

中国隐私风险指数 分析报告

2020 年度中国隐私风险指数分析报告

《中国隐私风险指数报告》已连续编纂 3 年，2020 年度相比前两年获取了更多的分析数据，包括 3670 万真实用户数据¹（2018-2019 年度约 3000 万）和约 40 万 App 数据²（2018-2019 年度约 30 万），能对当前移动应用场景下的用户隐私数据被收集的情况进行更准确的调研分析，同时导出了中国隐私风险指数 3 年的变化情况。在内容上，该报告继续沿用先前提出的中国隐私风险指数体系，旨在从用户、移动应用程序（Mobile Application，简称 App）和数据收集者（即 App 开发者）三个角度揭示当前用户隐私数据被收集的现状，及其产生的隐私风险状况。

基于数据拥有者（用户）量化模型，本报告对 2018-2020 年度中国总体隐私风险情况进行对比，并分析其原因。结论显示，2020 年度中国总体隐私风险指数相比 2019 年度下降 15.8%，接近于 2018 年度的中国总体隐私风险。而该隐私风险指数的下降伴随着用户平均安装 APP 数量的上升，该现象主要是 APP 平均请求权限敏感度下降所致，该平均敏感度相比 2019 年度同比下降 2.25%。此结论也揭示出各 APP 在 2020 年度国家各部门强力的 APP 治理政策的压力下，对敏感权限的收集有所收敛，说明了当前法律政策对 App 敏感数据收集的遏制作用。

表 1 2018-2020 中国总体隐私风险情况对比

	用户平均 App 安装量 (个)	用户平均权限数据泄露量 (份)	App 平均权限敏感度	总体隐私风险指数	增长率
2018 年	27	291	2.17	0.45	—
2019 年	31	336	2.22	0.57	26.7%
2020 年	56	532	2.17	0.48	-15.8%

2018-2020 中国总体隐私风险详细情况如表 1 所示。在该表中，总体隐私风险指数，是指基于数据拥有者量化模型中单个用户隐私风险指数，计算得到的所有用户样本隐私风险指数的均值。1 份权限数据指 1 个用户通过 1 次权限请求泄露的数据。对任一用户而言，1 个 App 通过其多个权限请求可获取多份用户权限数据，用户手机上的 1 个 API 可向多个 App 泄露多份用户权限数据。

在数据拥有者分析中，本报告继续基于移动用户的隐私风险指数，从区域、人群、行为三个角度对隐私风险进行评估，并进行 2018-2020 的三年对比。

2018-2020 年度区域隐私风险对比显示，旅游省份与经济发达省份与其他区域的隐私风险指数差异增大，前者的隐私风险愈加明显高于后者。从每年度的隐私风险排名变化上可发现，贵州省、云南省与吉林省的隐私风险排名连年上涨，吉林省三年间全国隐私排名更是提高了 14 个位次；而湖南省和江苏省的隐私风险排名连年下降，江苏省三年间累计下降了 11 个位次。从各省份的隐私风险值上，其总体的区域与中国总体隐私风险趋势保存一致，即 2019 年度的隐私风险值明显偏大，2020 年度相比之前有明显下降。

2018-2020 年度人群与行为隐私风险总体上无大变化。从人群隐私风险来看，男性人群的隐私风险指数首年超越女性人群。在其他属性上，经济能力较高人群的隐私风险亦较高，该结论愈加凸显；同时，已婚人群、青年人群（26-35 岁人群）的隐私风险相比同类别其他属性人群依旧保持着较低水平。从行为隐私风险来看，三年来出行方式与贷款方法对不同行

¹ 3670 万真实用户数据是由 AURORA 极光大数据公司提供的脱敏数据集，分析结果仅涉及统计信息，不涉及个体隐私。

² 40 万 App 数据由 WAMDM 实验室该项目参与者从应用宝、豌豆荚等应用市场爬取得到。

为人群的隐私风险影响一直较高，2020 年度“学生贷”和“团购”行为首次成为对隐私风险影响较大的偏好行为。

在数据收集者分析中，本报告对 2018-2020 年度中国的数据垄断状况进行分析对比，结论显示，10%的数据收集者依旧可收集 99%的权限数据，数据垄断的严峻形势仍居高不下。但从具体的数据来看，2018 至 2020 年，数据垄断的形势一直有着极其轻微的缓解趋势，数值上该缓解对应收集数据比例的变化幅度不足 0.1%。

在 2019 年新增的移动应用程序分析中，本报告进一步完成了 2019 与 2020 年度的结果对比。从 APP 的请求权限数量上，2020 年度各类 APP 请求权限的数量相比之前均有减少。在 APP 的权限设置隐私风险 P1 上，各类 APP 的 P1 隐私风险指数均有降低，且最高隐私风险级别对应的 APP 数量与用户量均有减少，相应地，这些 APP 与用户相对均匀地转移至其他较低的隐私风险级别中去；在 APP 的使用量隐私风险 P2 上，由于疫情期间用户对多种 APP 使用量的增加，各类别 APP 的 P2 隐私风险指数有明显上升，不同级别对应的 APP 数量与用户数量与之前的分布趋势相同，即使用量隐私风险越高，APP 数量与用户量越大。

综上，本报告展示了 2018 至 2020 三年来中国隐私风险与数据垄断局势居高不下的总体特征，从区域差异、用户差异、行为差异、APP 差异等各个角度展示了中国隐私风险的细微变化，亦证明了国家各部门对 APP 治理的阶段性成效，即各 APP 对用户敏感权限的请求均有所收敛。但目前这些举措仍旧不足以扭转当前较高的数据隐私风险与数据垄断局势，发掘有效的数据治理技术与体系势在必行，只有这样才能更好的响应中共中央国务院发布于 2020 年 4 月发布的《关于构建更加完善的要素市场化配置体制机制的意见》，加快培育数据要素市场！

一、中国隐私风险指数体系

中国隐私风险指数是一个反映我国在特定时段内数据拥有者（移动用户）因个人数据被收集者（App 开发者）获取而面临的隐私风险及数据收集者造成的隐私风险相对数值的宏观指标，用来反映不同移动用户个体或群体面临隐私风险的差异。该指数基于中国现有 334 个地级市分层抽样的 3670 万（36,722,417）真实用户的 App 使用数据、162 个维度的用户属性画像数据、通过爬取第三方应用网站得到 40 万（406,054）个 App 相关信息构成的数据集计算得到。

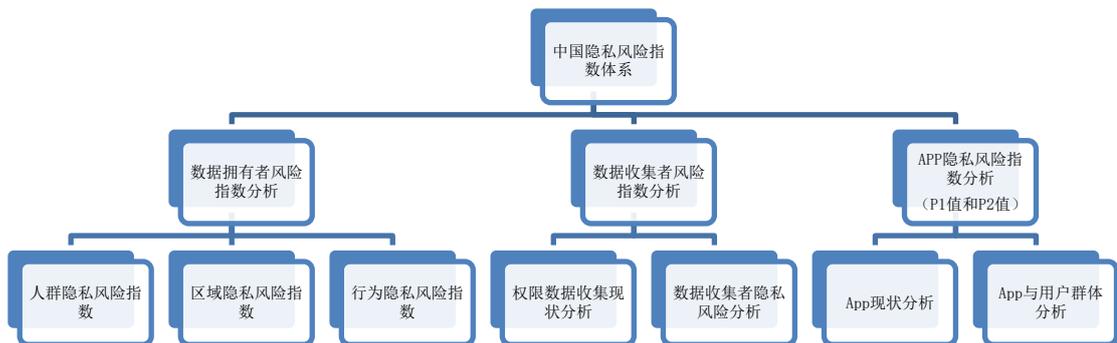


图 1-1 中国隐私风险指数体系

基于数据拥有者（移动用户）、数据收集者（App 开发者）、以及移动应用程序（APP）的隐私风险量化结果，本报告进一步制定中国隐私风险指数体系，如图 1-1 所示，从数据收集者角度揭示移动用户数据的流向，并从自然属性、社会属性、行为属性等维度揭示不同数据拥有者群体的隐私风险特征，最终分析汇总成《中国隐私风险指数分析报告》。

二、大规模数据收集现状分析

图 2-1 与图 2-2 分别展示了 2018-2020 年度不同比例的数据收集者收集数量及隐私风险占比变化情况。连续三年的分析显示：**10%的数据收集者可收集 99%的权限数据，数据垄断的严峻形势近两年只有极其轻微的缓解，总体上仍居高不下。**

如图 2-1 所示，2018-2019 年度，前 10%、5%、1%的数据收集者收集数量比例基本接近；2020 年，前 0.1%、0.0.1%的数据收集者收集数量比例相比 2019 年有所上涨，但仍低于 2018 年的收集数量比例。图 2-2 所示的不同比例的数据收集者的隐私风险占比变化趋势与之类似。因此可得出以下结论：

- 2019 年相比 2018 年数据垄断形势略有缓解；
- 2020 年相比 2019 年数据垄断形势有极其轻微的上涨，但相比 2018 年依旧有缓解的趋势。

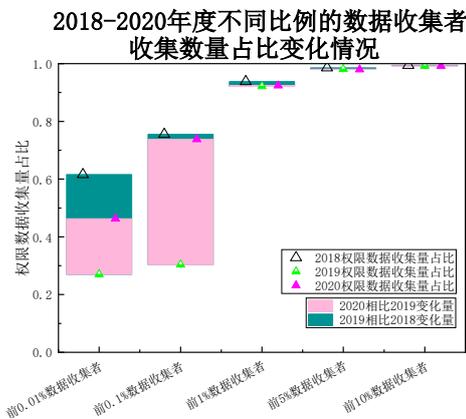


图 2-1 2018-2020 年度不同比例的数据收集者收集数量占比变化情况

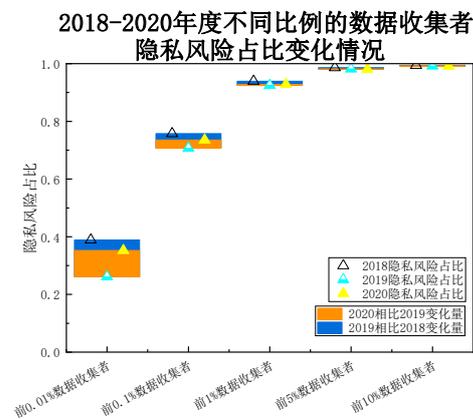


图 2-2 2018-2020 年度不同比例的数据收集者隐私风险占比变化情况

三、移动应用程序分析

本节对 2018-2020 年度各类别 App 的权限请求数量及隐私风险状况进行对比分析。

图 3-1 展示了 2018-2020 年度各类别 App 平均请求权限数量。由此可见：2020 年各类别 App 的权限请求数量均少于 2018 年、2019 年，说明**各类别 App 在 2020 年对用户权限的请求均有所收敛。**

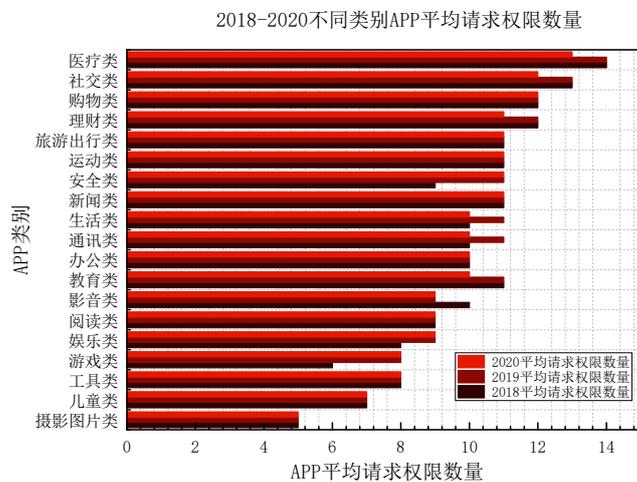


图 3-1 2018-2020 年度不同类别 App 平均请求权限数量

图 3-2、图 3-3 分别展示了 2019-2020 年度各类别 App 的 P1 隐私风险值和 P2 隐私风险值。可以看出：2020 年大部分类别 App 的 P1 隐私风险值相对 2019 年有所降低，P2 隐私风险值相对 2019 年提高较为明显。这与 2020 年“用户安装 APP 数量增加、但 APP 请求用户权限更为谨慎”结论一致。

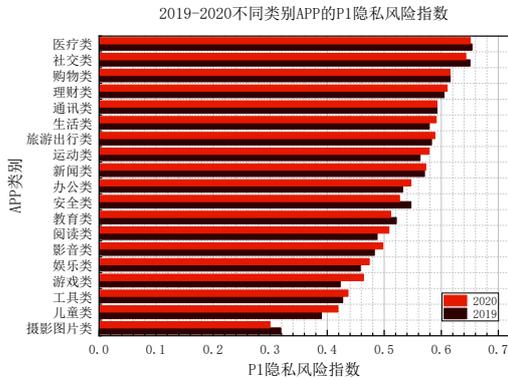


图 3-2 2019-2020 年度不同类别 App 的 P1 隐私风险指数

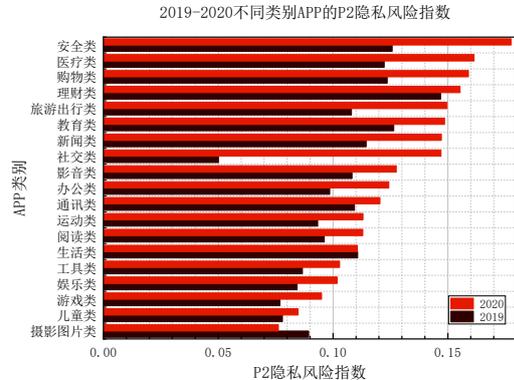


图 3-3 2019-2020 年度不同类别 App 的 P2 隐私风险指数

四、移动应用程序分级体系分析

本节对 2019-2020 年度不同隐私风险级别的 App 数量及用户量进行对比分析。由图 4-1 可得：

App 对高敏感度权限的请求有所收敛。2020 年，最高 P1-L10 隐私风险级别 App 的用户量减少，其他隐私风险级别 App 的用户量有所提高，而各隐私风险级别 App 数量主要呈现增长趋势。这与上节中“各类别 App 在 2020 年对用户权限的请求均有所收敛”的结论一致。

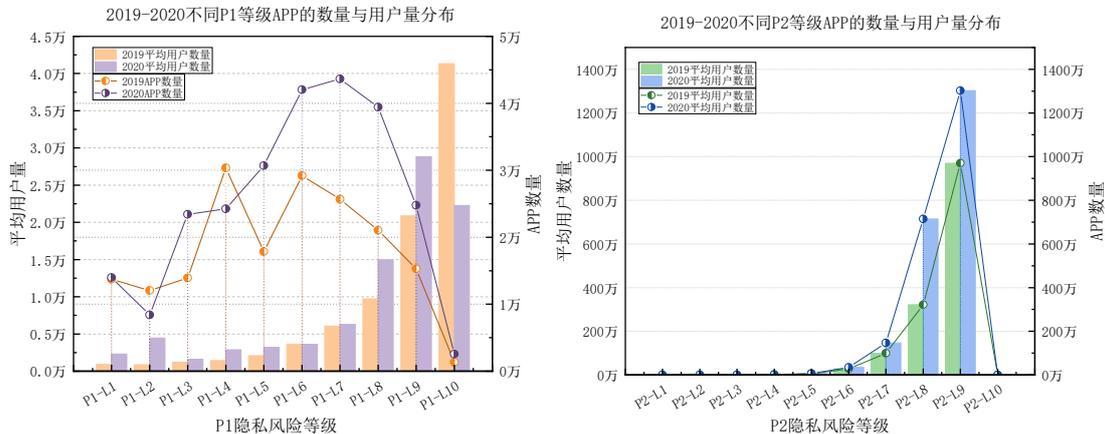


图 4-1 2019-2020 年度不同隐私风险级别的 App 数量与用户量分布

五、区域隐私风险指数

本节对 2018-2020 年度省级区域的隐私风险指数进行对比分析。由图 5-1 可以看出，省级区域隐私风险差异化趋势越来越明显：经济发达及旅游省份隐私风险较高，其他省份隐私风险较低。



图 5-1 2018-2020 年度中国隐私风险指数地图

六、人群隐私风险指数

本节将分别对职业、消费水平、收入能力、婚姻状况、年龄及性别与人群隐私风险指数的关系进行 2018-2020 年度对比分析。总体来看，除在性别基本属性上，男性人群的隐私风险首次略高于女性人群，其他基础属性对应的隐私风险分布并无明显变化。

6.1 职业

由图 6-1 可知：2018-2020 年，工程施工人员、旅游及健身娱乐场所服务人员、社会服务和居民生活服务人员的隐私风险指数相对较高，运输服务人员、安全保卫和消防人员、科学研究人员的隐私风险指数相对较低。

职业与人群隐私风险指数关系

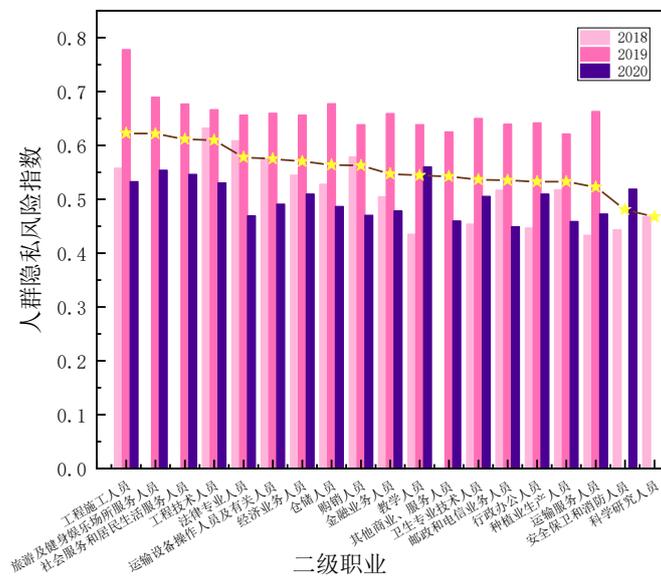


图 6-1 2018-2020 年度职业与人群隐私风险指数关系

6.2 消费水平

由图 6-2 可知：2018-2020 年，人群隐私风险指数与消费水平基本呈正比。

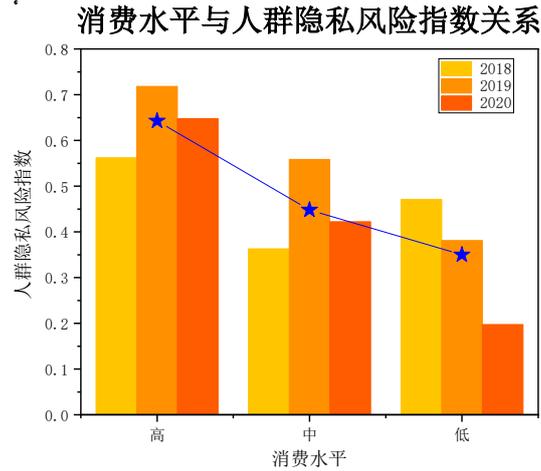


图 6-2 2018-2020 年度消费水平与人群隐私风险指数关系

6.3 收入能力

由图 6-3 可知：2018-2020 年，人群隐私风险指数与收入能力基本呈正比。

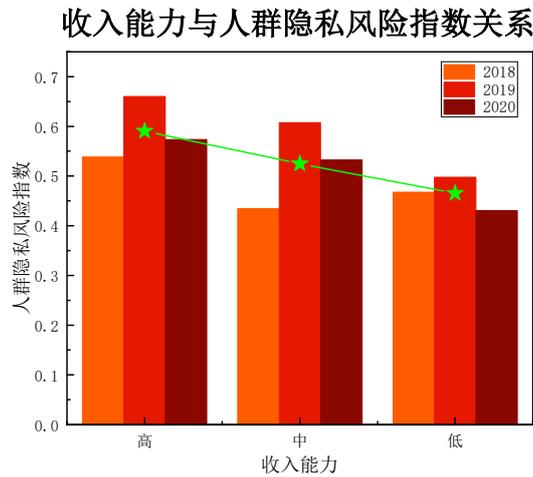


图 6-3 2018-2020 年度收入能力与人群隐私风险指数关系

6.4 婚姻状况

由图 6-4 可知：2018-2020 年，未婚人群相较已婚人群隐私风险指数高。

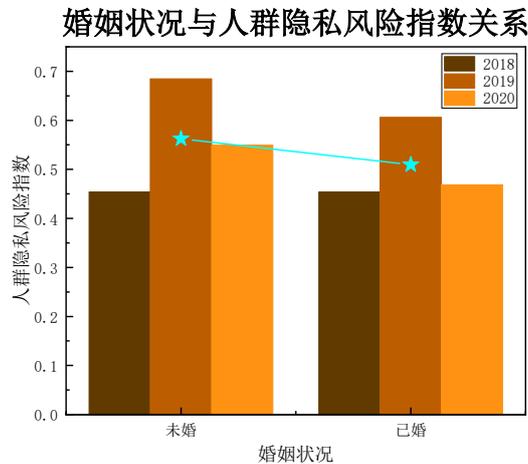


图 6-4 2018-2020 年度婚姻状况与人群隐私风险指数关系

6.5 年龄

由图 6-5 可知：2018-2020 年，26-35 岁人群相对其他年龄段人群隐私风险最低。

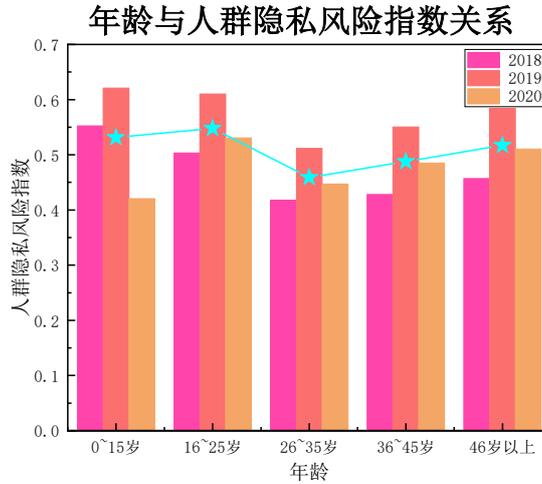


图 6-5 2018-2020 年度年龄与人群隐私风险指数关系

6.6 性别

由图 6-6 可知：2019 年之前，女性人群隐私风险偏高；但 2020 年，男性人群隐私风险略高。

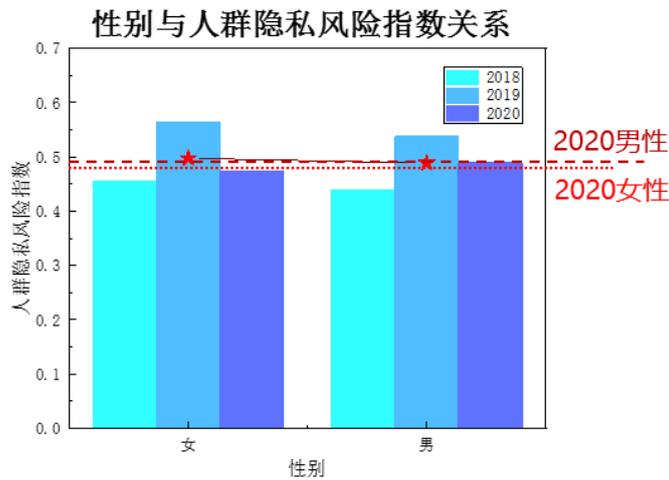


图 6-6 2018-2020 年度性别与人群隐私风险指数关系

七、行为隐私风险指数

相比 2018 年度数据集中各行为标签在全部用户中的分布较为均匀，2020 年同 2019 年的数据一样，部分标签的用户数呈现两级分化的特点，如教育行为中，超 38 万用户被标记为“儿童学习”，而“胎儿教育”的用户仅为 1 人。本报告仅对现有数据集中所体现的样本特征进行分析，并由此估计总体特征；同时，对基于各自的数据集以及分析模型得到的 2019、2020 年行为隐私风险指数的变化进行说明，并忽略数据集本身导致的系统误差。

对比 2019 年的分析结果，2020 年上述 12 类行为属性的隐私风险总体指数方差增大，隐私风险排名基本保持不变。在各类行为属性的具体分析中，不同倾向人群的人均 App 使用个数与隐私风险的变化呈现相同趋势，即该人群人均 App 使用个数越多，隐私风险指数越高。

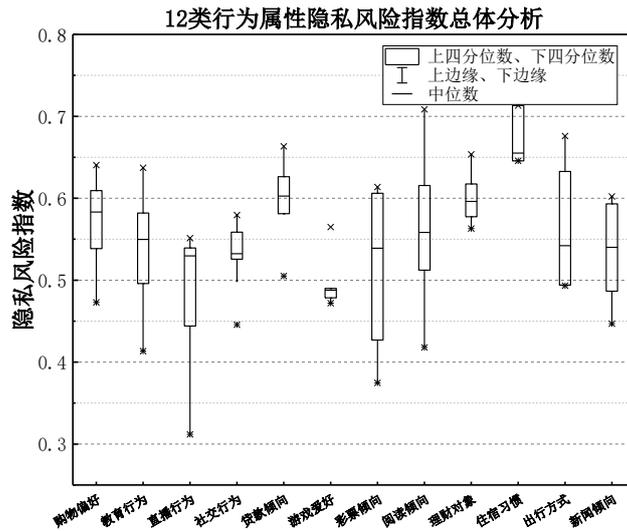


图 7-1 12 类行为属性隐私风险指数总体分析

如图 7-1 所示，2020 年度不同行为人群隐私风险分析得到如下结论：

- 不同购物偏好人群中，从事汽车专卖隐私风险由 2019 年的第七上升至第三。
- 不同教育行为中，从事 IT 教育隐私风险上升至最高，从事胎儿教育隐私风险降至最低。
- 不同直播行为中，喜欢收看美妆直播节目的人群隐私风险指数又降低至最低，收看明星直播节目的人群隐私风险指数提升至最高。
- 不同社交行为中，喜欢陌生人社交的人群隐私风险指数降低至最低。
- 不同游戏爱好行为中，偏好神话修真的人群隐私风险指数提升至最高，且明显高于其他类型。
- 不同阅读行为中，喜欢有声小说的人群隐私风险指数从 2019 年的第二降低至最低，喜欢金融知识的人群隐私风险仍然保持最高。
- 不同新闻倾向中，喜欢财经新闻的人群隐私风险仍然保持最高，喜欢综合新闻的人群隐私风险指数从 2019 年的第二降低至最低。

图 7-2 展示了 2018-2020 年度不同行为属性的隐私风险指数。分析可得，3 年来出行方式和贷款倾向对隐私风险的影响较大。

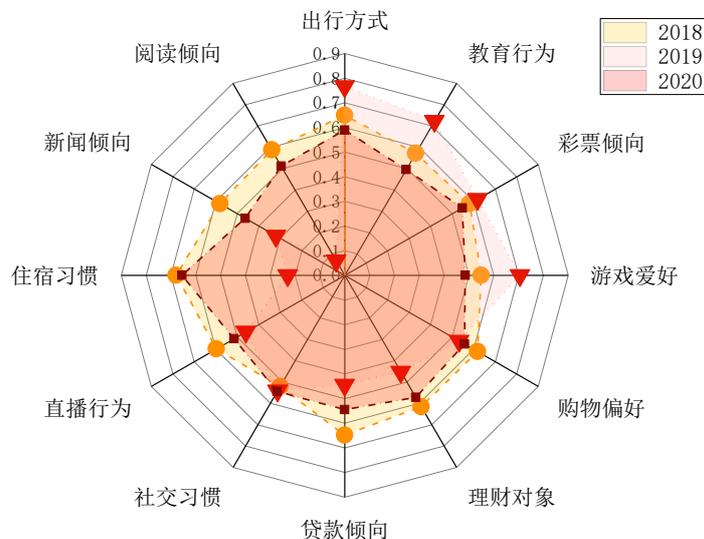


图 7-2 2018-2020 年度不同行为属性的隐私风险指数

图 7-3 展示了 2018-2020 年度的行为属性词云，字体越大表示隐私风险指数越高。分析可得以下结论：

- 2018 年，偏好阅读神话修真小说、喜欢民宿、大巴出行等行为人群隐私风险偏高；
- 2019 年，偏好星级酒店、飞机出行、喜欢返利购物、有车贷等行为人群隐私风险偏高；
- 2020 年，偏好星级酒店与民宿、大巴出行、阅读金融知识、有学生贷等行为的人群隐私风险偏高。



图 7-3 2018-2020 年度行为属性词云

八、总结

中国人民大学 WAMDM 实验室 2018、2019 年连续两年发布《中国隐私风险指数分析报告》，从数据拥有者（移动用户）、数据收集者（App 开发者）和移动应用（App）三个角度，分析了数据收集者获取的权限数据和造成的隐私风险，并从自然属性、社会属性、地域信息及行为属性多维度地分析数据拥有者（移动用户）的各群体隐私风险特征，从权限设置隐私 P1 分级和使用量隐私 P2 分级对应的不同级别 APP 分析其用户分布与隐私风险，从而构建中国隐私风险指数体系。本报告基于上述研究成果，基于 334 个地级市分层抽样的 3700 万用户和通过第三方网站爬取的 40 万 APP，进一步分析了 2020 年度中国隐私风险指数。

本报告已有成果主要如下：

- **大规模真实数据集：**约 3670 万用户集上的 179 个属性标签、40 万余 App 数据，相比前两年增加了约 670 万用户数据和 10 万 APP 数据。
- **四个主要贡献：**分析了 2018 至 2020 三年间中国隐私总体风险指数变化趋势、数据拥有者隐私风险变化趋势、数据垄断变化形势与 APP 发展趋势，揭示了当前中国居高难下的隐私风险状况与数据垄断形势。
- **三类分析对象：**数据拥有者（移动用户）、数据收集者（App 开发者）及移动应用程序（App）
- **六组分析结论：**大规模数据收集现状、移动应用程序分析、移动应用程序分级体系分析、区域隐私风险指数分析、人群隐私风险指数分析、行为隐私分析指数分析

后续将基于现有工作,进一步展开新一年度的分析,同时针对当前中国隐私风险指数体系中的不足加以改进,包括以下三个方面:

- 进一步开展 App 分级机制研究,扩宽分析维度,形成一套完整机制,促进其在实际应用程序市场落地;
- App 隐私风险预警系统构建;
- 和国家有关职能部门进一步沟通,为隐私相关政策的制定提供研究基础。

参考文献

- [1] Q Ye, H Hu, X Meng. PrivKV: Key-Value Data Collection with Local Differential Privacy[C]. IEEE Symposium on Security and Privacy (S&P). 2019:317-331.
- [2] M Zhu, Q Ye, X Yang, et al. Poster: AppPrivacy: Analyzing Data Collection and Privacy Leakage from Mobile Apps[C]. IEEE Symposium on Security and Privacy (S&P). 2019:41-42.
- [3] 孟小峰,朱敏杰,刘俊旭.大规模用户隐私风险量化研究[J].信息安全研究, Vol5(9):778-788, 2019.
- [4] 孟小峰,朱敏杰,刘立新,等.数据垄断与其治理模式研究[J].信息安全研究, Vol5(9):779-797, 2019.
- [5] 孟小峰,王雷霞,刘俊旭.人工智能时代的数据隐私、垄断与公平[J].大数据, Vol6(1): 35-46, 2020.
- [6] 中国消费者协会.《100款App个人信息收集与隐私政策测评报告》,2018-11-28
- [7] 朱敏杰,叶青青,孟小峰,等.基于权限的移动应用程序隐私风险量化[J].中国科学:信息科学, 2020. (Online)
- [8] J. Zang, K. Dummit, J. Graves, P. Lisker, and L. Sweeney, "Who Knows what About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps," Technology Science, 2015.
- [9] Hamed A, Ayed H K B. Privacy risk assessment and users' awareness for mobile Apps permissions[C]// Computer Systems and Applications. IEEE, 2017:1-8.
- [10] Hamed A, Ayed H K B. Privacy risk assessment and users' awareness for mobile Apps permissions[C]//Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference of. IEEE, 2016: 1-8.