

# The Problem of Balance between “Information Requester” and “Privacy Applicant” from a Double Perspective

Min Xu Yanwen Zheng Chen Wang

Anhui University of Finance and Economics, Bengbu, Anhui, 233000, China

## Abstract

With the development of mobile Internet and the rise of all kinds of mobile APPs, all kinds of apps bring convenience to users, while the use rights and privacy issues behind them also gradually cause consumers' close attention with the outbreak of Internet security events in recent years. In this paper, the degree of information collected by smartphone APP and the tolerance of privacy collected by users are studied, and then the method of privacy protection is proposed. Based on the above research, the author believes that compliance and legal processing of users' data “footprints” and the adoption of a new generation of information technology will eventually achieve the balance between “information requester” and “privacy applicant”.

## Keywords

tolerance; analytic hierarchy process; fuzzy quantification algorithm; comprehensive evaluation index system; new generation information technology

## 双视角下的“信息所求者”和“隐私诉求者”的关系平衡问题

许敏 郑延雯 王晨

安徽财经大学, 中国·安徽 蚌埠 233000

## 摘要

随着移动互联网发展以及各类手机 APP 的兴起, 各类 APP 在给使用者带来便利的同时, 其背后的使用权限和隐私问题也随着近年来互联网安全事件的陆续爆发而逐渐引起消费者的密切关注。论文针对智能手机 APP 收集用户信息程度以及用户对 APP 收集隐私容忍度问题进行研究, 进而提出用户隐私保护的方法。基于以上研究, 笔者相信合规、合法处理用户的数据“足迹”, 采用新一代信息技术终将能实现“信息索求者”和“隐私诉求者”的关系平衡问题。

## 关键词

容忍度; 层次分析法; 模糊量化算法; 综合评价指标体系; 新一代信息技术

## 1 研究选题的提出

### 1.1 研究背景

在移动互联网背景下, 手机 APP 用户一方面享受着 APP 带来的便捷服务, 另一方面又被隐私信息泄露的风险所困扰。APP 开发商需要根据个人信息来进行大数据演算, 优化升级 APP, 但是过度索取用户个人隐私信息也会导致用户的不满。中国消费者协会于 2018 年 11 月 28 日发布的《100 款 APP 个人信息收集与隐私政策测评报告》显示, 多达 91 款 APP 要求的权限涉嫌过界, 属于过度收集用户个人信息。在越来越多的手机 APP 上市的背后, 存在着用户隐私泄露的安全隐患。层出不穷的隐私泄露问题给用户带来了不良的体验, 也给社会造成了负面影响。随着用户隐私保护意识的逐渐加强, 移

动应用公司开始推出隐私保护软件, 政府立法部门也加强了相关的法制政策。2019 年 1 月 25 日, 中央网信办、工信部、公安部、市场监管总局等四部门联合发布《关于开展 App 违法违规收集使用个人信息专项治理的公告》; 2019 年 8 月 8 日中国信息安全标准化技术委员会秘书处发布《信息安全技术移动互联网应用 (App) 收集个人信息基本规范 (草案)》向社会征求意见。因此, 在大数据时代, 隐私信息与收集信息的矛盾已经得到多方关注, 在开发商与用户个人隐私信息保护之间寻求平衡点已成为社会热点问题。

### 1.2 研究意义

当下的销售模式以市场需求为导向, 使得 APP 开发商越

来越注重对用户需求的了解。云计算技术的应用与普及、大数据挖掘技术的日益发展,为APP开发商收集用户数据、分析用户喜好提供了条件。对于APP开发商而言,收集用户个人信息可用于市场趋势分析和消费者行为研究,有助于发现用户内在需求、锁定目标客户,进而增加应用下载量,使公司获得效益。但大量收集用户数据,会导致成本和风险的增加,同时有可能导致用户满意度、用户粘性的下降。过多的收集用户数据往往会带来道德和法律风险,一旦数据涉及用户敏感信息、侵犯了用户的隐私权,APP开发商的声誉就会受到影响,甚至需要承担法律责任。因此,基于开发商角度的新时代信息经济发展和用户隐私信息的保护,适当适量的收集用户数据、挖掘用户信息成了APP开发商亟需解决的问题之一。因此,本次研究的目的就是通过对十大类型APP针对用户的信息收集调查分析现下APP收集个人信息的程度和用户对APP收集隐私信息的容忍度,建立综合评价指标体系,为APP收集个人信息是否过度做出客观判断,同时建立用户对隐私信息披露程度的容忍度模型,寻找在大数据时代下用户与开发商的“信息索求者”和“隐私诉求者”的平衡<sup>[1]</sup>。

## 2 模型建立

### 2.1 信效度检验

#### 2.1.1 信度检验

信度 (Reliability) 即可靠性或可信度,指测量结果的一致性 or 稳定性,即测量工具确能否稳定地测量到它要测量的事项,与测量结果的正确与否无关。通常认为,信度系数应该在 0~1 之间,当信度系数在 0.5 及以下表示量表有些项目需要抛弃,当信度系数在 0.5~0.7 表示量表接近于真实情况,量表分析出的结果更接近于真实情况,当信度系数在 0.8 以上,表明量表数据接近于理想化,笔者用 SPSS 对问卷进行了 Cronbach's alpha 检验。

表 1 可靠性检验

| 可靠性统计 |             |    |
|-------|-------------|----|
| 克隆巴赫  | 基于标准化项的克隆巴赫 |    |
| Alpha | ALpha       | 项数 |
| 0.610 | 0.645       | 20 |

由表可知,“app 收集个人信息及用户隐私容忍度调查”问卷的克隆巴赫系数 alpha 系数为 0.610 可信度较好且更接近与真实情况。

#### 2.1.2 效度检验

效度越高表示测量结果越能显示出所要测量对象的真正特征。效度分为三种类型:内容效度、准则效度和结构效度。在问卷在设计过程中,笔者参考了大量市场调查理论和文献,并运用相关的统计分析方法,使设计出的问卷真实有效地达到欲测量的目的。

表 2 KMO 和巴特利特检验

| KMO 和巴特利特检验 |      |         |
|-------------|------|---------|
| KMO 取样适切性量数 |      | .698    |
| 巴特利特球形度检验   | 近似卡方 | 842.895 |
|             | 自由度  | 190     |
|             | 显著性  | 0.000   |

笔者采用因子分析法对问卷中 109 个项目的效度进行检验,得出了 KMO 和 Bartlett 球形检验图,如图所示:KMO 的值为 0.698, Bartlett 球形检验的 P 值小于 0.01,差异具有统计学意义,各变量存在相关性。

## 2.2 模型的建立

### 2.2.1 APP 对用户个人隐私信息搜集程度的多变量逐步回归

设 App 对用户的个人隐私收集程度隐私的收集程度为 Y,选取被要求打开蓝牙权限的频率、要求获取设备信息权限的频率、被要求打开摄像头权限的频率、被要求使用话筒录音权限的频率、被要求读取通话记录权限的频率、被要求获取位置信息权限的频率、被要求读取短信记录权限的频率作为解释变量,设它们分别为  $X_i(i=1,2,3,4,5,6,7)$ ,另外考虑到用户在注册前对其所使用的 APP 是否存在默认勾选同意用户协议或隐私政策的不良问题的知晓情况对结果也会产生影响,为了提高整个模型的精度,于是:

$$D_1 = \begin{cases} 1 & \text{有} \\ 0 & \text{其他} \end{cases} \quad D_2 = \begin{cases} 1 & \text{没有} \\ 0 & \text{其他} \end{cases}$$

利用 eviews 软件,导入数据,利用逐步回归法并不断剔除,逐步消除最小显著性的变量,得到结果如下:

表 3 eviews 得出的多元回归结果

| Variable | Coefficient | Std.Error | t-Statistic | Prob   |
|----------|-------------|-----------|-------------|--------|
| C        | 0.198329    | 0.224583  | 0.883098    | 0.3785 |
| X1       | 0.013706    | 0.084019  | 0.163134    | 0.8706 |
| X3       | 0.137003    | 0.081520  | 1.680605    | 0.0948 |
| X4       | 0.202339    | 0.081815  | 2.473133    | 0.0145 |

|                    |           |                       |           |          |
|--------------------|-----------|-----------------------|-----------|----------|
| X5                 | 0.342831  | 0.052828              | 6.489515  | 0.0000   |
| X6                 | 0.223253  | 0.078044              | 2.860609  | 0.0048   |
| D1                 | -0.488850 | 0.155204              | -3.149726 | 0.0020   |
| D2                 | -0.391296 | 0.130276              | -3.003588 | 0.0031   |
| R-squared          | 0.693014  | Mean dependent var    |           | 3.006098 |
| Adjusted R-squared | 0.679239  | S.D.dependent var     |           | 1.245837 |
| S.E.of regression  | 0.705589  | Akaike info criterion |           | 2.187984 |
| Sum squared resid  | 77.66559  | Schwarz criterion     |           | 2.339197 |
| Log likelihood     | -171.4147 | Hannan-Quinn criter.  |           | 2.249371 |
| F-statistic        | 50.30950  | Durbin-Watson stat    |           | 1.557325 |
| Prob(F-statistic)  | 0.000000  |                       |           |          |

由结果可知, 在  $prob>F$  表示拒绝原假设, 当  $prob<0.05$  的情况下, 有 95% 的把握认为可接受原假设, 即该变量对结果的影响是显著的; 其中 R-squared 与 Adjust R-squared 表示你和优度和修正后的拟合优度, 由上述结果可知,  $R^2 = 0.69$ ,  $\overline{R^2} = 0.68$  表明该结果拟合优度较好, 且不存在多重共线性和自相关性, 因此笔者可以得到该模型的回归方程:

$$Y = 0.0137X_1 + 0.1370X_3 + 0.2023X_4 + 0.3428X_5 + 0.2233X_6 - 0.4889D_1 - 0.3913D_2$$

从方程中可以看出 APP 对个人隐私信息搜集的程度深浅主要受被要求打开蓝牙权限频率、被要求打开摄像头权限频率、被要求使用话筒录音权限频率、被要求读取通话记录权限的频率及被要求获取位置信息权限频率, 以及用户使用协议和隐私政策的模糊程度 6 个变量的影响, 即被要求获取这些隐私权限的频率越高, 则用户认为 APP 对隐私信息搜集的程度越深<sup>[2]</sup>。

App 使用协议和隐私政策的模糊程度对 app 商家收集程度呈现负影响, 即 App 用户对隐私政策留意的越少, app 对隐私的收集程度越深, 从隐私信息的重要度来讲, 话筒录音、通话记录属于核心隐私信息, 因此对被解释变量的影响所占权重最大, 即用户面对此类隐私被要求使用权限时往往比其他信息更为敏感; 而蓝牙属于普通信息, 因此对解释变量的影响所占权重较小, 即用户面对普通信息被要求使用权限时表现的没有那么敏感。

### 2.2.2 容忍度模型建立

将用户对 app 收集个人隐私的容忍度测度指标分为信息敏感性 ( $S_1$ ), 接收者敏感性 ( $S_2$ ) 和使用者敏感性 ( $S_3$ ), 其中信息敏感性是指个人对某类信息敏感性的感知, 用来分

析用户对隐私泄露种类的容忍度; 接收者敏感性是用来分析用户对隐私泄露的传播范围和对象: 使用者用户可容忍的隐私使用范畴。将 app 收集个人隐私的容忍度进行量化, 令其为和 H, 即:

$$\max H = m_1S_1 + m_2S_2 + m_3S_3$$

并且满足:

$$m_1 + m_2 + m_3 = 1$$

在信息敏感性指标, 将隐私信息分为不同的种类, 进而衡量用户对不同种类隐私发生泄露时的态度。将不同种类隐私进行量化, 因此  $S_1$  的集合为:

$$S_1 = \{a_1, a_2, a_3, \dots, a_n\}$$

在接收者敏感性维度中, 根据移动互联网产业构成将手机隐私信息的企业进行分类, 用来探究用户对不同种类企业收集个人信息大的意愿度, 也从侧面反映用户对不同企业的信任度, 因此  $S_2$  的集合为:

$$S_2 = \{b_1, b_2, b_3, \dots, b_n\}$$

在使用者敏感度中, 笔者将 app 企业处理用户的实践行动进行分类, 用来用户对各类隐私实践的态度, 因此  $S_3$  的集合为:

$$S_3 = \{c_1, c_2, c_3, \dots, c_n\}$$

设用户对 app 收集个人隐私的容忍度的门限值为  $t$ , 在信息敏感性维度中, 设  $a_i(i=1,2,3)$  为其子集合, 则信息敏感性  $S_1$  的表达式为:

$$S_1 = p\{a_i \geq t | \text{信息收集后}\}$$

同理接收者敏感性  $S_2$  和使用者敏感性  $S_3$  的表达式分别为:

$$S_2 = p\{b_i < t | \text{信息发布后}\}$$

$$S_3 = P\{c_i \geq t | \text{信息被企业利用}\}$$

如果隐私信息没有发生被传播, 则用户的容忍度信息集合需要满足:

$$\sum_{i=1}^n f(a_i, b_i) = \sum_{i=1}^n f_i$$

$$\sum_{i=1}^n r_i \leq m \leq \sum_{i=1}^n r_{i\max}$$

$n$  表示可参考的指标数,  $r$  为指标的冗余度, 为的最大值。

容忍度门限值与可参考指标的映像关系为:

$$t=f(n)$$

综上, 可建立容忍度评估模型; 对于用户对 app 隐私收集程度中相同层次上的每个元素相对于上一层的准则进行两两对比, 构建判断矩阵如下:

$$B = (k_{ij})_{m \times n} = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_{nn} \end{bmatrix}$$

根据判断矩阵的解, 计算出各个元素相对于该准则的权重; 采用特征根法得到权重的表达式:

$$B = \mu \delta$$

其中  $\mu$  表示特征根,  $\zeta$  表示其对应的特征向量, 对判断矩阵进行一致性检验

$$CR = \frac{CI}{RI}$$

根据笔者所收集的数据可以计算得到  $CR = 0.088 < 0.1$ , 因此该判断矩阵的一致性可以被接受。根据上述步骤, 从上到下将模型中的单层元素的权重进行合成, 获得模型最底层相对于最高层的合成权重, 并进行相应参数求解<sup>[1]</sup>。

由于  $S_1, S_2, S_3$  收所求的三个主成分,  $\lambda_1, \lambda_2, \lambda_3$  是对应成分的特征根, 将特征根归一化:

$$m_i = \frac{\lambda_i}{\sum_{i=1}^3 \lambda_i} (i=1, 2, 3)$$

表4 公因子方差表

|                                 | 初始值   | 提取值   |
|---------------------------------|-------|-------|
| 在安装和使用 APP 时您是否会允许软件获取权限?       | 1.000 | 0.578 |
| 对于 APP 中存在强制同意某些隐私才成运行的霸王条款您会选择 | 1.000 | 0.449 |
| 您是否有过隐私泄露的经历?                   | 1.000 | 0.809 |
| 您是否同意 APP 开发商将个人隐私保护作为重要考虑因素    | 1.000 | 0.741 |
| 我不能容忍我的身体健康状况被商家知道              | 1.000 | 0.790 |
| 我不能容忍网络上的浏览记录被商家知道              | 1.000 | 0.650 |
| 我不能容忍我的通讯录被商家读取                 | 1.000 | 0.733 |

|                                   |       |       |
|-----------------------------------|-------|-------|
| 我不能容忍我的私人文件被商家读取                  | 1.000 | 0.693 |
| 我不能容忍我的通话记录, QQ 微信等社交软件的聊天记录被商家获取 | 1.000 | 0.741 |
| 我不能容忍我目前所在的地理位置被商家知道              | 1.000 | 0.770 |
| 我不能容忍我之前去过的场所被商家知道                | 1.000 | 0.746 |
| 我不能容忍个人信息被透露给应用服务提供商(百度公司等)       | 1.000 | 0.742 |
| 我不能容忍将个人信息透露给通讯服务公司               | 1.000 | 0.686 |
| 我不能容忍将个人信息透露给操作系统开发商              | 1.000 | 0.435 |
| 我不能容忍将信息透露给手机制造厂商                 | 1.000 | 0.746 |
| 我不能容忍企业大量的存储我的个人信息                | 1.000 | 0.809 |

将数据导入软件得到公因子方差表

由上述表达式求解得:

$$\max H = 0.316S_1 + 0.558S_2 + 0.128S_3$$

表5 解释变量回归系数

| 解释变量 | 信息敏感性 | 使用者敏感性 | 接受者敏感性 |
|------|-------|--------|--------|
| 回归系数 | 0.316 | 0.558  | 0.128  |

由于容忍度与可接受性成反比关系, 因此用户对接收者敏感性的容忍程度偏低, 因此用户更在意自己的隐私被用在哪以及后续产生的。

### 3 结论与局限

#### 3.1 结果分析

在样本各类型的 APP 收集个人信息中, 存在过度收集用户信息。对于蓝牙的收集程度电商平台类型 APP 达到了峰值为 3.81, 高于出行导航类 APP, 对于通话记录的收集程度电商平台类型 APP 达到了峰值为 3.84 高于通讯类 APP, 对于短信读取的收集程度资讯阅读类型 APP 达到了峰值为 3.72, 高于社交通讯类 APP。因而存在收集不相关信息的行为, 存在潜在可能的过度收集情况。

APP 收集信息程度分为主动和被动上, 存在用户主动披露信息和被动收集上, 受到用户对隐私安全风险的关注度影响上<sup>[4]</sup>。

用户对 APP 收集信息的容忍度从信息敏感性来说他们更介意收集聊天记录信息, 对位置信息的容忍度较高; 从信息终端的角度来说他们更愿意容忍个人信息被通讯服务公司收集, 介意商家收集; 从使用信息的角度来说他们排斥未经允许保存和访问。

用户对于 APP 收集信息的容忍度受到多方面的影响, 其中, 隐私技术保护方面影响最大, 此外默认勾选同意用户协

议和隐私政策对其容忍度的影响也颇大。

### 3. 2 研究贡献

用户对 APP 收集信息收集容忍度普遍集中在保留态度,并不是完全排斥权限请求,但会考虑收集不相关信息的行为,存在潜在可能的过度收集情况。

基于以上调查笔者相信合规、合法处理用户的数据“足迹”,采用新一代信息技术如论文给出的云模型做为探索道路终将能实现“信息索求者”和“隐私诉求者”的关系平衡问题。

### 4 未来展望

随着现代科技的快速发展,以云计算技术为代表的新一代信息技术正在快速得到应用,云计算在企业和个人工作领域得到普遍发展,使得数据处理效率获得了质的提升,使得企业个人的数据不在局限于传统的实践和地点,云计算

将在隐私信息保护中扮演越来越重要的角色,因此在这里论证云计算的对用户隐私信息的保护机制<sup>[5]</sup>。

### 参考文献

- [1] 李睿. 移动互联网环境下隐私泄露容忍度的测量与实证研究 [D]. 大连: 大连理工大学, 2014.
- [2] 李春忠, 靖稳峰, 徐健. 基于多尺度信息融合的层次聚类算法 [J]. 工程数学学报, 2019(03):245-255.
- [3] 李旭辉, 魏瑞斌. 长江经济带战略性新兴产业信息化水平动态测度体系研究 [J/OL]. 情报杂志: 1-9[2019-10-17].
- [4] 陈剩勇, 卢志朋. 互联网平台企业的网络垄断与公民隐私权保护——兼论互联网时代公民隐私权的新发展与维权困境 [J]. 学术界, 2018(07):39-52.
- [5] 史奉楚. 还有多少侵犯用户隐私的“默认勾选”? [N]. 人民邮电, 2018-01-16.