模型包含前向和后向两个独立的 LSTM 网络, 分别对输入序列的正序 w_f 和逆序 w_b 进行处理, 最终通过拼接双向输出向量获取融合上下文信息的序列特征。

(4) 将第三步中获得的数据传递到 EGP 算法中进行计算。EGP 将输入张量划分为 x 个矩阵输出, 其中 x 为实体类型数量。通过对每个矩阵进行计算, 最终确定输出的隐私实体类别。

3.2 改进 RoformerBERT

BERT 模型因其强大的上下文理解能力和双向建模特性, 在多语言和复杂任务上表现出色, 成为当前主流的隐私识别模型。BERT 模型采用多层 Transformer 编码器单元, 通过在大规模语料库上进行预训练, 使多层编码器能够学习到通用的文本表示。BERT 模型主要由嵌入层和多层 Transformer 变换层组成, 如图 2 所示。

在处理较长文本时, BERT 模型往往会面临编码溢出的问题, 且其对文本中位置变化的处理能力有限, 从而导致无法有效提取特征信息, 进而影响长文本隐私识别的准确性。为了解决这一问题, 并提高嵌套实体及长文本的识别精度, 本文在 BERT 模型的基础上, 融合了 SRPE 和注意力机制。SRPE 通过 RoPE 结合绝对位置和相对位置信息, 增强了模型对相对位置的建模能力, 同时结合跨度依赖信息, 提升了模型对序列中不同跨度依赖关系的感知能力, 从而增强了模型对全局特征的提取能力。SRPE 的具体步骤如下:

(1) 将绝对位置信息添加到 q 和 k 中, 其中 m 和 n 表示绝对位置信息, $f(\cdot)$ 是一个运算, 在 $f(\cdot)$ 运算后 q 和 k 将具有 m 和 n 的绝对位置信息。如公式(1)所示:

$$\tilde{q}_m = f(q, m) \quad \tilde{k}_n = f(k, n) \quad (1)$$

(2) 利用注意力机制中的内积计算和复数共轭计算, 将内积转化为只依赖于相对位置 $m-n$ ^[29] 的形

式, 巧妙地融合绝对位置信息和相对位置信息。同时, 引入与 $m-n$ 跨度相关的偏移量, 增强模型对序列中不同跨度依赖关系的感知能力。如公式(2)所示:

$$\begin{aligned} \langle \tilde{q}_m e^{im\theta}, \tilde{k}_n e^{in\theta} \rangle &= \text{Re} \left[\tilde{q}_m e^{im\theta} \tilde{k}_n e^{in\theta} \right] \\ &= \text{Re} \left[\tilde{q}_m \tilde{k}_n * e^{i(m-n)\theta} e^{ig(|m-n|)} \right] \end{aligned} \quad (2)$$

其中, $g(|m-n|) = \lambda \log(1 + |m-n|)$ $\text{Re}[\cdot]$ 表示考虑了结果的实数部分, g 是跨位置距离的权重调节函数, $e^{im\theta}$ 和 $e^{in\theta}$ 表示将虚部添加到 $\tilde{q}_m$ 和 $\tilde{k}_n$, $*$ 为共轭计算。

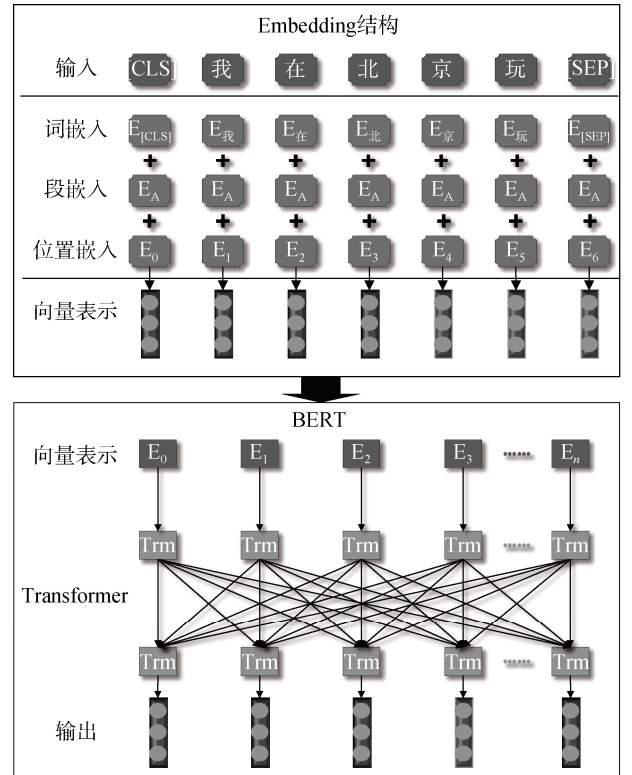


图 2 BERT 预训练模型

Figure 2 BERT pre-training model

(3) 为了更好地捕捉长序列中的结构性信息和依赖关系, 引入跨越不同距离的跨度注意力权重, 使模型能够感知标记之间的跨度信息。 $\odot$ 表示元素级的加权, G 是跨度权重矩阵, 如公式(3)所示:

$$\text{Attention}(Q_m, K_n, V) = \text{Softmax} \left(\frac{\text{Re}[Q_m, K_n]}{\sqrt{d_k}} \odot G \right) V \quad (3)$$

其中, $G = \frac{1}{1 + |m-n|^\lambda}$ 。

RoformerBERT 模型的每一层均包含 Bias 项, 这不仅增加了参数量, 还可能导致模型过度依赖 Bias 项, 影响泛化能力。因此, 提出的模型在结构上去除

了 Bias 项, 虽然 Bias 的参数数量较小, 但在大型模型中, 尤其是含有大量神经元的全连接层中, 去除 Bias 项可以有效降低计算成本, 简化模型结构, 从而使模型更易于分析和优化。此外, RoformerBERT 使用层归一化技术(Layer Normalization, Layer Norm), 在神经网络的每一层输出后, 对输出数据进行归一化处理, 通过将数据的均值调整为 0、方差为 1, 以稳定每一层输入数据的分布特征, 从而缓解内部协变量偏移问题。然而, 在处理复杂任务或需要模型适应不同输入分布的情况下, Layer Norm 会增加计算资源和时间开销。并且, 其严格的归一化方式可能会限制模型的学习效率, 导致模型的泛化能力较差。因此, 提出的模型将 Layer Norm 换成了更高效的均方根层归一化技术(Root Mean Square Layer Normalization, RMS Norm)^[30]。RMS Norm 是对 Layer Norm 的一种简化, 对每个时间步的特征进行归一化, 通过舍弃中心不变性来降低计算量, 在处理动态长度序列时更为灵活和有效, 有效提升训练速度的同时保持模型的泛化性。RMS Norm 计算公式如公式(4)所示:

$$\bar{a}_i = \frac{a_i}{\text{RMS}(a)} g_i, \text{ 其中 } \text{RMS}(a) = \sqrt{\frac{1}{n} \sum_{i=1}^n a_i^2} \quad (4)$$

其中, a_i 是模型或层的输入, n 为样本数量。

改进的 RoformerBERT 模型通过整合 SRPE 和注意力机制, 融合相对位置编码和绝对位置编码信息, 提高嵌套实体和长文本的识别准确率, 同时移除 Bias 项并采用 RMS Norm 层归一化技术, 有效提升训练速度和模型的泛化性。

3.3 BI-LSTM 模型

BERT 模型中的 Transformer 虽能通过自注意力机制有效捕捉输入序列的长距离依赖, 但在嵌套实体识别等需精细处理序列信息的任务中, 其特征提取能力仍显不足。因此, 采用 BERT 进行初步序列建模与特征提取, 再结合 BI-LSTM 模型进行上下文建模与细化处理。该组合方法能更精确地捕捉局部上下文信息及序列的动态变化, 从而提升模型性能。BI-LSTM 模型的结构如图 3 所示。

BI-LSTM 模型整体输出的计算公式如公式(5)所示:

$$\begin{aligned} \mathbf{wb}_i &= \overline{\text{LSTM}}(\mathbf{wb}_{i-1}, \mathbf{x}_i, \overline{\mathbf{C}_{i-1}}), i \in [1, T] \\ \mathbf{wf}_i &= \overline{\text{LSTM}}(\mathbf{wf}_{i-1}, \mathbf{x}_i, \overline{\mathbf{C}_{i-1}}), i \in [1, T] \\ \mathbf{w}_i &= \sigma(\mathbf{W}[\mathbf{wb}_i, \mathbf{wf}_i] + \mathbf{b}_n) \end{aligned}$$

其中, $\mathbf{wf}_i$ 是前向层的隐藏输出序列, $\mathbf{wb}_i$ 是后向层

的隐藏输出, $\mathbf{wf}_{i-1}$ 和 $\overline{\mathbf{C}_{i-1}}$ 表示前向 LSTM 层中第 $i-1$ 个输入点的隐藏状态和细胞状态, $\mathbf{wb}_{i-1}$ 和 $\overline{\mathbf{C}_{i-1}}$ 分别表示反向 LSTM 层中第 $i-1$ 个输入点的隐藏状态和细胞状态, $\mathbf{b}_n$ 为偏置项, σ 是激活函数。

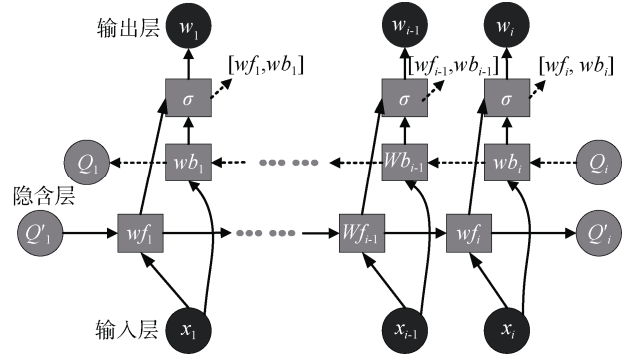


图 3 BI-LSTM 模型

Figure 3 BI-LSTM model

3.4 EGP 算法

目前大多数隐私实体识别任务中使用的 CRF 算法在序列标注和结构预测方面表现出显著优势。然而, CRF 算法复杂性高、计算资源需求大, 并且处理嵌套隐私方面表现较差。相比之下, 全局指针算法(Global Pointer, GP)采用全局归一化思想进行隐私实体识别, 在处理多标签序列时能够无差别地识别嵌套实体和非嵌套实体^[31]。GP 算法的打分函数如公式(6)所示:

$$\begin{aligned} \mathbf{q}_{i,\alpha} &= \mathbf{W}_{q,\alpha} \mathbf{h}_i, & \mathbf{k}_{i,\alpha} &= \mathbf{W}_{k,\alpha} \mathbf{h}_i \\ \text{GP} &= \text{Re}[\mathbf{q}_{i,\alpha}^T \mathbf{k}_{j,\alpha} e^{(m-n)\theta}] \end{aligned} \quad (6)$$

其中, $\mathbf{q}_{i,\alpha}$ 表示 Query 矩阵, $\mathbf{k}_{i,\alpha}$ 表示 Key 矩阵, $\mathbf{h}_i$ 表示通过 BI-LSTM 模型提取特征后的向量序列, $\mathbf{W}$ 表示全连接层的权重参数。

由于无差别识别可能会产生大量的实体, 且 GP 算法对实体的长度和跨度敏感性较低, 常常会错误地将任意两个实体的首尾组合识别为目标实体。针对这一问题, 本文提出了一种改进的 EGP 算法, 该算法通过使用 SRPE 嵌入相对位置信息和跨度依赖信息, 使得 GP 算法对实体的长度和跨度更加敏感, 从而能够更准确地区分真实实体和嵌套实体。此外, 为了提高参数的利用率, 使用了基于实体类型共享得分计算的方法, 将 NER 任务有效地分解为实体提取和实体分类两个子任务。在提取步骤中, 将文本片段提取为实体, 而在分类步骤中识别每个实体的具体类型。这种分解方法可将提取过程等同于仅包含一个实体类型的 NER 任务, 可以用评分矩阵

$(W_q h_i)^T (W_k h_j)$ 完成它。分类步骤可以理解为 $w_\alpha^T [h_i; h_j]$, 其中 w_α 是上下文相关的动态权重, 可以根据输入序列中上下文的变化动态调整权重, 捕获不同实体类型与上下文之间的依赖关系, 由 h_i 和 h_j 上下文特征生成, 新的评分函数如公式(7)所示:

$$s_\alpha(i, j) = (W_q h_i)^T (W_k h_j) + w_\alpha^T [h_i; h_j] \quad (7)$$

其中 $w_\alpha = f(h_i; h_j) = W_2 \cdot \sigma(W_1 [h_i; h_j] + b_1) + b_2$

其中 W_1 和 W_2 是权重矩阵, b_1 和 b_2 是偏置项。为了进一步减少参数, 使用 $[q_i; k_i]$ 替换 h_i 来代表 Token。最终的评分函数为公式(8)所示:

$$s_\alpha(i, j) = q_i^T k_j + w_\alpha(q_i; k_i; q_j; k_j) \quad (8)$$

其中 $w_\alpha \in \mathbb{R}^{4d}$ 。直观上看, 每个新实体类型增加的参数数量是 $4d$, d 为词嵌入的维度, 其中 $d = 64$ 。有效减少算法参数数量并提升训练速度。

EGP 算法采用全局归一化思想进行隐私实体识别, 针对每种实体类型共享得分计算, 大幅减少参数数量的同时有效提升嵌套实体和非嵌套实体的识别准确率。

3.5 损失函数

由于私有实体识别数据集中存在过多的非实体数据, 导致数据集中存在长尾现象。在本文中, 损失计算方法通过改进单目标多分类交叉熵损失函数, 使得私有标签类的得分高于非私有标签类^[32]。对于多标签分类任务, 目标是使每个目标类别的得分不低于非目标类别的得分。 P 是标签类, N 是非标签类, s 表示每个类的得分。损失值采用交叉熵函数计算, 如公式(9)所示:

$$\log \left(1 + \sum_{i \in N, j \in P} e^{s_i - s_j} \right) = \log \left(1 + \sum_{i \in N} e^{s_i} \sum_{j \in P} e^{-s_j} \right) \quad (9)$$

对于类别不固定的多标签分类, 需要一个阈值来确定输出哪些类。因此引入一个额外的 0 类, 希望目标类的分数都大于 s_0 , 非目标类的分数都小于 s_0 , 所以往 $\log$ 里边加入 $e^{s_i - s_0}$ 得到公式(10):

$$\begin{aligned} \log(1 + \sum_{i \in N, j \in P} e^{s_i - s_j} + \sum_{i \in N} e^{s_i - s_0} + \sum_{j \in P} e^{s_0 - s_j}) \\ = \log(e^{s_0} + \sum_{i \in N} e^{s_i}) + \log(e^{-s_0} + \sum_{j \in P} e^{-s_j}) \end{aligned} \quad (10)$$

当指定阈值为 0, 最后简化的 Loss 形式如下公式(11)所示:

$$\text{Loss} = \log \left(1 + \sum_{i \in N} e^{s_i} \right) + \log \left(1 + \sum_{j \in P} e^{-s_j} \right) \quad (11)$$

4 实验评价

4.1 隐私实体识别数据集

深度学习模型涉及大量训练数据, 在众多社交网络中, 新浪微博是中国最流行、使用最广泛的社交网络平台, 日均活跃用户达 2.11 亿人。实验采用了 Liu 等人^[33]基于新浪微博语料库构建的隐私数据集, 并通过扩充其中识别率低的样本, 构建了一个增强数据集, 用于训练 FPERM 模型。隐私数据集的收集标准如公式(12)所示。一个数据块中只要存在一个能标识为自然人的实体且存在位置、工作、教育和公司等信息的语句, 就将该句子标注为含有隐私信息的语句。本数据集通过对新浪微博数据集中的隐私数据进行提取, 最终获得 7203 条具有隐私实体的数据集。其中, 包含隐私实体 13415 个。在本文中, 隐私实体主要涵盖个人身份信息(如姓名、生日、身高、体重等)、住址(包括当前位置、工作地点、家庭地址等)、职业、教育背景和公司等方面的内容。具体分布如表 2 所示。由于一个数据块中可能涉及多个隐私实体, 所以各项隐私实体数之和大于总体的数据集条数。其中, 包括地址(LOC)有 2103 个隐私数据项, 个人信息(BI)有 1492 个隐私数据项, 职业(JOB)有 3032 个隐私数据项, 教育背景(EDU)有 2786 个隐私数据项, 公司(COM)有 4002 个隐私数据项。

$$(\exists \text{Person} \vee \exists \text{BI}) \wedge (\exists \text{LOC} \vee \exists \text{JOB} \vee \exists \text{EDU} \vee \exists \text{COM}) \rightarrow \text{隐私信息}$$

(12)

表 2 隐私实体类型的数量分布

Table 2 Distribution of the number of privacy entity types

实体类型	数量
BI	1492
JOB	3032
EDU	2786
COM	4002
LOC	2103
ALL	13415

4.2 评测指标

为了从多个方面验证该方法的效果和性能, 本文选取了四个重要的指标: 准确率、召回率、精确率和 F1 来评估隐私实体识别任务。各个指标的解释如下所示。

- Precision: 是指有多少正样本是真正样本。
- Recall: 表示在所有样本中, 有多少正例样本

被预测正确。

- Accuracy: 预测正确的结果占总样本的百分比。
- F1-score: F1-score 是一个综合指标, 是精度的调和平均值。

四个指标的计算方法如下列公式所示:

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

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

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (15)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

其中, TP 表示预测为正样本的正样本数量, TN 表示预测为负样本的负样本数量, FN 表示预测为负样本的正样本数量, FP 表示预测为正样本的实际负样本数量。

4.3 超参数选择

FPERM 模型采用 BERT+BI-LSTM+GP 的结构, 实验中的超参数选择如表 3 所示。主要的超参数有训练批量大小、循环次数和学习率, 其中训练批量大小和循环次数是影响模型训练速度的关键参数, 本文设置的训练批量大小和循环数分别为 16 和 35。学习率对模型效果起决定性作用。本文选取 AdaFactor 作为优化器, 结合自适应学习率和 Adam 算法, 保持学习率在训练过程中的稳定性, 有助于加速收敛和提高模型的性能。在 AdaFactor 算法中, 通过初始学习率初始化更新学习率大小。合适的初始学习率可以帮助模型快速收敛到较优解, 加快训练速度。

表 3 参数设置

超参数	参数值
max_seq_length	512
learning_rate	5e-4
hidden_size	768
batch_size	16
bi_lstm_units	64
optimizer	AdaFactor
epoch	35
min_dim_size_to_factor	1000000
beta1	0.9

设置合适的初始学习率可以避免学习率过大导致参数更新不稳定, 或者过小导致收敛速度缓慢的问题。本文对学习率进行了筛选从 $5e-1$ 降低到 $5e-10$, 每次降低了 0.1 的数量级。实验结果表明在训练过程中, 当学习率大于 $5e-3$, 小于 $5e-5$ 时, 在整个模型训练过程中都会发生梯度爆炸导致参数更新时出现异常情况。因此, 本文在 $5e-3$ 和 $5e-5$ 之间进行实验测试, 不同学习率模型训练的对应损失值变化和 F1 值变化结果如图 4 和图 5 所示。当学习速率为 $5e-4$ 时, 训练模型的损失值最小且 F1 值的最优值可达 96.15%。而当学习速率为 $5e-3$ 时, F1 值为 94.77%。当学习速率为 $5e-5$ 时, 训练模型的最优 F1 值为 82.76%。因此, 本实验的模型学习率为 $5e-4$ 。

4.4 模型时间复杂度

为了进一步说明 FPERM 模型的轻量化, 本文对 FPERM 模型与 RoformerBERT 基线模型的时间复杂度进行了对比分析。

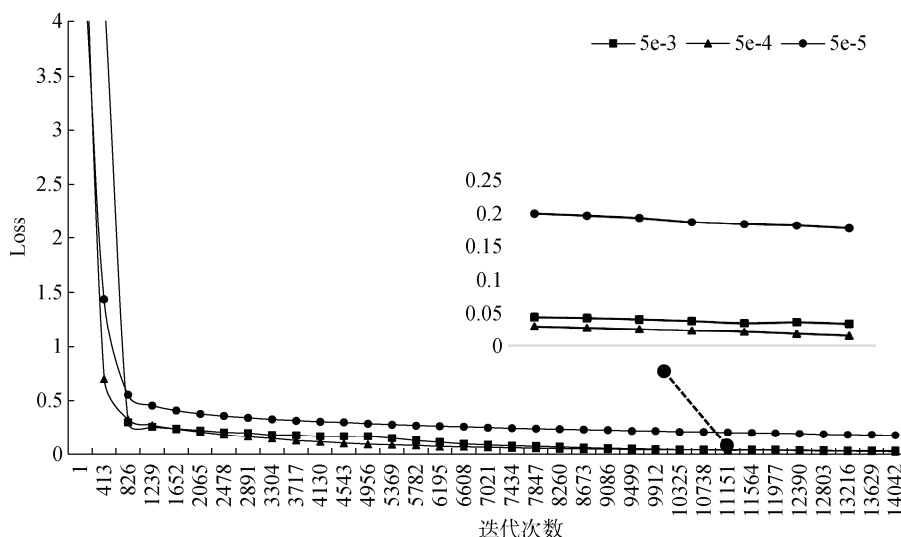


图 4 模型在训练过程中各种学习率所对应的损失值变化

Figure 4 Variation of loss values corresponding to various learning rates during training of the model

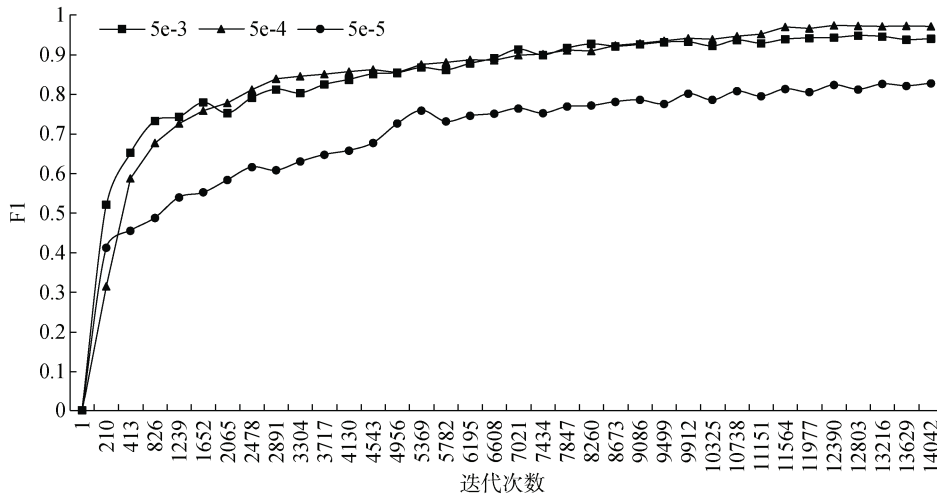


图 5 模型在训练过程中各种学习率所对应的 F1 值变化

Figure 5 Changes in F1 Scores corresponding to various learning rates during training of the model

ReformerBERT 模型由嵌入层、位置编码层、注意力层、密集层、加法层、归一化层和前馈全连接层组成。嵌入层的时间复杂度为 $O(E_{in} \times E_{out})$, 其中 E_{in} 表示嵌入层的输入, 由输入文本的长度 L 和词典维度 V 得到, E_{out} 表示嵌入层的输出, 由词典维度和模型输出的词向量维度 H 得到。位置编码层的时间复杂度为 $O(P_{in} \times P_{out})$, 其中 P_{in} 表示位置编码层的输入, P_{out} 表示位置编码层的输出。注意力层的时间复杂度为 $C \times O(A_{in} \times A_{out})$, 其中 C 表示注意力层中的层数, A_{in} 表示注意力层的输入, A_{out} 表示注意力层的输出。由于去除了 Bias 层, 全连接层的时间复杂度为 $O(M \times L \times H \times H - L \times H)$, 其中 M 表示全连接层的数量。由于使用 RMS Norm 进行归一化, 所以残差连接归一化层的时间复杂度为 $M \times [O(L \times H) + O(H \times 2 \times 2) - L \times d]$ 其中 d 为隐含层维度。前馈全连接层的时间复杂度为 $M \times O(L \times H \times I)$, 其中 I 表示前馈全连接层中神经元的数量。

FPERM 模型中 BI-LSTM 模型的时间复杂度计算如下。根据 LSTM 模型的原理, LSTM 模型的时间复杂度为 $O(H \times L + H \times H \times H)$, 其中 H 是隐藏层的数

量。本研究使用 BI-LSTM 模型, 它采用了双向 LSTM 模型, 因此模型复杂度为 $2 \times O(H \times L - H \times H \times H)$ 。

FPERM 模型中 EGP 算法的时间复杂度计算如下。在本研究中, GP 算法本质上是一个多头注意力层, 其时间复杂度与注意力层相似。所以 EGP 算法的时间复杂度为 $O(G_{in} \times G_{out} + n \times g)$, 其中 G_{in} 是 EGP 算法的输入, G_{out} 是 EGP 算法的输出, n 是实体的类别数量, g 为提取方法复杂度。

总体模型时间复杂度的对比结果如表 4 所示, 可以看出 FPERM 模型在时间复杂度上明显优于基线模型。

4.5 模型性能实验

本文构建了一个包含 608 条数据、1153 个隐私实体的测试集来评估 FPERM 模型。其中 LOC 隐私数据项有 456 个, BI 隐私数据项有 192 个, EDU 隐私数据项有 205 个, JOB 隐私数据项有 190 个, COM 隐私数据项有 110 个。通过准确率、F1 值、精确率和召回率四项指标对模型性能进行评估, 表 5 展示了模型的预测效果以及各项实验数据。从测试集的整体性能来看, 模型的总体预测准确率达到 94.01%, F1 值达到 96.15%, 精确率达到 98.32%, 召回率达到 94.06%。从各项隐私项识别结果来看, COM 和 BI 的

表 4 FPERM 模型与基线模型时间复杂度比较

Table 4 Time complexity comparison between FPERM model and baseline model

基线模型		FPERM
时间复杂度	i	$O(E_{in} \times E_{out}) + O(P_{in} \times P_{out}) + C \times O(A_{in} \times A_{out}) +$ $O(M \times L \times H \times H - L \times H) + M \times [O(L \times H) + O(H \times 2 \times 2) - L \times d] +$ $M \times O(L \times H \times I) + 2 \times O(H \times L - H \times H \times H) + O(G_{in} \times G_{out} + n \times g)$

表 5 FPERM 模型对隐私实体以及整体隐私实体识别的准确率及实验各项指标

Table 5 Accuracy and experimental metrics of FPERM model for recognizing private entities as well as overall private entities

隐私类型	LOC	JOB	EDU	COM	BI	ALL
预测正确数量	429	182	193	102	178	1084
隐私实体数量	456	190	205	110	192	1153
Accuracy	94.09	95.78	94.14	92.72	92.71	94.01
F1	94.31	95.97	94.18	95.23	91.96	96.15
Precision	97.54	97.68	96.75	96.84	93.74	98.32
Recall	91.28	94.31	91.75	93.67	90.25	94.06

隐私识别效果相较于其他四项隐私项的识别效果稍

差, 但总体预测准确率仍能达到92%以上, 能够有效感知隐私实体。

为了进一步验证提出模块对模型性能的影响, 本文在 BERT 模型的基础上进行了消融实验, 主要对比了加入不同模块模型的训练速度、模型参数和准确率, 消融实验结果如表 6 所示。可以看到加入 SRPE 模块后, 模型的精准率提升了1%, 训练速度和模型参数几乎不变。去除 Bias 项, 并加入 EGP 算法后, 模型的参数量有一定程度的减少, 精准率提升了 5%, 训练速度提升 30%。而同时加入 BI-LSTM 模型和 EGP 算法后, 精准率再次提升了 11%, 训练速度也提升了 20%左右。FPERM 模型在训练速度、模型参数和准确率都有了一定程度的提升。

表 6 FPERM 模型与 BERT 模型训练速度以及模型参数大小比较

Table 6 Comparison of FPERM model and BERT model training speed and model parameter size

	BERT	SRPE	Bias	BI-LSTM	EGP	模型参数	训练速度	Precision
BERT	√		√			101677056	1.0x	88.42
BERT+SRPE	√	√	√			102562944	1.0x	89.47
BERT+SRPE+EGP	√	√	×		√	95105792	1.3x	94.28
FPERM	√	√	×	√	√	94597376	1.2x	98.32

表 7 与其他模型的比较

Table 7 Comparison with other models

Dataset	BERT-CRF ^[34]	ALBERT-MogAtt_BI-LSTM-CRF ^[35]	GFNER ^[36]	Our model
Our dataset	F: 91.62/P: 91.11	F: 93.32/P: 95.53	F: 94.21/P: 96.25	F: 96.15/P: 98.32
CLUNER ^[37]	F: 78.14/P: 76.82	F: 81.42/P: 81.26	F: 82.64/P: 82.87	F: 84.74 /P: 84.63
IFLYTEK ^[38]	A: 60.88	A: 62.81	A: 62.54	A: 63.32

BERT 模型由多个嵌入层和注意力机制组成, 在处理长文本时会出现编码溢出问题, 无法提取有效的特征信息。FPERM 模型在 BERT 模型的基础上引入了 SRPE 位置编码、BI-LSTM 模型和 EGP 算法, 使得模型在提取特征过程中, 提取的特征更全面并加快网络训练速度。由此可见, FPERM 模型具有更好的特征提取能力及更强的泛化性。

4.6 对比实验

为了突出本文提出模型的优势, 将提出的模型与几种流行的 NER 模型进行了比较分析。在比较实验中, 研究使用了以下实验模型: BERT-CRF 模型^[34]、ALBERT-MogAtt_BI-LSTM-CRF 模型^[35]和 GFNER 模型^[36]。本文从 F1 值、精确率两个指标进行了对比实验, 对比实验结果如表 7 所示。从表 7 可以看出, 在隐私实体数据集中, 本文提出的模型的各项性能指标都高于 ALBERT-MogAtt_BI-LSTM- CRF 模型、BERT-CRF 模型和 GFNER 模型。从 F1 值可以看出,

本文提出的模型比 BERT-CRF 模型高出约 5%, 比 ALBERT-MogAtt_BI-LSTM-CRF 模型高出约 3%, 比 GFNER 模型高出约 2%。图 6 展示了模型中各项隐私实体的识别情况。从 F1 值来看, 本文提出的模型在各项隐私项上比 BERT-CRF 模型提升 2%~6%, 比 ALBERT-MogAtt_BI-LSTM-CRF 模型提升 1%~4%, 比 GFNER 模型提升 1%~4%。

同时在 CLUNER 细粒度命名实体识别数据集和 IFLYTEK 长文本数据集中^[37-38], 本文提出的模型在准确率和精确率方面均优于其他模型, 证明了其在隐私实体识别中的高效性和精准性。尽管现有方法在多个领域取得了良好的效果, 但由于隐私实体通常由复杂的结构组成, 并且包含大量嵌套实体, 传统的 CRF 方法在隐私实体的识别上表现不佳。为此, 本文提出的 FPERM 模型通过融合 SRPE 和注意力机制, 显著提高了对嵌套实体和长文本的识别准确性。在面对大量隐私实体类型的识别时, 结合 BI-LSTM

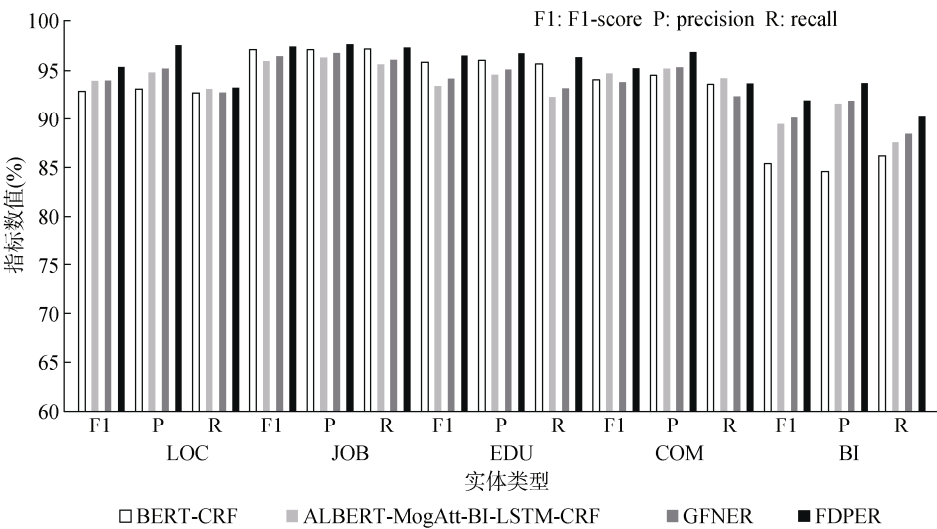


图 6 不同实体识别模型中每个隐私实体识别的精确率比较

Figure 6 Comparison of the accuracy rate of each privacy entity recognition in different entity recognition models

表 8 FPERM 模型对案例的预测结果
Table 8 FPERM model prediction results for cases

案例: 大三刚开始已经在海南绿盟公司实习一个月了		
	预测结果	输出结果
传统模型 ^[36]	ooooooooB_COMI_COMI_COMI_COMI_COMI_COMoooooooo	隐私类型: COM
本文提出的模型	Address: (9,10)	隐私类型: Address, COM
	COM: (9,14)	位置: 海南
		隐私类型: COM
		位置: 海南绿盟公司

模型和 EGP 算法不仅有效节省了计算资源和训练时间, 还能够提取复杂的全局特征, 从而增强了模型的特征提取能力。实验结果表明, 本文提出的隐私实体识别模型在实际应用中是可行且有效的。

4.7 案例分析结果

本文提出的 FPERM 模型解决了传统的社交网络自披露隐私识别模型仅能指出隐私泄露的类型, 无法提取嵌套隐私实体和无法准确定位具体的隐私泄露位置这一问题^[39-40], 实现对隐私信息的精准提取。表 8 展示了本文提出的模型与传统社交网络自我披露隐私识别模型之间的不同之处。如表中所示, 本文提出的模型让用户能够更清晰地了解哪些文字暴露了他们的隐私信息。以表中案例为例, 句子是: “大三刚开始已经在海南绿盟公司实习一个月了。” 该句子存在嵌套地址隐私, 即“海南”是一个地址, “海南绿盟公司”是公司实体。基于 CRF 的隐私实体识别模型只能识别出“海南绿盟公司”这个公司实体, 但无法将“海南”正确识别为一个地址实体。在表 8 中, B_COM 表示公司实体的开始, I_COM 是公司实体的中间和结束, o 不是实体。而本文提出的模型输

出了一个坐标 i , 表示数据中从 j 到 (9, 10) 包含隐私实体, 其中 i 表示实体的起始位置, j 表示结束位置。如表 8 所示, 坐标 (9, 10) 表示从句子中的第九个位置开始到第十个位置结束为地址实体。综上所述, 本文提出的模型可以自动感知社交网络中用户共享的隐私信息, 并准确定位文本泄露敏感信息的位置。

5 结论

针对传统隐私识别算法难以精确识别和定位隐私泄露信息的问题, 本文提出了一种基于高效深度神经网络的社交网络隐私自披露感知模型。FPERM 模型使用改进的 RoformerBERT 模型, 结合注意力机制和 SRPE 编码实现对社交网络隐私信息的自动感知, 并精准定位隐私信息的位置。在此基础上使用 BI-LSTM 模型对本文信息进行多次编码, 分层提取上下文特征并生成动态权重使模型更好地适应上下文变化。通过 EGP 算法, 将 NER 任务分解为提取和分类两个子任务, 有效提高参数利用率和模型的识别准确率。实验结果表明, FPERM 模型的总体准确率高达 94.01%, 各项隐私项的准确率均保持在 92%

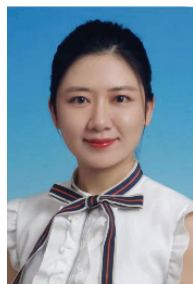
以上。FPERM 模型能够对用户的隐私泄露信息进行精准识别和定位, 并将这些隐私数据反馈给用户, 以增强他们的隐私意识, 有效保护隐私信息。

虽然本文提出的模型在直接隐私识别中取得了较好的结果, 但是在识别和推理间接隐私实体方面仍有不足。未来的研究将着重挖掘隐私实体之间的隐藏关系并进行度量, 让用户可以选择不同程度的隐私保护措施。

参考文献

- [1] Kim J, Hastak M. Social Network Analysis: Characteristics of Online Social Networks after a Disaster[J]. *International Journal of Information Management*, 2018, 38(1): 86-96.
- [2] Qiu M K, Gai K K, Xiong Z G. Privacy-Preserving Wireless Communications Using Bipartite Matching in Social Big Data[J]. *Future Generation Computer Systems*, 2018, 87: 772-781.
- [3] Xu L, Jiang C X, Chen Y, et al. A Framework for Categorizing and Applying Privacy-Preservation Techniques in Big Data Mining[J]. *Computer*, 2016, 49(2): 54-62.
- [4] Xu L, Jiang C X, He N Q, et al. Trust-Based Collaborative Privacy Management in Online Social Networks[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(1): 48-60.
- [5] Oukemeni S, Rifà-Pous H, Puig J M M. Privacy Analysis on Microblogging Online Social Networks: A Survey[J]. *ACM Computing Surveys*, 2020, 52(3): 1-36.
- [6] Liu K, Han Y L, Wang J J, et al. Location Privacy Protection Method Based on Geohash Coding and Pseudo-Random Sequence[C]. *2022 3rd Information Communication Technologies Conference*, 2022: 178-183.
- [7] Li Q X, Lopuhaä-Zwakenberg M, Heusdens R, et al. Two for the Price of One: Communication Efficient and Privacy-Preserving Distributed Average Consensus Using Quantization[C]. *2022 30th European Signal Processing Conference*, 2022: 2166-2170.
- [8] Zablit F, Osman I H. ReviewModus: Text Classification and Sentiment Prediction of Unstructured Reviews Using a Hybrid Combination of Machine Learning and Evaluation Models[J]. *Applied Mathematical Modelling*, 2019, 71: 569-583.
- [9] Adams S, Melanson D, De Cock M. Private Text Classification with Convolutional Neural Networks[C]. *The Third Workshop on Privacy in Natural Language Processing*, 2021: 53-58.
- [10] Srivastava A, Geethakumari G. Measuring Privacy Leaks in Online Social Networks[C]. *2013 International Conference on Advances in Computing, Communications and Informatics*, 2013: 2095-2100.
- [11] Liang K T, Liu J K, Lu R X, et al. Privacy Concerns for Photo Sharing in Online Social Networks[J]. *IEEE Internet Computing*, 2015, 19(2): 58-63.
- [12] Maximilien E M, Grandison T, Sun T, et al. Privacy-as-a-service: Models, algorithms, and results on the Facebook platform[C]. *Proceedings of Web*, 2009: 2.
- [13] Liu K, Terzi E. A Framework for Computing the Privacy Scores of Users in Online Social Networks[J]. *ACM Transactions on Knowledge Discovery from Data*, 2010, 5(1): 1-30.
- [14] Aghasian E, Garg S, Gao L X, et al. Scoring Users' Privacy Disclosure across Multiple Online Social Networks[J]. *IEEE Access*, 2017, 5: 13118-13130.
- [15] Pensa R G, Di Blasi G. A Privacy Self-Assessment Framework for Online Social Networks[J]. *Expert Systems with Applications*, 2017, 86: 18-31.
- [16] Alsarkal Y, Zhang N, Xu H. Your Privacy Is Your Friend's Privacy: Examining Interdependent Information Disclosure on Online Social Networks[J]. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2018: 1-10.
- [17] Serfontein R, Kruger H, Drevin L. Identifying Information Security Risks in a Social Network Using Self-Organising Maps[M]. *Information Security Education. Education in Proactive Information Security*. Cham: Springer International Publishing, 2019: 114-126.
- [18] Djoudi A, Pujolle G. Social Privacy Score through Vulnerability Contagion Process[C]. *2019 Fifth Conference on Mobile and Secure Services*, 2019: 1-6.
- [19] Li Y J, Bi R, Jiang N, et al. Methods and Challenges of Cryptography-Based Privacy-Protection Algorithms for Vehicular Networks[J]. *Electronics*, 2024, 13(12): 2372.
- [20] Yang H M, Huang Y F, Yu Y, et al. Privacy-Preserving Extraction of HOG Features Based on Integer Vector Homomorphic Encryption[M]. *Information Security Practice and Experience*. Cham: Springer International Publishing, 2017: 102-117.
- [21] Yu L J, Motipalli S M, Lee D, et al. My Friend Leaks My Privacy: Modeling and Analyzing Privacy in Social Networks[C]. *The 23rd ACM on Symposium on Access Control Models and Technologies*, 2018: 93-104.
- [22] Zhang F, Zhang Y Q, Zhang X H. Desensitization Method of Meteorological Data Based on Differential Privacy Protection[J]. *Journal of Cleaner Production*, 2023, 389: 136117.
- [23] Gu T, Yang L, Wang H. Data Privacy Protection Model Based on Graph Convolutional Neural Network[J]. *Mobile Networks and Applications*, 2024, 29(5): 1433-1440.
- [24] Tang B, Jiang D, Chen Q, et al. De-identification of clinical text via Bi-LSTM-CRF with neural language models[C]. *AMIA Annual Symposium Proceedings. American Medical Informatics Association*, 2019, 2019: 857.
- [25] Zhang T Z, Ma P P. Effects of Hyper-Parameters Setting in Bi-LSTM-CRF on Chinese Named Entity Recognition[C]. *Fourteenth International Conference on Graphics and Image Processing*, 2023: 109.
- [26] Qiu Q J, Xie Z, Wang S, et al. ChineseTR: A Weakly Supervised Toponym Recognition Architecture Based on Automatic Training Data Generator and Deep Neural Network[J]. *Transactions in GIS*, 2022, 26(3): 1256-1279.
- [27] Mehdy N, Kennington C, Mehrpouyan H. Privacy Disclosures Detection in Natural-Language Text through Linguistically-Motivated Artificial Neural Networks[M]. *Security and Privacy in New Computing Environments*. Cham: Springer International Publishing, 2019: 152-177.
- [28] Su J L, Ahmed M, Lu Y, et al. RoFormer: Enhanced Transformer with Rotary Position Embedding[J]. *Neurocomputing*, 2024, 568: 127063.
- [29] Shaw P, Uszkoreit J, Vaswani A. Self-Attention with Relative Po-

- sition Representations[EB/OL]. 2018: arXiv: 1803.02155. <https://arxiv.org/abs/1803.02155>.
- [30] Zhang B, Sennrich R. Root mean square layer normalization[J]. *Advances in Neural Information Processing Systems*, 2019: 32.
- [31] Su J L, Murtadha A, Pan S F, et al. Global Pointer: Novel Efficient Span-Based Approach for Named Entity Recognition[EB/OL]. 2022: arXiv: 2208.03054. <https://arxiv.org/abs/2208.03054>.
- [32] Sun Y F, Cheng C M, Zhang Y H, et al. Circle Loss: A Unified Perspective of Pair Similarity Optimization[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 6397-6406.
- [33] Liu G, Sun X T, Li Y R, et al. An Automatic Privacy-Aware Framework for Text Data in Online Social Network Based on a Multi-Deep Learning Model[J]. *International Journal of Intelligent Systems*, 2023, 2023: 1727285.
- [34] Phan H T, Nguyen N T, Tran V C, et al. An Approach for a Decision-Making Support System Based on Measuring the User Satisfaction Level on Twitter[J]. *Information Sciences*, 2021, 561: 243-273.
- [35] Fan S Y, Yu H, Cai X Y, et al. Multi-Attention Deep Neural Network Fusing Character and Word Embedding for Clinical and Biomedical Concept Extraction[J]. *Information Sciences*, 2022, 608: 778-793.
- [36] Chen J Y, Chen X, Pan S, et al. GFNER: A Unified Global Feature-Aware Framework for Flat and Nested Named Entity Recognition[J]. *IEEE Access*, 2023, 11: 55139-55148.
- [37] Xu L, Tong Y, Dong Q Q, et al. CLUENER2020: Fine-Grained Named Entity Recognition Dataset and Benchmark for Chinese[EB/OL]. 2020: arXiv: 2001.04351. <https://arxiv.org/abs/2001.04351>.
- [38] Xu L, Hu H, Zhang X W, et al. CLUE: A Chinese Language Understanding Evaluation Benchmark[EB/OL]. 2020: arXiv: 2004.05986. <https://arxiv.org/abs/2004.05986>.
- [39] Xu G S, Qi C H, Yu H, et al. Detecting Sensitive Information of Unstructured Text Using Convolutional Neural Network[C]. *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2019: 474-479.
- [40] Kutbi M. Named Entity Recognition Utilized to Enhance Text Classification while Preserving Privacy[J]. *IEEE Access*, 2023, 11: 117576-117581.



郭祯 于西安电子科技大学获得博士学位, 现任海南大学网络空间安全学院教授、副院长、博士和硕士研究生导师。CCF 会员。主要研究方向为数据安全、隐私保护、物联网安全等。Email: guozhen@hainanu.edu.cn



范晨阳 于宜春学院获得学士学位, 现于海南大学攻读硕士学位, 主要研究方向为区块链、隐私保护、深度神经网络等。Email: Fcy0912@hainanu.edu.cn



徐嘉 于重庆理工大学获得硕士学位, 现于海南大学攻读博士学位, 主要研究方向为医学图像处理、数字水印、深度神经网络和计算机视觉等。Email: jiaxu@hainanu.edu.cn



刘千 于云南大学获得硕士学位, 现于海南大学攻读博士学位, 主要研究方向为隐私计算、基于深度学习的隐私度量技术等。Email: 20083900110001@hainanu.edu.cn