УДК 621.395.7 (043.2)

В. В. Гуменюк[1]
К. Б. Безпалий[1]

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ З ПІДВИЩЕНОЮ ДОСТОВІРНІСТЮ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

[1] Вінницький національний технічний університет

Анотація

У роботі розглянуто проблеми аналізу кіберзагроз для забезпечення інформаційної безпеки у сучасному світі, значення інтелектуального аналізу даних, зокрема за допомогою нейронних мереж, у виявленні кіберзагроз та розробці стратегій захисту. Розглядаються переваги використання нейронних мереж для аналізу даних та кластеризації об'єктів для виявлення потенційно небезпечних об'єктів, а також питання покращення інтелектуального аналізу даних та їхнього практичного застосування для забезпечення інформаційної безпеки.

Ключові слова: кіберзагрози, інтелектуальний аналіз даних, нейронні мережі, дескриптивний аналіз, аномальні об'єкти, кластерний аналіз, карти Кохонена.

Abstract

The work examines the problems of analyzing cyber threats to ensure information security in the modern world, the importance of intelligent data analysis, in particular with the help of neural networks, in detecting cyber threats and developing protection strategies. The advantages of using neural networks for data analysis and clustering of objects to identify potentially dangerous objects are considered, as well as the issue of improving intelligent data analysis and their practical application to ensure information security.

Keywords: *cyber threats, intelligent data analysis, neural networks, descriptive analysis, anomalous objects, cluster analysis, Kohonen maps.*

Вступ

На сьогодні зростання кількості кіберзагроз є критичною щодо забезпечення інформаційної безпеки користувачів та бізнесу. Аналіз кіберзагроз стає необхідним інструментом для забезпечення безпеки в онлайн середовищі, дозволяє отримати інформацію про методи, які використовують зловмисники, та створює можливість ідентифікувати їх для подальшого захисту інформаційних ресурсів. Інтелектуальний аналіз даних, зокрема за допомогою нейронних мереж, може стати важливим інструментом у боротьбі з кіберзагрозами, і дозволяє виявляти правила та закономірності у великих наборах даних і ефективно виявляти загрози. Нейронні мережі широко використовуються для розпізнавання образів, прогнозування, управління та ідентифікації електронних систем. Вони дозволяють автоматизувати процеси аналізу даних і забезпечують високу точність виявлення кіберзагроз. Тому дослідження, спрямовані на покращення інтелектуального аналізу даних на основі нейронних мереж для виявлення загроз, зокрема веб-ресурсу, є актуальними і мають на меті підвищення достовірності та ефективності виявлення кіберзагроз. Інтелектуальний аналізу даних як засіб обробки інформації дозволить виявляти та застосовувати знання з наявних наборів даних, використовуючи такі техніки, як класифікація, асоціація, кластеризація, прогнозування та інші.

Інтелектуального аналізу даних з підвищеною достовірністю

Інтелектуальний аналіз даних з підвищеною достовірністю на основі математичного апарату нейронних мереж буде включати методи data mining, що дадуть змогу виявляти загрози безпеці веб-ресурсів, а також досліджуються можливості використання нейронних мереж для інтелектуального аналізу даних та виявлення загроз веб-ресурсу, інтелектуальний аналіз даних на основі нейронних мереж та кластеризації для виявлення загроз веб-ресурсу.

Для інтелектуального аналізу даних використовуються різні методи математичної статистики, теорії баз даних, теорії штучного інтелекту, серед яких слід виділити штучні нейронні мережі, генетичні алгоритми, асоціативний аналіз, дерева рішень та інші. Цей аналіз може бути використаний для виявлення загроз і вразливостей на веб-сайтах. Основні завдання включають дескриптивний аналіз, аналіз зв'язків, багатовимірний статистичний аналіз, аналіз тимчасових рядів, розпізнавання, кластеризацію, прогнозування та інші. Загалом інтелектуальний аналіз даних дозволяє виявляти загрози та вразливості, а також розробляти оптимальні стратегії для їхнього нейтралізації і кластеризації даних.

Кластеризація є важливим інструментом для розподілу сукупності об'єктів на групи за певними схожими параметрами. У даному випадку, застосування нейронних мереж для кластеризації може допомогти виявити групи аномальних чи потенційно небезпечних об'єктів, які можуть вказувати на наявність загроз для інформаційної системи. Дані про об'єкти спочатку подаються на вхід нейронної мережі, яка буде навчатися розпізнавати певні патерни або характеристики, які відокремлюють об'єкти на групи. Після навчання мережі здійснюється процес кластеризації, де об'єкти розділяються на групи відповідно до вивчених патернів. Нейронні мережі мають кілька переваг для кластеризації даних, вони можуть автоматично виявляти складні залежності та нелінійні взаємозв'язки між даними. Крім того, навчання нейронних мереж може відбуватися без необхідності вручну визначати алгоритми кластеризації або передбачати структуру даних наперед. Це робить їх ефективним інструментом для виявлення небезпеки в інформаційних системах, особливо, коли дані мають складну структуру або великий обсяг. Задача навчання з учителем включає у себе середовище, вчителя та машину, що навчається. Машина, що навчається, намагається навчитися функції, яка найкраще апроксимує бажаний відгук. Це досягається через мінімізацію функціоналу середнього ризику. Існує кілька способів вирішення цієї задачі, таких як відновлення розподілу ймовірностей, побудова рекурентної послідовності та заміна на функціонал емпіричного ризику.

Нейромережа створюється на основі попередньо оброблених даних, де кожен вхідний вектор представляє поведінку користувача за певний період часу у вигляді проміжних статистичних значень. Навчання мережі відбувається на основі "нормальної" поведінки користувачів, використовуючи дані безпечного періоду. Для навчання наявності загрози використовуються дані, що відповідають критеріям загрози. Після навчання мережі застосовується метод кластерного аналізу. Це допомагає описати властивості кластера на основі аналізу характеристик даних і поділу їх на групи. Визначається структура мережі, включаючи кількість нейронів у шарі Кохонена. Це може бути зроблено на основі характеристик даних та потреб застосування. Вагові коефіцієнти мережі ініціалізуються випадковими значеннями, що відповідають заданим обмеженням для забезпечення нормалізації. Виконується навчання мережі за алгоритмом Кохонена, включаючи визначення нейрона-переможця, корекцію вагових коефіцієнтів та умови закінчення циклу. Умови для завершення циклу навчання включають вичерпання заданої кількості епох, незначних змін вагових коефіцієнтів або вичерпання заданого часу навчання.

Для вдосконалення методу запропоновано проводити додатковий аналіз на основі відстаней між точками у кластері та міжкластерних відстаней. Після першого етапу кластеризації буде проведено аналіз неврахованих кластерів, що може покращити результати аналізу.

Алгоритмом роботи програмного засобу для визначення загроз веб-ресурсу є:

- визначення загроз для веб-ресурсу та запобігання їм;
- використання методу кластеризації на основі карт Кохонена або використання алгоритму вдосконалення методу кластеризації для детальнішого аналізу даних;
- проведення першого етапу кластеризації на основі карт Кохонена;
- аналіз неврахованих кластерів та внесення змін за допомогою вдосконаленого методу кластеризації;
- використання алгоритму аналізу відстаней між точками у кластері та міжкластерних відстаней;
- уточнення та внесення змін до кластерів на основі результатів аналізу відстаней;
- визначення загроз для веб-ресурсу на основі покращених результатів аналізу.

Модель інтелектуального аналізу даних для виявлення загроз веб-ресурсу базується на навчанні нейронної мережі Кохонена та методі кластеризації. Підготовка даних включає у себе збір і аналіз, зокрема таких параметрів як країна підключення, час сесії та об'єм трафіку. Розробка програмного забезпечення буде здійснюватися на основі описаного методу, з можливістю виведення висновків та

графічного представлення результатів аналізу. Після підготовки даних програмний засіб проводитиме навчання нейронної мережі та обробку отриманих результатів для виявлення ознак кіберзагроз.

Розроблено програмне забезпечення на основі описаного методу, використовуючи нейронну мережу Кохонена та метод кластеризації. Після завантаження даних програмний засіб проводитиме навчання нейронної мережі та аналіз отриманих результатів для виявлення ознак кіберзагроз. Результати аналізу представляються у вигляді висновків щодо виявлених ознак кіберзагроз. Після аналізу даних програмний засіб формує рішення про наявність кіберзагроз на веб-ресурсі. Якщо виявлено ознаки кіберзагроз, система генерує відповідне сповіщення, рішення базується на результаті аналізу, проведеному за допомогою навчання нейронної мережі та методу кластеризації. Графічні дані можуть включати у себе візуалізацію різних параметрів, таких як час сесії, обсяг трафіку тощо. Для забезпечення подальшої обробки даних програмний засіб надає можливість вивантажити опрацьовані результати у вигляді файлу. Ці дані можуть бути використані у інших програмах для подальшого аналізу або для прийняття подальших рішень щодо кібербезпеки.

Висновки

Проведено дослідження можливості покращення інтелектуального аналізу даних на основі нейронних мереж для виявлення загроз, зокрема веб-ресурсу. Здійснено розробку модуля аналізу даних на основі нейромережі та вдосконалення методу кластеризації, а також розробку алгоритму для виявлення загроз веб-ресурсу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Securing Web Applications: OWASP Top 10 Vulnerabilities and what to do about them. [Електронний ресурс] – Режим доступу до ресурсу: https://www.ptsecurity.com/ww-en/analytics/knowledge-base/securing-web-applications-top-owasp-threats-and-what-to-do-about-them/
2. Актуальні проблеми Data Mining : навчальний посібник для студентів факультету комп’ютерних наук та кібернетики / О. О. Марченко, Т. В. Россада. К. : КНУ ім. Т. Шевченка, 2017. – 150 с.
3. Data Mining : пошук знань в даних / А. Я. Гладун, Ю. В. Рогушина. – К. : ТОВ «ВД «АДЕФ-Україна», 2016. – 452 с.
4. Безпека веб-додатків: найкращі практики та вразливості [Електронний ресурс] – Режим доступу до ресурсу: https://itproger.com/ua/news/bezopasnost-veb-prilozheniy-luchshie-praktiki-i-uyazvimosti
5. Головні прогнози Check Point щодо кібербезпеки на 2024 рік [Електронний ресурс] – Режим доступу до ресурсу: https://my-itspecialist.com/check-point-top-cybersecurity-predictions-for-2024
6. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак – Чернівці: Технодрук, 2022. – 367 с.
7. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.

Гуменюк В’ячеслав Володимирович – асистент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, Вінниця, e-mail: hvv@vntu.edu.ua

Безпалий Кирило Валерійович – асистент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, Вінниця, e-mail: kyrylo.bezpalyi@vntu.edu.ua

Humeniuk Vyacheslav Volodymyrovych – associate assistant of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: hvv@vntu.edu.ua

Bezpalyi Kyrylo Valeriovitch – associate assistant of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: kyrylo.bezpalyi@vntu.edu.ua