Сергей В. Запечников^{1,2}

¹Национальный исследовательский ядерный университет «МИФИ»,
Каширское ш., 31, Москва, 115409, Россия

²Центр развития криптовалют и цифровых финансовых активов ВИНТИ РАН,
Усиевича ул., 20, Москва, 125190, Россия
e-mail: SVZapechnikov@mephi.ru, <https://orcid.org/0000-0002-7975-6040>

МОДЕЛИ И АЛГОРИТМЫ КОНФИДЕНЦИАЛЬНОГО МАШИННОГО ОБУЧЕНИЯ

DOI: <http://dx.doi.org/10.26583/bit.2020.1.05>

Аннотация. Статья посвящена недавно возникшей научной проблеме обеспечения конфиденциальности в машинном обучении. Актуальность проблемы определяется возрастающими потребностями в использовании методов машинного обучения для персональных данных, а также данных, составляющих коммерческую, медицинскую, финансовую и иные охраняемые законом виды тайны. Цели работы заключаются в систематизации моделей обеспечения безопасности машинного обучения, выявлении алгоритмических инструментов, которые могут быть использованы для обеспечения конфиденциальности в процессе обучения и применения моделей, сравнительном анализе систем конфиденциального машинного обучения. В статье приводятся основные понятия и определения, связанные с машинным обучением, дается систематизация задач машинного обучения и методов их решения, отмечаются современные и перспективные направления развития машинного обучения. Среди задач машинного обучения выделяются те, для которых актуально обеспечение конфиденциальности данных обучающей, тестовой и рабочей выборок. Решаемым задачам сопоставляются специальные криптографические методы и протоколы. Дается краткая характеристика известных из литературы систем конфиденциального машинного обучения, при этом отмечаются поддерживаемые ими методы машинного обучения, тип нарушителя, которому может противостоять система, используемые в реализации криптографические примитивы. Рассматриваются нерешенные проблемы в сфере конфиденциального машинного обучения и перспективы развития этой научной области.

Ключевые слова: интеллектуальный анализ данных, машинное обучение, глубокое обучение, федеративное обучение, адаптивное обучение, конфиденциальность, безопасные многосторонние вычисления, схемы разделения секрета, гомоморфное шифрование.

Для цитирования: ЗАПЕЧНИКОВ, Сергей В. МОДЕЛИ И АЛГОРИТМЫ КОНФИДЕНЦИАЛЬНОГО МАШИННОГО ОБУЧЕНИЯ. Безопасность информационных технологий, [S.l.], v. 27, n. 1, p. 51-67, 2020. ISSN 2074-7136. Доступно на: <<https://bit.mephi.ru/index.php/bit/article/view/1251>>. Дата доступа: 10 feb. 2020. DOI: <http://dx.doi.org/10.26583/bit.2020.1.05>.

Sergey V. Zapechnikov^{1,2}

¹National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Kashirskoye shosse, 31, Moscow, 115409, Russia

²Research Center for Cryptocurrencies and Digital Assets,
Usievicha str., 20, Moscow, 125190, Russia

e-mail: SVZapechnikov@mephi.ru, <https://orcid.org/0000-0002-7975-6040>

Models and algorithms of privacy-preserving machine learning

DOI: <http://dx.doi.org/10.26583/bit.2020.1.05>

Abstract. The paper is devoted to the recent scientific problem of privacy-preserving machine learning. The problem actuality is determined by the growing need to use machine learning for personal data, as well as for data that make up commercial, medical, financial and other types of information protected by law. The aim of the study is to systematize the security models of machine learning, to identify algorithmic tools that can be used to ensure the privacy of the learning process and application of models, as well as to analyze the privacy-preserving machine learning systems. The paper presents the major concepts and definitions related to machine learning, provides a systematization of machine learning

problems and methods of their solution, and pays attention to the modern and promising areas of development of machine learning. Among the tasks of machine learning are those for which it is important to ensure the privacy of data from training, test and work samples. Special cryptographic methods and protocols are correlated to the problems solved. A brief description of the known privacy-preserving machine learning systems is given. The machine learning methods supported by these systems, as well as the type of adversary that the system can resist, and the cryptographic primitives used for the implementation are described in the paper. Unsolved problems in the field of privacy-preserving machine learning and prospects for the development of this scientific field are considered.

Keywords: data mining, machine learning, deep learning, federated learning, transfer learning, privacy, secure multi-party computations, secret sharing scheme, homomorphic encryption.

For citation: ZAPECHNIKOV, Sergey V. Models and algorithms of privacy-preserving machine learning. IT Security (Russia), [S.l.], v. 27, n. 1, p. 51-67, 2020. ISSN 2074-7136. Available at: <<https://bit.mephi.ru/index.php/bit/article/view/1251>>. Date accessed: 10 feb. 2020. DOI: <http://dx.doi.org/10.26583/bit.2020.1.05>.

Введение

Машинное обучение, возникшее как самостоятельная научная дисциплина на стыке интеллектуального анализа данных, математической статистики и методов оптимизации, в последние годы стала одним из основных трендов в сфере компьютерных наук. Главными движущими факторами развития машинного обучения стали, во-первых, колоссальный рост вычислительных мощностей компьютерной техники, а во-вторых, взрывообразное увеличение объёма доступных для анализа информационных ресурсов – явление, получившее наименование *больших данных* (big data). Классические логико-алгебраические, численные, статистические, графовые и иные методы анализа данных подразумевают фиксированный алгоритм обработки и позволяют получать обобщенные характеристики массива данных, например, выборку объектов по условию, находить минимумы и максимумы, средние значения, дисперсию, расстояние, кратчайший путь и пр.

В новых условиях появляется потребность извлекать знания из доступных данных, т.е. не просто находить ответы на серию одиночных вопросов, считать точечные характеристики, а обрабатывать данные системно, извлекая из них новые закономерности, и затем уже на основе этих закономерностей находить ответы на новые, заранее неизвестные вопросы. Такие задачи получили название *интеллектуального анализа данных* (data mining). Некоторые из них могут успешно решаться комбинаторными, вероятностными, графовыми и иными методами. Так, например, можно отыскивать пересечения множеств объектов, т.е. общее подмножество двух множеств, часто встречающиеся в базе данных подмножества объектов (frequent itemset) и ассоциативные правила.

Однако подавляющее большинство задач интеллектуального анализа данных не может быть решено без *машинного обучения* (machine learning). К ним, прежде всего, относятся кластерный анализ, классификация объектов, поиск аномалий, ранжирование. Суть машинного обучения заключается в том, что для анализа данных используется не фиксированный алгоритм, а настраиваемый шаблон алгоритма, который изначально назначается экспертом, а затем подстраивается под конкретные задачи, типы и массивы данных. В простых случаях такой шаблон может представлять собой параметрическое семейство функций (например, полиномы не выше заданной степени), в котором настраиваемыми параметрами являются коэффициенты полинома. В сложных случаях шаблоном является некоторая конструкция вычислительной структуры (например, искусственной нейронной сети), а настраиваемыми параметрами – связи между нейронами и их веса. Алгоритм, получившийся в результате настройки шаблона, принято

называть *моделью*. Процесс машинного обучения всегда включает в себя два этапа: сначала, чтобы из шаблона получить модель, её нужно обучить на некоторой части доступных данных – *обучающей выборке*, затем обученная модель сможет применяться для анализа новых данных, ранее никогда не обрабатывавшихся этой моделью – первоначально *тестовой выборке* (для оценки адекватности и качества модели), а затем и *рабочей выборке* (для получения целевого результата). Процесс машинного обучения, конечно, не является самоцелью – в конечном итоге результаты анализа данных используются для принятия решений в бизнесе, госуправлении, научных исследованиях и других сферах деловой деятельности.

Современная теория и практика машинного обучения включает в себя как уже ставшие традиционными разделы: обучение с учителем (supervised learning) и обучение без учителя (unsupervised learning), – так и новые, такие как глубокое обучение (deep learning), обучение с подкреплением (reinforcement learning), федеративное обучение (federating learning), обучение передаче опыта (transfer learning).

Однако до недавних пор машинное обучение обходило стороной вопросы обеспечения информационной безопасности на этапах обучения и применения модели. Невнимание к вопросам информационной безопасности объясняется тем, что преобладающими конфигурациями применения машинного обучения до сих пор были локальные или клиент-серверные приложения.

В современных условиях машинное обучение используется повсеместно, методы становятся всё более сложными, объёмы обучающих, тестовых и рабочих выборок выросли на несколько порядков величины, что обусловило переход к распределённым конфигурациям с участием множества серверов и большого числа клиентов. Не редкость – тот случай, когда данные для анализа собираются из множества источников, и при этом они являются интеллектуальной собственностью разных лиц, а порою среди них могут быть персональные данные, а также сведения, составляющие коммерческую, финансовую, медицинскую тайну и иные охраняемые законодательством виды тайны. Тем не менее, владельцы, обладатели, операторы этих данных прямо или косвенно заинтересованы в использовании их в качестве обучающих, тестовых и рабочих выборок. Следует отметить, что к настоящему времени в целом ряде стран уже приняты или готовятся к вступлению в силу нормативные правовые акты, регламентирующие использование таких данных. В качестве примеров можно привести GDPR (The General Data Protection Regulation) в Евросоюзе, CCPA (California Consumer Privacy Act) в США, Закон о кибербезопасности в Китае. Таким образом, описанная ситуация во многих странах подпадает под нормативно-правовое регулирование. В связи с этим актуальной является проблема обеспечения безопасности информации, используемой в процессе машинного обучения для обучения, тестирования и применения моделей, и в первую очередь, в аспекте конфиденциальности.

1. Модели машинного обучения

Основными объектами нашего внимания являются процессы обучения и применения моделей. Для удобства их описания начнем с формальной модели машинного обучения.

Принципиально важной стороной машинного обучения является признаковое описание объектов. Считается, что каждый объект анализа описывается некоторым множеством *признаков* (features). Выделяют несколько типов признаков:

- номинальные – признаки, которые задаются путем перечисления всех возможных значений, и этот список является закрытым;

- бинарные – частный случай номинальных, когда значения могут быть только двух типов, условно $\{0, 1\}$;
- порядковые – признаки, которые могут принимать конечное множество значений, и на этом множестве введено отношение порядка;
- числовые – признаки, которые измеряются действительными числами.

Все признаки делятся на подмножество признаков, известных для всех объектов без исключения, и подмножество признаков, известных только для объектов обучающей выборки, – *ответов*. Для объектов тестовой и рабочей выборок их нужно спрогнозировать, используя обученную модель.

Пусть X – множество объектов, Y – множество ответов, $y: X \rightarrow Y$ – неизвестная зависимость. Тогда задача машинного обучения в самом общем виде формулируется следующим образом [1]. Дано: $\{x_1, \dots, x_l\} \subset X$ – обучающая выборка, $y_i = y(x_i)$, $i = \overline{1, l}$ – известные ответы, l – количество объектов обучающей выборки. Найти: $a: X \rightarrow Y$ – алгоритм, или решающую функцию, приближающую y на всем множестве X .

Как отмечалось выше, самый распространенный способ задания объектов – это их признаковое описание: $f_j: X \rightarrow D_j$, где f_j , $j = \overline{1, n}$ – признаки объектов, n – количество признаков объекта, D_j – области значений признаков, причём $D_j = \{0, 1\}$, если f_j – бинарный признак, $|D_j| < \infty$, если f_j – номинальный признак, $|D_j| < \infty$, и на D_j введено отношение порядка, если f_j – порядковый признак, $D_j = \mathbb{R}$, если f_j – количественный признак. Вектор $(f_1(x), \dots, f_n(x))$ – это признаковое описание объекта x . Задача обучения по прецедентам имеет дело с матрицей «объекты – признаки» (feature data):

$$F = \|f_j(x_i)\|_{l \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_l) & \dots & f_n(x_l) \end{pmatrix}.$$

1. Этап обучения модели заключается в том, что метод обучения $\mu: (X \times Y)^l \rightarrow A$ по выборке $X^l = (x_i, y_i)_{i=1}^l$ строит алгоритм $a = \mu(X^l)$:

$$\left[\begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_l) & \dots & f_n(x_l) \end{pmatrix} \xrightarrow{y} \begin{pmatrix} y_1 \\ \dots \\ y_l \end{pmatrix} \right] \xrightarrow{\mu} a.$$

2. Этап применения модели состоит в том, что алгоритм a для новых объектов $x'_1, \dots, x'_k$ выдает ответы $a(x'_i)$:

$$\begin{pmatrix} f_1(x'_1) & \dots & f_n(x'_1) \\ \dots & \dots & \dots \\ f_1(x'_k) & \dots & f_n(x'_k) \end{pmatrix} \xrightarrow{a} \begin{pmatrix} a(x'_1) \\ \dots \\ a(x'_k) \end{pmatrix}.$$

Чтобы оценить качество алгоритма a , т.е. близость a к y , используются функционалы качества. Для точечной оценки используется функция потерь $\mathcal{L}(a, x)$ – величина ошибки алгоритма $a \in A$ на объекте $x \in X$, для интегральной оценки – эмпирический риск

$$Q(a, X^l) = \frac{1}{l} \sum_{i=1}^l \mathcal{L}(a, x_i),$$

который служит функционалом качества алгоритма a на X^l .

Таким образом, цель машинного обучения можно сформулировать как минимизацию эмпирического риска:

$$\mu(X^l) = \arg \min_{a \in A} Q(a, X^l).$$

Рассматриваемая модель удобна тем, что она позволяет легко определить все важнейшие частные задачи машинного обучения.

Обучение с учителем полностью соответствует описанной модели, при этом $Y \neq \emptyset$. На практике обучению с учителем соответствуют задачи классификации, восстановления регрессии, отчасти – обнаружения аномалий и ранжирования.

Обучение без учителя – это ситуация, когда $Y = \emptyset$, т.е. ответы для объектов обучающей выборки неизвестны. В этом случае метод обучения μ должен сгруппировать объекты по какому-то принципу, который формируется в процессе обучения. На практике обучению без учителя соответствуют задачи кластеризации, отчасти – обнаружения аномалий и ранжирования.

Разумеется, объекты окружающего мира можно описать бесконечным множеством признаков, кроме того, в реальных задачах объекты часто бывают заданы не готовыми признаками, а «сырыми» данными, из которых ещё нужно выделить признаки. Во всех вышеупомянутых случаях речь идет лишь о тех признаках, которые имеют существенное значение для решаемой задачи. Для сложных объектов не всегда легко определить, какие признаки существенны. В традиционном машинном обучении для этого применяются различные приемы сокращения размерности признакового пространства, в частности, метод главных компонент. Однако с ростом сложности объектов анализа наступает такая ситуация, когда становится затруднительным предугадать, а тем более выделить какие-либо признаки объекта, которые будут иметь существенное значение для решаемой задачи. Необходимость разрешения этой проблемы породила новую концепцию машинного обучения – *глубокое обучение* – такой подход к реализации процедур машинного обучения, при котором существенные признаки формируются автоматически в процессе обучения модели [2]. Невозможно четко разграничить традиционное и глубокое обучение. Методы глубокого обучения развивались постепенно, их первоосновой стали искусственные нейронные сети. Попытки выстроить архитектуры нейронных сетей в максимальном подобии с открытыми физиологами механизмами работы центральной нервной системы человека и высших животных позволили нащупать механизмы автоматического формирования признаков в процессе обучения, которые и стали основой нового направления. Особенно успешно методы глубокого обучения применяются в компьютерном зрении (анализ изображений) и обработке естественных языков (анализ и синтез текстов, машинный перевод, диалоговые системы).

Федеративное обучение – такая модель машинного обучения, при которой данные, описывающие либо объекты, либо признаки, либо и то, и другое сразу, находятся во владении разных лиц и, чаще всего, пространственно распределены [3]. Абстрагируясь на время от вопросов информационной безопасности, возможно выделить разные архитектуры систем федеративного машинного обучения [4]:

- небольшое число участников, разделенные между ними данные, большие вычислительные ресурсы у каждого из них, все вычисления выполняются самими участниками в ходе выполнения протоколов, отсутствует третья сторона, которой передавались бы данные для вычислений, – это модель многосторонних вычислений;
- большое число клиентов, малое число обрабатывающих центров, данные поставляются от клиентов, обучение и применение моделей осуществляется в центрах обработки данных.

Поскольку процедуры обучения, в особенности, глубокого обучения, очень ресурсоемки, в обоих случаях обрабатывающие центры нередко виртуализируются, что

означает перенос модели в облака и физическое отделение источников данных, хранилищ данных и места размещения обучаемой (обученной) модели. В зависимости от того, каким образом между участниками системы федеративного обучения разделяются данные обучающихся (тестовых, рабочих) выборок, выделяют два частных случая. Рассмотрим их на примере разделения выборок между двумя участниками: A и B . Пусть $1 < i < j < n$, $1 < p < q < l$, участнику A известна матрица признаков

$$\begin{pmatrix} f_1(x_1), \dots, f_i(x_1), \dots, f_j(x_1) \\ \vdots \\ f_1(x_p), \dots, f_i(x_p), \dots, f_j(x_p) \\ \vdots \\ f_1(x_q), \dots, f_i(x_q), \dots, f_j(x_q) \end{pmatrix},$$

участнику B – матрица признаков

$$\begin{pmatrix} f_i(x_p), \dots, f_j(x_p), \dots, f_n(x_p), y(x_p) \\ \vdots \\ f_i(x_q), \dots, f_j(x_q), \dots, f_n(x_q), y(x_q) \\ \vdots \\ f_i(x_l), \dots, f_j(x_l), \dots, f_n(x_l), y(x_l) \end{pmatrix},$$

Тогда:

- при *горизонтальном федеративном обучении* есть некоторое достаточно большое подмножество признаков, известных обо всех объектах выборки, т.е. обучение модели происходит на матрице

$$\begin{pmatrix} f_i(x_1), \dots, f_j(x_1), \emptyset \\ \vdots \\ f_i(x_p), \dots, f_j(x_p), y(x_p) \\ \vdots \\ f_i(x_l), \dots, f_j(x_l), y(x_l) \end{pmatrix};$$

- при *вертикальном федеративном обучении* есть некоторое достаточно большое подмножество объектов, таких что в совокупности все их признаки известны поставщикам данных, т.е. обучение модели происходит на матрице

$$\begin{pmatrix} f_1(x_p), \dots, f_n(x_p), y(x_p) \\ \vdots \\ f_1(x_q), \dots, f_n(x_q), y(x_q) \end{pmatrix}.$$

Вертикальное федеративное обучение на практике встречается гораздо чаще. В настоящее время предпринимаются активные усилия по международной стандартизации федеративного машинного обучения [5].

Наконец, возможна еще более сложная ситуация, когда данные распределены между поставщиками таким образом, что нет ни достаточного количества объектов, ни достаточного количества признаков для качественного обучения модели ни по «вертикальному», ни по «горизонтальному» сценарию. Тогда приходим к модели, которую называют *обучением передаче опыта* (transfer learning). Её реализации основаны на глубоком обучении: многослойная нейронная сеть первоначально обучается на одной из выборок, а для решения задач на новых выборках берется лишь часть входных слоев обученной сети, отсеченная же часть выходных слоев заново дообучается на этих выборках. Таким образом, происходит «передача опыта», заключенного в параметрах входных слоев обученной сети.

2. Модели обеспечения безопасности машинного обучения

Перечисленная выше типология систем машинного обучения порождает большое количество ситуаций, в которых требуется обеспечение безопасности информации, большое число подходов к решению этой задачи. И наоборот, для решения задачи могут выдвигаться требования к архитектурам систем машинного обучения. По этим причинам не будем пытаться составить каталог всех возможных моделей безопасности, а лишь перечислим те факторы, выбор которых определяет «лицо» модели.

Первая важная часть нашей условной модели безопасности – это модель угроз, т.е. в данном случае номенклатура принимаемых во внимание потенциально опасных для системы машинного обучения внешних факторов:

- разглашение конфиденциальных данных, передаваемых между участниками системы;
- манипуляции данными, поступающими на вход моделей;
- частичная или полная реконструкция объекта анализа по признаковому описанию;
- деанонимизация субъекта по (возможно, обезличенным) персональным (медицинским, финансовым и т.п.) данным и иным признакам, используемым для обучения или тестирования модели;
- установление членства (или отсутствия членства) объекта с фиксированным набором признаков в выборке;
- инверсия обученной модели, т.е. восстановление значений признаков, на которых была обучена модель;
- для методов обучения, основанных на градиентном спуске, – восстановление объектов выборки по небольшой части градиентов и другие факторы.

Вторая часть – это категории нарушителей, которым будет противостоять система. Самый важный фактор – предположение о нарушителе либо как об условно пассивном, но «любопытном» (curious), либо как об активном. Условно пассивный нарушитель характеризуется теми же поведенческими признаками, что и классический пассивный нарушитель, т.е. может лишь прослушивать каналы связи и собирать все передаваемые по ним в открытом либо зашифрованном виде данные, а также собирать все открытые, общедоступные данные об участниках. Однако, когда мы говорим о системах машинного обучения, такому противнику предоставляется дополнительная возможность применять к собранным данным любые доступные ему методы и средства интеллектуального анализа данных, включая машинное обучение, и пользоваться в целях совершения атаки любыми полученными таким образом результатами. Активный нарушитель понимается в привычном смысле как нарушитель, имеющий возможность произвольным образом вмешиваться в ход выполнения протокола путем вставки, удаления, модификации сообщений в каналах связи, а также посредством компрометации некоторых легитимных участников протокола. По последнему из этих признаков возможна дальнейшая классификация: какие именно участники (клиенты, обрабатывающие центры) могут быть скомпрометированы, в каком количестве и т.д. Получается целое семейство моделей.

В рамках настоящей статьи, однако, не ставится цель охватить все аспекты информационной безопасности машинного обучения. Далее остановимся только на аспектах безопасности, связанных с конфиденциальностью информации. Из краткого обзора моделей машинного обучения заметно, что обеспечение безопасности информации становится актуальным для систем машинного обучения, как только в процессе обучения участвуют две или более стороны, в частности, либо клиенты и обрабатывающие центры, либо несколько одноранговых участников. Именно поэтому на ранних этапах развития

машинного обучения, когда оно мыслилось как локальный процесс, проблема конфиденциальности не рассматривалась.

Иными словами, проблема конфиденциальности возникает тогда, когда не совпадают по крайней мере две из трех ролей участников процесса машинного обучения: источника данных, вычислителя и получателя результатов. Как известно, основным инструментом обеспечения конфиденциальности в распределенных компьютерных системах является криптография.

Разумеется, каждое конкретное решение требует применения стандартной для современной доказательной криптографии схемы:

- формулирования определений безопасности в условиях решаемой задачи;
- выдвижения предположений о вычислительной сложности решения некоторых задач;
- конструирования криптографических алгоритмов, протоколов, схем;
- формулирования утверждений (теорем) о свойствах предложенных конструкций;
- доказательства утверждений (теорем) о соответствии свойств конструкций определениям.

Ввиду большого многообразия ситуаций, в которых требуется обеспечение конфиденциальности, опять-таки не будем стремиться перечислить ни все частные модели обеспечения конфиденциальности, ни пытаться строить какую-либо единую «сверхмодель». Ограничимся примерами постановок задач обеспечения конфиденциальности для упомянутых выше частных случаев машинного обучения.

При рассмотрении *горизонтального федеративного обучения* в большинстве случаев предполагается централизованная модель вычислений. Здесь условно пассивным нарушителем может быть сервер, а клиенты являются честными [6]. Требуется обеспечить неразглашение конфиденциальных данных «любопытному» серверу. В конце обучения модель и все ее параметры сообщаются всем участникам в одинаковом виде.

При рассмотрении *вертикального федеративного обучения* обычно предполагается модель многосторонних вычислений, при этом условно пассивными нарушителями может быть часть участников (очевидно, составляющая менее половины от общего числа участников) [6]. В самом простом случае с двумя участниками, например, предполагается, что они не состоят в сговоре между собой и не более чем один из них может быть скомпрометирован. Требуется, чтобы нарушитель не смог узнать какие-либо данные нескомпрометированного участника, за исключением общедоступных входных и выходных данных. В конце обучения каждый участник получает параметры обученной модели, обусловленные только переданной им самим частью обучающей выборки. Чтобы получить модель, обученную на полной выборке, участники должны дополнительно выполнить специальный протокол, а это будет невозможно, если компрометация обнаружена.

Случай *обучения передаче опыта* во многом аналогичен модели вертикального федеративного обучения [7].

Для обеспечения конфиденциальности в каждом из обозначенных случаев (и во всех остальных, которые могут иметь практическое значение) требуется создание специальных систем конфиденциального машинного обучения, включающих в себя комплект криптографических алгоритмов и протоколов. Примеры таких систем будут приведены ниже.

Архитектура и алгоритмическое наполнение этих систем в первую очередь зависят от метода машинного обучения, для которого обеспечивается защита. Однако системы

конфиденциального машинного обучения даже для одного метода могут строиться с использованием разных алгоритмических инструментов. При этом следует отметить, что все системы конфиденциального машинного обучения – не что иное как частные случаи еще более общей задачи – криптографической защиты процессов обработки данных [8, 9].

3. Инструменты конфиденциального машинного обучения

В качестве строительных блоков криптографических протоколов, реализующих конфиденциальное машинное обучение, используется целый ряд криптографических конструкций и концепций, среди них:

- безопасные многосторонние вычисления;
- GC-схемы (англ. garbled circuits);
- ОТ-протокол (англ. oblivious transfer), или протокол «забывающей» передачи;
- схемы разделения секрета;
- гомоморфное шифрование.

Далее кратко охарактеризуем их, опираясь на материалы [8, 10]. *Безопасные многосторонние вычисления* (англ. secure multi-party computations) – одна из фундаментальных задач в области криптографических протоколов. Постановка задачи и фундаментальные основы её решения известны из [11, 12] с середины 1980-х гг. Рассматривается множество из n участников криптосистемы. Каждый из участников обладает секретом $x_i, i = \overline{1, n}$. Цель выполнения протокола – в том, чтобы участники своими совместными действиями без привлечения доверенной третьей стороны смогли вычислить некоторую функцию F , аргументами которой являются их секреты, не разглашая при этом секреты друг другу и третьим лицам. Таким образом, протокол должен обладать двумя основными свойствами [13]:

- корректностью: каждый из участников должен получить значение $y = F(x_1, \dots, x_n)$, значения y должны совпасть у всех участников;
- конфиденциальностью: никакие иные сведения, кроме значения y , не будут разглашены в процессе выполнения протокола.

Для модели пассивного нарушителя классическими протоколами безопасных многосторонних вычислений считаются GC-схемы (англ. garbled circuits) [11]. Это двусторонний протокол для булевых функций, но он служит «строительным блоком» для множества других протоколов, в том числе и многосторонних. Широко известен протокол GMW [12], который решает эту задачу для многостороннего случая и применим как к булевым функциям, так и к арифметическим схемам. Выполняется он за количество раундов, пропорциональное глубине схемы. Альтернативным подходом может служить протокол BGW [14]. Есть и множество других подходов, отличающихся ограничениями, налагаемыми на вычисляемую функцию, числом раундов, вычислительной сложностью, различными нюансами модели нарушителя. Для модели активного нарушителя также известен целый ряд протоколов на основе GC-схем и не только, например, протокол, предложенный в [15].

Для реализации GC-схем, в свою очередь, необходим такой криптографический примитив как ОТ-протокол, или *протокол «забывающей» передачи*. По определению [16], это криптографический протокол, который выполняется двумя участниками: отправителем S и получателем R , где отправитель имеет два секрета x_0, x_1 , а получатель выбирает бит $b \in \{0, 1\}$. Протокол позволяет R получить x_b , не получая никаких сведений о другом секрете x_{1-b} . В то же самое время S не получает никаких знаний о том, какой именно из двух секретов выбрал R .

Следующий инструмент, используемый для обеспечения безопасности процедур

машинного обучения, – *схемы разделения секрета* (CPC) (англ. SS – secret sharing). Предположим, есть важная секретная информация, которую можно потерять, поэтому опасно доверять кому-то одному. Рассмотрим пути повышения надежности и безопасности ее хранения. Первый путь – сделать несколько копий этих данных и хранить их в разных местах. Резервирование обеспечивает высокую надежность хранения, но если скомпрометирована хотя бы одна копия, то секретность всей информации будет потеряна. Второй путь – разделить секрет на несколько частей и хранить их в разных местах, при необходимости собирая вместе. Такая схема обеспечивает высокую конфиденциальность (чтобы восстановить секрет, надо получить все его доли), но низкую надежность (если потеряна хотя бы одна доля, восстановить секрет уже будет невозможно). Желательно построить более гибкую схему. Пусть есть n участников криптосистемы. Мы хотим, чтобы любые t из них могли восстановить секрет, но никакие $t-1$ из них не смогли бы получить информацию о секрете. Число t ($t < n$) – параметр схемы, называемый порогом. Схема, обладающая такими свойствами, называется (t, n) -пороговой CPC. Её достоинство в том, что если кто-то из участников потеряет или разгласит свою долю секрета или не будет участвовать в восстановлении секрета, секрет все равно можно будет восстановить и сохранить его конфиденциальность.

Формально говоря, пусть D – область определения секрета, D_1 – область определения долей секрета, пусть $Shr: D \rightarrow D_1^n$ – (возможно, вероятностный) алгоритм разделения секрета, $Rec: D_1^k \rightarrow D$ – алгоритм восстановления секрета. (t, n) -пороговой CPC называется пара алгоритмов (Shr, Rec) , обладающая следующими двумя свойствами:

- корректностью: пусть $(s_1, s_2, \dots, s_n) = Shr(s)$, тогда

$$Pr[\forall k \geq t, Rec(s_{i_1}, \dots, s_{i_k}) = s] = 1;$$

- совершенной секретностью: любое подмножество из менее чем t долей секрета не разглашает никакой информации о секрете в теоретико-информационном смысле, т.е. для любых двух секретов $a, b \in D$ и любого допустимого вектора долей секрета $\vec{v} = v_1, v_2, \dots, v_k$, такого что $k < t$,

$$Pr[\vec{v} = Shr(a)|_k] = Pr[\vec{v} = Shr(b)|_k],$$

где символ $|_k$ обозначает проекцию в k -мерное подпространство.

Ещё один инструмент – *гомоморфное шифрование* – позволяет реализовать концепцию аутсорсинговых вычислений:

- одна из сторон протокола служит источником данных, и ей требуется получить результат вычисления некоторой функции на этих данных, не разглашая сами данные;
- вторая сторона получает эти данные от источника в зашифрованном виде, выполняет требуемые вычисления и возвращает зашифрованные результаты источнику, при этом не приобретая никаких знаний ни об исходных данных, ни о промежуточных результатах вычислений, ни об окончательном результате. Только источник данных, получив зашифрованный результат, может расшифровать его и получить открытый текст результата вычислений.

Выделяют два типа таких схем – частично и полностью гомоморфные. Частично гомоморфные схемы (англ. PHE – partly homomorphic encryption) позволяют реализовать аутсорсинговые вычисления только в отношении какой-либо одной арифметической операции (обычно сложения либо умножения). Однако практическое применение частично гомоморфных схем ограничено лишь теми задачами, которые могут быть сведены к вычислению функции, которая может быть выражена в терминах

поддерживаемой операции. Тем не менее, для задач машинного обучения в большинстве случаев их оказывается достаточно.

Полностью гомоморфное шифрование (англ. FHE – fully homomorphic encryption) обеспечивает требуемые свойства для любой функции, представимой в виде арифметической схемы. Формально схема полностью гомоморфного шифрования определяется как совокупность четырех алгоритмов (*Gen*, *Enc*, *Eval*, *Dec*) [17], где *Gen*, *Enc*, *Dec* соответствуют стандартной асимметричной схеме шифрования (генерация ключей, зашифрование и расшифрование), а алгоритм *Eval* принимает на вход вектор шифртекстов и функцию F , а выдает зашифрованный результат применения этой функции к вектору шифртекстов. При этом обеспечивается следующее свойство корректности:

$$Dec_{sk} \left(Eval \left(F, Enc_{pk}(x_1), \dots, Enc_{pk}(x_n) \right) \right) = F(x_1, \dots, x_n),$$

где $(pk, sk) \leftarrow Gen(1^k)$, k – параметр безопасности (длина секретного ключа схемы).

Практическая значимость полностью гомоморфных схем шифрования определяется тем, что они позволяют реализовать модель вычислений «вслепую», т.е. сторона вычислительного процесса, которой доверена обработка данных, сможет выполнять любые вычислительные операции, никогда не видя обрабатываемые данные.

Перечисленные теоретические принципы и концепции далее должны быть применены к реализации процедур машинного обучения.

4. Реализация систем конфиденциального машинного обучения

Рассмотрим примеры известных из литературы реализаций систем конфиденциального машинного обучения. Все известные на текущий момент реализации являются прототипами – систем, доведенных до стадии программного продукта, пока в мире нет. Сразу оговоримся, что автор не ставил целью провести полномасштабный обзор работ в этой области, а тем более проанализировать используемый в каждой из них математический аппарат – это не представляется возможным в рамках одной статьи. Задача состоит в том, чтобы провести некоторую систематизацию основных направлений исследований и привести примеры для каждого из них.

Системы конфиденциального машинного обучения для градиентных методов обучения. Как отмечалось выше, один из самых популярных способов реализации безопасных многосторонних вычислений – GC-схемы [11]. На их основе предложен целый ряд протоколов и схем конфиденциального машинного обучения.

Первый пример – система SecureML [18]. Она реализует процедуры конфиденциального машинного обучения методами линейной регрессии, логистической регрессии и обучения искусственных нейронных сетей. Система использует в качестве примитивов схемы разделения секрета, GC-схемы и протоколы «забывающей передачи» (oblivious transfer). Архитектура системы предполагает наличие большого количества клиентов, которые исполняют роли поставщиков данных и получателей результата, и двух серверов. Серверы считаются недоверенными, т.е. каждый из них по отдельности может быть скомпрометирован условно пассивным противником, но при этом они не взаимодействуют между собой. Система SecureML обеспечивает вычислительную стойкость к атакам такого противника, но при этом требует от участников достаточно высоких вычислительных и коммуникационных затрат. В системе реализованы протоколы безопасных двусторонних вычислений над действительными числами с фиксированной точностью. В этой системе была многократно улучшена по сравнению с ранее известными протоколами производительность конфиденциального машинного обучения для линейной

регрессии, впервые были предложены протоколы для конфиденциального обучения нейронных сетей.

Впоследствии на основе системы SecureML была разработана система ABY³ [19]. В ней модель конфиденциального обучения методами линейной регрессии, логистической регрессии и обучения искусственных нейронных сетей была расширена для трехсерверной архитектуры. Математическое обеспечение для этой системы было полностью обновлено по сравнению с SecureML, разработана схема для переключения безопасным трехсторонним вычислением булевых, арифметических и GC-схем. Протоколы системы ABY³ обеспечивают теоретико-информационную стойкость к воздействию условно пассивного противника. Обе системы реализованы на языке C++.

Наконец, в [20] предложена система CodedPrivateML, в которой возможности распределения вычислений обобщены для произвольного количества n серверов, но только для линейной и логистической регрессий. Протоколы также обеспечивают теоретико-информационную стойкость к воздействию условно пассивного противника.

Другой пример – система конфиденциального машинного обучения SecureNN, предложенная в [21]. Система поддерживает модель вычислений с тремя серверами, каждый из которых может быть скомпрометирован условно пассивным противником, и обеспечивает теоретико-информационную стойкость. Особенностью системы является то, что она поддерживает только конфиденциальное обучение искусственных нейронных сетей, но в ней предусмотрена возможность использования практически всех популярных нелинейных функций активации нейронов (ReLU, Maxpool и др.) и популярных архитектур нейронных сетей: многослойных полносвязных и сверточных сетей.

Еще один пример – система конфиденциального машинного обучения Trident, предложенная в [22]. Она обеспечивает конфиденциальность обучения для четырех методов: линейной регрессии, логистической регрессии, полносвязных и сверточных нейронных сетей. Архитектура системы – клиент-серверная, число параллельно работающих серверов – до 4-х, один из которых может быть скомпрометирован активным противником. По утверждению авторов, система является практичной как для локальных, так и для глобальных сетей и превосходит по производительности ранее известные системы, включая ABY³. Математической основой системы являются GC-схемы в совокупности со схемами разделения секрета.

В [23] описана стойкая к активному нарушителю система конфиденциального машинного обучения FLASH, которая функционирует в конфигурации с четырьмя серверами. Она поддерживает следующие методы машинного обучения: линейную регрессию, логистическую регрессию, многослойные нейронные сети, бинаризованные нейронные сети (сети, представленные в виде схемы из логических элементов). Алгоритмы системы основаны на протоколах безопасных многосторонних вычислений.

Системы конфиденциального машинного обучения для решающих деревьев и метода случайного леса. В [24] предложены протоколы конфиденциальных вычислений по решающим деревьям и случайному лесу, обеспечивающие стойкость к условно пассивному и активному нарушителям. При этом в качестве основных криптографических инструментов выступают схема шифрования, гомоморфная по операции сложения (и скалярного умножения), а также протокол «забывающей» передачи. Особо следует подчеркнуть, что в этой работе не предложено алгоритмов конфиденциального построения решающих деревьев – только вычислений по уже построенному ранее дереву.

И только в [25] предложен протокол конфиденциального обучения решающих деревьев в совокупности с протоколом конфиденциальной классификации по построенному таким образом дереву. Протоколы стойки к активному нарушителю.

Модель системы подразумевает наличие клиента, который является источником данных и получателем результатов, и одного недоверенного сервера, который и выполняет процедуры построения решающего дерева и вычислений по нему. В качестве основного криптографического инструмента используется полностью гомоморфная схема шифрования. Модель включает в себя многочисленные оптимизации с целью снижения высоких требований к вычислительным ресурсам, накладываемых гомоморфным шифрованием.

Системы конфиденциального машинного обучения для кластеризации по методу k средних. Исследование проблемы конфиденциального машинного обучения без учителя берет свое начало с [26], где предложено несколько вариантов протокола конфиденциальной кластеризации по методу k средних с использованием полностью гомоморфного шифрования. Конфигурация протокола подразумевает наличие клиента, который является источником данных и получателем результата, а также недоверенного сервера, выполняющего вычисления. Показано, что вычислительную сложность протокола можно уменьшить, если допустить снижение точности кластеризации.

В [27] предложена иная модель конфиденциальной кластеризации: несколько участников протокола с примерно равными вычислительными мощностями, каждый из которых является источником некоторой части обучающей выборки и получателем результата, совместно выполняют протокол кластеризации без привлечения третьей стороны. Основой конструкции протокола служат GC-схемы с применением схем разделения секрета и протокола «забывающей» передачи. Обеспечивается стойкость к условно пассивному противнику.

В табл. 1 приведены в сравнении основные характеристики рассмотренных схем.

Таблица 1. Сравнительная характеристика систем конфиденциального машинного обучения

Наименование системы	Поддерживаемые методы машинного обучения							Тип нарушителя		Используемые криптографические примитивы			
	ЛнР	ЛгР	ИНС	СНС	БНС	РД	K-means	Пасс.	Акт.	GC	SS	PHE	FHE
SecureML [18]	+	+	+					+		2PC			
ABY ³ [19]	+	+	+					+		3PC	+		
CodedPrivateML [20]	+	+						+		MPC	+		
SecureNN [21]			+	+				+		3PC	+		
Trident [22]	+	+	+	+				+	+	4PC	+		
FLASH [23]	+	+	+		+			+	+	4PC	+		
[24]						+		+	+			+	
[25]						+		+	+			+	
[26]							+	+					+
[27]							+	+		MPC	+		

В табл. 1 введены следующие обозначения: ЛнР – линейная регрессия, ЛгР – логистическая регрессия, ИНС – искусственные нейронные сети, СНС – сверточные нейронные сети, БНС – бинаризованные нейронные сети, РД – решающие деревья, K-means – метод k средних, «Пасс.» – условно пассивный нарушитель, «Акт.» – активный нарушитель, GC – garbled circuits (GC-схемы), SS – secret sharing (схемы разделения секрета), PHE – partly homomorphic encryption (частично гомоморфное шифрование), FHE – fully homomorphic encryption (полностью гомоморфное шифрование).

Заключение

Анализ существующих реализаций систем конфиденциального машинного обучения показывает, что на текущий момент алгоритмические решения предложены далеко не для всех методов машинного обучения. Основное внимание исследователи уделяют градиентным методам классификации и восстановления регрессии и, в первую очередь, искусственным нейронным сетям. Имеется достаточно большой набор решений для линейной, логистической, гребневой регрессии. На второе место по вниманию исследователей, пожалуй, можно поставить логические методы классификации, а именно, решающие деревья. Не обойдены вниманием также и методы обучения без учителя – здесь лидирует метод k средних. В то же время, насколько может судить автор, в литературе отсутствуют примеры систем конфиденциального машинного обучения для логических (байесовских) методов классификации, еще не затрагивалась задача ранжирования. Отсутствуют решения для композиций алгоритмов (бэггинг, бустинг), за исключением метода случайного леса. В связи с этим важнейшим направлением исследований в области конфиденциального машинного обучения на ближайшие годы остается расширение номенклатуры алгоритмических решений, чтобы охватить как можно более широкий класс методов.

Другая очень существенная проблема – это повышение производительности систем конфиденциального машинного обучения. Оба основных подхода к конструированию протоколов: на основе безопасных многосторонних вычислений и на основе гомоморфного шифрования – страдают от высоких требований к ресурсам компьютерной системы. В первом случае критическим фактором выступают высокие требования к пропускной способности каналов связи при умеренных требованиях к вычислительным ресурсам, во втором случае – наоборот. Ведущим направлением исследований в этой сфере будет оставаться поиск разумного компромисса между этими требованиями.

Создание систем конфиденциального машинного обучения, пригодных к широкому внедрению и способных обеспечивать безопасность информации при использовании широкого спектра методов машинного обучения, способно принести большую практическую пользу. Помимо очевидного эффекта: удовлетворения нормативных требований, повышения доверия пользователей к операторам персональных данных и т.п., оно позволит создавать персонализированные информационные сервисы. Лица, которым в нынешних условиях персональные (медицинские, финансовые и т.д.) данные пользователей предоставляется для исполнения их запросов, будут получать лишь результаты её обработки в точности в том объеме, который необходим для выполнения ими своих функций, без доступа к персональным данным кого бы то ни было, кроме их владельца. Со временем использование такой схемы обработки данных может быть закреплено законодательно. Появление персонализированных информационных сервисов способно дать толчок реализации новых бизнес-процессов и развитию новых сфер деловой деятельности.

СПИСОК ЛИТЕРАТУРЫ:

1. Воронцов, К.В. Математические методы обучения по прецедентам (теория обучения машин) / К. Воронцов // Сайт «Машинное обучение», курс лекций. 2011. – 141 с. URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения: 26.12.2019).
2. Zhang, A. Dive into deep learning. Release 0.7.1 / A. Zhang, Z.C. Lipton, M. Li et al. // 2019. – 900 p. URL: <https://d2l.ai/d2l-en.pdf> (дата обращения: 25.12.2019).
3. Liu, Y. Federated learning and transfer learning for privacy, security and confidentiality / Y. Liu, Q. Yang, T. Chen et al. // AAAI 2019 tutorial. – 101 p. URL: <https://img.fedai.org.cn/fedweb/1552916850679.pdf> (дата обращения: 26.12.2019)

4. Raykova M. Advanced cryptography on the way to practice / M. Raykova // Real World Cryptography 2019, Switzerland, Zurich, 2019. URL: <https://rwc.iacr.org/2019/slides/RWC-Raykova.pdf> (дата обращения: 29.09.2019)
5. IEEE P3652.1 Federated Machine Learning Working Group. Draft standard. 2019. URL: <https://sagroups.ieee.org/3652-1/documents-library/> (дата обращения: 26.12.2019).
6. Yang, Q. Federated machine learning: concept and applications / Q. Yang, Y. Liu, T. Chen et al. // arxiv.org site. 2019. – 19 p. URL: <https://arxiv.org/pdf/1902.04885.pdf> (дата обращения: 26.12.2019).
7. Liu, Y. Secure federated transfer learning / Y. Liu, T. Chen, Q. Yang // arxiv.org site. 2018. – 9 p. URL: <https://arxiv.org/pdf/1812.03337.pdf> (дата обращения: 26.12.2019).
8. Запечников, С.В. Криптографическая защита процессов обработки информации в недоверенной среде: достижения, проблемы, перспективы / С. Запечников // Вестник современных цифровых технологий. 2019. № 1. С. 6–18.
9. Lisin, N. Methods and Approaches for Privacy-Preserving Machine Learning / N. Lisin, S. Zapechnikov // Advanced Technologies in Robotics and Intelligent Systems. Proceedings of ITR 2019. Springer Verlag, 2020. URL: <https://www.springer.com/gp/book/9783030334901> (дата обращения: 26.12.2019).
10. Запечников, С.В. Криптографические протоколы и их применение в финансовой и коммерческой деятельности: Учебное пособие для вузов / С. Запечников. М.: «Горячая линия – Телеком», 2007. – 320 с.
11. Yao, C. How to generate and exchange secrets / C. Yao // Proc. of 27th Annual Symposium on Foundations of Computer Science, 1986. P. 162–167.
12. Goldreich O. How to play any mental game / O. Goldreich, S. Micali, A. Wigderson // Proceeding STOC '87 Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing. P. 218–229.
13. Damgard, I. Secure distributed systems / I. Damgard, J. B. Nielsen, C. Orlandi, Electronic book, 2018. – 332 p. URL: https://blackboard.au.dk/webapps/blackboard/content/listContent.jsp?course_id=_116846_1&content_id=_1801036_1&mode=reset (дата обращения: 04.10.2019).
14. Ben-Or, M. Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation (Extended Abstract) / M. Ben-Or, S. Goldwasser, A. Wigderson // Proc. of 20th Annual ACM Symposium on Theory of Computing. ACM Press. 1988. P. 1–10.
15. Mohassel, P. Fast and Secure Three-party Computation: The Garbled Circuit Approach / P. Mohassel, M. Rosulek, and Y. Zhang // Proc. of ACM CCS 15: 22nd Conference on Computer and Communications Security. Ed. by I. Ray, N. Li, and C. Kruegel. ACM Press. 2015. P. 591–602.
16. Rabin, M.O. How to exchange secrets with oblivious transfer / M. Rabin. Technical Report TR-81, Aiken Computation Lab, Harvard University, 1981. URL: <https://eprint.iacr.org/2005/187.pdf> (дата обращения: 26.12.2019).
17. Armknecht, F. A guide to fully homomorphic encryption / F. Armknecht, C. Boyd, C. Carr et al. 2017. – 35 p. URL: <https://pdfs.semanticscholar.org/e6bc/da98a3385f1a287175a152fe6aa636a97c17.pdf> (дата обращения: 04.10.2019).
18. Mohassel, P. SecureML: A system for scalable privacy-preserving machine learning / P. Mohassel, Y. Zhang // Cryptology ePrint Archive. 2017. – 38 p. URL: <https://eprint.iacr.org/2017/396> (дата обращения: 26.12.2019).
19. Mohassel, P. ABY³: A mixed protocol framework for machine learning / P. Mohassel, P. Rindal // Cryptology ePrint Archive. 2018. – 40 p. URL: <https://eprint.iacr.org/2018/403> (дата обращения: 26.12.2019).
20. So, J. CodedPrivateML: A fast and privacy-preserving framework for distributed machine learning / J. So, B. Guler, A.S. Avestimehr et al. // Cryptology ePrint Archive. 2019. – 14 p. URL: <https://eprint.iacr.org/2019/140> (дата обращения: 26.12.2019).
21. Wagh, S. SecureNN: Efficient and private neural network training / S. Wagh, D. Gupta, N. Chandran // Cryptology ePrint Archive. 2018. – 24 p. URL: <https://eprint.iacr.org/2018/442> (дата обращения: 26.12.2019).
22. Rachuri, R. Trident: Efficient 4PC framework for privacy preserving machine learning / R. Rachuri, A. Suresh // Cryptology ePrint Archive. 2019. – 26 p. URL: <https://eprint.iacr.org/2019/1315> (дата обращения: 26.12.2019).
23. Byali, M. FLASH: Fast and robust framework for privacy-preserving machine learning / M. Byali, H. Chaudhari, A. Patra et al. // Cryptology ePrint Archive. 2019. – 29 p. URL: <https://eprint.iacr.org/2019/1365> (дата обращения: 26.12.2019).
24. Wu, D. J. Privately evaluating decision trees and random forests / D. Wu, T. Feng, M. Naehrig et al. // Cryptology ePrint Archive. 2015. – 35 p. URL: <https://eprint.iacr.org/2015/386> (дата обращения: 26.12.2019).

25. Akavia, A. Privacy-preserving decision tree training and prediction against malicious server / A. Akavia, M. Leibovich, Y. Resheff et al. // Cryptology ePrint Archive. 2019. – 34 p. URL: <https://eprint.iacr.org/2019/1282> (дата обращения: 26.12.2019).
26. Jaschke, A. Unsupervised machine learning on encrypted data / A. Jaschke, F. Armknecht // Cryptology ePrint Archive. 2018. – 30 p. URL: <https://eprint.iacr.org/2018/411> (дата обращения: 26.12.2019).
27. Mohassel, P. Practical privacy-preserving k-means clustering / P. Mohassel, M. Rosulek, N. Trieu // Cryptology ePrint Archive. 2019. – 30 p. URL: <https://eprint.iacr.org/2019/1158> (дата обращения: 26.12.2019).

REFERENCES:

- [1] Vorontsov, K.V. Matematicheskie metody obucheniya po pretsedentam (teoriya obucheniya machin) K. Vorontsov (in Russian). Site “Machine learning”. 2011. – 141 p. URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (accessed: 26.12.2019).
- [2] Zhang, A. Dive into deep learning. Release 0.7.1 A. Zhang, Z.C. Lipton, M. Li et al. 2019. – 900 p. URL: <https://d2l.ai/d2l-en.pdf> (accessed: 25.12.2019).
- [3] Liu, Y. Federated learning and transfer learning for privacy, security and confidentiality Y. Liu, Q. Yang, T. Chen et al. AAAI 2019 tutorial. 101 p. URL: <https://img.fedai.org.cn/fedweb/1552916850679.pdf> (accessed: 26.12.2019).
- [4] Raykova M. Advanced cryptography on the way to practice M. Raykova. Real World Cryptography 2019, Switzerland, Zurich, 2019. URL: <https://rwc.iacr.org/2019/slides/RWC-Raykova.pdf> (accessed: 29.09.2019).
- [5] IEEE P3652.1 Federated Machine Learning Working Group. Draft standard. 2019. URL: <https://sagroups.ieee.org/3652-1/documents-library/> (accessed: 26.12.2019).
- [6] Yang, Q. Federated machine learning: concept and applications Q. Yang, Y. Liu, T. Chen et al. arxiv.org site. 2019. – 19 p. URL: <https://arxiv.org/pdf/1902.04885.pdf> (accessed: 26.12.2019).
- [7] Liu, Y. Secure federated transfer learning Y. Liu, T. Chen, Q. Yang. arxiv.org site. 2018. – 9 p. URL: <https://arxiv.org/pdf/1812.03337.pdf> (accessed: 26.12.2019).
- [8] Zapechnikov, S.V. Kriptograficheskaya zashchita protsessov obrabotki informatsii v nedoverennoj srede: dostizheniya, problem, perspektivy. S. Zapechnikov Vestnik sovremennykh tsifrovyykh tekhnologiy. 2019. No.1. P. 6–18 (in Russian).
- [9] Lisin, N. Methods and Approaches for Privacy-Preserving Machine Learning N. Lisin, S. Zapechnikov Advanced Technologies in Robotics and Intelligent Systems. Proceedings of ITR 2019. Springer Verlag, 2020. URL: <https://www.springer.com/gp/book/9783030334901> (accessed: 26.12.2019).
- [10] Zapechnikov, S.V. Kriptograficheskie protokoly i ih primeneniye v finansovoy i kommercheskoj deyatel'nosti: Uchebnoye posobie dlya vuzov S. Zapechnikov. Moscow, “Goryachaya liniya – Telekom”, 2007. – 320 p. (in Russian).
- [11] Yao, C. How to generate and exchange secrets C. Yao. Proc. of 27th Annual Symposium on Foundations of Computer Science, 1986. P. 162–167.
- [12] Goldreich O. How to play any mental game O. Goldreich, S. Micali, A. Wigderson. Proceeding STOC '87 Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing. P. 218–229.
- [13] Damgard, I. Secure distributed systems I. Damgard, J. B. Nielsen, C. Orlandi. Electronic book, 2018. – 332 p. URL: https://blackboard.au.dk/webapps/blackboard/content/listContent.jsp?course_id=_116846_1&content_id=_1801036_1&mode=reset (accessed: 04.10.2019).
- [14] Ben-Or, M. Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation (Extended Abstract) M. Ben-Or, S. Goldwasser, A. Wigderson. Proc. of 20th Annual ACM Symposium on Theory of Computing. ACM Press. 1988. P. 1–10.
- [15] Mohassel, P. Fast and Secure Three-party Computation: The Garbled Circuit Approach P. Mohassel, M. Rosulek, and Y. Zhang. Proc. of ACM CCS 15: 22nd Conference on Computer and Communications Security. Ed. by I. Ray, N. Li, and C. Kruegel. ACM Press. 2015. P. 591–602.
- [16] Rabin, M.O. How to exchange secrets with oblivious transfer M. Rabin. Technical Report TR-81, Aiken Computation Lab, Harvard University, 1981. URL: <https://eprint.iacr.org/2005/187.pdf> (accessed: 26.12.2019).
- [17] Armknecht, F. A guide to fully homomorphic encryption F. Armknecht, C. Boyd, C. Carr et al. 2017. – 35 p. URL: <https://pdfs.semanticscholar.org/e6bc/da98a3385f1a287175a152fe6aa636a97c17.pdf> (accessed: 04.10.2019).
- [18] Mohassel, P. SecureML: A system for scalable privacy-preserving machine learning P. Mohassel, Y. Zhang. Cryptology ePrint Archive. 2017. – 38 p. URL: <https://eprint.iacr.org/2017/396> (accessed: 26.12.2019).
- [19] Mohassel, P. ABY³: A mixed protocol framework for machine learning P. Mohassel, P. Rindal. Cryptology ePrint Archive. 2018. – 40 p. URL: <https://eprint.iacr.org/2018/403> (accessed: 26.12.2019).

- [20] So, J. CodedPrivateML: A fast and privacy-preserving framework for distributed machine learning J. So, B. Guler, A.S. Avestimehr et al. Cryptology ePrint Archive. 2019. – 14 p. URL: <https://eprint.iacr.org/2019/140> (accessed: 26.12.2019).
- [21] Wagh, S. SecureNN: Efficient and private neural network training S. Wagh, D. Gupta, N. Chandran. Cryptology ePrint Archive. 2018. – 24 p. URL: <https://eprint.iacr.org/2018/442> (accessed: 26.12.2019).
- [22] Rachuri, R. Trident: Efficient 4PC framework for privacy preserving machine learning R. Rachuri, A. Suresh. Cryptology ePrint Archive. 2019. – 26 p. URL: <https://eprint.iacr.org/2019/1315> (accessed: 26.12.2019).
- [23] Byali, M. FLASH: Fast and robust framework for privacy-preserving machine learning M. Byali, H. Chaudhari, A. Patra et al. Cryptology ePrint Archive. 2019. – 29 p. URL: <https://eprint.iacr.org/2019/1365> (accessed: 26.12.2019).
- [24] Wu, D. J. Privately evaluating decision trees and random forests D. Wu, T. Feng, M. Naehrig et al. Cryptology ePrint Archive. 2015. – 35 p. URL: <https://eprint.iacr.org/2015/386> (accessed: 26.12.2019).
- [25] Akavia, A. Privacy-preserving decision tree training and prediction against malicious server A. Akavia, M. Leibovich, Y. Resheff et al. Cryptology ePrint Archive. 2019. – 34 p. URL: <https://eprint.iacr.org/2019/1282> (accessed: 26.12.2019).
- [26] Jaschke, A. Unsupervised machine learning on encrypted data A. Jaschke, F. Armknecht. Cryptology ePrint Archive. 2018. – 30 p. URL: <https://eprint.iacr.org/2018/411> (accessed: 26.12.2019).
- [27] Mohassel, P. Practical privacy-preserving k-means clustering P. Mohassel, M. Rosulek, N. Trieu. Cryptology ePrint Archive. 2019. – 30 p. URL: <https://eprint.iacr.org/2019/1158> (accessed: 26.12.2019).

*Поступила в редакцию – 26 декабря 2019 г. Окончательный вариант – 04 февраля 2020 г.
Received – December 26, 2019. The final version – February 04, 2020.*