一种隐私保护的卷积神经网络预测方案

任艳丽[1),2),3)] 余凌赞[1)] 何 港[1)] 张新鹏[1)] 郭 筝[3)]

[1)](上海大学通信与信息工程学院 上海 200444)

[2)](郑州信大先进技术研究院 郑州 450000)

[3)](智巡密码(上海)检测技术有限公司 上海 201601)

摘 要 机器学习在图像目标识别、语音识别和图像处理等领域有广泛的应用.卷积神经网络是机器学习领域中广为流行的架构,训练模型所需计算代价对资源受限的用户来说难以负担,因此越来越多的模型所有者将预测服务托管在云平台上以供用户按需使用.在现有方案中,云端处理数据时可能会泄露用户数据和模型参数,预测准确度不高,且用户与云服务器交互需要大量通信开销.本文提出隐私保护的卷积神经网络预测方案,服务器基于密文模型对用户提供的密文数据进行预测,同时保护用户的隐私数据以及模型参数.而且,用户在上传加密数据之后即可离线等待预测结果,在预测任务执行期间与服务器间无需交互.理论和实验表明,所提方案提高了CNN预测方案的安全性,降低了用户的通信代价,最高可达到93%的预测准确率,与明文数据预测准确率近似相等.

关键词 卷积神经网络;数据预测;隐私保护;同态加密;非交互性

中图法分类号 TP309 **DOI号** 10.11897/SP.J.1016.2023.01606

A Scheme of Privacy-Preserving Convolutional Neural Network Prediction

REN Yan-Li[1),2),3)] YU Ling-Zan[1)] HE Gang[1)] ZHANG Xin-Peng[1)] GUO Zheng[3)]

[1)](*School of Communication and Information Engineering, Shanghai University, Shanghai* 200444)

[2)](*Zhengzhou Xinda Institute of Advanced Technology, Zhengzhou* 450000)

[3)](*ZhiXun Crypto Testing and Evaluation Technology Co., Ltd., Shanghai* 201601)

Abstract Machine learning is a technique of universal data processing, and extensively used in image object recognition, speech recognition and image processing. Neural networks identify the relationships behind data by referring to the neuronal tissue in the brain, which can be better applied to the processing of complex data and the problems related to data prediction. Convolutional neural network(CNN) is a popular architecture in the field of machine learning, which is demonstrated excellent performance in medical image analysis, image and audio recognition and classification. However, the model structure of CNN gradually deepens in recent years, which relies on powerful physical hardware and large training data sets and the computational costs of training models are too high for limited users. Therefore, model owners gradually deploy prediction services on cloud platforms in order to meet the needs of users. At present, there are three methods which are widely used in the field of machine learning for privacy protection. One is secure multi-party computing, whose main idea is multi-party collaborative computing under the premise of privacy protection, but it requires multiple interactions of multiple participants, and the communication cost is high. The second one is the technology of differential privacy, whose main idea is to conduct distributed

收稿日期:2022-05-30;在线发布日期:2023-03-16.本课题得到国家自然科学基金重点项目(U1936214)、上海市自然科学基金(20ZR1419700,22ZR1481000)、河南省网络空间态势感知重点实验室开放课题基金(HNTS2022011)资助.**任艳丽**(通信作者),博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为公钥密码学、区块链安全、人工智能安全等.E-mail:renyanli@shu.edu.cn.**余凌赞**,硕士,主要研究方向为外包计算、人工智能安全等.**何 港**,博士研究生,主要研究方向为外包计算、人工智能安全.**张新鹏**,博士,教授,国家杰出青年,主要研究领域为媒体内容安全、人工智能安全等.**郭 筝**,博士,高级工程师,主要研究领域为密码硬件系统安全分析检测等.

model training among multiple participants with local data, and add artificial noise to the parameters of the participants' local model for privacy protection, and finally aggregate them to get the global model. The third one is homomorphic encryption, but the computational complexity of operations on ciphertext is very high, and the encryption algorithm only has homomorphism for part of operations, so it cannot realize nonlinear functions in the model. Thus, user's private data and model parameters may be leaked during data processing, and the users need to interact with the cloud servers for many times, and the prediction accuracy is not high in the existing works. In this paper, we propose a privacy-preserving CNN prediction scheme based on homomorphic encryption which can realize data prediction on the encrypted data, rather than simply randomly splitting the data, and is able to preserve the user's privacy and keep the model private from any untrusted server. Moreover, after uploading encrypted data, the users and the model owner can wait for the prediction results offline, and there is no extra interaction between user and the cloud server during the prediction process. Also, the proposed scheme supports floating-point arithmetic and does not require approximate operation for nonlinear functions, which improves the accuracy of model prediction. The experimental results show that our proposed scheme improves security of the CNN prediction, greatly reduces the user's communication costs, and can achieve prediction accuracy to 93% at most, which is almost same to that of plaintext data prediction. It is well known that most of the existing schemes based on homomorphic encryption can only achieve data privacy, but cannot achieve model privacy, prediction accuracy and non-interaction of the users. Overall, the proposed scheme improves the security and prediction accuracy of CNN, and reduces the computing and communication costs of users compared with the previous ones.

Keywords　convolutional neural networks; data prediction; privacy-preserving; homomorphic encryption; non-interactive

1　引　言

机器学习[1]是一种通用性的数据处理技术，在数据挖掘、计算机视觉、自然语言处理等领域中表现良好，并得到了广泛的应用[2-3]. 神经网络通过参考模仿大脑中的神经元组织来识别数据背后的关系，可更好地适用于处理复杂数据以及数据预测相关问题，其中卷积神经网络(Convolutional Neural Network, CNN)被证明在医学图像分析、图像和音频的识别和分类中性能优异[4].

随着研究的深入，CNN 的模型结构逐渐加深. 由于其依赖于强大的物理硬件和庞大的训练数据集，使得个人或者小微企业无法承受. 为应对此情况，机器学习服务[5](Machine Learning as a Service, MLasS)作为一种新的服务被提出且受到广泛关注，可以将复杂的计算负载给云服务器，共享其强大的计算资源. 现阶段，人工智能技术仍在不断发展优化，未来会有越来越多的企业或者学者了解并且认可机器学习的能力与优势，机器学习服务势必将得到越来越多的应用，本文的目标旨在提出隐私保护的 CNN 预测方案.

CNN 预测一般包括两个阶段：(1) 训练阶段，CNN 通过不断对训练数据进行特征的学习从而优化模型的深层结构；(2) 预测阶段，利用训练后的模型对给定的数据进行分类或回归任务[6].

MLaaS 作为一种新的云计算服务，提供了这两个阶段的完整服务，其中基于 CNN 的数据预测服务存在许多实际应用场景，比如大型医院可以通过自身拥有患者的病历和诊断记录来训练神经网络模型，并将训练完成的神经网络模型通过 MLaaS 提供给小型诊所或者患者个体用于分析远程医疗数据诊断. 本文方案是针对 CNN 的预测阶段提出的.

机器学习预测服务在给资源受限的用户提供优秀的计算资源以及便利的同时，也面临一些新的难题[7-8]，如模型拥有者将训练好的模型或用户将待预测数据发送给云服务器时，由于提供的数据以及模型总是不可避免地包含敏感或者隐私的信息，这无

疑会带来极大的隐私和安全问题.本文将提出一种有效的预测服务,能对用户的隐私数据以及模型参数进行保护,无法被云服务器或者恶意攻击者获取,保证模型隐私性以及数据隐私性.用户在上传加密数据之后即可离线等待预测结果,在预测任务执行期间与服务器无需交互,且与明文数据预测相比,不降低数据预测的准确率.

1.1 研究现状

近年来,基于云计算的隐私保护机器学习研究逐步成为热点,已经取得了很多研究成果,接下来我们介绍隐私保护的机器学习相关工作.

Dowlin 等人在 2016 年提出了 Cryptonets[9] 的外包方案,文中提出的方案可应用于加密的输入数据,并基于层次同态加密[10](Leveled Homomorhic Encryption,LHE)构建完整方案.然而,由于 LHE 的局限性,只能应用于乘法、加法等算术运算,不能直接应用于神经网络中常用的激活函数.因此,该方案采用了多项式近似的方案对激活函数进行计算,但这种方法会导致预测结果准确性的损失,同时该方案也不能保证模型参数的数据隐私性.

之后 Liu 等人提出了一种基于安全多方计算的 MiniONN 协议[11],实现了用户与服务器之间安全秘密地共享数据[12].该协议可分为离线预计算阶段和在线预测阶段,虽然在线阶段的效率很高,但用户需要与服务器提前准备并计算大量乘法三元组,而这些乘法三元组的构建是基于同态加密完成的,故该方案存在大量预计算.而且用户在预测任务进行期间需保持在线状态,和服务器之间存在大量交互操作,因此用户的计算和通信代价很高.

XONN 协议[13]是一种基于混淆电路(Garbled Circuit,GC)的安全外包协议[14],深度学习模型中代价高昂的矩阵乘法操作被 XNOR 操作取代,计算函数编译成布尔电路的形式,利用二进制 XNOR 操作在 GC 混淆电路中的通用性以及便捷性进行外包计算且不泄露参与计算的双方私有信息.但该方法仅适用于二进位神经网络,其中的参数均为二进位,严重限制了神经网络预测结果精度.

SecureML[15] 和 CNN 预测外包方案[16] 比较类似,都是使用基于同态加密的秘密共享和安全多方计算技术[17],后者在两台服务器之间秘密共享用户提供的预测数据和模型参数,并将用户数据和模型参数随机拆分.虽然该协议大大降低了预计算阶段所需计算量,但在两台服务器之间随机拆分的方式可能被恶意攻击者截获并获取到隐私的模型参数.

目前,实现隐私保护的机器学习预测任务主要有三大挑战[18-19]:(1)确保外包方案中各参与方的数据隐私性,即用户本地数据、模型参数以及预测结果对于其他任一参与方来说都不可获取;(2)满足云服务器最终给出的预测结果的准确性;(3)保证整个过程的计算效率,即用户所需计算量需要尽可能地减小,并且在提交加密数据之后无需在线与云服务器进行交互.在已有研究中,CNNP 方案[16] 对用户输入数据以及模型提供者提供的模型参数只是进行简单的随机拆分处理,当拆分之后的数据被截获时,恶意攻击者可以恢复出用户隐私数据以及模型参数.

目前隐私保护机器学习领域中应用较为广泛的主要有三种办法:(1)安全多方计算[20],其主要思想是在隐私保护的前提下多方协同计算,但需要多个参与者进行多次交互,通信代价较高;(2)差分隐私技术,常用于联邦学习中[21],其主要思想是在多个拥有本地数据的参与方之间进行分布式模型训练,并在参与方的本地模型参数中添加人工噪声进行隐私保护,最终聚合得到全局模型;(3)同态加密,已有同态加密方案虽然保证数据安全,但加密后在密文上进行运算的计算复杂度很大,并且加密算法仅对部分运算具有同态性,无法实现模型中的非线性函数,因此已有研究大多采取近似的方法,会损失预测结果的精准度.在以上三种方法中,安全多方计算需要参与方的多次交互,而差分隐私更适用于联邦学习训练,因此本文所提方案采用 CKKS 同态加密算法[22],支持浮点向量在密文空间的加法和乘法运算并保持同态性,且无需近似操作,能够有效保证计算结果的精准度.

本文提出隐私保护的 CNN 预测方案,服务器可对用户提供的密文数据进行预测.方案采用同态加密对用户的隐私数据以及模型参数进行保护,实现了数据隐私性,此外,用户在上传加密数据之后即可离线等待预测结果,在预测任务进行期间与服务器间无需交互.实验表明所提预测方案能够达到 89%～93%的准确率,与明文数据预测准确率误差较小,即能保证数据预测的隐私性.

最近,Wei 等人基于 CKKS 算法,通过两个参与方的多次交互实现了隐私保护的逻辑回归方案[23],包括线性操作和指数形式的非线性操作,但只能实现数据的二分类.Lehmkuhl 等人[24] 和 Chandran 等人[25] 基于单个不可信的服务器,提出了隐私保护的神经网络预测方案,但服务器需要与用户进行多次

交互，计算和通信代价很高. 所提方案基于 CKKS 算法实现了隐私保护的卷积神经网络预测方案，包括线性操作和非线性操作，其中非线性操作是 ReLU 函数，不同于文献[23]中的指数函数. 而且，与文献[24-25]相比，所提方案无需用户间的多次交互，只需把加密数据上传到云服务器，预测结束后即可下载分类结果，降低了用户的通信代价. 所提方案适用于多分类场景，实验可对手写数字进行 0～9 的分类. 因此，与文献[23-25]相比，所提方案基于密文数据实现了非线性 ReLU 函数，无需用户和服务器之间的多次交互，且实现了加密数据的多分类.

1.2　本文贡献

我们提出了隐私保护的 CNN 预测方案，主要有以下创新点：

(1) 采用同态加密对用户隐私数据和模型参数进行保护，而不是仅仅对数据进行简单地随机拆分，同时确保用户数据和模型参数的隐私性，实现了秘密数据的隐私增强；

(2) 用户与模型拥有者在发送隐私数据后即可离线等待结果，预测阶段无需在线与服务器进行交互，降低了计算和通信代价；

(3) 所提方案支持浮点数运算，对于非线性函数无需采取近似操作，提高了模型预测的精准度. 实验表明所提方案预测准确率与明文数据预测准确率误差较小.

所提方案涉及到用户和模型提供者，双方均需要保护各自的隐私，直接应用同态加密并不能达到同时保护双方隐私的目标. 而为了实现用户在方案中的非交互性，所提方案基于两个不可信的服务器，基于密文数据和密文模型协作完成预测任务，并实现了预测结果的隐私性. 已有基于同态加密的方案至多只能实现数据隐私性，并不能实现模型隐私性、预测的精确性与用户的非交互性，而所提方案同时实现了用户数据、模型参数以及预测结果的隐私性，预测精度较高，且用户除上传数据外，只需离线等待预测结果，实现了非交互性. 综上，与已有基于同态加密的方案相比，所提方案提高了卷积神经网络预测的安全性和预测精度，降低了用户的计算和通信代价.

2　背景知识及定义

本节将介绍所提方案涉及的基础知识，包含 CKKS 同态加密、卷积神经网络介绍以及方案系统模型.

2.1　CKKS 同态加密

本文采用的同态加密方案是 2017 年提出的 CKKS(Cheon-Kim-Kim-Song)同态加密算法[22]. 该方案基于 BGV(Brakerski-Gentry-Vaikuntanathan)方案[26]，采用近似计算方式，通过放宽准确性的限制从而使得计算效率有较大提升. 方案的安全性基于 Lyubashevsky 等人提出的 RLWE(Ring Learning with Errors)问题[27]，支持浮点向量在密文空间的加减法和乘法运算，适用于隐私保护的机器学习等相关领域.

CKKS 方案建立在环 $R[x]=\mathbb{Z}[X]/(X^N+1)$ 上，N 代表明文空间对应环的多项式次数，密文空间建立在环 $R_Q^2[x]=\left(\frac{\mathbb{Z}_Q[X]}{X^N+1}\right)^2$ 上，Q 是密文空间的模数，同时也是对应最大层次中的模数.

我们用 $x\leftarrow D$ 表示从分布 D 中随机抽样产生 x 值，CKKS 方案是有限层次全同态加密(Fully Leveled Homomorphic Encryption)，每一个密文对应了一个深度，L 代表深度上限，λ 为安全参数，q 表示任一层的模数，其中 q 是 Q 的因子. 接下来我们将具体介绍 CKKS 加密算法及其同态性.

(1) 密钥生成算法

选取大素数 p 和整数 q_0，计算 $Q=q_0\cdot p^L$，生成 R 上的随机分布以及错误概率分布 $\mathcal{X}_{key}$，$\mathcal{X}_{err}$，$\mathcal{X}_{enc}$. 随机均匀选取 $s\leftarrow\mathcal{X}_{key}$，$a\leftarrow R_Q$，$e\leftarrow\mathcal{X}_{err}$，设置私钥为 $sk\leftarrow(1,s)$，以及公钥 $pk\leftarrow(b,a)\in R_Q^2$，其中 $b=-a\cdot s+e \bmod Q$.

之后随机均匀选取 $a'\leftarrow R_{PQ}$，$e'\leftarrow\mathcal{X}_{err}$，并选取一个模数 P 用于计算辅助密钥，输出 $evk=(b',a')$，其中 $b'=-a'\cdot s+e'+P\cdot s^2 \bmod P\cdot Q$.

(2) 加密算法

对于明文 $m\in R$，从用于加密的随机分布 $\mathcal{X}_{enc}$ 中均匀选取一个向量 $\boldsymbol{r}\leftarrow\mathcal{X}_{enc}$，以及 e_0，$e_1\leftarrow\mathcal{X}_{err}$，则加密之后输出的密文 $c=\boldsymbol{r}\cdot pk+(m+e_0,e_1) \bmod Q$.

(3) 解密算法

输入密文 $c\in R_q^2$ 以及私钥 $sk=(1,s)$，对应解密明文为 $m'=\langle c,sk\rangle \bmod Q$.

解密算法正确性验证如下：

$$
\begin{aligned}
m' &= \langle c,sk\rangle \bmod Q\\
&= \langle (b\cdot\boldsymbol{r}+m+e_0, a\cdot\boldsymbol{r}+e_1),(1,s)\rangle \bmod Q\\
&= ((-a\cdot s+e)\cdot\boldsymbol{r}+m+e_0)+(a\cdot\boldsymbol{r}\cdot s+e_1\cdot s) \bmod Q\\
&= e\cdot\boldsymbol{r}+m+e_0+e_1\cdot s \bmod Q\approx m.
\end{aligned}
$$

最后对 CKKS 的同态性进行介绍，主要包括加法以及乘法同态性.

(1) 加法同态. 假定 c 和 c' 分别为 m 和 m' 的密文，计算 $c_{Add}=c\oplus c' \bmod Q$，则 $Dec_{sk}(c_{Add})=Dec_{sk}(c\oplus c')\approx m+m'$.

(2) 乘法同态. 给定密文 c 和 c'，其中 $c=(c_0,c_1)$，$c'=(c_0',c_1')$.

计算 $c_{Mult}=c\otimes c' \bmod Q=(d_0,d_1)+\lfloor P^{-1}\cdot d_2\cdot evk\rceil \bmod Q$，其中 $c_{Mult}\in R_q^2$，$\lfloor\cdot\rceil$ 表示取整，$(d_0,d_1,d_2)=(c_0c_0',c_0c_1'+c_1c_0',c_1c_1')$.

则 $Dec_{sk}(c_{Mult})\approx(d_0,d_1)\cdot sk+s^2d_2 \bmod Q\approx m\cdot m'$.

2.2 CNN 预测算法

卷积神经网络(Convolutional Neural Network，CNN)是目前深度学习领域一个基础模型，因其在图像影音识别方面的优异表现被广泛研究和使用，也成为很多后续延伸拓展模型的基础架构. 本节我们将简单介绍基于卷积神经网络的预测是如何实现的.

图 1 是非常经典的 LeNet-5 模型结果，整个卷积神经网络由几部分构成，分别是卷积层、激活层、池化层以及全连接层.

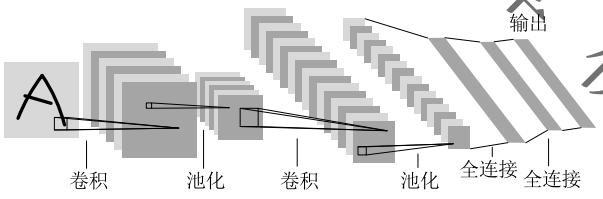

图 1 卷积神经网络预测流程图

(1) 卷积层

卷积层概括来说是利用卷积核来提取原始图像的特征，通过对图像与卷积核对应矩阵做内积运算，具体为将两个矩阵内所有对应位置的元素两两相乘并计算和值，并按照步长(stride)不断平移，直到原始图像中所有数据信息都被提取出特征相关信息. 不同大小及参数的卷积核可以提取到对应不同的特征(颜色深浅，轮廓大小). 随着卷积核的深度以及数量的提升，则更容易提取到更复杂的特征. 具体卷积操作可参考图 2.

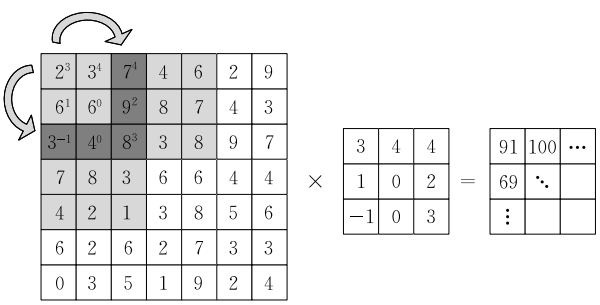

图 2 卷积层卷积运算示意图

(2) 激活层

激活层的具体作用是将卷积层的输出结果进行非线性映射，将输入数据通过某种函数映射到特定范围内，再往更深的层进行传递，目的是防止数据过大所造成的溢出风险，同时还可以使整个卷积神经网络更加强大，优化网络的性能. 本文采用的是常见的 ReLU 激活函数，即

$$f(x)=\max(0,x) \tag{1}$$

ReLU 激活函数没有复杂的指数运算，计算简单、有利于提升效率，且实际收敛速度较快，比较适用于我们的结构.

(3) 池化层

池化层的作用是使用平均池化法(mean-pooling)或最大池化(max-pooling)技术将特定范围内的特征点采取特定运算得到新的特征，缩减特征图的尺寸. 本文所提方案中采用的是平均池化法，其原理可概括为将每一个 $q\times q$ 窗口内的 q^2 个数据计算对应平均值作为平均池化的输出.

(4) 全连接层

在整个卷积神经网络中，全连接层类似于“分类器”的作用，将前层计算得到的特征空间通过线性变换映射到另一特征空间. 如图 3 所示，每一个神经元都会与前后层的所有节点相互连接，并且输入与输出都被延展成一维向量，故其核心操作就是矩阵与向量的乘积，数据在进入全连接层会先进行一维化处理，之后与对应的权重参数 w_i 以及偏置 b_i 进行矩阵与向量的计算. 全连接层中每个中间神经元的输出都可以表示为

$$output=\sum_k w_ix_i+b_i \tag{2}$$

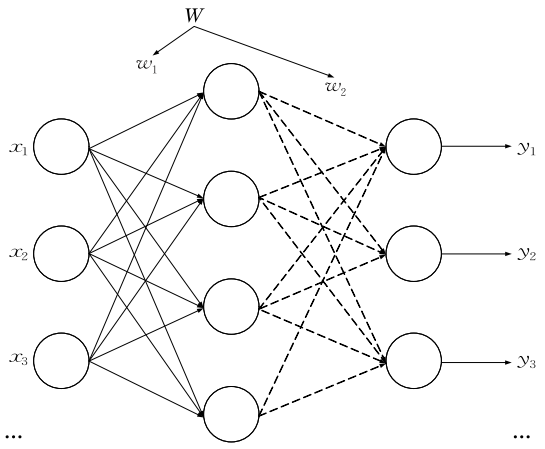

图 3 全连接层计算示意图

整个全连接层可以表示为

$$Y=WX+B \tag{3}$$

2.3 系统模型

图 4 展示了所提预测方案的系统模型. 为了实现非交互的预测方案，降低用户的计算和通信代价，所提方案设定了四个参与方：用户、模型拥有者、云服务器 1、云服务器 2. 在系统模型中，用户拥有输入图像 X；模型提供者拥有卷积核 C、全连接层参数 W 和 b，云服务器 1 和云服务器 2 将基于图像密文和加密模型进行交互计算，实现输入图像的预测服务.

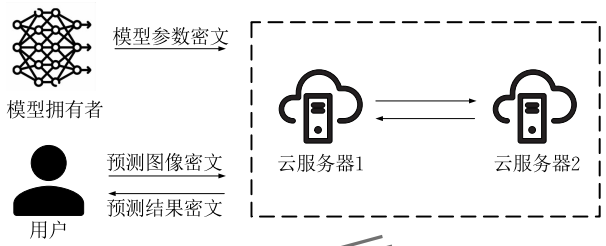

图 4　隐私保护的卷积神经网络预测方案

假设云服务器 1 和云服务器 2 均为诚实且好奇的，即两个参与方都是诚实地按照方案进行计算，但是会试图获取原始数据、模型参数及预测结果.

在所提方案中，用户和模型拥有者分别加密隐私数据和模型参数并发送给云服务器 1，之后云服务器 1 和云服务器 2 交互进行卷积神经网络预测，最终云服务器 1 将预测结果密文返回给用户，用户解密得到预测结果. 具体流程如下：

(1) 用户和模型拥有者加密本地数据和模型参数并将密文发送给云服务器 1；

(2) 云服务器 1 和云服务器 2 交互进行卷积神经网络预测方案中卷积层、激活层、池化层、全连接层算法；

(3) 云服务器 1 计算得到最终的预测结果密文，并将其返回给用户；

(4) 用户用私钥解密预测结果密文，得到最终的预测结果.

3 具体方案

本小节将具体介绍隐私保护的卷积神经网络预测方案，具体包括：初始化算法 *Initialization*、卷积层算法 *Convolution*、激活层算法 *Activation*、池化层算法 *Pooling*、全连接层算法 *Fully-Connection* 以及最终的解密算法 *Decryption*.

3.1 初始化算法 *Initialization*

假设用户持有 $n\times n$ 的本地图像 $X=(X_{1,1},X_{1,2},\cdots,X_{1,n},X_{2,1},\cdots,X_{n,n})$，模型拥有者持有卷积核 $C=(C_{1,1},C_{1,2},\cdots,C_{1,m},C_{2,1},\cdots,C_{m,m})$，全连接层参数 $W=(W_{1,1},W_{1,2},\cdots,W_{1,k},W_{2,1},\cdots,W_{l,k})$ 以及偏置参数 $B=(B_1,B_2,\cdots,B_k)$，其中 m,l,k 都是对应模型参数矩阵或向量的维度大小，且 $n>m$.

(1) 给定安全参数 λ，云服务器 1、云服务器 2 和用户分别生成各自的公私钥对 (PK_P,SK_P)，(PK_C,SK_C) 和 (PK_U,SK_U)，公开公钥 PK_P,PK_C,PK_U，并各自保管自己的私钥.

(2) 用户选择随机数 r_X，基于云服务器 2 的公钥 PK_C 加密数据 X 得到 $\boldsymbol{E}_{PK_C}(r_XX)$，加密后的结果为

$$\boldsymbol{E}_{PK_C}(r_XX_{1,1}),\boldsymbol{E}_{PK_C}(r_XX_{1,2}),\cdots,\boldsymbol{E}_{PK_C}(r_XX_{n,n}),$$

之后用自己的公钥 PK_U 加密 r_X^{-1} 得到 $\boldsymbol{E}_{PK_U}(r_X^{-1})$；

(3) 模型拥有者选择随机数 r_C,r_W，用云服务器 2 的公钥 PK_C 加密卷积核 C，得到加密后数据 $\boldsymbol{E}_{PK_C}(r_CC)$，即

$$\boldsymbol{E}_{PK_C}(r_CC_{1,1}),\boldsymbol{E}_{PK_C}(r_CC_{1,2}),\cdots,\boldsymbol{E}_{PK_C}(r_CC_{m,m}),$$

加密后的矩阵可见图 5. 同理，模型拥有者基于用户的公钥 PK_U 加密原始全连接层参数 W,B 得到 $\boldsymbol{E}_{PK_U}(r_WW),\boldsymbol{E}_{PK_U}(r_Cr_WB)$. 最后用云服务器 1 的公钥 PK_P 加密 $r_C^{-1}r_W^{-1}$ 得到 $\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1})$；

大小为 $n\times n$ 的图像密文

$\boldsymbol{E}_{PK_C}(r_XC_{1,1})$	$\boldsymbol{E}_{PK_C}(r_XC_{1,2})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{1,m})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{1,n})$
$\boldsymbol{E}_{PK_C}(r_XC_{2,1})$	$\boldsymbol{E}_{PK_C}(r_XC_{2,2})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{2,m})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{2,n})$
⋮	⋮		⋮		⋮
$\boldsymbol{E}_{PK_C}(r_XC_{m,1})$	$\boldsymbol{E}_{PK_C}(r_XC_{m,2})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{m,m})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{m,n})$
⋮	⋮		⋮		⋮
$\boldsymbol{E}_{PK_C}(r_XC_{n,1})$	$\boldsymbol{E}_{PK_C}(r_XC_{n,2})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{n,m})$	…	$\boldsymbol{E}_{PK_C}(r_XC_{n,n})$

大小为 $m\times m$ 的卷积密文

$\boldsymbol{E}_{PK_C}(r_CC_{1,1})$	$\boldsymbol{E}_{PK_C}(r_CC_{1,2})$	…	$\boldsymbol{E}_{PK_C}(r_CC_{1,m})$
$\boldsymbol{E}_{PK_C}(r_CC_{2,1})$	$\boldsymbol{E}_{PK_C}(r_CC_{2,2})$	…	$\boldsymbol{E}_{PK_C}(r_CC_{2,m})$
⋮	⋮		⋮
$\boldsymbol{E}_{PK_C}(r_CC_{m,1})$	$\boldsymbol{E}_{PK_C}(r_CC_{m,2})$	…	$\boldsymbol{E}_{PK_C}(r_CC_{m,m})$

图 5　初始化之后的密文矩阵

(4) 用户将密文数据 $\boldsymbol{E}_{PK_U}(r_X^{-1})$，$\boldsymbol{E}_{PK_C}(r_XX)$ 打包发送给云服务器 1，模型拥有者将密文数据 $\boldsymbol{E}_{PK_C}(r_CC)$，$\boldsymbol{E}_{PK_U}(r_WW)$，$\boldsymbol{E}_{PK_U}(r_Cr_WB)$，$\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1})$ 打包发送给云服务器 1.

3.2 卷积层算法 *Convolution*

云服务器 1 在接收到用户和模型拥有者发送的密文矩阵 $\boldsymbol{E}_{PK_C}(r_XX)$，$\boldsymbol{E}_{PK_C}(r_CC)$ 之后，进行卷积层算法. 具体操作为在 $\boldsymbol{E}_{PK_C}(r_XX)$ 上按步长为 1 逐个

选取 $m\times m$ 子图像与 $\boldsymbol{E}_{PK_C}(r_C C)$ 计算卷积运算，并遍历完整个密文图像，直到生成完整的卷积后的矩阵. 具体卷积操作可见算法 ENC-CON.

算法 1. 密文卷积算法 ENC-CON.

输入：$\boldsymbol{E}_{PK_C}(r_X X)$，$\boldsymbol{E}_{PK_C}(r_C C)$

输出：$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)]$

FOR $i=1$ TO $\lfloor n-m+1\rfloor$

FOR $j=1$ TO $\lfloor n-m+1\rfloor$

SET $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{i,j}]=0$

FOR $a=1$ TO m

FOR $b=1$ TO m

云服务器 1 计算

$\boldsymbol{E}_{PK_C}(r_X X_{i+a-1,j+b-1})\otimes\boldsymbol{E}_{PK_C}(r_C C_{a,b})$

$=\boldsymbol{E}_{PK_C}(r_X X_{i+a-1,j+b-1}\times r_C C_{a,b})$

$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{i,j}]$

$=\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{i,j}]\oplus$

$\boldsymbol{E}_{PK_C}(r_X X_{i+a-1,j+b-1}\times r_C C_{a,b})$

END

END

END

END

接下来我们具体解释密文卷积算法 ENC-CON. 首先云服务器 1 基于 $\boldsymbol{E}_{PK_C}(r_C C)$ 与大小为 $m\times m$ 的密文子图像进行同态乘法：

$$\boldsymbol{E}_{PK_C}(r_X X_{i,j})\otimes\boldsymbol{E}_{PK_C}(r_C C_{1,1})=\boldsymbol{E}_{PK_C}(r_X X_{i,j}\times r_C C_{1,1})$$

$$\boldsymbol{E}_{PK_C}(r_X X_{i,j+1})\otimes\boldsymbol{E}_{PK_C}(r_C C_{1,2})=\boldsymbol{E}_{PK_C}(r_X X_{i,j+1}\times r_C C_{1,2})$$

$$\cdots$$

$$\boldsymbol{E}_{PK_C}(r_X X_{i+m,j+m})\otimes\boldsymbol{E}_{PK_C}(r_C C_{m,m})=\boldsymbol{E}_{PK_C}(r_X X_{i+m,j+m}\times r_C C_{m,m}) \tag{4}$$

之后云服务器 1 基于密文同态乘法的计算结果 $\boldsymbol{E}_{PK_C}(r_X X_{i,j}\times r_C C_{1,1}),\cdots,\boldsymbol{E}_{PK_C}(r_X X_{i+m,j+m}\times r_C C_{m,m})$ 进行如下同态加法运算：

$$\begin{aligned}&\boldsymbol{E}_{PK_C}(r_X X_{i,j}\times r_C C_{1,1})\oplus\boldsymbol{E}_{PK_C}(r_X X_{i,j+1}\times r_C C_{1,2})\oplus\cdots\oplus\boldsymbol{E}_{PK_C}(r_X X_{i+m,j+m}\times r_C C_{m,m})\\&=\boldsymbol{E}_{PK_C}(r_X X_{i,j}\times r_C C_{1,1}+r_X X_{i,j+1}\times r_C C_{1,2}+\cdots+r_X X_{i+m,j+m}\times r_C C_{m,m})\\&=\boldsymbol{E}_{PK_C}[r_X r_C(X_{i,j}C_{1,1}+X_{i,j+1}C_{1,2}+\cdots+X_{i+m,j+m}C_{m,m})]\\&=\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{i,j}]\end{aligned} \tag{5}$$

云服务器 1 将子图像逐个进行密文卷积算法 ENC-CON，直到完成整张图像的卷积，详细流程图可见图 6. 最终得到完整卷积结果为大小为 $(n-m+1)\times(n-m+1)$ 的密文矩阵，其中包括 $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{1,1}]$，$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{1,2}]$，$\cdots$，$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{n-m+1,n-m+1}]$），记为 $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)]$，并将卷积层密文计算结果 $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)]$ 发送给云服务器 2.

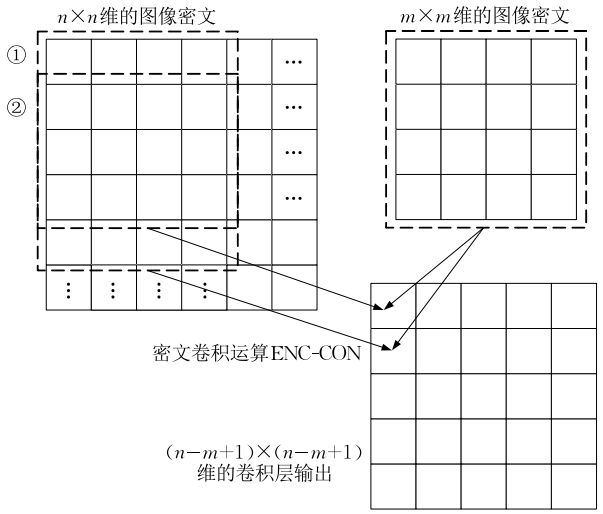

图 6 卷积层算法计算示意图

3.3 激活层算法 *Activation*

本方案采用的激活函数是 ReLU 激活函数，即 $f(x)=\max(0,x)$.

云服务器 2 接收云服务器 1 发送的 $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)]$，包含（$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{1,1}]$，$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{1,2}]$，$\cdots$，$\boldsymbol{E}_{PK_C}[r_X r_C(X*C)_{n-m+1,n-m+1}]$）.

云服务器 2 用自己的私钥 SK_C 对 $\boldsymbol{E}_{PK_C}[r_X r_C(X*C)]$ 进行解密：

$$D_{SK_C}[\boldsymbol{E}_{PK_C}(r_X r_C(X*C))]=r_X r_C(X*C) \tag{6}$$

由于 $r_X r_C$ 恒为正数，所以 $r_X r_C(X*C)$ 的正负性与 $X*C$ 保持一致. 因此只需判断解密之后得到的 $r_X r_C(X*C)$ 的正负性即可获得其经过 ReLU 激活函数后的输出，即

$$output=\begin{cases}r_X r_C(X*C), & r_X r_C(X*C)>0\\0, & \text{其他}\end{cases} \tag{7}$$

记子图像 $X_1=(X_{1,1},X_{1,2},\cdots,X_{m,m})$ 与卷积核 $C_1=(C_{1,1},C_{1,2},\cdots,C_{m,m})$ 的卷积结果为 $(X*C)_{1,1}$，我们提出的激活层算法见算法 2.

算法 2. 密文激活层算法 ENC-RELU.

输入：$\boldsymbol{E}_{PK_C}(r_X r_C(X*C))$，$SK_C$

输出：$r_X r_C(X*C)'$

FOR $i=1$ TO $\lfloor n-m+1\rfloor$

FOR $j=1$ TO $\lfloor n-m+1\rfloor$

云服务器 2 解密

$D_{SK_C}[\boldsymbol{E}_{PK_C}(r_X r_C(X*C))_{i,j}]=r_X r_C(X*C)_{i,j}$

IF $r_X r_C(X*C)_{i,j}>0$

THEN $r_X r_C(X*C)'_{1,1}=r_X r_C(X*C)$

ELSE

THEN $r_X r_C (X*C)'_{1,1}=0$

END

END

在经过密文激活层算法之后，云服务器 2 会得到 $r_X r_C (X*C)'$，其中包括 $r_X r_C (X*C)'_{1,1}$，$r_X r_C (X*C)'_{1,2}$，…，$r_X r_C (X*C)'_{n-m+1,n-m+1}$.

3.4　池化层算法 *Pooling*

本方案中我们选取平均池化作为池化层算法.

云服务器 2 在经过激活层算法 *Activation* 之后得到 $r_X r_C (X*C)'$，其中包括($r_X r_C (X*C)'_{1,1}$，$r_X r_C (X*C)'_{1,2}$，…，$r_X r_C (X*C)'_{n-m+1,n-m+1}$)，之后云服务器 2 基于 $r_X r_C (X*C)'$ 进行密文平均池化操作. 假设平均池化的窗口为 $q\times q$，其原理就是对 $r_X r_C (X*C)'$ 中的每一个 $q\times q$ 的区域进行聚合统计，输出区域内所有值的平均值，具体操作流程可见算法 3.

算法 3.　密文平均池化算法 ENC-AVG.

输入：$r_X r_C (X*C)'$

输出：$r_X r_C P$

FOR $i=1$ TO $\left\lfloor\frac{n-m+1}{q}\right\rfloor$

FOR $j=1$ TO $\left\lfloor\frac{n-m+1}{q}\right\rfloor$

SET $sum=0$

FOR $a=1$ TO q

FOR $b=1$ TO q

$sum=sum+r_X r_C (X*C)'_{(i-1)\times q+a,(j-1)\times q+b}$

END

END

$output=sum/q^2=r_X r_C P_{i,j}$

END

END

设经过平均池化后 $(X*C)'$ 中的第一个 $q\times q$ 窗口的结果为 $P_{1,1}$，则

$$P_{1,1}=\frac{[(X*C)'_{1,1}+(X*C)'_{1,2}+\cdots+(X*C)'_{q,q}]}{q^2} \tag{8}$$

我们以第一个 $q\times q$ 的区域对算法 3 进行说明. 云服务器 2 计算

$$\frac{r_X r_C (X*C)'_{1,1}+r_X r_C (X*C)'_{1,2}+\cdots+r_X r_C (X*C)'_{q,q}}{q^2}=\frac{r_X r_C[(X*C)'_{1,1}+(X*C)'_{1,2}+\cdots+(X*C)'_{q,q}]}{q^2} \tag{9}$$

由于每个具体的数据都用随机数 $r_X r_C$ 进行盲化，所以可得云服务器 2 在这个 $q\times q$ 窗口经过平均池化的输出 $r_X r_C\frac{[(X*C)'_{1,1}+(X*C)'_{1,2}+\cdots+(X*C)'_{q,q}]}{q^2}$，即 $r_X r_C P_{1,1}$. 云服务器 2 遍历完卷积层输出结果 $r_X r_C (X*C)'$ 上所有区域，流程示意图可参考图 7. 最终云服务器 2 会得到池化层输出 $r_X r_C P$，其中包括 $r_X r_C P_{1,1}$，$r_X r_C P_{1,2}$，…，$r_X r_C P_{\left\lfloor\frac{n-m+1}{q}\right\rfloor,\left\lfloor\frac{n-m+1}{q}\right\rfloor}$. 得到 $r_X r_C P$ 之后，云服务器 2 会使用用户公钥 PK_U 加密 $r_X r_C P$ 得到 $\boldsymbol{E}_{PK_U}(r_X r_C P)$. 之后云服务器 1 将 $\boldsymbol{E}_{PK_U}(r_X^{-1})$ 发送给云服务器 2，云服务器 2 基于二者计算

$$\boldsymbol{E}_{PK_U}(r_X r_C P)\otimes\boldsymbol{E}_{PK_U}(r_X^{-1})=\boldsymbol{E}_{PK_U}(r_C P) \tag{10}$$

之后云服务器 2 将最终计算结果 $\boldsymbol{E}_{PK_U}(r_C P)$ 返回给云服务器 1.

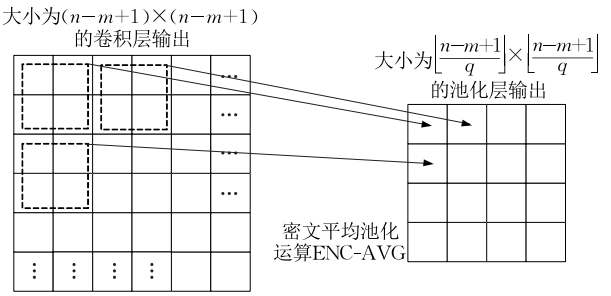

图 7　平均池化层算法流程示意图

3.5　全连接层算法 *Fully-Connection*

首先，云服务器 1 用自己保管的私钥 SK_P 解密 *Initialization* 阶段收到的密文 $\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1})$

$$D_{SK_P}[\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1})]=r_C^{-1}r_W^{-1} \tag{11}$$

然后使用用户公钥 PK_U 加密 $r_C^{-1}r_W^{-1}$ 得到 $\boldsymbol{E}_{PK_U}(r_C^{-1}r_W^{-1})$.

由于全连接层算法与卷积层算法较为类似，都是基于密文进行加法和乘法运算，故本小节不做详述. 概括来说，云服务器 1 接收到云服务器 2 在池化之后返回的 $\boldsymbol{E}_{PK_U}(r_C P)$ 后计算：

$$\boldsymbol{E}_{PK_U}(r_W W)\otimes\boldsymbol{E}_{PK_U}(r_C P)=\boldsymbol{E}_{PK_U}(r_W r_C WP) \tag{12}$$

并将结果与模型拥有者在 *Initialization* 阶段发送的 $\boldsymbol{E}_{PK_U}(r_C r_W B)$ 进行密文加法运算：

$$\boldsymbol{E}_{PK_U}(r_W r_C WP)\oplus\boldsymbol{E}_{PK_U}(r_C r_W B)=\boldsymbol{E}_{PK_U}[r_W r_C (WP+B)] \tag{13}$$

最后云服务器 1 计算

$$\boldsymbol{E}_{PK_U}(r_C^{-1}r_W^{-1})\otimes\boldsymbol{E}_{PK_U}[r_W r_C (WP+B)]=\boldsymbol{E}_{PK_U}(WP+B) \tag{14}$$

云服务器 1 得到最终预测结果的密文，并将其发送给用户.

3.6　解密算法 *Decryption*

用户接收到密文预测结果 $\boldsymbol{E}_{PK_U}(WP+B)$，用自己持有的私钥 SK_U 解密：

$$D_{SK_U}[\boldsymbol{E}_{PK_U}(WP+B)]=WP+B \tag{15}$$

完成解密算法后，最终用户可得明文状态下的预测结果 $WP+B$.

4 方案分析

本小节对所提预测方案进行严格的安全性分析,并与现有方案进行性能比较.

4.1 安全性分析

本文中用户数据的隐私性和模型参数的隐私性都是基于同态加密算法 CKKS 实现的.只要用户和模型提供者的私钥是保密的,两个云服务器均无法获取用户和模型提供者的隐私数据.

首先我们将具体分析所提方案中每个阶段的算法是如何保证实现数据隐私性的.

(1)在初始化算法 *Initialization* 中,可信第三方为云服务器 1、云服务器 2 和用户分别生成各自的公私密钥对(PK_P,SK_P)、(PK_C,SK_C)和(PK_U,SK_U).之后用户会选择随机数 r_X,并且用自己的公钥 PK_U对明文数据采用 CKKS 算法进行加密,然后将加密后的密文数据 $\boldsymbol{E}_{PK_C}(r_X X)$以及 $\boldsymbol{E}_{PK_U}(r_X^{-1})$发给云服务器 1.由于私钥 SK_U和 SK_C由用户和云服务器 2 分别保管,对云服务器 1 来说均是保密的,可以保证用户待预测数据的隐私性.用户将密文数据传输给云服务器 1 后,可离线等待最终预测结果.

(2)在初始化算法 *Initialization* 中,模型拥有者选择随机数r_C,r_W,并用云服务器 2 的公钥 PK_C加密卷积核参数 $\boldsymbol{E}_{PK_C}(r_C C)$.之后,基于用户的公钥 PK_U加密全连接层的权重和偏置矩阵得到 $\boldsymbol{E}_{PK_U}(r_W W)$,$\boldsymbol{E}_{PK_U}(r_C r_W B)$,并用云服务器 1 的公钥 PK_P加密随机数得到 $\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1})$.所有数据发送给云服务器 1 之后,云服务器 1 仅能用自己的私钥 SK_P解密得到 $r_C^{-1}r_W^{-1}$.由于随机数 r_C,r_W均是未知的,因此云服务器 1 无法获取到模型拥有者的模型参数 W,B.

(3)后续的卷积层算法 *Convolution* 等均是基于密文下的计算,在云服务器 1 和云服务器 2 之间交互进行.在卷积层算法 *Convolution* 中,云服务器 1 会将 $\boldsymbol{E}_{PK_C}[r_X r_C(X * C)]$发送给服务器 2.由于最终传输结果是基于云服务器 2 的公钥 PK_C加密得到的,所以云服务器 2 可以解密得到 $r_X r_C(X * C)$,而这只是整体预测过程中的中间结果.由于随机数 r_X,r_C分别由用户和模型拥有者选择,对攻击者来说均是保密的,因此云服务器 2 无法获得具体的 $X * C$ 的值,同时无法得到 X 和 C 中的具体参数值.

(4)云服务器 1 运行全连接层算法 *Fully-Connection* 之后,将运行之后结果 $\boldsymbol{E}_{PK_U}(WP+B)$返回给用户,在这一步骤中最终预测结果 $WP+B$ 被用户的公钥 PK_U加密保护,故除用户外,攻击者均无法获得最后的预测结果.

下面给出所提预测方案严格的安全性证明.

定理 1. 假设 CKKS 方案是(t,ε)安全的,那么本文中用户的加密方案也是(t,ε)安全的.

证明. 假设存在敌手 A 能够在时间 t 内以不可忽略的优势 ε 攻破用户加密方案,那么我们可以构建一个模拟者 S 在时间内 t 以不可忽略的优势攻破 CKKS 方案.证明通过模拟者与敌手进行以下交互.

初始化阶段:输入安全参数 λ,产生系统参数 $params=\{p,q_0,N,Q,l,s,a,b,e,r\}$,生成公私钥对$(PK_C,SK_C)$,$(PK_U,SK_U)$,并将公钥 PK_C,PK_U发送给 A 和模拟者 S.给定模拟者 S 一组 CKKS 明文密文对(m_0,m_1,C_τ),其中 $C_\tau=Enc_{PK_C}(m_\tau)$,$\tau\in\{0,1\}$,模拟者 S 的目标是区分 C_τ是 m_0还是 m_1的密文.

询问阶段:敌手 A 发送明文 m 给模拟者 S.S 选择随机数 r_X,计算密文 $C=(C_0,C_1)$,其中 $C_0=\boldsymbol{E}_{PK_C}(r_X m)$,$C_1=\boldsymbol{E}_{PK_U}(r_X^{-1})$,随后将密文 C 发送给敌手 A.

挑战阶段:模拟者 S 选择挑随机数 r'_X,并计算密文 $C_{\tau'}=(C_{\tau'_0},C_{\tau'_1})$,其中

$$C_{\tau'_0}=Enc_{PK_C}(m_\tau)\otimes Enc_{PK_C}(r'_X)=C_\tau Enc_{PK_C}(r'_X),\quad C_{\tau'_1}=Enc_{PK_U}(r'_X) \tag{16}$$

并将结果$(m_0,m_1,C_{\tau'})$发送给敌手 A.

猜测阶段:A 输出猜测值 $\tau'\in\{0,1\}$,则 S 输出 $\tau=\tau'$区分 C_τ是 m_0还是 m_1的密文.

假设敌手 A 能以不可忽略的优势 ε 区分所提预测方案的密文,则 S 能以不可忽略的优势 ε 区分 CKKS 方案的密文.由于 CKKS 方案已在文献[22]中被证明达到选择明文攻击下的密文不可区分性(IND-CPA)安全,则所提预测方案中用户加密方案也是 IND-CPA 安全的.

定理 2. 假设 CKKS 方案是(t,ε)安全的,那么本文中模型拥有者的加密方案也是(t,ε)安全的.

证明. 假设存在敌手 A 能够在时间 t 内以不可忽略的优势 ε 攻破模型拥有者加密方案,那么我们可以构建一个模拟者 S 在时间 t 内以不可忽略的优势攻破 CKKS 方案.证明通过模拟者与敌手进行以下交互.

初始化阶段:输入安全参数 λ,产生系统参数 $params=\{p,q_0,N,Q,l,s,a,b,e,r\}$,生成公私钥对$(PK_C,SK_C)$,$(PK_U,SK_U)$,$(PK_P,SK_P)$,并将公钥 PK_C,PK_U,PK_P发送给 A 和模拟者 S.给定模拟者 S 一组 CKKS 明文密文对(m_0,m_1,C_τ),其中 $C_\tau=$

$Enc_{PK_C}(m_\tau)$，$\tau\in\{0,1\}$，模拟者 S 的目标是区分 C_τ 是 m_0 还是 m_1 的密文.

询问阶段：敌手 A 发送明文 m 给模拟者 S. S 选择随机数 r_C，r_W，模型参数 W，B，计算密文 $C=(C_0,C_1,C_2,C_3)$，其中

$$C_0=\boldsymbol{E}_{PK_C}(r_C m),\ C_1=\boldsymbol{E}_{PK_U}(r_C W),$$
$$C_2=\boldsymbol{E}_{PK_U}(r_C r_W B),\ C_3=\boldsymbol{E}_{PK_P}(r_C^{-1}r_W^{-1}) \quad (17)$$

随后将 C 发送给敌手 A.

挑战阶段：模拟者 S 选择随机数 r'_C，r'_W，任意模型参数 W'，B'，计算密文 $C_{\tau'}=(C_{\tau'_0},C_{\tau'_1},C_{\tau'_2},C_{\tau'_3})$，其中

$$C_{\tau'_0}=Enc_{PK_C}(m_\tau)\otimes Enc_{PK_C}(r'_C)=C_\tau Enc_{PK_C}(r'_C),$$
$$C_{\tau'_1}=Enc_{PK_U}(r'_C W'),$$
$$C_{\tau'_2}=Enc_{PK_U}(r'_C r'_W B'),$$
$$C_{\tau'_3}=\boldsymbol{E}_{PK_P}[(r'_C)^{-1}(r'_W)^{-1}] \quad (18)$$

随后将 $(m_0,m_1,C_{\tau'})$ 发送给敌手 A.

猜测阶段：敌手 A 输出猜测值 $\tau'\in\{0,1\}$，则 S 输出 $\tau=\tau'$ 区分 $C_{\tau'}$ 是 m_0 还是 m_1 的密文.

假设敌手 A 能以不可忽略的优势 ε 区分所提预测方案的密文，则 S 能以不可忽略的优势 ε 区分 CKKS 方案的密文. 由于 CKKS 方案已在文献[22]中证明达到 IND-CPA 安全，则所提方案中模型拥有者加密方案也是 IND-CPA 安全的.

由定理 1 和定理 2，所提预测方案中用户和模型拥有者的图像数据和模型参数都可以实现隐私保护，在预测过程中均不会泄露给攻击者.

4.2　方案对比分析

本小节将所提方案与现有同类方案进行比较，将从用户数据隐私性、模型隐私性、方案准确性以及用户是否需要交互等方面进行比较分析，如表 1 所示. 在表 1 中，HE 代表同态加密(Homomorphic Encryption)，GC 代表混淆电路(Garbled Circuits)，SS 代表秘密共享(Secret Share).

表 1　现有方案功能对比

方案	方案构成	数据隐私性	模型隐私性	精确性	非交互
CryptoNets[9]	HE	√	×	×	√
XONN[13]	GC,SS	√	×	×	√
SecureML[15]	HE,GC,SS	×	×	×	×
MiniONN[11]	HE,SS	×	×	√	×
NN[28]	HE,GC,SS	×	×	√	√
CNNP[16]	HE,GC,SS	×	×	√	√
本方案	HE	√	√	√	√

如表 1 所示，简单地将用户隐私数据或者模型参数随机拆分成两部分是无法保证隐私性的. 以文献[28]为例，神经网络模型参数 W 被随机拆分为 W_1，W_2，满足 $W_1=W-r$，$W_2=r$，其中 r 为随机数，拆分之后将 W_1，W_2 分别发送给两个云服务器. 虽然方案假设两个服务器是不能共谋的，但当恶意攻击者同时截获到 W_1，W_2 时，即可通过计算 W_1+W_2 轻易恢复出模型参数，用户隐私数据同理也可恢复. 因此，SecureML[15] 方案、MiniONN[11]、NN[28] 和 CNNP[16] 方案均有可能泄露用户隐私数据和模型参数. 所提方案基于 CKKS 同态加密算法保护隐私数据，其安全性是基于 Lyubashevsky 等人提出的 RLWE 困难问题. 当私钥 SK_U 不公开时，云服务器或恶意攻击者均无法获取到用户和模型提供者的隐私数据，即同时实现用户数据隐私性以及模型隐私性，具体分析见 4.1 节. CryptoNets[9] 方案同样对原始数据采取了同态加密，实现了用户数据隐私性，但模型参数对服务器是公开的. XONN[13] 方案使用混淆电路实现了用户数据隐私性，但模型参数对服务器也是公开的.

接下来是模型精确性的比较，具体来说，CryptoNets[9] 方案设计了一个可以在加密数据上运行的神经网络框架，并使用 ReLU 激活函数的多项式近似进行预测，但加密方案不支持浮点数，具体位置使用固定精度的实数，通过适当的比例将其转换为整数，故准确性上存在一定缺陷. 此外，XONN[13] 方案使用“XNOR替换”矩阵乘法，但该方案仅适用于二元神经网络，参数限制为二进制，影响预测精度. SecureML[15] 方案关注简单的神经网络，以及线性/逻辑回归，不适用于更复杂的卷积神经网络，并且方案中用 x^2 近似代替 ReLU 激活函数. 故上述三种方案均在模型精准度上有不同程度的折损. MiniONN[11]、NN[28]、CNNP[16] 方案无需近似操作，不降低模型精准度. 虽然所提方案基于密文数据进行预测，但由于 CKKS 同态加密方案支持浮点数运算，故在模型精准度方面并无过多损失.

最后是方案中用户是否可离线的对比分析，如果用户在上传完数据之后离线等待最终结果，可大幅降低计算和通信代价. 在现有文献中，SecureML[15] 方案需要用户在预测阶段一直在线. 在 MiniONN[11] 方案中，用户需要和服务器协作完成网络前向计算过程，这会加重用户的计算和通信负担，并且用户需要与服务器多次交互. 其他文献均满足离线等待的要求.

对比现有文献，我们基于同态加密构造了完整的卷积神经网络预测方案，能够实现外包方案中的

用户数据隐私性以及模型参数的隐私性，并且所提方案对于 ReLU 激活函数以及池化操作都不需要近似操作，不降低模型的精准度. 并且在所提方案中，用户可以在加密并上传原始数据后离线等待预测结果，大幅降低了用户的计算和通信代价. 具体关于预测准确性的实验仿真效果可参考第 5 节.

5 实验仿真

本文所提隐私保护的卷积神经网络预测方案基于 Win10 系统实现，用户和模型提供者使用处理器为 Intel(R) Core(TM) i5-1135G7 CPU@2.40 GHz、内存为 16 GB 的计算机模拟；云服务器使用处理器为 AMD Rome 2.6 GHz，内存为 64 GB 的服务器模拟. 方案代码部署在 PyCharm Community Edition 2020.3 x64 软件上，编程语言为 Python 3.6. 我们使用微软的 SEAL 库中调用 CKKS 相关函数完成 CKKS 同态加密算法.

实验使用 MNIST 手写数字数据集和 CIFAR-10 数据集测试密文预测的准确率. 我们对整体数据集进行划分，其中 60000 条被用作训练数据，10000 条被用作测试数据. 实验基于两种数据不断提高模型效果并优化模型参数，最终得到一个具有较高准确率的预测模型. 我们将基于所提隐私保护的卷积神经网络预测方案，利用初始化算法 *Initialization* 中的加密方案对用户数据和模型参数进行预处理，即对用户提供的隐私数据以及卷积神经网络中的卷积核参数 C，全连接层参数 W，B 进行预处理，接着进行密文下的预测操作.

在实验仿真阶段，我们参考了图 1 所示模型结构，这是一个高效且非常适用于 MNIST 手写体字符识别的卷积神经网络模型，即模型拥有者提供的卷积神经网络模型中有包含两层卷积层，两层池化层以及两层全连接层. 为模拟实验效果，卷积核大小的设置上存在些许差异，具体的参数设置如表 2. 例如 5×5×4 代表此原模型的卷积核大小为 5×5，且深度为 4. 之后会经过两层 ReLU 激活运算以及平均池化层运算，此外在模型的预测部分还会设置两层的全连接层.

表 2 不同模型参数下的预测准确性

数据集	卷积核尺寸	原有CNN模型预测准确率/%	文献[16]预测准确率/%	所提方案预测准确率/%
MNIST	5×5×4	92.90	92.36	89.03
	5×5×16	98.29	96.74	92.87
	5×5×64	98.79	96.52	92.64
	7×7×2	83.42	83.35	82.50
	7×7×4	93.66	93.09	90.57
	7×7×16	98.02	96.91	93.32
CIFAR-10	5×5×6	87.37	87.21	85.53
	5×5×12	89.52	89.13	86.65

表 2 列出了不同模型参数下不同方案的预测准确率. 表中原模型预测准确性是指在训练集大小 batch 设置为 500 的条件下，对网络训练模型训练 20 次之后得到的预测准确性. 在本节实验中，选取 MNIST 手写数字数据集和 CIFAR-10 数据集中的 200 至 300 张图像，作为用户想要进行预测的原始图像，然后测试预测准确率，即预测正确的图像数量占总预测图像数据集的比例.

从表 2 可看出，所提方案的预测准确率最高可达 93%，与原有 CNN 模型和文献[16]中方法相比仅有小幅折损. 折损原因可能为当卷积核深度较深时，导致的计算量较大，并且模型参数中存在一些接近零的参数值. 我们采用的 CKKS 同态加密是一种近似同态加密算法，在处理计算精度较高(小数点位数较多)的数据时会产生一定误差. 文献[16]虽然在精确率上略高于所提方案，但不能实现用户数据和模型训练结果的隐私性，具体对比分析见表 1.

此外，我们还对用户和模型拥有者各自的计算开销进行了模拟测试，从具体方案中可知，用户和模型拥有者本地计算开销只涉及加解密. 我们在 SEAL 库调用 CKKS 相关函数时首先进行了参数设置，选取模多项式的次数为 $N=8192$，以及 $coeff\_mod\_bit\_sizes=[60,40,40,60]$，具体为设置了大小分别为 60，40，40，60 比特的四个素数，其中第一个素数代表 2.1 节 $KeyGen(\lambda)$ 算法中的 q_0 为 60 比特，最后一个素数代表用于生成 evk 的模数 P 同样为 60 比特，中间的两个素数代表 p_1，p_2 均为 40 比特，对应 CKKS 中第一层的模数为 $q_0 \cdot p_1$，第二层的模数为 $q_0 \cdot p_1 \cdot p_2$.

在表 3 和表 4 中，我们分别列出了用户和模型拥有者对图像数据和模型参数进行 CKKS 同态加密计算的具体耗时. 在表 3 中，用户输入图像大小中“28×28×1”中的 28×28 指的是固定大小为 28×28 像素的 MINST 数据集中的输入图像，1 代表需要进行预测的图像数量. 在表 4 中，模型参数大小中“5×5×16 的卷积核”代表卷积核大小为 5×5，且每一层卷积层中卷积核的深度为 16. 由表 3 和表 4 可以看出，所提方案中用户以及模型拥有者的本地计算耗时都较小，证明我们的方案虽然采用的是同态加密，不是对数据进行简单的随机拆分，同样能保证用户和模型拥有者的本地效率.

表 3　用户加密数据所需耗时

数据集	用户输入图像大小	所需耗时/s
MNIST	28×28×1	0.3313
	28×28×5	0.3469
	28×28×10	0.3615
CIFAR-10	32×32×3	0.3242
	32×32×6	0.3409

表 4　模型拥有者加密模型参数所需耗时

模型参数大小	所需耗时/s
5×5×16 的卷积核＋2 个全连接层	0.3927
5×5×64 的卷积核＋2 个全连接层	0.6271
7×7×16 的卷积核＋2 个全连接层	0.4263

此外，在表 5 中我们模拟并对比了云服务器在 MNIST 数据集上运行卷积神经网络预测时的计算耗时，并详细拆分并展示各个阶段计算耗时. 两个方案均存在两个服务器，因此统计的是两个服务器的计算总耗时. 由于所提方案无需在离线计算乘法三元组，故离线阶段中无需额外计算. 在线预测阶段中，所提方案在卷积层以及全连接层均基于密文数据进行同态计算，例如 $\boldsymbol{E}_{PK_C}(r_X X)$，$\boldsymbol{E}_{PK_C}(r_C C)$，故相较文献[16]来说所需耗时较高. 对于池化层，云服务器基于 $r_X r_C (X * C)'$ 进行平均池化计算，$(X * C)'$ 被随机数 $r_X r_C$ 盲化，但不是同态加密后的密文，即池化层中无需同态计算，故所提方案与文献[16]中的计算耗时相同. 与文献[16]相比，所提方案由于使用同态加密方案，服务器计算耗时相对较高，但是方案能同时实现数据和模型隐私性，以及预测结果的隐私性. 而且，由表 3 和表 4 可以看出，所提方案中用户以及模型拥有者的本地计算耗时都较小.

表 5　云服务器在各阶段所需耗时比较

阶段	运算	CNNP[16]/s	所提方案/s
离线阶段	乘法三元组	0.471	/
在线阶段	卷积层 1	0.055	29.245
	ReLU 激活层 1	0.425	3.278
	平均池化层 1	0.001	0.001
	卷积层 2	0.098	27.694
	ReLU 激活层 2	0.317	3.645
	平均池化层 2	0.001	0.001
	全连接层 1	0.006	2.634
	全连接层 2	0.001	1.283
总耗时		3.835	67.781

结合上节方案分析以及本节实验结果可知，本文所提方案采用同态加密保证了数据隐私性和模型隐私性，且所有复杂运算均由两个云服务器承担，用户端和模型提供者的计算开销很低，满足了终端用户计算的高效性. 实验结果表明，所提预测方案准确率最高可达 93%，与明文数据预测准确率相差较小. 此外，用户在上传密文数据后无需与云服务器进行交互，可离线等待预测结果，实现了非交互性. 综上，所提方案提高了隐私保护卷积神经网络预测的安全性和预测精度，降低了用户的计算和通信代价.

6　总结与展望

本文方案采用同态加密对用户的输入数据以及模型拥有者提供的模型参数进行隐私保护，最终的预测结果对服务器也是保密的. 相较于秘密共享方法，所提方案能够抵抗恶意攻击者截获数据后恢复用户原始数据和模型参数. 用户和模型提供者上传密文数据后无需参与预测，实现了非交互性. 实验表明，所提方案实现了预测准确性，与明文数据相比仅有小部分折损，是一个兼顾数据隐私性以及预测准确性的方案.

但是，由于同态加密的限制，所提方案并不适用于较深的卷积神经网络，如何实现深度隐私保护卷积神经网络预测是未来我们进一步研究的方向. 另外，现有的神经网络预测研究均未实现对预测结果的验证，无法检测出恶意服务器返回的错误预测结果，所以针对预测结果正确性的检验是需要进一步研究的问题.

参　考　文　献

[1] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Cambridge: MIT Press, 2016

[2] Zhao L, Wang Q, Zou Q, et al. Privacy-preserving collaborative deep learning with unreliable participants. IEEE Transactions on Information Forensics and Security, 2020, 15: 1486-1500

[3] Zou Q, Jiang H, Dai Q, et al. Robust lane detection from continuous driving scenes using deep neural networks. IEEE Transactions on Vehicular Technology, 2020, 69: 41-54

[4] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84-90

[5] Ribeiro M, Grolinger K, Capretz M. MLaaS: Machine learning as a service//Proceedings of the IEEE 14th International Conference on Machine Learning and Applications. Miami, USA, 2015: 896-902

[6] Haykin S, Lippmann R. Neural networks, a comprehensive foundation. International Journal of Neural Systems, 1994, 5(4): 363-364

[7] Botta A, De Donato W, Persico V, et al. Integration of cloud computing and Internet of Things: A survey. Future Generation Computer Systems, 2016, 56: 684-700

[8] Varghese B, Buyya R. Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 2018, 79: 849-861

[9] Gilad-Bachrach R, Dowlin N, Laine K, et al. CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016: 201-210

[10] Wang W, Chen Z, Huang X. Accelerating leveled fully homomorphic encryption using GPU//Proceedings of the 2014 IEEE International Symposium on Circuits and Systems. 2014: 2800-2803

[11] Liu J, Juuti M, Lu Y, et al. Oblivious neural network predictions via MiniONN transformations//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York, USA, 2017: 619-631

[12] Shamir A. How to share a secret. Communications of the ACM, 1979, 22(11): 612-613

[13] Riazi M, Samragh M, Chen H, et al. XONN: XNOR-based oblivious deep neural network inference//Proceedings of the 28th USENIX Security Symposium. Santa Clara, USA, 2019: 1501-1518

[14] Yao A. Protocols for secure computations//Proceedings of the 23rd Annual Symposium on Foundations of Computer Science. Chicago, USA, 1982: 160-164

[15] Mohassel P, Zhang Y. SecureML: A system for scalable privacy-preserving machine learning//Proceedings of the 2017 IEEE Symposium on Security and Privacy. San Jose, USA, 2017: 19-38

[16] Li M, Chow S, Hu S, et al. Optimizing privacy-preserving outsourced convolutional neural network predictions. IEEE Transactions on Dependable and Secure Computing, 2022, 19(3): 1592-1604

[17] Zhao C, Zhao S, Zhao M, et al. Secure multi-party computation: Theory, practice and applications. Information Sciences, 2019, 476: 357-372

[18] Hohenberger S, Lysyanskaya A. How to securely outsource cryptographic computations//Proceedings of the Theory of Cryptography Conference. Berlin, Germany, 2005: 264-282

[19] Gennaro R, Gentry C, Parno B. Non-interactive verifiable computing: Outsourcing computation to untrusted workers//Proceedings of the Annual Cryptology Conference. Berlin, Germany, 2010: 465-482

[20] Demmler D, Schneider T, Zohner M. ABY—A framework for efficient mixed-protocol secure two-party computation//Proceedings of the NDSS. San Diego, USA, 2015: 8-11

[21] Wei K, et al. Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469

[22] Cheon J H, Kim A, Kim M, et al. Homomorphic encryption for arithmetic of approximate numbers//Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security. Cham, Switzerland, 2017: 409-437

[23] Wei Q, Li Q, Zhou Z, et al. Privacy-preserving two-parties logistic regression on vertically partitioned data using asynchronous gradient sharing. Peer-to-Peer Networking and Applications, 2021, 14: 1379-1387

[24] Lehmkuhl R, Mishra P, Srinivasan A, Popa R. Muse: Secure inference resilient to malicious clients//Proceedings of the 30th USENIX Security Symposium. Boston, USA, 2021: 2201-2218

[25] Chandran N, Gupta D, Obbattu S. SIMC: ML inference secure against malicious clients at semi-honest cost//Proceedings of the 30th USENIX Security Symposium. Boston, USA, 2021: 1361-1378

[26] Brakerski Z, Gentry C, Vaikuntanathan V. (Leveled) fully homomorphic encryption without bootstrapping. ACM Transactions on Computation Theory, 2014, 6(3): 1-36

[27] Lyubashevsky V, Peikert C, Regev O. On ideal lattices and learning with errors over rings//Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin, Germany, 2010: 1-23

[28] Ma X, Chen X, Zhang X. Non-interactive privacy-preserving neural network prediction. Information Sciences, 2019, 481: 507-519

REN Yan-Li, Ph. D., professor. Her research interests include public key cryptography, blockchain security and AI security.

YU Ling-Zan, M. S. His research interests include outsourcing computation and AI security.

HE Gang, Ph. D. candidate. His research interests include outsourcing computation and AI security.

ZHANG Xin-Peng, Ph. D., professor. His research interests include multimedia security and AI security.

GUO Zheng, Ph. D., senior engineer. His research interests include security analysis and detection of cryptographic hardware system.

Background

The topic studied in this paper belongs to the privacy-preserving outsourcing schemes in the field of machine learning. At present, there exist some related outsourcing schemes for linear regression, matrix factorization, neural network prediction and so on. In recent years, along with the development of research, scholars began to gradually focus on more complex algorithms of machine learning. Especially, convolutional neural network (CNN) is a popular architecture in the field of machine learning, which is demonstrated excellent performance in medical image analysis, image and audio recognition and classification.

We propose a CNN prediction scheme that can realize data prediction on the encrypted data, and preserve privacy in the outsourced setting based on homomorphic encryption, which is able to preserve the user's privacy and keep the model private from any hosting server. Moreover, after uploading encrypted data, the users and the model owner can wait for the prediction results offline, and there is no extra interaction between user and the cloud server during the prediction process. Also, the proposed scheme supports floating-point arithmetic and does not require approximate operation for nonlinear functions, which improves the accuracy of model prediction. The experimental results show that our proposed scheme improves security of the CNN prediction, greatly reduces the user's communication cost, and can achieve prediction accuracy to 93% at most, which is almost same to that of plaintext data prediction. Therefore, the proposed scheme improves the security and prediction accuracy of CNN, and reduces the computing and communication costs of users compare with the previous ones.

The work described in this paper was supported by the Key Project of National Natural Science Foundation of China (No. U1936214), Research on Virtual User Shaping and Information Hiding in Social Networks.

The purpose of this project is to utilize resources such as massive users in social networks and social behaviors to realize the shaping of virtual users, improve the capacity and security of steganography, and realize transfer of secret messages privately based on the consensus reached by both parties.

Our research group is affiliated to the "Shanghai Advanced Data Communication Research Institute" and the "Sino-UK Digital Urban Innovation" International Joint Research Center established by the Shanghai IV peak characteristic disciplines. We have participated in more than 50 key projects in recent five years, including the projects of the National Natural Science Foundation of China and the provincial and ministerial-level projects. More than 200 papers have been published on high-level journals in the field of cryptography, and we have obtained some achievements in the field of secure outsourced computation and processing of big data. Some members in our group have make a theoretical research of covert communication in social networks, and have proposed innovative works such as covert communication detection based on deep learning and privacy preserving outsourcing algorithm of statistical features.

Overall, our proposed scheme in this paper realizes data prediction on the encrypted data and data privacy preservation in the outsourced setting. Our scheme has great significance for calculation based on ciphertext and copyright protection in social networks.