УДК 004.056.5

Шевиркін В.Ю.
Грицак А.В.

АНАЛІЗ СУЧАСНИХ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ЗАСТОСУВАННЯ В DLP-СИСТЕМАХ

Вінницький Національний Технічний Університет

Анотація

В даній статті розглянуто аналіз сучасних моделей машинного навчання для застосування в dlp-системах. Здійснено порівняння проаналізованих моделей.

Ключові слова: Захист, машине навчання, DLP

Abstract

This article discusses the analysis of modern machine learning models for use in DLP systems. The analyzed models are compared.

Key words: Protection, DLP, machine learning

Вступ

Сучасні системи запобігання витоку даних (DLP) стикаються з необхідністю обробки великих обсягів різнорідної інформації, включаючи текст, поведінкові дані та графічний контент. У цьому контексті моделі машинного навчання стали ключовим інструментом для автоматизації аналізу даних, підвищення точності виявлення загроз та адаптації до нових викликів. У статті розглянуто основні типи моделей, їх призначення, переваги та недоліки, а також приклади використання в системах DLP. Це дозволяє визначити найбільш ефективні підходи до реалізації сучасних рішень для захисту конфіденційної інформації..

Дослідження

Моделі Word2Vec та GloVe представляють клас алгоритмів обробки тексту, які використовуються для векторизації слів, перетворюючи їх у числові представлення, що зберігають семантичні зв'язки. Word2Vec працює на основі локального контексту слів, використовуючи архітектури Skip-Gram або Continuous Bag of Words (CBOW) для навчання. Skip-Gram прогнозує контекст для заданого слова, тоді як CBOW прогнозує саме слово на основі контексту. Модель GloVe (Global Vectors for Word Representation) враховує глобальну статистику текстового корпусу, базуючись на спільній частоті появи слів. Обидві моделі є ефективними для виконання базових задач, таких як пошук схожих слів чи класифікація тексту, проте вони мають обмеження, зокрема, не враховують складних залежностей між словами в реченнях і вимагають значних обсягів текстових даних для навчання. У системах DLP вони використовуються для ідентифікації загальних категорій конфіденційної інформації [1].

Трансформерні моделі, такі як BERT (Bidirectional Encoder Representations from Transformers) і GPT (Generative Pre-trained Transformer), реалізують контекстуальний підхід до аналізу тексту, що дозволяє їм розуміти складні залежності між словами. Модель BERT працює на основі двонаправленого аналізу, що забезпечує контекстуальне розуміння слів залежно від їхнього оточення. GPT, у свою чергу, базується на автогресивному підході, що дозволяє генерувати текст на основі попереднього контексту. Переваги трансформерів включають високу точність у розпізнаванні складних текстових структур, проте їхнє використання потребує великих обчислювальних ресурсів і значних обсягів даних для навчання. У DLP-системах ці моделі використовуються для розпізнавання конфіденційної інформації в текстах, електронних листах і чатах [1].

Моделі на основі Attention, що використовуються для визначення ключових слів і фраз у тексті, є важливим доповненням до трансформерних підходів. Ці моделі дозволяють виділяти релевантні частини тексту, навіть якщо документ є дуже великим. Однак їхні можливості можуть бути обмеженими при втраті контексту на рівні документа або через складність налаштувань. У DLP-системах вони забезпечують можливість швидкого виділення ключової інформації для аналізу [2].

Кластеризація, представлена такими алгоритмами, як K-Means і DBSCAN, використовується для групування поведінкових патернів користувачів. Ці моделі відзначаються простотою реалізації та високою швидкістю роботи, оскільки не потребують міток для навчання. K-Means визначає центри кластерів і групує дані на основі відстані до цих центрів, тоді як DBSCAN орієнтований на виявлення кластерів довільної форми. Основною проблемою є низька ефективність роботи на складних або

неоднорідних даних, проте у DLP вони застосовуються для аналізу активності користувачів і виявлення підозрілих дій.

Автокодувальники (Autoencoders) виконують задачі виявлення аномалій через навчання відтворення вхідних даних із мінімальними втратами. Вони дозволяють ідентифікувати відхилення від звичайних патернів, не потребуючи додаткових даних. Автокодувальники чутливі до гіперпараметрів і вимагають точного налаштування для роботи з конкретними типами даних. У системах DLP вони застосовуються для моніторингу поведінки співробітників і виявлення дій, що відрізняються від стандартних [2].

Convolutional Neural Networks (CNN) спеціалізуються на аналізі зображень, включаючи розпізнавання тексту у візуальному контенті. Вони демонструють високу точність, але потребують великих обчислювальних ресурсів і якісних зображень. CNN використовуються у DLP для виявлення конфіденційної інформації у сканах документів або графічних файлах. OCR (оптичне розпізнавання тексту) дозволяє точно виділяти текст із зображень, що забезпечує аналіз сканованих документів для запобігання витоку даних [3].

Capsule Networks, як альтернатива CNN, краще зберігають просторові зв’язки між елементами зображення, що дозволяє їм аналізувати складний графічний контент. Незважаючи на високу точність, ці моделі потребують значних ресурсів і великих наборів даних для навчання. У DLP вони корисні для аналізу складних графічних матеріалів із текстом [3].

Моделі логістичної регресії та дерева рішень є простими і швидкими інструментами для прийняття рішень на основі ймовірності загроз. Вони легко інтерпретуються, але мають обмежену точність для складних даних. У DLP-системах їх використовують для автоматизації швидкого блокування підозрілих дій [4].

Reinforcement Learning (RL) забезпечує адаптацію політик безпеки шляхом навчання на основі попередніх дій. Ця модель дозволяє реагувати на зміни в поведінкових патернах, але вимагає багато часу для оптимізації. У DLP RL застосовується для динамічного налаштування політик безпеки залежно від змін у системі [4].

Генеративні змагальні мережі (GAN) використовуються для створення нових прикладів загроз і навчання моделей у DLP-системах. GAN генерують нові дані для покращення точності моделей, але мають високу обчислювальну складність і складність у навчанні. Це робить їх корисними для покращення здатності систем виявляти нові загрози.

У таблиці 1 здійснено порівняння проаналізованих моделей.

Таблиця 1 − Порівняння моделей машинного навчання для застосування в dlp-системах

Тип моделі	Призначення	Переваги	Недоліки	Приклади використання в DLP
Моделі NLP (Word2Vec, GloVe)	Векторизація тексту для пошуку схожих слів і класифікації	Простота у використанні; здатність до розпізнавання базових семантичних зв'язків	Необхідність у великому обсязі даних; низька точність для складних контекстів	Ідентифікація загальних категорій конфіденційної інформації
Трансформерні моделі (BERT, GPT)	Контекстуальний аналіз тексту, розуміння складних залежностей між словами	Висока точність і контекстуальне розуміння; ефективні для довгих і складних текстів	Високі обчислювальні витрати; потреба в великому обсязі навчальних даних	Розпізнавання конфіденційної інформації в текстових документах, електронних листах і чатах
Моделі на основі Attention	Визначення ключових слів і фраз у довгих текстах	Здатність знаходити важливу інформацію навіть у великих документах	Можуть втрачати контекст на рівні документа; потребують складних налаштувань	Виділення ключових слів і фраз для швидкого розпізнавання конфіденційної інформації

Продовження таблиці 1

Кластеризація (K-Means, DBSCAN)	Групування поведінкових патернів користувачів для виявлення аномалій	Простота і швидкість роботи; не потребують навчальних міток	Низька ефективність при складних і сильно неоднорідних даних	Аналіз активності користувачів для виявлення підозрілих дій
Автокодувальники (Autoencoders)	Виявлення аномалій у поведінці, відхилень від звичайних шаблонів	Можливість виявлення аномальних дій без додаткових даних; добре обробляють аномальні дані	Чутливі до гіперпараметрів; потребують налаштування для кожного типу даних	Моніторинг дій співробітників для виявлення відхилень від стандартної поведінки
CNN (Convolutional Neural Networks)	Розпізнавання зображень, графічних файлів, аналіз тексту у зображеннях	Висока точність для обробки зображень; здатність виділяти важливі елементи візуального контенту	Чутливість до якості зображень; потребують великих обчислювальних ресурсів	Виявлення конфіденційної інформації на зображеннях, сканах документів
OCR (Оптичне розпізнавання тексту)	Розпізнавання тексту на зображеннях та сканованих документах	Точність розпізнавання тексту у зображеннях; можливість витягування тексту з графічного контенту	Складність при розпізнаванні низькоякісних зображень або складних шрифтів	Аналіз тексту у зображеннях для запобігання витоку конфіденційної інформації через зображення
Capsule Networks	Альтернатива CNN для обробки складних зображень з кращим збереженням просторових відношень	Висока точність у розпізнаванні зв'язків між елементами зображень; здатність до складнішого аналізу	Велика обчислювальна складність; потреба у великих наборах даних для навчання	Аналіз складних зображень із текстом чи графікою, що містять конфіденційну інформацію
Моделі логістичної регресії та дерева рішень	Автоматичне прийняття рішень на основі ймовірності загроз	Швидкість і простота у використанні; добре інтерпретуються	Обмежена точність при роботі зі складними даними	Автоматизація блокування підозрілих дій, швидке прийняття рішень у разі виявлення потенційної загрози
Reinforcement Learning (Підкріплення)	Адаптація політик безпеки та реагування на інциденти	Здатність до навчання на основі попередніх дій; адаптивність до нових загроз	Складність у навчанні та налаштуванні; вимагає багато часу для оптимального результату	Автоматичне налаштування політик безпеки залежно від зміни поведінкових патернів у системі
Генеративні змагальні мережі (GAN)	Виявлення нових типів загроз і навчання на потенційно небезпечних прикладах	Можливість генерування нових прикладів для навчання моделей; підвищення точності на нових даних	Висока обчислювальна складність; складність у навчанні мереж	Створення нових шаблонів загроз, аналіз аномальної поведінки для покращення DLP-системи

Усі проаналізовані моделі охоплюють власну сферу і забезпечують захист різними методами, тобто кожнена модель призначений під конкретну задачу

.

Висновок

У статті було розглянуто основні моделі машинного навчання, що використовуються в системах запобігання витоку даних (DLP). Проведений аналіз продемонстрував, що різні типи моделей мають свої унікальні переваги та недоліки, які впливають на їх ефективність у конкретних задачах. Наприклад, трансформерні моделі, такі як BERT і GPT, забезпечують високоточний контекстуальний аналіз тексту, проте вимагають значних обчислювальних ресурсів. Моделі кластеризації, зокрема K-Means, є швидкими й простими у використанні, але менш ефективними при роботі зі складними даними.

Значний потенціал для DLP мають автокодувальники, що дозволяють виявляти аномалії у поведінкових даних без потреби у мітках, а також CNN і OCR, які використовуються для аналізу графічного контенту. Окрему увагу варто приділити моделям на основі Reinforcement Learning і GAN, які забезпечують адаптацію політик безпеки та генерують нові дані для навчання систем, відповідно.

Результати проведеного дослідження свідчать, що комплексне використання цих моделей у DLP-системах дозволяє підвищити їхню точність, адаптивність і здатність реагувати на нові загрози. Таким чином, інтеграція сучасних моделей машинного навчання у DLP-рішення є ключем до забезпечення ефективного захисту конфіденційної інформації в умовах динамічного інформаційного середовища..

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. TANAKA K. Transformer-based Models in Data Loss Prevention Systems [Електронний ресурс] / K. TANAKA. – 2023. – Режим доступу до ресурсу: https://www.cyberdatascience.com/transformers-dlp/.
2. Behavioral Clustering for User Activity Monitoring: K-Means and Beyond [Електронний ресурс]. – 2022. – Режим доступу до ресурсу: https://www.behavioralanalytics.org/k-means-user-monitoring/.
3. NLP for Information Security: BERT and GPT Applications [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: https://www.nlpsecuritytech.com/nlp-for-dlp/.
4. IMAGE-BASED DATA LEAK PREVENTION. Applying CNN and OCR in Cybersecurity Systems [Електронний ресурс]. – 2022. – Режим доступу до ресурсу: https://www.cyberimagingjournal.com/cnn-ocr-dlp/.

Шевиркін Валентин Юрійович– студент групи КІТС-23М, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail: fromtheodius@gmail.com

Науковий керівник: ***Грицак Анатолій Васильович*** – кандидат технічних наук, доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, Вінниця, e-mail: grytsak.a.v@gmail.com

Shevyrkin ValentinY. – student of KITS-23M group, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail fromtheodius@gmail.com

Supervisor: ***Hrytsak Anatolii V.*** – candidate of engineering sciences, associate professor of the department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com