范文仲：新技术——隐私计算

意见领袖丨范文仲（北京金融控股集团董事长）

第二章 新技术

二、区块链与隐私计算

（二）隐私计算

伴随着云计算、大数据、人工智能等新一代信息技术的快速发展，数据作为战略性和基础性资源，不但是连接虚拟空间和实体空间的纽带，也是数字经济体系中技术创新、需求挖掘、效率提升的重要动能。但大数据在互联网时代蓬勃发展的同时，也面临着安全问题的挑战，既有公民个人信息和隐私安全的隐患，也有行业和企业数据安全的隐患。加密技术使参与主体在无法对原始数据进行浏览、复制、修改的前提下，完成对数据的计算，得出有价值的计算结果，从而避免人为因素对数据流动和使用的介入，逐渐成为一种被广泛认可的解决方案。这种技术被称作隐私计算技术，又被形象地称为可用不可见技术。

新金融书系
NEW FINANCE BOOKS

Digital Economy and Financial Innovation

数字经济与金融创新

范文仲◎编著

中国金融出版社

1． 隐私计算技术三大流派

（1）附带隐私保护的明文算法流派

明文算法增强流派主要包括联邦学习、差分隐私、数据脱敏等技术路线。主要特点是，通过改变数据的使用方式，或在一定程度上降低数据的精确性，换取数据安全性和隐私性的提升。

联邦学习（FederatedLearning）。联邦学习是一种分布式机器学习技术和系统，包括两个或多个参与方，这些参与方通过安全的算法协议进行联合机器学习， 可以在各方数据不出本地的情况下联合多方数据源建模和提供模型推理与预测服务。在联邦学习框架下，各参与方只交换密文形式的中间计算结果或转化结果，不交换数据，保证各方数据不泄露。联邦学习可以通过同态加密、差分隐私、秘密分享等提高数据协作过程中的安全性。根据联邦学习各参与方拥有的数据的情况，可以将联邦学习分为两类，即横向联邦学习和纵向联邦学习（见图 2-1）。

图 2 - 1 横向联邦学习和纵向联邦学习

在横向联邦学习中，参与方在各方数据的“数量”这个维度上进行合作，解决单个参与方的训练数据不足的问题。在纵向联邦学习中，参与方在数据的“特征”和“标签”这两个维度上进行合作，解决单个参与方的数据特征过少或者没有标签的问题。纵向联邦学习需要计算参与方共同拥有的样本 ID，可以通过安全多方计算中的隐私集合求交技术实现。

差分隐私（DifferentialPrivacy）。差分隐私原理是基于统计学，在数据中加入足够的噪声，使数据与其拥有者无法完全关联，从而确保隐私性。因此，差分隐私并非通过将数据隔离在安全的地方确保隐秘， 而是将其淹没在噪声的海洋中，通过统计学的方式保障安全。从本质上讲，差分隐私

是通过一定程度地降低数据的精确性，换取了数据安全性和隐私性的提升。加入的噪声方差越大，隐私保护程度越高，计算的精确性也越低。使用差分隐私，需要在使用数据的准确性和隐私安全性之间寻求平衡。

数据遮掩（DataMasking）。数据遮掩就是按照一定的脱敏规则对敏感数据进行变形，实现对数据的保护。常见的技术手段有遮盖、泛化、替换、乱序、加扰等。同时，随着脱敏后数据信息完整性的丧失，数据的分析价值将随之降低。数据脱敏从技术上可以分为静态数据脱敏和动态数据脱敏两种。静态数据脱敏一般应用于数据外发场景，例如需要将生产数据导出发送给开发人员、测试人员、分析人员等；动态脱敏一般应用于直接连接生产数据的场景，例如运维人员在运维的工作中直接连接生产数据库进行运维，客服人员通过应用直接调取生产中的个人信息等。

（2）密码学流派

密码学流派主要基于数学与密码学原理，数据可在加密状况下进行计算，且将得到与明文计算相同结果。通过对数据和算法进行加密，使数据始终在密文状态下运算。主要技术路线是安全多方计算及相关支撑性技术。