South Ural State University

National Research University

Южно-Уральский государственный университет

Высшая школа электроники и компьютерных наук

Международная лаборатория проблемно-ориентированных облачных сред

Семинар
Машинное обучение с сохранением конфиденциальности как услуга

Докладчик

Хорхе Марио Кортес-Мендоса

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

ISP RAS

Xi'an Jiaotong-Liverpool University 西交利物浦大學

Россия, декабрь 2020.

South Ural State University
National Research University
1943

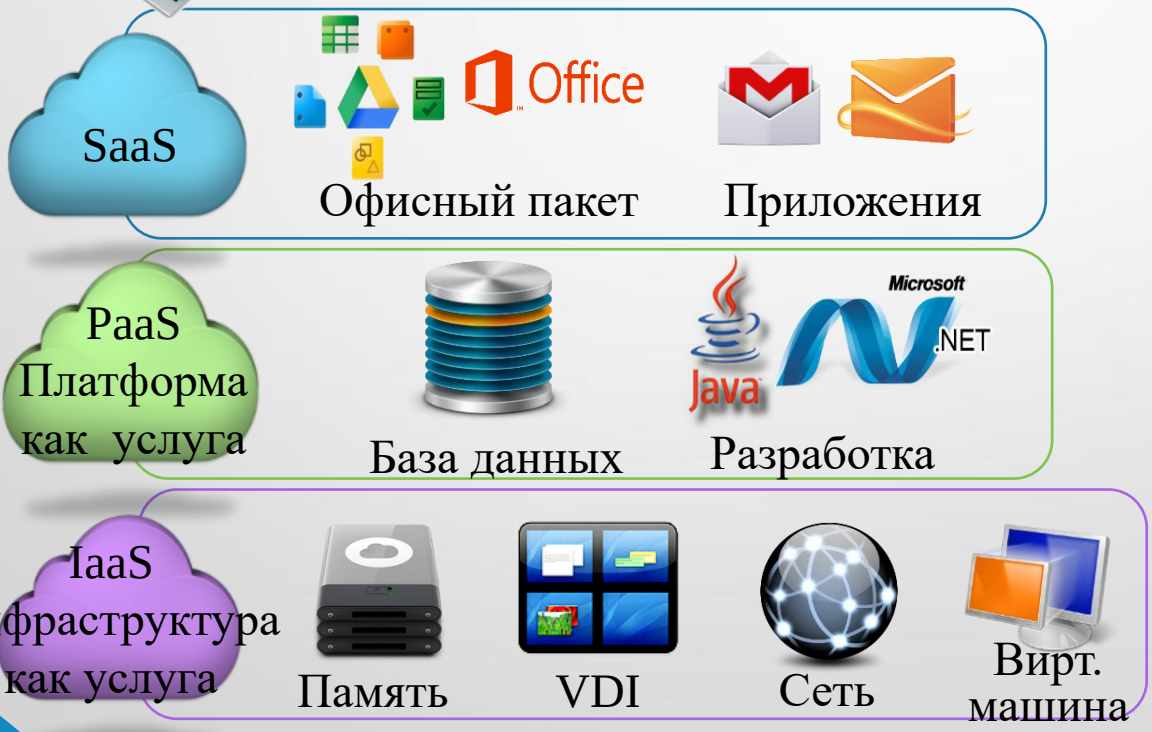
Содержание

- Обоснование
- Машинное обучение как услуга
- Гомоморфное шифрование
- Нейронные сети с сохранением конфиденциальности
- Логистическая регрессия с сохранением конфиденциальности
- Направление дальнейших исследований

Обоснование

Облачные вычисления получили широкое распространение, поскольку позволяют получать вычислительные ресурсы по запросу

Машинное обучение как услуга(MLaaS) является гибким и масштабируемым решением в облачных средах



Обоснование

MLaaS предлагает различные виды ресурсов и инструментов для обучения и развертывания моделей машинного обучения

Нейронные сети

Глубокое обучение

Обработка естественного языка

Процесс машинного обучения

Процесс обучения может потреблять много вычислительных ресурсов и времени

- Высокопроизводительные вычислительные ресурсы в облаке могут сократить время обучения и тестирования
- Удаленная инфраструктура облака уменьшает проблемы ресурсов и реализации, но создает ряд проблем конфиденциальности для закрытых данных

South Ural State University
National Research University
1943

Машинное обучение как услуга

Безопасность данных в облачных вычислениях обеспечивает их защиту от кражи, утечки, удаления на уровнях брандмауэров, тестирования на проникновение, обфускации, токенизации, виртуальных частных сетей (VPN) и т. д.

Использование сторонних сервисов может привести к нескольким рискам кибербезопасности

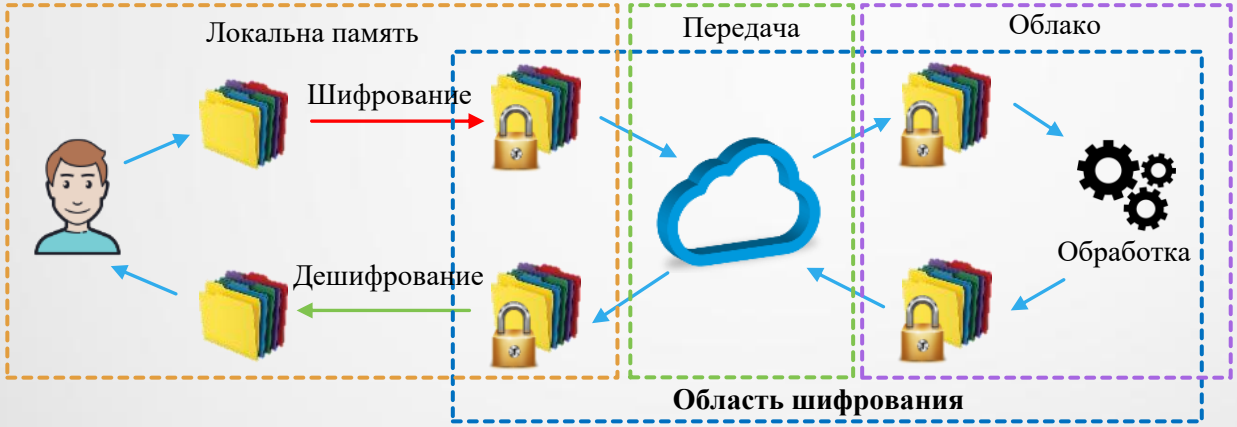
- Обычное шифрование не решает вопроса, поскольку модель МО требует полный доступ к конфиденциальным данным

Безопасность и конфиденциальность – серьезные задачи, поскольку данные могут быть расшифрованы для анализа

Гомоморфное шифрование

Гомоморфное шифрование (ГШ) и безопасные многопользовательские вычисления (БМВ) – это способы устранения уязвимостей обработки данных

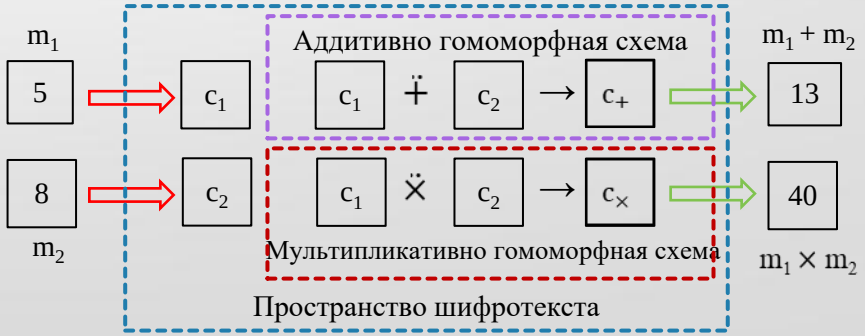
ГШ – система шифрования, позволяющая обрабатывать информацию о шифротекстах



Схемы ГШ усиливают ряд аспектов безопасности в облаке

Шифротексты c_1 и c_2 шифруют содержание сообщений m_1 и m_2

- c_+ создается с помощью c_1 и c_2 , а его расшифровка дает $m_1 + m_2$
- c_x шифрует m_1 x m_2



Гомоморфное шифрование

Термин “гомоморфный” относится к отображению функций в пространстве сообщений и шифротекстов

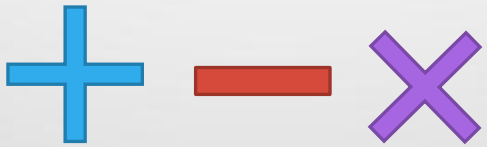
- Функция, применяемая к шифротекстам, дает тот же (зашифрованный) результат, что и ее гомоморфная функция, используемая в зашифрованных ими сообщениях

Система использует только общедоступную информацию без риска утечки данных

- Нет доступа к информации в шифротексте или какого-либо секретного ключа

Реализация ГШ имеет несколько ограничений, из которых три основных - это:

Низкая производительность



Малое число простейших операций

Нейронные сети Глубокое обучение

Модели МО

South Ural State University
National Research University
1943

Гомоморфное шифрование для МО

Исследования в сфере ГШ имеют большую значимость для МО, сосредоточив внимание, в частности, на повышении производительности, новых подходах, приложениях. Они дают фундаментальные знания и общую панораму исследователям, заинтересованным в применении и расширении методов ГШ.

Таблица 1. Основня проблематика статей о ГШ

Литература \ Проблематика	Техническая	Ограничения	Приложения	Инструменты	Облачные	Реализации
Vaikuntanathan [1]	●		●			
Armknecht et al. [2]	●	●	●	●		●
Naehrig et al. [3]	●	●	●		●	●
Archer et al. [4]			●		●	
Acar et al. [5]	●	●	●	●		
Martins et al. [6]	●	●	●			
Parmar et al. [7]	●		●			
Shunmuganathan [8]	●		●			
Gentry [9]	●	●				
Aguilar-Melchor [10]	●		●			●
Hrestak and Picek [11]	●			●	●	
Moore et al. [12]	●					●

South Ural State University
National Research University
1943

Гомоморфное шифрование для МО

Таблица 2. Сравнение методов ГШ

Год	Операции			Метод МО					Схемы					Показатель	
	Сложение	Умножение	Прочие	LR	NN	DNN	Decision Trees	Ideal Lattice	Integer-based	(R) LWE	NTRU	Two-party	Multi-party	Безопасность	Эффективность
1978		•										•		•	
1985		•										•		•	
1999	•											•		•	
2009	•	•						•				•		•	
2011	•	•		•						•		•		•	
2014	•	•									•	•			•
	•	•		•					•			•		•	
2015	•	•					•			•		•		•	
	•	•					•			•		•		•	
2016	•	•				•			•				•	•	•
	•	•			•				•			•		•	
	•	•		•						•		•		•	
	•	•				•			•				•	•	
2017	•	•								•		•		•	•
	•	•				•			•			•		•	
	•	•								•		•		•	
2018	•	•		•						•		•		•	
	•					•			•				•	•	•
	•	•				•			•			•		•	
	•	•				•			•			•		•	•
	•	•				•			•			•		•	
2019	•	•			•				•				•	•	
	•	•	•									•		•	•

Количество простейших операций для прогнозирования и классификации конфиденциальной информации с использованием схем ГШ невелико.

Основная цель – расширение парадигмы МОкУ.

South Ural State University
National Research University
1943

Гомоморфное шифрование для МО

Теоретические исследования в области ГШ должны быть дополнены качественными реализациями.

В последние годы промышленные и академические группы выпустили несколько библиотек ГШ

Таблица 3. Сравнение достоинств и недостатков основных библиотек ГШ общего назначения

Инструмент	Поддержка	Достоинства	Недостатки
SEAL	Microsoft	Хорошо документирован ; Простая настройка параметров безопасности	Малая гибкость; Ограничено количество поддерживаемых схем
HElib	IBM	Эффективные гомоморфные операции	Низкая производительность бутстрэппинга; Сложная настройка параметров безопасности
TFHE		Быстрая самозагрузка	Низкая производительность для простых задач
PALISADE	DARPA, MIT, UCSD, etc.	Многообразие схем ГШ; Многоплатформенный	Слабая документация и поддержка
cuHE		Массовый параллелизм и высокая пропускная способность памяти графических процессоров	
HEAAN	Сеульский Нац. Университет	Операции между рациональными числами	
HE-transformer	Intel	Интеграция с библиотеками глубокого обучекния	Расширение SEAL

South Ural State University
National Research University
1943

Гомоморфное шифрование для МО

Анализ показывает растущий интерес исследовательского сообщества к конструкциям ГШ при обработке особо секретных данных

- Модели машинного обучения для обработки зашифрованных данных

Рис. 1. Ключевые слова, относящиеся к понятиям ГШ и специальным приложениям в сфере ГШ (5 лет)

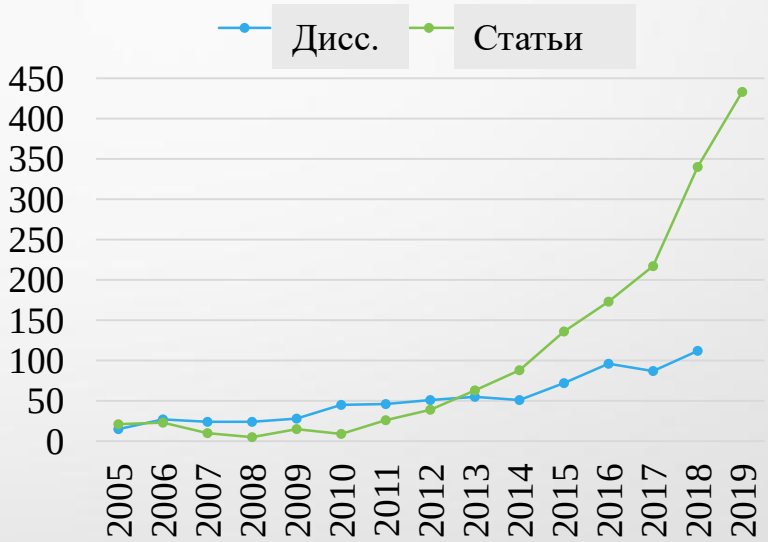


Рис 2. Количество публикаций, относящихся к ГШ

South Ural State University

National Research University

Южно-Уральский государственный университет
Высшая школа электроники и компьютерных наук
Международная лаборатория проблемно-ориентированных облачных сред

Нейронные сети с сохранением конфиденциальности

Jorge M. Cortés-Mendoza
Andrei Tchernykh
Mikhail Babenko
Luis B. Pulido-Gaytán
Gleb Radchenko
Arutyun Avetisyan
Alexander Yu. Drozdov

Россия, декабрь 2020.

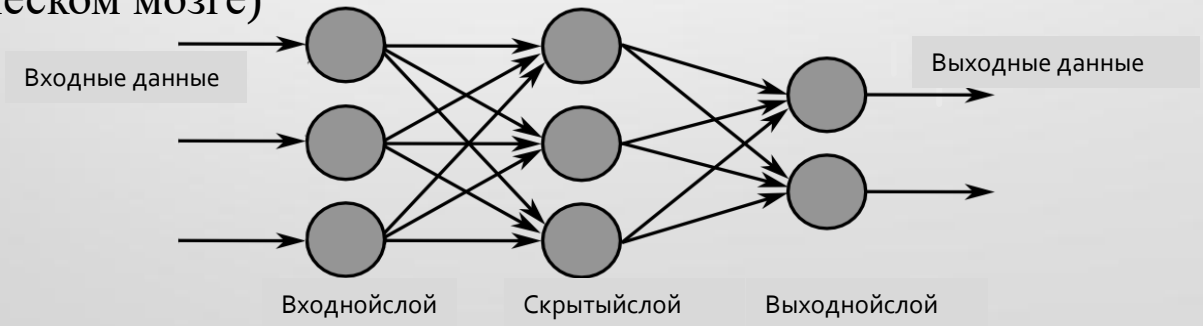
South Ural State University
National Research University
1943

Нейронные сети с сохранением конфиденциальности

Искусственные нейронные сети (ИНС), или нейронные сетии (НС), - это вычислительные системы, созданные по типу биологических нейронных сетей в мозгу живых существ.

НС – это совокупность связанных элементов, или узлов, называемых искусственными нейронами, являющихся примерными моделями нейронов в биологическом мозге

Каждое соединение может передавать сигнал другим нейронам (подобно синапсам в биологическом мозге)



Нейронные сети с сохранением конфиденциальности

Каждый нейрон состоит из n_I входов $x = (x_1, \ldots, x_{n_I})$ и выхода y

$$y = f\left(\sum_{i=1}^{n_I} w_i \times x_i + \beta\right)$$

Величина y определяет взвешенную сумму входных данных с учетом весов $w = (w_1, \ldots, w_{n_I})$, смещение β и нелинейную функцию активации f

ГШ-версия нейрона (НС-ГШ) замещает $+, \times,$ и f

$$\bar{y} \leftarrow \ddot{f}\left(\sum_{i=1}^{n_I} (\bar{w}_i \ddot{\times} \bar{x}_i) \ddot{+} \bar{\beta}\right)$$

где $\bar{x}, \bar{w},$ и $\bar{\beta}$ - это соответствующие шифротексты $x, w,$ и $\beta,$ а $\ddot{f}$ - гомоморфная версия f

$\ddot{f}$ - полиномиальная аппроксимация, состоящая только из операций $\ddot{+}$ and $\ddot{\times}$

$\bar{y}$ содержит зашифрованный выход нейронного вычисления, обеспечивая конфиденциальность результата даже при раскрытии

Структура сети определяет взаимодействие между слоями (совокупность нейронов)

НС-ГШ не производит никаких изменений в структуре НС

Нейронные сети с сохранением конфиденциальности

Функция активации важна в конструкции модели НС

- Определение $\ddot{f}$ остается открытой проблемой (стандартная функция активации использует операции, не поддерживаемые ГШ)

Полиномиальная аппроксимация вынуждена балансировать между сложностью и точностью.

- Полином высокой степени дает высокую точность при медленных вычислениях
- Полином низкой степени позволяет быстрые вычисления с низкой точностью

Таблица 4. Краткое описание аппроксимации функции активации

Функция	n	Модель LR	Модель NN	Метод аппроксимации
Sigmoid	2, 3		●	Полиномы Чебышева
	2	●		Ряды Тейлора, область
	1	●		Ряды Тейлора
	3, 5, 7	●		Ряды Тейлора
	9		●	Ряды Тейлора, Padé
Tanh	2, 3		●	Полиномы Чебышева
	9		●	Ряды Тейлора, Padé
	3, 4		●	Полиномы Чебышева
ReLU	2,3,4,5,6		●	Подбор полинома меотдом наименьш.квадратов (soft.)
	2, 3		●	Производная по ReLU
	1		●	Ряды Тейлора, Padé
	3, 4		●	Полиномы Чебышева
	2		●	Политопный метод
Swish	3,4		●	Полиномы Чебышева
	2		●	Политопный метод

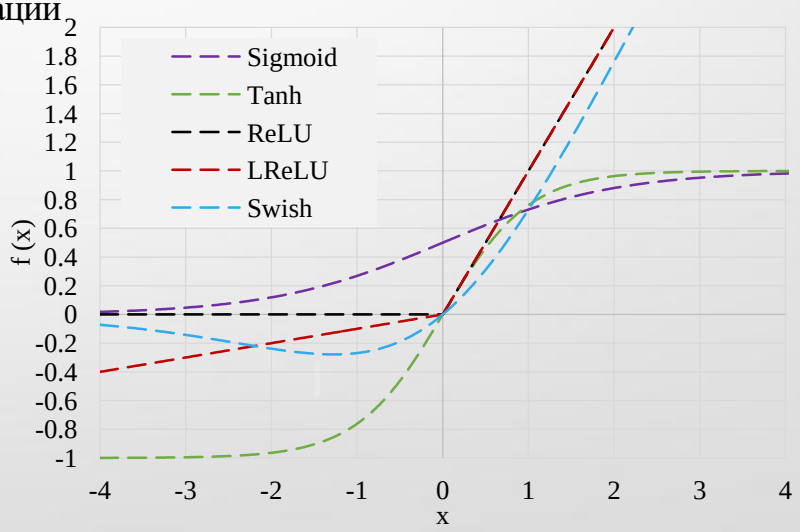


Рис. 3. Функции активации

Нейронные сети с сохранением конфиденциальности

Процесс обучения состоит из разработки преобразования из входного пространства в выходное на основе модификации *w* каждого нейрона.

- В домене ГШ процесс обучения подразумевает *большие шифрованные сообщения* и *несколько процедур бутстрэппинга*.

1. Бутстрэппинг снижает шум в шифротексте
2. Шум гарантирует определенную степень безопасности
3. Каждая операция усиливает исходный шум
4. Дешифрование сообщения невозможна, если шум превышает определенный порог

*“Вычислительные затраты семиуровневого обучения КНС – около **одного часа** с обычным ЦПУ, в то время как обучение той же КНС с ГШ требует около **года**[13]”*

Нейронные сети с сохранением конфиденциальности

Две опции являются общими в работе с бутстрэппингом при НС-ГШ обучении

1. Ускорение бутстрэппинга. Высокопроизводительные, распределенные и параллельные вычисления дают средства для обучения больших шифрованных массивов данных
   - Аппаратные ускорители GPU, FPGA, etc.) и конкретные микросхемы (ASIC)

2.1 Избегание операций бутстрэппинга направлено на дешифровании шифротекста внутри безопасного объекта (клиент-сервер, безопасный HPC и т.п.)
   - Гибридная модель между ГШ и SMC

2.2. Значимость предварительного обучения НС. Во избежание лишних расходов обучение производят над нешифрованными данными (оценка – над шифрованными)
   - Современная практика

Оценка НС-ГШ включает эффективные реализации взвешенной суммы и f
   - Операция умножения более медленная и добавляет большое количество шума
   - Операция бутстрэппинга необходима, если шифротекст содержит большое количество шума

South Ural State University

National Research University

Южно-Уральский государственный университет
Высшая школа электроники и компьютерных наук
Международная лаборатория проблемно-ориентированных облачных сред

Логистическая регрессия с сохранением конфиденциальности как облачный сервис на основе системы счисления остаточных классов

Jorge M. Cortés-Mendoza
Andrei Tchernykh
Mikhail Babenko
Luis B. Pulido-Gaytán
Gleb Radchenko
Franck Leprevost
Xinheng Wang
Arutyun Avetisyan
Sergio Sergio Nesmachnow

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Xi'an Jiaotong-Liverpool University
西交利物浦大學

ISP RAS

Russia, December 2020.

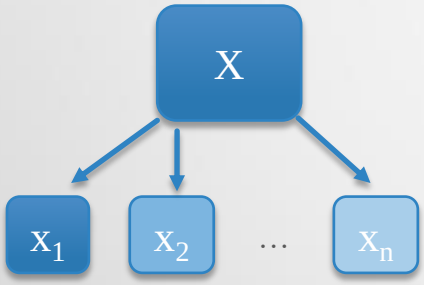
South Ural State University
National Research University
1943

Система счисления остаточных классов

Система остаточных классов (СОК) представляет собой вариацию изоморфизма конечного кольца, где исходные числа представлены в виде остатков.

Набор модулей парных взаимно простых чисел $\{p_1, p_2, \dots, p_n\}$ определяет представление значений в множестве $P = \prod_1^n p_i$

Целое число $X \in [0, P-1)$ определяется в СОК как последовательность $(x_1, x_2, \dots, x_n)$ где x_i является остатком деления X by p_i



$p_1 = 7, p_2 = 5, p_3 = 3, p_4 = 2,$

$8_{10} = (1, 3, 2, 0)_{RNS}$

делитель → 7 | 8 ← делимое

1 ← частное

1 ← остаток

$X \otimes Y = Z$

4-moduli

$x_1 \otimes y_1 \bmod m_1 = z_1$

$x_2 \otimes y_2 \bmod m_2 = z_2$

$x_3 \otimes y_3 \bmod m_3 = z_3$

$x_4 \otimes y_4 \bmod m_4 = z_4$

$\otimes$ обозначает $+, -, \times$

South Ural State University
National Research University
1943

Логистическая регрессия

Логистическая регрессия (ЛР) – статистический метод анализа информации, где:

Массив данных $X^{(i)} \in \mathbb{R}^d$ и их метки $Y^{(i)} \in \{0,1\}$ для $i = 1, 2, \dots, n$ используются для моделирования двоично зависимой переменной

- Предсказание бинарного исхода учитывает логистическую функцию

Логический вывод ЛР учитывает условие $h_\theta(X^{(i)}) = g(\theta^T X^{(i)})$ where

- Logistic function: $g(z) = \frac{1}{1+e^{-z}}$
- Weights: $\theta^T = [\theta_0, \theta_1, \dots, \theta_d]^T$
- Data: $X^{(i)} = [1, X_1{}^{(i)}, X_2{}^{(i)}, \dots, X_d{}^{(i)}]^T$

Фаза обучения ЛР направлена на нахождение θ^*, значений θ, минимизирующих количество ошибок в предсказании

- θ^* используется для оценки бинарной классификации новых данных
- Для $X' = [1, X_1, \dots, X_d] \in \mathbb{R}^{d+1}$ можно предсказать его двоичное значение $Y' \in \{0,1\}$ by

$$Y' = \begin{cases} 1 & if\ h_{\theta^*}(X') \geq \tau \\ 0 & if\ h_{\theta^*}(X') < \tau \end{cases}$$

τ определяет переменный порог в $0 < \tau < 1$, обычно с величиной, равной 0.5

Логистическая регрессия с сохранением конфиденциальности

Градиентный спуск (ГС) – это алгоритм оптимизации для минимизации функции ошибок или целевой функции θ

- Процесс оптимизации обновляет θ согласно $\nabla_\theta J(\theta)$, частной производной от $J(\theta)$
- Скорость обучения α определяет размеры шагов

Algorithm 1. Пакетный ГС

Input: $X, Y, \theta, \alpha, and\ maxIter$
Output: θ
1 For $i \leftarrow 1$ to $maxIter$
2 $\theta \leftarrow \theta - \alpha \nabla_\theta J(\theta, X, Y)$
3 Return θ

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n} y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

Мы предлагаем конфиденциальность данных ЛР для облачного сервиса с ГШ на базе СОК

Таблица 5. Основные характеристики схем ГШ для логистической регрессии

Шифрование	Степень полиномиал. аппроксимации	Градиентный спуск	Metrics	Library	Datasets	Ref.
Paillier, LWE, Ring-LWE	2	BGD	F-score, AUC	-	Pima, SPECTF	[14]
Ring-LWE	1	GD-FHN	ROC, accuracy	NFLlib	iDASH, financial data	[15]
Ring-LWE	3,5, 7	NAG	AUC, accuracy	HEAAN	iDASH, lbw, mi, nhanes3, pcs, uis	[16]
Ring-LWE, RNS	7	NAG	AUC, accuracy	HEAAN	Lbw, uis	[17]
Ring-LWE	5	NAG	AUC	HEAAN	MNIST, credit	[18]
-	-	BGD	AUC	-	NIDDK	[19]

Логистическая регрессия с сохранением конфиденциальности

Обычно в литературе используются четыре варианта основного ГС:

- Пакетный градиентный спуск (ПГС)
- Стохастический ГС (СГС)
- Импульсный градиентный спуск (ИГС)
- Ускоренный градиент Нестерова (УГН)

Алгоритм2. Стохастический ГС

Input: $X, Y, \theta, \alpha, and\ iters.$
Output: $\theta.$
For $i \leftarrow 1$ to $iters$
  Shuffle (X, Y)
  For $j \leftarrow 1$ to $length(X)$
    $\theta \leftarrow \theta - \alpha \nabla_\theta J(\theta, x^{(j)}, y^{(j)})$
Return θ

Алгоритм 3. Импульсный ГС

Input: $X, Y, \theta, \alpha, \beta, and\ iters.$
Output: $\theta.$
For $i \leftarrow 1$ to $iters$
  Shuffle (X, Y)
  For $j \leftarrow 1$ to $length(X)$
    $v_t \leftarrow \beta v_{t-1} - \alpha \nabla_\theta J(\theta, x^{(j)}, y^{(j)})$
    $\theta \leftarrow \theta + v_t$
Return θ

Алгоритм 4. Ускоренный градиент Нестерова

Input: $X, Y, \theta, \alpha, \beta, and\ iters$
Output: $\theta.$
For $i \leftarrow 1$ to $iters$
  Shuffle (X, Y)
  For $j \leftarrow 1$ to $length(X)$
    $v_t \leftarrow \beta v_{t-1} - \alpha \nabla_\theta J(\theta - \beta v_{t-1}, x^{(j)}, y^{(j)})$
    $\theta \leftarrow \theta + v_t$
Return θ

Логистическая регрессия с сохранением конфиденциальности

Каждая итерация алгоритма, все записи в обучающей последовательности используются для обновления значений θ.

hTheta обеспечивает полиномиальную аппроксимацию для логистической функции

HE.rescale устраняет накопившийся коэффициент пересчета, формирующийся после каждого умножения

Алгоритм 5. ГШ-СОК Пакетный ГС

Input: $X, Y, theta, alpha, maxIter$
Output: $theta$

```
1  For iter ← 1 to maxIter
2      For i ← 1 to X.size
3          parcialCost ← HE.sub ( hTheta ( X[i], theta), Y[i] ) )
4          For j ← 1 to theta.size
5              cost[j] ← HE.add ( cost[j], HE.mul ( parcialCost, X[i][j] ) )
6      For i ← 1 to theta.size
7          cost[i] ← HE.rescale ( HE.mul( average, HE.mul ( cost[i], alpha) ) )
8          theta[i] ← HE.sub ( theta[i], cost[i] )
9  Return theta
```

Мы предлагаем ЛР с сохранением конфиденциальности данных для облачного сервиса с ГШ на основе СОК

Настройка конфигурации

Экспериментальный анализ рассматривает 50 конфигураций для каждого набора данных для сравнения производительности и качества нашего решения с современными алгоритмами

- Шесть массивов данных из медицины и геномики
- Полиномиальная аппроксимация логистической функции
- 5-кратная перекрестная проверка
- Скалярный коэффициент 16 bits
- Семь попарных взаимно простых чисел
- Итерации: 5, 10, 15, 20, 25
- Скорость обучения: 1.6, 1.1, 0.6, 0.1, 0.06, 0.01, 0.006, 0.001, 0.0006, 0.0001

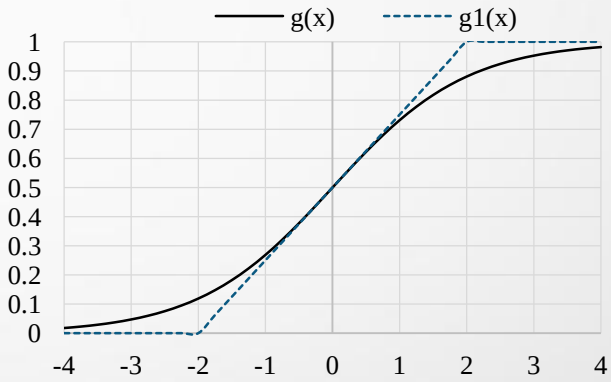


Fig. 5. Sigmoid and approximation functions

Показатели

- Точность (A) выражает систематическую ошибку для оценки величины
- Область под кривой ошибок (AUC) - показатель эффективности классификаторов

South Ural State University
National Research University
1943

Настройка конфигурации

Мы рассматриваем шесть массивов данных, широко используемых в источниках

1. Массив Низкая масса тела новорожденных (Lbw) состоит из данных о рождениях для пациенток акушерских клиник
2. Инфаркт миокарда (Mi) – массив данных по сердечным заболеваниям
3. Национальное исследование здоровья и питания (Nhanes3) включает базы данных экспосом и феномов человека
4. Массив данных Диабет в Индии (Pima)
5. Массив данных Исследование рака простаты (Pcs) – о пациентах с и без рака простаты
6. Массив данных Изучение последствий Умару (Uis) хранит информацию о лечении населения от наркомании

Таблица 6. Характеристики и размеры массивов данных

Массив данных	N	Features	N-Обучение	N-Тестирование
Lbw	189	9	151	38
Mi	1,253	9	1,002	251
Nhanes3	15,649	15	12,519	3,130
Pima	768	8	614	154
Pcs	379	9	303	76
Uis	575	8	460	115

Результаты

South Ural State University
National Research University
1943

В таблице 7 даны лучшие значения *AUC* и *A* для всех версий ГС; каждая величина является средним для 30 выполнений с различными исходными значениями θ

Таблица7. Средняя AUC (кривая ошибок) и точность A

Название	*AUC*							*A*(%)						
	Lbw	Mi	Nhanes3	Pcs	Pima	Uis	Average	Lbw	Mi	Nhanes3	Pcs	Pima	Uis	Average
ГШ-ПГС-СОК	0.7358	0.9388	0.8112	0.7445	0.6983	0.5483	0.7557	71.84	88.87	78.89	66.05	67.79	74.43	76.02
ПГС	0.7353	0.9357	0.7961	0.7406	0.6964	0.5458	0.7507	71.84	89.02	78.86	66.14	67.65	74.35	76.04
ГШ-SGD-RNS	0.7541	0.9421	**0.9029**	**0.8151**	0.8505	**0.6118**	0.8052	**73.42**	88.9	84.51	**66.32**	**74.7**	**74.81**	77.59
СГС	0.7618	**0.9445**	**0.903**	0.8162	0.8487	**0.6158**	0.8083	73.86	89.39	84.3	66.32	74.7	**74.75**	77.72
ГШ-ИГС-СОК	**0.7552**	**0.9445**	0.902	0.8143	**0.8508**	0.6116	0.8055	72.89	**88.95**	**84.53**	66.01	74.66	74.72	77.42
ИГС	**0.7634**	**0.9445**	**0.903**	0.8169	0.8488	0.6152	0.8086	73.86	**89.42**	**84.33**	**66.36**	74.77	74.72	77.74
ГШ-УГН-СОК	**0.7552**	**0.9445**	0.902	0.8143	**0.8508**	0.6115	0.8055	72.81	**88.95**	**84.53**	66.01	**74.7**	74.72	77.40
УГН	0.763	**0.9445**	**0.903**	**0.817**	**0.8489**	0.6154	0.8086	**74.04**	**89.42**	**84.33**	**66.36**	**74.79**	74.72	77.77
НЕ-NA-LR[16]	0.689	0.958	0.717	0.74	-	0.603	0.7414	69.19	91.04	79.22	68.27	-	74.44	76.43
ГШ-SS-ЛР [14]	-	-	-	-	0.8763	-	-	-	-	-	-	80.7	-	-

Для *AUC* лучшие решения в трех массивах данных у ГШ-СГС-СОК, ГШ-ИГС-СОК и ГШ-УГН-СОК.

Для *A* трижды лучшие значения для θ дают ГШ-СГС-СОК и ГШ-УГН-СОК.

Максимальная разница между алгоритмами СОК и негомоморфными:

- 1.51 % для *AUC* с ГШ-СГС-СОК и Nhanes3
- 1.23 % для *A* с ГШ-УГН-СОК и Lbw.

Results

В таблице 8 даны лучшие значения *AUC* и *A* для всех версий ГС; каждая величина представляет лучшее θ из 1,500 выполнений: скорость обучения x итерации x исходные

Таблица 8. Лучшие значения *AUC* и *A*

Name	*AUC*							*A*(%)						
	Lbw	Mi	Nhanes3	Pcs	Pima	Uis	Average	Lbw	Mi	Nhanes3	Pcs	Pima	Uis	Average
ГШ-ПГС-СОК	0.7981	0.9485	0.8509	0.8045	0.795	0.585	0.7974	78.95	90.44	79.74	77.63	74.66	76.52	80.66
ПГС	0.8013	0.947	0.8317	0.8061	0.7946	0.5846	0.7941	78.95	90.84	79.36	**77.63**	73.97	76.52	80.66
ГШ-SGD-RNS	0.7949	0.9536	**0.9039**	**0.8357**	0.8602	0.6604	0.8297	**81.58**	**91.24**	**86.01**	**78.95**	**79.45**	76.52	82.86
СГС	0.7949	0.9557	**0.9039**	0.8341	0.86	0.66	0.8297	81.58	**91.24**	**86.17**	**77.63**	**80.14**	76.52	82.63
ГШ-ИГС-СОК	**0.8125**	**0.9541**	0.9033	0.8341	**0.8608**	**0.6632**	0.8334	**81.58**	90.84	85.88	**78.95**	**79.45**	**79.13**	83.28
ИГС	0.8045	0.9562	**0.9039**	**0.8518**	0.8627	0.6596	0.8352	81.58	**91.24**	85.88	**77.63**	78.77	**77.39**	82.74
ГШ-УГН-СОК	0.8013	0.9536	0.9033	0.8349	0.8596	0.6584	0.8303	**81.58**	**91.24**	85.94	**78.95**	**79.45**	76.52	82.85
УГН	**0.8077**	**0.9574**	**0.9039**	0.8486	**0.8631**	**0.6616**	0.8358	**84.21**	**91.24**	85.97	**77.63**	79.45	76.52	83.11
НЕ-NA-LR[16]	0.689	0.958	0.717	0.740	-	0.603	0.7414	69.19	91.04	79.22	68.27	-	74.44	76.43
ГШ-SS-ЛР [14]	-	-	-	-	0.8763	-	-	-	-	-	-	80.7	-	-

Для *AUC* ГШ-ИГС-СОК дает лучшие значения θ в четырех из шести массивов данных; за ним следует ГШ-СГС-СОК с двумя.

Для *A* ГШ-СГС-СОК превосходит ГШ-ИГС-СОК и ГШ-УГН-СОК в пяти массивах.

Максимальная разница между алгоритмами СОК и негомоморфными:

- 1.92 % для *AUC* для ГШ-ПГС-СОК с массивом Nhanes3
- 2.63 % для *A* для ГШ-УГН-СОК с учетом массива Lbw.

South Ural State University
National Research University
1943

Направление дальнейшей работы

Нейронные сети с сохранением конфиденциальности

1. Полиномиальная аппроксимация функции активации $\ddot{f}$
2. Бутстрэппинг
   - Ускорение
   - Безопасные многопользовательские вычисления
   - Модели предварительно обученных НС

Логистическая регрессия с сохранением конфиденциальности с СОК

1. Уровень безопасности
2. Полиномиальная аппроксимация логистической функции

South Ural State University
National Research University
1943

Публикации

1. Luis Bernardo Pulido-Gaytan, Andrei Tchernykh, **Jorge M. Cortés-Mendoza**, Mikhail Babenko, Gleb Radchenko, Arutyun Avetisyan, and Alexander Yu. Drozdov. Privacy-Preserving Neural Networks via Homomorphic Encryption: Challenges and Opportunities. ***Peer-to-Peer Networking and Applications: Special Issue on Advances in Privacy-Preserving Computing***, Springer. IF 2.793, Q2. July 2020 (under review).
2. Andrei Tchernykh, Luis Bernardo Pulido-Gaytan, Mikhail Babenko, **Jorge M. Cortés-Mendoza**, Gleb Radchenko, Arutyun Avetisyan, Alexander Yu. Drozdov. Privacy-Preserving Toward Fast and Accurate Polynomial Approximations for Practical Homomorphic Evaluation of Neural Network Activation Functions. ***International Workshop on Security, Privacy and Performance of Cloud Computing*** (SPCLOUD 2020), Barcelona, Spain. December 2020 (accepted).
3. Luis Bernardo Pulido-Gaytan, Andrei Tchernykh, **Jorge Mario Cortés-Mendoza**, Mikhail Babenko, Gleb Radchenko. A Survey on Security-Preserving of Machine Learning with Fully Homomorphic Encryption. ***CARLA 2020 -The Latin America High Performance Computing Conference***. Cuenca, Ecuador. September 2020 (accepted).

Публикации

4. **Jorge M. Cortés-Mendoza**, Gleb Radchenko, Andrei Tchernykh, Luis Bernardo Pulido-Gaytan, Mikhail Babenko, Arutyun Avetisyan, Alexander Yu. Drozdov, and Sergio Nesmachnow. Privacy-Preserving Logistic Regression Solutions based on Residue Number System: Design and Analysis. ***2nd Workshop on Secure IoT, Edge and Cloud systems (SIoTEC) 2021***, Melbourne, Australia. May 2021 (under submission).
5. Mikhail Babenko, Andrei Tchernykh, Bernardo Pulido-Gaytan, Elena Golimblevskaia, **Jorge M. Cortés-Mendoza**, Arutyun Avetisyan. Experimental Evaluation of Homomorphic Comparison Methods. ***ISPRAS OPEN 2020 - Ivannikov ISP RAS Open Conference***, Moscow, Russia, December 10-11, 2020 (under review)
6. **Jorge M. Cortés-Mendoza**, Andrei Tchernykh, Mikhail Babenko, Luis Bernardo Pulido-Gaytán, and Gleb Radchenko. Privacy-Preserving Logistic Regression with Residue Number System as a Cloud Service. ***RuSCDays'20 - The Russian Supercomputing Days***. Moscow, Russia. September 2020 (accepted).

South Ural State University
National Research University
1943

Литература

[1] Vaikuntanathan, V.: Computing Blindfolded: New Developments in Fully Homomorphic Encryption. In: IEEE 52nd Annual Symposium on Foundations of Computer Science, Palm Springs. pp. 5–16 (2011).

[2] Armknecht, F., Boyd, C., Carr, C., Gjøsteen, K., Jäschke, A., Reuter, C.A., Strand, M.: A Guide to Fully Homomorphic Encryption, IACR Cryptology ePrint Archive, (2015).

[3] Naehrig, M., Lauter, K., Vaikuntanathan, V.: Can Homomorphic Encryption Be Practical? In: 3rd ACM Workshop on Cloud Computing Security Workshop - CCSW ’11. pp. 113– 124 (2011)

[4] Archer, D., Chen, L., Cheon, J.H., Gilad-Bachrach, R., Hallman, R.A., Huang, Z., Jiang, X., Kumaresan, R., Malin, B.A., Sofia, H., Song, Y., Wang, S.: Applications of Homomorphic Encryption. (2017)

[5] Acar, A., Aksu, H., Selcuk Uluagac, A., Aksu, H., Uluagac, A.S.: A Survey on Homomorphic Encryption Schemes: Theory and Implementation. ACM Comput. Surv. 51, (2018). https://doi.org/10.1145/3214303

[6] Martins, P., Sousa, L., Mariano, A.: A Survey on Fully Homomorphic Encryption: An Engineering Perspective. ACM Comput. Surv. 50, 33 (2017). https://doi.org/10.1145/3124441

[7] Parmar, P. V, Padhar, S.B., Patel, S.N., Bhatt, N.I., Jhaveri, R.H., S’ad Vidya, S., Shri S’ad, M., Mandal, V.: Survey of Various Homomorphic Encryption Algorithms and Schemes. Int. J. Comput. Appl. 91, (2014)

[8] Sobitha Ahila, S., Shunmuganathan, K.L.: State Of Art in Homomorphic Encryption Schemes. Int. J. Eng. Res. Appl. 4, 37–43 (2014)

[9] Gentry, C.: Computing on the Edge of Chaos: Structure and Randomness in Encrypted Computation. In: Proceedings of the International Congress of Mathematicians (2014)

[10] Aguilar-Melchor, C., Fau, S., Fontaine, C., Gogniat, G., Sirdey, R.: Recent Advances in Homomorphic Encryption: A Possible Future for Signal Processing in the Encrypted Domain. IEEE Signal Process. Mag. 30, 108–117 (2013). https://doi.org/10.1109/MSP.2012.2230219

South Ural State University
National Research University
1943

Литература

[11] Hrestak, D., Picek, S.: Homomorphic Encryption in the Cloud. In: 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO'14). pp. 1400–1404 (2014)

[12] Moore, C., O'Neill, M., Hanley, N., O'Sullivan, E.: Accelerating integer-based fully homomorphic encryption using Comba multiplication. In: IEEE Workshop on Signal Processing Systems, SiPS. pp. 1–6. IEEE (2014)

[13] Rondeau, T.: Data Protection in Virtual Environments (DPRIVE). (2020)

[14] Aono, Y., Hayashi, T., Trieu Phong, L., Wang, L.: Scalable and Secure Logistic Regression via Homomorphic Encryption. In: Proceedings of the Sixth ACM on Conference on Data and Application Security and Privacy - CODASPY '16. pp. 142–144. ACM Press, New York, New York, USA (2016). https://doi.org/10.1145/2857705.2857731.

[15] Bonte, C., Vercauteren, F.: Privacy-preserving logistic regression training. BMC Med. Genomics. 11, 86 (2018). https://doi.org/10.1186/s12920-018-0398-y.

[16] Kim, A., Song, Y., Kim, M., Lee, K., Cheon, J.H.: Logistic regression model training based on the approximate homomorphic encryption. BMC Med. Genomics. 11, 83 (2018).

[17] Cheon, J.H., Han, K., Kim, A., Kim, M., Song, Y.: A Full RNS Variant of Approximate Homomorphic Encryption. Presented at the (2019). https://doi.org/10.1007/978-3-030-10970-7_16.

[18] Cheon, J.H., Kim, D., Kim, Y., Song, Y.: Ensemble Method for Privacy-Preserving Logistic Regression Based on Homomorphic Encryption. IEEE Access. 6, 46938–46948 (2018).

[19] Yoo, J.S., Hwang, J.H., Song, B.K., Yoon, J.W.: A Bitwise Logistic Regression Using Bi-nary Approximation and Real Number Division in Homomorphic Encryption Scheme. Presented at the (2019). https://doi.org/10.1007/978-3-030-34339-2_2.

South Ural State University
National Research University
1943

Спасибо за внимание

Вопросы?

South Ural State University

National Research University

South Ural State University

School of Electronic Engineering and Computer Science

Problem-Oriented Cloud Computing Environment International Laboratory

Seminar

Privacy-Preserving Machine Learning as a Service

Speaker

Jorge Mario Cortés-Mendoza

Russia, December 2020.