从第 3.2.2 节加密数据比较协议可知,$z=x+r \bmod N$,其中,x 为 l 位整数,r 为 $\lambda+l$ 位整数.

因为本文的 $\lambda+l+1<\log_2^N$,所以有 $z=x+r$.在统计学中,z 和 $\tilde{z}$ 的区分度不大,因此有 $(SK_P,[\![\tilde{z}]\!])\equiv_s(SK_P,[\![z]\!])$;又 z 和 $\tilde{z}$ 的分布是独立于 $\tilde{t}'$ 和 $\tilde{r}_{l+1}$ 的,所以有:

$$(PK_{QR},SK_P,l,[\![\tilde{z}]\!];coins;[\tilde{t}'],[\tilde{r}_{l+1}])\equiv_s(PK_{QR},SK_P,l,[\![z]\!];coins;[\tilde{t}'],[\tilde{r}_{l+1}]).$$

因为 QR 是语义安全的,因此有:

$$(PK_{QR},SK_P,l,[\![z]\!];coins;[\tilde{t}'],[\tilde{r}_{l+1}])\equiv_c(PK_{QR},SK_P,l,[\![z]\!];coins;[t'],[\tilde{r}_{l+1}]),$$

$$S_B(PK_{QR},SK_P,l)\equiv_c V_B([\![a]\!],[\![b]\!],l,SK_{QR},PK_{QR},SK_P,PK_P).$$

本文使用模块化顺序组合来实现协议执行过程中的安全性衔接,使用改进的 DGK 协议取代了理想计算得到的[t'],使用定理 1 证明了其在半诚实模型下的安全性.综上所述,本文的比较协议在半诚实模型下是安全的.

3.3 点积协议

本文通过欧式距离计算测试样本与训练样本之间的距离.为了保证 PP-kNN 分类器中距离计算的安全性,本文设计了点积协议.

协议处理的是密文数据,所有操作皆在密文空间进行.它由 A 和 B 两方参与:A 表示客户端,输入测试样本,记作 x;B 表示服务器,输入训练样本,记作 y.协议 2 是点积协议的具体描述.

协议 2. 点积协议.

输入方 A:$x=(x_1,\ldots,x_d)\in\mathbb{Z}^d$,公钥 PK_{FHE}.

输入方 B:$y=(y_1,\ldots,y_d)\in\mathbb{Z}^d$,私钥 SK_{FHE}.

输出方 A:$[\![v]\!]$.

1: B 对向量 $y_1,\ldots,y_d$ 进行加密,然后将加密后的数据$[\![y]\!]$发送给 A

2: A 对向量 $x_1,\ldots,x_d$ 进行加密,得到加密数据$[\![x]\!]$

3: A 计算$[\![z]\!]\leftarrow[\![y]\!]\cdot[\![x]\!]^{-1} \bmod N^2$▷$z=y-x$

4: A 计算 $[\![v]\!]\leftarrow\prod_i[\![z]\!]^{[\![z]\!]}$,然后输出加密的 $[\![v]\!]$▷$v=\sum z_i^2$

协议 2 在半诚实模型下的安全性:在本协议中,B 未接收任何信息,仅提供了数据及用于加密的随机数,则模拟器 S_B 可表示为 $S_B(y,SK_{FHE})=(y,SK_{FHE};coins)=V_B(x,y,SK_{FHE},PK_{FHE})$,其中,$coins$ 为 B 方生成的随机数.A 的元组表示为 $V_A(x,PK_{FHE};r^A;[\![z_1]\!],\ldots,[\![z_n]\!])$,模拟器 S_A 的构造过程如下.

Step 1. 生成 n 个 FHE 加密的 0:$c_n,\ldots,c_0$.

Step 2. 生成随机数 $\widetilde{coins}$.

Step 3. 输出 $(x,PK_{FHE};\widetilde{coins};(c_n,...,c_0))$.

$coins$ 和 $\widetilde{coins}$ 的分布相同,因此,

$$\{(x,PK_{FHE};\widetilde{coins};(c_n,..,c_0));[\![\langle z,z\rangle]\!]\}\equiv_c\{(x,PK_{FHE};coins;(c_n,..,c_0));[\![\langle z,z\rangle]\!]\}.$$

由于 FHE 加密方案的语义安全,所以有:

$$\{(x,PK_{FHE};coins;(c_n,\ldots,c_0));[\![\langle z,z\rangle]\!]\}\equiv_c\{(x,PK_{FHE};coins;[\![z_1]\!],\ldots,[\![z_n]\!]);[\![v]\!]\}.$$

函数 f 满足 $f(x,y,SK_{FHE},PK_{FHE})=([\![\langle z,z\rangle]\!],\varnothing)$时,有:

$$\{S_A(x,PK_{FHE},[\![v]\!]);f(x,y,SK_{FHE},PK_{FHE})\}\equiv_c\{V_A(x,y,SK_{FHE},PK_{FHE});Output(x,y,SK_{FHE},PK_{FHE})\}.$$

本协议中,B 将加密好的数据$[\![y]\!]$发送给 A,A 没有私钥无法对其解密,保证了数据的安全性.加密方案是加法与乘法同态的,因此密文数据计算过程是安全且同态的.上述过程通过定义 1 证明了其在半诚实模型下的安全性.综上所述,本文的点积协议在半诚实模型下是安全的.

3.4 加密方案转换协议

PP-kNN 分类器由点积协议、比较协议通过顺序组合的方式构造而成,点积协议的输入与输出数据是 FHE

进行加密的密文数据,比较协议的输入数据是 Paillier 进行加密的密文数据.为了保证 PP-kNN 分类器的点积协议和比较协议可以进行模块化顺序组合,本文设计了加密方案转换协议,实现了从一种加密方案到另一种加密方案的转换.协议 3 是加密方案转换协议的描述.

协议 3. 加密方案转换协议.

输入方 A:$[\![c]\!]_1$,公钥 PK_1,PK_2.

输入方 B:密钥 SK_1,SK_2.

输出方 A:$[\![c]\!]_2$.

1: A 均匀随机选择一个数 $r \leftarrow M$

2: A 计算$[\![c']\!]_1 \leftarrow [\![c]\!]_1 \cdot [\![r]\!]_1$ 将$[\![c']\!]_1$ 发送给 B ▷Blind c

3: B 解密后得到 c',用 E_2 重新加密

4: B 将重新加密后的$[\![c']\!]_2$ 发送给 A

5: A 去除噪音$[\![r]\!]_2$ 得到 $[\![c]\!]_2 = [\![c']\!]_2 \cdot [\![r]\!]_2^{-1}$ ▷Blind c

6: A 输出$[\![c]\!]_2$

加密方案转换协议中,E_1,E_2 是两种不同的加密方案,A 通过 E_1 对随机数 r 进行加密,为$[\![c]\!]_1$ 添加噪音$[\![r]\!]_1$,将处理后的$[\![c']\!]_1$ 发送给 B,B 解密后无法获取 c 的真实值,保证了该值在解密-再加密过程中数据的安全性,其中,r 是 A,B 共享,M 表示 E_1 的信息空间.B 通过 E_2 对 c'重新加密,将$[\![c']\!]_2$ 发送给 A,A 去除噪音,得到 E_2 加密的真实值$[\![c]\!]_2$,实现了从加密方案 E_1 到 E_2 的转换.整个加密方案转换过程中,密文噪音并未增加,且中间解密时数据真实值亦无法获取,因此,密文数据在加密转换协议中是安全的.根据本文的应用需求,设置加密方案 E_1 是 FHE,加密方案 E_2 是 Paillier,通过协议 3 实现了从 FHE 向 Paillier 加密方案的转换.

4 PP-kNN 分类器构造及分类过程

4.1 浮点数据处理

本文的加密方案都是对整型数据进行加密,而原始数据中部分为浮点型数据,因此对数据进行处理前,需将浮点型数据转换为整型数据,处理方法是用一个足够大的常量 K 乘以浮点数.下面详细描述浮点数的处理过程:

首先,将浮点数据表示为 IEEE 754 双精度浮点数格式,即 $V=(-1)^s \cdot M \cdot 2^{E-1023}$,其中,$S$ 为符号位占 1 比特,M 为尾数,二进制表示为$(M)_2=1.d$,占 52 比特位,$1 \leqslant M \leqslant 2$,$M$ 可重新表示为

$$M = \frac{M_i'}{2^{52}},$$

其中, $M_i' \in \mathbb{N} \cap [2^{52}, 2^{53})$.

其次,寻找合适的常量 K,使得 K 满足:

$$K \cdot v_i \in \mathbb{N},$$

其中, $v_i = M_i' \cdot 2^{e_i - 52}$.

令 $e^*=\min_i e_i$,$\delta_i = e_i - e^* \geqslant 0$,则

$$v_i = M_i' \cdot 2^{\delta_i} \cdot 2^{e^* - 52}.$$

令 $K = 2^{52-e^*}$, 则

$$K \cdot v_i = M_i' \cdot 2^{\delta_i} \in \mathbb{N}.$$

因此,经过上述计算,可以得出 $K = 2^{52-e^*}$. 值得注意的是,在数据转换过程中,可能会出现空间溢出和精度损失问题,数据转换后可能会对基本操作的精准度产生影响,甚至影响分类结果.针对此类问题,本文给出如下说明.

- 首先,加密方案的明文空间大于 2^{52}, $K = 2^{52-e^*}$, 因此数据转换过程中不存在空间溢出和精度损失.

- 其次,kNN 分类器具有的基本操作只有加法、乘法和比较,因此对转换后的数据进行操作仍能得到相同的分类结果.
- 最后,在执行加、乘、比较操作时,为确保操作不会造成密文数据的精度丢失,需要设置计算和比较所需的比特位数.以比较协议为例:令 d 表示输入数据 x 的属性个数,则比较时所需的最大比特位数为 $l_{\max}=d+1+(52+\delta^*)$,其中,$\delta^*=\max\ \delta_i$,1 表示类别标识,$52+\delta^*$表示密文数据的二进制位数.因此,比较协议中比较位数必须大于 $l_{\max}$.此外,还要确保 $\log_2^N > l_{\max}+1+\lambda$,其中,$\lambda$为安全系数,$N$ 为 Paillier 加密方案明文空间的模量.为了获得高安全性,设 $\log_2^N \geqslant 1024$, 即$\lambda=100$.

本文训练集和测试集分别用 Y 和 X 表示,其中,x_i 表示第 i 个训练样本,则转换后的数据表示为

训练集 Y:$y_i=\lceil Ky_{ij} \rceil$, 测试集 X:$x_i=\lceil Kx_{ij} \rceil$,

其中,j 表示样本 x_i 的第 j 个特征值.

4.2 PP-kNN分类器的构造过程

本节利用第 3 节的协议来构造 PP-kNN 分类器,构造过程如下.

首先,将训练数据的类型由浮点数转换为整数,使用 FHE 对其加密;其次,通过协议 2 计算测试样本与所有训练样本的欧式距离,结果是 FHE 加密的密文数据.然后,通过协议 3 将结果转换为 Paillier 加密的密文数据,再通过 *get*MIN*n* 得到距离数组中的最小值.其思想为:先将数组中的值两两比较,得到两个中较小的值,将较大的赋值为 0,较小的值的下标记为两者下标中较小方的下标值,所有较小方组成新的数组.然后继续比较新的数组,直到数组个数为 1,该值即为最小值.其比较通过协议 1 实现,每次比较得到一个最小值,然后将最小值重新赋值为最大值.循环 k 次,得出 k 近邻样本.最后,使用第 4.2.1 节中介绍的方法来统计类别个数,得出分类结果.其中,将每个协议看作一个模块,通过模块化顺序组合进行模块衔接,构造 PP-kNN 分类器,使得客户端只能获知最后的分类结果,而不能知道测试样本与训练样本间的距离;使服务器无法获取客户端的输入 x(x 是测试样本的向量表示).

4.2.1 PP-kNN 分类器的近邻样本类别个数统计

计算测试样本 $x=(x_1,\ldots,x_d)$与训练样本 $y_i=(y_{i1},\ldots,y_{id})$的距离 $d(x_i,y_i)$,通过比较进行排序,获取前 k 个训练样本对应的分类标签.

设 $N=\{y_1,\ldots,y_k\}$表示包含 k 个训练样本的数据集,则 x 对应的分类 $c_x=\max_{v\in L}\sum_{y\in N} I(v=Class(c_y))$. 其中,$L=(c_1,\ldots,c_m)$是所有标记的集合,$I(\cdot)$是用来获取 k 个样本所属分类的函数,执行情况如下.

For y in N:

$Class(c_y)$得到样本 y 所属的分类

For v in L:依次与类别标签集 L 比较,若相同,则返回 1;否则,返回 0

If $v=Class(c_y)$:v+=1

Else v+=0

按上述步骤完成对 k 个样本所属分类个数的统计,类别个数最多的分类即为待测样本的预测分类 c_x.

4.2.2 PP-kNN 分类器的分类过程

kNN 分类器由服务器端与客户端两部分组成,其处理的数据是通过 FHE 加密的密文数据,PP-kNN 分类器,其实质是将 kNN 分类器针对明文数据的基本计算用第 3 节中的安全协议替换,使 kNN 分类器在分类过程中对密文数据进行操作,保证数据在分类过程中的安全性与同态性,最后得出分类结果.协议 4 是对 PP-kNN 分类协议的描述.

协议 4. PP-kNN 分类协议.

C 输入:测试样本 $x=(x_1,x_2,\ldots,x_d)\in\mathbb{Z}^d$,公钥 PK_P,PK_{FHE},私钥 SK_{QR}.

S 端输入:私钥 SK_P,SK_{FHE},公钥 PK_{QR},训练集 $D=(y_1,\ldots,y_m)$,标记 $L=(c_1,\ldots,c_m)$,近邻数 k.

C 输出:下标 i,c_i 是 k 个近邻样本中类别个数最多的类.

1: S 提供训练集 D,对训练集中的训练样本进行浮点数到整数的类型转换,然后通过 FHE 加密方案对训练

样本进行加密

2: S 将加密的$[\![D]\!]$和近邻数 k 发送给 C

3: 设样本容量为 m,**for** $1\leqslant i\leqslant m$,C 通过点积协议计算测试样本与训练样本的距离$[\![d(x_i,y_i)]\!]$

其中,$[\![d(x,y_i)]\!]=\sqrt{\sum_{j=1}^{d}([\![x_i]\!]-[\![y_{ij}]\!])^2}$,取结果的平方存放到数组 *dis_fhe* 中

4: C,S 通过加密方案转换协议将 FHE 转换为 Paillier 加密方案,存入 *dis_paillier* 中

5: C: **for** $0\leqslant M<k$:

(1) C 和 S 通过 *get*MIN*n* 获取数组 *dis_paillier* 中的最小值 min,并将其存入队列 queue 中

(2) 将 min 对应的值重新赋值为最大值

6: C 和 S 通过第 5 步得到 k 个近邻样本$[\![d_0]\!][\![d_1]\!]\ldots[\![d_k]\!]$

7: 统计 k 个近邻样本的类别数目,然后得到类别最多的类 c_i

4.2.3 安全性分析

由于点积协议、方案转换协议、比较协议在半诚实模型下是安全的,模块化线性组合在半诚实模型下也是安全的,因此,通过模块化线性组合对点积协议、方案转换协议、比较协议进行组合构造的 PP-kNN 分类器也是安全的.

- 首先,点积协议在半诚实模型下是安全的.在通过点积协议计算测试样本与训练样本间距离时,服务器仅发送加密后的密文数据给客户端,未接收来自客户端的任何输入,保证了客户端输入数据 x 的安全性.客户端不拥有私钥,无法解密来自服务器的输入数据;又因为 FHE 加密方案的加法、乘法同态性,保证计算过程的安全性,因此,距离计算时是安全的.
- 然后,加密方案转换协议在半诚实模型下是安全的.距离数据增加噪音干扰后发送给服务器,保证其解密再加密过程无法获知距离的真实值,因此不会推测出 x 的值;重新加密后,发送回客户端,客户端不具有 Paillier 私钥,无法解密.因此,加密方案转换过程是安全的.
- 最后,比较协议在半诚实模型下是安全的,因此调用比较协议获取最小值的运算过程中数据也是安全的,又因为模块化顺序组合在半诚实模型下是安全的.

综上所述,通过模块化顺序组合将点积协议、加密方案转换协议、比较协议进行组合构造的 PP-kNN 分类器也是安全的.

5 实 验

本文利用自定义加密数据对比较协议和加密方案转换协议进行性能评估,利用 4 种 UCI 数据集[30]对 PP-kNN 分类器的性能进行评估.

测试环境具体描述如下:CPU 为英特尔酷睿 i7 处理器(双核,3.4GHz);内存为 16GB.

实验在相同的网络下进行,因此,本文将一个数据包的往返时间记为 40ms 来模拟网络延迟.加密方案中的密钥长度为 1 024 位,统计安全参数λ=100.

5.1 比较协议的性能评估

首先,针对两种比特长度的加密数据,从客户端、服务器运行时间、交换数据量、交换次数这 4 个方面对比较协议进行了评估,实验结果见表 3.

Table 3 Evaluation of comparison protocol

表 3 比较协议评估

比较比特长度(bit)	客户端(ms)	服务器端(ms)	交换数据量(KB)	交换次数
64	13.15	15.47	27.91	11
128	21.16	23.04	54.91	11

然后,分别对 64 位、128 位、256 位、512 位、1 024 位比较比特长度的加密数据进行评估,如图 1 所示.

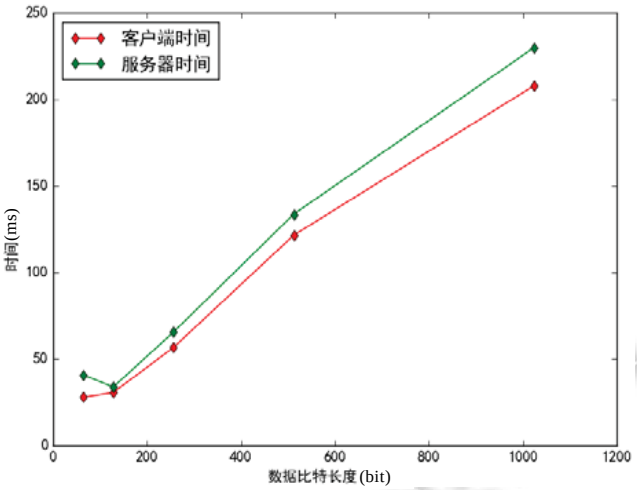

Fig.1 Performance of comparison protocol

图 1 比较协议性能

表 3 的实验结果表明,比较协议的运行时间与比较比特长度有关,比特长度越长,服务器和客户机运行时间越长,交换数据量越多.

5.2 kNN分类器性能评估

本节实验在 Iris、Wine、Zoo、Glass Identification 公共数据集上进行测试.这些数据集是 UCI 标准数据集[30],见表 4.

Table 4 Standard dataset

表 4 标准数据集

数据集	样本数	类别数	特征数	训练集样本数
Iris	150	3	4	69
Wine	178	3	13	66
Glass Identification	180	6	9	114
Zoo	101	7	16	65

测试数据与训练数据都是按一定比例随机进行抽取,各训练样本数见表 4,样本集中剩余样本作为测试集.本实验从客户端和服务端各自的计算、比较时间、交换数据总量及交换次数这几个方面进行评估,具体实验结果见表 5.

Table 5 Performance of PP-kNN classifier based on different test encrypted datasets

表 5 基于不同测试加密数据的 PP-kNN 分类器性能

数据集	k 值	比特位数	客户端(ms)			服务器(ms)			交换数据总量(MB)	交换次数
			距离计算	方案转换	k 近邻求解	距离计算	方案转换	k 近邻求解		
Iris	3	128	5 434.49	529.25	2 261.11	1 250.97	781.24	2 736.46	149.87	2 591
Wine	3	128	15 750.8	691.62	2 090.14	8 480.45	1 306.31	2 685.74	358.42	3 071
Glass	5	128	19 061.5	1 076.83	5 982.34	13 475.2	1 662.19	7 549.04	466.22	7 357
Zoo	3	64	19 394.3	679.93	2 119.72	71 080.7	1 394.69	2 751.55	423.59	3 219

表 5 的实验结果表明,PP-kNN 分类器的运行时间在几秒到几十秒不等,执行时间随着训练样本数与特征数的增加而增加.与贝叶斯分类器、线性分类器不同,kNN 分类器在训练阶段的消耗为 0,其计算全部集中在分类阶段,因此在密文数据处理的速度方面具有一定优势.

5.3 安全两方工具比较协议评估与对比

PP-kNN 中的基本操作协议都是两方协议,并且支持当前典型两方协议(TASTY[31]、Fairplay[32,33])所支持的全部功能,为了进一步说明 PP-kNN 的性能优势,将本文中的比较协议与 TASTY 进行比对.

本文基于不同比特位数的数据设置相应的比较位数,对 TASTY 安全两方计算工具中的两方比较协议进行了性能测试,实验结果如图 2 所示.TASTY 与本文的比较操作都是基于 Garble 电路的,但 TASTY 的数据传输消耗时间长,本文的数据传送时间较短,因此安全两方比较协议总的运行时间约为 TASTY 的 1/100.综上所述,本文的两方比较协议在性能上有所提高.

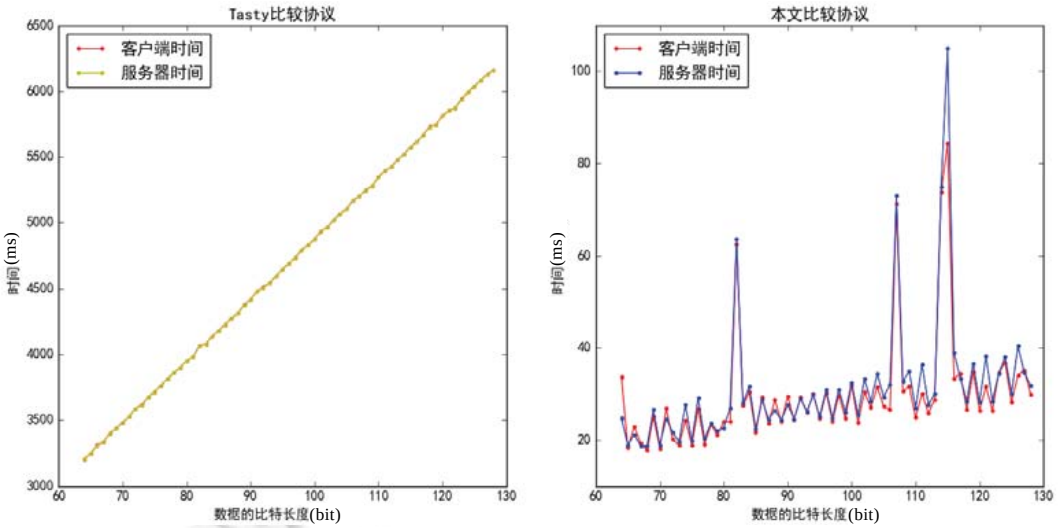

Fig.2 Secure two-party comparison protocol runtime distribution of Tasty and ours

图 2 Tasty 和本文的安全两方比较协议运行时间分布

6 结 论

本文设计了一种支持隐私保护的 kNN 分类器.首先,从 kNN 分类器中提取出了一些基本操作,包括加法、乘法、比较等;其次,选择了两种同态加密方案和一种全同态加密方案对数据进行加密,基于此,设计了针对基本操作的安全协议,并证明了协议在半诚实模型下的安全性;然后,通过将基本操作的安全协议按照模块化顺序组合的方式构造出了 PP-kNN 分类器;最后,在自定义加密数据及 4 种 UCI 标准数据集上,分别对安全协议及所构造的 PP-kNN 分类器进行了性能评估.实验结果表明,本文设计的安全协议是安全且高效的,分类器能够以较高的效率对密文数据进行分类,同时实现了对用户数据的隐私保护.

References:

[1] Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proc. of the ACM SIGSAC Conf. on Computer and Communications Security. New York: ACM, 2015. 1322−1333.

[2] Fredrikson M, Lantz E, Jha S, Lin S, Page D, Ristenpart T. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In: Proc. of the USENIX Security Symp. National Center for Biotechnology Information, U.S. National Library of Medicine, 2014. 12−32.

[3] Alotaibi K, Rayward-Smith VJ, Wang WJ, de la Lglesia B. Non-linear dimensionality reduction for privacy-preserving data classification. In: Proc. of the ASE/IEEE Int'l Conf. on Privacy, Security, Risk and Trust. IEEE, 2012. 694−701. [doi: 10.1109/SocialCom-PASSAT.2012.76]

[4] Agrawal R, Srikant R. Privacy-preserving data mining. In: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM, 2000. 439−450.

[5] Bayardo RJ, Agrawal R. Data privacy through optimal *k*-anonymization. In: Proc.of the Int'l Conf. on Data Engineering (ICDE). Institute of Electrical and Electronics Engineers Computer Society, 2005. 217−228.

[6] Evfimievski A, Srikant R, Agrawal R, Gehrke J. Privacy preserving mining of association rules. Information Systems, 2004,29(4): 343−364. [doi: 10.1016/j.is.2003.09.001]

[7] Lindell Y, Pinkas B. Privacy preserving data mining. In: Proc. of the Advances in Cryptology-crypto. Springer-Verlag, 2000. 36−54.

[8] Hu HB, Xu JL, Ren CS, Choi BR. Processing private queries over untrusted data cloud through privacy homomorphism. In: Proc. of the IEEE Int'l Conf. on Data Engineering (ICDE). IEEE, 2011. 601−612.

[9] Kantarcıoğlu M, Clifton C. Privately computing a distributed *k*-NN classifier. In: Boulicaut JF, ed. Proc. of the Lecture Notes in Artificial Intelligence. Springer-Verlag, 2004. 279−290.

[10] Xiong L, Chitti S, Liu L. *k* nearest neighbor classification across multiple private databases. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management (CIKM). New York: Association for Computing Machinery, 2006. 840−841. [doi: 10.1145/1183614.1183757]

[11] Kung SY, Chanyaswad T, Chang JM, Wu PY. Collaborative PCA/DCA learning methods for compressive privacy. ACM Trans. on Embedded Computing Systems, 2017,16(3):1−18. [doi: 10.1145/2996460]

[12] Liu XM, Lu RX, Ma JF, Chen L, Qin BD. Privacy-preserving patient-centric clinical decision support system on Naïve Bayesian classification. IEEE Journal of Biomedical & Health Informatics, 2016,20(2):655−668. [doi: 10.1109/JBHI.2015.2407157]

[13] Jia Q, Guo LK, Jin ZP, Fang YG. Privacy-preserving data classification and similarity evaluation for distributed systems. In: Proc. of the IEEE Int'l Conf. on Distributed Computing Systems. Los Alamitos: IEEE Computer Society, 2016. 690−699. [doi: 10.1109/ICDCS.2016.94]

[14] Ligier D, Carpov S, Fontaine C, Sirdey R. Privacy preserving data classification using inner-product functional encryption. In: Proc. of the Int'l Conf. on Information Systems Security and Privacy. SciTePress, 2017. 423−430.

[15] Du WL, Han YS, Chen SG. Privacy-preserving multivariate statistical analysis: Linear regression and classification. In: Proc. of the SIAM Int'l Conf. on Data Mining. Society for Industrial and Applied Mathematics Publications, 2004. 222−233.

[16] Graepel T, Lauter K, Naehrig M. ML confidential: Machine learning on encrypted data. In: Proc. of the Information Security and Cryptology (ICISC). LNCS 7839, Springer-Verlag, 2013. 1−21. [doi: 10.1007/978-3-642-37682-5_1]

[17] Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. Journal of Biomedical Informatics, 2014,50: 234−243. [doi: 10.1016/j.jbi.2014.04.003]

[18] Barni M, Failla P, Lazzeretti R, Paus A, Sadeghi AR, Schneider T, Kolesnikov V. Efficient privacy-preserving classification of ECG signals. In: Proc. of the IEEE Int'l Workshop on Information Forensics and Security (WIFS). IEEE, 2009. 91−95. [doi: 10.1109/WIFS.2009.5386475]

[19] Barni M, Failla P, Kolesnikov V, Lazzeretti R, Sadeghi AR, Schneider T. Secure evaluation of private linear branching programs with medical applications. Computer Security (ESORICS), 2009,5789:424−439.

[20] Barni M, Failla P, Lazzeretti R, Sadeghi AR, Schneider T. Privacy-preserving ECG classification with branching programs and neural networks. IEEE Trans. on Information Forensics & Security, 2011,6(2):452−468. [doi: 10.1109/TIFS.2011.2108650]

[21] GUL KSQ, Yin JZ, Pan LM, *et al*. Research on the algorithm of named entity recognition based on deep neural network. Information and Network Security, 2017,(10):29−35 (in Chinese with English abstract). [doi: 10.3969/j.issn.1671-1122.2017.10.005]

[22] Goldwasser S, Micali S. Probabilistic encryption and how to play mental poker keeping secret all partial information. In: Proc. of the ACM Symp. on Theory of Computing. Association for Computing Machinery, 1982. 365−377.

[23] Paillier P. Public-key cryptosystems based on composite degree residuosity classes. In: Proc of the Int'l Conf. on the Theory and Application of Cryptographic Techniques. Springer-Verlag, 1999. 223−238. [doi: 10.1007/3-540-48910-X_16]

[24] Halevi S, Shoup V. Algorithms in HElib. In: Proc. of the Advances in Cryptology-Crypto. LNCS 8616, Springer-Verlag, 2014. 554−571. [doi: 10.1007/978-3-662-44371-2_31]

[25] Goldreich O. The Foundations of Cryptography—Volume 2, Basic Applications. Cambridge University Press, 2004.

[26] Canetti R. Security and composition of multi-party cryptographic protocols. Journal of Cryptology, 2000,13(1):143−202.

[27] Damgard I, Geisler M, Kroigard M. Homomorphic encryption and secure comparison. Int'l Journal of Applied Cryptography, 2008, 1(1):22−31. [doi: 10.1504/IJACT.2008.017048]

[28] Veugen T. Improving the DGK comparison protocol. In: Proc. of the IEEE Int'l Workshop on Information Forensics and Security. IEEE, 2012. 49−54.

[29] Veugen T. Comparing encrypted data. 2011. http://siplab.tudelft.nl/sites/default/files/Comparing%20encrypted%20data.pdf

[30] 2017. http://archive.ics.uci.edu/ml/datasets.html

[31] Henecka W, Kögl S, Sadeghi AR, Schneider T, Wehrenberg I. Tasty: Tool for automating secure two-party computations. In: Proc. of the ACM Conf. on Computer and Communications Security (CCS). Association for Computing Machinery, 2010. 451−462. [doi: 10.1145/1866307.1866358]

[32] Malkhi D, Nisan N, Pinkas B, Sella Y. Fairplay—A secure two-party computation system. In: Proc. of the Usenix Security Symp. Berkeley: USENIX Association, 2004. 287−302.

[33] Ben-David A, Nisan N, Pinkas B. Fairplaymp: A system for secure multi-party computation. In: Proc. of the ACM Conf. on Computer and Communications Security. ACM, 2008. 257−266.

附中文参考文献:

[21] GUL KSQ,尹继泽,潘丽敏,等.基于深度神经网络的命名实体识别方法研究.信息网络安全,2017,(10):29−35. [doi: 10.3969/j.issn.1671-1122.2017.10.005]

徐剑(1978－),男,山东胶南人,博士,副教授,CCF 专业会员,主要研究领域为网络与信息安全,云计算安全,机器学习与隐私保护.

毕猛(1982－),男,博士,工程师,主要研究领域为网络与信息安全.

王安迪(1996－),女,硕士生,CCF 学生会员,主要研究领域为机器学习,隐私保护.

周福才(1964－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络与信息安全,可信计算,电子商务基础理论及关键技术.