

应用大数据防范互联网号贩子的实践

李超峰 马嘉潜 何 荣

肖华锋

(中山大学附属肿瘤医院信息科 广州 510080)

(中山大学精准医学中心 广州 510080)

何彩升

(中山大学附属肿瘤医院信息科 广州 510080)

〔摘要〕 以中山大学附属肿瘤医院利用大数据防范号贩子的实践为基础,从挂号异常、退号异常、异常抢号、异常用户绑定几方面分析技术型号贩子恶意占号的行为特征,制定对应的技术方案和应对措施,限制号贩子利用预约系统恶意挂号,确保医疗资源公平公正地分配到患者手中。

〔关键词〕 互联网;大数据;号贩子

〔中图分类号〕 R-056 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2017.02.014

Practice of Applying Big Data for Prevention of Web Registration Traffickers LI Chao-feng, MA Jia-qian, HE Rong, Information Department of Sun Yat-sen University Cancer Center, Guangzhou 510080, China; XIAO Hua-feng, Precision Medicine Center of Sun Yat-sen University, Guangzhou 510080, China; HE Cai-sheng, Information Department of Sun Yat-sen University Cancer Center, Guangzhou 510080, China

〔Abstract〕 Based on Sun Yat-sen University Cancer Hospital's practice of applying big data for prevention of registration traffickers, the paper analyzes the behavioral characteristics of technology-based registration traffickers' malicious occupation from the aspects of abnormal registration, abnormal withdrawal, abnormal grab and abnormal user binding, makes the corresponding technical plans and countermeasures, so as to limit the malicious registration of registration traffickers using the reservation system and guarantee the medical resources can be distributed to patients fairly and impartially.

〔Keywords〕 Internet; Big data; Web registration trafficker

1 引言

我国大城市的大医院尤其是北京、上海和广州等地的全国知名医院周围存在这样一个群体,他们

不为看病而来,但对医院就诊流程、诊疗特色和专家出诊时间等信息了如指掌。这个群体内部有着明确的角色分工和职责,运用现代与传统的方式实施专业化、集团化运作^[1],目的是垄断紧俏的专家号源,转手卖给急于就诊但苦于挂不上号的、被疾病折磨的患者,从而赚取高额差价^[2]。这群人被社会俗称为“号贩子”,号贩子的存在严重扰乱了正常医疗秩序,损害患者利益和医院声誉,浪费医疗资

〔修回日期〕 2016-10-22

〔作者简介〕 李超峰,硕士,发表论文 10 余篇。

源, 危及医疗安全^[3]。为解决号贩子问题, 国家卫生和计划生育委员会(卫计委)采用与公安机关联合打击、推进分级医疗体制、推行网上预约等方法进行治理^[4], 其中网上预约是各级卫计委、医院普遍推行的做法。政府和医院推出预约挂号原本秉持方便市民的初衷, 甚至希望能彻底解决号贩子问题; 未曾想却事与愿违, 挂号 App 等非但没有让号贩子望而却步, 反而被号贩子用作“抢号利器”^[5], 滋生了一批新的技术号贩子。

中山大学附属肿瘤医院现有 App、微信公众号、支付宝服务窗、自助机等多渠道挂号, 虽然医院很早就实行全面实名制预约挂号, 但也面临着相同的问题, 还有不法号贩子在倒卖医院热门专家号。为创建公平公正的就诊环境, 维持良好的就诊秩序, 维护患者权益, 提升患者就医感受, 医院基于诊疗过程的医疗大数据^[6], 详细分析号贩子数据的潜在特征, 提前进行预防干预, 从技术层面尽可能封堵号贩子可以利用的漏洞。

2 医疗大数据防范号贩子的思路

如何从预约挂号、诊疗过程的全部数据中, 准确地分析出哪些是恶意的占号行为, 是防范号贩子的基础。大数据的核心思想是尽可能利用全量数据, 进行细节分析, 挖掘数据的潜在价值^[7]。中山大学附属肿瘤医院充分利用全量的挂号、诊疗数据, 以大数据分析技术为工具, 将数据特征和业务逻辑结合, 反复进行数据建模分析和论证, 通过异常值分析确定了“挂”、“抢”、“退”、“绑”各种异常行为及应对措施: (1) 分析挂号数据, 识别异常挂号账户, 限制异常账户挂号。(2) 分析退号数据, 甄别异常退号行为, 追踪非法账户, 限制其挂号。(3) 分析抢号数据, 识别机器抢号和人工抢号, 限制号贩子抢号。(4) 制定退号策略, 将退号产生的二次号源, 在一定时间段内随机释放, 让号贩子退号再抢占失去效用, 确保号源信息公平地分配到真正需要的患者手中。

3 技术型号贩子挂号行为特征分析

3.1 挂号异常行为

在肿瘤医院就诊的患者主要包含两大类: (1) 治疗过程中的患者, 表现为集中在一小段时间内的频繁挂号。(2) 治疗结束后的患者, 基本上进行治疗后的定期复诊, 其挂号频率不会很高, 基本 3 个月到半年 1 次, 其挂号专家相对固定, 都是就诊的主治医师或教授。因此, 医院以单个患者年度挂号数据为计量区间, 通过大数据的可视化技术进行分析, 可以明显看到挂号异常的患者。图 1 是中山大学附属肿瘤医院 2015 年全年的挂号数据, 从挂号数量分布可以看出: 年挂号量在 50 ~ 100 之间的, 属于轻度可疑对象, 需要进一步结合诊疗行为信息; 年挂号量在 100 ~ 150 之间的, 属于中度可疑行为, 分析其挂号科室等信息可以基本判断其属性; 年挂号量超过 150 个的, 基本就能判定为恶意抢占号源账户。

3.2 退号异常行为

正常的挂号都以到医院就诊看病为目的, 因此患者频繁退号也是技术型号贩子的重要特征。异常退号行为在数据上表现为: (1) 1 个月或一段时间出现多次退号记录。(2) 判定医疗行为, 挂号的目的是为了就诊, 如果挂号中大部分没有产生诊疗行为就退号, 即可判定为恶意抢占资源。图 2 是中山大学附属肿瘤医院 2015 年全年的退号数据, 从单个患者退号数量分布可以看出: 年退号量在 50 ~ 100 个之间的属于高度怀疑对象, 需要结合医疗情况进行甄别分析; 年退号量超过 100 个的, 基本可以判定为号贩子的异常退号、抢号行为。

3.3 异常抢号行为

中山大学附属肿瘤医院每周日早上 8:00 准时释放未来两周的专家号源。一些号贩子为了保证抢号, 配置专用服务器、高速网, 甚至专门的抢号软件, 在放号的瞬间进行抢号。系统在抢号过程中记录抢号的主机、IP 等相关设备信息, 通过对单个设

备的挂号时间数据进行分析, 可以获得异常抢号行为。如果单次放号的前几个预约号都来自同一地址或主机, 或者多次放号的前几个号都被同一地址或设备抢到, 显然就是异常的数据。图 3 是中山大学附属肿瘤医院 2015 年全年的抢号数据, 从单个患者 8:00 的抢号数量分布可以看出: 1 年内连续在 8:00 抢号超过 80 个明显偏离正常范围, 有显著的号贩子特征。

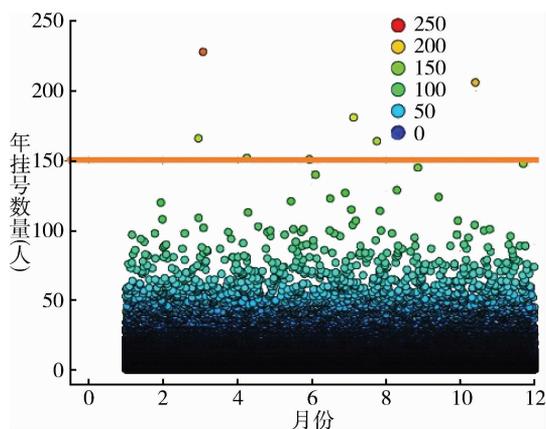


图 1 挂号大数据分析

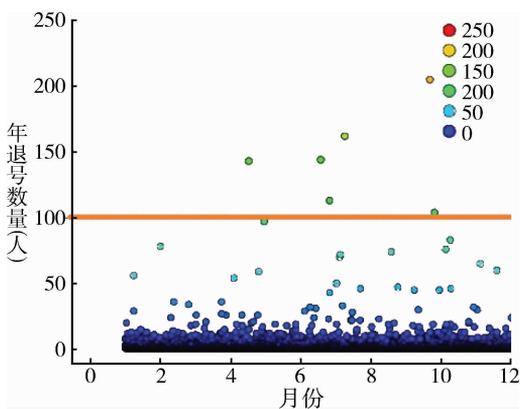


图 2 退号大数据分析

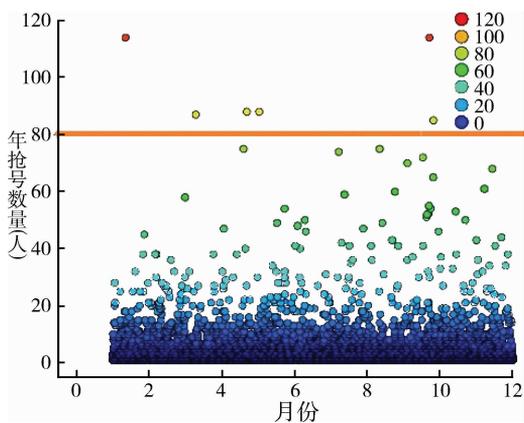


图 3 抢号大数据分析

3.4 异常用户绑定

为了便于老人、小孩等特殊患者就诊, 医院制定允许绑定亲友身份信息进行代为挂号的政策。这些便民的功能也给号贩子钻了空子, 号贩子利用从其手中购买过挂号票的真实患者的信息, 恶意绑定患者信息并抢占号源 (系统设计初期未能记录每个用户绑定和解绑的历史记录, 仅能分析当前绑定患者数量)。图 4 是中山大学附属肿瘤医院 2015 年全年的账号绑定数据, 从单个用户绑定患者数据分布可以看出: 绑定超过 5 个患者身份的账户, 具有高度异常特征; 其中有个别账户绑定 16 个身份证之多, 而且身份证信息天南地北, 明确显示出号贩子特征。

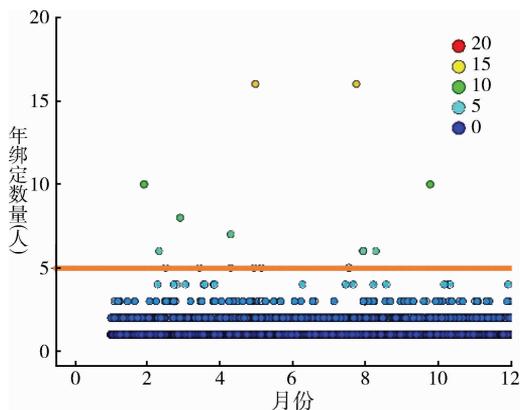


图 4 单个账号绑定患者大数据分析

4 防范号贩子的应对措施

基于大数据分析发现, 互联网预约环境下号贩子主要利用退号、绑定、抢号等环节的漏洞进行技术抢号。医院管理和技术部门针对这些情况制定应对方案, 在单个账号管理、号源管理、退号管理等维度上进行封堵和限制, 在号源统一管理、客户端身份认证和限制等技术层面进行管控。主要措施包括: (1) 完善挂号管理, 对单个用户进行挂号数量控制, 如每个月最多只能挂 8 个专家号等措施, 控制单个账户的抢占行为, 有针对性地改善放号管理, 提供更多的预约机会。(2) 完善退号管理, 建立随机二次放号规则。所有退掉的号源, 在当天随机放出, 避免退号 - 抢号模式的号贩子出现。(3)

完善账号绑定规则, 每个账号最多绑定 5 个患者、累积不超过 10 个患者, 超过上限必须经过人工验证、确认特殊情况后才能继续绑定。(4) 对各种异常账号进行黑名单管理, 报各主管部门进行处理。(5) 完善挂号验证体系, 患者在注册时与公安部门联网验证患者身份信息的正确性; 患者挂号时记录客户端信息, 根据客户端的使用频次、模式进行操作行为分析, 对异常操作模式进行记录和人工核实; 挂号确认时增加注册码验证, 保证非机器自动操作; 增加挂号页面和服务的日志记录和访问次数预警, 防范号贩子通过程序刷号方式进行攻击。(6) 完善现场、电话等多种预约方式, 优化预约挂号流程, 使多种挂号方式有机整合, 避免号贩子钻漏洞。经过改造, 医院所有预约使用统一号源、统一的 Web 服务, 可通过对后台 Web 服务进行监控分析各种挂号方式的异常情况, 进行防范。(7) 完善大数据分析方法和持续监控和识别异常行为并及时处理。针对已经发现的异常模式, 开发可视化的服务运行监控和报警程序, 对于被异常调用的服务, 通过可视化界面、短信等方式通知管理部门。信息部门也应成立数据分析小组, 定期分析业务数据, 识别新的风险模式。

5 结语

在“互联网+”医疗的潮流面前, 越来越多的医疗服务正在通过互联网实现, 这些新的应用在带来便利性的同时也带来了一些新的问题, 如新的技术型号贩子的产生。随着号贩子等逐步利用高科技手段, 医院也需要提高新技术的应用水平。保证医疗资源的公平合理分配和改善就医体验是一条漫长的道路, 任重而道远, 医院管理和信息部门唯有持续投入、严格管理, 利用大数据等新技术进行多维度的监控, 才能使号贩子的行为无处藏身。

参考文献

- 1 王霞, 俞光岩. 呼吁重拳打击“号贩子” [J]. 中国当代医药, 2012, 19 (8): 4.
- 2 李颖. 医院号贩子的灰色产业链条 [J]. 中国质量万里行, 2010, (7): 72-73.
- 3 陈虹, 贾小溪. 医院号贩子的成因、危害及遏制措施 [J]. 中国医院, 2014, 18 (11): 54-56.
- 4 国家卫计委: 近期出台方案严打号贩子 [EB/OL]. [2016-08-10]. <http://business.sohu.com/20160225/n438421916.shtml>.
- 5 冷蓝. “挂号 APP” 为何成号贩子抢号利器 [J]. 时代金融, 2015, (22): 12.
- 6 张振, 周毅, 杜守洪, 等. 医疗大数据及其面临的机遇与挑战 [J]. 医学信息学杂志, 2014, 35 (6): 2-8.
- 7 维克托·迈尔·舍恩伯格, 周涛. 大数据时代—生活、工作与思维的大变革 [J]. 人力资源管理, 2013, (3): 174.

(上接第 51 页)

- 5 Mitchell DJ. A Model of Information Therapy Definition and Empirical Application [D]. Denton: University of North Texas, 1999: 15.
- 6 Kemper D W, Mettler M. Information Therapy: prescribing the right information to the right person at the right time [J]. Managed Care Quarterly, 2002, 10 (4): 43-46.
- 7 ‘Information Therapy’ Fuels Wise Consumer Healthcare Choice. Consumer Driven Healthcare, v3 (11) [EB/OL]. [2016-05-10]. <http://www.nhionline.net>.
- 8 US National Library of Medicine. The Health Information Prescriptions [EB/OL]. [2016-05-10]. http://www.nlm.nih.gov/news/press_releases/GAhealthRX03.html.
- 9 Mettler M, Kemper D W. Information Therapy: prescribed

- information as a reimbursable medical service [M]. Idaho: Healthrise, 2002: 3.
- 10 Williams MD, Gish KW, Giuse NB, et al. The Patient Informatics Consult Service (PICS): an approach for a patient-centered service [J]. Bull Med Libr Assoc, 2001, (89): 185-193.
- 11 Mettler M, Kemper DW. Information Therapy: the strategic role of prescribed information in disease self-management [J]. APLAR Journal of Rheumatology, 2005, (8): 69-76.
- 12 Koop C. Plenary Address. Delivered at the American Medical Informatics Association’s Eighteenth Annual Symposium on Computer Application in Medical Care [M]. Washington DC: Hanley&Belfus, 1994.