

“最小必要”原则视角检测移动应用程序的隐私合规性

余佩厚, 徐天辰, 孙雯倩, 陈云芳, 于乐, 张伟

(南京邮电大学计算机学院, 江苏 南京 210023)

摘要: 为了满足法律法规对个人数据隐私保护的要求, 移动应用程序开发商常以隐私政策的形式向用户披露其对用户个人数据的收集行为。如何精准理解这些冗长且复杂的法律文件, 研究人员提出了一系列基于自然语言处理 (natural language processing, NLP) 的方法分析隐私政策文本并进行合规性检测。然而, 目前的研究大多针对的是“透明、公开、合法”等原则, 对“最小必要”原则进行检测的研究尚存在空白。为此, 提出了一种基于数据收集的“最小必要”原则视角对应用程序进行自动化合规性检测方法 MNPD (minimum necessary principle detection)。通过多标签文本分类模型对待测 App 进行服务类型上的分类, 确定不同类别 App 的“最小必要”信息范围; 构造提示词指导大语言模型从隐私政策文本中提取出该 App 在基本业务功能模式下的数据收集行为, 转化为隐私声明三元组并进行规范化; 最后, 通过合规性检测模型对待测 App 进行文本表述上的一致性检测以及“最小必要”原则检测。实验结果表明, 所提方法对华为应用市场获取的 101 个“在线影音类”App 自动化分析结果达到了 86.20% 的 F1 分数。

关键词: App; 隐私政策; 大语言模型; 最小必要原则

中图分类号: TP309.2

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024084

Detecting privacy compliance of mobile applications from the perspective of the "minimum necessary" principle

YU Peihou, XU Tianchen, SUN Wenqian, CHEN Yunfang, YU Le, ZHANG Wei

School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: To comply with legal requirements for personal data privacy protection, mobile App developers typically disclose their data collection practices to users through privacy policies. Researchers have proposed various methods using natural language processing (NLP) techniques to analyze privacy policy texts and perform compliance checks. However, most existing studies focus on principles like transparency, openness, and legality, leaving a gap in the evaluation of the 'minimum necessary' principle. For this purpose, a framework called MNPD (minimum

收稿日期: 2024-05-23; 修回日期: 2024-10-10

通信作者: 张伟, zhangw@njupt.edu.cn

基金项目: 国家自然科学基金 (62202406)

Foundation Item: The National Natural Science Foundation of China(62202406)

引用格式: 余佩厚, 徐天辰, 孙雯倩, 等. "最小必要"原则视角检测移动应用程序的隐私合规性[J]. 网络与信息安全学报, 2024, 10(6): 109-122.

Citation Format: YU P H, XU T C, SUN W Q, et al. Detecting privacy compliance of mobile applications from the perspective of the "minimum necessary" principle[J]. Chinese Journal of Network and Information Security, 2024, 10(6): 109-122.

necessary principle detection) was proposed for automated compliance checking of applications from the perspective of the ‘minimum necessary’ principle. Initially, a multi-label text classification model categorized the target App based on its service type to determine the range of ‘minimum necessary information’ for different App categories. Then, prompt words were constructed to guide the large language model in extracting data collection practices of the App under its basic business functionality mode, transforming them into privacy statement triples and standardizing them. Finally, the compliance checking model conducted consistency checks on the text representation of the target App and evaluated its adherence to the ‘minimum necessary’ principle. The experimental results show that the proposed method achieves 86.20% F1 score in the automated analysis of 101 ‘Online Audio-Visual’ Apps obtained from Huawei’s application market.

Keywords: App, privacy policy, large language model, minimum necessary principle

0 引言

截至 2023 年 9 月, 中国移动应用市场上监测到的活跃 App 数量为 261 万款 (包括安卓和苹果商店), 安卓应用商店在架应用累计下载量 542 亿次, 为全球第一大市场^[1]。为了向用户提供更加精准的个性化服务, 移动应用程序往往会收集各种类型的个人数据。随着隐私泄露事件的频发和用户隐私意识的不断增强, 人们越来越关注应用程序的个人数据收集行为。为了规范个人数据的收集, 全球多个国家或国际组织实施了隐私保护法律。欧盟于 2018 年实施了《通用数据保护条例》^[2] (general data protection regulation, GDPR)。美国 2020 年实施了《加州消费者隐私法》^[3] (California consumer privacy act, CCPA)。2021 年实施了《中华人民共和国个人信息保护法》(Personal Information Protection Law, PIPL)^[4]。这些隐私保护法律的一个共同特性就是要求移动应用程序必须清楚地向用户披露其任何有关个人数据的收集行为, 这些信息通常以隐私政策的形式告知用户。

无论是国内的 PIPL, 还是 GDPR, 都要求数据处理行为遵循“公平、透明和合法”“最小必要”等原则。然而, 目前的研究大多聚焦对数据收集行为的“公平、透明和合法”原则进行检测, 即检测移动应用程序是否在隐私政策中向用户明确披露了任何有关个人数据的收集行为以及该应用程序实际执行的数据收集行为是否与其隐私政策中声明的数据收集行为一致。从目前了解的情况来看, 尚未有针对“最小必要”原则的 App 合规性研究。“最小必要”原则指的是移动应用程序收集的用户个人数据应与处理目的直接

相关, 并应限制在为实现特定目的所必不可少的范围内。

假设给定一个移动应用程序, 它的主要功能是视频剪辑, 其隐私政策中声明该应用程序在使用过程中会收集用户的地理位置信息用于个性化广告服务。而对于用户来说, 个性化广告服务不是他们使用该应用程序的主要目的, 并且地理位置信息对于视频剪辑功能来说显然不是必要条件, 因此该程序出现了明显的过度收集用户数据行为, 不符合“最小必要”原则的要求。但是以往的研究只是针对数据收集行为是否公开、透明进行检测, 如果隐私政策中不存在表述上的逻辑矛盾并且与实际行为相符合, 那么即使出现了明显的过度收集行为, 也不会判定该应用程序违规。

分析一个移动应用程序隐私政策中声明的数据收集行为是否遵循“最小必要”原则, 需要法律专家综合考虑应用功能、隐私政策、合规要求等多个方面, 这需要投入大量的人力成本并且很难控制分析结果的一致性。因此, 有必要设计一种自动化分析方法对应用程序进行初步筛选, 再将分析结果提供给法律专家进行专业审查。“最小必要”原则的自动化检测存在以下两个难点。

1) 如何界定数据收集的“最小必要”信息范围。目前大多数隐私法律、法规都没有明确定义 App 如何收集用户数据才能满足“最小必要”原则, 而是授权开发人员可以实现自己的业务功能收集他们认为“必要”的信息, 导致在“最小必要”原则检测过程中不同 App 所必须收集的个人信息数据的范围模糊。

2) 如何从隐私政策文本中提取出应用程序在主要业务场景下对用户数据的收集行为信息。隐私政策文本中的语句结构多样且复杂, 包含复杂语句和交叉引用。一方面难以区分隐私政策中声明的哪些数据收集行为是服务于该应用程序的主要业务; 另一方面难以从自然语言编写的隐私政策文本中提取出准确的隐私声明信息用于合规性检测。

2022年, 国家市场监督管理总局、国家标准化管理委员会正式发布《信息安全技术 移动互联网应用程序(App) 收集个人信息基本要求(GB/T 41391-2022)》(以下简称《基本要求》)[5], 该标准明确规定了39种常见服务类型App必要个人信息范围和使用要求, 这为“最小必要”原则的检测工作提供了官方依据。

针对当前研究的不足, 本文提出了一种从数据收集的“最小必要”原则视角检测移动应用程序的隐私合规性方法。本文工作主要贡献如下。

1) 提出了一种从数据收集的“最小必要”原则视角对应用程序进行自动化合规性检测的框架MNPD, 对应用程序的隐私政策文本进行文本表述上的一致性检测, 并基于对等比较思想对应用程序是否过度收集个人数据进行检测。

2) 由于隐私政策文本中的语句结构多样且复杂, 传统的基于依存句法树[6]或者语义角色标注(semantic role labelling, SRL)[7]等自然语言处理(natural language processing, NLP)技术的方法需要消耗大量人力并且很难处理复杂句和交叉引用, 本文基于大语言模型对隐私政策文本进行分析, 能够高效地提取出隐私政策中对用户个人数据的收集行为信息。

3) 为了验证本文方法的有效性, 基于华为应用市场上101个“在线影音类”App的人工标注结果进行了全面的实验验证和分析。实验结果表明, 本文提出的合规性检测方法能够达到86.20%的F1分数。

1 背景

1.1 隐私政策

隐私政策指网络服务提供者就其如何处理用户个人信息所作出的书面声明, 用以告知用户个人信息如何被收集、使用以及与第三方共享的情

况。Amos等[8]通过分析一个包含了数百万个隐私政策的数据集指出, 隐私政策已经变得更加冗长, 而且内容复杂、难以阅读, 造成了用户在没有完全理解隐私政策的情况下, 往往倾向于点击“同意”的现状。而且即使用户仔细阅读了数千字的隐私政策, 也很难判断隐私政策所声明的数据收集行为是否与应用程序的实际行为一致。此外, 一些应用市场监管审核不严, 开发者可能会由于忽视或者是特殊目的在隐私政策中声明较少的数据收集行为, 但在实际行为中违规收集未被声明覆盖的用户个人信息。因此, 研究人员开发了各种工具来帮助用户理解隐私政策中的个人数据收集情况[9-10]。

1.2 大语言模型及提示工程

大语言模型(large language model, LLM)是一类具有巨大参数数量的自然语言处理模型, 可以学习和预测自然语言文本的规律和模式。这种模型使用大量文本数据进行训练, 因此可以很好地理解和生成人类语言。大语言模型(如GPT-3[11]、PaLM[12]和GPT-4[13]等)具有很强的泛化能力和学习能力, 被广泛应用在语音识别、机器翻译、自动问答、聊天机器人等领域。

虽然大语言模型具有极高的泛化能力, 但其输出往往缺乏确定性和可控性。为了解决这个问题, 研究人员开始尝试通过提供特定的提示信息来引导模型的输出, 即提示工程。具体来说, 提示工程是指系统地设计和优化输入提示, 以指导大语言模型的响应, 确保所生成输出的准确性、相关性和一致性。这个过程对充分发挥大语言模型的潜力, 使它们更易于访问和适用于不同的领域至关重要。

1.3 App服务类型

App服务类型是指移动互联网应用程序提供的业务功能分类, 实现用户主要使用目的的业务功能所属的服务类型即为App的服务类型。基本要求将常见App的服务类型分为39类, 包括地图导航类、网络约车类以及即时通信类等。App所属服务类型不同, 所必需的个人信息类型也有所不同。目前应用市场上对App类型的分类粒度较粗, 并且不同的应用市场有自己独立的分类体系和分类标准, 对App类型的判断标准不统一。

这种情况导致普通用户难以根据应用市场的分类结果判断所使用的 App 的服务类型, 从而影响了用户对该 App 的使用期望。

1.4 基本业务功能和最小必要信息

App 的功能可以划分为基本业务功能和扩展业务功能。基本业务功能是指实现用户主要使用目的的业务。扩展业务功能则是指基本业务功能之外的功能, 包括仅为实现改善服务质量、提升使用体验、定向推送信息以及研发新产品目的的业务功能。通常用户使用一个 App 的主要目的是使用该 App 的基本业务功能来满足自己的需求, 而该 App 的扩展业务功能不是用户的主要关注点。移动应用程序为了保障基本业务功能的实现所收集的用户数据才被称为“最小必要”信息。

2 相关工作

2.1 隐私政策文本分析

目前移动应用程序的隐私政策文本分析主要集中在英文隐私政策方面, 研究人员通常通过机器学习 (machine learning, ML) 技术和 NLP 技术从隐私政策中提取有关个人数据收集行为的声明用于合规性检测, 通常包括数据接收者、数据类型以及数据使用目的等信息。在以往的研究中, 研究人员将从隐私政策中提取信息表征为文本分类^[14-15]、文本对齐^[16]等, 尽管这些方法能够有效地识别隐私政策文档中与隐私实践相关的语句或片段, 但缺乏提取细粒度结构化信息的能力。Ahmad 提出了 PolicyIE^[9], 这是一个包含 31 个隐私策略的英语语料库, 涵盖了 5 250 个意图类别和 11 788 个有关具体隐私数据收集行为的文本段, 可用于训练提取细粒度个人数据的模型。PolicyLint^[6]基于依存句法树从隐私政策的句子中提取隐私声明, 并使用特定领域的命名实体识别 (named entity recognition, NER) 模型提取语句中的实体和数据类型, 同时构造个人数据本体, 从而从语句中提取出表示个人数据收集行为的三元组信息 (包括数据接收者、数据类型和是/否收集)。Bui 等^[7]使用 SRL 技术分析隐私政策中的每个语句, 提取以数据收集、使用、分享动词为谓词的语句, 并分析其语义参数, 将语义参数映射到表示数据收集目的的五元组中。但是, 由于

该方法使用的 SRL 模型是在通用数据集上进行训练的, 无法完全适用于隐私政策分析领域, 使得模型的 F1 分数和召回率偏低。尽管在隐私政策的文本分析方面已经开展了大量的研究工作, 但是现有的 NLP 技术依然难以实现高精确率和召回率的隐私政策分析, 提取其中的隐私声明信息; 基于少量自定义规则对大量具有不同表述形式的隐私政策语句进行匹配很难做到全覆盖, 从而导致理解偏差。此外, 这些半自动化的方法需要专业人员的大量人力投入。

2.2 数据收集的必要性

确定数据收集合理范围的一种常用方法是通过与同类应用程序进行比较来分析应用程序中权限请求和隐私披露的必要性。Peddinti 等^[17]设计了一种算法机制来减少移动应用中的权限请求。Jana 等^[18]发现 Google Play 商店中某些 Chrome 扩展和应用程序存在最小权限侵犯。Zhou 等^[19]从对等比较的角度自动检查应用程序隐私政策中的数据收集行为是否过于宽泛, 同时通过分析隐私政策中的语句, 确定过度收集的原因。Zhou 等的基本观点是: 如果一个应用程序在其隐私政策中声称收集的个人信息比大多数同类应用程序要多得多, 那么它就更有可能会进行过度收集。但是应用程序的功能复杂多样, 开发者为了丰富自己的应用程序功能, 在基本功能之外开发附加功能, 导致会比同类型应用程序收集更多的信息, 这种做法在对用户进行了告知的情况下实际上并不违规。

2.3 提示工程与隐私合规

提示工程包含一系列技术, 例如角色提示^[20]、少样本提示^[21]等基本方法, 以及思维链 (chain of thought, CoT) 提示^[22]等更复杂的方法。Tang 等^[23]提出了一个基于大语言模型的隐私政策文本分析框架 PolicyGPT, 对两个隐私政策领域的数据集^[24-25]进行了测试分析, 并运用了一种称为“前缀提示”的提示词构造方法, 以增强模型理解隐私定义和分类复杂性的能力。PolicyGPT 在两个数据集集中的表现均超过了基线机器学习和神经网络模型, 证实了可以利用大语言模型中固有的高级语言理解能力来完成隐私政策文本处理任务的潜力。

3 系统设计

3.1 总体流程

本文提出一种基于大语言模型从数据收集“最小必要”原则视角检测App隐私合规性的方法——MNPDP (minimum necessary principle detection)。MNPDP的整体框架如图1所示, 包含3个部分, 分别是App服务类型分类、隐私声明提取和合规性检测。总体流程为: 首先, 从应用市场中爬取App相关信息(包括简介/描述信息和隐私政策文本等), 使用基于预训练模型的多标签文本分类方法, 根据App的描述信息对其进行服务类型分类。然后, 依据所属服务类型, 分析App隐私政策中有关基本业务功能的描述段落, 利用大模型提取这些段落中有关用户隐私数据的收集行为信息, 转化为三元组的形式并进行规范化。最后, 通过对隐私政策进行文本表述上的一致性检测并比对该App为实现基本业务功能所收集的信息是否符合“最小必要”信息的范围来判断其合规性。

3.2 App服务类型分类

目前移动应用市场中的部分App会提供多种基本业务, 包含多个服务类型。例如, 在腾讯应用宝上3.5亿次下载量的“去哪儿旅行”App, 同时提供交通票务购买、酒店预订等多种服务, 实际上这都属于“去哪儿旅行”的基本业务功能。用户对一个App的主要业务功能的判断往往基于该App的简介或者描述信息。基于此现象,

本文基于App的描述信息, 并采用多标签文本分类算法对App进行服务类型上的分类。

多标签文本分类是自然语言处理中的一种任务, 其目标是为给定的文本文档分配多个标签。与传统的单标签分类任务不同, 多标签文本分类要求模型能够处理文本中同时存在多个主题、类别或属性的情况。具体来说, 给定一个标签集合 $L = \{l_1, l_2, l_3, \dots, l_n\}$, 其中 l_i 表示一个标签, n 表示标签的个数。对于一个文本集合 $T = \{t_1, t_2, t_3, \dots, t_i\}$ 中的每一个文本 t_i , 分类结果 $L = \{l_a, l_b, \dots, l_c\} \subseteq L$ 。

随着预训练模型的快速发展以及在下流任务中的优越表现, 不同于依赖行业专家进行特征工程的传统机器学习方法以及需要大量人力进行样本标注工作的深度学习方法, 预训练+微调的方法通过在大规模数据上进行训练, 能够学习到更丰富、更深层次的语义信息, 在相对较小的目标任务数据上能够实现更好的性能。本文基于在中文领域内模型效果和计算效率具有较好评价的ERNIE轻量级模型^[26], 使用预训练模型微调的方法进行多标签文本分类。具体来说, 将预训练模型中[CLS]标记的输出特征表示作为输入句子的全局语义特征, 然后一个线性层对其进行处理, 以完成文本分类任务的训练。通过对待测App进行服务类型上的多标签文本分类, 可以确定该App的基本业务功能, 从而界定该App可以收集的用户个人信息范围。

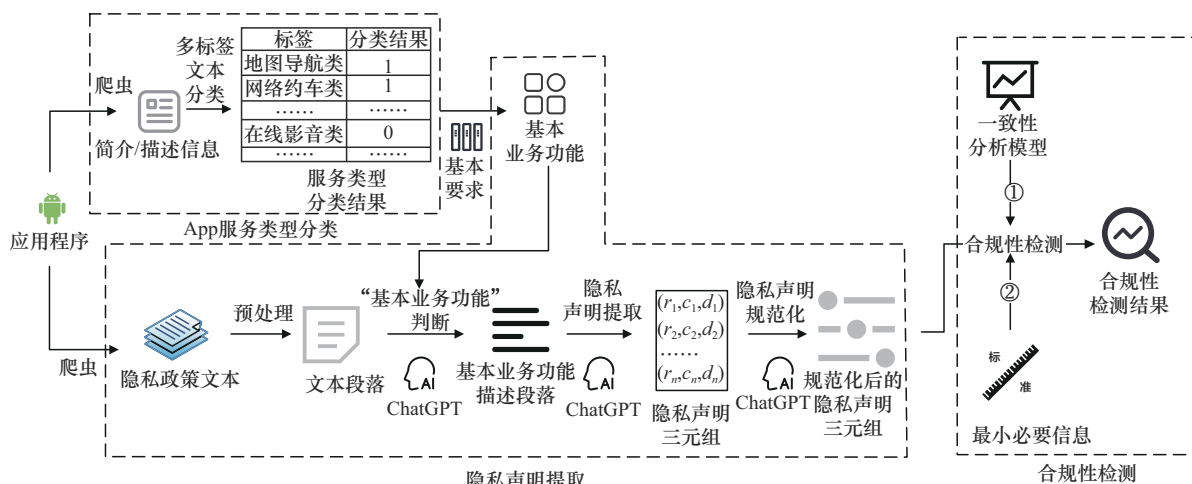


图1 MNPDP的整体框架

Figure 1 MNPDP overall framework

3.3 隐私声明提取

隐私声明提取模块的主要功能是依据 App 所属的服务类型，在具体服务类型基本业务功能模式下，从 App 的隐私政策中提取该 App 的数据收集行为，并形式化为三元组便于后续合规性检测。为了提高隐私声明信息的提取准确率以及可解释性，隐私声明提取模块细分为 4 个部分：隐私政策文本预处理、基本业务功能判断、隐私声明提取和隐私声明规范化。

由于隐私政策中语句结构的多样性和复杂性，基于传统 NLP 技术的方法一方面需要大量的人工工作，包括特定领域数据集的构建、匹配或者映射规则的完善等，流程烦琐；另一方面，由于这些方法都是基于语句级别的 NLP 技术，只能处理单一语句，在语句间存在交叉引用或者语句包含复杂子句的情况下难以精确提取隐私声明。而大语言模型通常在大规模的文本数据集上进行训练，使用的深度神经网络架构和自注意力机制能够捕获长距离的依赖关系和复杂的语言结构，能够更好地理解和利用全局上下文信息，在众多 NLP 任务中表现优越。因此，本文使用大语言模型来完成隐私政策文本分析任务。

为了指导大模型准确地完成隐私政策文本分析任务，本文使用 CoT 方法指导 ChatGPT 分步骤分析每一段文本。Wei^[22]等通过在多个数据集上的对比试验表明，CoT 明显优于标准提示方法，在某些数据集上的表现甚至超过特定领域微调后的模型。

3.3.1 隐私政策文本预处理

应用程序的隐私政策通常通过应用市场链接到开发者网站，并以 HTML 的形式呈现。为了从

隐私政策网页中提取出隐私政策文本，本文使用 Python 第三方库 html2text 将 HTML 文档转化为 Markdown 格式。以往的研究大多使用“。”“！”等句子分隔符将隐私政策分割成一个个语句，然后使用语句级别的 NLP 技术解析每个语句。但是，这种方法会丢失语句的上下文信息，且无法处理语句间的交叉引用，导致遗漏很多隐私声明信息。本文方法保留 Markdown 格式的段落结构，对每个段落进行分析以保持完整语义信息。

对于隐私政策文本中的列表，分别合并列表开头文本和每一个子句形成多个完整段落，在保持完整语义的同时区分各个子句的具体含义。最后，通过去除隐私政策文本中每一个段落中的列表编号、标题标记等字符来规范化隐私政策文本。列表文本预处理示例如图 2 所示。

3.3.2 “基本业务功能”判断

该模块的主要目的是基于基本业务功能对隐私政策中的大量文本进行预过滤，精确定位到有关“基本业务功能”的描述段落，便于后续隐私声明的提取。

为了分辨出隐私政策中的每段文本是否与基本业务功能相关，本模块使用的提示工程方法为 Zero-shot-CoT^[27]，提供具体分析步骤指导 ChatGPT 分析每一段文本。首先，让 ChatGPT 判断该段文本是否描述了 App 的相关业务功能。这是因为隐私政策中包含了许多与 App 业务功能无关的文本，会影响大模型的后续判断。然后，让 ChatGPT 判断该文本是否与基本业务功能直接相关。最后，对输入的每一段文本输出一个判断结果。

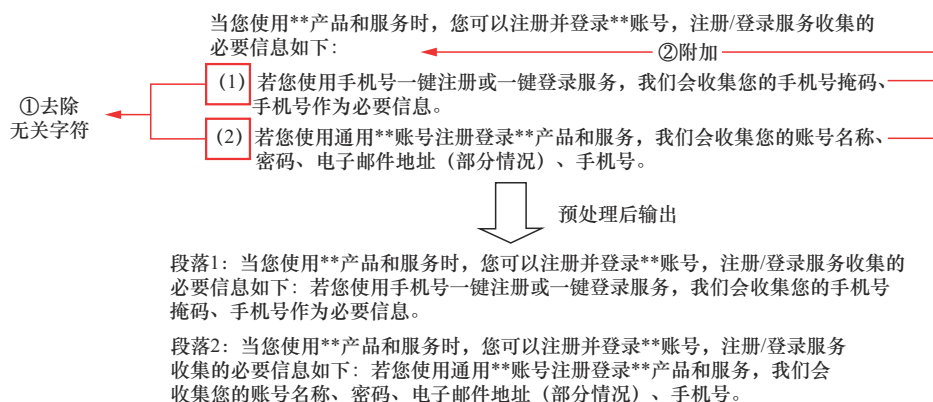


图2 列表文本预处理示例

Figure 2 List text preprocessing example

3.3.3 隐私声明提取

隐私声明提取的目的是获得隐私政策语句的简明表示,以便对该语句进行自动推理。即将隐私政策中描述的有关个人数据收集的语句转化成便于推理的三元组 (r,c,d) 形式。其中, r 表示数据接收者,通常为第一方(App本身)或者第三方(包括具体的服务提供商等); c 表示是/否收集; d 表示收集的数据类型,如用户姓名、年龄等敏感信息。对于语句“当您使用我们的音视频、电子读物(书籍、漫画)等服务内容的浏览、播放功能时,我们需要记录您的设备信息和上网记录。”该模块会将此语句转换为两个三元组,即(我们,是,设备信息)和(我们,是,上网记录)。

由于本模块需要完成的任务相对复杂,因此本模块采用Few-shot-CoT方法使用ChatGPT完成隐私声明的提取任务,在CoT的基础上给大模型提供少量样本进行学习,同时让大模型能够按照规定的格式进行输出,方便后续自动化处理。具体来说,首先让ChatGPT分析该行语句是否描述了个人数据的收集、使用、共享行为,如果是,就从句子中提取出数据收集行为的参数信息,返回一个json数据,键分别为数据接收者、数据类型、是/否收集。

3.3.4 隐私声明规范化

对于第3.3.3节提取出的隐私声明三元组,其中的数据接收者和数据类型都需要进行规范化处理,以满足后续自动化分析的需要。给定这样两个隐私声明三元组:(App,是,设备信息);(我们,否,设备基础信息)。该模块会将以上两个隐私声明三元组分别转化为:(第一方,是,设备基础信息);(第一方,否,设备基础信息)。

对数据接收者和数据类型进行术语上的统一,便于后续进行隐私声明三元组之间的比较。

对于数据接收者,隐私政策中通常使用“我们”“本App”或者该App名称等表示App本身,而使用“第三方”或者具体服务商名称来表示除了App本身之外的数据接收者。因此,使用少量的关键词即可对数据接收者进行规范化。具体来说,对于表示App本身的词语,将其转化为“第一方”,如“我们”“我”等;对于表示具体服务提供商并且与该App名称无关的名词,直接保留,如“百度联盟SDK”等;对于表示第三方的宽泛术语,如“第三方提供商”“服务提供商”等,将其转化为“第三方”标准表示。需要注意的是,在隐私声明三元组之间的比较过程中,具体服务提供商如“百度联盟SDK”被视作“第三方”的一个特例,形式化为“百度联盟SDK”<“第三方”,即“百度联盟SDK”从属于“第三方”。

对于数据类型,以往研究的做法通常是先规定数据类型的基准名词,再通过同义词列表将隐私政策中出现的不规范的数据类型名词映射到基准名词上。这种方法依赖于同义词列表的完备性。但是,由于隐私政策中的数据类型术语复杂且多样,同义词列表很难覆盖所有情形。得益于大语言模型在自然语言处理上的巨大优势,本模块调用ChatGPT完成不规范的数据类型到基准名词的映射。

结合《中华人民共和国个人信息保护法》、《信息安全技术 个人信息安全规范》等隐私法规,本文定义了7种敏感数据类型以及28种一般数据类型,如表1所示。使用ChatGPT将不规范的数据类型映射到基准名词实际上就是对隐私政策中的数据类型进行分类。提示词结构如图3所示,

表1 数据类型类别
Table1 Categories of data types

类别	数据类型
敏感数据类型	种族、生物识别信息、宗教信仰、医疗健康信息、金融账户信息、行踪轨迹信息、性取向
一般数据类型	姓名、年龄、出生日期、性别、住址、手机号码、邮箱地址、身份账号信息、国籍、民族、教育经历 设备标识符/设备唯一识别码、IP地址、音视频、照片、应用列表、设备基础信息 浏览和搜索历史记录、联系人列表、播放记录、购买记录、支付信息、交易信息、上网记录、地理位置信息、通信录、通信记录、日历

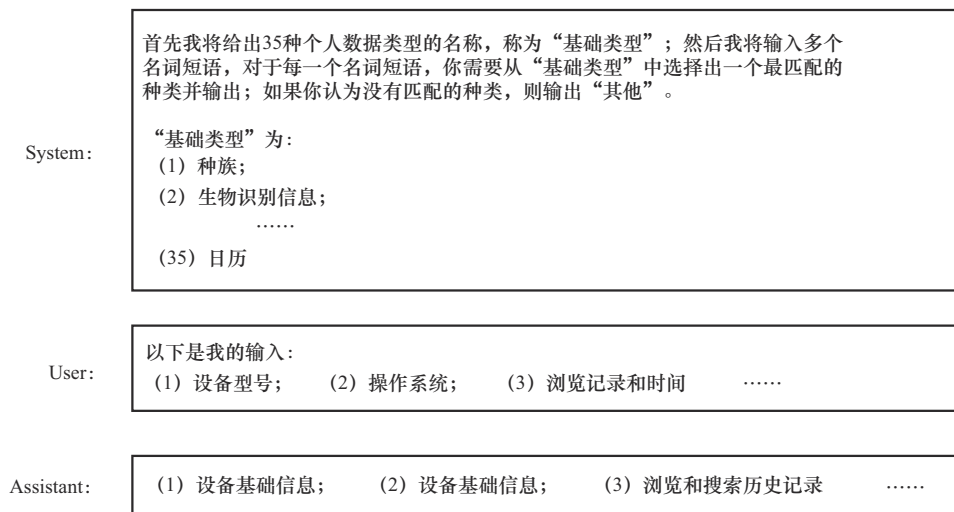


图3 提示词结构

Figure 3 Prompt word structure

通过 System 角色设定 AI 的行为、角色和背景,使 ChatGPT 更好地理解其在对话中的角色;通过 User 角色输入待规范化的数据类型名词;Assistant 则是 ChatGPT 的分析结果。同时,为了方便合规性检测,本文还规定了个人信息作为表 1 列出的 35 种数据类型的父级节点术语,即这 35 种数据类型术语从属于“个人信息”。

3.4 合规性检测

3.4.1 “最小必要”信息界定

为了判断 App 收集的个人信息是否合规,首先需要界定数据收集的“最小必要”信息范围。基本要求中提供了偏向于消费侧的“最小必要”信息范畴,但是这对于大多数 App 实现其基本业务功能来说是不够的,还需要增加服务供给侧用户个人信息。由于没有统一的法规明确规定“最小必要”信息的范围,本文参考 POLICYCOMP^[19]中提出的方法,使用对等比较的思想来界定“最小必要”信息的范围。使用这种方法的直觉假设是,大多数实现类似目的的应用程序均会收集的个人信息数据更有可能是 App 为了实现这些目的所必需的,从而为确定“最小必要”信息提供了潜在的标准。

POLICYCOMP 通过计算待测应用程序描述和候选应用程序描述之间的语义相似度,并进行排序来提取出对等应用,以生成该类 App 普

遍收集的个人数据类型用于合规性检测。但是,这种方法忽略了部分应用缺乏高度相似的对等应用的现象,有些待测 App 甚至只能提取出 3 个对等应用,对等应用过少导致对等比较可信度偏低;而且,当待测 App 与其对等应用普遍收集了过于宽泛的数据类型时,对等比较也会失去意义。因此,本文在对待测 App 进行服务类型分类的基础上,选择该服务类型中下载量最高的 Top- k 个应用程序作为该服务类型的基准 App 集。这是因为应用程序的下载量高表明该应用程序相对该类别中的其他应用程序而言更被用户认可和信任,其数据收集行为符合规范的概率更高。

为了衡量一个数据类型是否属于“最小必要”信息的范围,给出以下定义。

定义 1 给定一个服务类型 S , 其“最小必要”信息表示为 $N(S) = \{d_1, d_2, \dots, d_s\}$ 。其中, s 表示服务类型, S 表示“最小必要”信息中的数据类型数量, d_i 表示规范化后的数据类型名词。

定义 2 给定一篇隐私政策 P , 其声明收集的数据类型集合表示为 $D(P) = \{d_1, d_2, \dots, d_n\}$ 。其中, n 表示隐私政策 P 中收集的数据类型种类数量, d_i 表示规范化后的数据类型名词。

定义 3 给定服务类型 S 中的 Top- k 篇隐私政策, $d_i \in D(P_1) \cup D(P_2) \cup \dots \cup D(P_k)$, 即 d_i 表示在这 k 篇隐私政策中出现的某一数据类型,

其属于“最小必要”信息范围的可信度定义为

$$L(d_i) = \frac{1}{k} \sum_{m=1,2,\dots,k} \begin{cases} 1, d_i \in D(P_m) \\ 0, \text{其他} \end{cases}$$

对于一个给定阈值 α ，当 $L(d_i) \geq \alpha$ 时，即认定 $d_i \in N(S)$ 。

3.4.2 合规性检测

假设给定一篇隐私政策中的两个句子：①在您使用我们的产品时，我们会收集您的手机号码。②我们不收集用户的任何个人信息。对于这样两个句子，按照第3.3节所述提取出各自的隐私声明三元组信息：①（第一方，是，手机号码）；②（第一方，否，个人信息）。由于手机号码属于个人信息的一种，因此该隐私政策显然存在表述上的逻辑矛盾。而对于另外一种情况，当两个隐私声明三元组为（第一方，否，手机号码）和（第一方，是，个人信息）时，可以将第一方不收集手机号码的行为视为第一方对个人信息的收集行为的特例，则并不存在表述上的矛盾。也就是说，当隐私政策使用宽泛的术语表示数据收集行为而使用精确的术语表示不收集行为时，不认定为存在矛盾。

为了分析隐私政策本身是否存在逻辑上的矛盾，本文对隐私政策文本一致性分析模型进行了定义。

定义4 给定一篇隐私政策 P ，经过第3.3节所述的处理后提取出的隐私声明三元组集合表示为 $PS_p = \{ps_1, ps_2, \dots, ps_n\}$ 。其中， $ps = (r, c, d)$ 表示隐私声明， r 表示数据接收者， c 表示是/否收集， d 表示收集的数据类型， n 表示三元组的数量。同时，给出了以下两个集合的定义。

1) $PS_{p_not} = \{ps_1, ps_2, \dots, ps_m\}$ ，对于所有 $ps_i = (r_i, c_i, d_i) \in PS_{p_not}, c_i = \text{'否'}$ ；

2) $PS_{p_collect} = \{ps_a, ps_b, \dots, ps_z\}$ ，对于所有 $ps_j = (r_j, c_j, d_j) \in PS_{p_collect}, c_j = \text{'是'}$ 。

其中， PS_{p_not} 表示隐私政策 P 中声明的不收集行为的集合， $PS_{p_collect}$ 表示隐私政策 P 中声明的收集行为的集合。

定义5 表述一致：对于隐私政策 P 中提取出的所有隐私声明三元组 $PS_p = PS_{p_collect} \cup PS_{p_not}$ ，

对于任意 $ps_j = (r_j, c_j, d_j) \in PS_{p_collect}$ 和 $ps_i = (r_i, c_i, d_i) \in PS_{p_not}$ ，不存在 $r_j \leq r_i$ 并且 $d_j \leq d_i$ 的情况，其中 $r_j < r_i$ 表示 r_j 从属于 r_i ， $d_j < d_i$ 表示 d_j 从属于 d_i 。

定义6 表述不一致：对于隐私政策 P 中提取出的所有隐私声明三元组 $PS_p = PS_{p_collect} \cup PS_{p_not}$ ，存在 $ps_j = (r_j, c_j, d_j) \in PS_{p_collect}$ 和 $ps_i = (r_i, c_i, d_i) \in PS_{p_not}$ ，满足 $r_j \leq r_i$ 并且 $d_j \leq d_i$ ，其中 $r_j < r_i$ 表示 r_j 从属于 r_i ， $d_j < d_i$ 表示 d_j 从属于 d_i 。

如果待测App的隐私政策存在表述上的不一致，则直接判定其不合规；对于表述逻辑上不存在矛盾的隐私政策，则分析其声明要收集的数据类型是否符合“最小必要”信息的范围要求。

为了显示App隐私政策违反“最小必要”原则的严重程度，本文基于表1敏感数据类型和一般数据类型定义，定义了3个风险等级。

定义7 高风险等级：过度收集了敏感数据类型，即对于一个待测App，其所属的服务类型为 $S = \{S_1, S_2, \dots, S_n\}$ ，隐私政策表示为 P ，隐私政策中声明收集的数据类型为 $D(P) = \{d_1, d_2, \dots, d_n\}$ ，存在 $d_i \in D(P)$ 且 d_i 属于敏感数据类型，而 $d_i \notin N(S_1) \cup N(S_2) \cup \dots \cup N(S_n)$ 。

定义8 中风险等级：过度收集了一般数据类型，即对于一个待测App，其所属的服务类型为 $S = \{S_1, S_2, \dots, S_n\}$ ，隐私政策表示为 P ，隐私政策中声明收集的数据类型为 $D(P) = \{d_1, d_2, \dots, d_n\}$ ，存在 $d_i \in D(P)$ 且 d_i 属于一般数据类型而 $d_i \notin N(S_1) \cup N(S_2) \cup \dots \cup N(S_n)$ ，不存在 $d_i \in D(P)$ ，且 d_i 属于敏感数据类型而 $d_i \notin N(S_1) \cup N(S_2) \cup \dots \cup N(S_n)$ 。

定义9 低风险等级：没有出现过度收集数据的情况，即对于一个待测App，其所属的服务类型为 $S = \{S_1, S_2, \dots, S_n\}$ ，隐私政策表示为 P ，隐私政策中声明收集的数据类型为 $D(P) = \{d_1, d_2, \dots, d_n\}$ ，对于任意 $d_i \in D(P)$ ， $d_i \in N(S_1) \cup N(S_2) \cup \dots \cup N(S_n)$ 。

4 实验

4.1 实验环境

本文从华为应用市场 16 个大类别、77 个子类别中抓取了 6 125 个应用程序的相关信息, 包括应用程序的简介、隐私政策链接以及应用下载量等信息。其中超过 75% 的应用程序在华为应用市场上的下载量超过 10 万次。

为了训练多标签文本分类模型, 本文从科大讯飞的长文本分类数据集^[28]中选取了 3 915 个样本进行重新标注, 其中每个类别的样本数量约为 100。使用“ernie-1.0-large-zh-cw”作为预训练模型, 将分词器 tokenizer 使用的最大序列长度设为 256, 批处理大小设为 16, 训练轮次设为 100, 并使用早停法 (early stopping, ES), 即模型在验证集经过一定 epoch 后精度表现不再上升, 训练终止。选择在验证集上表现最好的参数作为最终模型参数。

对于隐私声明提取模块, 本文使用的大语言模型为 OpenAI 推出的“GPT-3.5-Turbo”模型, 调用该模型的 API 进行自动化分析, 并设置 temperature 参数为 0.1。

为了验证本文方法的有效性, 从 6 125 个应用程序中提取出 107 个服务类型仅为在线影音类的 App, 去除了无法提取隐私政策的 App 后形成了包含 101 篇 App 隐私政策的数据集。为了减少

噪声, 删除了每一篇隐私政策中字符长度小于 10 的段落, 最终形成了包含 18 009 个隐私政策段落的语料库。

4.2 实验结果

4.2.1 MNPD 分析结果

对于 App 服务类型的分类, 本文使用基于“ernie-1.0-large-zh-cw”微调后的模型对 6 125 个应用程序进行分类, 分类结果如图 4 所示。其中 162 个应用程序具有多个标签, 即具有多个服务类型。

由于在线影音类 App 在日常生活中使用较广泛, App 数量在中位数附近 (图 4 红色部分), 因此本文选择对在线影音类 App 进行自动化分析测试, 具体包含 101 个样本, 提取出 467 个与基本功能相关的隐私政策段落, 形成 2 517 个隐私声明元组, 排除掉 5 篇存在表述上逻辑错误的隐私政策后, 剩下的 96 篇隐私政策声明其在基本功能模式下一共收集了 366 个数据类型。该类 App 收集的数据类型数量和对每个数据类型的收集数量的分布分别如图 5 和图 6 所示。从图 5 可以看出, 大部分 App 在基本功能模式下收集的数据类型数量小于 15 个, 有超过 25% 的应用程序在隐私政策中声明其在基本功能模式下不收集用户的个人数据。从图 6 可以看出, 对于敏感数据, 只有少部分应用程序进行了收集, 这表明大部分应用程序对于敏感数

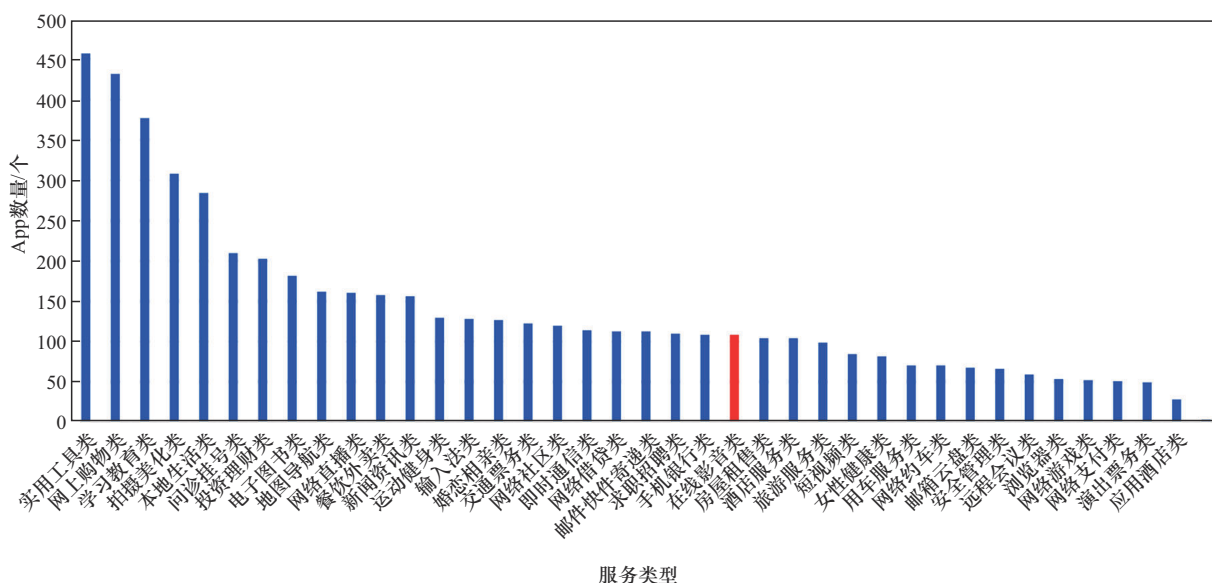


图 4 不同服务类型 App 的数量

Figure 4 Number of different types of service App

据的收集是很谨慎的。有超过40%的应用程序声明了其音视频、设备基础信息和浏览和搜索历史记录一般类型数据进行了收集。

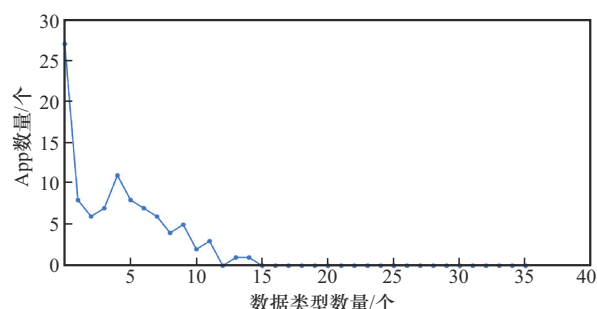


图5 在线影音类App收集的数据类型数量

Figure 5 The number of data types collected by online audio-visual App

本文选择在线影音类App中下载量最高的Top- k 个应用程序作为基准App，并将 $L(d_i) \geq \alpha$ 的数据类型置为“最小必要”信息。为了选取合

适的Top- k 和 α 值，本文对96篇隐私政策（排除了存在表述上逻辑错误的5篇隐私政策）进行了人工标注，提取每篇隐私政策声明的在基本功能模式下收集的数据类型并分析每个数据类型是否属于过度收集，最终提取出369个数据类型，其中有145个数据类型的收集被判定为过度收集。选取的Top- k 和 α 值以及对应的F1分数如表2所示。依据表2可以看出当 $k = \lceil \text{App_count} * 0.2 \rceil = 20$ 并且 $\alpha = 0.6$ 时，F1分数达到了最高。

以 $k = \text{App_count} * 0.2 = 20$ 和 $\alpha = 0.6$ 为参数对“在线影音类”App进行合规性检测，结果如表3所示。基于该结果有以下发现。

1) 目前大部分App的隐私政策不存在表述上的不一致现象，只有4.95%的App隐私政策存在表述不一致的逻辑矛盾问题。这表明大部分App的隐私政策在对其数据收集行为进行声明时都不

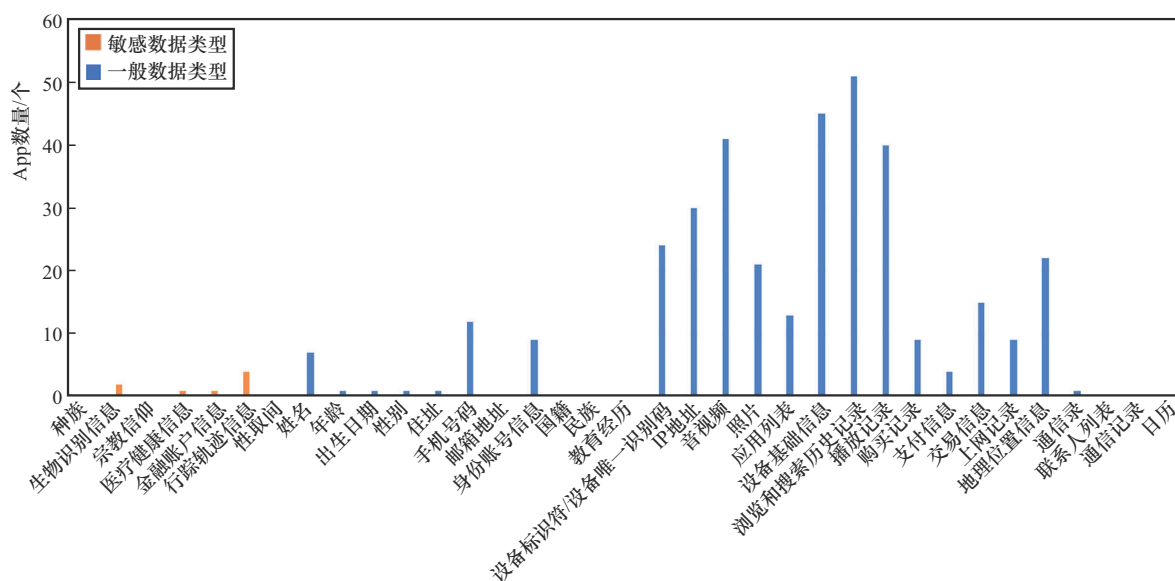


图6 “在线影音类”App中对每个数据类型的收集数量

Figure 6 The number of data types collected by each online audio-visual App

表2 App_count = 96时，不同 k 值和 α 值下的F1分数

Table 2 F1 scores for different k and α when App_count = 96

	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.66$	$\alpha = 0.75$
$k = \lceil \text{App_count} * 0.1 \rceil = 10$	0.760 1	0.766 4	0.895 4	0.794 2
$k = \lceil \text{App_count} * 0.2 \rceil = 20$	0.867 7	0.895 4	0.794 2	0.711 6
$k = \lceil \text{App_count} * 0.3 \rceil = 29$	0.895 4	0.895 4	0.627 0	0.627 0
$k = \lceil \text{App_count} * 0.4 \rceil = 39$	0.895 4	0.791 9	0.711 6	0.562 6

会产生歧义,不会导致用户阅读理解上的障碍。Andow 等^[6]声称其从谷歌应用市场上 11 430 个应用程序中识别出 14.2% 的应用程序隐私政策存在逻辑上的表述矛盾。与之相比,华为应用市场中的 App 表现较为良好。这表明国内应用程序开发者和应用商店更加重视 App 隐私政策,更加重视隐私合规建设。

表 3 在线影音类 App 合规性检测结果
Table 3 Compliance test results of “Online Audio-Visual” App

表述不一致	表述一致		
	低风险	中风险	高风险
4.95%	34.65%	54.46%	5.94%

2) 超过 60% 的 App 隐私政策存在基本功能模式下的个人数据过度收集行为。这表明应用市场上大部分 App 的隐私政策还存在问题,其中超过 90% 的 App 被识别为中风险,显然应用程序开发者还不够重视一般数据类型,没有意识到对一般数据类型如 IP 地址、应用列表等信息的过度收集实际上也是对用户隐私的侵犯。

4.2.2 实验评价

为了对 MNPD 的性能进行评估,本文将所有样本标注为三类,即表述不一致、表述一致且没有过度收集行为、表述一致但存在过度收集行为。对 101 个在线影音类 App 进行人工标注,最终数据集包含 4 个表述不一致的样本、47 个表述一致且没有过度收集行为的样本、50 个表述一致但存在过度收集行为的样本。使用 MNPD 对这 101 个样本进行整体检测并分析,最终达到了 86.20% 的 F1 分数。

接下来对 App 服务类型分类以及隐私声明提取和合规性检测部分分别进行实验分析。

(1) App 服务类型分类

本文使用宏平均 (macro-average) 和微平均 (micro-average) 作为多标签文本分类模型的评价指标。宏平均是对每个类别单独计算性能指标 (如准确率、召回率和 F1 分数),然后对所有类别的指标取平均,每个类别被视为同等重要,不考虑类别的样本数量。微平均是将所有类别的真阳性、假阳性和假阴性汇总,然后计算准确率、

召回率和 F1 分数。因为所有类别的预测结果都被汇总到一个混淆矩阵中,所以微平均更关注较大类别的性能。指标计算公式如下所示。

$$\text{Macro Precision} = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$\text{Macro Recall} = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$\text{Macro F1} = \frac{2 \times \text{Macro Precision} \times \text{Macro Recall}}{\text{Macro Precision} + \text{Macro Recall}} \quad (3)$$

$$\text{Micro Precision} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i)} \quad (4)$$

$$\text{Micro Recall} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FN_i)} \quad (5)$$

$$\text{Micro F1} = \frac{2 \times \text{Micro Precision} \times \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}} \quad (6)$$

人工标注的样本数据按照 8:2 划分训练集和开发集,实验评价指标结果如表 4 所示。

表 4 App 服务类型分类效果

服务类型	Precision	Recall	F1
Macro	92.86%	84.92%	88.71%
Micro	92.90%	85.62%	88.48%

为了验证本文所使用的基于预训练模型的多标签文本分类算法的有效性,本文与传统的机器学习方法、深度学习方法进行了对比实验,实验结果如表 5 所示。可以看出,基于预训练的方法相比机器学习方法、深度学习方法在 F1 分数上均有超过 10% 的提升。

表 5 F1 分数对比

模型	Macro F1	Micro F1
SVM	71.45%	72.25%
TextCNN	75.87%	77.87%
ERNIE	88.71%	88.48%

(2) 隐私声明提取和合规性检测

同样,本文使用宏平均 (macro-average) 和微平均 (micro-average) 作为系统性能的评价指标,用来检测隐私声明提取模块和合规性检测模块的整

体性能。合规性检测系统评价指标如表6所示。从实验结果可以看出, MNPD能够以较高的准确度从“最小必要”原则视角对应用程序进行合规性检测。

表6 合规性检测系统评价指标

Table 6 Evaluation metrics for compliance detection system

类别	Precision	Recall	F1
表述不一致	80.00%	100.00%	88.89%
表述一致且没有过度收集行为	100.00%	74.47%	85.37%
表述一致但存在过度收集行为	80.33%	98.00%	88.29%
Macro-average	86.78%	90.82%	88.75%
Micro-average	87.13%	87.13%	87.13%

通过分析实验结果, 发现表述一致且没有过度收集行为类别的召回率偏低, 即存在部分该类别样本没有被正确预测为表述一致且没有过度收集行为。通过分析错误样本, 发现这部分样本被误分类为表述一致但存在过度收集行为, 这也是表述一致但存在过度收集行为类别精准率过低的原因。通过进一步分析发现, 在对这些App的隐私政策进行自动化分析过程中, 基本业务功能判断模块误将部分与基本业务功能无关的段落进行了提取, 导致后续分析过程中误提取了过多的数据类型。考虑在后续研究中通过使用更加先进的提示工程方法(如least-to-most prompting^[29]等)或者性能更加优秀的大语言模型(如GPT-4等)进行改善。

5 结束语

本文基于中国应用市场上的移动应用程序进行“最小必要”原则检测研究的空白, 提出了一种基于对等比较思想并利用大语言模型对隐私政策进行分析的合规性检测框架, 并给出了合规性检测模型的形式化定义, 从“最小必要”原则视角对华为应用市场上的App进行自动化检测。实验结果表明, 本文提出的方法能够准确地对移动应用程序进行合规性检测, 分析其是否存在对用户数据的过度收集行为。

在下一步工作中, 将基于“最小必要”原则,

结合移动应用程序的实际运行行为, 将其与隐私政策中所声明的数据收集行为进行对比, 分析该应用程序隐私政策是否存在刻意“说谎”现象。

参考文献:

- [1] 中华人民共和国工业和信息化部. 2023年前三季度互联网和相关服务业运行情况[EB]. 2023.
Ministry of Industry and Information Technology of the People's Republic of China. Operation status of the internet and related services in the first three quarters of 2023 [EB]. 2023.
- [2] VOIGT P, VON DEM BUSSCHE A. The EU general data protection regulation (GDPR) A Practical Guide[M]. Berlin: Springer International Publishing 2017, 10(3152676): 10.
- [3] GOLDMAN E. An introduction to the California consumer privacy act (CCPA) [N]. Santa Clara Univ. Legal Studies Research Paper, 2020.
- [4] 中华人民共和国个人信息保护法[J]. 中华人民共和国全国人民代表大会常务委员会公报, 2021, (6): 1117-1125.
Personal Information protection law of the People's Republic of China[J]. Gazette of the Standing Committee of the National People's Congress of the People's Republic of China, 2021, (6): 1117-1125.
- [5] 国家市场监督管理总局. 信息安全技术 移动互联网应用程序(App)收集个人信息基本要求: GB/T 41391-2022[S]. 北京: 中国标准出版社, 2022.
State Administration for Market Regulation. Information security technology-basic requirements for collecting of personal information in mobile Internet applications: GB/T 41391-2022[S]. Beijing: Standards Press of China, 2022.
- [6] ANDOW B, MAHMUD S Y, WANG W, et al. PolicyLint: investigating internal privacy policy contradictions on google play[C]//28th USENIX Security Symposium (USENIX Security 19). 2019: 585-602.
- [7] BUI D, YAO Y, SHIN K G, et al. Consistency analysis of data-usage purposes in mobile Apps[C]//Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. 2021: 2824-2843.
- [8] AMOS R, ACAR G, LUCHERINI E, et al. Privacy policies over time: Curation and analysis of a million-document dataset[C]//Proceedings of the Web Conference 2021. 2021: 2165-2176.
- [9] AHMAD W U, CHI J, LE T, et al. Intent classification and slot filling for privacy policies[J]. arXiv preprint arXiv:2101.00123, 2021.
- [10] BUI D, SHIN K G, CHOI J M, et al. Automated extraction and presentation of data practices in privacy policies[J]. Proc Priv Enhancing Technol, 2021, 2: 88-110.
- [11] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
- [12] CHOWDHURY A, NARANG S, DEVLIN J, et al. Palm: scaling language modeling with pathways[J]. arXiv Preprint arXiv: 2204.02311, 2022.
- [13] OpenAI. GPT-4 technical report [R]. 2023.
- [14] WILSON S, SCHAUB F, RAMANATH R, et al. Crowdsourcing annotations for websites' privacy policies: can it really work[C]//

Proceedings of the 25th International Conference on World Wide Web. 2016: 133-143.

- [15] 李昕, 唐鹏, 张西珩, 等. 面向 GDPR 隐私政策合规性的智能化检测方法[J]. 网络与信息安全学报, 2023, 9(6): 127-139.
- LI X, TANG P, ZHANG X H, et al. GDPR-oriented intelligent checking method of privacy policies compliance[J]. Chinese Journal of Network and Information Security, 2023, 9(6): 127-139.
- [16] LIU F, RAMANATH R, SADEH N, et al. A step towards usable privacy policy: automatic alignment of privacy statements[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 884-894.
- [17] PEDDINTI S T, BILOGREVIC I, TAFT N, et al. Reducing permission requests in mobile Apps[C]//Proceedings of the Internet Measurement Conference. 2019: 259-266.
- [18] JANA S, ERLINGSSON Ú, ION I. Apples and oranges: detecting least-privilege violators with peer group analysis[J]. arXiv Preprint arXiv:1510.07308, 2015.
- [19] ZHOU L, WEI C, ZHU T, et al. {POLICYCOMP}: counterpart comparison of privacy policies uncovers overbroad personal data collection practices[C]//32nd USENIX Security Symposium (USENIX Security 23). 2023: 1073-1090.
- [20] SHANAHAN M, MCDONELL K, REYNOLDS L. Role play with large language models[J]. Nature, 2023: 1-6.
- [21] LOGAN IV R L, BALAŽEVIĆ I, WALLACE E, et al. Cutting down on prompts and parameters: simple few-shot learning with language models[J]. arXiv Preprint arXiv:2106.13353, 2021.
- [22] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
- [23] TANG C, LIU Z, MA C, et al. PolicyGPT: automated analysis of privacy policies with large language models[J]. arXiv preprint arXiv:2309.10238, 2023.
- [24] WILSON S, SCHAUB F, DARA A A, et al. The creation and analysis of a website privacy policy corpus[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 1330-1340.
- [25] LIU S, ZHAO B, GUO R, et al. Have you been properly notified? automatic compliance analysis of privacy policy text with GDPR article 13[C]//Proceedings of the Web Conference 2021. 2021: 2154-2164.
- [26] SUN Y, WANG S H, LI Y, et al. ERNIE: enhanced representation through knowledge integration[J]. arXiv Preprint arXiv:1904.09223, 2019.
- [27] KOJIMA T, GU S S, REID M, et al. Large language models are zero-shot reasoners[J]. Advances in Neural Information Processing Systems, 2022, 35: 22199-22213.
- [28] XU L, HU H, ZHANG X, et al. CLUE: a Chinese language understanding evaluation benchmark[J]. arXiv Preprint arXiv:2004.05986, 2020.
- [29] ZHOU D, SCHÄRLI N, HOU L, et al. Least-to-most prompting enables complex reasoning in large language models[J]. arXiv Preprint arXiv:2205.10625, 2022.

[作者简介]



余佩厚 (2000-), 男, 安徽滁州人, 南京邮电大学硕士生, 主要研究方向为隐私合规。



徐天辰 (2000-), 男, 江苏盐城人, 南京邮电大学硕士生, 主要研究方向为隐私合规、自然语言处理。



孙雯倩 (2000-), 女, 江苏连云港人, 南京邮电大学硕士生, 主要研究方向为隐私合规。



陈云芳 (1976-), 男, 江苏镇江人, 博士, 南京邮电大学副教授, 主要研究方向为人工智能算法、特定应用领域的功能分析、使用智能系统的应用开发。



于乐 (1990-), 男, 江苏扬州人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为移动应用安全与隐私合规检测、物联网和车联网协议逆向。



张伟 (1973-), 男, 江苏泰兴人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为智能感知与认知、隐私保护与人工智能安全。