RAQAMLI IQTISODIYOT VA AXBOROT TEXNOLOGIYALARI
2024
№4
oktabr-dekabr
http://dgeconomy.tsue.uz/
ILMIY ELEKTRON JURNAL

ОБНАРУЖЕНИЕ ФИШИНГОВЫХ АТАК С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ: СИСТЕМАТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ

Мирзаахмедов Дилмурод
Старший преподаватель Ташкентского государственного экономического университета
mirzaakhmedovdilmurod@gmail.com

Аннотация
Многие исследователи используют методы глубокого обучения для обнаружения фишинга. Однако предложенные методы все еще имеют недостатки в производительности, особенно при выявлении неизвестных атак, несмотря на их продвинутую разработку. Для получения более полного понимания текущего состояния исследований по использованию глубокого обучения для обнаружения фишинга необходим систематический обзор литературы (SLR). Цель данного SLR заключается в выявлении методов глубокого обучения, показателей их эффективности, методов предотвращения переобучения, используемых наборов данных, параметров, типов фишинга, а также рекомендаций для будущих исследований в области обнаружения фишинга.

Ключевые слова
Обнаружение, Глубокое обучение, Фишинг, Систематический обзор литературы.

Введение

Фишинг представляет собой атаку, наносящую вред организациям и отдельным пользователям в области кибербезопасности. Многие исследователи используют методы глубокого обучения для обнаружения фишинга. Однако предложенные методы все еще имеют недостатки в производительности, особенно при выявлении неизвестных атак, несмотря на их продвинутую разработку. Для получения более полного понимания текущего состояния исследований по использованию глубокого обучения для обнаружения фишинга необходим систематический обзор литературы (SLR). Цель данного SLR заключается в выявлении методов глубокого обучения, показателей их эффективности, методов предотвращения переобучения, используемых наборов данных, параметров, типов фишинга, а также

рекомендаций для будущих исследований в области обнаружения фишинга. Методология SLR включает постановку исследовательских вопросов и целей, стратегию поиска, критерии включения и исключения, а также извлечение и анализ данных. За последние пять лет SLR выявил 12 качественных статей, посвященных обнаружению фишинга с использованием глубокого обучения. Вклад данного SLR заключается в предоставлении глубокого анализа текущего состояния исследований и определении направлений для будущих исследований в области обнаружения фишинга с применением методов глубокого обучения.

Обзор литературы

Фишинг представляет собой опасную атаку, направленную на отдельных лиц, организации и даже страны [1]. Фишинг — это вид мошенничества, при котором злоумышленники пытаются получить конфиденциальную информацию, такую как данные для входа или учетной записи, выдавая себя за авторитетное лицо или организацию с высокой репутацией через такие сервисы или каналы связи, как электронная почта, социальные сети и другие [2]. Подобные фишинговые действия подпадают под определение преступлений согласно уголовному праву, поскольку они являются недобросовестными поступками, совершаемыми с целью получения личной выгоды или подрыва репутации другого лица [3]. В связи с этим исследователи разработали различные методы предотвращения фишинга, включая машинное обучение и глубокое обучение. Однако в последние годы ученые начали совершенствовать методы обнаружения фишинга, применяя методы глубокого обучения. Это связано с тем, что машинное обучение требует значительных временных затрат, особенно на этапах ручной разработки признаков [4]. Существует множество исследований по обнаружению фишинга на основе глубокого обучения, однако лишь немногие из них тщательно оценивали способность глубокого обучения выявлять фишинговые атаки. Исследователи представляют всесторонний систематический обзор, который охватывает типы фишинга, методы глубокого обучения, оценку эффективности, проблему переобучения, используемые параметры и наборы данных.

Таблица 1

Обзор последних исследований по обнаружению фишинга с использованием глубокого обучения

Исследователь	Период	Типы/Источники фишинга	Метод	Оценка производительности	Параметры эксперимента
[5]	2017-2023		✔	✔	
[6]	2018-2021	✔	✔		✔
[7]	2016-2020	✔	✔		✔
Это исследование	2020-2024	✔	✔	✔	✔

Лишь немногие исследования использовали эту техническую схему для проведения обзора, что связано с тем, что применение методов глубокого обучения для обнаружения фишинговых атак всё ещё является относительно

новым. В таблице 1 представлено сравнение нашего исследования с другими работами, посвящёнными обнаружению фишинга с использованием глубокого обучения. Связанные работы рассмотрены с технической точки зрения, а именно: типы фишинга, используемые методы, оценка производительности, оптимизация параметров, переобучение и наборы данных. Как видно из таблицы 1, большинство исследователей сосредотачиваются на методах, оценке производительности и используемых наборах данных для обнаружения фишинга. Работа [5] была сосредоточена исключительно на обзоре фишинга по электронной почте с использованием глубокого обучения, в то время как в [6] был добавлен обзор параметров, используемых для оптимизации моделей обнаружения фишинга. При этом ни один из этих трёх исследований не рассматривал критически важный параметр модели для обнаружения фишинга — переобучение. Каждый исследователь, занимающийся проблемой фишинга, должен учитывать методы борьбы с переобучением при создании моделей для его обнаружения. Таким образом, целью данного исследования является предоставление глубокого анализа использования методов глубокого обучения для обнаружения фишинга с применением подхода систематического обзора литературы (СОЛ) с акцентом на такие аспекты, как тип/источник фишинга, метод, оценка производительности, параметры оптимизации, переобучение и набор данных.

Проведённый в рамках данного исследования систематический обзор литературы (СОЛ) стремится устранить некоторые недостатки предыдущих исследований. Этот СОЛ вносит фундаментальный вклад в область исследований по обнаружению фишинга с использованием глубокого обучения. Основные вклады данного СОЛ заключаются в следующем:

1. СОЛ описывает текущее состояние исследований, посвящённых обнаружению фишинговых атак с использованием глубокого обучения.
2. СОЛ исследует методы глубокого обучения, которые используют учёные, с особым акцентом на проблему переобучения.
3. СОЛ помогает выявить ограничения возможностей каждого метода глубокого обучения, применяемого для обнаружения фишинга.

Методы

Этот систематический обзор литературы (СОЛ) использует методологию, предложенную в работе [8], с некоторыми модификациями, адаптированными для исследований по обнаружению фишинга. Методология, предложенная в [8], широко известна в исследованиях, основанных на систематическом обзоре литературы. На рисунке 1 процесс СОЛ включает 4 этапа: постановка исследовательских вопросов и целей, стратегия поиска, критерии включения и исключения, а также извлечение и анализ данных.

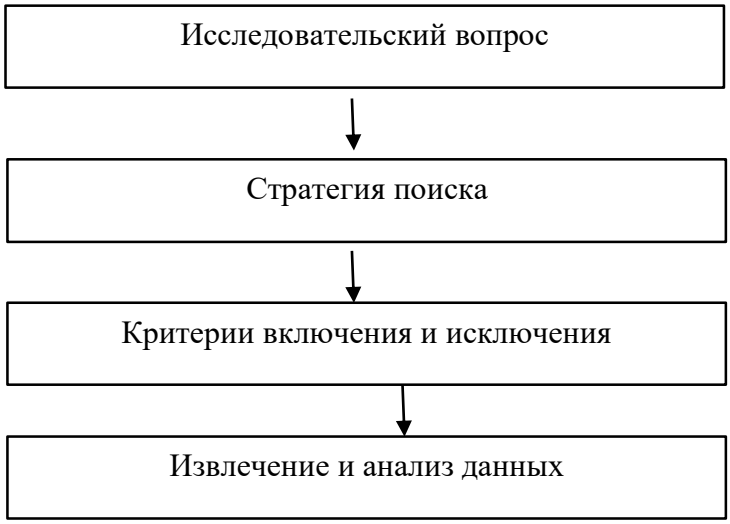

Рисунок 1. Методология систематического обзора литературы (СОЛ)

Исследовательский вопрос (ИВ)

В рамках данного систематического обзора литературы (СОЛ) сформулировано шесть исследовательских вопросов (ИВ), которые поддерживают цели исследования. Основной задачей является предоставление глубокого анализа использования методов глубокого обучения для обнаружения фишинга с использованием подхода систематического обзора литературы (СОЛ) с акцентом на следующие аспекты: тип/источник фишинга, метод, оценка производительности, параметры оптимизации, переобучение и набор данных. Ниже приведены исследовательские вопросы, использованные в рамках (СОЛ):

ИВ1. Какие типы фишинга широко используются в качестве объектов исследования?

ИВ2. Какие методы глубокого обучения чаще всего применяются для обнаружения фишинга?

ИВ3. Какие метрики оценки используются для измерения производительности моделей глубокого обучения?

ИВ4. Какие параметры используют исследователи для повышения эффективности моделей глубокого обучения в обнаружении фишинга?

ИВ5. Какие методы предотвращения переобучения были использованы для повышения качества моделей обнаружения фишинга?

ИВ6. Какие наборы данных использовались для обучения и тестирования предложенных моделей?

Стратегия поиска

В рамках данного исследования стратегия поиска предполагает использование авторитетных научных баз данных, таких как IEEE, ACM, Science Direct, SpringerLink и другие. Для извлечения статей из этих источников в данном СОЛ применяется поисковая система Web of Science (WOS). В качестве ключевых слов для поиска использованы термины "Phishing" (Фишинг), "Detection" (Обнаружение) и "Deep Learning" (Глубокое обучение), которые наиболее полно отражают исследуемую область.

Критерии включения и исключения

В рамках данного процесса критерии, используемые в SLR, включают тип документа и, в частности, наличие в заголовках статей ключевых слов, таких как "Phishing" (Фишинг), "Detection" (Обнаружение) и "Deep Learning" (Глубокое обучение). Кроме того, ограничения по типу документа используются для отбора статей высокого качества, соответствующих целям SLR. Критерии включения предусматривают отбор доступных статей, материалы, не относящиеся к исследовательским вопросам (ИВ), а также статьи, представленные на конференциях, главы из книг, монографии, диссертации и другие публикации, не удовлетворяющие требованиям.

Извлечение и анализ данных

Данный систематический обзор литературы (СОЛ) ограничивает анализ статьями, опубликованными в период с 2020 по 2024 годы, чтобы включить только самые актуальные исследования. После этого статьи будут извлечены и классифицированы по нескольким темам, связанным с исследовательскими вопросами (ИВ), а именно: тип/источник фишинга, метод, оценка производительности, параметры оптимизации, переобучение и наборы данных. Результаты извлечения будут подвергнуты количественному анализу для получения выводов по каждой из указанных тем обсуждения.

Результаты и обсуждение

В рамках данного исследования были проанализированы данные из трёх обзорных статей и 9 исследовательских работ.

Рисунок 2. Процесс отбора наиболее актуальных статей

На рисунке 2 представлен процесс отбора наиболее актуальных статей, посвящённых обнаружению фишинга с использованием методов глубокого обучения. Поисковые запросы были настроены с учётом тематики статей, заголовков и публикаций за период с 2020 по 2024 годы. Затем выбранные статьи оценивались по критериям включения и исключения, таким как доступность статей, их актуальность исследовательским вопросам (ИВ), а также исключались материалы в виде конференционных тезисов, монографий, диссертаций и других публикаций.

Определение поисковых запросов:

Настройка поисковых запросов с учётом тематики, заголовков и периода публикаций (2020–2024 гг.).

Поиск и сбор начального набора статей:

Проведение поиска в научных базах данных.

Сбор первоначального набора статей (3 обзорных и 9 исследовательских работ).

Оценка статей по критериям включения:

Проверяется возможность получить полный текст статьи.

Оценивается соответствие содержания статьи исследовательским вопросам.

Исключение по критериям исключения:

Удаляются материалы, которые являются конференционными тезисами, монографиями, диссертациями и другими несоответствующими публикациями.

Формирование финального набора наиболее актуальных статей:

Составляется окончательный список статей для детального анализа в исследовании.

Анализ научных статей и эффективность метода: Анализ отобранных статей показал высокую эффективность используемого метода. Критерии включения, такие как доступность статей и их актуальность исследовательским вопросам (ИВ), обеспечили фокусировку на наиболее значимых и качественных работах. Исключение материалов по критериям исключения (конференционные тезисы, монографии, диссертации и другие несоответствующие публикации) позволило избежать избыточности и повысить качество обзора.

Заключение

В результате проведенного систематического обзора литературы были изучены современные исследования по обнаружению фишинговых атак с использованием методов глубокого обучения. Несмотря на достижения в этой области, обнаружены недостатки в производительности моделей, особенно при выявлении новых или неизвестных атак. Проблема переобучения остается актуальной и требует дальнейшего изучения.

Строгий отбор статей позволил сосредоточиться на наиболее релевантных и качественных работах за период 2020–2024 годов. Определены ключевые метрики оценки эффективности моделей и параметры, используемые для их оптимизации. Недостаточное внимание к методам предотвращения

переобучения подчеркивает необходимость разработки новых подходов для повышения устойчивости моделей.

Рекомендуется расширить исследования, включив использование разнообразных наборов данных и интеграцию различных методов глубокого обучения. Полученные результаты способствуют углублению понимания текущего состояния исследований и служат основой для разработки более эффективных систем защиты от фишинговых атак. Дальнейшее развитие в этом направлении имеет важное значение для усиления кибербезопасности и противодействия растущим угрозам фишинга.

Список литературы

1. N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” Sensors, vol. 24, no. 7, p. 2077, Mar. 2024, doi:10.3390/s24072077.
2. O. K. Sahingoz, E. BUBEr, and E. Kugu, “DEPHIDES: Deep Learning Based Phishing Detection System,” IEEE Access, vol. 12, pp. 8052–8070, 2024, doi: 10.1109/ACCESS.2024.3352629.
3. R. Brindha, S. Nandagopal, H. Azath, V. Sathana, G. Prasad Joshi, and S. Won Kim, “Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification,” Comput. Mater. Contin., vol. 74, no. 3, pp. 5901–5914, 2023, doi: 10.32604/cmc.2023.030784.
4. M. K. Prabakaran, P. Meenakshi Sundaram, and A. D. Chandrasekar, “An enhanced deep learning based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders,” IET Inf. Secur., vol. 17, no. 3, pp. 423–440, May 2023, doi: 10.1049/ise2.12106.
5. K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, “A Systematic Review on DeepLearning-Based Phishing Email Detection,” Electronics, vol. 12, no. 21, p. 4545, Nov. 2023, doi:10.3390/electronics12214545.
6. N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, “Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions,” IEEE Access, vol. 10, pp. 36429–36463, 2022, doi: 10.1109/ACCESS.2022.3151903.
7. C. Catal, G. Giray, B. Tekinerdogan, S. Kumar, and S. Shukla, “Applications of deep learning for phishing detection: a systematic literature review,” Knowl. Inf. Syst., vol. 64, no. 6, pp. 1457–1500, Jun. 2022, doi: 10.1007/s10115-022-01672-x.
8. B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering – A systematic literature review,” Inf. Softw. Technol., vol.51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.