计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications [doi: 10.15888/j.cnki.csa.009653]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于多模态融合的移动应用细粒度用户意图理解①

张逸涵, 洪　赓, 杨哲慜

(复旦大学 计算机科学技术学院, 上海 200433)
通信作者: 洪　赓, E-mail: ghong@fudan.edu.cn

摘　要: 随着移动应用功能日益复杂, 现有基于用户意图的隐私泄露检测方法面临更大挑战. 一方面, 传统隐私泄露检测基于应用级别的用户意图, 只关注应用的隐私收集行为是否与应用的核心功能需求相符合, 不适用于现如今具有广泛功能和多元用户意图的移动应用安全检测, 亟需粒度更细的用户意图分类; 另一方面, 现行研究大多集中于评估图标等界面小部件触发的隐私收集行为是否与用户意图一致, 然而, 图标不当设计和滥用现象十分普遍, 这限制了仅依赖小部件用户意图进行隐私风险评估的有效性, 因此当前仍需要对整体用户界面的意图进行理解. 针对以上问题, 本文首先从中文隐私政策中提取总结出常见的、适用于隐私合规判断的细粒度用户意图列表; 之后结合移动应用界面设计特点, 设计并实现了多模态特征融合的多分类模型对整个移动界面反应的用户意图进行识别. 评估结果表明, 本文隐私政策意图提取工具精确率与召回率均达到 83%, 用户意图识别工具精确率与召回率分别达到了 80% 与 83%, 具有较好的检测效果与实际可用性.

关键词: 移动应用; 用户意图; 隐私合规

引用格式: 张逸涵,洪赓,杨哲慜.基于多模态融合的移动应用细粒度用户意图理解.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9653.html

Fine-grained User Intention Understanding for Mobile Applications Based on Multi-modality Fusion

ZHANG Yi-Han, HONG Geng, YANG Zhe-Min

(School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract: With the increasing complexity of mobile applications, existing privacy leak detection methods based on user intent face greater challenges. On the one hand, traditional privacy leak detection, which is based on APP-level user intent, only focuses on whether the privacy collection behavior of the application aligns with its core functional requirements. This approach is not suitable for today's mobile APP security detection, which has broad functionalities and diverse user intents, necessitating a more fine-grained user intent classification. On the other hand, current research mainly focuses on evaluating whether the privacy collection behaviors triggered by interface widgets, such as icons, are consistent with user intent. However, the improper design and misuse of icons are very common, which limits the effectiveness of privacy risk assessments that rely solely on widget-based user intents. Therefore, a comprehensive understanding of user intent at the overall interface level is still needed. In response to the above issues, this study first extracts and summarizes a fine-grained user intent list suitable for privacy compliance detection based on Chinese privacy policies. Then, based on the characteristics of mobile application interface design, a multi-classification model with multi-modal feature fusion is designed and implemented to identify the user intent reflected by the entire mobile interface. Evaluation results show that the intent extraction tool in this study has achieved 83% in both precision and recall, and the user intent classification model reaches 80% and 83% in precision and recall, respectively, demonstrating good detection

① 基金项目: 工信部专项 (TC220H079)
收稿时间: 2024-04-07; 修改时间: 2024-05-06; 采用时间: 2024-05-14; csa 在线出版时间: 2024-09-24

effectiveness and practical usability.

Key words: mobile application; user intention; privacy compliance

随着全球步入移动互联网时代, 移动应用 (APP) 走进千家万户, 从在线购物到视频观看, 移动应用为用户提供多样化的服务, 覆盖生活的方方面面, 成为人们日常生活的重要部分. 为了提供更好的服务, 这些移动应用程序会收集用户的个人信息, 包括身份证、面部信息等敏感个人数据, 而其中部分应用会在用户不知情的情况下收集过多或非必要的隐私数据, 造成用户隐私泄露, 如 2023 年知名购物软件拼多多就被爆出秘密收集用户隐私信息, 并引发了广泛的隐私侵犯担忧. 事实上, 根据中华人民共和国工业和信息化部统计[1], 在 2024 年第 1季度对不良手机应用有效投诉中, 超半数投诉与个人信息和权限问题及信息安全问题相关, 这说明移动隐私安全问题已是目前移动应用的主要问题, 当前需要对移动应用的隐私安全进行有效管理和监控, 以确保用户的移动隐私安全.

为了保护个人隐私数据, 学术界根据应用隐私收集行为是否和用户意图一致判断该隐私收集行为是否可能构成隐私泄露. 例如, 拍摄视频使用摄像头, 登录注册收集用户名和密码, 这些应用功能涉及的隐私收集行为和人们日常生活常识一致, 符合用户意图, 可视为合理的隐私收集; 相反, 如果移动应用在登录注册时强制要求填写生日和性别, 则该行为违反了用户意图, 可能导致隐私泄露.

但随着移动设备性能的不断提升和用户需求的多样化, 移动应用开发呈现出多样化和复杂化的趋势, 现有基于用户意图的隐私泄露检测方法已不能适应当前移动应用隐私安全检测需要. 首先, 传统隐私泄露检测基于应用级别的用户意图, 使用应用描述或应用用户评论建模用户意图[2–5], 仅检测应用的隐私收集行为是否与应用的核心功能需求相符合, 而当前日益复杂的移动应用往往具有多种功能以满足用户多样需求, 因此, 传统应用级的粗粒度用户意图分类已经不能满足现如今功能多样化、意图复杂化的移动应用安全检测需求. 目前亟需更细粒度用户意图分类, 支持更细粒度的用户意图分析, 精确评估和监控移动应用在执行各项功能时的隐私收集行为, 保证用户隐私安全. 此外, 现有许多工作忽视应用功能, 仅关注隐私收集行为触发点, 研究用户界面中类似图标等的可交互控件的用户意图和其触发的隐私收集行为是否一致[6–10]. 然而, 一方面, 图标本身存在设计不当或被滥用的情况, 另一方面, 图标有限信息含量并不能准确反应实际应用功能, 同一图标表示多种意图的情况并不罕见[11], 这限制了依赖图标进行隐私风险评估的准确性和有效性, 仅使用用户界面小部件的意图评估应用功能所需的隐私收集行为会导致误报, 因此, 当前仍需要对能全面反应用户意图的整体用户界面进行理解以准确推断当前用户意图. 综上, 当前不仅需要更细粒度的用户意图分类, 也需要可用的用户意图推断方法以全面且准确地识别细粒度的用户意图, 以支持更细粒度的用户意图分析.

针对以上问题, 本文首先从隐私政策中的移动应用隐私使用实践提取总结出常见的、适用于隐私合规判断的细粒度用户意图列表; 之后, 本文结合移动应用界面设计特点, 设计并实现了多模态特征融合的多分类模型对整个移动界面反应的用户意图进行识别. 具体的, 本文结合语义特征、词性特征和上下文特征构建多特征实体识别模型从大量隐私政策文本中自动识别并提取隐私收集使用目的描述语句中的目的表述, 并对提取的短语进行聚类, 整理得到常见的隐私收集使用目的作为细粒度用户意图分类; 之后, 本文根据安卓界面设计特点, 提取界面控件的位置、类别和文本特征, 结合空间注意力和协同注意力机制构建整体界面多模态融合特征图, 对细粒度用户意图进行识别, 准确理解移动应用运行时用户意图, 弥合细粒度检测需求和具体执行之间的差距, 为监管提供技术支持.

综上, 本文提出并实现了细粒度用户意图类型自动化提取方法, 得到常见细粒度用户意图列表, 提供更细粒度的隐私收集使用分析视角, 比之前工作的用户意图更直接地反映了应用程序的目的和用户使用预期; 此外, 本文设计并实现了细粒度用户意图识别模型, 能够准确地理解界面整体反应的细粒度用户意图. 本文在使用真实世界的应用收集得到的数据集上对细粒度用户意图类型提取方法和用户意图识别模型进行有效性测试和评估, 实验结果显示自动化用户意图提取技术整体准确性达到 83%; 用户意图识别模型达到 80%,

具有良好的可用性.

1 相关工作

本节首先对当前基于用户意图的隐私泄露检测相关研究进行介绍, 并阐述现有工作在用户意图建模上的局限性, 之后介绍现有用户意图分类方法并分析其不足

1.1 用户意图与应用行为一致性检测

早期应用隐私泄露检测不区分敏感数据的传输是否是用户有意的, 这些早期的静态污点分析方法, 如TaintDroid[12], 和动态污点跟踪方法, 如 FlowDroid[13], 将所有敏感数据传输都视为违规. 因此, 为了区别正常应用的合理数据传输和真正的隐私泄露行为, 用户意图概念被学者引入作为标准用以检测应用程序中真正的异常行为和权限滥用.

许多学者都从不同角度研究了如何提取和建模用户意图以检测应用异常行为. 常见的用户意图建模来源包括应用描述、应用评论、隐私政策和用户界面.

(1) 基于应用描述和应用评论的异常行为检测. AutoCog[2]、WHYPER[3]和 CHABADA[4]是最早关注移动应用的应用描述的工作, 它们检测描述文本是否和授权的权限匹配以识别意外行为. 但 SmartPI[5]发现应用程序描述很容易被开发者用来欺骗用户和检测系统. 因此, SmartPI[5], ReviewSolver[14]和 CHAMP[15]选择使用 NLP 技术提取关键字, 通过聚类模型以真实的用户评论推断应用实际的权限使用. 不过, 由于应用描述和应用评论仅提供对应用的整体综合描述信息, 因此只适用于应用整体进行粗粒度用户意图合规检测.

(2) 基于隐私政策的异常行为检测. 隐私政策包含应用如何收集和使用用户信息的声明, 需要得到用户授权, 同样反应用户意图. Slavin 等人[16] 是最早研究隐私政策和移动应用代码 API 行为是否一致的学者, 之后, GUILeak[17], PoliCheck[18]和 Purpliance[19]分别检测了隐私政策和用户输入隐私、数据流以及网络流量的一致性. 这些工作通过规则匹配提取隐私政策中对隐私的使用说明和不同的应用行为进行匹配. 但隐私政策本身存在可能编写错误[20], 影响检测时的准确性.

(3) 基于用户界面的异常行为检测. 根据最小惊讶原则, 用户界面元素的行为方式应与其用户期望的行为方式一致[21], 因此 UI 推断的意图被广泛用于检测移动应用可疑行为. AppIntent[22]是最早利用图形用户界面事件序列通过人工识别推断用户意图判断隐私泄露的研究; 之后, 学者开始使用 NLP 技术研究界面语义是否和隐私行为一致[23,24].

因为 UI 部件常作为应用行为触发入口, UI 部件的意图推断和异常检测受到关注. 许多工作关注点击按钮意图和应用行为是否一致[6–10,25–27]. 早期小部件研究主要使用文本语义信息[21,25], 如 AsDroid[25]使用文本分析技术检查敏感 API 的文本和 UI 小部件的描述性文本是否匹配. 随着计算机视觉技术的发展, 图像特征被用于检测应用图标的预测权限使用和实际权限使用之间的差异[6,7,9,26,27]. 最早的 IconIntent[7]会区分图片图标和文本图标并用图像尺度不变特征和文本关键字编辑距离分别分类. 随后的 DeepIntent[6]首次通过注意力网络融合图片像素特征和图标文本特征综合推断图标意图. 随后图标意图推断工作大都沿用相似模型分类思路只是改进提取模型[26,27]或匹配关联方法[9]. 但图标本身并不能很好反映实际用户意图, 许多图标存在歧义[11]. 而经典图像和文本多类分类深度学习模型对整体移动界面意图推断, 因为截图界面复杂多样, 其准确率仅有 66%[28].

1.2 用户意图分类

在上述用户意图和应用行为一致性检测工作中, 仅有 CHABADA[4]、 AsDroid[25]和 Purpliance[19]提供了具体的用户意图类别, 而非只提供应用行为或权限与用户意图是否一致的二元检测结果. 而研究工作表明[29]数据收集和使用/共享的目的对用户决定是否披露其个人信息的有非常重大的影响, 分类、识别并告知用户意图, 能帮助用户和监管者更好地理解和管理隐私.有学者尝试分类用户意图. Lin 等人[29]归纳总结九种第三方库目的类别; Han 等人[30]和 Jain 等人[31]组织专家小组确定自定义代码的编写目的.但他们都是人工手动归纳总结意图也并未提供自动化检测工具.

在自动意图分类工作中, CHABADA[4]使用主题模型对应用的描述进行聚类得到适用于应用级检测的意图分类, 如“design and art”. 之后, Purpliance[19]通过句法规则匹配从大量英文隐私政策提取的移动应用隐私使用目的. 隐私政策含有应用真实行为的描述, 这样提取的应用目的具备实践价值, 文本挖掘的方法相比人工归纳, 也更加客观, 但该工作得到的意图, 如“provide service”和“personalize service”, 非常模糊, 这是因为英文隐私政策中隐私使用目的描述不具体, 导致其提取

的意图类别实际可用性不足.

除了移动应用隐私安全, 其他领域比如软件改进[32]、推荐系统[33–36]或文本分析[37,38]等也引入用户意图作为其研究的标准, 但不同领域之间对用户意图粒度的需求并不相同, 这些领域的用户意图类别, 比如“功能请求”[32]、“家庭”[34]等并不适用于移动应用隐私合规分析, 分类方法差异也较大.

本文将从移动应用安全检测视角出发, 从中文隐私政策文本中挖掘具体可用的细粒度移动应用用户意图类别, 并利用图像、布局和语义的多模态特征结合特征融合技术提供有效的用户意图识别工具, 辅助用户决策软件收集行为的可接受性, 也为现在更规范和常态化数据监管提供技术支持.

2 整体框架

本文总共分为两个阶段: 首先, 确定实用可行的细粒度用户意图类别; 其次, 设计并实现对整体应用界面的细粒度用户意图的准确识别. 具体框架结构如图 1 所示.

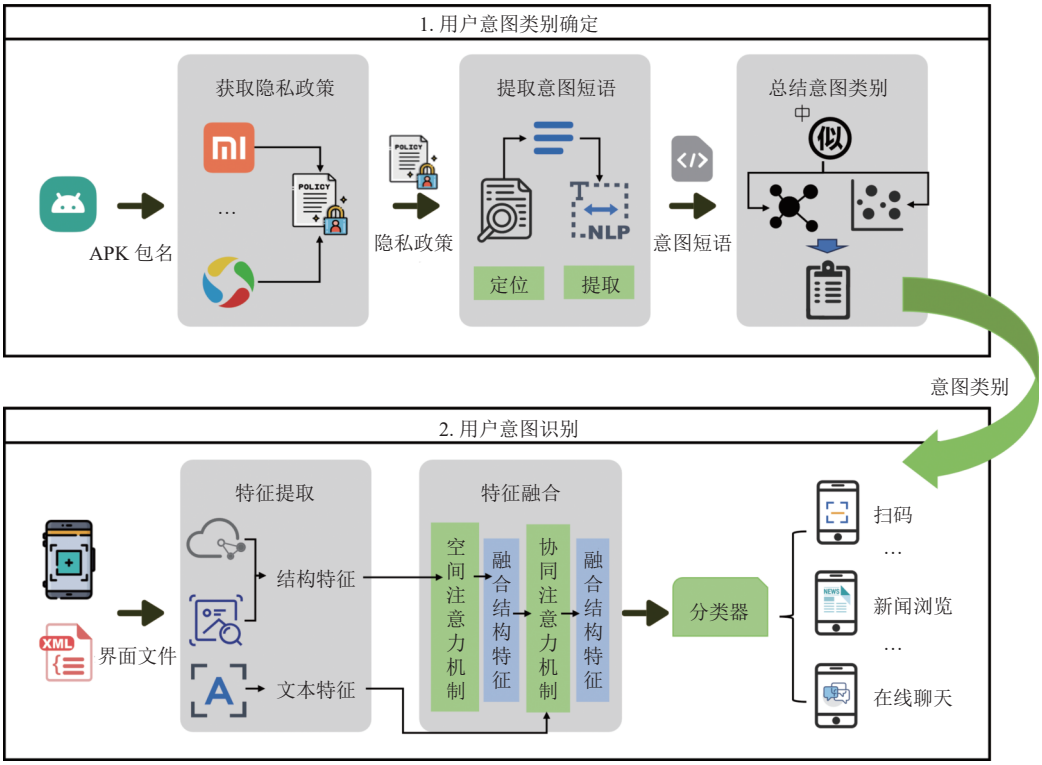


图 1 整体架构

为了得到适用于移动应用隐私安全分析的细粒度用户意图类别, 本文首先从各大中文应用商店中批量收集移动应用隐私政策, 接着定位隐私政策中的数据收集使用规则语句并提取目的短语, 根据短语相似度进行主题聚类, 整理得到细粒度用户意图列表.

确定用户意图类别, 首先需要确定合适的意图来源, 中文应用商店隐私政策受到中国政府部门长期监管和应用商店审查, 相比英文隐私政策对个人信息描述更为规范和具体且容易获取.

然而中文文本语法较具有语法结构的英文文本更加灵活复杂, 且具有更多多义字词, 适用于英文隐私政策文本的规则匹配方法并不能有效识别中文的目的短语. 此外, 隐私政策中大量相似但不属于目的收集语句描述性文本, 同样干扰用户意图的提取. 对此, 本文分析隐私收集使用语句的成分结构, 确定个人数据为切入口, 先定位含有个人数据的语句, 收集大量有效数据, 训练目的语句识别模型, 过滤无关文本; 再结合语义特征、词语特征和上下文特征, 构建多特征命名实体识别模型, 提取意图短语.

最后, 因为意图表述方式多种多样, 所以本文再通过聚类对意图短语进行合并, 而为了降低短语提取带来的噪音, 选择凝聚层次聚类按照簇大小过滤得到最

终细粒度意图类别.

之后, 本文设计了基于多模态特征融合的多分类模型对应用整体用户界面反应的细粒度用户意图进行识别. 目前传统图像识别技术在移动应用用户界面分类领域的效果不如人意. 由于现代应用程序的设计和实现各不相同, 同样意图下的用户界面可供分析的像素差异较大, 使得经典图像分类模型效果不佳. 本文由通用设计模式得到灵感, 提出使用设计结构特征分类应用程序界面. 本文创新性地使用目标检测模型和点云模型从位置大小和控件结构提取布局设计特征, 同时设计空间注意力和协同注意力网络减少空间特征和文本特征融合过程中的损失. 最后, 采用焦点损失函数, 设计类别权重, 平衡不可能均衡的样本数量.

2.1 用户意图类别确定

本节主要介绍: (1) 用于提取细粒度用户意图类别的中文个人数据收集使用规则; (2) 本文设计的适用于中文隐私政策文本的细粒度移动应用用户意图类别确定的方法; (3) 本文得到的细粒度用户意图.

2.1.1 用户意图类别来源

隐私政策, 作为应用运者商如何收集和使用个人信息的声明, 需要得到用户授权, 反映了用户意图. 但目前英文隐私政策中使用规则匹配方法的意图挖掘工作只能提取出如“provide service”“improve service”等模糊类别[18], 这是因为英文隐私政策文本中对使用个人信息的描述并不清晰.

中国于 2021 年 11 月实施的《中华人民共和国个人信息保护法》第 7 条明确规定: “处理个人信息应当遵循公开、透明原则, 公开个人信息处理规则, 明示处理的目的、方式和范围.” 且不同于如欧盟仅出台了法律法规, 中国在近些年出台系列指导性政策文件《网络安全标准实践指南》, 同时展开长期 APP 专项治理工作, 并重点监管了“未公开收集使用规则”和“未明示收集使用个人信息的目的、方式和范围”的违规行为, 因此, 相比于英文隐私政策文本, 中文隐私政策会真实、准确、完整地向个人告知应用使用个人信息的目的、方式、范围等内容. 应用目的和用户见到界面后的意图一体两面, 即是本文关注的用户意图类别.

表 1 列出了中文隐私政策中个人信息的收集使用规则的示例和分析. 典型的收集使用规则由三要素组成: 收集者 (APP 主体或软件开发套件 (SDK) 名称)、收集的个人信息和收集的目的. 但组成要素的缺少或是要素由复合结构组成是常见现象, 且相比于英文相对规则的语法和语义标记结构, 中文语法更加多样, 简单的规则匹配会带来较多漏报, 因此本文使用深度学习技术训练多特征识别模型提取目的短语.

表 1 个人数据收集使用规则示例

收集者	个人信息	目的	例子
—	手机号码	号码识别、骚扰识别	为使用号码识别和骚扰识别功能, 您需要提供通话记录中非联系人的手机号码
公安人脸SDK	摄像头、麦克风、人脸影像	人脸识别、实名认证	公安人脸SDK, 需要申请摄像头、麦克风权限, 收集您的人脸影像信息, 提供人脸识别、实名认证功能
本文	验证码	注册、登录、支付	我们会读取其中的验证码等信息用于您进行注册、登录或支付的验证

2.1.2 用户意图类别挖掘方法

● 数据集准备. 首先需要准备进行意图提取所需的数据集, 即应用程序的隐私政策文本. 本文编写爬虫脚本从各大应用商店中批量下载应用的隐私政策. 为响应国家工信部等对应用分发平台自查自纠保护用户隐私的要求, 当前许多分发平台都在应用下载详情页向用户展示应用隐私政策.

以应用宝为例, 本文通过使用 BeautifulSoup4[39]处理详情页网页结构获取标签为“privacy_button”的超链接网址, 即可访问隐私政策网页. 之后, 本文使用 jusText[40]库排除网页中如导航栏、广告等无关信息, 并将网页超文本转换为纯文本文档. 本文根据超文本标签进行分段, 将得到的段落文本内容作为之后意图提取的输入.

● 意图短语提取. 意图短语提取分为定位和提取两步. 隐私政策中除个人数据收集使用规则外还有其他内容, 比如联系方式、数据存储条款等, 为了降低无关文本带来的误报, 本文首先利用基于 BERT[41]的文本分类模型定位隐私政策中的目的语句, 之后针对收集规则语法多样、目的短语不具有明确的语法和语义标记的问题, 结合语义、词性和上下文多种特征, 通过条件随机场 (CRF) 模型[42]对目的短语进行预测.

(1) 个人使用规则定位

分析表 1 中隐私收集使用规则语句的句子组成, 个人信息相比于收集者一定存在, 而较收集目的又相对种类有限变化不多, 因此本文选择个人数据作为切入口, 定位规则语句.

本文首先根据《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》附录 C《特定类型

个人信息收集要求》[43]和随机抽取的隐私政策文件中涉及的个人信息和表述方式构建常见个人数据关键词列表, 通过正则表达式判断语句中是否含有隐私项, 定位可能的收集规则语句. 之后人工标注目的语句作为训练集, 使用 BERT 中文预训练模型对语句进行词嵌入, 再将特征送入 ALBERT[44]模型预测句子是否与表述目的相关, 一旦相关, 该语句及周围文本都用于后续目的短语提取, 以避免遗漏具有特殊格式、文本不连续的收集规则.

(2) 目的短语提取

针对收集规则语法多样、目的短语不具有明确的语法和语义标记的问题, 本文使用多特征的命名实体识别模型对使用规则中目的短语进行预测和提取, 其具体流程如图 2 所示.

本文首先使用 BIO 格式对使用规则中的收集者、个人信息和目的这 3 类实体进行标注. O 字符表示语句中不相关的部分, B-Subject、I- Subject、B-person-Data、I-personData、B-purpose、I-purpose 这 6 个标注属性分别代表收集主体、收集的个人隐私以及收集的目的. B-和 I-开头的标注属性代表命名实体的开始部分和其他部分, 以实现对相邻的实体进行区分, 处理多个组成成分的问题. 之后, 通过训练的多特征命名实体识别模型预测语句中的⟨收集者, 隐私项, 目的⟩三元组.

具体地, 对于输入文本, 本文首先使用 NLP 技术进行常见文本预处理, 比如去除停用词等, 再对文本成分进行提取. 本文提取模型以 BERT 模型为基础, 通过自定义 BERT 嵌入层, 使用语义、上下文和词性这 3 个维度特征进行分类预测. 对于语义特征, 本文通过将词汇映射到向量空间进行提取, 对于上下文特征, 本文通过位置编码保存各个词在整个句子中的绝对位置, 以此提取各个词之间的上下文特征. 考虑到动词或名词短语可能更常用于表示目的, 本文使用百度 LAC 工具[45]提取词性特征. 最后, 语义特征、上下文特征和词性特征相加生成的词嵌入向量输入后续 BERT 模型, 使用线性层将 BERT 模型输出转换为每个实体标签的分数.

训练阶段, 使用 CRF 模块计算标签损失, 反向传播优化模型参数; 推理阶段, 使用 CRF 解码预测标签序列. CRF 是一种常用于序列预测任务的模型, 它能够考虑到标签之间的依赖性, 适合于命名实体识别任务. 由此, 本文得到了⟨收集者, 隐私项, 目的⟩三元组, 实现了对收集使用规则中目的短语的提取.

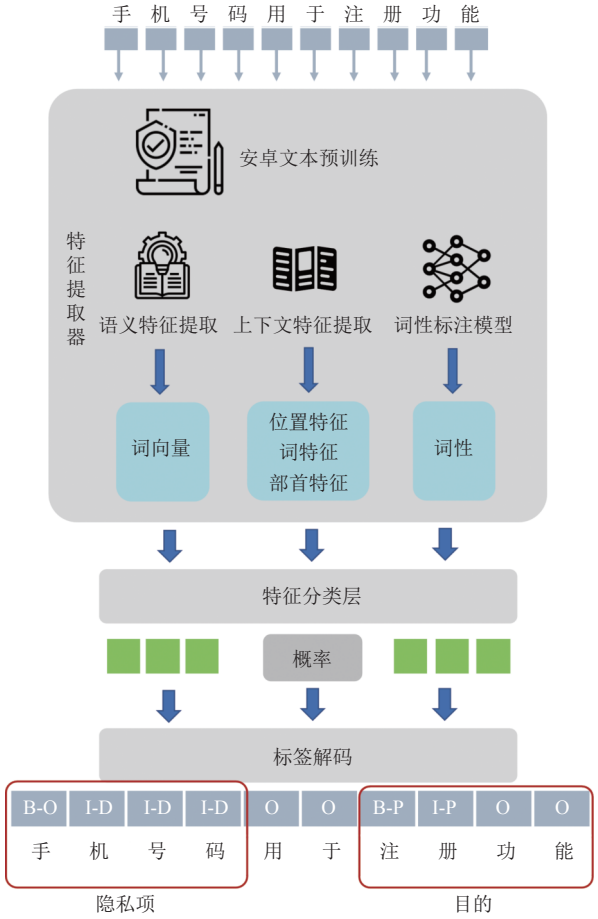


图 2 目的短语提取方法架构

● 意图主题聚类. 因为意图常有多种表述方式, 比如美颜、美妆、美化不同词汇均表示美颜意图, 得到目的短语后, 本文再通过聚类对短语进行合并, 又因为上阶段分词和预测都可能导致提取的关键词范围具有差异, 聚类文本中含有大量噪音, 所以本文取聚类结果中出现频率高的簇作为本文需要的细粒度用户意图, 这些簇也是常见用户意图类别.

在聚类相似度计算上, 本文选择 paraphrase-multilingual-MiniLM-L12-v2 模型[46]进行特征提取, 这是一个开源的多语言预训练模型, 可以将句子和段落映射到 384 维密集向量空间. 因为并不预先知晓意图类别的数量且上一阶段结果可能得到噪点数据, 所以本文选择使用凝聚层次聚类 (agglomerative clustering) 算法进行无监督、自动确定簇数量的聚类. 凝聚层次聚类是一种基于距离的聚类算法, 它逐步将数据点合

并为更大的簇, 直到所有数据点被包含在一个簇中, 对噪声点和初始化不敏感[47], 适用于本文未知意图类别数量且样本含有大量噪音的情况.

得到聚类结果后, 按照每簇内短语的出现次数, 由大到小进行排序, 选择前 100 个, 进行人工整理, 最终得到意图类别列表.

2.1.3 用户意图类别展示

表 2 展示了最终得到部分意图 (完整列表见 https://github.com/zyhlx/UserIntent_DataSet/blob/main/Intents.json. 本文数据集代码均开源于该仓库, 后文不再赘述). 统计意图出现频率, 最常见的意图是登录和注册, 这也和人们日常生活常识相符. 在实践中这两个意图经常合并出现, 比如“如果您是登录时未注册, 将自动注册”, 二者的数据收集方式和隐私要求也非常相似, 因此在后续意图识别工具中, 本文也将其合并识别.

表 2 意图列表概览 (部分)

意图	短语
登录	登录, 一键登录
注册	注册, 创建用户账号, 创建账号
身份认证	身份证识别认证, 身份认证, 身份证识别, 身份认证服务
分享内容	分享内容, 分享的内容, 共享, 分享, 资讯分享, 信息分享
浏览新闻	浏览新闻, 查看消息, 推荐新闻, 搜索新闻, 本地新闻, 推送新闻, 新闻推送, 新闻报道, 新闻
广告	广告软文, 推送程序化广告, 推送定向广告, 广告统计, 监察广告投放的情况, 广告内容推送, 广告投放, 定位广告, 广告业务, 广告发布, 广告推广…
位置共享	位置共享, 位置分享, 分享自己的定位
视频播放	视频观看, 视频展示, 播放视频, 视频播放

本文的意图列表区分位置共享和分享内容这两个在之前的工作中是没有被区分的意图. 单纯的内容分享并不应该发生位置信息的传输, 但在特定位置共享中, 将位置发送给共享者是合理的. 这样的意图列表说明本文的目的提取工作和聚类工作能够将名词实体的语义区别纳入考虑, 综合动作和动作涉及的个人信息全面的判断用户意图类别, 切实提供了更细粒度的隐私分析视角.

本文的意图列表广泛扩展了在之前工作中被粗略地归类为“Service”[18]的意图, 也涵盖了如广告、统计等传统工作[3]的意图, 提供了更全面、更细粒度的隐私合规检测视角. 本文的成果涵盖了人们在日常生活中遇到的大部分意图, 包括分享、新闻浏览和登录, 能帮助用户更全面地了解自己使用的应用功能时应用会收集什么数据, 以助于用户就此类收集实践的可接受性做出决策.

2.2 用户意图识别

本节介绍本文提出的基于多模态特征融合模型识别用户界面用户意图的方法.

2.2.1 用户意图识别模型架构

如图 3 所示, 本文由隐私政策文本提取得到的用户意图类别可对应实际应用中特定用户界面.

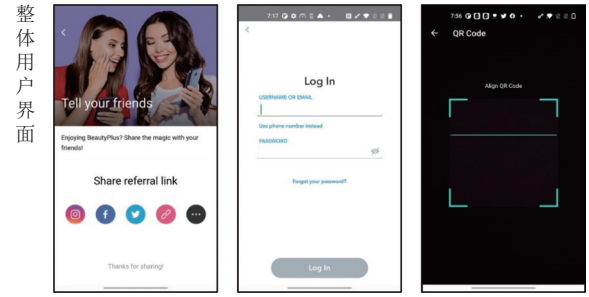


图 3 用户界面反应的用户意图

而准确地理解应用程序界面反映的用户意图并非易事. 目前传统图像识别技术在移动应用用户界面分类领域的效果不如人意, 这是因为开发人员通常以各种方式实现用户界面, 导致可供分析的视觉线索复杂多样, 简单的图像分类模型效果不佳[28], 为了能准确识别用户意图, 要寻找整体页面有效特征对界面进行建模, 同时需要考虑多模态特征的如何有效进行后续分类任务. 此外, 因为工具仅识别特定关注的用户意图, 大量非关注意图的样本 (负样本) 如何处理也需要纳入考虑.

本文由通用设计模式得到灵感, 使用结构特征对整体界面意图进行识别. 首先, 通过目标检测模型和点云模型提取界面设计结构特征, BERT 模型提取界面文本特征, 之后, 使用空间注意力机制得到融合布局特征, 再通过协同注意力机制融合文本和结构特征, 最后, 采用焦点损失函数 (focal loss)[48], 设计类别权重, 平衡不同类别样本数量. 本文用户意图识别工具包含 3 个关键组成部分: 多模态特征提取、多模态特征融合以及多分类器. 整体结构如图 4 所示, 下面将详细介绍这些组件的设计和实现.

2.2.2 特征提取

现代应用程序界面复杂多样, 用户意图相同但界面风格各异, 页面配色、明暗等视觉差距较大, 常见像

素特征不能有效表征相同用户意图的共同之处. 虽然移动应用相同用户意图界面的视觉像素和代码都有不同, 但是根据设计原则, 相似功能的应用往往采取类似的设计模式, 以降低用户学习成本, 则设计模式是移动应用程序共有特征, 因此本文尝试对构成设计的控件的位置、大小和类型进行建模.

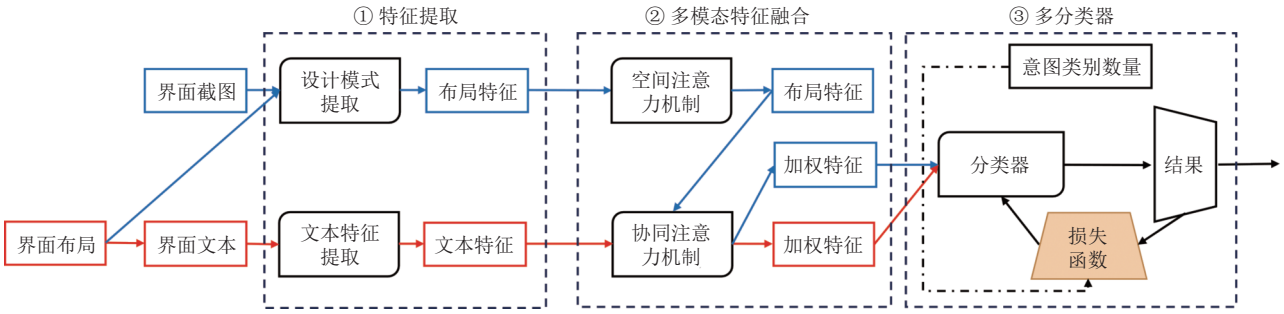


图 4 用户意图识别模型结构

常见的图片定位和类型判断的方法是使用目标检测, 通过输入图片, 目标检测模型输出目标类型、大小和位置, 得到来自图像的结构特征. 但安卓界面是以节点构成的层次结构表示, 每个节点代表一个组成控件, 具有特定属性表示节点状态, 该结构适应于三维空间中目标检测常用的点云模型.

因此本文选择 Faster R-CNN[49]这一常见目标检测模型从图像视觉对设计模型进行理解, 并创新性结合点云模型 PointNet++[50]分析安卓布局文件带语义的节点从控件构成理解设计模式作为特征补充.

为提高 Faster RCNN 模型提高控件识别效果, 本文通过提取布局文件中的控件类别和位置属性, 实现自动化的页面控件标注, 并以此为数据集预训练目标检测模型. 本文的 Faster RCNN 模型使用 MobileNetV2[51]作为骨干网络提取结构特征, 卷积层输出通道数设置为 1 280, 经过 RPN (region proposal network) 模型预测候选框的数量和位置, 再经 ROIPooling 层将结果映射为 $N\times N$ 个特征图, 表示为 $f_u=\mathrm{RCNN}(u), u\in R^{(512\times512\times3)}$, $f_u\in R^{(N\times N\times D_m)}$, u表示图像, D_m表示特征图维度, 此时为 1 280, 具体提取结构如图 5(a) 所示. 接着, 为从控件构成理解设计模式, 本文提取布局文件中节点的坐标和类型, 并额外分析节点组成的嵌套关系作为 PointNet++ 模型输入, 经全连接层得到点云提取的结构特征, 表示为 $f_v=\mathrm{PointNet{+}{+}}(v), f_v\in R^{256}$, v表示坐标类型和嵌套关系, 提取流程如图 5(b) 所示.

对于语义特征, 本文使用 BERT, 当前最流行的自然语言处理模型进行提取, 提取流程如图 5(c) 所示.

首先, 本文从布局文件提取界面文本, 主要关注节点的“text”和“content-desc”属性的内容. 节点的“text”属性表示控件显示内容, “content-desc”属性表示开发者对该控件的介绍.

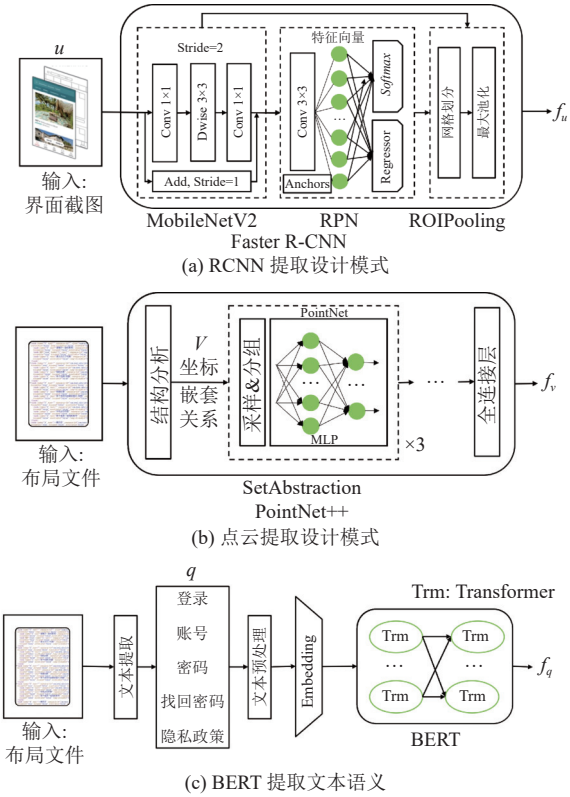


图 5 特征提取示意图

获取文本内容后, 本文对内容进行了简单的预处理, 包括对含有下划线或者驼峰命名法等不规范用词进行拆分、对缩写或具有特定格式的特殊词语进行替换、过滤非法字符并截取过长文本. 对于输入文本 q,

将 q 输入 BERT 模型得到语义向量表示 $f_q = \text{BERT}(q)$, $f_q \in R^{(K\times F)}$. K 和 F 取值由具体实现决定. 出于资源限制, 本文实验中选择较为基础的 BERT-BASE-CASED 模型[41]作为文本提取模型.

2.2.3 特征融合

为使多源空间特征能够有效互相补充, 本文搭建空间注意力模块对布局特征进行融合. 此外, 设计结构特征和文本特征维度差异较大, 为了利用好布局特征和文本特征多模态特性, 本文使用协同注意力机制融合界面设计结构特征和语义多模态特征, 让不同模态特征同时相互更新, 以综合判断界面意图, 本文特征融合模块如图 6 所示.

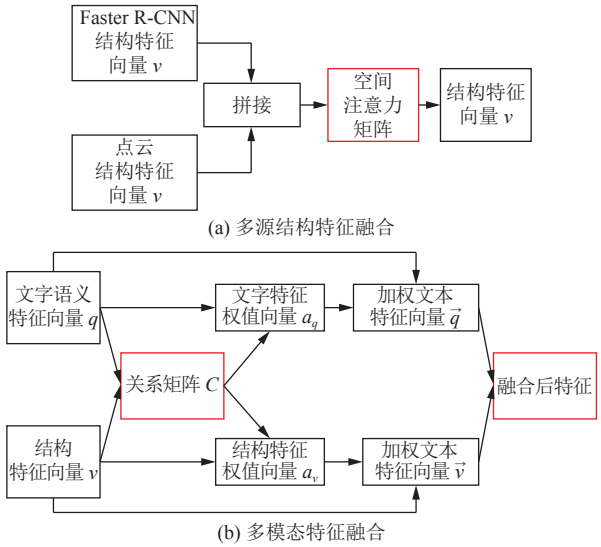


图 6 特征融合示意图

由前述可知, 一方面, 模型要使多源结构特征能够有效互相补充, 另一方面, 不同模态特征尺寸差异较大, 模型需要对结构特征向量进行压缩, 为了尽可能保留结构特征且充分利用好多源结构特征, 本文设计空间注意力机制, 使用 1×1 的卷积核计算每个区域注意力分数, 设置激活函数为*Sigmoid*函数将注意分数值归一化, 由此得到注意力图后再对原始特征图进行加权, 之后对加权特征在空间维度上进行求和并归一化, 得到每个通道的聚合特征作为融合后的结构特征$out(b,c)$. 具体公式如下:

$$A'_u = Sigmoid(Conv(f_u))$$

$$f'_u = f_u \times A'_u$$

$$sum(f'_u) = \sum_{w=1}^{W}\sum_{h=1}^{H} f_u(b,c,w,h)$$

$$sum(A) = \sum_{w=1}^{W}\sum_{h=1}^{H} A'_u(b,c,w,h)$$

$$out(b,c) = sum(f'_u)/sum(A)$$

其中, f_u表示提取的结构特征, f'_u表示注意力加权后的特征, A'_u表示注意力图, b,w,c,h分表示批次, 通道, 宽, 高, sum表示求和.

为充分利用多模态特征且尽量降低特征信息计算损耗, 本文采用协同注意力机制来对多模态特征向量进行融合以得到综合性的界面特征向量. 协同注意力机制是一种双向的注意力机制, 可以让不同模态特征同时相互更新, 帮助模型在不同类别判定中察觉有利于分类的特征, 提高召回率和精确度.

首先, 定义关系矩阵 C、u 代表文字特征向量, v 表示结构特征, 具体公式如下:

$$C = \tanh(q^{\mathrm{T}} W_c u)$$

其中, W_c为权重.

之后使用关系矩阵 C 分别和特征向量计算得到特征权重, 以布局权值向量a^v为例, 生成公式为:

$$H^v = \tanh(W_v v + (W_u u)C^{\mathrm{T}})$$

$$a^v = Softmax(W_{hv}^{\mathrm{T}} H^u)$$

得到权值向量后, 再通过加权计算得到特征融合后的布局特征向量. 以布局特征 v 为例, 得到加权语义特征的布局特征$\vec{v}$, 公式如下:

$$\vec{v} = \sum_{n=1}^{N} a_n^v v_n$$

其中, N 表示区域数量.

2.2.4 分类模型

得到融合特征后, 本文构建多分类器以识别用户意图类型, 分类器由一个全连接层和 *Softmax* 函数组成, 激活函数为 tanh, 反向传播优化器为 Adm. 为了缓解不均匀的样本数量带来的问题, 在损失计算时, 本文使用 focal loss, 计算每个类别的训练数据占比, 以为不同类别设置不同的损失权重. 损失计算公式如下: 使用 $loss_{\text{focal}}(p,y)$表示对于标签为 y 的样本, 模型预测概率向量为 p 时的损失, n 为训练集样本总数, $m(y)$ 为训练集中 y 标签样本的数量.

$$loss_{\text{focal}}(p,y) = -\alpha_y(1-p_y)^2\log(p_y)$$

$$\alpha_y = 1 - m(y)/n$$

3 整体框架

本节将评估本文细粒度用户意图类型提取方法和实时细粒度用户意图识别工具的有效性. 本文的实验设计主要为回答以下两个问题.

RQ1: 细粒度用户意图类型提取方法在隐私政策中提取短语的准确度如何?

RQ2: 用户意图识别模型有效性如何? 针对移动应用界面特点设计的多模态特征和特征融合方法是否具有优势?

3.1 实验环境

本文实验使用的环境具体配置如表 3 所示.

表 3 实验环境

实验环境	具体配置
操作系统	Ubuntu 22.04
显卡	A100
内存	256 GB
编程语言	Python 3.9
深度学习框架	PyTorch 2.2.1

3.2 用户意图提取方法有效性评估实验

● 数据集准备. 本文按照第 2.1.2 节所述隐私政策获取方法, 从应用宝和小米商店批量爬取隐私政策, 通过个人隐私列表定位个人收集使用规则语句. 按照第 2.1.2 节所述实体标注格式对收集到的 3 714 个隐私收集使用规则语句进行标注, 在收集者标注时, 本实验只考虑第三方 SDK, 最终标注得到标签和对应数量如表 4 所示. 随机选择 100 个收集使用规则语句作为测试集, 其余作为训练集.

表 4 用户意图提取方法实验数据集

标签	数量	标签	数量
B-purpose	4 143	I-purpose	16 127
B-Subject	2 405	I-Subject	10 602
B-personData	8 389	I-personData	32 396

● 评估指标. 本实验采用精确度 (*Precision*)、召回率 (*Recall*) 和 *F*1 分数 (*F*1-score) 作为指标评估收集者、个人信息或意图这 3 类标注预测结果.

精确度 (*Precision*) 表示预测正确的正样本占所有预测为正例的样本数的百分比, 计算公式如下:

$$Precision = \frac{TP}{TP+FP}$$

召回率 (*Recall*) 表示预测正确的正样本占所有正样本数的百分比, 计算公式如下:

$$Recall = \frac{TP}{TP+FN}$$

*F*1 分数用以综合表示精确度和召回率的效果, 计算公式如下所示:

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} = \frac{2\times TP}{2TP+FP+FN}$$

其中, *TP* (true positive) 表示标签预测正确的正样本, *TN* (true negative) 表示标签预测正确的负样本, *FP* (false positive) 表示标签预测错误的负样本, *FN* (false negative) 表示标签预测错误的正样本. 在本多分类场景中, 对每类命名实体的预测结果分别进行评估时, 属于该命名实体的样本定义为正样本, 而非该实体的短语定义为负样本.

● 实验结果. 本实验结果如表 5 所示. 3 类命名实体识别的精确度、召回率及 *F*1 分数均达到了 0.8 以上. 具体分析 3 类实体识别效果, 意图类别识别效果相对较差, 这是因为意图类型多样且语法形式灵活多变, 对比类型有限的个人信息和经常是主语且有 SDK 标志的收集者, 识别难度较高.

表 5 命名实体识别效果

实体	*Precision*	*Recall*	*F*1
收集者	0.98	0.94	0.96
个人信息	0.97	0.93	0.95
意图	0.83	0.83	0.83

3.3 用户意图识别模型有效性评估实验

● 数据集准备. 本文邀请两位隐私保护研究人员手动扩展特定用户意图类别的数据到 150 份样本, 一份样本包含屏幕截图和对应的布局文件. 数据集准备分为获取样本和样本标注两个阶段. 本文人工手动触发移动应用, 使用安卓自动化测试工具 ADB 在触发过程中收集屏幕截图和相应的布局文件. 数据收集完后进入标注阶段, 本文先向隐私保护研究人员讲解意图类型及其关键字列表, 之后, 两位隐私保护研究人员依照用户意图类型共识对截图表示的用户意图进行标签, 每个类别标注完成后, 进行讨论, 解决可能存在的分歧. 如果他们无法达成共识, 本文会邀请第 3 位隐私保护研究人员做出最终决定.

本文确保每个用户意图类别中数据来自至少 30 个不同的应用, 以尽可能保证样本的多样性. 最后, 本文总共收集到 3 417 样本, 14 个用户意图类型每类约 150 张样本, 1 153 张为其他 (other) 负样本. 随机按

照类别样本数量比抽取样本作为测试集, 共抽取680 份样本, 平均每个有效类别 30 张, 其余样本构成训练集. 在数据集划分时, 本文确保训练集中每类别样本来源应用与测试集中对应类别样本来源应用没有交集, 即特定用户意图类别训练集或者验证集中样本来源应用不会重叠, 以保证测试的公正.

● 评估指标. 本实验采用精确度 (*Precision*)、召回率 (*Recall*) 和 *F*1 分数 (*F*1-score) 作为多分类模型的评估指标对每个用户意图类别的分类效果进行评估, 公式同第 3.2 节所列, 在本实验中, 对每类意图分类效果进行评估时, 该意图样本定义为正样本, 非该类别样本定义为负样本.

为了衡量多分类模型的整体效果, 本文使用宏观精确度 (*MacroPrecision*)、宏观召回率 (*MacroRecall*) 和宏观 *F*1 分数 (*MacroF*1) 表示模型多个类别分类的平均情况. 具体公式如下所示, 其中 N 表示类别的总数量. $Precision_i$、$Recall_i$、$F1_i$ 分别表示第 i 个类别的准确度、召回率和 *F*1 分数.

$$MacroPrecision = \frac{1}{N}\sum_{i=1}^{N} Precision_i$$

$$MacroRecall = \frac{1}{N}\sum_{i=1}^{N} Recall_i$$

$$MacroF1 = \frac{1}{N}\sum_{i=1}^{N} F1_i$$

3.3.1 模型整体有效性评估

本文用户意图识别模型在测试集上的精确度、召回率和 *F*1 分数如表 6 所示. 模型整体宏观精确度和召回率为 0.80 和 0.83. 其中 14 个类别的 *F*1 分数大于等于 0.7, 说明本文多分类模型具有良好的检测效果.

表 6 用户意图识别模型和相关技术有效性实验结果

类别	本文工具			不同特征对比模型						不同特征融合方式					
				Text_Only		Image_Only		PC_Only		Add			Concat		
	Precision	*Recall*	*F*1	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*F*1	*Precision*	*Recall*	*F*1
扫码	0.78	0.97	**0.87**	0.18	0.23	0.77	0.67	0.24	0.33	0.76	0.63	0.69	0.87	0.67	0.76
文件管理	0.81	0.73	**0.77**	0.62	0.27	0.73	0.53	0.43	0.50	0.85	0.77	0.81	0.71	0.73	0.72
登录	0.88	0.87	**0.88**	0.68	0.70	0.77	0.90	0.67	0.58	0.77	0.82	0.79	0.80	0.88	0.84
地图	1.00	0.80	**0.89**	0.55	0.73	0.73	0.63	0.48	0.33	0.78	0.70	0.74	0.79	0.73	0.76
编辑	0.63	0.80	**0.70**	0.38	0.10	0.64	0.60	0.46	0.57	0.65	0.73	0.69	0.51	0.67	0.58
音乐	0.77	1.00	0.87	0.89	0.57	0.76	0.93	0.88	0.77	0.74	0.93	0.82	0.90	0.93	0.92
浏览新闻	0.72	0.90	**0.80**	0.56	0.62	0.67	0.83	0.38	0.28	0.62	0.86	0.72	0.63	0.83	0.72
在线聊天	0.83	0.63	**0.72**	0.72	0.60	0.65	0.50	0.24	0.27	0.75	0.30	0.43	0.85	0.57	0.68
拍摄	0.92	0.77	**0.84**	0.45	0.61	0.79	0.87	0.36	0.71	0.83	0.77	0.80	0.92	0.74	0.82
发帖	0.93	0.43	0.59	0.86	0.20	0.56	0.33	0.50	0.17	0.64	0.23	0.34	0.71	0.57	0.63
搜索	0.67	0.80	0.73	0.53	0.30	0.72	0.87	0.43	0.50	0.79	0.87	0.83	0.77	0.77	0.77
分享	0.77	0.77	0.77	0.83	0.50	0.95	0.70	0.58	0.47	0.9	0.87	0.89	0.96	0.77	0.86
视频播放	0.87	0.67	**0.76**	0.75	0.40	0.49	0.77	0.26	0.30	0.6	0.83	0.70	0.64	0.77	0.70
天气	1.00	1.00	**1.00**	0.77	0.34	0.97	1.00	1.00	0.90	1.00	1.00	1.00	0.97	0.97	0.97
其他	0.90	0.93	0.92	0.53	0.76	0.94	0.92	0.90	0.87	0.90	0.93	0.92	0.92	0.94	0.93
宏观平均	0.83	0.80	0.81	0.62	0.46	0.74	0.74	0.52	0.50	0.77	0.75	0.74	0.80	0.77	0.78

具体分析模型对每意图类别的分类结果, 本文发现相比于其他类别, 本文模型发帖类别的召回率较低, 编辑、搜索类别的精确度较低. 在分析漏报和误报后, 可能的原因是, 发帖类别中存在的发帖形式多样, 比如图片、投票、公式、视频、纯文本等各种样式, 如图 7 所示, 但当前样本量较小, 训练集并未覆盖, 而文本上因为动态测试随机输入, 导致有效文本信息稀释, 导致漏报, 未来考虑增加训练集样本.

编辑类别和搜索类别同样存在有效文本信息较少的问题, 编辑类别中大量样本界面上不含有效文本信息, 仅能通过图标标识按键功能, 而单纯结构特征又和拍照比较类似, 造成误报, 使精确度不佳, 如图 8(a) 所示, 右图即为错误分为编辑类别的拍照样本, 其结构与左图实际编辑样本非常相似. 未来可以通过引入图标特征提高召回率. 搜索类别由搜索框加列表形式呈现的搜索结果组成, 如图 8(b) 所示, 大量无关搜索结果使得有效文本被稀释, 文本框特征也和登录或在线聊天类别等相似, 导致搜索类别被错分为以上类别,

未来可以通过词频文档过滤等方式过滤无关文本提高精确度.

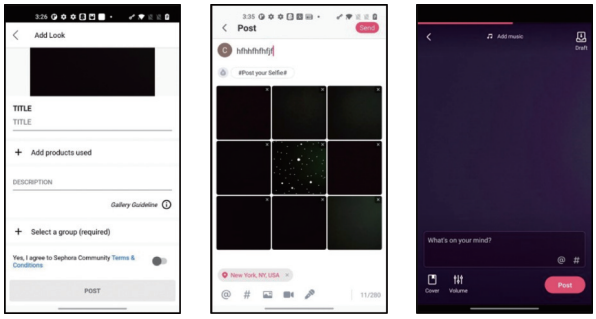
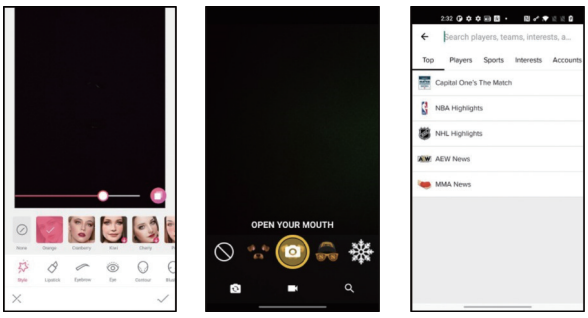


图 7 发帖: 结构多样导致漏报



(a) 编辑: 与拍照结构类似 (b) 搜索示意图

图 8 低精确度意图示意图

3.3.2 联合特征学习的有效性

为了探究本文设计的特征和特征融合方案在用户意图识别任务中的有效性, 本文设计了如下对比模型进行消融实验.

(1) Text_Only: 仅使用 BERT 模型提取文本特征进行分类.

(2) Image_Only: 仅使用 RCNN 模型提取结构特征进行分类.

(3) PC_Only: 仅使用点云模型提取结构特征进行分类.

(4) Add: 使用多模态特征直接相加得到的特征进行分类.

(5) Concat: 使用多模态特征在通道维度上的合并的特征进行分类.

本文对比模型在测试集上的精确度、召回率和 $F1$ 分数如表 6 所示. 由实验数据可知, 本文模型在准确度、召回率和 $F1$ 分数上均优于其他对比模型, 以 $F1$ 分数为例, 本文模型的平均 $F1$ 分数分别比 Text_Only、Image_Only、PC_Only、Add 和 Concat 模型高了 31%、8%、31%、7%、3%, 单特征模型中效果最好的 Image_Only 模型以及效果最好的 Concat 特征融合方案仍分别有 6 个和 4 个用户意图类别准确率低于 0.7.

● 特征有效性. 比较文本特征和结构特征模型效果, 观察目前实验数据可以发现, 在本实验中结构特征效果优于文本特征. 本文认为其可能原因是, 本实验选择的大部分用户意图类别使用结构文本特征可以更直观预测用户意图, 仅依靠文本特征不能有效识别的意图超过一半. 具体分析文本特征不佳的原因, 首先, 一部分意图的本身文本信息就较少如编辑、扫码、文件管理, 可见图 8(a), 其次则是部分意图界面中具有许多和意图本身无关的动态文本信息, 比如音乐、天气、视频、搜索、发帖, 音乐中歌曲、天气中的地点和温度数字、视频、搜索和发帖的文章内容都会造成实际有效文本的稀释, 可见图 8(b), 因此寻找除文本信息以外的其他有效信息是非常有必要的, 而本文使用结构特征由实验可证属于有效特征.

比较两类结构特征模型的效果, 使用目标检测模型提取特征的 Image_Only 模型效果要优于使用点云模型的 PC_Only 模型. 点云模型本身多用于 3D 物体, 本文实现的点云模型仅考虑控件坐标, 效果弱于考虑类型和大小的 Image_Only 模型符合预期. 本文通过将安卓布局结构转化为点云模型可用的物体坐标, 使用点云模型提取控件结构特征补充目标检测模型图像结构特征. PC_Only 模型在 15 类别分类中平均精确度和召回率达到 0.52 和 0.50, $F1$ 分数持平 Text_Only 模型, 证明点云模型提取的结构特征具备有效性.

● 特征融合有效性. 本文的特征融合模型在 14 个具体意图类别中的 10 个意图 $F1$ 分数都最优 (表 6 中粗体). 这意味着本文的特征融合方案确实有助于增强模型对相对更难的用户意图的预测. 扫码、地图和编辑类别在所有其他的特征组合方法的召回率都很低, 其中最好仅 0.73, 但在本文特征融合模型上却取得了不错的效果, 最低也达到 0.8. 扫码、地图和编辑类别的共同点是结构特征和文本特征都并不显著, 通过单一特征并不能准确识别意图. 以扫码类别为例, 扫码和编辑均和拍摄类别的结构相似, 且文本信息都较少, 仅有通过综合结构特征和文本特征才能有效判断出该类别. 在实际特征融合过程中如果类似 Concat 模型仅直接拼接不同特征向量, 会使尺寸较小的特征被大尺寸特征稀释, 扫码类别的召回率也和特征尺寸最大的 Image_Only 相同; 如果如 Add 模型压缩大尺寸特征或

放大小尺寸特征也会破坏特征中所携带的信息, 表现在实际模型效果中, Add 模型扫码类别召回率低于 Image_Only 模型.

4 结论与展望

当前应用级别粗粒度用户意图检测已不能满足日益复杂的移动应用的安全检测需求. 而仅依赖图标等界面组件的用户意图进行的隐私风险评估由于现实中图标的不当设计和错误使用实际有效性受限. 为了解决上述问题, 本文首先从中文隐私政策中提取总结出常见的适用于隐私合规判断的细粒度用户意图列表; 再结合移动应用界面设计特点, 设计并实现了多模态特征融合的多分类模型对整个移动界面反应的用户意图进行识别. 实验评估表明, 本文隐私政策意图提取工具和用户意图识别工具能准确提取用户意图并正确理解当前界面用户意图.

参考文献

1 中华人民共和国工业和信息化部. 工业和信息化部关于2024年第一季度电信服务质量的通告. https://www.miit.gov.cn/zwgk/zcwj/wjfb/tg/art/2024/art_6e0a9c81d2d24e71bd017cb2b61820a0.html. [2024-04-26].

2 Qu ZY, Rastogi V, Zhang XY, *et al*. AutoCog: Measuring the description-to-permission fidelity in Android applications. Proceedings of 2014 ACM SIGSAC Conference on Computer and Communications Security. Scottsdale: ACM. 2014. 1354–1365. [doi: 10.1145/2660267.2660287]

3 Pandita R, Xiao XS, Yang W, *et al*. WHYPER: Towards automating risk assessment of mobile applications. Proceedings of the 22nd USENIX Security Symposium. Washington: USENIX Association, 2013. 527–542.

4 Gorla A, Tavecchia I, Gross F, *et al*. Checking APP behavior against APP descriptions. Proceedings of the 36th International Conference on Software Engineering. Hyderabad: ACM, 2014. 1025–1035. [doi: 10.1145/2568225.2568276]

5 Wang R, Wang ZB, Tang BX, *et al*. SmartPI: Understanding permission implications of Android APPs from user reviews. IEEE Transactions on Mobile Computing, 2020, 19(12): 2933–2945. [doi: 10.1109/TMC.2019.2934441]

6 Xi SQ, Yang S, Xiao XS, *et al*. DeepIntent: Deep icon-behavior learning for detecting intention-behavior discrepancy in mobile APPs. Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. London: ACM, 2019. 2421–2436. [doi: 10.1145/3319535.3363193]

7 Xiao XS, Wang XY, Cao ZH, *et al*. IconIntent: Automatic identification of sensitive UI widgets based on icon classification for Android APPs. Proceedings of the 41st International Conference on Software Engineering. Montreal: IEEE Press, 2019. 257–268. [doi: 10.1109/ICSE.2019.00041]

8 Liu J, He DJ, Wu DY, *et al*. Correlating UI contexts with sensitive API calls: Dynamic semantic extraction and analysis. Proceedings of the 31st International Symposium on Software Reliability Engineering. Coimbra: IEEE Press, 2020. 241–252. [doi: 10.1109/ISSRE5003.2020.00031]

9 Nguyen TT, Nguyen DC, Schilling M, *et al*. Measuring user perception for detecting unexpected access to sensitive resource in mobile APPs. Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security. Hong Kong: ACM, 2021. 578–592. [doi: 10.1145/3433210]

10 Zhang SK, Lei HW, Wang YP, *et al*. How Android APPs break the data minimization principle: An empirical study. Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering. Luxembourg: IEEE Press, 2023. 1238–1250. [doi: 10.1109/ASE56229.2023.00141]

11 Li LL, Wang RF, Zhan X, *et al*. What you see is what you get? It is not the case! Detecting misleading icons for mobile applications. Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis. Seattle: ACM, 2023. 538–550. [doi: 10.1145/3597926.3598076]

12 Enck W, Gilbert P, Han S, *et al*. TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. ACM Transactions on Computer Systems, 2014, 32(2): 5. [doi: 10.1145/2619091]

13 Arzt S, Rasthofer S, Fritz C, *et al*. FlowDroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android APPs. ACM SIGPLAN Notices, 2014, 49(6): 259–269. [doi: 10.1145/2666356.2594299]

14 Yu L, Chen JC, Zhou H, *et al*. Localizing function errors in mobile APPs with user reviews. Proceedings of the 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Luxembourg: IEEE Press, 2018. 418–429. [doi: 10.1109/DSN.2018.00051]

15 Hu YY, Wang HY, Ji TT, *et al*. CHAMP: Characterizing undesired APP behaviors from user comments based on

market policies. Proceedings of the 43rd International Conference on Software Engineering. Madrid: IEEE Press, 2021. 933–945. [doi: 10.1109/ICSE43902.2021.00089]

16 Slavin R, Wang XY, Hosseini MB, *et al*. Toward a framework for detecting privacy policy violations in Android application code. Proceedings of the 38th International Conference on Software Engineering. Austin: ACM, 2016. 25–36. [doi: 10.1145/2884781.2884855]

17 Wang XY, Qin X, Hosseini MB, *et al*. GUILeak: Tracing privacy policy claims on user input data for Android applications. Proceedings of the 40th International Conference on Software Engineering. Gothenburg: ACM, 2018. 37–47. [doi: 10.1145/3180155.3180196]

18 Andow B, Mahmud SY, Whitaker J, *et al*. Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with POLICHECK. Proceedings of the 29th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2020. 56.

19 Bui D, Yao Y, Shin KG, *et al*. Consistency analysis of data-usage purposes in mobile APPs. Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2021. 2824–2843. [doi: 10.1145/3460120.3484536]

20 Andow B, Mahmud SY, Wang WY, *et al*. Policylint: Investigating internal privacy policy contradictions on Google play. Proceedings of the 28th USENIX Conference on Security Symposium. Santa Clara: USENIX Association, 2019. 585–602.

21 Avdiienko V, Kuznetsov K, Rommelfanger I, *et al*. Detecting behavior anomalies in graphical user interfaces. Proceedings of the 39th International Conference on Software Engineering Companion. Buenos Aires: IEEE Press, 2017. 201–203. [doi: 10.1109/ICSE-C.2017.130]

22 Yang ZM, Yang M, Zhang Y, *et al*. AppIntent: Analyzing sensitive data transmission in Android for privacy leakage detection. Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. Berlin: ACM, 2013. 1043–1054. [doi: 10.1145/2508859.2516676]

23 Fu H, Zheng ZZ, Das AK, *et al*. FlowIntent: Detecting privacy leakage from user intention to network traffic mapping. Proceedings of the 13th Annual IEEE International Conference on Sensing, Communication, and Networking. London: IEEE Press, 2016. 1–9. [doi: 10.1109/SAHCN.2016.7732993]

24 Pan X, Cao YZ, Du XC, *et al*. FlowCog: Context-aware semantics extraction and analysis of information flow leaks in Android APPs. Proceedings of the 27th USENIX Security Symposium. Baltimore: USENIX Association, 2018. 1669–1685. [doi: 10.5555/3277203.3277328]

25 Huang JJ, Zhang XY, Tan L, *et al*. AsDroid: Detecting stealthy behaviors in Android applications by user interface and program behavior contradiction. Proceedings of the 36th International Conference on Software Engineering. Hyderabad: ACM, 2014. 1036–1046. [doi: 10.1145/2568225.2568301]

26 Li YX, Feng RT, Chen S, *et al*. IconChecker: Anomaly detection of icon-behaviors for Android APPs. Proceedings of the 28th Asia-Pacific Software Engineering Conference. Taipei, China: IEEE Press, 2021. 202–212. [doi: 10.1109/APSEC53868.2021.00028]

27 Qi CH, Shao S, Guo YH, *et al*. An efficient method for analyzing widget intent of Android system. Proceedings of the 9th International Conference on Communications and Broadband Networking. Shanghai: ACM, 2021. 78–85. [doi: 10.1145/3456415.3456428]

28 Malviya VK, Tun YN, Leow CW, *et al*. Fine-grained in-context permission classification for Android APPs using control-flow graph embedding. Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering. Luxembourg: IEEE Press, 2023. 1225–1237. [doi: 10.1109/ASE56229.2023.00056]

29 Lin JL, Liu B, Sadeh NM, *et al*. Modeling users' mobile APP privacy preferences: Restoring usability in a sea of permission settings. Proceedings of the 10th Symposium on Usable Privacy and Security. Menlo Park: USENIX Association, 2014. 199–212.

30 Han S, Jung J, Wetherall D. A study of third-party tracking by mobile APPs in the wild. Technical Report UW-CSE-12-03-01, Washington: University of Washington.

31 Jain V, Gupta SD, Ghanavati S, *et al*. PAcT: Detecting and classifying privacy behavior of Android applications. Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks. San Antonio: ACM, 2022. 104–118. [doi: 10.1145/3507657.3528543]

32 陈瀚, 赵春蕾, 蒋昊达, 等. 基于融合模型与语义网络的APP用户意图识别研究. 计算机工程, 2024, 50(8): 50–63. [doi: 10.19678/j.issn.1000-3428.0068206]

33 姜超. 基于语义的用户意图领域多分类算法分析 [硕士学位论文]. 武汉: 武汉大学, 2018.

34 贺国秀, 张晓娟. 查询意图自动分类的方法改进探讨. 数字图书馆论坛, 2018(1): 53–60.

35 马莹雪. 基于用户意图和时序偏好特征的兴趣点推荐方法研究 [博士学位论文]. 北京: 北京科技大学, 2022. [doi: 10.26945/d.cnki.gbjku.2022.000251]

36 钱忠胜, 张丁, 李端明, 等. 结合用户共同意图及社交关系的群组推荐方法. 计算机科学与探索, 2024, 18(5): 1368–1382.

37 杜思佳. 基于深度神经网络的法律咨询用户意图理解研究与实现 [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2019. [doi: 10.27061/d.cnki.ghgdu.2019.002955]

38 张春英, 李春虎, 兰思武. 基于多粒度特征融合的用户意图分类. 华北理工大学学报 (自然科学版), 2019, 41(3): 127–134.

39 Richardson L. Beautiful soup. https://www.crummy.com/software/BeautifulSoup/. [2024-01-17].

40 Pomikálek J. Removing boilerplate and duplicate content from Web corpora [Ph.D. Thesis]. Brno: Masaryk University, 2011.

41 Devlin J, Chang MW, Lee K, *et al*. BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: 10.18653/V1/N19-1423]

42 Lafferty JD, McCallum A, Pereira FCN. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the 18th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2001. 282–289.

43 国家市场监督管理总局, 中国国家标准化管理委员会. GB/T 41391-2022 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. 北京: 中国标准出版社, 2022.

44 Lan ZZ, Chen MD, Goodman S, *et al*. ALBERT: A lite BERT for self-supervised learning of language representations. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.

45 Jiao ZY, Sun SQ, Sun K. Chinese lexical analysis with deep Bi-GRU-CRF network. arXiv:1807.01882, 2018.

46 Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: ACL, 2019. 3982–3992.

47 Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? Journal of Classification, 2014, 31(3): 274–295. [doi: 10.1007/s00357-014-9161-z]

48 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017. 2980–2988.

49 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

50 Qi CR, Yi L, Su H, *et al*. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 5105–5114.

51 Sandler M, Howard A, Zhu ML, *et al*. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.

(校对责编: 张重毅)