研究论文

基于机器学习的医疗健康APP隐私政策合规性研究*

赵　杨[1,2]　严周周[1]　沈棋琦[1]　李钟航[1]

[1](武汉大学信息管理学院 武汉 430072)

[2](武汉大学国家保密学院 武汉 430072)

摘要:【目的】采用机器学习集成方法对我国医疗健康APP隐私政策的合规性进行测评,提高隐私政策合规性测评的效率与精准性。【方法】依据国家相关政策法规构建医疗健康APP隐私政策合规性测评指标体系,基于硬投票分类器,综合应用卷积神经网络、循环神经网络、长短期记忆人工神经网络三种机器学习算法建立合规性检测模型,通过采集安卓手机应用市场中1 210款医疗健康APP数据,验证模型的有效性并进行隐私政策合规性测评。【结果】我国医疗健康APP隐私政策整体合规性较差,在6项测评维度上均存在较多违规问题,在线医疗、医药服务、健康管理、医学资讯4类细分领域APP的隐私政策合规性得分分别为0.63、0.59、0.61、0.66。【局限】由于标注的隐私政策数据量有限,合规性检测模型无法充分学习测评指标特征。【结论】基于机器学习集成方法的检测模型能够对APP隐私政策的合规性进行大规模、细粒度自动测评,为政府部门科学监管和APP运营商自检自查提供了新的思路与方法。

关键词: 医疗健康APP　隐私政策　机器学习　合规性测评

分类号: TP391　G250

DOI: 10.11925/infotech.2096-3467.2021.0897

引用本文: 赵杨, 严周周, 沈棋琦等. 基于机器学习的医疗健康APP隐私政策合规性研究[J]. 数据分析与知识发现, 2022, 6(5): 112-126.(Zhao Yang, Yan Zhouzhou, Shen Qiqi,et al. Evaluating Privacy Policy for Mobile Health APPs with Machine Learning[J]. Data Analysis and Knowledge Discovery, 2022, 6(5): 112-126.)

1　引　言

随着互联网医疗行业的快速发展,各类医疗健康APP不断涌现,在传统医疗行业数字化转型中发挥了关键作用[1]。根据艾媒咨询调查数据显示,2020年我国移动医疗用户规模已达6.61亿人,受到新冠肺炎疫情影响,医疗健康APP的市场需求进一步提升,发展前景广阔[2]。然而,由于涉及大量个人敏感信息的采集与使用,医疗健康APP也面临诸多隐私风险。为有效保护用户隐私,规范APP的个人信息收集、保存、使用等行为,我国采取APP运营商自律为主、政府监管为辅的模式,由运营商按照政府部门颁布的相关政策法规,自行制定隐私政策来明确对用户隐私保护的责任和义务[3]。近年来,全国信息安全标准化技术委员会、中华人民共和国工业和信息化部等先后颁布了《信息安全技术个人信息安全规范》、《中华人民共和国网络安全法》、《APP违法违规收集使用个人信息行为认定方法》等一系列政策法规,为APP运营商有效制定隐私政策提供了必要依据[4]。但在实际运作中,大量医疗健康APP的隐私政策仍存在重要条款缺失、内容晦涩难懂、权利义务界定不清等问题,未能对用户隐私保护起到

通讯作者(Corresponding author):赵杨(Zhao Yang),ORCID:0000-0003-1784-2733,E-mail:yangzhao_0813@whu.edu.cn。

*本文系武汉大学人文社会科学青年学术团队项目(项目编号:201909)和武汉大学国家保密学院自主科研项目(项目编号:2021017)的研究成果之一。

The work is supported by Wuhan University Humanities and Social Sciences Youth Academic Team Project (Grant No. 201909), Wuhan University National Secrecy School Independent Scientific Research Project (Grant No.2021017).

应有的约束和引导作用。鉴于此，如何科学评估APP隐私政策的合规性，建立安全可靠的移动医疗服务环境，已成为政府部门、APP运营商和广大用户共同关注的焦点。

本研究以我国颁布实施的个人信息保护相关政策法规为依据，在建立医疗健康APP隐私政策合规性测评指标体系的基础上，针对传统评价方法存在的局限性，采用机器学习集成方法对APP隐私政策进行大规模、细粒度测评，进而精准定位存在的关键问题，并提出相应改进建议，旨在为APP运营商规范制定隐私政策、有效保护个人信息安全提供新的思路与方法，同时也为政府部门进行科学监管提供有益参考。

2　文献综述

APP隐私政策是APP运营商基于其提供的服务，对合法采集、保存、共享和利用个人信息的行为进行说明并做出承诺的一种自律性文件，一般在APP中以“隐私权条款/声明”、“隐私保护指引”、“隐私政策协议”等形式列出[4]。医疗健康APP汇集了大量个人敏感信息，比其他APP有着更高的隐私保护要求。近年来，国内外学者围绕医疗健康APP隐私政策的发展现状、监管机制、合规性评价、框架优化等关键问题展开了系统研究[5-8]。其中，合规性评价作为指导APP隐私政策规范制定的重要基础，是学者们研究的热点。已有成果在建立相关评价指标体系的基础上，主要采用内容分析、比较分析、案例分析等方法，对医疗健康APP的隐私政策进行抽样评估，考察其在个人信息采集、保存、使用、共享和转让等权利与义务执行上是否严格遵守国家相关政策法规。研究结果显示，大量APP均存在隐私政策或关键条款缺失、内容模糊不清、可读性差、更新不及时甚至加入强制性条款等问题，严重阻碍了互联网医疗行业的有序发展，如表1所示。

纵观国内外研究现状，学者们已对医疗健康APP隐私政策测评展开了相关理论与实践探索，但现有研究大多面向海外移动医疗应用市场，缺乏我国政策背景下的合规性分析。在研究方法上则以内容分析、案例分析为主，采用抽样调查方式对有限数量的APP进行测评，在研究结果的精准性、全面性和时效性上存在较大局限。随着机器学习和自然语言处理（Natural Language Processing，NLP）技术的广泛应用，基于政策文本自动分类的定量评价方法被逐渐引入隐私政策合规性研究，如Contissa等应用机器学习算法对隐私政策展开定量分析[20]；Harkous等采用众包方式进行隐私政策数据标注，构建了自动化政策条款测度框架[21]。但在医疗健康APP隐私政策合规性测评中，相关研究几乎是空白，缺乏对政策评价算法设计、优化以及实证检验的深入分析。鉴于此，本文在综合应用卷积神经网络（Convolutional Neural Networks，CNN）、循环神经网络（Recurrent Neural Network，RNN）、长短期记忆（Long Short-Term Memory，LSTM）人工神经网络三种机器学习算法的基础上，通过构建隐私政策合规性自动检测模型，对我国医疗健康APP进行全面评估，从而精准定位隐私政策中存在的违规问题，为政府部门进行科学监管和APP运营商自检自查提供重要参考依据。

3　研究方法

本研究在构建APP隐私政策合规性测评指标体系的基础上，综合应用CNN、RNN、LSTM机器学习算法对我国医疗健康APP隐私政策的合规性进行自动定量测评，总体研究框架如图1所示。

（1）以我国颁布实施的个人信息保护相关政策法规为参考依据，建立医疗健康APP隐私政策合规性测评指标体系；

（2）综合应用CNN、RNN、LSTM算法构建APP隐私政策合规性检测模型；

（3）采集我国移动应用市场中的医疗健康APP隐私政策数据并进行预处理，根据（1）中建立的测评指标对一部分预处理后的隐私政策进行语料标注，构建用于合规性检测模型训练和测试的语料库；

（4）应用训练集语料对APP隐私政策合规性检测模型进行训练，利用测试集语料对模型进行测试与参数调节，提高隐私政策检测的准确率；

（5）利用训练好的检测模型对所采集的APP样本进行隐私政策合规性测评，并计算合规性得分；

（6）根据检测结果，分别对医疗健康APP隐私政策的整体合规性和各细分领域APP的隐私政策合规

表1　医疗健康APP隐私政策评价相关研究

Table 1　Related Research on Privacy Policy Evaluation of Mobile Health APP

研究对象	研究方法	评价维度	主要研究结论	文献
20款医疗健康APP	比较分析法	应用权限获取、可读性测试分析	隐私政策未能充分考虑用户的阅读理解水平，缺乏标准的格式和术语	[9]
24款电子健康档案类APP	内容分析法	访问、变更通知、认证机制、数据加密、安全标准和法律规范约束、第三方授权	电子健康档案的安全性和隐私性得分低于3.5分，37%的APP没有隐私政策变更通知	[10]
19款怀孕监控类APP	统计调查法	隐私政策存取管理、数据和访问管理、标准或规定是否遵守法规	抽样APP的隐私政策均不符合HIPAA法，只有47%的APP会直接通知用户隐私政策更新，隐私政策的框架结构和内容均需改进	[11]
70款糖尿病和心理健康类APP	比较分析法	字数、句子、单词、单词字符、阅读轻松度、Gunning Fog得分、SMOG指数、自动可读性指数等	糖尿病APP的隐私政策与心理健康APP的隐私政策在可读性上没有显著差异；隐私政策的复杂性可能是知情决策的障碍	[12]
31款癌症APP	案例分析法	信息收集者的身份、信息处理的目的、信息的种类、跨境传输、存储周期、个人主体的权利等	29%的APP没有制定隐私政策；现有的隐私政策缺乏公平性；只有45%的APP会告知用户主体权利	[13]
61款心理健康类APP	内容分析法	隐私权政策的可访问性和可读性、数据收集使用和共享、安全和信息泄露的投诉以及GDPR法规对隐私政策影响	41%的APP隐私政策没有告知用户如何收集、保留或与第三方共享个人信息；政策内容难以阅读和理解	[14]
116款心理健康类APP	内容分析法	信息存储/共享、密码保护、服务器加密、信息编辑、信息删除	大多数APP隐私政策内容模糊，缺乏数据加密的细节、密码保护等重要信息，可读性不强	[15]
72款老年痴呆APP	内容分析法	隐私政策一般特征、信息采集、共享、出售、披露、用户是否可以删除/修改个人信息	46%的APP有可用的隐私政策；现有的隐私政策内容不清晰	[16]
388款心理健康类APP	比较分析法	隐私政策可读性、信息的性质、种类、使用、第三方共享信息性质、第三方类型、披露信息原因、安全措施	大多数APP不包含隐私政策或协议条款；29%的APP隐私政策没有描述收集信息的性质；仅有10%的隐私政策写明信息共享前需征得用户同意；仅有16%的隐私政策写明收集的个人信息可以被用户删除	[17]
600款健康APP	内容分析法	信息收集类型、理由、共享和用户控制	仅有30.5%的APP拥有隐私政策；66.1%的隐私政策没有专门针对APP本身；隐私政策的阅读需要用户具有大学水平的读写能力	[18]
104款健康APP	实证研究法	隐私政策属性、个人信息的收集、存储、使用、共享、咨询与反馈	隐私政策整体评价不高，平均得分为44.58分；在规范性和完备性上均需完善；部分APP存在过度收集和滥用用户数据的情况	[19]
15款健康APP	内容分析法	信息收集、信息保存、信息使用、信息共享转让披露、信息咨询与反馈	14款APP在隐私政策上存在不足，仅4款APP告知了使用Cookie技术，6款APP提供了运营商联系方式，14款APP未写明响应时限	[6]
20款健康APP	案例分析法	个人信息采集与利用、Cookie/Web Beacon及相关技术的提醒、个人信息储存及保护、个人信息共享转让与披露、个人信息处理权益保护	大多数APP的隐私政策未达到个人信息安全规范标准，未成年人信息保护是国内健康APP隐私保护政策中的薄弱环节	[7]

性进行深入分析，形成最终测评结果。

3.1　医疗健康APP隐私政策合规性测评指标体系构建

测评指标是判断医疗健康APP隐私政策是否符合国家相关政策法规的必要依据。目前，已有研究成果中构建的指标体系基本是以全国信息安全标准化技术委员会2017年颁布的《信息安全技术个人信息安全规范》（GB/T 35273-2017）（简称《安全规范》）为参考依据。而最新版《安全规范》（GB/T 35273-2020）已于2020年正式颁布实施，对新一代信息技术环境下的个人隐私保护提出了新的要求[22]。与此同时，国家互联网信息办公室、工信部、公安部、市场监督管理总局以及电信终端产业协会等部门也相继出台了一系列专门针对APP个人信息安全的政策条例与行业标准，为APP隐私政策制定和规范运作提供了政策指导。

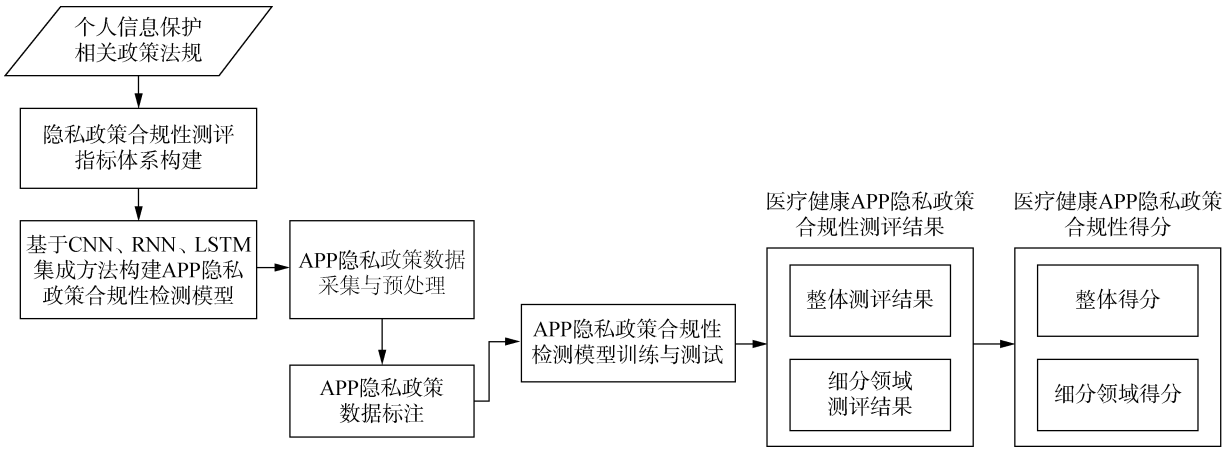

图 1　总体研究框架

Fig.1　Research Framework

为保证测评指标的权威性、科学性和全面性，本研究以2020年版《安全规范》[22]为主要依据，同时参考《中华人民共和国网络安全法》[23]、《App违法违规收集使用个人信息行为认定方法》[24]、《App违法违规收集使用个人信息自评估指南》[25]、《APP收集使用个人信息最小必要评估规范》[26]、《常见类型移动互联网应用程序必要个人信息范围规定》[27]等重要政策法规，分别从“APP隐私政策基本信息”、“个人信息收集与使用规则”、“个人信息存储规则”、“个人信息对外共享、转让、公开披露规则”、“APP运营商对个人信息的保护义务”、“个人信息主体的权利”等6个维度构建医疗健康APP隐私政策测评指标体系，如表2所示。

3.2　基于机器学习集成方法的合规性检测模型构建

在建立APP隐私政策测评指标体系的基础上，进一步采用机器学习集成方法对隐私政策文本进行自动分类，从而判断隐私政策在各项测评指标上的合规性。目前常用的文本分类方法包括Logistic回归、人工神经网络、支持向量机等，但大多采用单一模型，训练数据集的细微变化即会造成模型性能上的显著差异，导致隐私政策文本分类的鲁棒性和泛化能力较差。因此，本研究在综合利用CNN、RNN、LSTM三种神经网络算法的基础上，提出基于投票集成方法的隐私政策文本分类模型，利用不同算法的优势提高分类结果的准确性。常用的投票集成方法包括硬投票和软投票两种。其中，硬投票依据多数优先的原则，选择输出最多的类别结果作为输入文本变量的类别；软投票方法则是依据基础学习器中每种算法输出的类别及其概率，将概率作为权重，综合计算加权平均值，再输出每个类别结果的概率[28]。由于硬投票的输出为分类结果，软投票的输出为各分类结果的概率，因此硬投票方法在类别较多的情况下更为直观高效，便于测试模型效果，所以本文选择硬投票方法进行APP隐私政策指标分类，直接输出分类结果。

具体原理如图2所示。检测模型假定每个基础学习器（CNN、RNN、LSTM算法）为投票者，使用训练集和测试集对基础学习器分别进行语料训练和测试，将未分类的隐私政策语句作为输入变量输入到基础学习器中，学习器根据测评指标自动对其合规性进行判断，确定该语句符合哪项指标。三个学习器分别输出各自的判断结果，再将基础学习器输出的分类值送入集成分类器，由集成分类器对分类值进行投票，得票数过半的指标即为未分类政策语句对应的测评指标，从而逐一确定该APP的隐私政策是否达到各项测评指标要求。

3.3　APP隐私政策合规性得分计算

为进一步量化医疗健康APP隐私政策的合规程度，本研究以隐私政策中包含的测评指标完整性作为判断标准，计算各项APP隐私政策的合规性得分。如果隐私政策条款涵盖的测评指标越完整，则其合

表2　医疗健康APP隐私政策合规性测评指标体系

Table 2　Compliance Evaluation Index System of Mobile Health APP Privacy Policy

一级指标	二级指标	合规性说明
A1 APP隐私政策基本信息	B1 APP运营商情况	是否明确了APP名称、开发方及版权所有方、注册地址
	B2 隐私政策适用范围	是否明确了适用的APP产品和功能服务
	B3 隐私政策的时效	是否明确了政策生效和失效时间
	B4 隐私政策的修订与更新	是否明确了政策修订和更新方式
A2 个人信息收集与使用规则	B5 收集使用个人信息的目的、方式、范围	是否明确了个人信息收集目的、收集方式和收集内容
	B6 各项业务功能收集的信息类型	是否明确了APP提供的业务功能和业务功能需要收集的对应信息类型
	B7 Cookie及其同类技术的使用	是否明确了APP如何使用Cookie及其同类技术
	B8 第三方代码插件的使用	是否明确了第三方代码插件信息、第三方代码和插件处理个人信息的类型和方式
	B9 收集使用个人信息的授权同意	在个人信息的收集和使用过程中是否征得个人信息主体的授权同意
	B10 征得用户授权同意的例外	是否明确根据相关法律法规、监管要求及国家标准，APP可能会收集、使用个人信息而无需征求授权的情形
A3 个人信息存储规则	B11 个人信息存储地点	是否明确了个人信息存储的服务器、地理位置
	B12 个人信息存储期限	是否明确了个人信息存储的时间期限及其依据、延长存储期限的特殊情况
	B13 超期或服务停用处理方式	是否包含当个人信息超出相关保留期限或产品服务发生停止运营的情况时，APP对个人信息的处理机制
A4 个人信息对外共享、转让、公开披露规则	B14 对外共享、转让、公开披露个人信息的情形	是否明确了对外共享、转让、公开披露个人信息的目的、场景
	B15 涉及对外共享、转让、公开披露的个人信息类型	是否明确了对外共享、转让、公开披露的个人信息类型以及对应的APP服务内容
	B16 个人信息接收方类型或身份	是否明确了对外共享、转让、公开披露个人信息的接收方身份信息
	B17 共享、转让、公开披露个人信息时事先征得授权同意的例外	是否明确了根据相关法律法规、监管要求及国家标准，APP共享、转让、公开披露个人信息而无需征求授权的情形
	B18 个人信息跨境传输	是否包含个人信息跨境传输的目的、接收方，是否征得用户的授权同意
A5 APP运营商对个人信息的保护义务	B19 安全保护措施	是否包含个人信息安全影响评估、APP采取的个人信息保护措施与技术、对个人敏感数据的处理
	B20 安全事件应急处置和报告	是否包含为应对个人信息泄露、损毁和丢失等可能出现的风险制定的制度，是否明确了安全事件、安全漏洞的分类分级标准及相应的处理流程
	B21 未成年人信息保护	是否包含收集未成年人个人信息的情况前父母或监护人的授权同意，是否明确了处理和保护未成年人个人信息的机制
A6 个人信息主体的权利	B22 个人信息查询、访问	是否包含个人信息主体查询、访问个人信息的方式
	B23 个人信息更正	是否包含个人信息主体更正个人信息的方式
	B24 个人信息删除	是否包含个人信息主体删除个人信息的方式
	B25 个人账户注销	是否包含个人信息主体注销个人信息的方式
	B26 改变授权同意的范围	是否包含个人信息主体给予或收回授权同意的方式
	B27 个人信息主体请求的响应方式	是否包含个人信息主体的请求申请流程以及运营方响应方式
	B28 隐私问题投诉渠道及反馈机制	是否包含投诉反馈方式(办公地址、联系方式等)、投诉反馈流程

规性得分越高。据此，对各项测评指标进行赋值，并给予每项二级指标相同权重，当隐私政策包含该项指标考察的内容时，对其赋值为1，不包含则赋值为0。通过统计APP隐私政策中包含的完整指标个数，对结果进行归一化处理，得到每项隐私政策的合规性评分。具体计算方法如公式(1)所示。

$$E=\frac{\sum_{i=1}^{n}C_i}{n}\quad C_i=\begin{cases}1, \text{隐私政策包含第}i\text{项指标}\\0, \text{隐私政策不包含第}i\text{项指标}\end{cases} \tag{1}$$

其中，C_i为隐私政策在第i项测评指标上的得

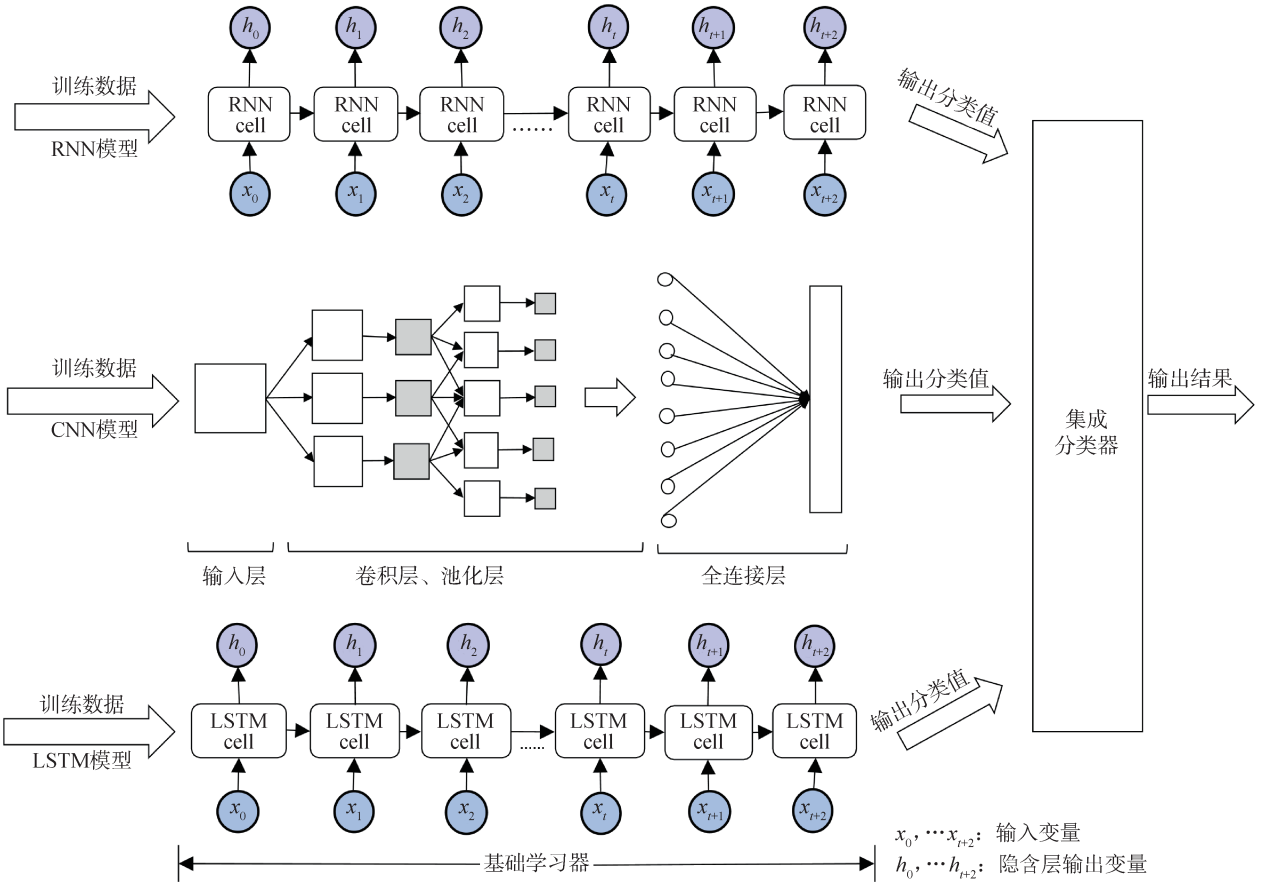

图 2　基于机器学习集成方法的 APP 隐私政策合规性测评原理

Fig.2　Principle of APP Privacy Policy Compliance Evaluation with Machine Learning Integration Method

分，n 为二级测评指标个数，本研究中 $n = 28$。

4　实证研究

4.1　研究样本采集

以华为应用市场、小米应用商店两个安卓手机应用市场上架的医疗健康类 APP 为实证研究对象（截至 2021 年 3 月 31 日，共计 2 236 款）。利用 Python 爬虫采集 APP 的基本信息和隐私政策文本数据（包括“隐私权条款/声明”、“隐私保护指引”、“隐私政策协议”等），剔除隐私政策缺失和内容完全重复的 APP，最终筛选得到符合要求的 1 210 款 APP。

通过 LDA 主题模型对所采集的 APP 简介信息进行聚类分析，按照服务功能将其细分为在线医疗、医药服务、健康管理和医学资讯 4 大类。其中：在线医疗类 APP 的数量为 432 款（35.70%），主要提供挂号、导诊、在线问诊、自诊自查等服务，如春雨医生、好大夫在线等；医药服务类 APP 的数量为 63 款（5.21%），主要提供药品查询、使用说明、销售和配送等服务，如 1 药网、叮当快药等；健康管理类 APP 的数量为 677 款（55.95%），主要提供慢性疾病和运动健康管理服务，如优健康、蜗牛睡眠等；医学资讯类 APP 的数量为 38 款（3.14%），主要提供专业医学资讯及优质健康教育内容服务，如掌上华医、丁香园等。

4.2　APP 隐私政策文本标注

为对所构建的 APP 隐私政策合规性检测模型进行训练和测试，需要建立相应的语料库。根据所采集的隐私政策文本数据，利用 Jieba 分词工具导入自定义词典，并调用哈工大停用词表去除没有实际意义的噪音词，得到隐私政策的未标注初始段落语料。在此基础上，采用众包任务分发模式，招募 15 名医疗和法律专业的大学生，根据表 2 中的隐私政策合

规性测评指标，共同对初始语料进行人工标注。经过领域专家严格培训，在确保每位标注者充分理解测评指标及其判断标准的基础上，向所有标注者开放智能化数据标注工具BRAT入口，对隐私政策进行详细标注。任务完成后，采用Cohen's Kappa系数检验语料标注的一致性，保证标注内容具有高可信度，最终形成包含52 960条医疗健康APP隐私政策标注语句的数据集，部分标注示例如表3所示。

表3　医疗健康APP隐私政策标注示例

Table 3　Example of Mobile Health APP Privacy Policy Annotation

编号	政策分句	对应测评指标
1	本政策适用于平台提供的所有服务，您访问平台网站或登陆相关客户端使用平台提供的服务，均适用本隐私政策。	B2 隐私政策适用范围
2	我们依照法律法规的规定，将在境内运营过程中收集和产生的您的个人信息存储于中华人民共和国境内。	B11 个人信息存储地点
3	公司注册地址：北京市海淀区西三旗建材城内4幢一层115号常用办公地址：北京市海淀区建材城中路27号金隅智造工厂N5在线客服：400-001-8855 。	B1 APP运营商情况
4	我们仅会在以下情况下，且采取符合业界标准的安全防护措施的前提下，才会公开披露您的个人信息。	B14 对外共享、转让、公开披露个人信息的情形
……	……	……
52957	如您对本政策内容有任何疑问、意见或建议、或发现个人信息可能被泄露的，请您在【我的】-【帮助与反馈】-【意见反馈】中留下反馈问题和您的联系方式，方便我们及时与您联系并处理问题。	B28 隐私问题投诉渠道及反馈机制
52958	支付功能：支付功能由与我们合作的第三方支付机构（支付宝、微信支付、银行卡支付，以下统称"支付机构"）向您提供服务，支付机构可能需要收集您的姓名、银行卡类型及卡号、有效期及手机号码。	B15 涉及对外共享、转让、公开披露的个人信息类型
52959	我们可能为以下情形需要披露您的个人信息：（1）遵守法院命令或其他法律程序的规定；（2）遵守相关政府机关的要求；（3）为遵守适用的法律法规、维护社会公共利益，或保护我们的客户、我们、其他用户的人身和财产安全或合法权益所合理必需的用途。	B17 共享、转让、公开披露个人信息时事先征得授权同意的例外
52960	如您使用跨境交易服务，向境外传输您的个人信息完成交易，我们会单独征得您的授权同意并要求接收方按照我们的说明、本隐私政策以及其他任何相关的保密和安全措施来处理这些个人信息。	B18 个人信息跨境传输

4.3　APP隐私政策合规性检测模型训练与测试

为对本研究提出的基于硬投票集成方法的APP隐私政策文本分类模型进行有效训练与测试，将隐私政策标注语料库中的数据按照8∶2分为训练集和测试集，并使用BERT作为预训练模型。与Word2Vec、GloVe预训练模型相比，BERT能够充分考虑词上下文的信息，根据上下文信息的不同动态获得更精确的词向量，同时具有易于迁移学习的特点，只需加载预训练好的BERT模型作为当前任务的词嵌入层，无需对代码做大量修改或优化。在训练过程中，BERT模型的训练参数包括：训练集循环次数为Epoch=3，每次循环训练的数据块大小为200，句子的最长长度为100，学习率为2e-5。

为准确判断模型检测效果，设置7个模型进行对比实验（CNN、RNN、LSTM、CNN-RNN、CNN-LSTM、RNN-LSTM、CNN-RNN-LSTM）。将BERT预训练模型输出的文本向量作为7个模型的输入，增加模型的深度。对于模型中的超参数设置，经过多次运行后观察各模型训练效果，其中，CNN中卷积层C1的过滤器个数是256，卷积核的大小为5；池化层S2的大小为2，步长为1；卷积层C3的过滤器个数为128，卷积核的大小为5；池化层S4的大小为3，步长为3；卷积层C5的过滤器个数为64，卷积核的大小为3。RNN由多个循环计算层构成，每个循环计算层中记忆体的个数值为100，返回每一个输入时间步长的隐藏状态。LSTM由记忆单元、输入门、输出门、遗忘门组成，记忆单元个数为50个，参数Dropout的值为0.1，用于控制输入线性变换的神经元断开比例；参数Recurrent_Dropout的值为0.2，用于控制循环状态线性变换的神经元断开比例。在配

置训练方法时，三种算法将 Softmax 作为激活函数，利用 Adam 优化器动态调整每个参数的学习率，采用多分类交叉熵损失函数（Categorical_Crossentropy）评估当前训练得到的概率分布与真实分布的差异。

本研究使用精确率、召回率和 F1 值对上述模型的检测效果进行比较，结果如表 4 所示。

表 4　APP 隐私政策合规性检测模型性能比较

Table 4　Models Performance of APP Privacy Policy Compliance Evaluation

模型算法	精确率（%）	召回率（%）	F1 值（%）
CNN	91.94	89.10	90.50
RNN	89.57	88.13	88.84
LSTM	86.79	84.21	85.48
CNN-RNN	93.28	91.50	92.38
CNN-LSTM	92.67	89.95	91.30
RNN-LSTM	90.42	87.16	88.76
CNN-RNN-LSTM	95.84	94.07	94.95

从比较结果来看，采用三种神经网络集成算法（CNN-RNN-LSTM）的隐私政策检测效果明显优于基于单一或两种神经网络算法的检测效果，在精确率、召回率和 F1 值上均达到 94% 以上。虽然 CNN 能够提取文本的主要特征，但其主要依靠要素匹配的方式，无法充分发挥文本的连续性和逻辑性特征；RNN 可以学习一定长度的序列文本，保持其连续性，但当长度过长时，会出现“梯度消失”现象，导致其只能学习有限序列长度数据；LSTM 也能保持文本的连续性特征，并有一定长期记忆功能，但无法突出文本的主要特征。因此，随着实验中 APP 隐私政策文本输入量的增加和文本结构复杂性的提升，单一算法和两两算法的组合应用均无法充分满足隐私政策精准检测的要求。而基于硬投票集成方法的 CNN-RNN-LSTM 模型在分类中既保持了文本的有序性特征，又突出了文本的主要特征，能够保证医疗健康 APP 隐私政策合规性检测结果的准确性与可靠性。

4.4　APP 隐私政策合规性测评结果

（1）整体测评结果

利用所构建的机器学习检测模型对所采集的医疗健康 APP 进行合规性测评，在各项一级指标维度上的测评结果如图 3 所示。

①隐私政策基本信息的合规性

在该测评维度，有 1 108 款（91.6%）医疗健康 APP 都未能明确说明隐私政策的生效与失效时间（B3），不符合《安全规范》中提出的个人信息保护政策生效时间说明的规定，不利于隐私政策的有效执行，应引起运营商的高度重视。此外，有 287 款（23.7%）APP 未介绍运营商的基本信息（B1），难以有效提高用户信任度。在隐私政策适用范围（B2）和政策修订与更新（B4）两项二级指标上，大部分 APP 表现良好，仅有 42 款（3.5%）和 95 款（7.9%）APP 未达到合规标准。

②个人信息收集与使用规则的合规性

该测评维度上，多达 1 144 款（94.5%）APP 的隐私政策都未对征得用户授权同意使用其个人信息的例外情况（B10）进行说明。根据《安全规范》要求，在与国家安全、公共安全、重大公共利益相关，或学术研究机构基于公共利益开展统计、学术研究等特殊情况下，可以无需征得用户授权同意。由于医疗健康 APP 能够采集大量与公共卫生安全、学术研究等相关的高价值个人数据，因此隐私政策中应尽可能全面地列出授权例外情况，使用户拥有必要的知情权，从而更好地发挥医疗健康大数据在国家安全战略和重大科学研究中的关键作用。此外，分别有 236 款（19.5%）和 392 款（32.4%）APP 未说明 Cookie 及其同类技术（B7）、第三方代码插件（B8）在收集用户个人信息时的具体方式与采集内容，由于这些技术手段较为隐蔽，往往未能引起用户充分重视，容易使部分运营商钻政策漏洞。

③个人信息存储规则的合规性

在该测评维度，医疗健康 APP 隐私政策的整体合规性较差。其中，有 367 款（30.3%）APP 未说明个人信息存储地点（B11），877 款（72.5%）APP 未说明个人信息存储期限（B12），1 044 款（86.3%）APP 未说明个人信息超出保存期限或服务停止运营后的处理方式（B13）。这说明目前各大运营商普遍没有高度重视用户信息的安全存储，缺少对存储地点、期限以及 APP 停止运营后如何处理其保存的个人信息等关键事项的明确说明。由于医疗健康 APP 在提供多元化医疗服务的同时也积累了海量用户行为数据，对医学领域的科学研究和产品研发具有重要支撑作

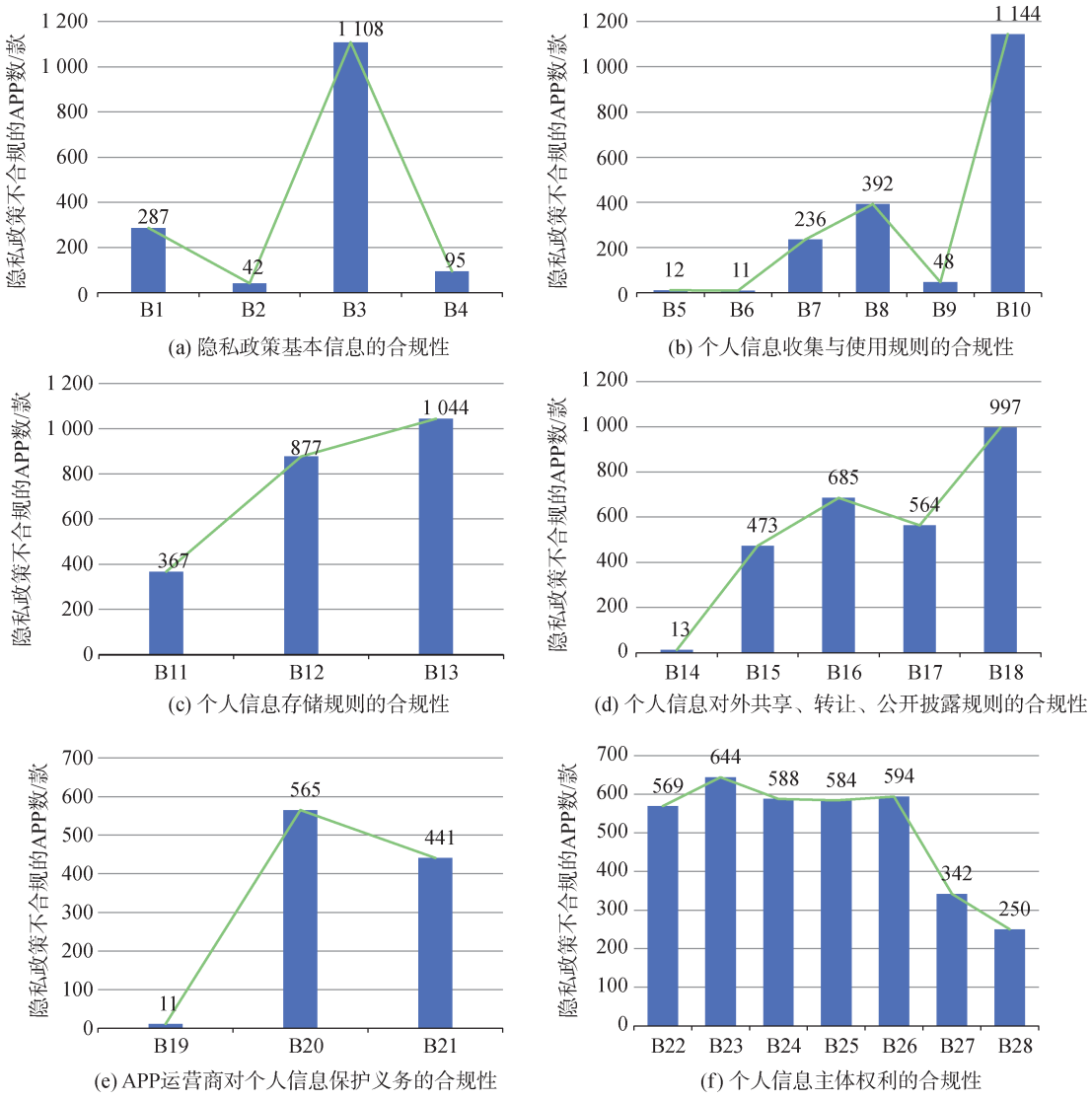

图 3　医疗健康 APP 隐私政策合规性整体测评结果

Fig.3　Evaluation Results of Mobile Health APP Privacy Policy Compliance

用，因此，安全存储和长期保存用户数据是深度挖掘数据价值的必要前提，也是运营商维护用户信息安全的根本义务，应规范制定相关隐私政策条款。

④个人信息对外共享、转让、公开披露规则的合规性

在该测评维度，医疗健康 APP 的整体合规性也较差，分别有 473 款（39.1%）、685 款（56.6%）、564 款（46.6%）、997 款（82.4%）APP 未明确说明对外共享、转让、公开披露的个人信息类型（B15），个人信息接收方的类型或身份（B16），征得用户授权同意的例外（B17）和个人信息跨境传输的相关事项（B18），导致个人信息对外输出和使用方面存在较大安全隐患。随着全球经济一体化进程加快，用户数据跨境流动日趋频繁，在推动医疗健康行业全球合作和资源开放共享的同时，也加剧了用户隐私泄露的风险，甚至威胁到国家安全。《中华人民共和国国家安全法》、《中华人民共和国网络安全法》、《中华人民共和国数据安全法》均对数据跨境流动中的个人信息安全做出了相应法律规定，因此，APP 运营商应高度重视与用户数据跨境传输相关的隐私条款制定。

⑤APP 运营商对个人信息保护义务的合规性

该测评维度上，APP 基本都明确说明了所采取

的个人信息保护措施(B19),仅有 11 款(0.9%)APP 未达到合规标准。但在安全事件应急处置和报告(B20)、未成年人信息保护(B21)两项指标上,则分别有 565 款(46.7%)和 441 款(36.4%)APP 未提及相关条款。由此可见,大部分运营商仅注意到信息安全风险的事前防控,未能充分考虑安全事件发生后的应急响应。与此同时,也没有专门针对未成年用户的个人信息保护采取特殊措施。近年来,随着 APP 隐私泄露事件频发,以及移动医疗服务的青少年用户群体不断增长,及时制定个人信息安全风险应急响应机制和面向未成年人的信息保护专项条款,已成为提高医疗健康 APP 用户使用意愿的当务之急。

⑥个人信息主体权利的合规性

在该测评维度,有超过半数的医疗健康 APP 在诸多二级指标上(B22-B26)均存在隐私政策不合规现象,说明许多运营商并未对个人信息管理的基本权利给予完整、明确的说明,不利于用户有效行使自身权利来保障个人信息安全。此外,分别有 342 款(28.3%)和 250 款(20.7%)APP 在隐私政策中未能明确提供用户请求的相应方式(B27)和隐私问题投诉渠道与反馈机制(B28),这反映出部分运营商未高度重视对用户隐私保护诉求的及时响应,不利于与用户建立良好的沟通机制与互信关系。

(2) 细分领域测评结果

为对医疗健康各细分领域 APP 隐私政策的合规性作进一步详细比较,分别统计在线医疗、医药服务、健康管理和医学资讯 4 类 APP 的不合规数量及其占该类样本的比重,如表 5 所示。

①在线医疗类 APP 隐私政策合规性分析

在线医疗类 APP 的隐私政策整体合规性表现较好,在大部分测评指标上,不合规 APP 的数量占该类样本的比重均低于 40%。特别是在 B2、B5、B6、B14、B19 等指标上,仅有个别 APP 未符合国家相关政策法律要求,说明该类 APP 运营商对用户隐私保护有较高的责任意识。如“春雨医生”的隐私政策中明确表示会使用加密技术保障用户数据安全,并会部署访问控制机制确保只有授权人员才能访问用户信息。但值得注意的是,在 B3、B10、B13 三项指标上,不合规 APP 的占比均超过 85%。由于在线医疗 APP 的主要功能包括用户注册、创建健康档案、在线问诊等,关联了大量用户真实信息(姓名、出生日期、身份证件号码、既往病史等),因此,除了在隐私政策中需要详细说明收集使用个人信息的目的、范围、方式之外,更应让用户清楚获知隐私政策对个人信息保护的时效以及无需征得授权使用时的特殊情况,并对 APP 停止服务后的用户数据处理方式进行明确说明,从而增强用户使用在线医疗服务的信心。

②医药服务类 APP 隐私政策合规性分析

医药服务类 APP 的隐私政策合规性表现并不理想,在 B3、B10、B13 等多项测评指标上的不合规 APP 数量占比均为最高。通过详细研读隐私政策内容发现,大部分医药服务类 APP 均未写明个人信息存储地点和期限,也未对用户个人信息更正、删除和账户注销的权利进行清晰表述,更缺乏对征得用户授权同意的例外和超期或服务停用后用户信息的处理方式给予详细说明。与其他三类 APP 相比,医药服务类 APP 起步较晚,相关隐私政策制定工作相对滞后,普遍存在条款不完整、表述不清晰、关键内容缺失等情况,导致整体合规性较差。由于医药服务类 APP 的主要功能是药品销售、送药上门、用药咨询,能够获取个人身份信息、交易信息、位置信息等重要隐私数据,因此,需要格外重视隐私政策制定工作,为用户信息安全提供更有力的政策保障。

③健康管理类 APP 隐私政策合规性分析

健康管理类 APP 的隐私政策在各项测评维度上的合规率与在线医疗类 APP 相似。目前,我国移动应用市场中的健康管理类 APP 发展迅猛,占医疗健康移动应用市场的比重最大,已形成较为成熟的生态体系,在隐私政策的制定和执行上也积累了丰富经验。值得注意的是,在“个人信息对外共享、转让、公开披露规则”和“个人信息主体的权利”两个测评维度上,该类 APP 的整体合规率较低,基本在 50.0% 以下,存在较大隐私风险隐患。由于健康管理已逐渐深入渗透到人们的日常生活,该类 APP 可以采集大量用户生活作息数据和身体指标数据,比其他医疗健康 APP 拥有更丰富、完整的用户个人信息,因此需要根据数据生命周期,制定从个人信息采集、存储、共享、使用直到 APP 停止服务后合规处理的完整隐私保护条款。

表5　各细分领域APP隐私政策合规性测评结果

Table 5　APP Privacy Policy Compliance Evaluation Results in Various Segments

一级指标	二级指标	在线医疗	医药服务	健康管理	医学资讯
A1 APP隐私政策基本信息	B1	87(20.1%)	16(25.4%)	179(26.4%)	5(13.2%)
	B2	5(1.2%)	3(4.8%)	33(4.9%)	1(2.6%)
	B3	391(90.5%)	60(95.2%)	623(92.0%)	34(89.5%)
	B4	28(6.5%)	5(7.9%)	59(8.7%)	3(7.9%)
A2 个人信息收集与使用规则	B5	2(0.5%)	2(3.2%)	8(1.2%)	0(0.0%)
	B6	2(0.5%)	0(0.0%)	9(1.3%)	0(0.0%)
	B7	88(20.4%)	8(12.7%)	136(20.1%)	4(10.5%)
	B8	135(31.3%)	24(38.1%)	218(32.2%)	15(39.5%)
	B9	16(3.7%)	0(0.0%)	28(4.1%)	4(10.5%)
	B10	410(94.9%)	60(95.2%)	638(94.2%)	36(94.7%)
A3 个人信息存储规则	B11	111(25.7%)	23(36.5%)	228(33.7%)	5(13.2%)
	B12	323(74.8%)	45(71.4%)	482(71.2%)	27(71.1%)
	B13	371(85.9%)	60(95.2%)	584(86.3%)	29(76.3%)
A4 个人信息对外共享、转让、公开披露规则	B14	2(0.5%)	1(1.6%)	9(1.3%)	1(2.6%)
	B15	145(33.6%)	30(47.6%)	283(41.8%)	15(39.5%)
	B16	258(59.7%)	35(55.6%)	376(55.5%)	16(42.1%)
	B17	197(45.6%)	32(50.8%)	318(47.0%)	17(44.7%)
	B18	337(78.0%)	50(79.4%)	578(85.4%)	32(84.2%)
A5 APP运营方对个人信息的保护义务	B19	4(0.9%)	0(0.0%)	7(1.0%)	0(0.0%)
	B20	195(45.1%)	28(44.4%)	327(48.3%)	15(39.5%)
	B21	153(35.4%)	27(42.9%)	248(36.6%)	13(34.2%)
A6 个人信息主体的权利	B22	216(50.0%)	30(47.6%)	307(45.3%)	16(42.1%)
	B23	213(49.3%)	39(61.9%)	374(55.2%)	18(47.4%)
	B24	221(51.2%)	36(57.1%)	319(47.1%)	12(31.6%)
	B25	188(43.5%)	36(57.1%)	345(51.0%)	15(39.5%)
	B26	186(43.1%)	37(58.7%)	355(52.4%)	16(42.1%)
	B27	103(23.8%)	15(23.8%)	214(31.6%)	10(26.3%)
	B28	77(17.8%)	19(30.2%)	147(21.7%)	7(18.4%)

④医学资讯类APP隐私政策合规性分析

医学咨询类APP隐私政策的整体合规性最高，在大部分测评指标上的合规率均高于60%。但在B3、B10、B18三项测评指标上，存在较严重的违规情况。作为以信息服务为主的医疗健康APP，其最先占领移动医疗服务市场，在隐私政策制定方面已积累较多成熟经验。由于医学资讯类APP主要为用户提供医疗健康知识、行业新闻、专业报告等服务，较少涉及个人身份信息、健康信息、交易信息等隐私数据的采集，因此其隐私政策制定的重点应放在对用户个人信息检索记录、咨询记录、浏览行为日志等数据的合理使用和公开披露方面，从而在保证隐私安全的前提下为用户画像构建和内容精准推送提供有效数据支撑。

4.5　医疗健康APP隐私政策合规性得分

根据医疗健康APP隐私政策合规性测评结果，本研究应用公式(1)计算得到所有样本APP隐私政策的整体合规性得分和各细分领域得分，从而判断隐私政策合规程度。具体分数分布情况如图4和图5所示。

由计算结果可知，医疗健康APP隐私政策的整体合规性得分为0.62(满分为1)，只有35款(2.89%)

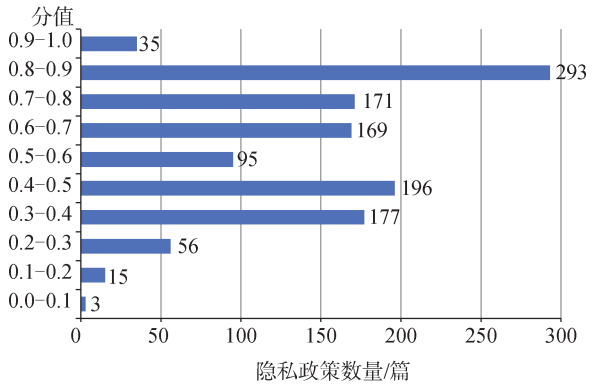

图 4　医疗健康 APP 隐私政策合规性得分分布

Fig.4　Score Distribution of Mobile Health APP Privacy Policy Compliance

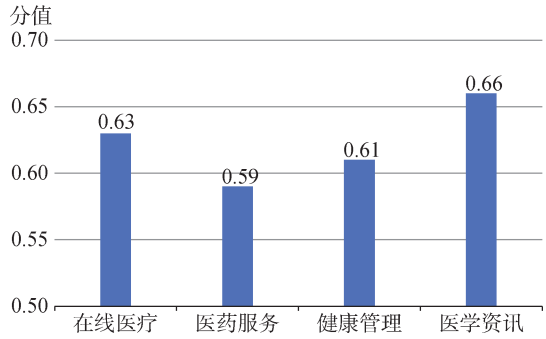

图 5　各细分领域 APP 隐私政策合规性得分情况

Fig.5　APP Privacy Policy Compliance Scores in Various Segments

APP 的得分在 0.90 以上，包括丁香园、薄荷健康、平安健康等。这些 APP 具有上线时间长、用户规模大等特点，说明发展越成熟的 APP 其隐私政策的合规性越高。而 44.8% 的 APP 合规性得分都在 0.60 以下，说明将近一半的 APP 在合规性方面都未能严格贯彻落实国家个人信息保护相关政策法规，使用户隐私面临诸多潜在风险，这将在很大程度上制约我国移动医疗健康服务的创新发展。

从各细分领域 APP 隐私政策的合规性得分来看，4 类 APP 的平均得分均低于 0.70。其中，医药服务类 APP 的得分最低，仅为 0.59，应引起政府监管部门和运营商的高度关注；医学资讯类 APP 的得分最高，为 0.66，可为其他三类 APP 的合规运作提供有益参考。在线医疗类 APP 和健康管理类 APP 作为我国医疗健康移动应用市场中的主流应用，合规性得分也较低，分别为 0.63 和 0.61，需要进一步加强政府监管与行业自律，提高其规范化水平。

5　医疗健康 APP 隐私政策合规性建议

5.1　建立医疗健康 APP 隐私政策标准框架

从测评结果来看，我国医疗健康 APP 隐私政策的不合规问题普遍存在。尽管许多运营商一直致力于隐私政策的修订与完善，但由于缺乏统一的内容标准和格式规范，隐私政策的合规性参差不齐，严重影响了用户对医疗健康 APP 的整体认知。对此，应在严格遵守国家相关政策法规的基础上，针对医疗健康 APP 的服务功能与用户个人信息特点，建立完整的隐私政策标准框架，明确政策中应包含的基本内容和关键条款，为隐私政策制定提供参考范本，并有效采用隐私标签，以标签化、瀑布流的形式清晰展示隐私条款内容，使用户能准确理解隐私政策包含的个人信息保护权利与义务，掌握隐私保护主动权。与此同时，针对互联网医疗行业发展中面临的隐私保护新问题、新风险，运营商应及时根据国家最新颁布的政策法规，对已有隐私政策进行更新与完善，特别是在征得用户授权同意的例外、个人信息数据跨境传输、未成年人信息保护等焦点问题上，制定更加规范详细的隐私条例，引导医疗健康 APP 的合规运作。

5.2　加强医疗健康 APP 隐私政策与服务功能之间的针对性

医疗健康 APP 包含众多细分领域，各领域 APP 涉及的用户个人信息采集范围、类型、途径、方式以及其他信息处理环节均存在较大差异。从本研究对各细分领域 APP 合规性测评结果来看，尽管不同领域 APP 的隐私政策合规程度有高低之分，但在具体规则制定上却较为笼统，并未明显体现不同移动医疗服务间的隐私保护差异。因此，运营商应充分根据自身 APP 的功能特点，有针对性地制定相应条款细则，切实将隐私保护工作落实到 APP 使用的各项环节。譬如，对在线医疗类 APP 而言，其收集用户个人信息设置的最小必要范围应限于注册用户的手机号、姓名、性别、证件类型及号码、病情描述等基本信息；而健康管理类 APP 则无须详细收集上述个人信息，由此降低用户隐私信息的泄露与滥用风险。

5.3 实现政府监管与行业自律的协同发展

有效发挥隐私政策在APP用户个人信息保护与合法利用中的关键作用，需要政府监管部门和移动医疗行业主体的共同努力。一方面，国家互联网信息办公室、工信部、公安部、市场监管总局和卫健委等政府部门应依据相关政策法规，针对互联网医疗行业的数据安全与治理要求，共同建立具有权威性、科学性的APP隐私政策合规性测评指标和相应的监管机制，据此对APP运营商制定的隐私政策及其落实情况进行严格监督和定期审查，及时对违规APP进行下架或通报处理；另一方面，各大APP运营商、内容提供商、服务提供商等主体应不断提高行业自律意识，在有条件的情况下设立专门的用户信息安全管理部门（团队），定期开展用户信息收集、存储、利用、共享等行为的合规性自检自查，在服务过程中严格按照隐私政策要求，积极履行服务主体的责任和义务，不断增强与用户之间的信任关系，推动互联网医疗行业的可持续发展。

（**致谢：**感谢武汉大学图书情报国家级实验教学示范中心对本研究的支持。）

参考文献：

[1] 王天屹, 刘爱萍. 大数据环境下医疗数据隐私保护对策研究[J]. 信息技术与网络安全, 2019, 38(8): 28-32.(Wang Tianyi, Liu Aiping. Research on Privacy Protection of Medical Information in Big Data[J]. Information Technology and Network Security, 2019, 38(8): 28-32.)

[2] 艾媒咨询. 2020-2021中国互联网医疗行业发展白皮书[EB/OL]. [2021-06-10]. https://www. iimedia. cn/c400/77397. html. (iMedia Research. 2020-2021 China Internet Medical Industry Development White Paper[EB/OL]. [2021-06-10]. https://www.iimedia.cn/c400/77397.html.)

[3] 何培育, 王潇睿. 智能手机用户隐私安全保障机制研究: 基于第三方应用程序“隐私条款”的分析[J]. 情报理论与实践, 2018, 41(10): 40-46. (He Peiyu, Wang Xiaorui. Security Assurance Mechanism of Smart Phone Users’ Privacy Based on “Privacy Clause” of Third Party Application[J]. Information Studies: Theory & Application, 2018, 41(10): 40-46.)

[4] 李延舜. 我国移动应用软件隐私政策的合规审查及完善——基于49例隐私政策的文本考察[J]. 法商研究, 2019, 36(5): 26-39. (Li Yanshun. The Compliance Review and Improvement of China’s Mobile App Privacy Policy: A Text Review on 49 Cases of Privacy Policy[J]. Studies in Law and Business, 2019, 36(5): 26-39.)

[5] 郭清玥, 吴丹. 基于文本分析的APP隐私政策框架优化研究[J]. 信息资源管理学报, 2021, 11(1): 18-29.(Guo Qingyue, Wu Dan. Research on Optimization of APP Privacy Policy Framework Based on Text Analysis[J]. Journal of Information Resources Management, 2021, 11(1): 18-29.)

[6] 刘乾坤, 刘昊鹏, 秦子昂, 等. 基于内容分析法的健康APP用户隐私保护政策发展现状研究[J]. 中国医院, 2019, 23(9): 20-23. (Liu Qiankun, Liu Haopeng, Qin Ziang, et al. Research on Users’ Privacy Protection Policy of M-Health Application Based on Content Analysis[J]. Chinese Hospitals, 2019, 23(9): 20-23.)

[7] 付少雄, 赵安琪. 健康APP用户隐私保护政策调查分析——以《信息安全技术 个人信息安全规范》为框架[J]. 图书馆论坛, 2019, 39(12): 109-118. (Fu Shaoxiong, Zhao Anqi. Research on User Privacy Protection Policies of Health Apps——Based on Information Security Technology—Personal Information Security Specification[J]. Library Tribune, 2019, 39(12): 109-118.)

[8] 王晰巍, 相甍甍, 张长亮, 等. 新媒体环境下信息隐私国内外研究动态及发展趋势[J]. 图书情报工作, 2017, 61(15): 6-14. (Wang Xiwei, Xiang Mengmeng, Zhang Changliang, et al. Research on the Development Trend of Domestic and Foreign Information Privacy Under New Media Environment[J]. Library and Information Service, 2017, 61(15): 6-14.)

[9] Rowan M, Dehlinger J. A Privacy Policy Comparison of Health and Fitness Related Mobile Applications[J]. Procedia Computer Science, 2014, 37: 348-355.

[10] Zapata B C, Hernández Niñirola A, Fernández-Alemán J L, et al. Assessing the Privacy Policies in Mobile Personal Health Records [C]//Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014: 4956-4959.

[11] Bachiri M, Idri A, Fernández-Alemán J L, et al. Evaluating the Privacy Policies of Mobile Personal Health Records for Pregnancy Monitoring[J]. Journal of Medical Systems, 2018, 42 (8): 144.

[12] Powell A C, Singh P, Torous J. The Complexity of Mental Health App Privacy Policies: A Potential Barrier to Privacy[J]. JMIR MHealth and UHealth, 2018, 6(7): e158.

[13] Benjumea J, Ropero J, Rivera-Romero O, et al. Assessment of the Fairness of Privacy Policies of Mobile Health Apps: Scale Development and Evaluation in Cancer Apps[J]. JMIR MHealth and UHealth, 2020, 8(7): e17134.

[14] Parker L, Halter V, Karliychuk T, et al. How Private is Your Mental Health App Data? An Empirical Study of Mental Health App Privacy Policies and Practices[J]. International Journal of Law and Psychiatry, 2019, 64: 198-204.

[15] O’Loughlin K, Neary M, Adkins E C, et al. Reviewing the Data Security and Privacy Policies of Mobile Apps for Depression[J].

Internet Interventions, 2019, 15: 110-115.

[16] Rosenfeld L, Torous J, Vahia I V. Data Security and Privacy in Apps for Dementia: An Analysis of Existing Privacy Policies[J]. The American Journal of Geriatric Psychiatry, 2017, 25(8): 873-877.

[17] Robillard J M, Feng T L, Sporn A B, et al. Availability, Readability, and Content of Privacy Policies and Terms of Agreements of Mental Health Apps[J]. Internet Interventions, 2019, 17: 100243.

[18] Sunyaev A, Dehling T, Taylor P L, et al. Availability and Quality of Mobile Health App Privacy Policies[J]. Journal of the American Medical Informatics Association, 2014, 22(e1): e28-e33.

[19] 马骋宇, 刘乾坤. 移动健康应用程序的隐私政策评价及实证研究[J]. 图书情报工作, 2020, 64(7): 46-55.(Ma Chengyu, Liu Qiankun. Research on the Privacy Policy's Evaluation and Empirical Study of Mobile Health Applications[J]. Library and Information Service, 2020, 64(7): 46-55.)

[20] Contissa G, Docter K, Lagioia F, et al. Claudette Meets GDPR: Automating the Evaluation of Privacy Policies Using Artificial Intelligence[J]. SSRN Electronic Journal, 2018: 1-59.

[21] Harkous H, Fawaz K, Lebret R, et al. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning[C]//Proceedings of the 27th USENIX Security Symposium. 2018:531-548.

[22] 全国信息安全标委会. 信息安全技术个人信息安全规范(GB/T 35273-2020)[EB/OL]. [2021-06-10]. http://openstd.samr.gov.cn/bzgk/gb/newGbInfo? hcno=4568F276E0F8346EB0FBA097AA0CE05E. (National Information Security Standardization Technical Committee. Information Security Technology—Personal Information Security Specification(GB/T 35273-2020)[EB/OL]. [2021-06-10]. http://openstd. samr. gov. cn/bzgk/gb/newGbInfo?hcno=4568F276E0F8346EB0FBA097AA0CE05E.)

[23] 中华人民共和国网络安全法[EB/OL]. [2021-06-10]. http://www. cac. gov. cn/2016-11/07/c_1119867116. htm. (The Cybersecurity Law of the People's Republic of China[EB/OL]. [2021-06-10]. http://www.cac.gov.cn/2016-11/07/c_1119867116.htm.)

[24] App违法违规收集使用个人信息行为认定方法[EB/OL]. [2021-06-10]. http://www.cac.gov.cn/2019-12/27/c_1578986455686625.htm. (Identification Method of APP's Illegal Collection and Use of Personal Information[EB/OL]. [2021-06-10]. http://www.cac.gov.cn/2019-12/27/c_1578986455686625.htm.)

[25] App违法违规收集使用个人信息自评估指南[EB/OL]. [2021-06-10]. https://pip.cybersac.cn/jbxt/privacy/detail/2019030211460093427 7. (Self-Assessment Guide of APP's Illegal Collection and Use of Personal Information[EB/OL]. [2021-06-10]. https://pip.cybersac.cn/jbxt/privacy/detail/20190302114600934277.)

[26] 电信终端产业协会. APP收集使用个人信息最小必要评估规范[EB/OL]. [2021-06-10]. http://www.taf.org.cn/StdDetail.aspx?uid=8EBE18CA-10C0-4300-B425-FDC43A9305ED&stdType=TAF. (Telecommunication Terminal Industry Forum Association. Application Software User Personal Information Collection and Usage Minimization and Necessity Evaluation Specification General Principle[EB/OL]. [2021-06-10]. http://www.taf.org.cn/StdDetail.aspx?uid=8EBE18CA-10C0-4300-B425-FDC43A9305ED&stdType=TAF.)

[27] 常见类型移动互联网应用程序必要个人信息范围规定[EB/OL]. [2021-06-10]. http://www. cac. gov. cn/2021-03/22/c_1617990997054277.htm. (Provisions on the Scope of Necessary Personal Information for Common Types of Mobile Internet Applications[EB/OL]. [2021-06-10]. http://www.cac.gov.cn/2021-03/22/c_1617990997054277.htm.)

[28] 赵宇鑫, 努尔布力, 艾壮. 基于集成学习投票算法的Android恶意应用检测[J]. 计算机工程与应用, 2020, 56(22): 74-82.(Zhao Yuxin, Nurbol, Ai Zhuang. Android Malware Detection Based on Ensemble Learning Voting Algorithm[J]. Computer Engineering and Applications, 2020, 56(22): 74-82.)

作者贡献声明:

赵杨:提出研究思路,设计研究方案,论文撰写与修订;
严周周:模型构建,实验实施,论文起草;
沈棋琦:数据采集和分析,实验实施,论文起草;
李钟航:数据采集,实验实施。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储,E-mail: 1476979185@qq.com。
[1] 沈棋琦, 严周周. 医疗健康APP隐私政策原始数据集.xlsx. 安卓手机应用市场1 210款医疗健康APP隐私政策原始数据.
[2] 沈棋琦, 严周周. 医疗健康APP隐私政策标注数据集.xlsx. 医疗健康APP隐私政策文本信息预处理及标注后的数据集.

收稿日期:2021-08-24
收修改稿日期:2021-11-15

Evaluating Privacy Policy for Mobile Health APPs with Machine Learning

Zhao Yang[1,2] Yan Zhouzhou[1] Shen Qiqi[1] Li Zhonghang[1]

[1](School of Information Management, Wuhan University, Wuhan 430072, China)

[2](School of National Secrecy, Wuhan University, Wuhan 430072, China)

Abstract: [**Objective**] This paper analyzes privacy policies for mobile health APPs in China with machine learning, aiming to improve the efficiency and accuracy of compliance evaluation. [**Methods**] First, we constructed the evaluation system for the privacy policy compliance of mobile health APPs according to relevant policies and regulations. Then, based on the hard voting classifier, we established the compliance evaluation model integrating three machine learning algorithms: CNN, RNN and LSTM. Finally, we examined our model using 1210 mobile health APPs from the Android APP market, and evaluated the compliance of their privacy policies. [**Results**] The overall compliance of the privacy policies for mobile health APPs was poor. There are many violations in the six evaluation criteria. The compliance scores of online medical APPs, medical service APPs, health management APPs, and medical information APPs were 0.63, 0.59, 0.61and 0.66. [**Limitations**] Due to the limited amount of annotated privacy policy data, the proposed model may not be able to fully learn the features of evaluation indicators. [**Conclusions**] This proposed model could conduct large-scale, fine-grained automatic evaluation of the compliance of APPs privacy policies. It also provides new ideas and methods for the government agencies and APP operators to improve decision making.

Keywords: Mobile Health APP Privacy Policy Machine Learning Compliance Evaluation