

基于同态加密的隐私保护主成分分析方法

张金斗 陈经纬 吴文渊 冯 勇

中国科学院重庆绿色智能技术研究院 重庆 400714

中国科学院大学重庆学院 重庆 400714

(zhangjindou@cigit.ac.cn)

摘 要 在现实生活中,不同的行业之间,甚至同行业不同部门之间的数据并不互通,随着计算机算力的提升,制约模型训练效果的不是算力而是数据量。因此,想要得到更好的算法模型,仅靠某一方的数据是不够的,需要两方或者多方的参与,这就要求对各方的数据进行隐私保护。除此之外,随着收集的数据越来越详细,数据的维数也越来越大。面对高维的数据,数据降维是不可缺少的环节,而在数据降维方面,主成分分析(Principal Component Analysis, PCA)是常用的手段。当拥有数据的两方想要合作进行隐私保护的数据降维时,同态加密技术是一种解决办法。同态加密技术可以在保护数据隐私的前提下对加密数据进行计算,可以用在加密数据的 PCA 上。针对上述应用场景,利用 CKKS 同态加密方案,通过幂法迭代的 SVD 技术设计了一种两方加密数据进行 PCA 的方案,在保护两方数据隐私的前提下实现数据降维的目的;通过改进传统幂法迭代步骤,避免了代价高昂的同态密文除法运算,使得在选取较小的加密参数时,也能支持更多的幂法迭代次数,从而在缩短同态计算时间的同时提高计算精度。在公共数据集上进行测试,并与现有方案进行对比,该方案在计算耗时上缩短了约 80%,与明文计算结果的均方误差缩减到 1% 以内。

关键词: 同态加密; 隐私保护; 主成分分析; 奇异值分解; 幂法

中图分类号 TP309.7

Privacy-preserving Principal Component Analysis Based on Homomorphic Encryption

ZHANG Jindou, CHEN Jingwei, WU Wenyuan and FENG Yong

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China

Chongqing School, University of Chinese Academy of Sciences, Chongqing 400714, China

Abstract In real life, data is not interconnected between different industries, or even between different departments within the same industry. With the improvement of computer computing power, it is not computing power but data volume that restricts the effectiveness of model training. Therefore, in order to obtain a better algorithm model, relying solely on one party's data is not enough. It needs the participation of two or more parties, which requires privacy protection for all parties. In addition, as data collection becomes more detailed, the data dimension also increases. For high dimension data, dimension reduction is an indispensable step. And in terms of dimension reduction, principal component analysis (PCA) is a commonly used method. Homomorphic encryption is a solution when two parties want to collaborate on privacy protection data dimension reduction. Homomorphic encryption can compute encrypted data while protecting data privacy, and can be used to compute the PCA on encrypted data. In this paper, a two party encrypted data PCA scheme is designed using the CKKS homomorphic encryption scheme and the power method for dominant eigenvectors, achieving the goal of dimension reduction while protecting the privacy of both parties' data. By improving the traditional power method iteration steps, the expensive homomorphic ciphertext division is avoided, allowing for more iterations with small encryption parameters, thereby reducing the computing time and improving the accuracy of the computed results. Through testing on public datasets and comparing it with some existing schemes, the scheme reduces the computational time by about 80%, and reduces the mean squared error to within 1% compared to the plaintext computation results.

到稿日期: 2023-08-28 返修日期: 2024-05-31

基金项目: 科技部重点研发计划(2020YFA0712300); 重庆市自然科学基金(cstc2021jcyj-msxmX0821, CSTB2023NSCQ-MSX0441, cstc2021yszx-jcyjX0004, 2022YSZX-JCX0011CSTB, CSTB2023YSZX-JCX0008); 中国科学院西部青年学者项目

This work was supported by the Ministry of Science and Technology Key Research and Development Programs(2020YFA0712300), Natural Science Foundation of Chongqing(cstc2021jcyj-msxmX0821, CSTB2023NSCQ-MSX0441, cstc2021yszx-jcyjX0004, 2022YSZX-JCX0011CSTB, CSTB2023YSZX-JCX0008) and Western Light of the West Young Scholars program of the Chinese Academy of Sciences.

通信作者: 陈经纬(chenjingwei@cigit.ac.cn)

Keywords Homomorphic encryption, Privacy preserving, Principal component analysis, Singular value decomposition, Power method

1 引言

1.1 研究背景

随着大数据、云计算和人工智能技术的迅速发展,出现了许多新型应用和服务模式,例如网约车平台、外卖平台和医疗咨询平台。这些服务利用机器学习技术提供个性化推荐、搜索和医疗诊断等服务,为人们的生活带来了便利。然而,这些服务需要大量数据来训练和预测,而一些数据涉及个人隐私,如个人喜好、身份信息、地理位置和健康数据。如果服务方直接分析和处理收集到的信息,而没有采取任何保护措施,则很容易导致用户数据泄露,严重威胁用户隐私安全。

此外,每年都会发生重大的隐私泄露事件,例如,2023 年 3 月 20 日,Vans 公司由于系统的安全漏洞,用户面临欺诈和隐私信息被盗窃的风险^[1];2 月 21 日,一家俄罗斯顶级托管服务提供商超过 5400 万的用户资料被暴露,包括敏感数据如电子邮件地址和电话号码^[2];2 月 12 日,Microsoft 披露了 Exchange 服务器中的一个关键零日漏洞,该漏洞导致 Azure 平台的数据泄露,数百个高管账户被窃取^[3];2 月 15 日,在 LockBit 勒索软件组织对银行服务提供商的网络攻击中,大约 57000 名美国银行客户的个人信息被暴露^[4]。这些事件表明,科技公司收集用户信息而未进行有效的保护,对用户隐私产生了巨大威胁。

另一方面,隐私权作为一项基本人权,对个人和企业来说都极其重要。欧盟于 2018 年实施的《通用数据保护条例》^[5]和美国最严隐私法案《加利福尼亚消费者隐私法案》^[6]都要求企业加强用户隐私和数据安全保护,违反规定的企业将受到严厉惩罚。我国在 2017 年实施的《中华人民共和国网络安全法》^[7]也指出,任何个人和组织不得窃取或以其他非法方式获取个人信息,未经被收集者同意,不得向他人提供个人信息。这些法规的建立对人工智能传统的数据处理模式提出了新的挑战。

在云计算环境中,传统的加密方法以加密形式存储敏感数据。但是,要对加密的数据执行计算,服务方要么需要在云端解密数据,这可能会导致安全问题,要么下载数据、解密数据并执行计算,这样代价很大也不安全;且出于行业隐私,不同的行业之间,甚至同行业不同部门之间存在数据壁垒,导致无法安全共享数据,而仅用各部门的私有数据进行机器学习模型的训练,难以实现最优的性能。因此,如何在保护用户隐私的前提下对隐私数据在云端进行计算是十分值得研究的内容。此外,随着量子计算机理论快速发展,不久的将来必然迎来量子计算机应用的大潮,量子计算机能够解决基于整数分解、离散对数问题的常用密码学方案,会对目前的隐私安全产生很大的影响。

在云计算环境中,传统密码加密数据存储,但存在不支持计算、不能抵御量子计算攻击的缺陷。因此,在实际的计算需求与对隐私问题的担忧的影响下,隐私计算技术发展了起来。

隐私计算(Privacy-preserving Computation)是为了让多个数据拥有者在不暴露数据本身的前提下实现数据的共享、互通、计算、建模,最终产生超出自身数据的价值,同时保证数据不泄露给其他参与方^[8]。目前隐私计算的技术大致可分为联邦学习^[9]、可信执行环境^[10]和安全多方计算^[11]。由于量子计算技术的兴起,现有的隐私保护技术很可能会受到较大冲击,因此如何设计能抵抗量子计算的安全协议是十分重要的。同态加密(Homomorphic Encryption)技术是基于格上的困难问题设计的,具有抵抗量子计算攻击的特性,能够在密文上进行计算后将得到的结果解密,与直接对明文进行计算得到的结果一致,且需要的通信开销较小,具有广阔的应用前景。

1.2 相关工作

目前,隐私计算主要通过机器学习方法来实现模型的训练和预测,进而完成计算任务。基于同态加密的机器学习在密文预测方面已有许多工作。Durga 等^[12]提出了一种适用于离散和连续数据的朴素贝叶斯分类器;微软研究院的 Gilad-Bachrach 等基于部分同态加密技术,提出了一种近似神经网络模型 CryptoNets^[13];Hesamifard 等提出了一种对密文数据进行分类的深度神经网络 CryptoDL^[14];Sanyal^[15]提出了一种并行化和加速加密数据计算的算法技巧,用于实现神经网络同态的预测;Bourse 等^[16]利用同态加密方案,通过深度离散神经网络在 1.7 s 内对 MNIST 数据集中的加密图像进行分类,准确率超过 96%。

如果只进行密文的预测,对公有的数据集效果可能很好,但由于模型参数不易修改,难以得到适用于每个用户的模型,对私有数据的预测结果无法估计,因此在密文上进行训练是十分重要的。使用全同态加密只能训练非常小的模型,而模型的好坏非常依赖数据量,因此如果要利用大规模的数据进行密文的训练,则必须要对数据进行降维处理。在数据降维方面,主成分分析(PCA)是常用的手段。

在基于同态加密的隐私保护主成分分析方法方面,目前相关的研究还较少,且主要是通过幂法迭代的奇异值分解(Singular Value Decomposition, SVD)技术来实现加密数据的 PCA 方法。Lu 等^[17]、Rathee 等^[18]均利用 BGV^[19]方案在较小的数据集上完成了第一个主成分的提取,但 BGV 仅支持整数运算,而幂法本身是数值算法,数据往往也是浮点的数据,在应用时没有支持浮点运算的 CKKS^[20]方案高效;Panda^[21]利用 CKKS 方案,同样采取幂法进行了多个主成分的提取,但方案中幂法迭代引入的同态除法操作会导致计算效率降低。此外,上述方案都是在用户-服务器模式下实现的,如图 1 所示。而现实生活中,不同的行业之间,甚至同行业不同部门之间的数据并不互通,且随着计算机算力的提升,用户依靠自身的算力即可完成主成分的求解,制约模型训练效果的不再是算力而是数据量。想要得到更好的算法模型,仅靠某一方的数据是不够的,因此利用两个或多个用户的数据进行主成分分析更能满足当下的需求。

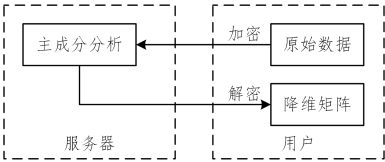


图 1 用户-服务器模式
Fig. 1 Client-Server mode

1.3 本文的贡献

1)利用 CKKS 同态加密方案,通过幂法迭代的 SVD 技术实现了两方加密数据的 PCA 方法,在保护两方数据隐私的前提下实现数据降维的目的,如图 2 所示。

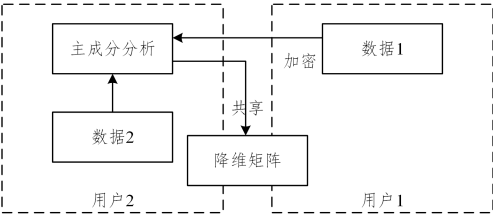


图 2 双用户共用数据模式
Fig. 2 Client-Client mode

2)避免了幂法迭代中的同态除法过程。传统的幂法迭代过程需要对向量做标准化处理,即要对加密的向量做同态的除法运算,而目前的同态加密方案做除法的代价很大。利用密文-明文的除法(实际运算时转化为乘法)代替密文-密文的除法,从而避免了同态的除法运算,实现向量的近似标准化。这项改进既保证了精度,也大大减少了同态计算的时间消耗。

3)增加了每次交互时幂法迭代的次数。在不解密密文的前提下,进行同态乘法的次数受限于 CKKS 方案的安全参数。本研究在做近似标准化的密文-明文除法时,通过选取合适的明文数值,将可迭代次数翻倍,使得精度得到进一步提高。

相比现有方案,本文方案的优势在于:1)利用两方数据进行数据降维,提高了数据质量;2)使用支持浮点运算的加密方案的同时,避免了同态的除法操作,使得计算效率与计算精度均得到较大提升。

2 背景知识

2.1 主成分分析方法

主成分分析(PCA)是一种常用的数据降维和特征提取方法。其目标是将高维数据转换为低维空间,并找到能够最大程度保留原始数据方差的主要特征。PCA 的主要思想是通过线性变换将原始特征空间转换为新的特征空间,其中新的特征是原始特征的线性组合。这些新特征被称为主成分,其选择是基于数据集中的方差最大化的原则。在传统的 PCA 中,首先计算数据集的协方差矩阵,然后对协方差矩阵进行特征值分解。特征值分解将协方差矩阵分解为特征值和对应的特征向量。这些特征向量就是数据集的主成分,而特征值表示数据集中的方差。

奇异值分解(SVD)是一种更一般的矩阵分解方法,可以应用于任意矩阵,而不仅仅是协方差矩阵。对于一个矩阵 A , SVD 将其分解为 3 个矩阵的乘积,即 $A=U\Sigma V^T$,其中 U 和 V

是正交矩阵, Σ 是一个对角矩阵,对角线上的元素称为奇异值。在 SVD 中, U 的列向量是 A 的左奇异向量, V 的列向量是 A 的右奇异向量, Σ 的对角线上的元素是 A 的奇异值。PCA 可以通过 SVD 来实现。对于一个数据矩阵 X ,其中每一行表示一个样本,每一列表示一个特征,PCA 的步骤可以按照以下方式与 SVD 联系起来。

- 1)对数据矩阵 X 进行中心化,使得每个特征的均值为 0。
- 2)对中心化后的数据矩阵 X 进行 SVD 分解: $X=U \cdot \Sigma \cdot V^T$ 。
- 3)选取 U 的前 k 列作为主成分,构成新的特征矩阵,其中 k 是降维后的维度。
- 4)计算投影矩阵 $Y=X \cdot W$,其中 W 为特征矩阵 U 的前 k 列。

这样,通过 SVD,可以得到数据矩阵的主成分和降维后的表示。在这种情况下,主成分就是数据矩阵 X 的左奇异向量,奇异值对应于 PCA 中的特征值的平方。

幂法(Power Method)是一种迭代算法,用于计算矩阵的特征值和对应的特征向量中的最大特征值和特征向量。幂法的基本思想是通过反复迭代,逐步逼近矩阵的最大特征值和对应的特征向量。算法的核心是利用特征向量与矩阵的乘积的特性,不断对一个初始向量进行矩阵乘法和标准化操作,使得向量逐渐收敛到矩阵的最大特征值所对应的特征向量。利用这个性质,可以通过幂法迭代来近似计算协方差矩阵的最大特征值和对应的特征向量,进而实现 PCA。

利用幂法实现 PCA 的伪代码如算法 1 所示。

算法 1 基于幂法的 PCA

输入:数据矩阵 A ,迭代次数 N

输出:降维后的矩阵 C

```
1. for i from 1 to k do //降低到 k 维
2.   计算协方差矩阵  $X:=A^T A$ 
3.   初始化一个随机向量  $u$ 
4.   for j from 1 to N do
5.      $t:=u$ 
6.      $u:=X \cdot u$ 
7.      $v:=\frac{u}{\|u\|_2}$ 
8.   end for
9.   将每轮计算得到的  $v$  存储到  $V$  中
10.  $\lambda:=\frac{\|u\|_2}{\|t\|_2}$ 
11.  $A:=A-\lambda \cdot v \cdot v^T$ 
12. end for
13. return  $C:=A \cdot V$ 
```

2.2 同态加密技术

2.2.1 同态加密技术的发展历程

1978 年,Rivest 和 Adleman(“RSA”加密算法中的“R”和“A”)以及 Dertouzos 提出了全同态加密的构想^[22],自此成为了密码学研究领域的一个公开难题。目前,同态加密算法主要分为部分同态加密和全同态加密两大类。满足有限运算同态性而不满足任意运算同态性的加密算法被称为部分同态加密。典型的部分同态加密特性包括乘法同态、加法同态、有限次数全同态等。部分同态加密主要包括以 RSA 算法^[23] 和

ElGamal 算法^[24]为代表的乘法同态加密、以 Paillier^[25]算法为代表的加法同态加密和支持任意次加法同态和一次乘法同态运算的 Boneh-Goh-Nissim 方案^[26]。

全同态加密算法主要包括以 Gentry 方案^[27]为代表的第一代方案、以 BGV 方案^[19]和 BFV 方案^[28]为代表的第二代方案、以 GSW 方案^[29]为代表的第三代方案以及支持浮点数近似计算的 CKKS 方案^[20]等。

其中 CKKS 同态加密方案是一种近似计算同态加密方案,它支持实数和复数上的近似算术,并具有预定义精度,其安全性依赖于 RLWE 问题^[30]。CKKS 算法的目标是做近似计算,因此不追求解密结果和明文完全一致,该技术的核心思想是将同态加密方案中的噪声项看作近似计算过程中误差的一部分。允许误差存在使得 CKKS 相于其他基于 LWE/RLWE 问题的同态方案,在细节上有了较大的简化,计算效率也有了很大提升,被广泛应用于机器学习相关的隐私保护。

2.2.2 CKKS 同态加密方案

CKKS 同态加密方案的明文空间是复向量空间 $\mathbb{C}^{N/2}$,复向量明文空间使得 CKKS 方案支持复数域上的运算。

对于一个 2 的整数幂 M ,设 $\phi_M(x)$ 是第 M 个次数为 $N = \phi_M$ 的分圆多项式, $\phi_M(x) = (x - \zeta) \cdots (x - \zeta') \cdots (x - \zeta^{M-1})$,其中 $\zeta = e^{\frac{2\pi i}{M}}$,记环 $R = \mathbb{Z}[X]/\phi_M(x)$,密钥分布为 χ_{key} ,噪声分布为 χ_{error} ;加密时所需分布为 χ_{enc} ,选取整数 M 和 P ,模数为 q_L 。CKKS 方案包含如下算法。

1) 密钥生成 $KeyGen(1^\lambda)$: 采样 $s \leftarrow \chi_{\text{key}}, a \leftarrow R_{q_L}, e \leftarrow \chi_{\text{error}}$,生成私钥 $sk = (1, s)$,公钥 $pk = (b, a)$,其中 $b = -as + e \pmod{q_L}$ 。采样 $a' \leftarrow R_{q_l}^2, e' \leftarrow \chi_{\text{error}}$ 。令 $evk = (b', a')$,其中 $b' = -a's + e' + Ps^2$ 。

2) 加密 $Enc_{pk}(m)$: 对于明文多项式 $m \in R$,采样 $v \leftarrow \chi_{\text{enc}}, e_0, e_1 \leftarrow \chi_{\text{error}}$,输出一个密文 $c = v \cdot pk + (m + e_0, e_1)$ 。

3) 解密 $Dec_{sk}(c)$: 输入一个处于 l 层的密文 c ,计算 $m' = m + e = \langle c, sk \rangle \pmod{q_l}$ 。

4) 加法 $Add(c_1, c_2)$: 输入密文 $c_1, c_2 \in R_{q_l}^2$,输出 $c_{\text{add}} = c_1 + c_2 \pmod{q_l}$ 。

5) 乘法 $Mult(c_1, c_2)$: 对于密文 $c_1, c_2 \in R_{q_l}^2$,令 $(d_0, d_1, d_2) = (b_1 b_2, a_1 b_2 + a_2 b_1, a_1 a_2) \pmod{q_l}$,输出 $c_{\text{mult}} = (d_0, d_1) + \lfloor P^{-1} \cdot d_2 \cdot evk \rfloor \pmod{q_l}$ 。

6) 重缩放 $R_{s_l \rightarrow r}(c)$: 输入密文 $c \in R_{q_l}^2$,输出 $c' = \lfloor q_r / q_l \cdot c \rfloor \pmod{q_r}$ 。

(7) 旋转 $Rot_{gk}(c; r)$: 输入密文 $c \in R_{q_l}^2$ 和旋转密钥 gk ,输出 c 的明文向量旋转 r 个位置后对应的密文 c' 。

2.3 密文打包技术

由于 CKKS 方案中有 $N/2$ 个位置可以存放数值,因此可以支持 SIMD(Single Instruction Multiple Data)操作。对于向量的点积、求和等运算, SIMD 可以直接对向量的所有元素进行运算,不需要单独对每个分量进行逐一计算。

密文打包技术允许将多个加密的数值放入一个密文里,并用 SIMD 的方式进行计算。通过选择适当的编码、解码方法,通过映射 $\mathbb{C}^n \rightarrow \mathbb{R}$ 可以将明文空间 M 中的 n 维向量打包成

单个明文多项式,每个分量对应一个明文槽,这样在密文计算时可以对每个分量同时执行算术运算,使计算并行化,优化了时间复杂度和空间复杂度。

2.4 对角线编码技术

一般的编码方案在计算密文的矩阵-向量乘法时较为复杂, Halevi 等提出了一种对角线编码的技术^[31],可以简化矩阵-向量乘法的过程。

对于一个矩阵 $U \in R^{n \times n}$ 和一个向量 $m \in R^n$,定义 U 的第 i 条对角线组成的向量为 $u_i = (U_{0,i}, U_{1,i+1}, \dots, U_{n-i-1,n-1}, U_{n-i,0}, \dots, U_{n-1,i-1})$,那么可以将矩阵向量乘法表示为旋转操作与标量乘法的组合: $U \cdot m = \sum_{i=0}^{n-1} (u_i Rot(m; i))$,表示向量之间每个分量的乘法。算法 2 列出了利用对角线编码技术对加密向量 c 进行矩阵-向量乘法的过程。

算法 2 矩阵-向量乘法 LinTrans(ct; U)

输入: 矩阵 U , 密文向量 c

输出: 矩阵-向量乘法后的密文 c'

- 1. $c' \leftarrow Mult(c, u_0)$
- 2. for i from 1 to $n-1$ do
- 3. $c' \leftarrow Add(c', Mult(Rot(c; i); u_i))$
- 4. end for
- 5. return c'

如算法 2 所示,矩阵-向量乘法主要包含加法、常数乘法及旋转操作,由于旋转需要进行密钥交换操作,与其他的运算过程相比代价高很多,因此可以认为算法 2 的时间复杂度约为 $O(n)$ 次旋转操作。通过小步大步法的思想,做恒等变形得到:

$$\begin{aligned} U \cdot m &= \sum_{i=0}^{n-1} (u_i Rot(m; i)) \\ &= \sum_{i=0}^{l-1} \sum_{j=0}^{k-1} (u_{ki+j} Rot(m; ki+j)) \\ &= \sum_{i=0}^{l-1} (Rot(\sum_{j=0}^{k-1} (Rot(u_{ki+j}; -ki) Rot(m; j)); ki)) \end{aligned}$$

将原来的 n 次旋转操作转化为 $l+k$ 次旋转操作,其中 $n = l \cdot k$,当 $l=k$ 时,时间复杂度变为 $O(2\sqrt{n})$,当 n 较大时,有效提高了算法的计算效率。

2.5 半诚实敌手模型

半诚实敌手(Semi-honest Adversaries)会遵守协议的规则,但可能会尝试从协议的消息中推断其他方的隐私信息。这是一种比较弱的敌手模型,如果一个协议对半诚实的敌手是安全的,那么它不允许参与方从协议中学习到任何额外的信息。

本文设计的协议主要针对半诚实敌手模型进行安全防范,参与方为用户 1 和用户 2 两方。在半诚实模型下,两方都会诚实地执行协议,不会进行伪造、篡改等主动攻击,但会尝试通过诚实执行协议所获得的消息进一步推断出其他信息。

假定用户 1 和用户 2 要完成的计算任务 $f = (f_1, f_2)$ 是一个概率多项式时间函数。 π 是用于计算 f 的两方协议,其安全参数为 λ 。定义用户 i 在执行协议 $\pi(x, y, \lambda)$ 时的视角为 $view_i^\pi(x, y, \lambda)$,即协议 π 的参与方 i 能接触到的所有信息。用户 i 在执行协议 $\pi(x, y, \lambda)$ 后,从自己视角得到的输出为 $output_i^\pi(x, y, \lambda)$ 。两方的输出:

$$output^{\pi}(x, y, \lambda) = (output_1^{\pi}(x, y, \lambda), output_2^{\pi}(x, y, \lambda))$$

由此可以得到半诚实敌手下安全性的定义,如定义 1 所示^[32]。

定义 1 令函数 $f = (f_1, f_2)$, 如果存在概率多项式时间的算法 S_1 和 S_2 使得:

$$\{(S_1(1^{\lambda}, x, f_1(x, y)), f(x, y))\}_{x, y, \lambda} \equiv \{(view_1^{\pi}(x, y, \lambda), output^{\pi}(x, y, \lambda))\}_{x, y, \lambda} \quad (1)$$

$$\{(S_2(1^{\lambda}, x, f_2(x, y)), f(x, y))\}_{x, y, \lambda} \equiv \{(view_2^{\pi}(x, y, \lambda), output^{\pi}(x, y, \lambda))\}_{x, y, \lambda} \quad (2)$$

则称协议 π 在半诚实敌手模型下安全计算了 f 。其中, \equiv 表示左右两个式子是计算上不可区分的, x 表示用户 1 的输入, y 表示用户 2 的输入。

定义 1 中等式左边对应的是理想世界和模拟器, 右边对应的是现实世界。在理想世界中, 假设存在可信第三方 T , 由 T 计算 $f(x, y)$ 后将结果发送给用户 1 和用户 2。为了证明协议 π 的安全性, 需要在理想世界中为用户 1 和用户 2 分别建立一个与 π 等效的模拟器, 即 S_1 和 S_2 。如果 S_1 和 S_2 的输出 $f(x, y)$ 与协议 π 的输出 $output^{\pi}(x, y, \lambda)$ 是计算不可区分的, 则可以说明协议 π 是安全的。

本文中, f 为确定性函数, 当 $f_1(x, y) \equiv output_1^{\pi}(x, y)$, $f_2(x, y) \equiv output_2^{\pi}(x, y)$ 时, 有 $f(x, y) \equiv output^{\pi}(x, y, \lambda)$ 。为了方便后续证明, 将式(1)、式(2)分别简记为 $S_1(x; f_1(x, y)) \equiv View_1(x; y; out_1)$, $S_2(x; f_2(x, y)) \equiv View_2(x; y; out_2)$ 。

3 隐私保护的主成分分析方法

方案考虑有用户 1 和用户 2, 每个用户各拥有一部分数据, 同时用户 2 也可以提供计算服务。在用户 1 和用户 2 均不知道对方所持有数据的前提下, 用户 1 和用户 2 想利用全部数据来进行数据的降维, 使得降维的效果与直接对用户 1 和用户 2 合并后的数据进行训练的效果一致。方案的流程如图 3 所示, 用户 1 首先利用公钥将自身的数据进行加密, 之后发送给用户 2。用户 2 通过双方的数据计算出协方差矩阵, 之后对协方差矩阵进行幂法迭代。迭代完成后, 将得到的特征向量发送给用户 1 进行解密。之后用户 1 将特征向量解密并做标准化操作, 随后发送给用户 2。双方利用标准化后的特征向量更新自身的数据, 之后重复整个过程, 直至完成所有特征向量的计算。将所有特征向量组合为一个矩阵, 即为主成分分析方法中的降维矩阵。至此, 双方利用全部的数据完成了主成分分析的过程。

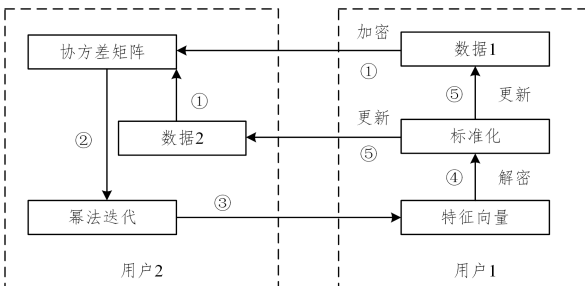


图 3 隐私保护的两方 PCA 流程

Fig. 3 Two-party PCA process for privacy protection

由 2.1 节中的幂法迭代过程可以看到, 如果要在密文上

进行主成分分析, 绝大部分运算都会集中在算法 1 的第 6 和第 7 步中, 即矩阵-向量乘法及向量的标准化操作。因此, 如何在密文上设计高效的算法来实现这两个操作是极为重要的。

3.1 幂法迭代中的矩阵-向量乘法

利用 3.2 节中提到的对角线编码技术, 可以很方便地在加密的数据上实现矩阵-向量乘法。具体而言, 可以将协方差矩阵 \mathbf{X} 的每一条对角线进行编码。定义 \mathbf{X} 第 i 条对角线组成的向量 $\mathbf{x}_i = (\mathbf{X}_{0,i}, \mathbf{X}_{1,i+1}, \dots, \mathbf{X}_{n-i-1,n-1}, \mathbf{X}_{n-i,0}, \dots, \mathbf{X}_{n-1,i-1})$, 则 $\mathbf{X} \cdot \mathbf{v} = \sum_{i=0}^{l-1} (\text{Rot}(\sum_{j=0}^{k-1} (\text{Rot}(\mathbf{x}_{i+j}; -ki) \text{Rot}(\mathbf{v}; j)); ki))$ 。

3.2 近似标准化操作

由于同态加密技术难以处理同态的密文除法及开根号等非线性运算, 因此如何实现幂法迭代过程中加密向量的标准化操作是方案最重要的问题。在实际的幂法迭代过程中, 每一次迭代都在进行乘法运算, 如果不对数值加以限制, 那么经多次迭代后, 存在数据溢出的可能性, 因此标准化操作的主要意图就是限制数据的膨胀。

考虑两方拥有的数据分别为 $\mathbf{A} \in R^{m_1 \times n}$, $\mathbf{B} \in R^{m_2 \times n}$, 其中 n 为数据的特征维数, m_1 和 m_2 为两方各自所用的样本数。在实际的密文计算过程中, 不必在每次迭代时进行严格的向量标准化操作, 而是选取与 $\|\mathbf{u}\|$ 相同量级的数值 P , 使得 $\frac{\mathbf{u}}{P}$ 也可以起到限制数据膨胀的作用。

引理 1 令矩阵 $\mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}$, $P_A = \|\mathbf{A}^T \mathbf{A}\|_{\infty}$, $P_B = \|\mathbf{B}^T \mathbf{B}\|_{\infty}$, $\mathbf{X} = \mathbf{C}^T \mathbf{C} \in R^{n \times n}$, $P = P_A + P_B$, 则对任意 $\mathbf{u} \in R^{n \times n}$, 有 $\frac{\|\mathbf{X}\mathbf{u}\|_{\infty}}{P} \leq \|\mathbf{u}\|_{\infty}$ 。

证明:

$$\begin{aligned} \|\mathbf{X}\mathbf{u}\|_{\infty} &\leq \|\mathbf{X}\|_{\infty} \|\mathbf{u}\|_{\infty} \\ &\leq \|\mathbf{A}^T \mathbf{A} + \mathbf{B}^T \mathbf{B}\|_{\infty} \|\mathbf{u}\|_{\infty} \\ &\leq (\|\mathbf{A}^T \mathbf{A}\|_{\infty} + \|\mathbf{B}^T \mathbf{B}\|_{\infty}) \|\mathbf{u}\|_{\infty} \\ &= P \|\mathbf{u}\|_{\infty} \end{aligned}$$

即 $\frac{\|\mathbf{X}\mathbf{u}\|_{\infty}}{P} \leq \|\mathbf{u}\|_{\infty}$, 证明完毕。

幂法迭代时, 若用明文 P 替代密文的模来实现近似标准化操作, 不仅可以解决迭代时的数据膨胀问题, 还能够避免同态的除法操作, 这会大大提升在密文上幂法迭代的效率。

3.3 协方差矩阵的计算

根据 2.1 节中的幂法, 首先要计算出整体数据的协方差矩阵 \mathbf{X} , 与此同时也要保护两方数据的隐私, 即 \mathbf{X} 是加密的。

记 \mathbf{A} 的均值向量为行向量 \mathbf{u}_A^T , \mathbf{B} 的均值向量为行向量 \mathbf{u}_B^T , 则全部数据的均值向量为:

$$\mathbf{u}^T = (m_1 \mathbf{u}_A^T + m_2 \mathbf{u}_B^T) / (m_1 + m_2)$$

那么协方差矩阵为:

$$\begin{aligned} \mathbf{X} &= (\mathbf{A}^T - \mathbf{u} \cdot \mathbf{1}^T, \mathbf{B}^T - \mathbf{u} \cdot \mathbf{1}^T) \cdot \begin{pmatrix} \mathbf{A} - \mathbf{1} \cdot \mathbf{u}^T \\ \mathbf{B} - \mathbf{1} \cdot \mathbf{u}^T \end{pmatrix} \\ &= \mathbf{A}^T \mathbf{A} + \mathbf{B}^T \mathbf{B} - (m_1^2 \mathbf{u}_A \mathbf{u}_A^T + m_2^2 \mathbf{u}_B \mathbf{u}_B^T + m_1 m_2 (\mathbf{u}_A \mathbf{u}_B^T + \mathbf{u}_B \mathbf{u}_A^T)) / (m_1 + m_2) \end{aligned} \quad (3)$$

假定拥有数据 \mathbf{B} 的一方为计算方,即拥有数据 \mathbf{A} 的用户一方无法获取计算方有关 \mathbf{B} 的信息,同样计算方也无法获取 \mathbf{A} 的相关信息。在此条件下,式(3)中 $\mathbf{A}^T \mathbf{A}, \mathbf{B}^T \mathbf{B}, m_1^2 \mathbf{u}_A \mathbf{u}_A^T, m_2^2 \mathbf{u}_B \mathbf{u}_B^T$ 这 4 部分可在明文下完成计算,之后进行加密,并由计算方进行处理。

为了利用 4.1 节中矩阵-向量乘法的对角线编码方式,需要将 \mathbf{X} 也按照对角线进行编码。因此,对于 $\mathbf{u}_A \mathbf{u}_B^T$,有:

$$\mathbf{u}_A \mathbf{u}_B^T = \begin{bmatrix} u_{A_1} u_{B_1} & u_{A_2} u_{B_1} & u_{A_3} u_{B_1} & \cdots & u_{A_n} u_{B_1} \\ u_{B_1} u_{B_2} & u_{A_2} u_{B_2} & u_{A_3} u_{B_2} & \cdots & u_{A_n} u_{B_2} \\ u_{A_1} u_{B_3} & u_{A_2} u_{B_3} & u_{A_3} u_{B_3} & \cdots & u_{A_n} u_{B_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{A_1} u_{B_n} & u_{A_2} u_{B_n} & u_{A_3} u_{B_n} & \cdots & u_{A_n} u_{B_n} \end{bmatrix}$$

不难得到 $\mathbf{u}_A \mathbf{u}_B^T$ 的第 i 条对角线为: $(u_{A_1} u_{B_i}, \cdots, u_{A_n} u_{B_{n-i+1}}, u_{A_1} u_{B_{n-i+2}}, \cdots, u_{A_{i-1}} u_{B_n}) = Rot(\mathbf{u}_A; i) \odot \mathbf{u}_{B_i}$ 。通过明文的旋转操作,可以将对角线编码运用到矩阵 $\mathbf{u}_A \mathbf{u}_B^T$ 中。同样地,可以得到 $\mathbf{u}_B \mathbf{u}_A^T = Rot(\mathbf{u}_B; i) \odot \mathbf{u}_{A_i}$ 。至此,可以在保护两方数据隐私的情况下得出整体数据的协方差矩阵,如协议 1 所示。

协议 1 计算加密的协方差矩阵 \mathbf{c}_X

输入:数据矩阵 \mathbf{A}, \mathbf{B} , 两方的样本数 m_1, m_2

输出:加密的协方差矩阵 \mathbf{c}_X

用户 1

- 计算 $\mathbf{u}_A \mathbf{u}_A^T, -m_1^2 \mathbf{u}_A \mathbf{u}_A^T, \mathbf{A}^T \mathbf{A}, -Rot(\mathbf{u}_A; 1 \rightarrow n)$, 将这 4 部分按对角线编码方式分别进行编码,将每部分对应的 n 条明文记为 $p_{A_1}, p_{A_2}, p_{A_3}, p_{A_4}$ 。
 - 加密 $p_{A_1}, p_{A_2}, p_{A_3}, p_{A_4}$, 得到加密后的 4 部分各 n 条密文 $c_{A_1}, c_{A_2}, c_{A_3}, c_{A_4}$ 。
 - 将加密后的 $c_{A_1}, c_{A_2}, c_{A_3}, c_{A_4}$ 及用户 1 的样本数 m_1 发送给计算方。
- 用户 2(计算方)

- 计算 $-\frac{m_2^2 \mathbf{u}_B \mathbf{u}_B^T}{m_1 + m_2}, \mathbf{B}^T \mathbf{B}, -Rot(\mathbf{u}_B; 1 \rightarrow n)$, 将这 3 部分按对角线编码方式分别进行编码,每一部分的 n 条明文记为 $p_{B_2}, p_{B_3}, p_{B_4}$ 。
- 通过明-密文乘法得到 $c_{A_2} = c_{A_2} / (m_1 + m_2), c_{A_4} = m_1 m_2 c_{A_4} / (m_1 + m_2)$
- $c_1 = m_1 m_2 c_{A_1} u_{B_1 \rightarrow n} / (m_1 + m_2), c_2 = m_1 m_2 p_{B_4} c_{A_1} / (m_1 + m_2)$
- 加密 p_{B_2}, p_{B_3} , 得到 c_{B_2}, c_{B_3}
- 协方差矩阵共有 6 部分密文组成,每一部分均有按对角线编码方式得到的 n 条密文, $\mathbf{c}_X = c_{A_3} + c_{B_3} + c_{A_2} + c_{B_2} + c_1 + c_2$

3.4 基于同态加密的两方 PCA 协议

基于前文的算法,可以构建一个基于同态加密的两方主成分分析方法协议,使得可以在保护两方数据隐私的前提下,实现对两方数据的联合降维,最终得到由特征向量组成的降维变换矩阵 \mathbf{V}_k 。如协议 2 所示。

协议 2 基于同态加密的两方 PCA 协议

输入:密文协方差矩阵 \mathbf{c}_X ,降低到的维数 k

输出:降维变换矩阵 \mathbf{V}_k

- for i from 1 to k do
- 明文计算 P_{A_i} , 发送给用户 2
- 明文上计算 $P := P_{A_i} + P_{B_i}$
- 初始化一个随机向量 \mathbf{u}

- for j from 1 to N do
- $\mathbf{c}_u := \text{LinTrans}(c_u; c_X)$
- $\mathbf{c}_v := \frac{1}{P} \times c_u // \text{明-密文乘法}$
- end for
- 将 \mathbf{c}_v 发回给用户 1
- 用户 1
- 将 \mathbf{c}_v 解密得到明文 \mathbf{v} , 计算 $\mathbf{v} := \frac{\mathbf{v}}{\|\mathbf{v}\|_2}$
- 将每一轮迭代得到的 \mathbf{v} 保存到 \mathbf{V}_k 中
- 更新矩阵 $\mathbf{A} := \mathbf{A} - \mathbf{A} \cdot \mathbf{v} \cdot \mathbf{v}^T$
- 将 \mathbf{v} 发送给用户 2
- 用户 2(计算方)
- 更新矩阵 $\mathbf{B} := \mathbf{B} - \mathbf{B} \cdot \mathbf{v} \cdot \mathbf{v}^T$
- end for
- return \mathbf{V}_k

3.5 同态 PCA 协议分析

3.5.1 协议复杂度分析

协议可以在 k 次交互后完成 k 个特征向量的计算,即通过 k 次交互实现两方原始数据维度降低到 k 维的操作。若原始数据为 n 维,则每次交互两方发送的密文条数、密文-明文乘法 CP_{mult} 、密文-密文乘法 CC_{mult} 以及密文旋转 C_{rotate} 的条数如表 1 所列,其中 N 为迭代次数。

在实验中发现,由于 CKKS 方案的限制,进行 $\frac{1}{P} \times c_u$ 的明-密文乘法时,编码过程需要将明文 $\frac{1}{P}$ 及密文 \mathbf{u} 乘以一个相同的缩放因子 Δ (一般为 2^{10}),这会导致在进行明文-密文乘法操作时,也需要重缩放操作,并消耗一层可用于乘法计算的层数,造成了浪费。由于 P 本身是一个估计值,对 P 进行向上取整的操作,使得 $\lfloor P \rfloor \geq P \geq \|\mathbf{C}\|_\infty$ 。这样可以使 $\left\lfloor \frac{1}{P} \right\rfloor$ 的缩放因子 Δ 取成 1,这样在乘法结束后进行重缩放操作时,得到的结果即为需要的结果。该方式充分利用了 CKKS 方案可进行同态乘法运算的层数,增加了单次交互内幂法迭代的次数,从而提升了结果的精度。

表 1 单次交互的开销
Table 1 Cost of a single interaction

参与方	发送	CP_{mult}	CC_{mult}	C_{rotate}
1	$4n$	0	0	0
2	2	$(N+4)n$	Nn	$2N\sqrt{n}$

3.5.2 协议安全性分析

协议中,用户 1 和用户 2 进行联合 PCA 的目的都是将两方数据进行合作降维。在半诚实模型下,协议 1 是安全的,其安全性证明如下:

- 在执行协议 1 时,用户 1 与用户 2 的视角为:
 $View_1 = (pk, sk, \mathbf{A}, m_1, \mathbf{u}_A; \emptyset; c_{A_1-A_4})$
 $View_2 = (pk, \mathbf{B}, m_2, \mathbf{u}_B; m_1, c_{A_1-A_4}; \mathbf{c}_X)$
- 用户 1 的视角中,并没有获取到任何有关用户 2 的信息,因此容易验证 $View_1 \equiv S_1$ 。
- 向模拟器 S_2 输入 $(pk, \mathbf{B}, m_2, \mathbf{u}_B, m_1, c_{A_1-A_4})$, 使用 pk 加密 n 个 n 维随机向量,组合在一起得到一个加密的矩阵

$\mathbf{c}_X', S_2 = (pk, sk, \mathbf{A}, m_1, \mathbf{u}_A; m_1, c_{A_1 - A_i}; \mathbf{c}_X')$ 。由 CKKS 方案及 RLWE 问题的困难型可知, \mathbf{c}_X 与 \mathbf{c}_X' 在计算上是不可区分的, 因此 $View_2 \equiv S_2$ 。

综上所述, 协议 1 在半诚实模型下是安全的。
同样地, 在半诚实模型下, 协议 2 也是安全的, 其安全性证明如下。

1) 在执行协议 2 时, 用户 1 与用户 2 的视角为:
 $View_1 = (pk, sk, \mathbf{A}, m_1, \mathbf{u}_A, P_{A_i}, \emptyset; \mathbf{V}_k)$
 $View_2 = (pk, \mathbf{B}, m_2, \mathbf{u}_B, P_{B_i}, \mathbf{c}_X; m_1, P_{A_i}, \mathbf{c}_v)$
2) 用户 1 的视角中, 并没有获取到任何有关用户 2 的信息, 其输出为协议的输出, 即双方共享的降维矩阵 \mathbf{V}_k , 因此容易得到 $View_1 \equiv S_1$ 。

3) 向模拟器 S_2 输入 $(pk, \mathbf{B}, m_2, \mathbf{u}_B, P_{B_i}, \mathbf{c}_X, m_1, P_{A_i})$, 使用公钥 pk 加密一个 n 维的随机向量 \mathbf{c}_v' , 可以得到 $S_2 = (pk, \mathbf{B}, m_2, \mathbf{u}_B, P_{B_i}, \mathbf{c}_X, m_1, P_{A_i}; \mathbf{c}_v')$ 。由 CKKS 方案及 RLWE 问题的困难型可知, \mathbf{c}_v 与 \mathbf{c}' 在计算上是不可区分的, 因此 $View_2 \equiv S_2$ 。

综上所述, 协议 2 在半诚实模型下是安全的。

4 实验结果

本方案选择的多项式次数为 16 384 及 32 768, 对应 $\log q$ 分别为 360 和 520, 误差分布的标准差取的是默认值 3. 19。通过 lattice estimator^[33] 计算可得, 在这样的参数设置下, 方案至少可以达到 128 bit 的安全性。用 SEAL 库与 Panda 方案进行了对比实验, 计算了相同数据集的前 3 个主成分, 选择

的多项式次数均为 16 384, 结果如表 2、表 3 所列。可以看出, 本文方案在精度上有显著的提升且时间开销大幅减少。

表 2 红酒质量数据集^[34]上的实验结果

数据规模		1 599 × 11			
主成分	明文结果	Panda 方案 密文结果	本文方案 密文结果	Panda 方案密文 计算耗时/s	本文方案密文 计算耗时/s
1	2. 991 200	2. 991 277	2. 991 200		
2	0. 676 158	0. 172 909	0. 676 158	86. 9	9. 8
3	0. 156 440	3. 693 241	0. 156 389		

表 3 人脸数据集^[35]上的实验结果

数据规模		165 × 256			
主成分	明文结果	Panda 方案 密文结果	本文方案 密文结果	Panda 方案密文 计算耗时/s	本文方案密文 计算耗时/s
1	1. 713 911	1. 623 415	1. 667 288		
2	1. 587 648	2. 667 312	1. 589 215	758. 5	145. 4
3	1. 351 156	7. 258 823	1. 397 309		

方案在其他的数据集上也进行了实验, 在维数较小的数据集上求解每轮主成分时进行了 3 次幂法迭代, 在维数较高的数据集上选择了 5 次迭代。对解密后的前三个主成分与标准的明文结果计算了均方误差 (Mean-square Error, MSE), 如表 4 所列。其中前 3 行数据集所用多项式次数为 32 768, 其余数据集选用多项式次数为 16 384。结合表 2—表 4 可以看到, 相比 Panda 方案, 本文方案在精度上获得了显著的提升且时间开销大幅减少。

表 4 不同数据集上的实验结果
Table 4 Experiment results on different datasets

数据集	样本数	维数	N	迭代次数	均方误差	密文计算耗时/s
MNIST ^[36]	200	256	32 768	5	2.25×10^{-3}	147. 5
Fashion-MNIST ^[37]	200	256	32 768	5	3.76×10^{-3}	147. 7
Yale ^[35]	165	256	32 768	5	1.44×10^{-3}	145. 4
Winequality-white ^[34]	4 898	11	16 384	3	8.62×10^{-10}	10. 2
Winequality-red ^[34]	1 599	11	16 384	3	1.13×10^{-10}	9. 8
Air Quality ^[38]	9 357	13	16 384	3	2.27×10^{-7}	11. 9
Parkinsons ^[39]	197	16	16 384	3	6.69×10^{-5}	13. 7

结束语 基于同态加密技术, 提出了一种两方参与的隐私保护 PCA 的解决方案, 并对方案的准确性、效率和安全性进行了评估。与现有方案相比, 其不仅支持了浮点运算, 还避免了低效的同态除法操作, 在计算耗时上缩短了约 80%, 与明文计算结果的均方误差缩减到 1% 以内; 能够在较小的通信开销下完成两方数据的联合降维, 同时有效地保护了数据的隐私。

目前的方案只考虑了两方共用数据的模式, 没有扩展到多用户的场景。在未来的研究工作中, 希望通过结合安全多方计算或多密钥全同态加密技术来对方案进行拓展, 以实现多用户的联合数据降维。对本文方案而言, 也要进一步减少迭代次数, 从而降低计算开销; 通过减少交互次数来进一步降低通信开销, 使方案的整体效率达到最优; 寻找更好的数据打包方式以优化存储空间。在隐私保护的数据降维方面, 未来将尝试更多的数据降维方法, 并将方案应用到实际的机器学习算法预测中, 以实现更大范围的应用。

参 考 文 献

[1] TRAUTMAN L J, ORMEROD P C. Corporate directors’ and officers’ cybersecurity standard of care: The Yahoo data breach [J]. AM UL Rev, 2016, 66: 1231.

[2] PAULINA O. Russian Web hosting provider exposes data of more than 54M users [EB/OL]. [2024-02-21]. <https://cybernews.com/security/web-hosting-ucouz-uid-data-leak>.

[3] KRISHI C. Microsoft Azure Hit With The Largest Data Breach In Its History; Hundreds Of Executive Accounts Compromised [EB/OL]. [2024-02-21]. <https://techreport.com/news/microsoft-azure-hit-with-the-largest-data-breach-in-its-history-hundreds-of-executive-accounts-compromised>.

[4] NATALIE C. Bank of america customers left in the dark about data breach for 90 days [EB/OL]. [2024-02-15]. <https://www.forbes.com/advisor/personal-finance/data-breach-affects-bank-of-america-customers>.

- [5] VOIGT P, VON DEM BUSSCHE A. The eu general data protection regulation(gdpr)[M]. Springer, 2017.
- [6] BARRETT C. Are the EU GDPR and the California CCPA becoming the de facto global standards for data privacy and protection? [J]. Scitech Lawyer, 2019, 15(3): 24-29.
- [7] PARASOL M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams [J]. Computer law & security review, 2018, 34(1): 67-98.
- [8] LI F H, LI H, JIA Y, et al. Research Scope and Development Trends in Privacy Computing [J]. Journal on Communications, 2016, 37(4): 4.
- [9] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: Challenges, methods, and future directions [J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.
- [10] SABT M, ACHEMLAL M, BOUABDALLAH A. Trusted execution environment: What it is, and what it is not [C] // 2015 IEEE Trustcom/BigDataSE/Ispa. IEEE, 2015, 1: 57-64.
- [11] YAO A C. Protocols for secure computations [C] // 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). IEEE, 1982: 160-164.
- [12] DURGA PRASAD K, ADI NARAYANA REDDY K, VASUMATHI D. Privacy-Preserving Naive Bayesian Classifier for Continuous Data and Discrete Data [C] // First International Conference on Artificial Intelligence and Cognitive Computing (AICC 2018). Springer Singapore, 2019: 289-299.
- [13] GILAD-BACHRACH R, DOWLIN N, LAINE K, et al. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy [C] // International Conference on Machine Learning. PMLR, 2016: 201-210.
- [14] HESAMIFARD E, TAKABI H, GHASEMI M. Cryptodl: Deep neural networks over encrypted data [J]. arXiv: 1711. 05189, 2017.
- [15] SANYAL A, KUSNER M, GASCON A, et al. TAPAS: Tricks to accelerate(encrypted) prediction as a service [C] // International Conference on Machine Learning. PMLR, 2018: 4490-4499.
- [16] BOURSE F, MINELLI M, MINIHOLD M, et al. Fast homomorphic evaluation of deep discretized neural networks [C] // Advances in Cryptology (CRYPTO 2018): 38th Annual International Cryptology Conference. Santa Barbara, CA, USA, Part III 38. Springer International Publishing, 2018: 483-512.
- [17] LU W, KAWASAKI S, SAKUMA J. Using fully homomorphic encryption for statistical analysis of categorical, ordinal and numerical data [J]. Cryptology ePrint Archive, 2016.
- [18] RATHEE D, MISHRA P K, YASUDA M. Faster PCA and linear regression through hypercubes in HELib [C] // Proceedings of the 2018 Workshop on Privacy in the Electronic Society. 2018: 42-53.
- [19] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) fully homomorphic encryption without bootstrapping [J]. ACM Transactions on Computation Theory (TOCT), 2014, 6(3): 1-36.
- [20] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers [C] // Advances in Cryptology (ASIACRYPT 2017): 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, Part I 23. Springer International Publishing, 2017: 409-437.
- [21] PANDA S. Principal component analysis using CKKs homomorphic scheme [C] // Cyber Security Cryptography and Machine Learning, 5th International Symposium (CSCML 2021). Be'er Sheva, Israel, Springer International Publishing, 2021: 52-70.
- [22] RIVEST R L, ADLEMAN L, DERTOUZOS M L. On data banks and privacy homomorphisms [J]. Foundations of Secure Computation, 1978, 4(11): 169-180.
- [23] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems [J]. Communications of the ACM, 1978, 21(2): 120-126.
- [24] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms [J]. IEEE Transactions on Information Theory, 1985, 31(4): 469-472.
- [25] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes [C] // International Conference on the Theory and Applications of Cryptographic Techniques. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999: 223-238.
- [26] BONEH D, GOH E J, NISSIM K. Evaluating 2-DNF formulas on ciphertexts [C] // Theory of Cryptography: Second Theory of Cryptography Conference (TCC 2005). Cambridge, MA, USA, Springer Berlin Heidelberg, 2005: 325-341.
- [27] GENTRY C. Fully homomorphic encryption using ideal lattices [C] // Proceedings of the forty-first annual ACM Symposium on Theory of Computing. 2009: 169-178.
- [28] FAN J, VERCAUTEREN F. Somewhat practical fully homomorphic encryption [J/OL]. <https://eprint.iacr.org/2012/144.pdf>.
- [29] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based [C] // Advances in Cryptology-CRYPTO 2013: 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, Part I. Springer Berlin Heidelberg, 2013: 75-92.
- [30] LYUBASHEVSKY V, PEIKERT C, REGEV O. On ideal lattices and learning with errors over rings [C] // Advances in Cryptology (EUROCRYPT 2010): 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques. French Riviera, Springer, 2010: 1-23.
- [31] HALEVI S, SHoup V. Algorithms in helib [C] // Advances in Cryptology-CRYPTO 2014: 34th Annual Cryptology Conference, Santa Barbara, CA, USA, Part I 34. 2014: 554-571.
- [32] LINDELL Y. How to simulate it—A tutorial on the simulation proof technique [M] // Tutorials on the Foundations of Cryptography Information Security and Cryptography. 2017: 277-346.
- [33] ALBRECHT M R, PLAYER R, SCOTT S. On the concrete

hardness of learning with errors [J]. Journal of Mathematical Cryptology, 2015, 9(3):169-203.

[34] CORTEZ P, CERDEIRA A, ALMEIDA F, et al. Modeling wine preferences by data mining from physicochemical properties [J]. Decision Support Systems, 2009, 47(4):547-53.

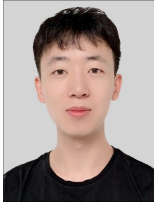
[35] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):711-720.

[36] LECUN Y, CORTES C, BURGESS C. MNIST handwritten digit database. AT&T Labs (2010) [EB/OL]. <http://yann.lecun.com/exdb/mnist>.

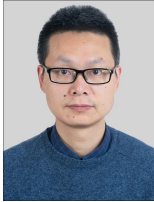
[37] XIAO H, RASUL K, VOLLGRAF R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms [J]. arXiv:1708.07747, 2017.

[38] DE VITO S, MASSERA E, PIGA M, et al. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario [J]. Sensors and Actuators B: Chemical, 2008, 129(2):750-757.

[39] TSANAS A, LITTLE M, MCSHARRY P, et al. Accurate tele-monitoring of Parkinson's disease progression by non-invasive speech tests [J/OL]. Nature Proceedings, 2009. <https://doi.org/10.1038/npre.2009.3920.1>.



ZHANG Jindou, born in 1997, postgraduate. His main research interests include homomorphic encryption and information security.



CHEN Jingwei, born in 1984, Ph.D, professor. His main research interests include error-free numerical computation and lattice-based cryptography.

(责任编辑:喻黎)

2024CCF 会士提名启动

根据《中国计算机学会会士条例》的规定,2024 年度 CCF 会士候选人提名工作即日启动。CCF 会士和 CCF 杰出会员具有提名权,请主提名人在 2024 年 11 月 1 日前将“CCF 会士提名表”逐项填妥后通过 CCF OA 系统提交。

- 一、CCF 会士候选人的资格
1. 候选人在提名截止日前(2024 年 11 月 1 日)在计算机或相关领域从业 15 年以上(受高等教育期间的从业时间按如下方式计算:学士 2 年、硕士 4 年、博士 6 年,按最高学历计算,不累计)、本学会会龄 5 年以上,并在计算机及相关领域有重大发明创造及有重要贡献、或对本学会发展有重要贡献的人士。

2. 候选人须得到 1 名主提名人和 2 名附议人的提名(附议提名人仅需 2 名)。
- 二、提名人的资格
- CCF 会士和杰出会员为有效提名人(如会员资格已失效,则提名无效)。
- 三、提名要求
1. 每位提名人作为主提名人提名的会士候选人不得超过 2 人,作为附议提名人不得超过 3 人(即主提名+附议提名总共 5 人)。

2. 作为主提名人提名时,应保证所提名的会士候选人获得另外 2 名附议提名人的提名。
- 四、重要提示
- 本次提名方式为线上提名,由主提名人填写提名表后登录 CCF OA 系统提交,附议提名人收到提醒邮件通过链接登录 CCF OA 系统提交意见后完成附议提名。如提名人及附议提名人系统预留邮箱有变更,请联系会员部更新。

联系人:刘老师 xliu@ccf.org.cn/010-6264 8654

据 CCF 微信公众号