SEQUENTIAL КАК ОСНОВА ЗАЩИТЫ ИНФОРМАЦИОННЫХ СИСТЕМ

Гончаренко Ю.Ю.[1], Погуляй Г.С.[2]

Ключевые слова: *нейронные сети, обучение, обнаружение угроз, математическое моделирование, Python, tensorflow, pandas.*

Аннотация

Цель исследования: выявление возможностей модели нейронной сети Sequential в качестве основного инструмента по обнаружению угроз информационным системам, статистических зависимостей между структурами моделей, построенных на основе Sequential, характеристик, которые могут описывать подозрительную активность.

Методы исследования: математическое моделирование, визуализация данных и программирование на языке программирования Python при помощи библиотек tensorflow, keras, sklearn, numpy, pandas.

Результаты исследования: модель Sequential может выступать в роли основного инструмента для обнаружения подозрительной активности за счёт высоких показателей на основе тестовых данных. Были выявлены статистические зависимости между структурами моделей, где в качестве описывающих метрик выступают точность предсказания и потери. Были определены характеристики, позволяющие классифицировать вид взаимодействия с информационной системой с высокой точностью и минимальной потерей.

DOI: 10.21681/1994-1404-2024-2-44-48

Введение

Неуклонный рост технического прогресса обеспечивает ускорение рабочих процессов, что приводит к дальнейшему экономическому развитию. Одной из важных составляющих технологических процессов являются информационные системы [1, 2], которые накапливают, объединяют, хранят информацию и обеспечивают необходимым функционалом взаимодействие сотрудников без их личного присутствия, что позволяет увеличивать эффективность выполнения работы [3]. Однако сосредоточение больших объемов конфиденциальной информации является чуть ли не главной целью злоумышленников, которые для достижения своих целей могут находить и использовать различные уязвимости информационных систем, что обуславливает необходимость обеспечения их безопасности [4—7]. С этой целью следует создавать не только инструменты, представляющие собой барьеры, но и средства, позволяющие реагировать в режиме реального времени на действия пользователя в информационной системе, определять и классифицировать угрозы, предлагать регламент действий по их предотвращению [8, 9].

Целью исследования является определение возможностей нейронных сетей в области определения и классификации действий внутри информационных систем на основе использования модели нейронных сетей Sequential [10, 11]. Для достижения поставленной цели необходимо решить следующие задачи.

Во-первых, подготовить набор данных для обучения и тестирования нейронной сети, направленной на процесс классификации. Важность данной задачи обуславливается поисками метрик, которые могут описывать выбранную характеристику, а совокупность таких метрик представляет собой набор данных, позволяющий создать математическую модель, описывающую характеристики [12].

Во-вторых, создать модели путём определения её слоёв, а также провести тестирование полученной мо-

[1] ***Гончаренко Юлия Юрьевна****, доктор технических наук, доцент, профессор кафедры «Информационная безопасность» СевГУ, г. Севастополь, Российская Федерация.*

E-mail: yygoncharenko@sevsu.ru

[2] ***Погуляй Геннадий Сергеевич****, студент СевГУ, г. Севастополь, Российская Федерация.*

E-mail: gena.pogulyay.0000@mail.ru

дели для последующего сравнения различных вариантов ее поведения. В качестве статистических данных оценки модели выступают показатели потери и точности предсказания.

Основная часть

Для проведения исследовательской работы были использованы язык программирования Python, в том числе такие библиотеки как

tensorflow, которая представляет собой фреймворк для создание нейронных сетей на основе обучения различных моделей,

- keras, представляющую собой высокоуровневое АПИ для взаимодействия с tensorflow,
- sklearn, аналог tensorflow, необходимый для машинного обучения,
- pandas и numpy для работы с данными, их хранения и обработки [13, 14].

В работе также использовался набор данных NSL-KDD [15], который был целенаправленно создан для использования в качестве набора данных для обучения нейронных сетей. Этот набор данных представляет собой обширный набор различных метрик, которые собираются из журнала событий и сетевого трафика.

Следующий этап представляет собой обработку этих данных, исключение лишних метрик, не несущих никакого веса для определения целевых характеристик, приведение всех параметров к общему виду, приемлемому для обработки нейронной сетью [16, 17]. В результате визуального просмотра и определения смысловой нагрузки каждой отдельной метрики было решено исключить из данных последний столбец. Такое решение обусловлено тем, что этот столбец олицетворяет метрику, описывающую сложность действий пользователя, которая не несёт в себе возможности охарактеризовать действие пользователя. Далее необходимо произвести классификацию действий или же обобщить используемые инструменты в разные виды атак. Ярким примером обобщенной классификации является объединение атак “udp storm” и “tear drop” как атаку DDOS [18, 19]. Следующий шаг обработки этого набора данных представляет собой нормализацию, стандартизацию и трансформацию данных с использованием инструментов в виде функций, представленных библиотекой sklearn. В результате произведённых действий получается набор данных, готовый для использования в обучении нейронных сетей. Таким образом, следующий этап в исследовательской работе представляет собой формирование модели и выстраивание её уровней. Количество и вид уровней непосредственно влияют на качество работы нейронной сети [20, 21]. Такую задачу необходимо решать методом проб и ошибок, так как изначально невозможно однозначно определить оптимальное количество уровней, их вид и очерёдность. В результате тестирования различных комбинаций уровней, представляющих собой Embedding, LSTM, RepeatVector, Dropout, Dense и Flatten, было определено, что оптимальной является следующая последовательность. Первый уровень представляет собой LSTM, который имеет 64 единицы, возврат последовательности и входную форму. Второй уровень — это Dropout с коэффициентом 0,2. Следующий уровень повторяет первый уровень, с таким же количеством единиц и возвратом последовательности, однако уже без входной формы. Четвёртый уровень повторяет собой второй уровень с такой же величиной коэффициента. Пятый уровень — это LSTM с количеством единиц, равным 32, и возвратом последовательности. Шестой уровень — это Flatten без использования каких-либо дополнительных параметров. Седьмой уровень — Dense, который имеет 50 единиц. Последний уровень повторяет предыдущий, однако количество единиц соответствует 4 и в качестве дополнительного параметра уровня используется метод активации, равный softmax.

Таким образом получается готовая модель со всеми необходимыми уровнями для оптимальной работы нейронной сети, которую необходимо скомпилировать. При компиляции модели используется параметр потери в виде категориальной кросс-энтропии, среди параметров также используется оптимизация с помощью алгоритма, называющегося adam.

Таблица 1

Результаты предсказания

Номер	Точность	Сигнал к возврату	F1-оценка	Помощь
0	1.00	1.00	1.00	10635
1	0.97	0.98	0.98	2737
2	0.88	0.84	0.86	740
3	0.99	0.99	0.99	15568

Скомпилированная модель обучается на подготовленном наборе данных. В качестве дополнительных параметров для обучения используется количество эпох, равное 30. В результате обучения получается на выходе набор параметров, включающий в себя показатель потери, равный 0.0429, правильность, составляющую 0.986, и точность, равную 0.9854, что является высоким результатом для нейронных сетей; остаётся провести дополнительное тестирование путём предсказания классификационной характеристики на данных, которые нейронная сеть ещё не вида́ла. Таким образом, после проведения дополнительного тестирования получаем следующие значения параметров: показатель потери — 0.044, правильность — 0.9873, точность — 0.9877, что, в свою очередь, является подтверждением статистики, полученной в ходе обучения нейронной сети.

Используя встроенные возможности библиотеки sklearn, можно получить детальное описание результатов предсказания классификации с помощью функции classification_report. Результаты представлены в табл. 1.

Для понимания результатов необходимо определить соответствие между классификацией, используемой в наборе данных, и упрощенной формой этой классификации. Таким образом, DDOS равен номеру 0, Probe соответствует 1, R2L равен 2 и нормальное поведение равняется 3.

Вывод

В результате практического тестирования Sequential с помощью языка программирования Python было определено, что подобная модель может выступать в качестве основного инструмента по обеспечению защиты информационных систем за счёт больших возможностей по классификации угроз, что позволяет вовремя среагировать на потенциальную угрозу. Была также определена структура слоёв модели, которые позволяют достичь максимальных результатов при работе с реальными данными, полученными из журнала событий. Таким образом, структура модели имеет восемь различных слоёв, которые в результате обучения позволят в кратчайшее время осуществлять работу по классифицировании действий пользователя в информационной системе с наибольшей точностью.

Литература

1. Барабанов А.В., Дорофеев А.В., Марков А.С., Цирлов В.Л. Семь безопасных информационных технологий / Под ред. А.С. Маркова. М. : ДМК Пресс, 2017. 224 с.
2. Карцан И.Н., Жуков А.О. Механизм защиты промышленной сети // Информационные и телекоммуникационные технологии. 2021. № 52. С. 19—26.
3. Горшков Ю.Г. Тестирование средств засекречивания речи // Вопросы кибербезопасности. 2015. № 2 (10). С. 26—30.
4. Марков А.С., Матвеев В.А., Фадин А.А., Цирлов В.Л. Эвристический анализ безопасности программного кода // Вестник МГТУ имени Н.Э. Баумана. Серия: Приборостроение. 2016. № 1 (106). С. 98—111. DOI: 10.18698/0236-3933-2016-1-98-111 .
5. Наумова Е.Г. Государственные информационные системы. К вопросу об информационной безопасности / Е.Г. Наумова, Н.И. Кечкина // Современное государственное и муниципальное управление: в поисках ресурсов и технологий общественного развития : сборник научных трудов Всероссийской научно-практической конференции, Дзержинск, 15 апреля 2021 г. / Редакционная коллегия: А.И. Егоров (председатель), И.Ю. Первухина (зам. председателя). Дзержинск : Аджика, 2021. С. 86—88.
6. Лебеденко А.В., Гончаренко Ю.Ю, Нестеренко В.Р. Система биометрической идентификации сверточной нейронной сети для небольшой организации // Доклады XXIII пленума ФУМО ВО ИБ и Всероссийской научной конференции / Отв. редактор: В.И. Петренко. 2019. С. 22—25.
7. Карцан И.Н. Биометрические данные: новые возможности и риски // Современные инновации, системы и технологии. 2023. Т. 3. № 3. С. 201—211.
8. Бажутова Д.А., Зыков Д.С., Маслова М.А., Олешко А.Ю., Куликов М.А. Поведенческие особенности виртуальной личности на просторах интернета // Инженерный вестник Дона. 2021. № 5 (77). С. 97—107.
9. Маслова М.А. Анализ и определение рисков информационной безопасности // Научный результат. Информационные технологии. 2019. Т. 4. № 1. С. 31—37.
10. Ожиганова М.И. Применение машинного обучения в защите веб-приложений / М.И. Ожиганова, Э.С. Куртаметов // НБИ технологии. 2020. Т. 14. № 2. С. 16—20.
11. Маслова М.А. Обзор существующих методов автоматической генерации тестовых заданий на естественном языке // Computational Nanotechnology. 2023. Т.10. № 4. С. 46—55.
12. Егорова А.О. Математическая модель адаптивной системы защиты информации от утечки по техническим каналам / А.О. Егорова, Е.Н. Тищенко // Вестник УрФО. Безопасность в информационной сфере. 2022. № 2 (44). С. 37—42.
13. Попов А.Ю., Ремез М.В., Жилина Е.В., Ожиганова М.И. Парсинг электронных ресурсов. Библиотека selenium или fake useragent? // Информатизация в цифровой экономике. 2022. Т. 3. № 4. С. 197—210.
14. Петренко А.С., Петренко С.А., Ожиганова М.И. О киберустойчивости и безопасности изобразительных нейросетей // Защита информации. Инсайд. 2023. № 6 (114). С. 50—54.
15. Григорьева Н.М. Атаки на модели данных систем машинного обучения / Н.М. Григорьева, С.А. Петренко, М.И. Ожиганова // Защита информации. Инсайд. 2023. № 4 (112). С. 29—33.
16. Маслова М.А., Дмитриев А.С., Холкин Д.О. Методы распознавания именованных сущностей в русском языке // Инженерный вестник Дона. 2021. № 7 (79). С. 93—105.
17. Герасимов В.М., Маслова М.А. Необходимость комплексной системы защиты биометрического голосового отпечатка от воздействия кибермошенников в сети интернет // Вестник Луганского государственного университета им. В. Даля. 2022. № 5 (59). С. 95—102.

18. Кузьминых Е.С., Маслова М.А. Анализ роста кибератак и рынка информационной безопасности РФ // Научный результат. Информационные технологии. 2023. Т. 8. № 2. С. 11—17.
19. Нестеренко В.Р., Маслова М.А. Современные вызовы и угрозы информационной безопасности публичных облачных решений и способы работы с ними // Научный результат. Информационные технологии. 2021. Т. 6. № 1. С. 48—54.
20. Нестеренко В.Р., Лебеденко А.В. Сравнительный анализ сверточных и импульсных нейронных сетей в контексте задачи распознавания образов // Современные проблемы радиоэлектроники и телекоммуникаций. 2019. № 2. С. 152.
21. Нуриев С.А., Карцан И.Н. Обеспечение безопасности конфиденциальной информации компании при удаленном доступе сотрудника // Современные инновации, системы и технологии. 2023. Т. 3. № 2. С. 234—242.

SEQUENTIAL AS A BASIS FOR INFORMATION SYSTEM PROTECTION

Iuliia Goncharenko, *Dr.Sc. (Technology), Associate Professor, Professor at the Information Security Department of the Sevastopol State University, Sevastopol, Russian Federation.*
E-mail: *yygoncharenko@sevsu.ru*

Gennadii Poguliai, *student at the Sevastopol State University, Sevastopol, Russian Federation.*
E-mail: *gena.pogulyay.0000@mail.ru*

Keywords: *neural networks, learning, threat detection, mathematical modelling, Python, tensorflow, pandas.*

Abstract

Purpose of the study: determining the capabilities of the Sequential neural network model as the main tool for detecting threats for information systems, statistical dependencies between the structures of models based on Sequential, and characteristics capable of describing suspicious activity.

Methods used in the study: mathematical modelling, data visualisation, and programming in the Python programming language using the libraries tensorflow, keras, sklearn, numpy, pandas.

Study findings: the Sequential model can act as the main tool for detecting suspicious activity due to its high performance based on test data. Statistical dependencies between the structures of models where prediction accuracy and losses are the describing metrics, were found. Characteristics allowing to categorise the type of interaction with the information system with high accuracy and minimal loss were identified.

References

1. Barabanov A.V., Dorofeev A.V., Markov A.S., Tsirlov V.L. Sem' bezopasnykh informatsionnykh tekhnologii. Pod red. A.S. Markova. M. : DMK Press, 2017. 224 pp.
2. Kartsan I.N., Zhukov A.O. Mekhanizm zashchity promyshlennoi seti. Informatsionnye i telekommunikatsionnye tekhnologii, 2021, No. 52, pp. 19–26.
3. Gorshkov Iu.G. Testirovanie sredstv zasekrechivaniia rechi. Voprosy kiberbezopasnosti, 2015, No. 2 (10), pp. 26–30.
4. Markov A.S., Matveev V.A., Fadin A.A., Tsirlov V.L. Evristicheskii analiz bezopasnosti programmnogo koda. Vestnik MGTU imeni N.E. Baumana. Seriia: Priborostroenie, 2016, No. 1 (106), pp. 98–111. DOI: 10.18698/0236-3933-2016-1-98-111 .
5. Naumova E.G. Gosudarstvennye informatsionnye sistemy. K voprosu ob informatsionnoi bezopasnosti. E.G. Naumova, N.I. Kechkina. Sovremennoe gosudarstvennoe i munitsipal'noe upravlenie: v poiskakh resursov i tekhnologii obshchestvennogo razvitiia : sbornik nauchnykh trudov Vserossiiskoi nauchno-prakticheskoi konferentsii, Dzerzhinsk, 15 aprelia 2021 g. Redaktsionnaia kollegiia: A.I. Egorov (predsedatel'), I.Iu. Pervukhina (zam. predsedatelia). Dzerzhinsk : Adzhika, 2021, pp. 86–88.

6. Lebedenko A.V., Goncharenko Iu.Iu, Nesterenko V.R. Sistema biometricheskoi identifikatsii svertochnoi neironnoi seti dlia nebol'shoi organizatsii. Doklady XXIII plenuma FUMO VO IB i Vserossiiskoi nauchnoi konferentsii. Otv. redaktor: V.I. Petrenko. 2019, pp. 22–25.
7. Kartsan I.N. Biometricheskie dannye: novye vozmozhnosti i riski. Sovremennye innovatsii, sistemy i tekhnologii, 2023, t. 3, No. 3, pp. 201–211.
8. Bazhutova D.A., Zykov D.S., Maslova M.A., Oleshko A.Iu., Kulikov M.A. Povedencheskie osobennosti virtual'noi lichnosti na prostorakh interneta. Inzhenernyi vestnik Dona, 2021, No. 5 (77), pp. 97–107.
9. Maslova M.A. Analiz i opredelenie riskov informatsionnoi bezopasnosti. Nauchnyi rezul'tat. Informatsionnye tekhnologii, 2019, t. 4, No. 1, pp. 31–37.
10. Ozhiganova M.I. Primenenie mashinnogo obucheniia v zashchite veb-prilozhenii. M.I. Ozhiganova, E.S. Kurtametov. NBI tekhnologii, 2020, t. 14, No. 2, pp. 16–20.
11. Maslova M.A. Obzor sushchestvuiushchikh metodov avtomaticheskoi generatsii testovykh zadanii na estestvennom iazyke. Computational Nanotechnology, 2023, t. 10, No. 4, pp. 46–55.
12. Egorova A.O. Matematicheskaia model' adaptivnoi sistemy zashchity informatsii ot utechki po tekhnicheskim kanalam. A.O. Egorova, E.N. Tishchenko. Vestnik UrFO. Bezopasnost' v informatsionnoi sfere, 2022, No. 2 (44), pp. 37–42.
13. Popov A.Iu., Remez M.V., Zhilina E.V., Ozhiganova M.I. Parsing elektronnykh resursov. Biblioteka selenium ili fake useragent? Informatizatsiia v tsifrovoi ekonomike, 2022, t. 3, No. 4, pp. 197–210.
14. Petrenko A.S., Petrenko S.A., Ozhiganova M.I. O kiberustoichivosti i bezopasnosti izobrazitel'nykh neirosetei. Zashchita informatsii. Insaid, 2023, No. 6 (114), pp. 50–54.
15. Grigor'eva N.M. Ataki na modeli dannykh sistem mashinnogo obucheniia. N.M. Grigor'eva, S.A. Petrenko, M.I. Ozhiganova. Zashchita informatsii. Insaid, 2023, No. 4 (112), pp. 29–33.
16. Maslova M.A., Dmitriev A.S., Kholkin D.O. Metody raspoznavaniia imenovannykh sushchnostei v russkom iazyke. Inzhenernyi vestnik Dona, 2021, No. 7 (79), pp. 93–105.
17. Gerasimov V.M., Maslova M.A. Neobkhodimost' kompleksnoi sistemy zashchity biometricheskogo golosovogo otpechatka ot vozdeistviia kibermoshennikov v seti internet. Vestnik Luganskogo gosudarstvennogo universiteta im. V. Dalia, 2022, No. 5 (59), pp. 95–102.
18. Kuz'minykh E.S., Maslova M.A. Analiz rosta kiberatak i rynka informatsionnoi bezopasnosti RF. Nauchnyi rezul'tat. Informatsionnye tekhnologii, 2023, t. 8, No. 2, pp. 11–17.
19. Nesterenko V.R., Maslova M.A. Sovremennye vyzovy i ugrozy informatsionnoi bezopasnosti publichnykh oblachnykh reshenii i sposoby raboty s nimi. Nauchnyi rezul'tat. Informatsionnye tekhnologii, 2021, t. 6, No. 1, pp. 48–54.
20. Nesterenko V.R., Lebedenko A.V. Sravnitel'nyi analiz svertochnykh i impul'snykh neironnykh setei v kontekste zadachi raspoznavaniia obrazov. Sovremennye problemy radioelektroniki i telekommunikatsii, 2019, No. 2, p. 152.
21. Nuriev S.A., Kartsan I.N. Obespechenie bezopasnosti konfidentsial'noi informatsii kompanii pri udalennom dostupe sotrudnika. Sovremennye innovatsii, sistemy i tekhnologii, 2023, t. 3, No. 2, pp. 234–242.