
数据匿名化或不足以保护个人隐私 《自然-通讯》

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/6199.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

数据匿名化或不足以保护个人隐私 《自然-通讯》。《自然-通讯》发表的一篇论文 Estimating the success of re-identifications in incomplete datasets using generative models 介绍了一种可以评估一个人的身份是否能够从一个不完整的匿名化数据库中被重新识别出来的方法。该论文认为目前的匿名化和数据共享方法可能不足以保护个人隐私或满足数据保护法律法规的要求，如欧盟的《通用数据保护条例》(GDPR)。



图1图源：VisualHunt

数据科学和人工智能有望变革我们日常生活的方方面面，如医疗、卫生保健、商业和治理。这些方法依赖于大规模的详细个人数据，但是收集和共享个人数据引发了有关个人隐私的担忧。针对

这个问题，目前的解决方法包括匿名化处理和公布不完全的数据集。但是，近期发生了利用匿名数据集，包括浏览历史记录、手机和信用卡数据，成功重新识别出个体身份的情况，这表明上述措施可能还不够充分。

英国帝国理工学院的Yves-Alexandre de Montjoye及同事开发了一种统计方法，能够准确估算通过匿名数据集正确地重新识别个体身份的可能性。作者发现，只需要知道少数几个属性，如邮政编码、出生日期、性别和子女数量，一般就能够以高可信度重新识别出个体身份——即使数据集是不完整的。已知属性越多，识别的可能性越大。例如，99.98%的马萨诸塞州人口可以通过15个人口统计学属性识别出来。因此，他们总结认为只公布取样数据集或不完全数据集不足以保护个人隐私。

论文标题：Estimating the success of re-identifications in incomplete datasets using generative models

期刊：Nature Communications

作者：Luc Rocher, Julien M. Hendrickx, Yves-Alexandre de Montjoye

发表时间：2019/07/23

数字识别码：10.1038/s41467-019-10933-3

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发