5 隐私保护

引言

互联网的迅猛发展改变了人们的生活方式，诸如电子商务、社交网络、电子政务、智能出行、网络教育等领域互联网产品的蓬勃发展以及网络的日益普及，将人们的日常生活和网络融为一体。当用户逐渐对购物推荐、智能打车、路线导航、信息搜索等日常便捷服务产生依赖的同时，已经将大量个人数据保存在各互联网平台上，网络空间中的数据呈现爆炸式增长。但是近年来，信息泄露事件频发，非法采集、窃取、贩卖和滥用网络个人信息已经趋于产业化和智能化，人们在享受大数据带来便利的同时也承担着隐私泄露的风险。2018 年，Facebook 被爆出有 8700 万用户私人信息被剑桥分析公司不正当使用，而这或许只是数据被滥用的冰山一角；数月后，Facebook 又接连被暴露出两个可被利用进行隐私信息获取及未授权访问的安全漏洞。不难发现，互联网上的类似事件数不胜数。一旦隐私被泄露，用户的人身财产将直接受到威胁，合法权益可能受到严重侵犯，甚至思维方式也可能被分析利用。

如今，互联网上隐私被广泛侵犯的问题已经引发互联网用户的普遍担忧，并成为大众关注的焦点。实际上，互联网广告等商业模式之所以取得巨大的成功，就是因为**可以免费利用用户搜索关键词等隐私信息，从而可以给用户精准推送广告。**可以说，所有能够取得成功的互联网商业模式，都或多或少侵犯了用户隐私。近年来，随着用户隐私保护意识的增强，各类隐私保护技术开始迅速发展，例如，利用可信执行环境（Trusted Execution Environment，TEE）等实现“数据可用不可见”的隐私计算最近成为了大家关注的焦点。本章中，我们将对数据中隐私信息的保护问题展开讨论，带领读者了解多种隐私保护技术。

在全球互联网公司中，排名前列的 Google 公司 2020 年的广告收入达到 1469.2 亿美元，占该年总收入的 80.9%，可以说是名副其实的广告公司。

我们假设有以下几个场景：医院想要发布一个医疗数据库数据供其他机构研究，显而易见，医疗信息一定是用户的隐私信息。那么医院该如何处理数据

来保证患者信息不被泄露呢？最简单的想法是将容易推测出患者身份的属性删去，使攻击者无法分析出某条数据的归属者。但这样就一定能保护患者的隐私吗？如果该数据库还对外提供统计服务，例如用户可以查询某种疾病的患病人数，这种情况下是否会泄露患者信息呢？如果医院因为自己没有足够的计算能力需要委托其他机构对数据库中的信息进行计算，又该如何保证患者信息不被泄露呢？通过对本章的学习，相信读者可以找到这些问题的答案。

本章概述数据隐私保护技术，对多种隐私保护技术的思想与基础知识进行讲解。全章分为 5 节。5.1 节从隐私的概念出发，揭示隐私泄露的危害并介绍数据隐私保护技术分类。接着，5.2 节 ～ 5.4 节分别介绍了匿名化、差分隐私和同态加密技术。匿名化通过隐藏用户身份和数据的对应关系来保护隐私，但却没有严格的数学证明作为保证。差分隐私则提供了严格的可证明的隐私保护模型，通过在查询结果中添加噪声来保护个体隐私。同态加密可以直接处理加密数据，保护了数据计算过程的安全。5.5 节介绍了安全多方计算技术，它是一种解决互不信任的参与方之间进行协同计算的隐私保护框架。最后回顾本章内容并进行小结。表 5.1 为本章的主要内容及知识框架。

表 5.1 隐私保护技术概览

隐私保护	知识点	概要
隐私保护技术初探	• 隐私的概念 • 隐私泄露的危害 • 隐私保护技术分类	隐私泄露带来的巨大危害促进了隐私保护数据挖掘的发展
匿名化	• 匿名化隐私保护模型 • 匿名化方法	限制发布：隐藏用户身份和数据的对应关系
差分隐私	• 差分隐私基础 • 数值型差分隐私 • 非数值型差分隐私	数据失真：在查询结果中加入噪声来保护个体隐私
同态加密	• 同态加密基础 • 半同态加密 • 全同态加密	数据加密：处理加密数据，将在密文上进行的操作映射到明文上
安全多方计算	• 安全多方计算基础 • 百万富翁协议	隐私保护框架：实现保护隐私的协同计算

5.1 隐私保护技术初探

简单来说，**隐私是个人或者团体不愿被他人知晓的信息，**这一概念是主观且动态变化的。网络空间中的隐私包括个人数据、网络行为数据和通信内容数据。近年来频频发生的隐私泄露事件使得隐私保护逐渐成为了大众关注的焦点。本节我们将从隐私这一概念讲起，带领读者了解隐私泄露的危害，并介绍数据隐私保护技术的分类。

读者可以自己想一想，哪些信息是自己的隐私信息。身份证号？每月的收入？最爱去的餐馆？还有最爱喝的饮料？其实这些都应该算是用户隐私信息。

5.1.1 网络空间中的隐私

在原始社会时期，面对残酷的环境，人类是没有隐私可言的，但在不断进化中，当人类开始用树皮、动物皮毛等来遮挡身体的隐私部位时，隐私意识的萌芽就诞生了。此后，隐私以各种各样的形式存在于人们的生活中，比如将屏风放在房间中遮挡出个人生活空间，使用床幔封闭睡觉的空间等。虽然人类早就意识到了隐私的重要性，然而将隐私作为人的一项基本权利的意识直到西方自由主义观念盛行后才得到重视。**1890 年，当时的美国报纸行业为了迎合大众口味，大量报道各种犯罪、丑闻、名人私生活等新闻，**美国学者塞缪尔·沃伦（Samuel D. Warren）由于不满报纸对其家庭生活的报道，和路易斯·布兰代斯（Louis D. Brandeis）共同执笔在 *Harvard Law Review* 期刊上发表了一篇名为 *The Right To Privacy* 的文章，并在该文中使用了“隐私权”（The Right To Privacy）一词，这被公认为是隐私权的首次公开提出。

这一时期被称为美国新闻事业的“黄色新闻时期”。

随着社会科技和媒介形式的不断发展，网络已经成为了存储隐私数据的聚集地，人类隐私的范围已经从现实空间发展到了虚拟空间。由此可见，隐私的内容与形式是随着社会发展不断变化的。如今，保护隐私已被当作一项权利，世界各国都已颁布了多部关于保护公民隐私权的法律。

在古汉语中，“隐”的主要含义为隐藏、隐避，“私”的主要含义为私人、私下。《汉语词典》对隐私的解释是不愿告人或不愿公开的个人的事。百度百科则解释为与公共利益、群体利益无关，不愿告人或不愿公开的个人的私事。简单来说，隐私就是个人或者团体不愿被他人知晓的信息，这一概念是主观且动态变化的，每个人甚至不同国家的民众对隐私的定义都存在差异。比如，不同的人可能对身高、体重是否属于隐私有不同的理解。再比如，根据波士顿咨询公司针对多个国家如何看待个人数据隐私这一问题的调查结果来看，亚洲（如日本）民众的隐私观念相比其他国家（如美国、英国、法国等）要淡薄一些[1]。

生活的强烈和复杂，伴随着逐渐进步的文明，让人们偶尔需要从外界逃离。在不断完善的文化影响之下，人类对公众的感知越来越敏感，所以独处与隐私成为了个体的基本需求。
——《隐私权》

在现实生活中，个人的身份信息、工作信息和生活信息等都是隐私，而在

网络空间中，隐私主要包括如下数据。

（1）个人数据：与个人身份有关的数据，比如个人证件信息、账号信息等。

（2）网络行为数据：个人在使用网络提供的服务时产生的数据，比如进行网络社交产生的信息、浏览网页产生的信息等。

（3）通信内容数据：个人使用网络进行通信时产生的数据，比如电子邮件、在线选举等。

这些携带着网民个人隐私的数据一旦被泄露，势必会为网民带来损害。因此，为了保护人们在使用网络服务时的隐私权，目前已经有 120 多个国家或地区制定了有关隐私保护的法律规定。早在 1970 年，德国就颁布了《联邦数据保护法》，美国在 1974 年颁布了《隐私权法》，英国在 1984 年颁布了《数据保护法》。近几年，我国也相继颁布了多部法律法规：2017 年颁布了《网络安全法》，2020 年发布了新版《个人信息安全规范》，2021 年 6 月召开的第十三届全国人大常委会上表决通过了我国第一部保护数据安全的法律——《数据安全法》。另外，《个人信息保护法》也已于 2021 年 11 月起正式施行。截至目前，欧盟在 2018 年 5 月 25 日生效的《通用数据保护条例》（*General Data Protection Regulation*，*GDPR*）是在隐私保护领域应用范围最广以及最受关注的一部法律。GDPR 规定，就算企业没有直接在欧洲开展业务，也没有在欧洲设立任何的分支机构，只要是处理的数据涉及欧洲公民的个人数据，就必须遵守其法律条款。**一旦企业违反了该数据保护条例，就要面临被罚款的风险，**罚款最高可达到 2000 万欧元或者企业上一年全球营业总额的 4%，以其中较高者为准。

截至 2019 年 9 月 24 日，欧洲数据保护机构已经开出了约 3.7 亿欧元（约合 29 亿元人民币）的罚单，涉及的知名企业包括英国航空公司、万豪集团、Google 公司等。

5.1.2 隐私泄露的危害

《2018 年数据泄露水平指数》调查报告显示，仅在 2018 年上半年发生的较大型数据泄露事件就多达 945 起（其中有 6 次重大泄露事件发生在社交媒体领域），这些数据泄露事件涉及的数据记录总量超过了 45 亿条。与 2017 年同期数据显示的 19 亿条相比，2018 年上半年的数据泄露条数涨幅已经超过 130%。近些年来，全球数据泄露的数目正在呈现逐年上涨的趋势，频频发生的隐私泄露事件已经引起了公众对于数据隐私的高度关注。通过回顾几个备受瞩目的隐私泄露事件，让我们一起思考隐私泄露带来的危害。

陷入丑闻的剑桥分析公司在 2018 年 5 月正式启动破产程序，停止了所有运营。而 Facebook 则股价暴跌，最终被美国联邦贸易委员会罚款 50 亿美元。

2018 年 3 月，**美国纽约时报称，Facebook 用户的个人信息未经用户许可就被剑桥分析公司擅自使用，随后 Facebook 承认有 8700 万用户个人信**

息被该公司进行了不正当使用。媒体称这些数据被用来精准地向用户投放广告从而影响其政治倾向，帮助 2016 年特朗普团队参选美国总统，并且 Facebook 早已知晓但并没有将此事告知公众。同年 9 月份，Facebook 称黑客可利用漏洞获得 3000 万 Facebook 用户的账号信息，12 月份，Facebook 又宣布了可利用第三方应用程序访问近 700 万用户未公开发布的照片的漏洞。

2018 年 6 月，美国市场营销和数据聚合公司 Exactis 被发现将一个数据库放在了可公开访问的服务器上。该数据库包含的 3.4 亿条记录涉及上亿美国成年人和数百万公司的信息，总量超过 2TB。这些被暴露在网络上的个人信息非常详细，包括电话号码、住址、邮箱，以及诸如个人爱好、习惯等与个人特征高度相关的信息。它们的泄露使公民、家庭和企业可以轻松地被画像、模仿或追踪。

类似的隐私泄露事件数不胜数，它们对用户、企业甚至国家造成了巨大的危害。首先，用户人身财产安全受到威胁，犯罪分子可利用用户信息进行电信诈骗、非法讨债甚至绑架勒索等犯罪活动。其次，用户思想可能被操控，剑桥分析公司通过分析收集的信息对用户实现精准广告投放，旨在干预美国总统大选，这毫无疑问侵犯了用户的合法权益。除此以外，隐私泄露可能对企业甚至国家造成危害，以 2016 年发生的土耳其国民信息数据库泄露事件为例，在该事件中，近 5000 万公民（包括现任总统）的姓名、身份证号、父母亲姓名、性别、出生日期、居住地详细地址等隐私被泄露。要知道 2016 年该国总人口只有不到 8000 万，该事件使土耳其近三分之二的人口面临欺诈和身份盗用等风险。

5.1.3 隐私保护技术介绍

计算机技术的飞速发展使网络空间中的数据呈现指数级增长，在这个“掌握了数据就是掌握了未来”的时代，迫切地想要从海量数据中找到有价值信息的需求推动了数据挖掘（Data Mining，DM）技术的发展，这是一种从大量数据中发现有用知识（模型或规则）的技术。然而，**频频发生的数据泄露事件所造成的巨大危害让人们看到了隐私保护的重要性，**因此如何在保证用户隐私不被泄露的情况下从数据中挖掘出更多的价值，即实现隐私保护数据挖掘（Privacy Preserving Data Mining，PPDM），已经成为了数据挖掘领域的研究热点之一。在进行数据挖掘时实现隐私保护要从两方面进行考虑：一方面，要对诸如姓名、身份证号这种直接泄露用户隐私的数据进行修改或者删除；另一方面，还要限制能够用挖掘算法挖掘出的敏感知识，因为此类知识也会泄露用户隐私。

在继续往下阅读之前，建议读者可以停下来稍作思考，哪些技术可以用来保护隐私？特别要记住，保护隐私的同时我们还需要从数据中挖掘价值。

根据使用的隐私保护技术的不同，隐私保护数据挖掘主要分为三类，分别是基于限制发布、基于数据失真和基于数据加密的隐私保护技术[2–4]。

（1）基于限制发布的隐私保护：这类技术的思想是有选择地发布原始数据，不发布或者发布精度较低的敏感数据，通过将发布敏感数据带来的隐私风险控制在一定范围内来实现隐私保护，并且在此过程中要保证隐私泄露风险和数据可用性的平衡。目前该类技术的研究集中于“数据匿名化”：一是研究匿名化原则，使遵循此原则发布的数据在具有较大利用价值的同时可以保护隐私；二是研究如何针对匿名化原则设计更高效的匿名化算法[2]。

（2）基于数据失真的隐私保护：这类技术的思想是对原始数据进行扰动，但要保证在修改数据的同时数据的某些统计性质保持不变，以便之后进行数据挖掘时能够从失真数据中得到有价值信息。另外，用该类技术处理后的数据还要保证攻击者无法利用失真后的数据重构出原始数据。基于数据失真的隐私保护技术主要包括随机化、交换、阻塞、变形等技术。随机化的思想是通过在原始数据中加入随机噪声来隐藏真实的敏感数据。数据交换的思想是在保证某些统计性质不变的情况下通过在记录之间交换数值来扰动真实数值。

（3）基于数据加密的隐私保护：这类技术的思想是对原始数据进行加密，通过将明文转化成无意义的密文来防止数据挖掘过程中出现隐私泄露。由于加密技术可用于解决通信的安全问题，因此这类技术被广泛应用于分布式应用环境中。分布式应用采取两种存储数据的模式，分别为垂直划分的数据模式和水平划分的数据模式。垂直划分数据是指分布式环境中的每个参与者只存储数据的部分属性，所有参与者存储的属性不重复；水平划分数据则是指每个参与者存储具有完整属性的数据记录，但是所有参与者存储的数据不重复。这两种存储模式的相同之处是每个参与者只掌握了部分数据，因此在实现隐私保护数据挖掘的过程中会涉及保证数据的安全传递以及互不相通的问题。表 5.2 是这三类隐私保护技术的对比。

表 5.2 三类隐私保护技术的对比

隐私保护技术	优点	缺点
基于限制发布	发布的数据真实可靠	数据丢失部分信息
基于数据失真	算法效率较高	数据丢失部分信息
基于数据加密	数据安全性和准确性较高	计算开销大

实现隐私保护数据挖掘与使用的具体的挖掘算法有关，我们还可以**根据使用的挖掘算法的不同，将隐私保护数据挖掘分为隐私保护的关联规则挖掘、隐私保护的分类挖掘和隐私保护的聚类挖掘**[5–8]。

（1）隐私保护的关联规则挖掘：关联规则反映了事件之间的关联，可以利用关联规则从一件事情的发生推测另一件事情的发生。关联规则挖掘则是从大量数据中挖掘数据项之间隐藏的关系，发现数据集中项集之间的关联和规则的过程。

项集是关联规则中的基本概念，数据中的属性称为项，多个项的集合称为项集。

关联规则挖掘已经被广泛应用于人们的生活中，帮助人们进行决策和管理。比如**作为其应用之一的“购物篮分析”，可帮助商家通过分析用户购物篮中的商品研究用户的购买行为，找到购买的商品之间隐藏的关联规则，**这对帮助商家进行决策起到了至关重要的作用。隐私保护的关联规则挖掘是要在保证非敏感规则容易被发现的同时抑制对敏感规则的挖掘。根据隐私保护对象的不同，隐私保护的关联规则挖掘可分为对敏感数据的保护和敏感规则的保护。

沃尔玛超市对顾客的购物行为进行“购物篮分析”后发现，跟尿布一起购买最多的商品竟是啤酒！这是因为奶爸们下班后经常要到超市去买尿布，而他们中有一部分人会随手买些啤酒来犒劳自己。

（2）隐私保护的分类挖掘：分类挖掘利用有标签（即数据类别）的数据集构建分类模型，通过将数据项映射到给定的类别中以实现类别预测。但是由于分类结果可能会暴露数据集中的隐私信息，因此需要实现隐私保护的分类挖掘，即在降低敏感数据分类准确性的同时不影响非敏感数据的分类准确性，最终得到一个分类准确且无隐私泄露的分类模型。

（3）隐私保护的聚类挖掘：聚类挖掘利用无标签的数据集构建分类模型，该模型利用数据相似性将数据分为多个类别或者多个簇，在最后的分类结果中，同一类别中的数据相似性越高越好，不同类别中的数据相似性越低越好。但是由于聚类结果同样可能会暴露数据集中的隐私信息，因此需要实现隐私保护的聚类挖掘，即在保证准确的聚类结果的同时保护数据隐私，达到准确性和隐私保护之间的平衡。

学习完本节，我们已经对隐私保护技术有了初步了解，接下来我们将具体介绍三种典型的隐私保护技术：匿名化、差分隐私和同态加密，以及一种隐私保护框架——安全多方计算，它可以确保多个参与方在保护隐私的前提下进行协同计算。

5.2 匿名化

如果医院想要安全地发布病患的医疗数据集供其他机构研究，如何处理数据才能保证用户隐私不能被泄露呢？能想到的最直接的做法是不是将数据集中

容易推测出用户身份的属性删除或者替换，使攻击者无法根据处理后的数据推测出拥有该数据的用户呢？从表面上看，这种方法比较简单而且似乎能够达到一定的效果，可是如果我们拥有另一个与此数据集部分属性重叠的数据集，这种方法还能保护用户隐私吗？显然是不能的，我们称这种攻击方式为链接攻击。**为了抵抗链接攻击，更好地实现数据匿名化，隐藏用户身份和数据的对应关系，研究人员提出了一系列匿名化隐私保护模型，而实现匿名化的操作主要有两种，分别是泛化和抑制。**

5.2.1 匿名化隐私保护模型

在介绍匿名化技术之前，我们先来了解数据表中涉及的几个概念：

（1）标识符（Identifiers）：唯一标识个体身份的属性或者属性的集合，如姓名、身份证号等。

本章中的病患信息是完全虚构的，如有雷同，纯属巧合。

（2）准标识符（Quasi-Identifiers，QID）：与其他数据表进行链接以标识个体身份的属性或属性组合，如性别、出生日期、邮政编码等。准标识符的选择由进行链接的外部数据表决定。

（3）敏感属性（Sensitive Attributes，SA）：发布时需要保密的属性，如工资、健康状况等。

表 5.3 是一个病患的医疗数据集，其中，姓名为标识符，疾病为敏感属性，其他属性的组合为准标识符。若想安全地发布该医疗数据集供其他机构研究，需要隐藏用户身份和数据的对应关系，我们能想到的一个简单方法是删去表中容易关联到患者本人且研究价值不大的属性，即姓名和家庭住址，或者将姓名替换为假名，如表 5.4 所示。但是这种操作能保证病患的隐私不被泄露吗？稍作思考我们就知道答案是否定的，因为**准标识符可能被攻击者用来与其他能够获得的公共数据集联系起来，获得隐私信息。**例如，表 5.5 是攻击者掌握的选民信息表，我们发现攻击者可以根据选民信息表中的性别、年龄、出生年月和邮政编码链接发布的表 5.4 匿名病患信息，来确定王宇患糖尿病。也就是说，如果攻击者掌握的两个数据集中有重叠的属性，就有可能通过链接攻击推断出用户的信息。

该研究曾用美国麻省总医院出院数据匹配选举投票注册数据，链接出了麻省某议员的住院信息。

2000 年，卡内基-梅隆大学教授拉坦亚 • 斯威尼（Latanya Sweeney）在 *Simple demographics often identify people uniquely* 报告中提到，在已知美国选举人公共注册信息的基础上，基于邮编、性别和出生日期就有可能识别出 87% 的美国人的个人身份，因此这些数据字段的公开可能会导致隐私泄露。2006 年，

美国在线（AOL）为了学术研究公开了 65 万名用户在 3 个月内的 2000 万次搜索请求。虽然公开数据中的用户名都被随机 ID 代替，但是实践表明，如果对某一用户的搜索记录进行分析，仍然有可能判断出其身份和行为。这两起真实的案例说明**简单地删除敏感字段或者将姓名替换为假名并不能保证个人隐私的安全。**

请思考：既然简单地删除或者替换标识符无法抵抗链接攻击，那该如何处理数据来保证匿名呢？

表 5.3 病患的医疗数据集

姓名	性别	年龄	出生年月	邮政编码	家庭住址	疾病
张艳	女	36	1984-12	235023	海河市泾县	流感
李磊	男	42	1978-06	235152	江宁市凤城	脂肪肝
王宇	男	35	1985-02	152030	丰宁市沙县	糖尿病
赵静	女	34	1986-02	154263	江宁市安县	哮喘

表 5.4 匿名后病患的医疗数据集

性别	年龄	出生年月	邮政编码	疾病
女	36	1984-12	235023	流感
男	42	1978-06	235152	脂肪肝
男	35	1985-02	152030	糖尿病
女	34	1986-02	154263	哮喘

表 5.5 攻击者掌握的选民信息表

姓名	性别	年龄	出生年月	邮政编码	登记日期
钱晶	女	24	1996-05	256230	2020-07
王宇	男	35	1985-02	152030	2020-07
周平	男	32	1988-12	152630	2020-07
秦桦	女	29	1991-03	152410	2020-07
吴沛	男	31	1989-08	256230	2020-07

为了防御链接攻击，研究人员对匿名化展开了深入的研究，下面介绍 3 种匿名化隐私保护模型，分别是 *k-anonymity*[9–11]、*l-diversity*[12] 和 *t-closeness*[13]。这些匿名化隐私保护模型都是以信息损失为代价，逐渐提高隐私保护效果。

1. *k-anonymity*

k-anonymity 隐私保护模型由彼得兰格拉 • 萨马拉蒂（Pierangela Samarati）和拉坦亚 • 斯威尼（Latanya Sweeney）在 1998 年正式提出，有效解决了链接攻击问题。匿名化后的数据表中具有相同准标识符的若干记录被称为一个等价类，**将 k 条记录放入一个等价类中，要求任意一条记录与其他至少 $k-1$ 条记录是不可区分的，**这样数据中的每一条记录都能找到与之相似的记录，降低了数据的识别度，这就是 *k-anonymity*。

如果由于数据太少，一条记录无法找到与之相似的 $k-1$ 条记录，那么这条记录就不应该被纳入数据集。

k-anonymity 可以防止敏感属性值的泄露。当攻击者进行链接攻击时，由于对任意一条记录的攻击，都会同时关联到等价类中的其他 $k-1$ 条记录，因此攻击者无法确定特定用户，达到了保护用户隐私的目的。k 值越大，关联的记录越多，隐私保护效果越好，但数据丢失也越严重。这里有两个简化的数据表，表 5.6 展示了公布的病患数据，表 5.7 为病患数据的 3-*anonymity* 版本，其中 { 年龄，邮政编码 } 为准标识符，疾病为敏感属性。即使攻击者掌握选民信息表，由于表中某选民的年龄和邮政编码只能与表 5.7 中的一个等价类对应，而每一个等价类中有 3 名患者，因此攻击者无法确定究竟对应的是哪一位患者。

看到这里，我们不禁产生疑问，符合 *k-anonymity* 的要求就一定能保证隐私不被泄露吗？仔细观察表 5.7，我们发现第一个等价类中的敏感属性值全部

表 5.6 公布的病患数据

年龄	邮政编码	疾病
52	123023	心脏病
32	120156	糖尿病
59	123152	心脏病
30	120162	糖尿病
56	123485	心脏病
35	120154	哮喘

表 5.7 3-*anonymity* 版病患数据

年龄	邮政编码	疾病
5*	123***	心脏病
5*	123***	心脏病
5*	123***	心脏病
3*	1201**	糖尿病
3*	1201**	糖尿病
3*	1201**	哮喘