ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ В ЗАДАЧАХ КИБЕРБЕЗОПАСНОСТИ. ЧАСТЬ 1

Гайфулина Д.А.[1], Котенко И.В.[2]

Цель статьи: *аналитический обзор в области глубокого обучения для задач кибербезопасности.*

Метод исследования: *системный анализ современных методов глубокого обучения, применяемых в задачах кибербезопасности, разработка их классификации по используемой архитектуре. Анализ релевантных обзоров в области глубокого обучения.*

Полученный результат: *приводится описание методов глубокого обучения, предложена их классификация. Определены области применения описанных методов в различных приложениях кибербезопасности, включая обнаружение вторжений и вредоносного программного обеспечения, анализ сетевого трафика и некоторые другие задачи. Представлена сравнительная характеристика релевантных обзоров в области глубокого обучения, позволяющая определить дальнейшие направления исследований.*

Основной вклад авторов в область исследования методов глубокого обучения для задач кибербезопаности заключается в классификации предметной области, проведении общего и сравнительного анализа существующих подходов, отражающих современное состояние научных исследований.

Ключевые слова: *наука о данных, машинное обучение, глубокое обучение, глубокие нейронные сети, кибербезопасность, обнаружение вторжений, обнаружение вредоносного программного обеспечения.*

DOI: 10.21681/2311-3456-2020-03-76-86

Введение

Глобальная информатизация различных сфер жизнедеятельности человека, таких как наука, производство или экономика, способствует развитию информационных технологий для хранения, обработки и передачи большого объема данных. Вместе с тем, растет количество угроз информационной безопасности, и появляются новые виды кибератак. В настоящее время для обеспечения безопасности компьютеризированных инфраструктур используются различные методы проектирования средств защиты информации, оценки защищенности, выявления аномальной и подозрительной активности.

Системы киберзащиты, использующие данные методы, генерируют объемные данные для аналитики безопасности, оповещении об инцидентах и генерации контрмер. Использование науки о данных в области кибербезопасности может помочь в выявлении закономерностей и обнаружении аномального поведения для повышения эффективности работы систем киберзащиты. В настоящее время технологии машинного обучения повсеместно применяются для решения множества задач классификации, прогнозирования и принятия решений. Все чаще в основе этих направлений лежит класс методов, называемых глубоким обучением. Алгоритмы глубокого обучения делают большие успехи в решении задач искусственного интеллекта, которые на протяжении многих лет ставились перед научным сообществом благодаря возможности анализировать сложные структуры в многомерных данных. Алгоритмы глубокого обучения имеют большой потенциал, так как требуют намного меньше ручной работы по сравнению с традиционными методами машинного обучения и их эффективность увеличивается с развитием современных вычислительных систем.

Данная статья представляет собой аналитический обзор в области глубокого обучения для задач кибербезопасности. В связи с этим, описываются методы глубокого обучения, их архитектуры и цели применения для обеспечения безопасности информационно-коммуникационных систем.

Статья имеет следующую структуру. В *разделе 1* представлены основные понятия глубокого обучения, приводится классификация и описание наиболее распространенных методов и моделей глубокого обучения. В *разделе 2* анализируются существующие исследовательские работы в области методов глубокого обучения для задач кибербезопасности.

1. Основные понятия глубокого обучения

Концепция глубокого обучения возникла из исследований искусственных нейронных сетей. Нейронные

1 Гайфулина Диана Альбертовна, аспирант, младший научный сотрудник лаборатории проблем компьютерной безопасности, ФГБУН Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Россия. E-mail: gaifulina@comsec.spb.ru

2 Котенко Игорь Витальевич, доктор технических наук, профессор, главный научный сотрудник и заведующий лабораторией проблем компьютерной безопасности, ФГБУН Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Россия. E-mail: ivkote@comsec.spb.ru

сети состоят из отдельных блоков (нейронов), формирующих отдельные слои, первый из которых является входным, а последний – выходным слоем [1]. Слои, расположенные между ними, называются скрытыми слоями. Каждый слой, кроме первого, принимает на вход результаты работы предыдущего слоя, и общее количество слоев, таким образом, формирует глубину сети. Выходной слой представляет результат работы нейросети, его предсказания основаны на данных, поступающих на вход нейросети.

1.1. Базовые понятия и классификация методов глубокого обучения

Количество нейронов на каждом слое может быть различным, но при этом переходы из одного измерения в другое должны быть согласованы. Каждый нейрон по своей сути является функцией, выполняющей нелинейное преобразование входного сигнала, и обладает двумя параметрами – вектором весов и коэффициентом смещения [2]. На выход он подает результат функции активации, аргументом которой является скалярное произведение входного и вектора весов, смещенное на заданное расстояние. Функция активации определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения. Существует ряд различных нелинейных функций, которые можно использовать; однако наиболее распространенными являются сигмоида, гиперболический тангенс, функция softmax и линейный выпрямитель (Rectified linear unit, ReLU) [3]. Также применяются несколько методов контроля емкости нейронной сети, позволяющих предотвратить переобучение, такие как L1 и L2 регуляризация, метод ограничения нормы вектора весов и дропаут (dropout) [4]. Пример общей архитектуры глубокой нейронной сети представлен на рис. 1. Здесь и далее на изображениях архитектуры нейронных сетей используются следующие обозначения:

- $x_1, ..., x_n$ – нейроны входного слоя сети;
- $y_1, ..., y_m$ – нейроны выходного слоя сети;
- ω_{ij} – нейроны скрытых слоев сети, где i – порядок слоя, j – порядок нейрона;
- φ – функция активации.

Процесс обучения нейронной сети состоит в подборе оптимальных параметров для нейронов. Данный процесс может происходить как с наличием «учителя», и такой способ обучения подразумевает предоставление сети обучающего набора данных с правильными откликами (классами, показателями), так и без «учителя», при котором используется неразмеченный обучающий набор данных. Главной целью обучения нейронной сети является достижение баланса между способностью выдавать правильные результаты классификации, как для обучающей выборки данных, так и для схожих данных, неидентичных обучающей выборке [1]. При инициализации нейронной сети случайными весами получаются некоторые выходные значения.

Обучение с учителем предполагает, что веса сети изменяются до тех пор, пока для каждого входного вектора значений не будет получен приемлемый уровень отклонения от целевого выходного вектора. Функция потерь возвращает некоторую метрику их несоответствия, например перекрестную энтропию или среднюю квадратичную ошибку. Для минимизации потерь необходимо найти минимум функции потерь с помощью вычисления ее градиента, например с помощью методов пошагового или стохастичного градиентного спуска [5]. Обучение без учителя заключается в настройке весов нейронной сети таким образом, чтобы предъявление достаточно близких входных векторов давало одинако-

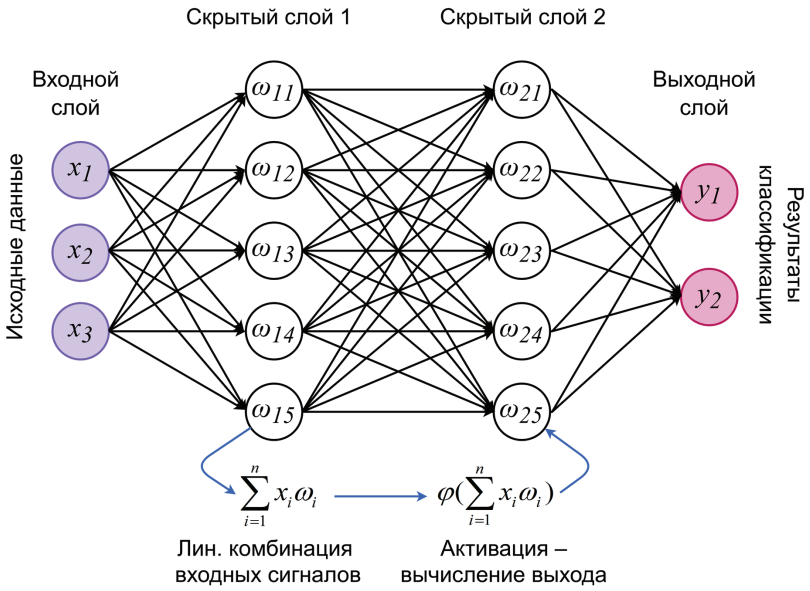

Рис. 1. Архитектура глубокой нейронной сети

вые выходные значения. Также возможна организация нейронной сети с использованием гибридизации способов обучения с учителем и без учителя.

Классификация наиболее распространенных методов глубокого обучения представлена на рис. 2.

Далее рассмотрим представленные методы глубокого обучения и их особенности.

1.2. Методы обучения без учителя

Глубокая сеть доверия (Deep Belief Network, DBN) является нейронной сетью, состоящей из нескольких скрытых слоев, в которых нейроны внутри одного слоя не связаны друг с другом, но связаны с нейронами соседнего слоя. Глубокую сеть доверия можно рассматривать как композицию таких сетей, как автокодировщики или ограниченные машины Больцмана [6], в которых скрытый слой каждой подсети служит видимым слоем для следующей.

Автокодировщик (Autoencoder) является специализированной искусственной нейронной сетью, применяющей обучение без учителя с использованием метода обратного распространения ошибки [1]. Автокодировщик содержит входной, скрытый и выходной слой, при этом количество нейронов во входном и выходном слое должно совпадать. Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к ее входам и получении на выходе наиболее близкого отклика к входному [7]. При этом чтобы результат не был тривиальным, на скрытый слой накладывается ограничение по количеству нейронов: его размерность должна быть меньше размерности входного и выходного слоев. Автокодировщик состоит из двух частей, осуществляющих кодирование данных (энкодер) и последующее декодирование (декодер) (рис. 3). Кодирование переводит входной сигнал $(x_1,...,x_n)$ в его представление на скрытом слое ω, а декодирование восстанавливает сигнал $(x'_1,...,x'_n)$.

При использовании *разряженного автокодировщика* (Sparse Autoencoder) ограничивается число одновременно активных нейронов скрытого слоя [8]. Такие

Рис. 2. Классификация основных методов глубокого обучения

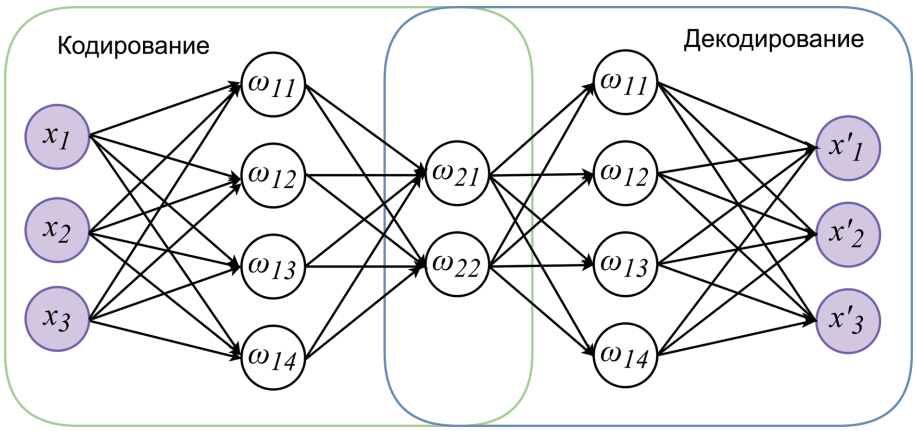

Рис. 3. Автокодировщик

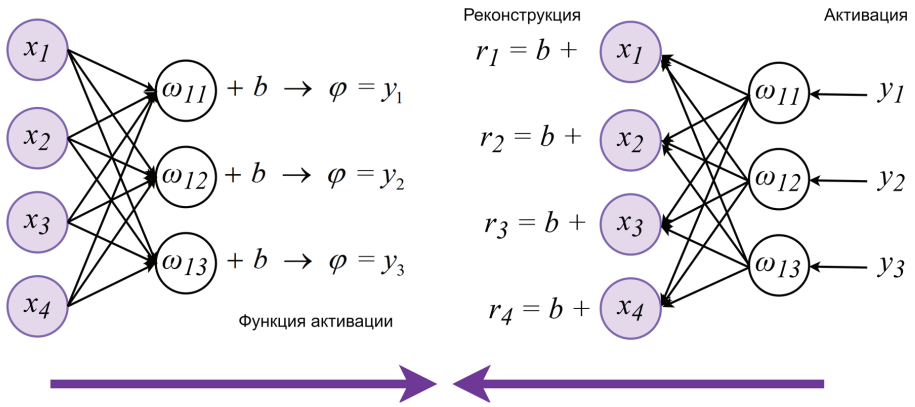

Рис. 4. Ограниченная машина Больцмана

ограничения заставляют сеть искать закономерности в исходных данных, выполняя их сжатие, то есть выделять общие признаки, кодируемые весовыми векторами сети. Автокодировщик может быть спроектирован так, чтобы устранять шум и быть более устойчивым, позволяя реконструировать входные данные из зашумленной версии ввода. Данная архитектура называется *шумоподавляющим автокодировщиком* (Denoising Autoencoder) [9]. *Многоуровневый автокодировщик* (Stacked Autoencoder) представляет собой нейронную сеть, состоящую из нескольких уровней, в котором выходное значение каждого слоя подключается к входам следующего уровня.

Ограниченная машина Больцмана (Restricted Boltzmann Machine, RBM) является стохастической нейронной сетью, которая определяет распределение вероятности на входных образцах данных. Ограниченная машина Больцмана базируется на бинарных элементах с распределением Бернулли, составляющих видимый и скрытый слои сети. Соответственно, нейроны подразделяются на видимые и скрытые, и связи допустимы только между нейронами разного типа [10]. Объединение нескольких каскадов ограниченных машин Больцмана формирует глубокую сеть доверия. Скрытый слой RBM содержит глубокие признаки данных, которые выявляются в процессе обучения. На рис. 4 показана двухслойная ограниченная машина Больцмана. Связи между слоями задаются с помощью матрицы весов, а также смещений для видимого слоя и для скрытого слоя. Результативное значение каждого узла сети получается умножением каждого входного параметра x на соответствующий ему вес ω и суммированием всех произведений. Данная сумма затем добавляется к смещению b, и, наконец, результат передается через функцию активации φ для получения выходных данных узла y. При реконструкции данные y умножаются на веса ω и суммирования произведений добавляются к некоторым новым смещениям b, которые производят окончательный результат r, являющийся реконструкцией исходных данных.

Восстановленные значения не совпадают с исходными, так как реконструкция делает предположение о распределении вероятностей входных данных. Глобальная цель RBM состоит в том, чтобы минимизировать ошибку между входными данными и реконструкцией, произведенной скрытыми элементами и их весами.

Рекуррентная нейронная сеть (Recurrent Neural Network, RNN) является классом нейронных сетей, в которой связи между нейронами образуют направленный цикл [1]. Это позволяет демонстрировать динамическое поведение сети. RNN может использовать свою внутреннюю память для обработки произвольных последовательностей входов x для получения результата y. Рекуррентная нейронная сеть создается путем применения того же набора весов ω рекурсивно над графоподобной структурой путем перемещения структуры в топологическом порядке (рис. 5).

Такие сети, как правило, обучаются обратным режимом автоматической дифференциации (Backpropagation Through Time, BPTT) [11]. Основное отличие рекуррентных сетей от традиционных заклю-

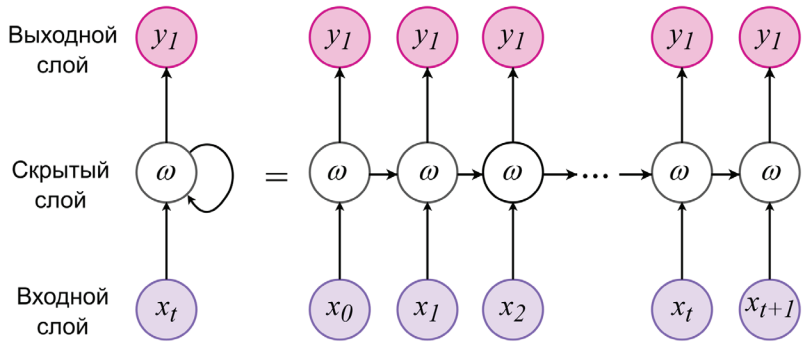

Рис. 5. Рекуррентная нейронная сеть

чается в логике работы сети, при которой каждый нейрон взаимодействует сам с собой. Каждый элемент исходной последовательности поочерёдно передаётся одним и тем же нейронам, возвращающим себе предсказание вместе со значением следующего нейрона, до тех пор, пока последовательность не закончится. Кроме входного сигнала нейрон использует также своё дополнительное скрытое состояние. Каждый нейрон отрабатывает несколько раз подряд, являясь своеобразной ячейкой памяти. Основное отличие разных типов рекуррентных нейронов друг от друга заключается в способе обработки ячейки памяти, расположенной внутри них.

Традиционный подход подразумевает сложение двух векторов (сигнала и памяти) с последующим вычислением активации от суммы. Но память, реализованная подобным образом, получается весьма короткой. Чтобы преодолеть этот недостаток, был придуман *блок долгой краткосрочной памяти* (Long Short-Term Memory, LSTM) [12], который имеет возможность удалять и добавлять информацию в состояние ячейки. Этот процесс регулируется специальными структурами, называемыми гейтами или «вентилями» (gate) (рис. 6). Гейт реализует функцию активации, например сигмоиды (σ), и операцию поточечного умножения $\times$, а его выходом является число в диапазоне [0,1]. Первый слой, гейт забывания (forget gate), вычисляет, как скоро на данном шаге ему нужно «забыть» предыдущую информацию, и определяет множители к компонентам вектора памяти в диапазоне от 0 до 1, где единица обозначает полное сохранение, а ноль – полное удаление. Второй слой, входной гейт (input gate), вычисляет, насколько интересна новая информация, пришедшая с сигналом, используя такой же множитель, но уже для наблюдения. При этом одна функция активации σ определяет, какие значения будут обновлены, а другая φ создает вектор новых значений. На третьем слое, выходном гейте (output gate), вычисляется линейная комбинация (+) памяти и наблюдения с только что вычисленными весами для каждой из компонент, образуя новое состояние памяти. Одна из функций активации σ решает, какая часть сигнала важна для дальнейших решений, затем другая φ проецирует вектор памяти на отрезок [-1;1], и в конце вычисляется скалярное произведение данных векторов.

Еще одной единицей RNN, которая была разработана для длинной памяти, является *управляемый рекуррентный блок* (Gated Recurrent Units, GRU) [13] (рис. 7).

В этом варианте гейт забывания и входной гейт объединены в один гейт обновления (update gate), а также вводятся гейт сброса состояния (reset gate) и контейнер памяти (memory container). Два гейта представляют собой функции от входного вектора и скрытого состояния на предыдущем шаге и используют свои веса. Эффективность GRU сопоставима с LSTM, но GRU обладает меньшим числом параметров (гейтов), что облегчает обучение нейронной сети.

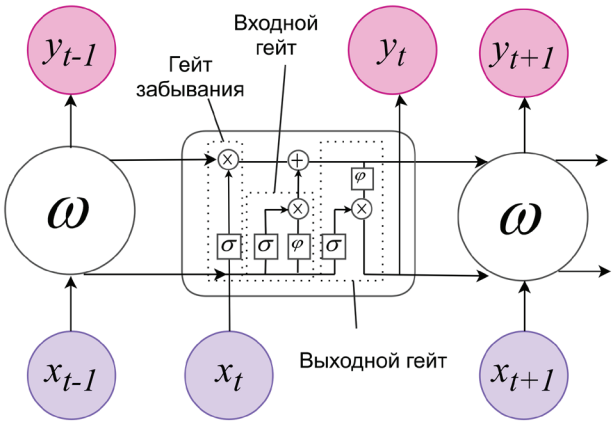

Рис. 6. Блок долгой краткосрочной памяти

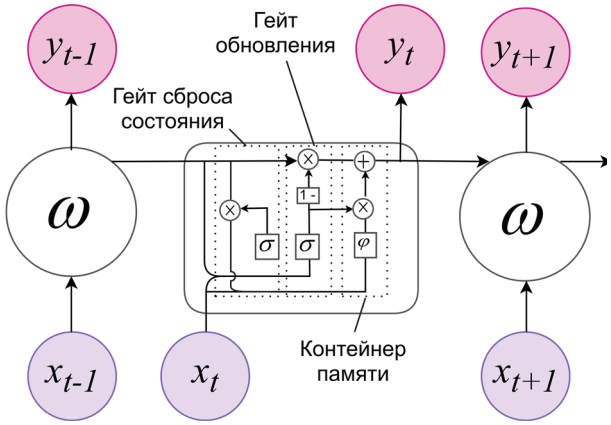

Рис. 7. Управляемый рекуррентный блок

1.3. Методы обучения с учителем

Нейронные сети прямого распространения (Feed Forward Neural Network, FFNN) передают информацию от входа к выходу прямолинейно [2]. Такие нейронные сети состоят из множества слоев, где каждый слой состоит из входных, скрытых или выходных нейронов. Нейроны одного слоя не связаны между собой, а соседние слои обычно полностью связаны. Нейронная сеть прямого распространения обычно обучается по методу обратного распространения ошибки. Если у сети есть достаточное количество скрытых нейронов, она теоретически способна смоделировать взаимодействие между входным и выходными данными.

Сверточная нейронная сеть (Convolutional Neural Network, CNN) в большинстве случаев применяется для обработки входных данных, хранящихся в массивах [14], например, изображений, спектрограмм звука и видео. Идея свёрточных нейронных сетей заключается в чередовании свёрточных слоев (convolution layers) и субдискретизирующих слоев (pooling layers) [10]. Сеть имеет однонаправленную многослойную структуру (рис. 8).

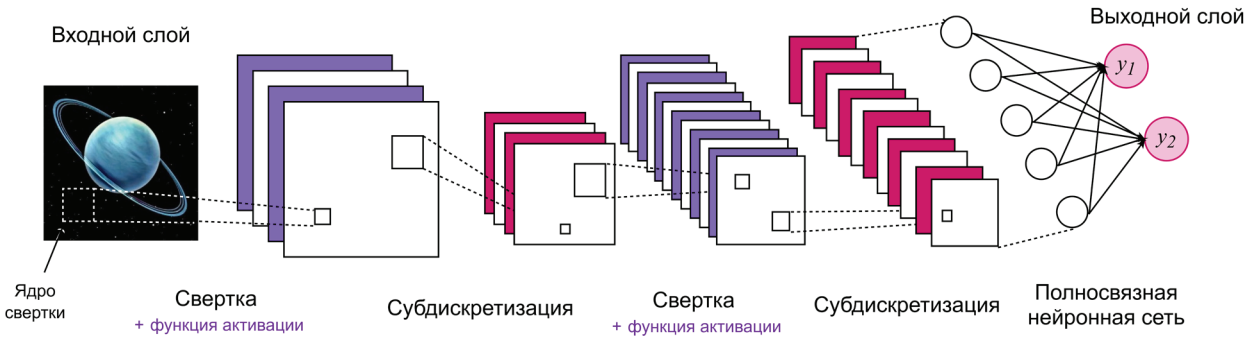

Рис. 8. Сверточная нейронная сеть

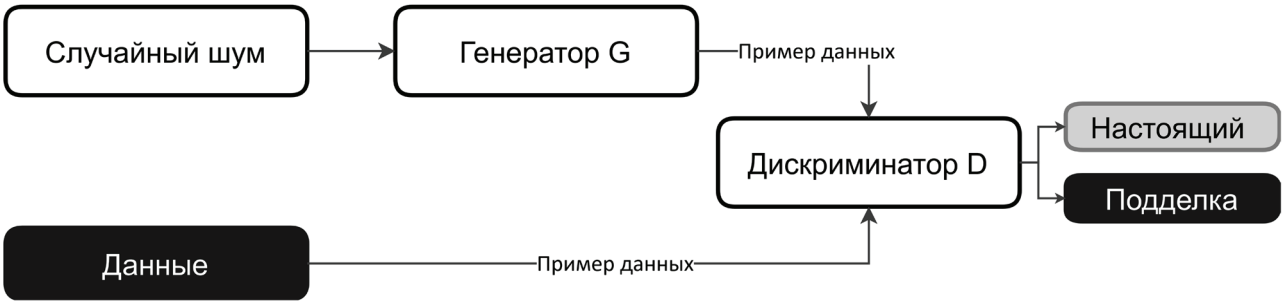

Рис. 9. Генеративная состязательная сеть

Сверточные слои являются ядром сверточной нейронной сети. Суть операции свертки заключается в том, что каждый фрагмент входного слоя поэлементно умножается на матрицу свертки, также называемую ядром свертки или рецептивным полем, а результат суммируется и записывается в аналогичную позицию выходного слоя. В качестве матрицы свертки используется ограниченная матрица весов небольшого размера, которая позволяет учитывать близкие физические или временные отношения в данных и помогает уменьшить объем памяти. Матрица свертки перемещается по входному слою и формирует после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. Результат затем передается через нелинейную функцию, обычно ReLU, и называется картой признаков. Операция субдискретизации (пулинг) выполняет уменьшение размерности сформированных карт признаков на соответствующих слоях. Информация о наличии определенного признака важнее его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотнённой карты признаков меньшей размерности. Это обеспечивает высокую степень инвариантности к исходным данным. После нескольких прохождений свертки изображения и уплотнения с помощью субдискретизации система перестраивается от конкретной сетки пикселей к более абстрактным картам признаков. Эти данные объединяются и передаются на обычную полносвязную нейронную сеть, которая тоже может состоять из нескольких слоёв. При этом полносвязные слои уже утрачивают пространственную структуру пикселей и обладают сравнительно небольшой размерностью по отношению к количеству пикселей исходного изображения.

1.4. Гибридные методы обучения

Генеративные состязательные сети (Generative Adversarial Networks, GAN) первоначально были разработаны Гудфеллоу [15] для оценки генеративных моделей с использованием состязательного процесса. Идея, лежащая в основе GAN, вытекает из теории игр, в которой две конкурирующие модели, а именно генератор и дискриминатор, противопоставлены друг другу. Генератор *G* изучает распределение реальных данных, а дискриминатор *D* оценивает вероятность того, что входные данные поступают из реальных данных или из генератора (рис. 9). Участники игры должны совершенствовать свои методы, пока распределение данных не станет неотличимым от модельного распределения, т. е. найти равновесие по Нэшу между двумя участниками.

Сеть дискриминаторов представляет собой стандартную сверточную сеть, которая может классифицировать изображения, подаваемые на нее с помощью биномиального классификатора, распознающего изображения как реальные или как поддельные.

Рекурсивные нейронные сети (Recursive neural network, ReNN) работают с данными переменной длины и используют иерархические структуры образцов при обучении [1]. В рекурсивных сетях нейроны с одинаковыми весами активируются рекурсивно в соответствии со структурой сети. В процессе работы формируется модель предсказания для структур как переменной размерности, так и скалярных структур. В самой простой архитектуре узлы сети сходятся к родителям через ма-

трицу весов скрытого слоя, используемую многократно через всю сеть, и нелинейную функцию активации типа гиперболического тангенса. ReNN успешно применяются при обучении последовательных структур и деревьев в задачах обработки естественного языка, при этом фразы и предложения моделируются через векторное представление слов.

2. Релевантные работы

Рассмотрев различные методы глубокого обучения, можно отметить, что их основными преимуществами, по сравнению с методами традиционного машинного обучения являются высокая производительность и масштабируемость для возрастающего количества данных. Предварительная подготовка глубокой нейронной сети позволяет использовать ее для решения различных задач в конкретной предметной области, что облегчает обучение всей модели. По сравнению с другими формами машинного обучения, глубокое обучение имеет намного меньше ограничений и требует меньше ручного программирования. При решении традиционных задач машинного обучения аналитик должен самостоятельно определять наиболее значимые признаки данных для обучения алгоритма. Алгоритмы на основе глубокого обучения позволяют осуществлять автоматический отбор информативных признаков и работать напрямую с необработанными данными, извлекая из них общие представления на разных уровнях детализации. Такая особенность делает их привлекательными для анализа больших данных, в том числе для решения различных задач кибербезопасности. Наиболее распространенные области применения глубокого обучения в кибербезопасности представлены на рис. 10.

В данном разделе анализируются аналитические обзорные статьи, посвященные методам глубокого обучения для решения задач кибербезопасности с точки зрения охвата области исследований и используемых методов.

Работа [16] рассматривает большое количество академических исследований в задаче обнаружения вторжений, основанных на машинном обучении и глубоком обучении. Приводится описание различных наборов данных, применяемых исследователями данной области, таких как DARPA Intrusion Detection Data Sets, KDD Cup 99, NSL-KDD и ADFA. Статья также описывает сходства и различия в машинном обучении и глубоком обучении. Авторы выявляют некоторые проблемы в этой области исследований, связанные с наборами эталонных данных и отсутствием единообразных метрик оценки.

В [17] представлен анализ различных алгоритмов машинного и глубокого обучения, направленных на решение задач обнаружения вторжений, анализа вредоносных программ и обнаружения спама. Авторы приходят к выводу, что глубокое обучение находится на ранней стадии развития в области кибербезопасности и нуждается в дальнейших исследованиях.

В [18] приводится краткий обзор методов регуляризации для алгоритмов глубокого обучения, используемых для защиты киберфизических систем. Регуляризация представляет собой любую модификацию, которая вносится в алгоритм, чтобы уменьшить ошибку классификации данных, не входящих в обучающую выборку сети. Также рассматриваются проблемы в этой области, обсуждаются будущие направления.

Работа [19] затрагивает методы машинного и глубокого обучения для защиты технологий Интернета вещей. Исследование охватывает широкий спектр типов кибератак и подходы к их обнаружению с использованием методов глубокого обучения, включая сверточные нейронные сети, рекуррентные нейронные сети и генеративно-состязательные сети. Для каждого метода описываются возможности, преимущества и недостатки.

Работа [20] посвящена структурированному и всестороннему обзору методов исследования в области обнаружения аномалий на основе глубокого обучения. Авторы сгруппировали современные методы исследования обнаружения аномалий по различным категориям, основываясь на принятом подходе. В рамках каждой категории описывается базовая техника обнаружения аномалий и ее возможные варианты. Для каждой категории представляются преимущества и недостатки, а также вычислительная сложность методов в реальных прикладных областях.

В [21] представлено краткое описание распространенных методов глубокого обучения, и приведен анализ их применения в области кибербезопасности. Авторы данного обзора рассматривают широкий спектр типов атак, включая вредоносные программы, спам, внутренние угрозы, вторжения в сеть, ложные инъекции данных и вредоносные доменные имена, используемые ботнетами.

Рис. 10. Области применения глубокого обучения в области кибербезопасности

Работа [22] также охватывает широкую область приложений глубокого обучения в безопасности, а именно обнаружение вторжений и вредоносных программ, выявление фишинга и спама, обнаружение фальсификаций веб-сайтов. Цитируемые статьи обобщаются и анализируются с точки зрения используемых методов, применимости модели и степени детализации.

В [23], в отличие от других обзоров, выделяется два важных аспекта: атаки на различные сети глубокого обучения и атаки на сеть с использованием самих моделей глубокого обучения.

В табл. 1 приводится сравнительная характеристика приведенных обзорных работ, в том числе сравнение проводится и с данной статьей.

Для обозначения области рассмотренных приложений кибербезопасности используются следующие сокращения: ОВ – обнаружение вторжений, ОВП – обнаружение вредоносного ПО, АТ – анализ сетевого трафика, РУ – раскрытие утечек данных, С – выявление спама, ВС – выявление вредоносных сайтов. Также указываются годы издания цитируемых источников.

В представленных работах приводится сравнение рассмотренных методов глубокого изучения по различным параметрам. Из них выделим следующие параметры: применение предлагаемого метода, число цитирований источника, архитектура сети, функция активации, используемых набор данных и оценка эффективности метода.

В табл. 2 отражены параметры, по которым проводят сравнительную характеристику авторы представленных обзорных работ, а также предлагаемые к анализу в данном исследовании.

Таблица 1.

Анализ релевантных работ

Обзорная работа	**Годы**	**Область исследования**						**Методы глубокого обучения**					
		ОВ	ОВП	АТ	РУ	С	ВС	DBN	RNN	CNN	FFNN	GAN	ReNN
Xin et al, 2018 [16]	2010-2017	✓	✓	✓	–	–	–	✓	✓	✓	–	–	–
Apruzzese et al, 2018 [17]	2007-2016	✓	✓	–	–	✓	–	✓	✓	✓	✓	–	–
Wickramasinghe et al, 2018 [18]	2014-2016	✓	✓	✓	–	–	–	✓	✓	✓	✓	–	–
Al-Garadi, 2018 [19]	2012-2017	✓	✓	✓	–	–	–	✓	✓	✓	–	✓	–
Chalapathy&Chawla, 2019 [20]	2015-2018	✓	–	–	–	–	–	✓	✓	✓	–	✓	–
Berman et al., 2019 [21]	2014-2018	✓	✓	✓	✓	✓	–	✓	✓	✓	–	✓	–
Mahdavifar&Ghorbani, 2019 [22]	2013-2018	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	–
Sharma&Mangrulkar, 2019 [23]	2009-2018	✓	✓	–	–	–	✓	✓	✓	✓	–	✓	✓
Гайфулина&Котенко, 2020	**2017-2020**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Таблица 2.

Содержание анализа методов глубокого обучения в кибербезопасности

Обзорная работа	Метод	Приме-не-ние	Число цити-рований	Архитек-тура сети	Функция активации	Набор данных	Оценка
Xin et al, 2018 [16]	✓	✓	–	–	–	✓	✓
Apruzzese et al, 2018 [17]	✓	✓	–	–	–	–	–
Wickramasinghe et al, 2018 [18]	✓	✓	–	–	–	–	–
Al-Garadi, 2018 [19]	✓	✓	–	–	–	–	–
Chalapathy&Chawla, 2019 [20]	✓	✓	–	–	–	✓	–
Berman et al., 2019 [21]	✓	✓	✓	–	–	✓	–
Mahdavifar&Ghorbani, 2019 [22]	✓	✓	–	–	–	✓	✓
Sharma&Mangrulkar, 2019 [23]	✓	✓	–	–	–	–	–
Гайфулина&Котенко, 2020	✓	✓	–	✓	✓	✓	✓

Таким образом, в части 2 нашего исследования мы планируем охватить применение большего числа методов глубокого обучения во всех выделенных областях кибербезопасности, проведя анализ работ последних лет. В этом заключается основное отличие от существующих релевантных работ, и выражается новизна проводимого исследования.

Заключение

В данной работе представлено краткое описание распространенных методов глубокого обучения и их классификация по используемой архитектуре. Определены области применения описанных методов в различных приложениях кибербезопасности, включая обнаружение вторжений и вредоносного программного обеспечения, анализ сетевого трафика и некоторые другие задачи. Приведена сравнительная характеристика релевантных обзоров в области глубокого обучения. Таким образом, определен понятийный аппарат исследования, его методологическая основа и актуальность. Способность алгоритмов машинного обучения иерархически извлекать разноуровневые представления из данных большого объема делает их привлекательными для решения задач кибербезопасности.

В продолжение данного исследования будет проведен аналитический обзор методов глубокого обучения в выделенных областях кибербезопасности со ссылками на соответствующие работы. Планируется составить сравнительную характеристику методов по целям применения, используемым моделям, их характеристикам, обучающим наборам данных и эффективности предлагаемого подхода. Подобный анализ позволит сделать вывод о том, какие архитектуры глубоких сетей применимы для конкретной задачи кибербезопасности и какой конфигурацией они обладают.

Рецензент: *Марков Алексей Сергеевич, доктор технических наук, профессор МГТУ им.Н.Э.Баумана, г.Москва, Россия. E-mail: markov@bmstu.ru*

Работа выполнена при частичной финансовой поддержке РФФИ (проект 18-29-22034) и бюджетной темы 0073-2019-0002.

Литература

1. Goodfellow I., Bengio Y., Courville A. Deep learning // MIT press. 2016. 800 p.
2. Nielsen M.A. Neural networks and deep learning // San Francisco, CA, USA: Determination press. 2015. 200 p.
3. Nwankpa C., Ijomah W., Gachagan A., Marshall S. Activation functions: Comparison of trends in practice and research for deep learning // arXiv preprint arXiv:1811.03378. 2018. 20 p.
4. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdino R. Dropout: a simple way to prevent neural networks from overfitting // The journal of machine learning research. 2014. Vol. 15. No. 1. P. 1929-1958.
5. Sutskever I., Martens J., Dahl G., Hinton G. On the importance of initialization and momentum in deep learning // International conference on machine learning, 2013. P. 1139-1147.
6. Le Roux N., Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks // Neural computation. 2008. Vol. 20. No 6. P. 1631-1649.
7. Cilimkovic M. Neural networks and back propagation algorithm // Institute of Technology Blanchardstown, Blanchardstown Road North Dublin. 2015. Vol. 15. P. 1-12.
8. Ranzato M.A., Boureau Y.L., Cun Y.L. Sparse feature learning for deep belief networks // Advances in neural information processing systems. 2008. P. 1185-1192.
9. Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P.A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion // Journal of machine learning research. 2010. Vol. 11. No. Dec. P. 3371-3408.
10. Fischer A., Igel C. Training restricted Boltzmann machines: An introduction // Pattern Recognition. 2014. Vol. 47. No. 1. P. 25-39.
11. Guo J. Backpropagation through time // Unpubl. ms., Harbin Institute of Technology. 2013. Vol. 40. P. 1-6.
12. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Vol. 9. No. 8. P. 1735-1780.
13. Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation // arXiv preprint arXiv:1406.1078. 2014. P. 1-15.
14. LeCun Y., Boser B.E., Denker J.S., Henderson D., Howard R.E., Hubbard W.E., Jackel L.D. Handwritten digit recognition with a back-propagation network // Advances in neural information processing systems. 1990. P. 396-404.
15. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets // Proceedings of the Advances in Neural Information Processing Systems (NIPS), 2014. P. 2672-2680.
16. Xin Y., Kong L., Liu Z., Chen Y., Li Y., Zhu H., Wang C. Machine learning and deep learning methods for cybersecurity // IEEE Access, 2018. Vol. 6. P. 35365-35381.
17. Apruzzese G., Colajanni M., Ferretti L., Guido A., Marchetti M. On the effectiveness of machine and deep learning for cyber security // 2018 10th International Conference on Cyber Conflict (CyCon. IEEE, 2018. P. 371-390.
18. Wickramasinghe C.S., Marino D.L., Amarasinghe K., Manic M. Generalization of deep learning for cyber-physical system security: A survey // IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2018. P. 745-751.
19. Al-Garadi M. A., Mohamed A., Al-Ali A., Du X., Guizani M. A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) security // arXiv preprint arXiv:1807.11023, 2018. P. 1-42.
20. Chalapathy R., Chawla S. Deep Learning for Anomaly Detection: A Survey // arXiv preprint arXiv:1901.03407. 2019. P. 1-50.
21. Berman D.S., Buczak A.L., Chavis J.S., Corbett C.L. A survey of deep learning methods for cyber security // Information, 2019. Vol. 10. No. 4. P. 122.
22. Mahdavifar S., Ghorbani A.A. Application of deep learning to cybersecurity: A survey // Neurocomputing. 2019. Vol. 347. P. 149-176.
23. Sharma B., Mangrulkar R. Deep learning applications in cyber security: a comprehensive review, challenges and prospects // International Journal of Engineering Applied Sciences and Technology. 2019 Vol. 4. Iss 8. P. 148-159.

APPLICATION OF DEEP LEARNING METHODS IN CYBERSECURITY TASKS

Gaifulina D.A.[3], Kotenko I.V. [4]

Abstract

The purpose of the article: *analytical review in the field of deep learning for cybersecurity tasks.*

Research method: *analysis of modern methods of deep learning applied to the issues of cyber security, development of their classification by the used architecture. Analysis of relevant reviews in deep learning.*

3 Diana Gaifulina, Ph.D., Associate Researcher, Computer Security Laboratory, FGBUN St. Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia.
E mail: gaifulina@comsec.spb.ru

4 Igor Kotenko, Dr.Sc., Professor, Head of Laboratory of Computer Security Problems at St. Petersburg Institute for Informatics and Automation of Russian Academy of Science, St. Petersburg, Russia.
E-mail: ivkote@comsec.spb.ru

DOI:10.21681/2311-3456-2020-03-76-86

The result obtained: *the description of common methods of deep learning and their classification. The definition of the application areas of the described methods in various cybersecurity tasks, including detection of intrusions and malicious software, analysis of network traffic, and some other. Comparative characteristics of relevant reviews in the field of deep learning, which allows you to determine further research directions.*

The main contribution of the authors to the research of deep learning methods for cybersecurity tasks is the classification of the subject area, conducting a general and comparative analysis of existing approaches that reflect the current state of scientific research.

Keywords: *data science, machine learning, deep learning, deep neural networks, cybersecurity, intrusion detection, malware detection.*

References

1. Goodfellow I., Bengio Y., Courville A. Deep learning // MIT press. 2016. 800 p.
2. Nielsen M.A. Neural networks and deep learning // San Francisco, CA, USA: Determination press. 2015. 200 p.
3. Nwankpa C., Ijomah W., Gachagan A., Marshall S. Activation functions: Comparison of trends in practice and research for deep learning // arXiv preprint arXiv:1811.03378. 2018. 20 p.
4. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdino R. Dropout: a simple way to prevent neural networks from overfitting // The journal of machine learning research. 2014. Vol. 15. No. 1. P. 1929-1958.
5. Sutskever I., Martens J., Dahl G., Hinton G. On the importance of initialization and momentum in deep learning // International conference on machine learning, 2013. P. 1139-1147.
6. Le Roux N., Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks // Neural computation. 2008. Vol. 20. No 6. P. 1631-1649.
7. Cilimkovic M. Neural networks and back propagation algorithm // Institute of Technology Blanchardstown, Blanchardstown Road North Dublin. 2015. Vol. 15. P. 1-12.
8. Ranzato M.A., Boureau Y.L., Cun Y.L. Sparse feature learning for deep belief networks // Advances in neural information processing systems. 2008. P. 1185-1192.
9. Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P.A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion // Journal of machine learning research. 2010. Vol. 11. No. Dec. P. 3371-3408.
10. Fischer A., Igel C. Training restricted Boltzmann machines: An introduction // Pattern Recognition. 2014. Vol. 47. No. 1. P. 25-39.
11. Guo J. Backpropagation through time // Unpubl. ms., Harbin Institute of Technology. 2013. Vol. 40. P. 1-6.
12. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Vol. 9. No. 8. P. 1735-1780.
13. Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation // arXiv preprint arXiv:1406.1078. 2014. P. 1-15.
14. LeCun Y., Boser B.E., Denker J.S., Henderson D., Howard R.E., Hubbard W.E., Jackel L.D. Handwritten digit recognition with a back-propagation network // Advances in neural information processing systems. 1990. P. 396-404.
15. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets // Proceedings of the Advances in Neural Information Processing Systems (NIPS), 2014. P. 2672-2680.
16. Xin Y., Kong L., Liu Z., Chen Y., Li Y., Zhu H., Wang C. Machine learning and deep learning methods for cybersecurity // IEEE Access, 2018. Vol. 6. P. 35365-35381.
17. Apruzzese G., Colajanni M., Ferretti L., Guido A., Marchetti M. On the effectiveness of machine and deep learning for cyber security // 2018 10th International Conference on Cyber Conflict (CyCon. IEEE, 2018. P. 371-390.
18. Wickramasinghe C.S., Marino D.L., Amarasinghe K., Manic M. Generalization of deep learning for cyber-physical system security: A survey // IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2018. P. 745-751.
19. Al-Garadi M. A., Mohamed A., Al-Ali A., Du X., Guizani M. A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) security // arXiv preprint arXiv:1807.11023, 2018. P. 1-42.
20. Chalapathy R., Chawla S. Deep Learning for Anomaly Detection: A Survey // arXiv preprint arXiv:1901.03407. 2019. P. 1-50.
21. Berman D.S., Buczak A.L., Chavis J.S., Corbett C.L. A survey of deep learning methods for cyber security // Information, 2019. Vol. 10. No. 4. P. 122.
22. Mahdavifar S., Ghorbani A.A. Application of deep learning to cybersecurity: A survey // Neurocomputing. 2019. Vol. 347. P. 149-176.
23. Sharma B., Mangrulkar R. Deep learning applications in cyber security: a comprehensive review, challenges and prospects // International Journal of Engineering Applied Sciences and Technology. 2019 Vol. 4. Iss 8. P. 148-159.