СРАВНЕНИЕ РАБОТЫ КОДИРОВЩИКОВ, НАПИСАННЫХ С ПОМОЩЬЮ PYTORCH И CRYPTEN

COMPARISON OF THE WORK OF CODERS WRITTEN WITH PYTORCH AND CRYPTEN

УДК 004.056.55+ 003.26

Желанкина Анна Сергеевна, магистрант, Московский государственный университет имени Н.Э. Баумана, г. Москва
Кислов Павел Михайлович, магистрант, Московский государственный университет имени Н.Э. Баумана, г. Москва
Нурлыева Дана Джалилевна, магистрант, Московский государственный университет имени Н.Э. Баумана, г. Москва
Федорова Антонина Алексеевна, магистрант, Московский государственный университет имени Н.Э. Баумана, г. Москва

Zhelanlina A.S. nusja1298@mail.ru
Kislov P.M. pmk24031997@gmail.com
Nurlyeva D.D. danamanchester@mail.ru
Fedorova A.A. tonya.fedorova1604@gmail.com

Аннотация

В статье рассмотрены различные области использования личных данных людей. Обусловлена необходимость их кодирования с помощью криптографии. Также произведён анализ различных решений по кодированию данных с помощью различных методов криптографии. В статье рассмотрены способы применения библиотек машинного обучения PyTorch и CrypTen для кодирования информации. Были сформулированы правила выбора именно той информации, которую необходимо закодировать. Произведён сравнительный

анализ работы кодировщиков, написанных с помощью данных библиотек и произведена проверка их эффективности работы. Сформулированы направления дальнейших исследований в области применения библиотек машинного обучения для кодирования информации.

Annotation

The article discusses various areas of using people's personal data. The necessity of their coding using cryptography is determined. We also analyzed various solutions for data encoding using various cryptography methods. This article discusses how to use the PyTorch and CrypTen machine learning libraries to encode information. The rules for choosing exactly the information that need to be encoded were formulated. A comparative analysis of the work of encoders written with the help of these libraries was carried out and their efficiency was checked. Directions for further research in the field of using machine learning libraries for coding information are formulated.

Ключевые слова: машинное обучение, CrypTen, PyTorch, защита информации.

Keywords: machine learning, CrypTen, PyTorch, information security.

Введение

Сейчас любой человек не представляет своей жизни без использования технологий. Они окружают нас повсюду: дома, на работе, в транспорте и т.д. Некоторые учёные предсказывают, что всего лишь через пару десятков лет мир будет находиться в состоянии технологической сингулярности [1], то есть технологическое развитие станет настолько быстрым и неконтролируемым, что будут происходить радикальные и непредвиденные изменения человеческой жизни и цивилизации.

В настоящий момент времени же различные технические устройства сопровождают нас везде и всегда. Люди всегда носят с собой мобильные

телефоны, которые стали не просто способом связи по сотовой связи, но и хранилищем банковских карт, записной книжкой и прочим. Почти во всей технике, которая помогает создавать «умные дома» [2], упрощать ежедневную рутину, делать жизнь удобнее, хранится большое количество личной информации. Все эти умные технические устройства знают, где Вы живёте, ваши привычки, номер Вашего счёта, куда Вы любите ходить. Невозможно избежать поступления всё новой и новой информации, потому что жизнь не стоит на месте. Но встаёт необходимость её хранить и беречь, чтобы она не попала в руки злоумышленников. В целях сохранения неприкосновенности персональных данных установлены требования к обеспечению их безопасности на законодательном уровне [3]. Однако потеря конфиденциальных данных продолжает оставаться серьезной проблемой как для организаций, так и для отдельных лиц, чья информация может быть подвержена риску нарушения [4].

Таким образом, в современном мире большое значение имеет сохранность личных данных людей в любой области жизнедеятельности человека. На любом уровне передачи личных данных остро стоит необходимость строго следить за оборотом данных, их хранением и использованием в частных и коммерческих целях [5]. Часто в этих целях прибегают к маскированию данных. Маскирование данных - это процесс маскировки конкретных элементов данных в хранилищах данных [6].

Необходимо использовать новые методы машинного обучения, которые помогут работать с личной информацией таким образом, чтобы она находилась в безопасности. В статье предлагается сравнение работы двух кодировщиков, написанных с помощью разных фреймворков машинного обучения.

Обзор существующих решений

Одним из методов защиты информации, который широко применяется в наши дни, является криптография. Криптография представляет собой науку о

методах преобразования информации для защиты от нежелательного использования [7].

Известны различные подходы к классификации методов криптографического преобразования информации. По виду воздействия на исходную информацию методы криптографического преобразования информации могут быть разделены на четыре группы: шифрование, кодирование, сжатие, стенография [8].

Шифрование, как правило, обратимо и для его применения необходим математический алгоритм и ключ. При кодировании используются специальные таблицы и словари и с их помощью происходит замена смысловых конструкций исходной информации кодами. Сжатие работает на алгоритмах, обеспечивающих возможность обратного преобразования. Целью сжатия является сокращение объёма информации. В отличие от других методов криптографического преобразования информации, методы стеганографии позволяют скрыть не только смысл хранящейся или передаваемой информации, но и сам факт хранения или передачи закрытой информации. Изначально стенография была связана с разработкой методов скоростного письма [9].

Так как для данной задачи необходимо обеспечить полную защиту чувствительных данных, метод сжатия не является подходящим. Так как нет необходимости скрывать факт наличия закрытой информации, метод стенографии так же не подходит. Для решения нашей задачи из двух оставшихся методов: шифрование и кодирование - был выбран метод кодирования.

Предлагаемое решение

Анализ поставленной задачи показал, что на сегодняшний день существует огромное количество разнообразных методов кодирования информации, однако существующие методы не дают приемлемое качество их работы [10]. Исходя из этого, задачу кодирования информации можно

характеризовать как актуальную задачу, наилучшие результаты по решению которой показывают различные аналитические методы. Для реализации кодировщиков необходимо выбрать способ кодирования информации и подготовить данные для обработки.

Таким образом, подготовка множества данных состоит из следующих шагов:

1. Необходимо избавиться от избыточной информации. Также необходимо обработать пропуски в данных путём из замены на что-либо или на удаление данных параметров.

2. Нужно выбрать, какую именно информацию необходимо скрывать и кодировать, то есть что именно содержит персональные данные, которые не подлежат разглашению [11].

Следующим этапом является реализация простейших кодировщиков, написанных с помощью двух различных фреймворков: CrypTen и PyTorch. CrypTen является фреймворком машинного обучения для сохранения конфиденциальности и позволяет обучать и использовать модели с шифрованием данных [12]. PyTorch – это библиотека, написанная для работы с тензорами, которая может использовать как графические, так и центральный процессор [13]. Было решено исходные данные, которые необходимо закодировать, сложить с вектором такой же, как они, размерности из случайных чисел, а затем поделить на вектор также из случайных чисел. Данные были специально переформатированы в тензоры обеих библиотек для измерения времени работы кодировщиков.

В результате работы кодировщиков стало ясно, что кодировщик, написанный на PyTorch с использованием GPU работает быстрее, чем написанный на CrypTen, который может работать только на CPU. Но при работе с самими данными кодировщик на CrypTen начинает маскировать данные с самого начала обработки, что повышает безопасность персональной информации.

Все вышеперечисленные методы были опробованы на одинаковых данных. Полученные результаты позволили сделать выводы о том, что наиболее быстрым является метод кодирования с помощью PyTorch, но он не подходит для решения задачи из-за недостаточного сокрытия данных во время работы. Данный метод не кодирует обрабатываемую информацию во время работы, что не позволяет гарантировать полную её защиту.

Также для оценки работы кодировщиков были выбраны дополнительные критерии. Адаптивность – возможность дальнейшей работы с полученными тензорами с использованием разных библиотек для написания нейронных сетей. Так как существуют специальные кодировщики тензоров из Pytorch в тензоры для других библиотек – данный кодировщик является более адаптивным [14]. Реализация кодировщика проще в случае с PyTorch. Также PyTorch немного проще преобразовать под других вид входных данных, чем CrypTen.

В итоге была получена следующая таблица критериев:

Таблица 1. Сравнение критериев

Критерии сравнения	Весовой коэффициент	CrypTen	PyTorch
Скорость	0,333	0,1	1
Безопасность	0,333	1	0,6
Адаптивность	0,083	0,4	0,8
Простота реализации	0,167	0,6	0,8
Гибкость под разные данные	0,083	0,8	0,9

Для определения лучшего метода применим метод взвешенной суммы:

$Y_j = \sum_i^n a_i k_{ij}$, где наилучшим метод является метод $Y_l = \max_{j \in m} Y_j$ [15]

После подсчета взвешенных сумм получаем следующий ранжированный ряд кодировщиков данных:

1) PyTorch (Y = 0,8075)
2) CrypTen (Y = 0,5661)

Таким образом, для наиболее полной защиты персональных данных подходит кодировщик, написанный на CrypTen. Но если в процессе обработки информации безопасность не так важна, как закодированный результат, то наиболее быстрым является кодировщик с использованием PyTorch.

Направление дальнейших исследований

Дальнейшая работа по данной теме будет заключаться в расширении методов кодирования информации, проверки сочетаемости работы кодировщиков, написанных с помощью нескольких фреймворков одновременно и в создании более гибкой программной реализации, способного функционировать заданным образом.

Выводы

В данной работе была решена задача сравнения работы двух кодировщиков, написанных с помощью разных фреймворков машинного обучения. С помощью фреймворков машинного обучения: CrypTen, PyTorch - были определены время работы кодировщиков и её результат, что позволило сделать вывод о наиболее оптимальном методе. Для полной безопасности данных им оказался кодировщик, написанный с помощью CrypTen, а наиболее быстрым на PyTorch.

Литература

1. Васюгова С. А. Исследование перспектив и проблем интеграции человека с компьютером: искусственный интеллект, робототехника, технологическая сингулярность и виртуальная реальность / Васюгова С. А., Остроух А.В., Краснянский М. Н., Самаратунга А. // Перспектива науки. -2011 №4(19). - С. 109-114.

2. Чибисова И. С. Технологии умного дома в России / Чибисова И. С., Чибисов Е. Б. // Эпоха науки. -2018 №16. - С. 206-207.

3. Федеральный закон от 27 июля 2006 г. N 152-ФЗ "О персональных данных" //Российская газета. – 2006. - №165.

4. Фотопулос К. Управление катастрофической потерей чувствительных данных. – Elsevier, 2008.

5. Пыск Д. А. Проблемы защиты и обработки конфиденциальных персональных данных потребителей / Пыск Д. А. // Инновации и инвестиции. -2019 №3 – С. 118-120.

6. Арчана Р. А. Разработка архитектуры маскирования данных и анализ методов маскирования данных для тестирования / Арчана Р. А., Равикумар Д.К., Хегади Р.С. // International Journal of Engineering Science and Technology (IJEST) – 2011. - №6. – Vol.3

7. Иванова М.М. Криптография. Основные задачи / Иванова М.М. // Аллея науки. -2018. №8(24) – Том II. -С. 794 – 796.

8. Сладкова М. Ю. Основные понятия, термины и определения криптографии. Криптографические методы защиты информации / Сладкова М. Ю., Турмуханова Г. Б., Истомина Н. Г. // Актуальные проблемы современного образования. -2018 №1 (24). -С. 322 – 326.

9. В. Е. Фирстов Из истории стенографии и интерпретации закона Ципфа−Мандельброта в учебном процессе / В. Е. Фирстов, Р. А. Иванов // Ярославский педагогический вестник. – 2013 №4 - Том III. – С. 14 – 20.

10. Чекмарев В. В. Некоторые оценки методов кодирования свойств объектов / Чекмарев В. В., Чекмарев В. В. // Инфраструктурное обеспечение бизнес-процессов в региональных социально-экономических системах: сб. науч. тр. КГУ. – Кострома -2016, -С. 115-125.

11. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение.-СПб.: Питер, 2017.-336 с

12. CrypTen Documentation [Электронный ресурс]. – Электрон. дан. - URL: https://crypten.readthedocs.io/en/latest/ (дата обращения: 24.03.2020)

13. PYTORCH DOCUMENTATION [Электронный ресурс]. – Электрон. дан. – URL: https://pytorch.org/docs/stable/index.html (дата обращения: 24.03.2020)

14. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей// СПБ.: Питер. 2019.

15. Постников В. М., Спиридонов С. Б. Методы выбора весовых коэффициентов локальных критериев. М.: Изд-во: Национальный Электронно-Информационный Консорциум, 2015, с. 267-287

Literature

1. Vasyugova S. A. Issledovanie perspektiv i problem integratsii cheloveka s komp'yuterom: iskusstvennyy intellekt, robototekhnika, tekhnologicheskaya singulyarnost' i virtual'naya real'nost' / Vasyugova S. A., Ostroukh A.V., Krasnyanskiy M. N., Samaratunga A. // Perspektiva nauki. -2011 №4(19). - S. 109-114.

2. Chibisova I. S. Tekhnologii umnogo doma v Rossii / Chibisova I. S., Chibisov E. B. // Epokha nauki. -2018 №16. - S. 206-207.

3. Federal'nyy zakon ot 27 iyulya 2006 g. N 152-FZ "O personal'nykh dannykh" //Rossiyskaya gazeta. – 2006. - №165.

4. Fotopulos K. Upravlenie katastroficheskoy poterey chuvstvitel'nykh dannykh. – Elsevier, 2008.

5. Pysk D. A. Problemy zashchity i obrabotki konfidentsial'nykh personal'nykh dannykh potrebiteley / Pysk D. A. // Innovatsii i investitsii. -2019 №3 – S. 118-120.

6. Archana R. A. Razrabotka arkhitektury maskirovaniya dannykh i analiz metodov maskirovaniya dannykh dlya testirovaniya / Archana R. A., Ravikumar D.K., Khegadi R.S. // International Journal of Engineering Science and Technology (IJEST) – 2011. - №6. – Vol.3

7. Ivanova M.M. Kriptografiya. Osnovnye zadachi / Ivanova M.M. // Alleya nauki. -2018. №8(24) – Tom II. -S. 794 – 796.

8. Sladkova M. Yu. Osnovnye ponyatiya, terminy i opredeleniya kriptografii. Kriptograficheskie metody zashchity informatsii / Sladkova M. Yu., Turmukhanova G. B., Istomina N. G. // Aktual'nye problemy sovremennogo obrazovaniya. -2018 №1 (24). -S. 322 – 326.

9. V. E. Firstov Iz istorii stenografii i interpretatsii zakona Tsipfa−Mandel'brota v uchebnom protsesse / V. E. Firstov, R. A. Ivanov // Yaroslavskii pedagogicheskii vestnik. – 2013 №4 - Tom III. – C. 14 – 20.

10. Chekmarev V. V. Nekotorye otsenki metodov kodirovaniya svoystv ob"ektov / Chekmarev V. V., Chekmarev V. V. // Infrastrukturnoe obespechenie biznes-protsessov v regional'nykh sotsial'no-ekonomicheskikh sistemakh: sb. nauch. tr. KGU. – Kostroma -2016, -S. 115-125.

11. Brink Kh., Richards D., Feverolf M. Mashinnoe obuchenie.-SPb.: Piter, 2017.-336 s

12. CrypTen Documentation [Elektronnyy resurs]. – electronic data. - URL: https://crypten.readthedocs.io/en/latest/ (data obrashcheniya: 24.03.2020)

13. PYTORCH DOCUMENTATION [Elektronnyy resurs]. – electronic data. – URL: https://pytorch.org/docs/stable/index.html (data obrashcheniya: 24.03.2020)

14. Nikolenko S., Kadurin A., Arkhangel'skaya E. Glubokoe obuchenie. Pogruzhenie v mir neyronnykh setey// SPB.: Piter. 2019.

15. Postnikov V. M., Spiridonov S. B. Metody vybora vesovykh koeffitsientov lokal'nykh kriteriev. M.: Izd-vo: Natsional'nyy Elektronno-Informatsionnyy Konsortsium, 2015, s. 267-287