



计算机科学

COMPUTER SCIENCE

基于卷积神经网络的 APP 用户行为分析方法

陈泳全, 姜瑛

引用本文

陈泳全, 姜瑛. 基于卷积神经网络的 APP 用户行为分析方法[J]. 计算机科学, 2022, 49(8): 78-85.

CHEN Yong-quan, JIANG Ying. Analysis Method of APP User Behavior Based on Convolutional Neural Network[J]. Computer Science, 2022, 49(8): 78-85.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于注意力机制的医学影像深度哈希检索算法](#)

Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism

计算机科学, 2022, 49(8): 113-119. <https://doi.org/10.11896/jsjcx.210700153>

[基于图卷积神经网络的文本分类方法研究综述](#)

Review of Text Classification Methods Based on Graph Convolutional Network

计算机科学, 2022, 49(8): 205-216. <https://doi.org/10.11896/jsjcx.210800064>

[基于可变形图卷积的点云表征学习](#)

Deformable Graph Convolutional Networks Based Point Cloud Representation Learning

计算机科学, 2022, 49(8): 273-278. <https://doi.org/10.11896/jsjcx.210900023>

[全局信息引导的真实图像风格迁移](#)

Photorealistic Style Transfer Guided by Global Information

计算机科学, 2022, 49(7): 100-105. <https://doi.org/10.11896/jsjcx.210600036>

[基于 DNGAN 的磁共振图像超分辨率重建算法](#)

Super-resolution Reconstruction of MRI Based on DNGAN

计算机科学, 2022, 49(7): 113-119. <https://doi.org/10.11896/jsjcx.210600105>

基于卷积神经网络的 APP 用户行为分析方法

陈泳全 姜 瑛

云南省计算机技术应用重点实验室 昆明 650500

昆明理工大学信息工程与自动化学院 昆明 650500

(1977162878@qq.com)

摘 要 随着移动互联网的快速发展,智能终端已经成为人们日常生活和工作中不可或缺的一部分。在使用智能终端的过程中,会产生大量的 APP 操作过程记录,对用户 APP 操作过程记录进行分析,可以获取到操作过程记录中用户的行为,从而获得用户的行为模式,以帮助开发人员有针对性地维护和改进 APP 软件。现有的用户行为分析偏向操作分析,缺少对用户操作的行为提取,因此提出了一种基于卷积神经网络的 APP 用户行为分析方法。该方法首先进行 APP 操作分析,提取出原始 APP 操作记录信息中的用户操作;然后挖掘 APP 操作与 APP 用户行为之间的关联性,构建 APP 操作与 APP 用户行为之间的相似度矩阵;最后提取 APP 用户行为。实验结果表明,该方法能够有效地提取和识别 APP 操作过程记录中用户的行为,有助于深层次地挖掘 APP 用户行为的含义。

关键词: APP 软件;用户 APP 操作过程记录;APP 用户行为;操作分析;卷积神经网络

中图法分类号 TP311

Analysis Method of APP User Behavior Based on Convolutional Neural Network

CHEN Yong-quan and JIANG Ying

Yunnan Key Lab of Computer Technology Application, Kunming 650500, China

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Abstract With the rapid development of mobile Internet, smart terminal has become an indispensable part of people's daily life and work. In the process of using smart terminal, a large number of APP operation process records will be generated. By analyzing the user's APP operation process records, the user's behaviors in the operation process and the user's behavior pattern can be obtained, which can help developers maintain and improve the APP software. Existing user behavior analysis is biased towards operation analysis and the behaviors extraction for user's operation is lacked. An APP user behaviors analysis method based on convolution neural network is proposed. At first, the APP operations are analyzed, and the user operations in the original APP operation record information are extracted. Then the correlation between APP operation and user's behaviors is mined, and the similarity matrix between APP operations and APP user's behaviors is constructed. Finally, the behaviors of users will be extracted. Experiments show that this method can extract and identify the user's behaviors in the records of APP operation process effectively, which will be helpful to explore the deep meaning of user's behaviors.

Keywords APP software, Operation process record of user APP, APP user behavior, Operation analysis, Convolutional neural network

1 引言

近年来,智能终端已经成为人们生活和工作中的不可或缺

的一部分^[1]。APP 是智能终端提供服务的载体和具体表现形式,用户可以通过智能终端上的 APP 实现日常生活和工作的许多目的,如购物、阅读、学习、社交、办公等。用户根据

到稿日期:2021-07-12 返修日期:2021-12-10

基金项目:国家自然科学基金(62162038,61462049,61063006,60703116);国家重点研发计划(2018YFB1003904);云南省应用基础研究计划重点项目(2017FA033);云南省计算机技术应用重点实验室开放基金(2020101)

This work was supported by the National Natural Science Foundation of China(62162038,61462049,61063006,60703116), National Key Research and Development Program of China(2018YFB1003904), Key Project of Yunnan Applied Basic Research(2017FA033) and Open Foundation of Yunnan Key Laboratory of Computer Technology Application(2020101).

通信作者:姜瑛(jy_910@163.com)

自己的需求、使用习惯、兴趣爱好等在智能终端上安装和使用 APP,这使得不同智能终端上安装的 APP 和用户日常使用智能终端时产生的使用行为存在很大的差异。即使是同一个 APP,不同用户的使用行为也会存在差异,这些差异体现在使用的频率和时长等方面。用户日常使用智能终端会产生大量的 APP 操作过程记录,由于一台智能终端一般由一个用户使用,因此这些 APP 操作过程记录体现的是用户真实的使用行为,对用户的 APP 操作过程记录信息进行挖掘和分析,可提取出数据背后隐藏的行为模式和使用习惯等用户个人行为信息^[2-3],有助于理解用户的行为、使用习惯、兴趣爱好,从而为完善 APP 提供指导意见,帮助开发人员有针对性地完善和改进 APP^[4-5]。

现有的用户行为分析方法大多是针对具体的某一软件或某一类型软件的用户行为进行分析^[6-7]。此外,目前部分研究中分析的是用户的操作,而不是用户行为^[8-10]。本文认为,行为是对操作的进一步抽象,操作是行为的具体体现。现有的分析方法仅对操作进行分析,并未明确其中的用户行为,无法有效地抽取操作的共性;操作的行为提取则是通过对操作进行分析,来明确其中的用户行为,对操作进行进一步的抽象,抽取操作的共有属性,以便于对用户行为的分析。针对上述问题,本文提出了一个 APP 用户行为分析方法,该方法基于卷积神经网络,能有效地提取和识别用户 APP 操作过程记录信息的特征,从而提取出 APP 操作过程记录信息中用户的行为,有助于深层次挖掘 APP 用户行为的含义。

2 相关工作

针对用户行为分析的方法,国内外学者展开了大量的研究,部分学者使用机器学习的方法对用户的行为进行分析。Jiang 从智能终端的 APP 使用行为记录出发,对用户使用 APP 的行为进行分析并进行深层次的理解,基于 Doc2Vec 模型提出了 APPUsage2Vec 模型,利用深度学习模型对用户 APP 使用行为进行建模,并利用 APPUsage2Vec 模型完成 APP 使用预测和 APP 协同行为理解^[11]。Jiang 设计并实现了一种基于用户行为的 APP 推荐系统,该系统根据从移动端采集到的用户使用数据对用户的行为进行分析,其在 APP 推荐部分使用了基于用户行为和时间衰减的协同过滤算法为用户推荐 APP^[12]。针对在大数据环境下的用户行为分析,Chen 提出了 Hadoop 大数据平台下的用户行为分析系统,通过输入法的 SDK 收集用户的日常日志信息,从而获得用户的行为数据,并对用户的行为数据进行常规的数据统计计算,包括用户的新增、活跃、留存分析、渠道分析以及自定义事件和错误的分析等。此外,其还通过用户的输入数据对用户进行情感色彩走向的分析^[13]。Guo 等提出了针对网页浏览日志的查询过程划分方法,并在划分的基础上对大量真实用户的浏览行为习惯做了统计分析^[14]。

Fujio 等针对私人聊天系统,分析了 LINE PLAY 提供的私聊日志数据,明确了私聊用户的行为,并使用聚类算法,根据用户行为分析的结果,将用户分为 15 类^[15]。Ren 提出了

面向移动终端用户路径分析的应用程序模型,并通过应用程序页面中的数据埋点,来收集用户数据、会话数据、页面访问数据,结合深度遍历来改进最大正向算法,输出每个会话期间用户访问的最大正向路径,最后结合 ApacheSpark 数据处理平台计算应用程序各功能模块的数据到达率,其研究的路径为移动终端应用用户行为路径,并提出了用户行为路径算法^[16]。Wang 等为了提高移动应用用户的服务质量,分析用户的需求和行为特征,设计了一个移动用户行为分析系统架构,在系统的底层,使用 SDK 和业务逻辑数据库来收集和保存用户行为数据,并使用 Spark 平台对数据进行快速的清理、提取、转换、合并和聚类。同时,他们还对比 Apriori 算法进行了改进,分析了用户行为的相关性,并在 APP 上进行了验证测试^[17]。

在上述研究中,由于对用户行为和用户操作的定义和界限都比较模糊,不易区分用户的行为和操作,因此文献^[11-14]提及的用户行为实际上是用户在使用软件时的操作,行为分析实际上为操作分析。文献^[15-17]针对用户的行为进行分析基本上是针对某一个或某一类软件的用户行为数据进行分析的,且并未明确地提取出用户的行为。因此,针对上述问题,本文提出了一种基于卷积神经网络的 APP 用户行为分析方法,该方法通过分析用户 APP 操作过程、挖掘 APP 操作与 APP 用户行为的关联性、提取 APP 用户行为等步骤,从 APP 操作记录信息中挖掘出用户的行为特征,提高了 APP 用户行为提取的准确率。

3 APP 用户行为分析方法

3.1 APP 操作与 APP 用户行为的关系分析

由于需要对用户的 APP 操作过程记录信息进行分析,从而获取用户使用 APP 的行为模式,因此对 APP 操作与 APP 用户行为之间的关系分析尤为重要。文献^[18]认为,用户操作是一个基于操作内容的抽象事件描述,是对用户特征的理解和表达,也是对人机交互过程的抽象描述。本文认为,用户的 APP 操作指用户使用 APP 达到既定目的时在 APP 内进行的动作。此外,文献^[19]认为,用户行为是用户行为序列中描述一个操作的字符串。本文认为,APP 用户行为是由用户执行的 APP 操作体现出来的,是操作的集成与抽象。

例如,用户希望在“支付宝”APP 中进行转账这一目的,需要完成“打开支付宝 APP”“点击转账选项”“输入转账账号”“输入转账金额”“输入支付密码”“点击确认按钮”等一系列活动,这一系列活动即用户在使用 APP 时的一系列操作。对于上述的一系列操作,体现的是“支付”的 APP 用户行为。

本文通过分析大量的 APP 操作记录发现,APP 使用中常见的用户行为有“购物”“看视频”“阅读”“玩游戏”“听音乐”“社交”“支付”“拍照”“查看图片”“处理文件”“管理应用”“监听”“系统行为”“学习”“其他行为”等。用户使用 APP 过程中可能产生的常见 APP 用户行为及其 APP 操作示例如表 1 所列。

表 1 常见 APP 用户行为及其 APP 操作示例

Table 1 Common APP user behaviors and APP operations examples

APP 用户行为	APP 用户行为特征	APP 操作示例
购物	在购物类、非购物类的 APP 中执行与购物相关的操作	打开京东 APP 查看和购买商品 打开哔哩哔哩 APP 点击会员购选项查看和购买商品
看视频	在视频类、非视频类的 APP 中执行与看视频相关的操作	打开哔哩哔哩 APP 看视频 打开搜狗搜索 APP 点击视频选项看视频
阅读	在阅读类、非阅读类的 APP 中执行与阅读相关的操作	打开搜狗搜索 APP 点击小说书架选项阅读小说 打开微信 APP 进入公众号阅读文章
玩游戏	在游戏类、非游戏类的 APP 中执行与玩游戏相关的操作	打开王者荣耀 APP 玩游戏 打开 QQ APP 点击小游戏选项玩游戏
听音乐	在音乐类、非音乐类的 APP 中执行与听音乐相关的操作	打开酷狗音乐 APP 播放音乐 打开录音 APP 播放音频
社交	在社交类、非社交类的 APP 中执行与社交相关的操作	打开微信 APP 发送消息 打开支付宝 APP 点击朋友选项发送消息
支付	在支付类、非支付类的 APP 中执行与支付相关的操作	打开支付宝 APP 扫码付款 打开微信 APP 扫码付款
拍照	在拍照类、非拍照类的 APP 中执行与拍照相关的操作	打开相机 APP 拍摄照片 打开微信 APP 点击拍摄照片选项拍摄照片
查看图片	在图片类、非图片类的 APP 中执行与查看图片相关的操作	打开图库 APP 查看图片 打开 QQ APP 点击空间相册选项查看图片
处理文件	在文件类、非文件类的 APP 中执行与处理文件相关的操作	打开 WPS Office APP 编辑文件 打开 QQ 邮箱 APP 编辑邮件的文件
管理应用	在管理应用类、非管理应用类的 APP 中执行与管理应用相关的操作	打开华为应用市场 APP 下载软件 打开支付宝 APP 点击版本选项更新版本
监听	在监听类、非监听类的 APP 中执行与监听相关的操作	打开 APP 操作过程分析软件启动监听服务 打开 360 卫士 APP 点击流量监控查看流量使用情况
系统行为	在系统类、非系统类的 APP 中执行与系统设置相关的操作	打开系统设置 APP 连接蓝牙 打开主题商店 APP 设置壁纸和主题
学习	在学习类、非学习类的 APP 中执行与学习相关的操作	打开学习通 APP 查看老师布置的作业 打开腾讯会议 APP 进入网课房间
其他行为	不在上述类别的 APP 中执行相关的操作	打开天气预报 APP 查看天气详情 打开闹钟 APP 设置闹钟

由于用户的 APP 操作记录信息和 APP 用户行为具有很强的领域特性,操作记录中的很多领域性词语在一定程度上可以体现用户的行为。例如“商品”可以体现“购物”的 APP 用户行为,“转账”可以体现“支付”的 APP 用户行为。文献 [20]认为,用户行为知识是业务系统正常运行的操作规范、行为规范以及相关的法律法规制度要求等。本文认为,APP 用户行为知识指与 APP 用户行为相关的领域性词汇,可以体现

APP 用户行为的特征词以及对 APP 用户行为识别起启发性的相关知识。因此,本文将与 APP 操作及 APP 用户行为相关的领域性词语进行整合,构建了可动态更新的 APP 用户行为知识库。APP 用户行为知识库如表 2 所列。通过明确用户 APP 操作与 APP 用户行为之间的关联性,建立两者的关系模型,抽取其中的 APP 用户行为知识,为提高 APP 用户行为分析方法的准确率和效率提供了支持。

表 2 APP 用户行为知识库

Table 2 APP user behaviors knowledge base

APP 用户行为	APP 用户行为知识
购物	商品,物品,取件,寄件,规格,型号,订单,购买,标配,好货,广告,外卖,礼品,专场,发货,收货,快递,物流,拼单,地址
看视频	快进至,后退至,直播,追番,影视,热门,投币,点赞,TV
阅读	书架,小说,新闻,看点,历史,订阅,书城,动态,文章,浏览器,说说,好书,评书,经典
玩游戏	网络游戏
听音乐	歌曲,歌手,下一首,乐库,歌,乐曲
社交	发送,消息,邮件,分享,联系人,拨号,电话,信息,好友,聊天,朋友,表情,头像,挂断,群聊,接收,朋友圈,表情,呼叫,通话,通信
支付	红包,付款,转账,收款,卡包,乘车码,付款码,收款码,余额,充值,银行,收益,账单,财富
拍照	相机,摄影,屏幕录制
查看图片	照片,截屏录屏,时刻,相册,图库,画廊,图像,头像
文件	邮箱,批注,复制,粘贴,剪切,文档,审阅,编辑,标注,演示,粘贴,附件,网盘
管理应用	升级,更新,卸载,版本
监听	频次,内存,运行,日志
系统行为	设置,通知框,无障碍,移动数据,系统导航,存储,关于手机,安全,辅助,权限,锁定屏幕,启动分屏,视图,屏
学习	作业,班级,签到,考试,成绩,笔记,课程,单词
其他行为	时钟,秒表,计时,闹钟,镜子,出行,骑行,地图,附近

3.2 APP 用户行为分析方法

APP 用户行为分析方法通过分析收集到的用户 APP 操作过程记录信息,来提取其中的 APP 用户行为。为了实现这一目标,首先需要将原始的用户 APP 操作过程记录作为输入,进行 APP 操作分析;其次进行 APP 操作与 APP 用户行为之间的关联性挖掘;最后通过分析 APP 操作与 APP 用户行为之间的关联性来提取 APP 用户行为。APP 用户行为分析方法的流程图如图 1 所示。

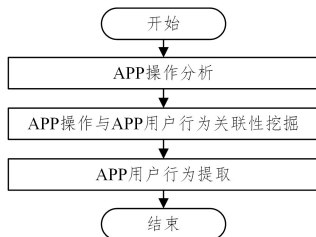


图 1 APP 用户行为分析方法流程图

Fig. 1 Flow chart of APP user behavior analysis method

3.2.1 APP 操作分析

本文通过分析大量的用户 APP 操作过程记录发现,用户 APP 操作过程记录信息主要有以下几类:1)APP 具体操作数据,用户打开 APP 的记录、用户在 APP 内的具体操作记录,如在“微信”APP 中点击了“朋友圈”这一操作记录;2)APP 使用数据,用户在一段时间内使用各个 APP 的时长、次数等;3)日志数据,用户在使用 APP 过程中产生的日志信息。

例如:操作记录数据“com. eg. android. AlipayGphone; android. widget. RelativeLayout; 1; 2020-11-29 08: 27: 13; 8, 朋友”(其中,“com. eg. android. AlipayGphone”是包名,“android. widget. RelativeLayout”是活动名,“朋友”是相关描述),用户的操作是在“2020-11-29 08: 27: 13”这个时间点打开了“支付宝”APP,在“相对布局”页面点击了“朋友”选项。其操作对象是“支付宝”APP,操作活动是“在相对布局页面点击了朋友选项”。由此可见,操作对象、操作活动这两方面在挖掘操作记录信息背后的 APP 用户行为时都至关重要。由上述分析可知,用户 APP 操作过程记录信息蕴含着用户的操作,但用户具体的操作并不明显,APP 操作记录中存在许多冗余无关信息,会对后续的 APP 用户行为提取产生干扰。为了提高 APP 用户行为分析的效率和相关结果的有效性,需要对这些操作记录进行操作分析,从操作数据中提取操作对象和操作活动,以进一步明确、提取其中的操作,从而获得操作记录的核心操作内容,排除无关内容的干扰。因此,本文设计了一种操作分析的算法对 APP 操作过程记录信息进行操作提取,以获取明确的用户操作。操作分析算法如算法 1 所示。

算法 1 操作分析算法

输入:CDL /* 原始 APP 操作过程记录列表 */;AML /* Activity 的对应信息列表 */

输出:analysedCDL /* 操作分析后的操作数据列表 */; /* 操作分析步骤 */

1. 初始化:K←CDL.length;
2. 初始化:J←AML.length;
3. for k ←1 to K do

4. AN←GetAPPName(k);
5. Activity←GetActivity(k);
6. Description←GetDescription(k);
7. for j←0 to J do
8. if AML.get(j) == Activity then
9. AM←GetActivityMessage(j);
10. else
11. AM←QueryActivityMessage(j);
12. 更新 AML;
13. end if
14. end for
15. if Description 是空值 then
16. Capture←AN+AM;
17. else
18. DKW←GetKeyWords(Description);
19. Capture←AN+DKW+AM;
20. analysedCDL.add(Capture);
21. end if
22. end for
23. return analysedCDL.

通过以上算法对用户 APP 操作过程记录信息进行操作分析,可以有效地提取出原始 APP 操作记录信息中的用户操作,为下一步 APP 操作与 APP 用户行为关联性挖掘提供明确的操作数据。

3.2.2 APP 操作与 APP 用户行为关联性挖掘

由于 APP 用户行为分析方法的最终目的是通过分析用户的操作来提取用户的行为,而在操作分析阶段明确的是操作记录的 APP 操作,APP 操作与 APP 用户行为之间的关联性并未建立,因此需要对 APP 操作与 APP 用户行为之间的关联性进行挖掘。当前,语义相似度的方法在挖掘关联性方面具有较为显著的效果和优势,因此,本文提出了一种基于语义相似度的 APP 操作与 APP 用户行为关联性挖掘方法。通过该方法可以有效地将 APP 操作与 APP 用户行为之间的关联性挖掘出来,以构建 APP 操作与 APP 用户行为之间的相似度矩阵,为下一步的 APP 用户行为提取提供关联性数据。

通过分析大量的 APP 操作记录发现,APP 操作中的操作对象和操作内容与常见的 APP 用户行为之间的关联性较为密切。因此,本文针对每一条经过操作分析的操作记录,将其操作对象和操作内容提取出来,使用语义相似度的方法挖掘操作对象、操作内容与 APP 用户行为之间的关联性。此外,由于 APP 操作中蕴含很多领域性的词语,这些词语与常见 APP 用户行为之间的关联性不明显,但与 APP 用户行为知识库中的行为相关知识的关联性较为密切。因此,本文进行关联性挖掘时,需要使用 APP 用户行为知识库中的 APP 用户行为相关知识。关联性挖掘方法的主要思想是依次计算操作对象、APP 类别、操作内容与 APP 用户行为相关知识的语义相似度,分别取相似度值最大的作为当前 APP 用户行为与上述 3 个属性的相似度值,最终构建 APP 操作与 APP 用户行为之间的相似度矩阵,从而获得 APP 操作与 APP 用户行为之间的关联性。其中,APP 类别信息是作为启发性知识参与到 APP 操作与 APP 用户行为关联性挖掘中。构建 APP

操作与 APP 用户行为之间的相似度矩阵的计算公式如式(1)所示:

$$S_{ij} = \text{MAX}\{S_{ij1}, \dots, S_{ijk}, \dots, S_{ijk}\} (1 \leq i \leq g, k \leq k) \quad (1)$$

其中, S_{ij} 表示相似度矩阵中的第 i 行、第 j 列, 其意义为第 i 个属性与第 j 种 APP 用户行为的语义相似度; S_{ijk} 的意义为第 i 个属性与第 j 种 APP 用户行为的第 k 个 APP 用户行为相关知识的语义相似度。

算法 2 APP 操作与 APP 用户行为关联性挖掘算法

输入: analysedCDL/* 操作分析后的操作数据列表 */; APPCL/* APP 的类别数据列表 */; APPUBK/* APP 用户行为知识列表 */; APPNS/* 存储 APPName 与每一种行为的语义相似度的列表 */; APPCS/* 存储 APPClass 与每一种行为的语义相似度的列表 */; CaptureS/* 存储 Capture 与每一种行为的语义相似度的列表 */

输出: Similarity/* 相似度结果列表 */; /* 关联性挖掘步骤 */

```

1. 初始化: I ← analysedCDL.length;
2. 初始化: J ← APPCL.length;
3. 初始化: K ← APPUBK.length;
4. for i ← 1 to I do
5.   Capture ← GetCapture(i);
6.   AN ← GetAPPName(i);
7.   for j ← 0 to J do
8.     if APPCL.get(j) == AN then
9.       AC ← GetAPPClass(j);
10.    else
11.      AC 置为空;
12.    end if
13.  end for
14. for k ← 0 to K do
15.   计算 AN 与 APPUBK[k] 的语义相似度;
16.   取其最大值添加到 APPNS;
17.   计算 AC 与 APPUBK[k] 的语义相似度;
18.   取其最大值添加到 APPCS;
19.   计算 Capture 与 APPUBK[k] 的语义相似度;
20.   取其最大值添加到 CaptureS;
21. end for
22. Similarity.add(APPNS, APPCS, CaptureS);
23. end for
24. return Similarity.

```

通过算法 2 可以有效地将 APP 操作与 APP 用户行为之间的关联性挖掘出来, 并量化成相似度矩阵, 为下一步的 APP 用户行为提取做准备。

3.2.3 APP 用户行为提取

在 APP 操作与 APP 用户行为关联性挖掘阶段, 通过计算语义相似度的方法将 APP 操作与 APP 用户行为之间的关联性进行挖掘并量化, 初步获得了 APP 操作与 APP 用户行为之间的对应关系。为了进一步挖掘操作数据的特征和 APP 操作与 APP 用户行为之间更深层次的关联, 需要对相似度矩阵进行特征提取和特征识别, 其中涉及大量数据的处理。当前的卷积神经网络在对大量数据的特征提取、特征识别和分类问题中具有独特的优势^[21]。因此, 在 APP 用户行为提取阶段, 本文使用了卷积神经网络对 APP 操作与 APP

用户行为构建成的语义相似度矩阵进行特征提取和识别, 构建了专门用于提取 APP 用户行为的卷积神经网络模型。

APP 用户行为提取卷积神经网络模型是一个具有 APP 操作输入层、APP 操作与 APP 用户行为关联性特征提取层、APP 用户行为提取层的卷积神经网络。其中, APP 操作输入层是 APP 操作与 APP 用户行为关联性挖掘阶段获得的 APP 操作与 APP 用户行为之间的相似度矩阵, 尺寸为 $3 * 15 * 1$, 其中隐含 APP 操作与 APP 用户行为关联性的特征, 需要使用卷积和池化操作将相关特征进行提取。由于 APP 操作与 APP 用户行为的相似度矩阵尺寸较小, 因此, 为避免多次卷积和池化导致细节的损失, 在 APP 操作与 APP 用户行为关联性特征提取层仅使用了 2 层卷积层和 1 层池化层来提取特征。其中, 根据输入层的尺寸, 卷积层 1 选择了尺寸较小的 $3 * 3 * 1$ 卷积核, 卷积核的数量为 32, 卷积步长为 1, 进行了填充; 池化层的池化窗口尺寸为 $3 * 3$, 池化步长为 1; 根据卷积层 1 和池化层的结果, 卷积层 2 选择了尺寸为 $2 * 2 * 32$ 的卷积核, 卷积核的数量为 64, 卷积步长为 1, 进行了填充。经实验验证, 此结构的特征提取层能有效地提取 APP 操作与 APP 用户行为关联性的相关特征。其次, APP 用户行为提取层由两层全连接神经网络构成, 第一层全连接神经元的数量为 320, 第二层全连接神经元的数量为 1024。全连接神经网络将提取到的关联性特征进行学习、识别, 最终使用 Softmax 函数进行归一化分类, 提取出 APP 用户行为。APP 用户行为提取卷积神经网络模型的结构如图 2 所示。

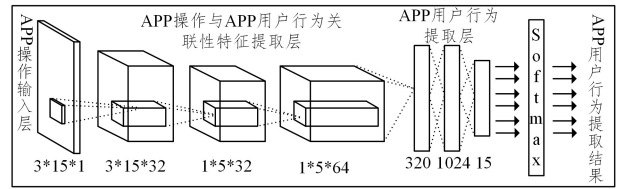


图 2 APP 用户行为提取卷积神经网络模型的结构

Fig. 2 Structure of convolutional neural network model for APP user behavior extraction

APP 用户行为提取卷积神经网络能有效地提取操作记录与 APP 用户行为之间的关联性特征并进行学习, 最终能准确地提取 APP 操作记录中的 APP 用户行为。本文提出的 APP 用户行为分析算法的主要核心步骤包括操作分析(算法 1)、APP 操作与 APP 用户行为关联性分析(算法 2)、APP 用户行为提取(APP 用户行为提取卷积神经网络模型), 该方法的时间复杂度为 $O(n^2)$ 。

4 实验及分析

由于目前未有 APP 用户操作分析的开源数据集, 本文为验证方法的有效性, 使用 Android Studio 开发平台开发了一个 APP 操作过程分析软件, 用于收集用户 APP 操作过程记录数据。本文使用该 APP 收集了 10 个用户在 5 个月内的 163562 条用户 APP 操作过程记录。此外, 为了获得更加准确的 APP 用户行为分析结果, 在从用户 APP 操作过程中的信息提取 APP 用户行为的过程中加入了 APP 类别特征。即从 360 应用市场爬取了 APP 的类别信息, 将 APP 类别信息

加入 APP 操作与 APP 用户行为关联性挖掘阶段,提高了 APP 用户行为提取模型的准确率和有效性。

由于收集到的记录数据是用户在一段时间内真实的 APP 操作过程记录,其中必然会存在冗余的操作数据,因此需要对收集到的大量操作记录进行分析和筛选。经过对操作记录进行筛选后,最终获得了 151 268 条用户操作记录,然后使用上述操作分析的算法对其进行分析,进一步明确、提取其中的操作,验证操作分析算法的有效性,并为下一步的实验构建数据集。

实验后统计发现,对 151 268 条操作记录数据进行操作分析后,提取出来的操作明确的数据有 133 115 条,提取出来的操作不明确的数据有 18 153 条。其中,提取出来的操作明确的数据指经过操作分析后,提取出来的操作对象和操作活动很明确的数据,并且 APP 操作与 APP 用户行为之间的关联性也比较明确。例如,操作数据“‘com..tencent.mobileqq’ ‘com..tencent.mobileqq.activity.ChatActivity’ ‘com..tencent.mobileqq+聊天’”,经过操作分析后提取的操作是“com.tencent.mobileqq+聊天”,即表示在“QQ”APP 中进行聊天,此操作与“社交”这一 APP 用户行为之间的关联性也较为明确。提取出来的操作不明确的数据指经过操作分析后,提取出来的操作对象和操作活动为不明确的数据,APP 操作与 APP 用户行为之间的关联性也不太明确。例如,操作数据“‘com.chaoxing.mobile’ ‘android.widget.Button’ ‘com.chaoxing.mobile+按钮’”经过操作分析后提取出来的操作是“com.android.systemui+按钮”,其中操作的含义并不明确,并不知道用户在 APP 中点击的具体的按钮,因此无法准确获得具体的 APP 操作。此外,其与 APP 用户行为的关联性也不明显,需要进一步的挖掘才能发现。操作分析的实验结果如表 3 所列。

表 3 操作分析的实验结果

Table 3 Experimental results of operation analysis

类别	数据数量	占比
操作明确的数据	133 115	0.88
操作不明确的数据	18 153	0.12
合计	151 268	1.00

使用操作分析算法对操作记录数据进行操作分析后,操作记录数据在原始操作记录数据的基础上,增加了提取出来的操作这一数据项,进一步明确了用户 APP 操作记录的操作。操作分析后的操作记录数据示例如表 4 所列。

表 4 操作分析后的操作记录数据示例

Table 4 Example of operations record data after operation analysis

操作记录数据	操作分析后的操作记录数据
‘cn.wps.moffice_eng’ ‘cn.wps.moffice.writer.multiactivity.Writer1’ ‘WPS Writer1’ ‘WPS 文字’	‘cn.wps.moffice_eng’ ‘cn.wps.moffice.writer.multiactivity.Writer1’ ‘WPS 文字’ ‘WPS Office+WPS 文字编辑器’
‘com.android.gallery3d’ ‘com.huawei.gallery.recycle.APP.RecycleAlbumActivity’ ‘图库’	‘com.android.gallery3d’ ‘com.huawei.gallery.recycle.APP.RecycleAlbumActivity’ ‘图库’ ‘图库+图库照片回收’
‘com.eg.android.AlipayGphone’ ‘android.widget.Button’ ‘确认转账’	‘com.eg.android.AlipayGphone’ ‘android.widget.Button’ ‘确认转账’ ‘支付宝+确认转账按钮’
‘com.tencent.mm’ ‘com.tencent.mm.ui.LauncherUI’ ‘微信’	‘com.tencent.mm’ ‘com.tencent.mm.ui.LauncherUI’ ‘微信’ ‘微信+微信启动界面’

由于最终的目的是使用卷积神经网络对操作数据进行特征提取、特征学习和识别,从而构建 APP 用户行为提取模型,因此需要对实验数据进行人工标注,为其标注 APP 用户行为标签。根据操作分析提取出来的 APP 操作和 APP 用户行为的定义对经过操作分析的 151 268 条数据进行了人工标注,构建了 APP 用户行为提取模型的基础数据集。

根据 APP 用户行为提取模型算法,将经过 APP 操作与 APP 用户行为关联性挖掘的相似度矩阵及其 APP 用户行为标签输入 APP 用户行为提取卷积神经网络模型中,以验证 APP 用户行为提取模型的有效性。实验使用表 3 中的数据,训练数据用于训练模型,测试数据用于测试模型的有效性。实验从操作明确及操作不明确的数据中随机抽取数据组成训练集和测试集,其数据组成比例为 7:3。其中,准确率是测试数据中识别准确的数据与当次实验中测试数据的比值。本次实验中,APP 用户行为提取实验的结果如表 5 所列。

表 5 APP 用户行为提取实验的结果

Table 5 Experimental results of APP user behaviors extraction

序号	训练数据/条	训练数据类型	测试数据/条	测试数据类型	是否打乱顺序	准确率/%	性能/s
1	93 180	操作明确	39 935	操作明确	是	92.4	2 143.00
2	12 707	操作不明确	5 446	操作不明确	是	89.8	61.72
3	105 888	操作明确(88.00%) 操作不明确(12.00%)	45 380	操作明确(88.00%) 操作不明确(12.00%)	是	92.3	3 154.66
4	42 350	操作明确	18 153	操作不明确	是	87.5	729.57
5	63 845	操作明确	27 362	操作明确(70.00%) 操作不明确(30.00%)	是	91.0	2 473.82
6	26 194	操作明确(73.56%) 操作不明确(26.44%)	11 226	操作不明确	是	88.4	305.80
7	12 707	操作不明确	5 446	操作明确(50.00%) 操作不明确(50.00%)	是	88.8	61.81
8	12 707	操作不明确	5 446	操作明确	是	87.7	62.43
9	36 306	操作明确(60.00%) 操作不明确(40.00%)	15 560	操作明确	是	90.6	575.61

表 5 中,“操作明确”指操作分析阶段提取出来的操作是明确的实验数据;“操作不明确”指操作分析阶段提取出来的操作是不明确的实验数据。通过实验发现,针对用户操作记录中操作明确和操作不明确的情况,本文方法均能提取其中的 APP 用户行为,平均准确率在 90%左右。模型提取 APP 用户行为的性能如表 5 所列,测试数据数量较少时,模型提取 APP 用户行为的速度较快,随着数据量的增加,模型的速度有所下降,但整体提取 APP 用户行为的速度较快。因此,基于卷积神经网络的 APP 用户分析方法在提取 APP 用户行为方面是有效的,且具备一定程度的稳定性和高效性。

表 6 对比实验及人工比对结果

Table 6 Comparative experiment and manual comparison results

序号	APP 用户行为	数据来源	方法名称	识别正确数量/条	识别错误数量/条	准确率/%
1	购物	购物类 APP	文献[22]中的方法	964	238	80.2
			本文方法	1 106	96	92.0
		非购物类 APP	文献[22]中的方法	0	977	0.0
			本文方法	890	87	91.1
2	支付	支付类 APP	文献[22]中的方法	105	23	82.0
			本文方法	116	12	90.6
		非支付类 APP	文献[22]中的方法	0	55	0.0
			本文方法	48	7	87.3
3	社交	社交类 APP	文献[22]中的方法	8 006	1 841	81.3
			本文方法	9 021	826	91.6
		非社交类 APP	文献[22]中的方法	0	1 209	0.0
			本文方法	1 087	122	89.9
4	阅读	阅读类 APP	文献[22]中的方法	870	187	82.3
			本文方法	927	130	87.7
		非阅读类 APP	文献[22]中的方法	0	710	0.0
			本文方法	619	91	87.2
5	看视频	视频类 APP	文献[22]中的方法	1 767	391	81.9
			本文方法	1 978	180	91.7
		非视频类 APP	文献[22]中的方法	0	928	0.0
			本文方法	840	88	90.5
6	查看图片	图片类 APP	文献[22]中的方法	603	137	81.5
			本文方法	665	75	89.9
		非图片类 APP	文献[22]中的方法	0	429	0.0
			本文方法	383	46	89.3

由于文献[22]中的方法主要是根据 APP 使用数据中的 APP 类别与 APP 用户行为之间的关系来提取用户的行为,如果数据来源与 APP 用户行为关联的 APP 类别不一致,则无法从该数据中提取出准确的 APP 用户行为(见表 6 中的第 5 列)。本文方法可以较好地处理这一情况。此外,本文方法的 APP 用户行为提取结果与人工比对结果的平均准确率为 90%,说明实验结果与用户真实情况的吻合度高。

为了更直观地展示两种方法有效性的差异,从上述表 6 中挑选部分具有代表性的实验,绘制图表,对比两种方法在处理数据来源差异时的效果。实验对比结果一如图 3 所示。

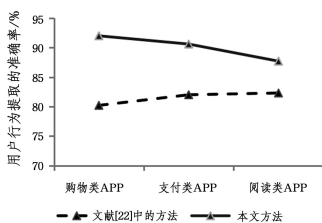


图 3 实验对比结果一

Fig. 3 Experimental comparison results 1

为了进一步验证本文方法的有效性,针对上述实验的实验数据,本文采用文献[22]提出的基于 Apriori 算法的用户行为分析方法进行对比实验。此外,为验证 APP 用户行为提取的客观准确性,实验结束后,邀请了 6 位用户针对上述 9 个实验中前 6 个实验提取出的 APP 用户行为进行人工比对,每一个实验随机选取一种 APP 用户行为对比两种方法提取出来的 APP 用户行为与人工比对的 APP 用户行为之间的差异。实验结果中的准确率是根据人工比对的结果计算所得。对比实验及人工比对结果如表 6 所列。

从图 3 可以看出,文献[22]中的方法和本文方法均可处理数据来源与 APP 用户行为关联的 APP 类别一致的情况。在提取 APP 用户行为的准确率方面,本文方法的准确率明显高于文献[22]中的方法的准确率。

实验对比结果二如图 4 所示。

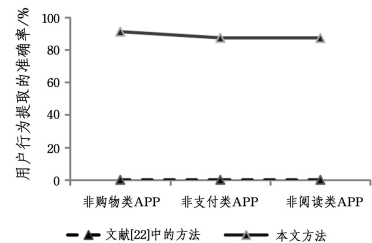


图 4 实验对比结果二

Fig. 4 Experimental comparison results 2

通过图 4 可以看出,文献[22]中的方法无法处理数据来源与 APP 用户行为关联的 APP 类别不一致的情况,准确率均为 0。本文方法能有效地处理这一情况,且准确率较高。

此外,针对表 5 中所有的测试数据(9 个实验中的部分

测试数据是重复的),实验结束后统计文献[22]中的方法与本文方法整体的准确率,实验的整体准确率如表 7 所列。

表 7 实验的整体准确率

方法名称	测试数据	准确率/%
文献[22]中的方法	173954	40.5
本文方法	173954	89.8

通过上述实验结果可得,与文献[22]的方法相比,本文方法在提取用户 APP 操作记录数据中的 APP 用户行为时效果更加显著。

结束语 针对目前用户行为分析中所分析的用户行为大多是用户的操作,未对操作进行进一步抽象以提取出其中的用户行为的问题,本文提出了一种基于卷积神经网络的 APP 用户行为分析方法,通过对用户的 APP 操作记录信息进行操作分析、APP 操作与 APP 用户行为关联性挖掘、APP 用户行为提取等处理,最终提取出 APP 操作信息中的 APP 用户行为,为进一步的 APP 用户行为分析提供服务。

实验结果表明,本文方法能够有效地提取和识别 APP 操作过程记录中用户的行为,具有较高的有效性。但是,本文并未对 APP 用户行为内部的深层含义进行挖掘。因此,下一步将对使用 APP 用户行为分析方法提取出来的 APP 用户行为的深层含义进行研究,挖掘 APP 用户行为体现的用户特征、用户的使用习惯和思维模式等。

参考文献

- [1] AHMED N. Generation Z's Smartphone and Social Media Usage: A Survey [J]. *Journalism and Mass Communication*, 2019, 9(3):101-122.
- [2] CAO H, LIN M. Mining smartphone data for app usage prediction and recommendations: A survey [J]. *Pervasive and Mobile Computing*, 2017, 37: 1-22.
- [3] ZHAO S, XU F, XU Y, et al. Investigating smartphone user differences in their application usage behaviors: an empirical study [J]. *CCF Transactions on Pervasive Computing and Interaction*, 2019, 1(2): 140-161.
- [4] SHEN K, YE X J, LIU X N, et al. Android APP behavior-intention inference based on API usage analysis [J]. *Journal of Tsinghua University (Science and Technology)*, 2017, 57(11): 1139-1144.
- [5] LIU Y Q, CEN R W, ZHANG M, et al. Automatic Search Engine Performance Evaluation Based on User Behavior Analysis [J]. *Journal of Software*, 2008, 19(11): 3023-3032.
- [6] KE Y, LIU Y, LIN B Q, et al. Positive and Unlabeled Learning for User Behavior Analysis Based on Mobile Internet Traffic Data [J]. *IEEE Access*, 2018, 6: 37568-37580.
- [7] LIAO Z F, LI S J, HE D Y, et al. Analysis of Key User Behavior in GitHub Open Source Software Development [J]. *Journal of Chinese Computer Systems*, 2019, 40(1): 164-168.
- [8] LI B. Research and Implementation of APP User Behavior Analysis Based on Android [D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [9] ZHANG L, ZHAO N. Internet Users' Browsing Behaviors Analysis [J]. *Computer Systems & Applications*, 2016, 25(6): 260-264.
- [10] LI Q, DAN L. Research of Music Recommendation System Based on User Behavior Analysis and Word2vec User Emotion Extraction [C] // *International Conference on Intelligent & Interactive Systems & Applications*. Cham: Springer, 2017.
- [11] JIANG Z W. AppUsage2Vec: user behavior modeling and application based on mobile app usage record [D]. Hangzhou: Zhejiang University, 2018.
- [12] JIANG Y. Design and implementation of APP recommendation system based on user behavior [D]. Nanjing: Southeast University, 2018.
- [13] CHEN D. Design and implementation of user behavior analysis department based on big data platform [D]. Wuhan: Huazhong University of Science & Technology, 2016.
- [14] GUO J X, GAO C, XU N S, et al. User behavior analysis based on Web browsing logs [J]. *Computer Science*, 2014, 41(3): 110-115.
- [15] FUJIO T, TAKAFUMI N, MITSUTERU T, et al. Analysis of User Behavior on Private Chat System [C] // *2015 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology Workshops*. 2016: 1-4.
- [16] REN Q. Design of Mobile APP User Behavior Analysis Engine Based on Cloud Computing [J/OL]. *Journal of Physics: Conference Series*, 2020, 1533(2). <https://iopscience.iop.org/article/10.1088/1742-6596/1533/2/022092>.
- [17] WANG R, QIN X, WANG B. Design of Mobile User Behavior Analysis System Based on Big Data [J]. *China Computer & Communication*, 2019(11): 88-90.
- [18] LING Y X, LI R, LAO S Y. Command Space User Operation Model Based on Multi-fingers Touch [J]. *Computer Engineering*, 2009, 35(10): 1-3.
- [19] GUO W Y. The Design and Implementation of Personalized Recommender System Based on the Analysis of User's Behaviors [D]. Nanjing: Nanjing University, 2012.
- [20] SU H, WAN G G. Electronic Forensics System Based on User Behaviors Correlation Analysis [J]. *Telecommunication Science*, 2010, 26(12): 72-78.
- [21] LI Y D, HAO Z B, LEI H. Survey of convolutional neural network [J]. *Journal of Computer Applications*, 2016, 36(9): 2508-2515.
- [22] LIU X S. Study on Mining Methods and Implementations of Smartphone APP Logs for Understanding User Behaviors [D]. Lanzhou: Northwest Minzu University, 2020.



CHEN Yong-quan, born in 1998, post-graduate, is a member of China Computer Federation. His main research interests include software engineering and so on.



JIANG Ying, born in 1974, Ph.D, professor, Ph.D supervisor, is a distinguished member of China Computer Federation. Her main research interests include software quality assurance and testing, cloud computing, big data analysis and intelligent software engineering.