Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження та розробка методів виявлення атак користувачів на рейтинги в рекомендаційних системах
(тема)

Виконав:
студент 2 курсу, групи СШІм-21-2
Іщенко А.І.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник проф. Чалий С.Ф.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ В.О. Філатов
(підпис) (прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту (СШІ)
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____»_____20____р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Іщенко Антону Іллічу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження та розробка методів виявлення атак користувачів на рейтинги в рекомендаційних системах

затверджена наказом університету від 31 березня 20 23 р. № 306Ст
2. Термін подання студентом роботи до екзаменаційної комісії 17 травня 20 23 р.
3. Вихідні дані до роботи наукові публікації та дані Інтернет-джерел

4. Перелік питань, що потрібно опрацювати в роботі
1) Аналіз підходів до побудови рекомендацій з урахуванням шилінг-атак
2) Метод виявлення шилінг-атак з урахуванням неявного зворотного зв'язку від користувачів
3) Експериментальна перевірка отриманих результатів

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри)_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Вивчення предметної області дослідження	05.04.2023	виконано
2	Дослідження літератури	08.04.2023	виконано
3	Вивчення проблеми, що потребує рішення	11.04.2023	виконано
4	Формування постановки задачі дослідження	12.04.2023	виконано
5	Аналіз існуючих підходів	15.04.2023	виконано
6	Концептуальне проектування методу рішення задачі	17.04.2023	виконано
7	Проектування компонентів ПЗ	20.04.2023	виконано
8	Розробка компонентів ПЗ	23.04.2023	виконано
9	Тестування працездатності розробленого ПЗ	28.04.2023	виконано
10	Оформлення пояснювальної записки	05.05.2023	виконано
11	Попередній захист	12.05.2023	виконано
12	Захист перед ЕК	18.05.2023	

Дата видачі завдання 3 квітня 2023 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Чалий С.Ф.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 61 с., 7 рис., 4 табл., 1 дод., 35 джерел.

ЕЛЕКТРОННА КОМЕРЦІЯ, ПЕРСОНАЛІЗАЦІЯ РЕКОМЕНДАЦІЙ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ШИЛЛІНГ-АТАКИ

Об'єкт дослідження – виявлення та запобігання шилінг-атаки в системах рекомендацій.

Предмет дослідження – методи виявлення та аналізу шилінг-атаки у системах рекомендацій.

Мета роботи – розроблення ефективного алгоритму виявлення шилінг-атаки та визначення їх впливу на якість рекомендаційної системи.

Методи дослідження – аналіз існуючих методів виявлення шилінг-атаки, розробка нових методів та їх тестування на відповідних даних, проведення експериментальних досліджень та аналіз отриманих результатів.

Під час виконання роботи було проведено аналіз літературних джерел щодо методів виявлення шилінг-атаки, наукових публікацій щодо їх впливу на якість рекомендаційної системи. Були виділені основні недоліки існуючих методів та запропоновано новий алгоритм виявлення шилінг-атаки на основі аналізу соціальних мереж користувачів та їх взаємодії з рекомендаційною системою.

Розроблений алгоритм був впроваджений у відповідну систему рекомендацій та протестований на реальних даних. На основі отриманих результатів був проведений аналіз впливу шилінг-атаки на якість рекомендацій та обчислені метрики для оцінки ефективності запропонованого методу.

ABSTRACT

Explanatory note: 61 p., 7 fig., 4 tabl., 1 ann., 23 sources.

COLLABORATIVE FILTERING, ECOMMERCE, PERSONALIZATIO N OF RECOMMENDATIONS, RECOMMENDER SYSTEMS, SHILLING ATTACKS.

The object of the research is detection and prevention of shilling attacks in recommendation systems.

The subject of the study is methods of detection and analysis of shilling attacks in recommendation systems.

The purpose of the work is to develop an effective algorithm for detecting shilling attacks and determine their impact on the quality of the recommender system.

Research methods – analysis of existing methods of detecting shilling attacks, development of new methods and their testing on relevant data, conducting experimental studies and analysis of the obtained results.

During the performance of the work, an analysis of literary sources on methods of detecting a shilling attack, scientific publications on their impact on the quality of the recommender system was carried out. The main shortcomings of the existing methods are highlighted and a new algorithm for detecting shilling attacks based on the analysis of users' social networks and their interaction with the recommendation system is proposed.

The developed algorithm is implemented in the relevant recommendation system and tested on real data. Based on the obtained results, an analysis of the impact of the shilling attack on the quality of recommendations was performed and metrics were calculated to evaluate the effectiveness of the proposed method.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ГМ – гібридні методи;

ДР – метод виявлення аномалій на основі динамічного розбиття;

КФ – коллаборативна фільтрація;

РНОК – рекомендації на основі контенту;

РС – рекомендаційна система;

СФ – спільна фільтрація;

СК – спільне кластеризування;

ЧІ – метод виявлення аномальних елементів на основі часових інтервалів.

ВСТУП

Шилінг-атаки – це один з типів атак на системи рекомендацій, що полягає в намаганні вплинути на рекомендації, додавши до системи фальшиві об'єкти, які можуть бути оцінені позитивно або негативно. Це може бути зроблено з метою зміни сприйняття системою об'єкта, або для отримання певного вигоди.

Одним з найпоширеніших прикладів шилінг-атаки є додавання багатьох позитивних оцінок для продукту, який не має високої якості. Це може вплинути на рекомендації, що надаються іншим користувачам, і збільшити продажі продукту, незважаючи на його недоліки.

Для запобігання шилінг-атаки, системи рекомендацій використовують різноманітні методи аналізу даних, такі як виявлення аномальних оцінок та кореляцій між користувачами. Також можуть використовуватися методи аутентифікації користувачів та контролю над внесенням нових об'єктів до системи.

Шилінг-атаки є важливою темою для дослідження в галузі рекомендаційних систем, і розуміння цієї проблеми допоможе зробити такі системи більш ефективними та надійними.

Зважаючи на те, що шилінг-атаки є загрозою для різних видів систем, можна розглянути аналітику на прикладі соціальної мережі Twitter, де такі атаки досить поширені.

За даними аналітичної компанії SEMrush, у 2020 році було виявлено понад 55 тисяч акаунтів, які використовувалися для проведення шилінг-атаки на Twitter. Відповідно до дослідження, проведеного University College London, приблизно 9% активних користувачів Twitter є шахраями, які намагаються впливати на думку та поведінку інших користувачів.

Щодо методів, які використовуються для проведення шилінг-атаки на Twitter, можна виділити кілька. Зокрема, це можуть бути платні послуги, де злоумисники наймають інших користувачів для того, щоб вони твітили

певні повідомлення або впливали на обговорення деяких тем. Також можуть використовуватися боти та автоматизовані скрипти, які дозволяють збільшувати кількість лайків, ретвітів та коментарів, що підвищує популярність певного контенту.

Однією з основних причин, які спричиняють шилінг-атаки, є намагання вплинути на поведінку інших користувачів, що може бути використано для здійснення різних маніпуляцій, таких як збільшення продажів або вплив на громадську думку. Це може бути особливо небезпечно, якщо такі маніпуляції виконуються у контексті виборів або політичних кампаній.

Щоб попередити шилінг-атаки, системи рекомендацій можуть використовувати різноманітні підходи, такі як аналіз оцінок користувачів, виявлення аномальних поведінок і використання методів машинного навчання для виявлення ознак шилінг-атаки. Також можуть використовуватися методи аутентифікації користувачів та контролю над внесенням нових об'єктів до системи.

Шилінг-атаки – це важлива проблема, яка може вплинути на ефективність і надійність систем рекомендацій. Розуміння цієї проблеми допоможе розробити ефективніші і більш надійні алгоритми рекомендацій та забезпечити захист від таких атак.

Шляхом дослідження шилінг-атаки можна розробити різні методи захисту від цих атак. Одним з них є використання складних алгоритмів розпізнавання аномальних поведінок, які можуть виявляти незвичні маловірогі дії, такі як надмірне внесення оцінок або незвичні зміни поведінки користувачів. Також можливо використовувати методи аутентифікації користувачів та контролю над внесенням нових об'єктів до системи.

Сам факт суттєвості шкоди, нанесеної цією атакою, показує, наскільки серйозними можуть бути наслідки для системи рекомендацій, що в більш комерційних системах може завдати неабияких збитків бізнесам, що

стали жертвами шиллінг-атак, тому є доцільним розгляд потенційних методів боротьби з ними..

В цілому, розуміння шилінг-атаки є важливим для розробників систем рекомендацій та для тих, хто користується цими системами, оскільки це допоможе забезпечити їх ефективність та надійність. .

1 АНАЛІЗ ПІДХОДІВ ДО ПОБУДОВИ РЕКОМЕНДАЦІЙ З УРАХУВАННЯМ ШИЛІНГ–АТАК

1.1 Аналіз характеристик рекомендаційних систем

Рекомендаційні системи – це програмні засоби, що аналізують дані про користувачів та їх взаємодію з різними об'єктами (такими як товари, статті, музика, фільми тощо), з метою рекомендації об'єктів, які користувачі можуть бути зацікавлені в перспективі. Рекомендаційні системи можуть бути використані у багатьох сферах, включаючи електронну комерцію, медіа, соціальні мережі та багато інших.

Рекомендації стають одним з найважливіших методів для надання документів, товарів та співробітників для задоволення потреб користувачів в інформації, торгівлі та послугах для суспільства (соціальних послуг), будь то через мобільний пристрій чи в Інтернеті [1].

Кількість даних та інформації зростає щоденно, що призводить до перевантаження інформацією та даними. У такий момент пошук потреб та тенденцій клієнтів став важливою проблемою.

Один з інноваційних засобів, були пошукові системи, але вони не можуть персоналізувати інформацію. Розробники систем запропонували рішення цієї проблеми – систему рекомендацій.

Ця система використовується для сортування та фільтрації інформації, даних та об’єктів. Системи рекомендацій використовують ідеї користувачів про соціум або спільноту, щоб допомогти ефективно реалізувати потреби та попит користувачів у суспільстві з можливо складним набором варіантів вибору [2].

Основною метою системи рекомендацій є створення значущих рекомендацій та пропозицій щодо інформації, товарів або об'єктів для суспільства користувачів, які можуть їх зацікавити. Наприклад, на сайті Amazon рекомендують книги, Netflix рекомендує фільми, використовуючи

системи рекомендацій для визначення тенденцій користувачів та приваблюючи їх все більше й більше [3].

Існує багато різних методів та алгоритмів, які можуть допомогти рекомендаційним системам створювати персоналізовані рекомендації. Усі підходи можна поділити на три категорії (таблиця 1.1):

– рекомендації на основі контенту: цей метод рекомендує об'єкти та інформацію, які схожі за змістом на об'єкти, які користувачі вже цікавились раніше, або порівнюються та відповідають характеристикам користувачів;

– СФ: системи рекомендують об'єкти та інформацію користувачеві на основі історичної оцінки всіх користувачів спільно;

– ГБ є поєднанням методів РНОК та СФ.

Таблиця 1.1 – Методи надання рекомендацій користувачам

Категорія	Рекомендації на основі контенту	СФ	ГМ
Опис	Рекомендації засновані на схожості за змістом об'єктів та інформації, що вже цікавили користувачів.	Рекомендації засновані на історичній оцінці всіх користувачів.	Поєднання методів рекомендацій на основі контенту та спільної фільтрації.
Переваги	Легко зрозуміти та пояснити, ефективні при обмеженій кількості даних про користувача.	Ефективний при великій кількості даних, враховує інформацію про користувачів.	Зменшує недоліки окремих методів та використовує їх переваги, ефективний в більшості випадків.
Недоліки	Обмежений у виборі рекомендацій, не враховує інформацію про користувача, яка не пов'язана з контентом.	Проблеми з холодним стартом, неефективний при нових користувачах, складний у використанні.	Складний у реалізації та обчислювально витратний.

СФ є ефективним методом категоризації об'єктів, який довів свою ефективність у передбаченні вибору клієнтом об'єктів за їхніми перевагами. Цей метод розроблено для роботи з величезними базами даних, і зарекомендував себе з поширенням систем рекомендацій в середині 1990-х років, таких як Netflix, Amazon та Elsevier.

СФ автоматизує процедуру рекомендацій за попередніми вподобаннями інших користувачів з подібними інтересами, замінюючи «вустами до вуст» рекомендації, що забезпечуються іншими користувачами. Goldberg та інші вчені використали CF для створення системи фільтрації, яка надавала користувачам можливість пояснювати свої електронні листи та документи. Ці методи автоматизують процес визначення близьких сусідів користувача, які можуть допомогти у пошуку документів, що цікавлять.

Найпоширенішою технікою у рекомендаціях є колаборативна рекомендація. Зазвичай корисності базуються на оцінках, які користувачі надали для предметів, з якими вони вже знайомі.

Головною перевагою колаборативної рекомендації є її простота. Проблема обчислення корисності перетворюється на проблему екстраполяції відсутніх значень у матриці оцінок, де кожен користувач - окремий рядок, кожен предмет - окремий стовпець, а значення - відомі оцінки. Це уявлення може бути реалізоване різними способами. Спочатку використовувалися методи найближчих сусідів для знаходження груп користувачів з схожими смаками.

Алгоритми КФ допомагають користувачам ділитися інформацією та документами, які є схожими, використовуючи патерни, що демонструють їхні переваги та взаємодії. Після визначення можливого збігу, алгоритми генерують рекомендації. Для визначення переваг користувачів використовуються патерни, які можуть бути отримані безпосередньо від них. Наприклад, на сайті Amazon користувачі можуть відсортувати об'єкти

від A до E, а КФ використовує матрицю для оцінки об'єктів після збору неявної або явної думки користувачів.

Алгоритми КФ зазвичай розділяються на дві частини:

– алгоритми, основані на моделі;

– алгоритми, основані на пам'яті.

Заснований на пам’яті КФ. Інший термін для алгоритмів на основі пам'яті – це «ліниві» рекомендаційні алгоритми. Вони відкладають розрахунки для передбачення пріоритетів клієнтів щодо об'єкта до моменту, коли клієнти запитують колекцію рекомендацій. Етап навчання алгоритму на основі пам'яті включає зберігання усіх рейтингів клієнтів в пам'яті. Існують два різних рекомендаційних алгоритми на основі пам'яті, які базуються на алгоритмі k-найближчих сусідів:

– фільтрування на основі об'єктів;

– фільтрування на основі клієнтів.

Фільтрування на основі об'єктів запропонував Sarwar в 2001. Воно складається зі знаходження найбільш подібних об'єктів. Об'єкти вважаються схожими, коли на них високі рейтинги або коли їх купували ті самі клієнти. Для кожного об'єкта, який належить активному клієнту, визначається його найбільш ймовірне середовище. Кожен найкращий k-сусід стає кандидатом із вказівкою на його подібність до об'єкту активного користувача. Співставляються балами схожості об'єктів, які зустрічаються декілька разів у списку кандидатів. Кандидати сортуються за цими накопиченими балами подібності, і верхні N рекомендацій представляються клієнту.

Фільтрація на основі користувачів/клієнтів відповідає активному користувачеві/клієнту, порівнюючи матрицю рейтингу, для пошуку сусідів активного користувача, з якими користувач мав спільне минуле співпраці. Спочатку визначаються всі сусіди, всі об'єкти у профілі яких належать до сусідів, які є незнайомими для активного користувача, розглядаються як

можливі рекомендації та класифікуються у сусідній квартал за їх частотою. Рекомендації генеруються на основі накопиченого значення цих частот.

Алгоритми КФ мають перевагу в тому, що вони допомагають уникнути повторення рекомендацій, які були вже запропоновані користувачеві. Це досягається завдяки використанню історії взаємодії користувача з системою, де зберігаються відомості про його попередні вибори та предмети, які він вже переглядав або купив. Застосування CF дозволяє системі враховувати цю інформацію та уникнути рекомендацій тих самих або схожих об'єктів, які користувач вже знайомий.

Хоча алгоритми КФ мають декілька переваг і з часом покращують якість рекомендацій, однією з основних проблем є «холодний старт», коли у системі є багато об'єктів і товарів, але мало клієнтів і рейтингів. Цю проблему можна вирішити шляхом використання інших наборів даних для заповнення системи та застосування алгоритмів, які не мають проблеми «холодного старту». Навіть після отримання більшої кількості рейтингів, недостатня кількість матриці «клієнт-об'єкт» може бути проблемою для КФ. Іншою проблемою є «сірі овечки», коли рекомендаційна система складна для людей, які не належать до очевидної групи. КФ працює добре для клієнтів, які належать до певної групи з багато схожими сусідами [4].

Масштабованість – наступний виклик для КФ. Коли кількість об'єктів та користувачів зростає, традиційна форма КФ стикається зі значними проблемами масштабованості. Наприклад, при великій кількості користувачів та об'єктів збільшується складність КФ. У цей час нам потрібні багато систем, щоб відповісти на запити в режимі реального часу, тому що ми потребуємо високого рівня масштабованості КФ.

Ще однією викликом, якому стикається КФ, є проблема синонімів. Ця проблема пов'язана з тим, що багато дуже схожих об'єктів мають різні назви. Рекомендаційні системи зазвичай не в змозі розпізнати цю проблему і тому розглядають ці об'єкти як різні. Фактично, продуктивність КФ буде зменшуватись за наявності синонімів. Ще одним викликом для

рекомендаційних систем є шилінг-атаки. Це означає, що кожен об'єкт або товар може бути оцінений кожним клієнтом порівняно з іншими об'єктами, які належать іншим людям. Клієнти можуть надавати більш високі оцінки своїм власним об'єктам або навіть негативні оцінки конкуруючим продуктам. Тому в багатьох випадках системи КФ повинні встановлювати заходи безпеки для запобігання шилінг-атакам.

Інший відомий алгоритм рекомендацій – це алгоритми РНОК. Ці алгоритми можна розглядати як розширену роботу, яка виконується при фільтрації інформації. Зазвичай методи фільтрації на основі контенту спрямовані на побудову кількох типів представлення контенту в системі, а потім навчання профілю переваг клієнтів. Потім, представлення контенту порівнюють з профілем переваг клієнта, щоб знайти об'єкти, які найбільше пов'язані з цим клієнтом. Як і в КФ представлення профілю переваг клієнта є моделями, які є довгостроковими, і ми також можемо оновлювати профіль переваг, що робить цю роботу більш доступною [5].

Одним із ключових питань у рекомендаціях на основі вмісту є якість ознак. Об'єкти, які планується рекомендувати, повинні бути описані так, щоб відбулося значуще навчання користувацьких вподобань. Ідеально, кожен об'єкт повинен бути описаний на одному рівні деталей, а набір ознак повинен містити описи, які корелюють з розрізненнями, зробленими користувачами. Незадовільно, це часто не відбувається. Описи можуть бути неповними, або деякі частини простору об'єктів можуть бути описані докладніше, ніж інші.

Важливою є також відповідність між набором ознак та функцією корисності користувача.

Загалом, метод рекомендацій на основі контенту має проблему, де представлення документа повинні бути зіставлені з представленням клієнта за схожістю тексту, коли контент, який є текстом представлень, є однорідним і використовується для навчання алгоритму передбачення.

Один з найбільш очевидних переваг алгоритмів фільтрації на основі контенту полягає в тому, що для їх використання не потрібно спеціальних знань з домену. Достатньо зібрати відгуки від користувачів про їхні переваги. Ще однією перевагою алгоритмів фільтрації на основі контенту є те, що вони краще за алгоритми КФ знаходять місцево подібні об'єкти. Це тому, що явний фокус алгоритмів фільтрації на основі контенту спрямований на схожість тексту.

Гібридні системи рекомендацій адаптовані для поєднання РНОК та КФ, які збільшують переваги та зменшують недоліки обох технік. Тому гібридні системи рекомендацій працюють з характеристиками, які пов'язані з обома методами. Дійсно, є багато підходів, які можна об'єднати РНОК та КФ [6].

Декілька підходів до поєднання, які використовуються для створення гібридних систем рекомендацій, є наступні:

– змішаний метод. Метод, що передбачає, що пропозиції та рекомендації, які рекомендуються з різних систем рекомендацій, представлені одночасно;

– зважений метод. Метод, що полягає в тому, що одну рекомендацію виробляють, використовуючи голоси та оцінки, що надаються різними системами рекомендацій;

– комбінація ознак. Характеристики, що стосуються різних джерел даних рекомендацій, об'єднуються в один алгоритм системи рекомендацій.

– каскадний підхід. Одна система рекомендацій фільтрує пропозиції та рекомендації, які надаються іншою системою рекомендацій.

– посилення ознак. Результати одного методу використовуються як вхідні дані та характеристики для іншого методу рекомендацій.

– підхід мета-рівня. Підхід, в якому метод, навчений однією системою рекомендацій, використовується як вхід для іншої системи рекомендацій.

– підхід перемикання. У цьому методі система рекомендацій перемикається між різними методами рекомендацій залежно від поточної ситуації.

За останні часи, зі зростанням технологій та кількості даних, нам потрібна система, яка допоможе людям знайти свої інтереси та предмети [7]. Ці підходи мають кілька переваг та недоліків, на які це дослідження спрямоване, зокрема на методи рекомендацій та їх слабкі місця. Незважаючи на те, що системи рекомендацій з такими умовами допомагають користувачам виявляти їхні уподобання, вони повинні бути постійно вдосконалювані.

Процес роботи рекомендаційної системи складається з трьох основних етапів (рис. 1.1): збору даних, їх аналізу та формуванню рекомендацій для користувачів.

На етапі збору даних рекомендаційна система збирає інформацію про користувачів та їх взаємодії з платформою, таку як перегляди сторінок, оцінки продуктів, дії на сайті тощо. Ці дані можуть бути як імовірнісними описами взаємодії користувача з платформою, так і структурованими даними.

Після збору даних наступним кроком є їх аналіз. На цьому етапі використовуються методи машинного навчання та статистичний аналіз для виявлення закономірностей в поведінці користувачів. Наприклад, можуть застосовуватися алгоритми кластеризації для виявлення груп користувачів зі схожими інтересами.

Останнім етапом є формування рекомендацій для користувачів. На основі результатів аналізу даних рекомендаційна система намагається зробити прогноз щодо того, які продукти або послуги можуть бути цікавими користувачу.

Для цього можуть використовуватися різні методи побудови рекомендацій, такі як рекомендації на основі контенту, спільної фільтрації або гібридні методи.

Рисунок 1.1 – Процес роботи РС

Аналіз рекомендаційних систем зазвичай включає в себе дослідження різних аспектів, таких як ефективність, точність та стійкість до різних видів атак. Деякі з основних аспектів аналізу рекомендаційних систем описані нижче.

Метрики ефективності. Одним з ключових аспектів аналізу рекомендаційних систем є оцінка їх ефективності. Це може бути зроблено за допомогою різних метрик, таких як точність, покриття, диверсифікація та інші. Наприклад, точність системи може бути виміряна за допомогою метрики середньої або медіанної абсолютної помилки (MAE або MRE) або коефіцієнта кореляції Пірсона.

Аналіз атак на систему. Рекомендаційні системи можуть бути підвернуті різним видам атак, таким як атаки підробки (зірвання легітимного користувача), атаки внесення змін (зміна рейтингів об'єктів) та інші. Для аналізу стійкості системи до таких атак, можуть бути використані різні методи, такі як аналіз відповідності (conformance analysis), аналіз вразливостей (vulnerability analysis) та аналіз ризиків (risk analysis).

Визначення рекомендаційного алгоритму. Рекомендаційні системи можуть використовувати різні алгоритми для прогнозування зацікавленості користувачів в об'єктах. Деякі з найпоширеніших алгоритмів включають

колаборативний фільтр, контент-базований фільтр, гібридний фільтр тощо. Для аналізу рекомендаційних систем може бути важливим визначення того, який алгоритм використовується та які параметри використовуються для підгонки моделі.

Обробка даних та візуалізація. Рекомендаційні системи зазвичай працюють з великою кількістю даних, тому обробка даних та їх візуалізація можуть бути важливими етапами аналізу. Це може включати очищення даних, обробку та агрегацію даних, статистичний аналіз та візуалізацію даних за допомогою графіків та діаграм.

Розгляд різних варіантів рекомендаційних систем. Оскільки рекомендаційні системи можуть використовувати різні алгоритми та методи, важливо розглянути різні варіанти рекомендаційних систем. Наприклад, системи, які використовують колаборативний фільтр, можуть бути порівняні з системами, які використовують контент-базований фільтр. Також можна розглядати гібридні системи, які використовують комбінацію різних методів.

Отже, аналіз рекомендаційних систем може бути важливим етапом вивчення проблеми шилінг-атаки. Це може допомогти зрозуміти, які фактори можуть вплинути на ефективність рекомендаційних систем та їх вразливість до атак.

Оцінка ефективності рекомендаційних систем. Оцінка ефективності рекомендаційних систем може включати різні метрики, такі як точність, повнота, F-міра тощо. Ці метрики можуть допомогти визначити, наскільки добре рекомендаційна система працює та наскільки корисні рекомендації для користувачів.

Аналіз впливу шилінг-атаки. Однією з важливих проблем, які можуть виникати в рекомендаційних системах, є шилінг-атаки. Ці атаки можуть бути використані для зламу системи та зміни рекомендацій на користь певних об'єктів або користувачів. Аналіз впливу шилінг-атаки може

допомогти виявити слабкі місця рекомендаційних систем та знайти способи їх підвищення.

Розробка заходів захисту. На основі аналізу рекомендаційних систем та їх вразливостей можна розробити заходи захисту, щоб зменшити вплив Шилінг-атаки та інших атак. Ці заходи можуть включати в себе використання алгоритмів, які зменшують вплив фальшивих даних, перевірку та аналіз поведінки користувачів тощо.

Експериментальна перевірка. Для перевірки ефективності заходів захисту та оцінки ефективності рекомендаційних систем можуть бути проведені експериментальні дослідження. Наприклад, можна провести порівняльний аналіз різних алгоритмів та методів захисту для рекомендаційних систем на основі метрик ефективності, таких як точність, повнота, F-міра тощо.

Розвиток рекомендаційних систем. Розвиток рекомендаційних систем є постійним процесом, який включає в себе вдосконалення алгоритмів та методів, що використовуються для рекомендацій, а також розширення функціональності системи. Наприклад, можуть бути додані нові функції, такі як підтримка різних типів контенту, підтримка різних форматів даних тощо.

Перевірка коректності та етики. Рекомендаційні системи можуть мати значний вплив на поведінку та вибір користувачів. Інтелектуальність РС зображено на рисунку 1.2.

Тому важливо перевіряти коректність та етику рекомендацій, щоб забезпечити, що вони не мають шкідливих наслідків для користувачів або для суспільства в цілому.

Наприклад, можна проводити експертні оцінки рекомендацій для перевірки їх етичності та безпеки [8].

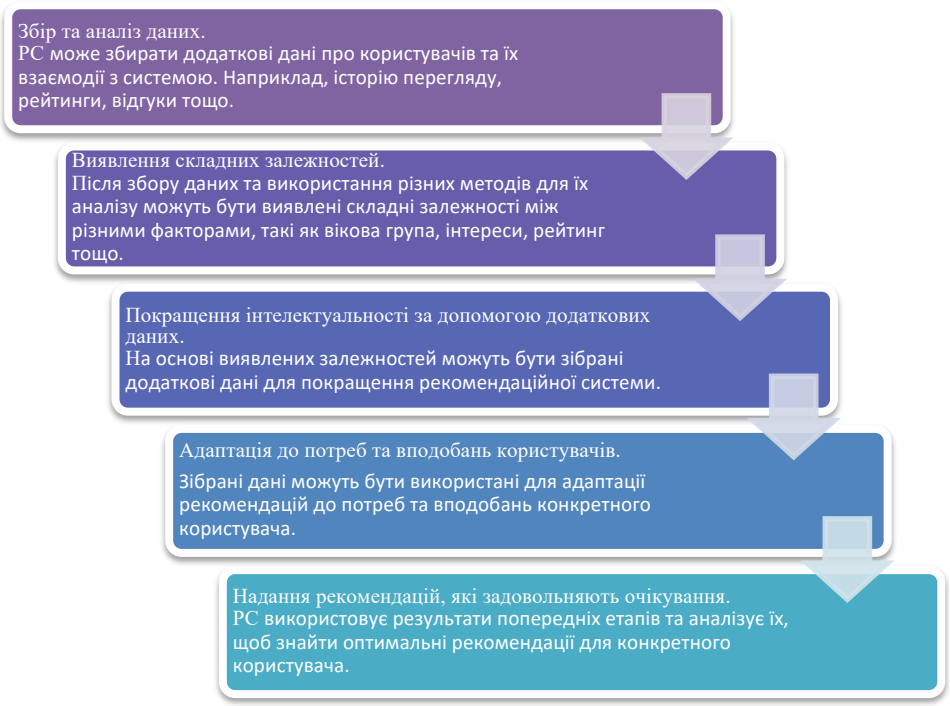

Рисунок 1.2 – Інтелектуальність РС

Загалом, аналіз рекомендаційних систем включає в себе дослідження різних аспектів системи, таких як алгоритми, джерела даних, метрики ефективності, вразливості (рис. 1.3) до атак та заходи захисту.

При аналізі рекомендаційних систем важливо мати на увазі їхню конкретну мету та контекст використання.

Наприклад, рекомендаційна система для онлайн-магазину може мати інші вимоги до ефективності та безпеки, ніж рекомендаційна система для медичної діагностики.

Зважаючи на те, що рекомендаційні системи стають все більш поширеними та впливовими, їхній аналіз та вдосконалення стають надзвичайно важливими завданнями. Нехтування цими аспектами може мати шкідливі наслідки для користувачів та суспільства в цілому [9].

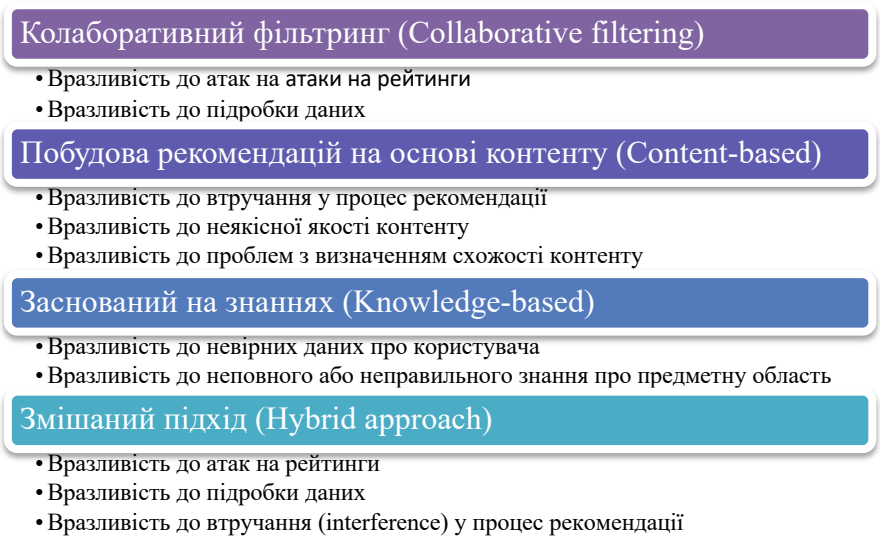

Рисунок 1.3 – Вразливості методів побудови РС

Для аналізу рекомендаційних систем важливо звернути увагу на різноманітні аспекти, такі як:

– типи рекомендацій. існують різні типи рекомендацій, такі як персоналізовані, популярні, нові, схожі тощо. Для кожного типу рекомендацій необхідні свої алгоритми та методи.

– моделі рекомендацій. існує багато різних моделей рекомендацій, таких як колаборативні фільтри, контентні фільтри, гібридні моделі тощо. Кожна з них має свої переваги та недоліки, тому важливо вибрати найбільш підходящу для конкретного випадку використання.

– оцінка ефективності. для аналізу рекомендаційної системи необхідно використовувати показники ефективності, такі як точність, покриття, ранжування тощо.

– захист від Шилінг-атаки. Шилінг-атаки є серйозною проблемою для рекомендаційних систем, тому важливо розуміти, як працюють ці атаки та як захистити систему від них.

– етика використання. важливо забезпечити етичне використання рекомендаційних систем та захистити користувачів від можливих наслідків, таких як дискримінація або порушення приватності.

Усі ці аспекти пов'язані між собою та взаємопов'язані. Аналіз рекомендаційної системи повинен бути комплексним та охоплювати всі ці аспекти, щоб забезпечити оптимальну ефективність та захист користувачів [10].

Навіть прості атаки можуть впливати на поведінку найпоширеніших алгоритмів рекомендаційних систем. В таких системах існує кілька видів атак, які можуть завдати різних шкод незахищеним системам в залежності від намірів зловмисників.

По-перше, одна з небезпек атак полягає у несправедливому представленні користувачів у рекомендаційних системах. Це означає, що деякі користувачі можуть бути недооцінені або ігноруватися, тоді як інші можуть бути перебільшені або надмірно виділені.

По-друге, атаки можуть призводити до неправильних або недостатніх рекомендацій для користувачів. Це може погіршити репутацію рекомендаційних систем, які стали жертвами таких атак. Також варто зазначити, що в певних умовах багато атак можуть викликати збої у роботі системи, призводячи до її некоректної роботи або навіть повного паралізу. Уникнення вводу фальшивих даних (профілів) у систему є складною задачею при боротьбі з недобросовісними користувачами. Для забезпечення надійності рекомендаційних систем необхідно точно виявляти та видаляти атакуючі профілі.

В цілому, розуміння та захист від шилінг-атак є важливими аспектами для підвищення надійності, ефективності та довіри до рекомендаційних систем.

1.2 Структуризація атак користувачів на рейтинги

Шилінг-атаки – це вид атак на рекомендаційні системи, при якому зловмисники намагаються змінити рекомендації, що надходять користувачеві, шляхом створення фальшивих облікових записів або

збільшення активності вже існуючих облікових записів. Зловмисники використовують ці фальшиві облікові записи, щоб впливати на алгоритми рекомендацій, що призводить до некоректних рекомендацій для користувачів.

Щоб зрозуміти, які властивості використовуються в Шилінг-атаки, давайте розглянемо деякі з їх основних аспектів (таблиця 1.2).

Таблиця 1.2 – Порівняння аспектів в шилінг-атаках

Аспекти	Фальшиві облікові записи	Створення шуму	Специфічна поведінка
Опис властивостей	Зловмисники створюють фальшиві облікові записи для зміни поведінки РС.	Зловмисники можуть намагатися створити шум, змінюючи вагу деяких властивостей.	Зловмисники можуть намагатися вплинути на поведінку рекомендованих облікових записів.
Перевага властивостей	Збільшення ваги певних облікових записів, які відповідають інтересам зловмисників, в алгоритмах рекомендацій.	Зменшення впливу певного контенту на рішення системи шляхом збільшення невизначеності та непередбачуваності результатів.	Збільшення впливу певного контенту на рішення системи шляхом спеціальної поведінки користувачів, що відповідає інтересам зловмисників.

Фальшиві облікові записи: зловмисники створюють фальшиві облікові записи, щоб змінити поведінку рекомендаційних систем. Ці облікові записи можуть бути створені за допомогою різноманітних методів, включаючи автоматизоване створення облікових записів або купівлю готових облікових записів на ринку.

Псевдо-відгуки: зловмисники можуть створювати псевдо-відгуки або коментарі, щоб впливати на рішення рекомендаційної системи. Ці відгуки можуть бути написані за допомогою ботів або людьми, які отримують винагороду за написання відгуку.

Створення шуму: зловмисники можуть намагатися створити шум, щоб змінити результати рекомендаційної системи. Наприклад, зловмисники можуть намагатися змінити вагу деяких властивостей, щоб змінити результати рекомендаційної системи.

Специфічна поведінка: зловмисники можуть намагатися вплинути на поведінку рекомендованих облікових записів, щоб змінити результати рекомендацій.

Наприклад, зловмисники можуть спеціально відвідувати певні сторінки або певний контент, щоб збільшити його вагу в алгоритмах рекомендацій [11].

Для того, щоб протистояти шилінг-атакам, рекомендаційні системи використовують різні методи. Наприклад, вони можуть використовувати алгоритми машинного навчання, які виявляють фальшиві облікові записи або відфільтровують неправдиві відгуки та коментарі.

Крім того, можуть застосовувати різні методи аналізу трафіку та виявлення аномальної активності, щоб виявити спроби Шилінг-атаки.

Таким чином, відкритість рекомендаційних систем до введених користувачами даних надає змогу для використання шилінг-атак, які полягають у створенні фальшивих профілів користувачів.

Це стає можливим, оскільки рекомендаційна система отримує достатній зворотній зв'язок від користувачів, що використовується для навчання моделей та надання рекомендацій.

Загалом, аналіз властивостей шилінг-атаки є важливим кроком для розуміння цього типу атак та розробки ефективних методів протидії. Аналіз властивостей шилінг-атаки дозволяє розуміти її потенційні наслідки та негативний вплив на рекомендаційні системи.

Це сприяє розробці ефективних методів виявлення та захисту від таких атак, забезпечуючи надійність та довіру до систем рекомендацій.

На рисунку 1.4 зображені типи шилінг-атак.

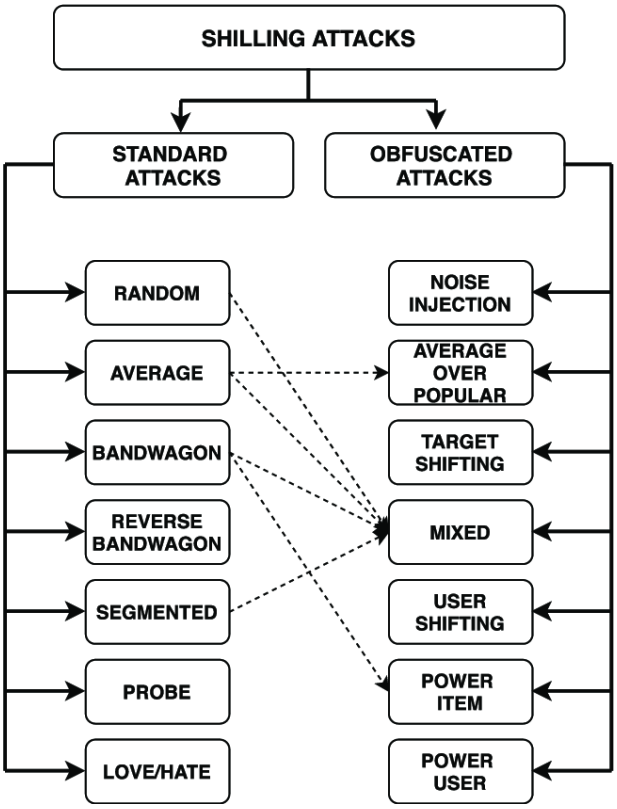

Рисунок 1.4 – Типи шилінг-атак

Random Attack, також відома як атака RandomBot, є найпростішою формою шилінг-атак. У цій моделі предмети, які були оцінені профілем атаки, вибираються випадковим чином, за винятком цільового предмета. Оцінки для цих предметів знаходяться навколо загального середнього значення системи. Цільовий предмет отримує максимальну або мінімальну оцінку в залежності від того, чи це атака push або nuke. Деякі атаки призначені для порушення довіри до системи рекомендацій, відомі як випадковий вандалізм. Найпростіша атака є найменш ефективною. Метою випадкової атаки зазвичай є більш ефективне порушення роботи системи рекомендацій, ніж просування цільового предмета. Легкість виконання випадкових атак полягає в їх низькій потребі в знаннях. Все, що потрібно атакеру, це загальне середнє значення системи, яке може бути легко

емпірично обчислене. Оскільки ця атака є найпростішою, вона не є дуже ефективною [12].

Атака з середнім рейтингом подібна до випадкової атаки за процесом вибору елементів. Випадково вибрані елементи оцінюються на основі розподілу рейтингів окремих елементів. Кожному елементу-заповнювачу присвоюється середній рейтинг для цього елемента. Ця атака є можливою лише у випадку, якщо нападник має великі знання про набір даних, на якому побудована система рекомендацій. Ефективність цієї моделі пропорційна знанням нападника. Хоча єдиною відмінністю між випадковою атакою та атакою з середнім рейтингом є рейтинги-заповнювачі, ефективність атаки з середнім рейтингом значно краща.

Атака «Приєднуйся до вагона» є типом атаки, при якій профілі, створені нападниками, заповнені популярними елементами з високими рейтингами. Профілі атаки природно ближчі до великої кількості користувачів. Цільовому елементу присвоюється найвищий рейтинг. Цю атаку можна поділити на дві підкатегорії: «Приєднуйся до вагона випадкової атаки» та «Приєднуйся до вагона з середнім рейтингом», залежно від схеми рейтингування для елементів-заповнювачів. «Приєднуйся до вагона» також входить до категорії атак з низьким рівнем знань, оскільки нападник потребує лише загальнодоступних даних.

Атака Reverse Bandwagon – це точний зворотний варіант атаки Bandwagon. Ця атака використовується для знищення цільового продукту, надавши низькі рейтинги товарам з високими негативними відгуками і найнижчий рейтинг цільовому продукту. Це також атака з низьким рівнем знань, так само як і атака Bandwagon. Хоча вона дуже схожа на атаку Bandwagon, ефективність атаки Reverse Bandwagon трохи краща.

Атака по сегментах націлена на конкретну групу користувачів, які ймовірно придбають цільовий товар у електронному комерційному середовищі. Сегментні атаки зазвичай застосовуються у КФ на основі товарів. Оцінювані товари та їх рейтинги ґрунтуються на знаннях нападника

про сегмент. Основна перевага цього методу полягає в його здатності досягати потенційних клієнтів. Наприклад, якщо цільовий товар - книга у жанрі наукової фантастики, то вибрані товари також будуть з того ж жанру. Такий вибір збільшує шанси того, що цільова книга потрапить до більшої кількості шанувальників наукової фантастики. Оскільки атака застосовується тільки в одному сегменті системи, вплив високий.

Атака Probe не є універсальною і не може бути застосована до всіх систем. Деякі рекомендаційні системи передбачають прогнозований рейтинг для кожного з елементів. Атакуючий використовує цей детальний рейтинг для оцінки елементів, дозволяючи йому бути подібним до інших користувачів. Атакуючий дає відповідні рейтинги деяким вихідним елементам. Потім, коли система пропонує більше елементів, атакуючий формує список оцінених елементів на основі цих елементів. Ця схема забезпечує те, що атакуючі профілі залишаються близькими до сусідніх. Вона також дозволяє атакуючому дізнатися більше про систему.

Атака Love/Hate – це високоефективна атака нюк, де атакуючий випадковим чином вибирає вхідні елементи та надає їм найвищі рейтинги, а цільовому елементу дає найменший рейтинг. Незважаючи на простоту цієї моделі, її ефективність дивовижно висока. Хоча вона була створена переважно для атак нюк, її також можна використовувати для атак штурм, змінюючи рейтинги. Атака штурм не є такою ж ефективною, як нюк атака. Таблиця 1. детально узагальнює різниці в різних моделях атак [13].

Атака з введенням шуму додає до кожного рейтингу для деякої частини профілів, до яких вона застосовується, випадкове число, розподілене за Гауссом, помножене на певну константу. Ступінь затуманювання залежить від тієї константи, яку множать на випадкове число. Це можна успішно застосовувати для всіх стандартних методів атак для затуманювання їхнього підпису. Оскільки рейтингова схема стає залежною від введеного шуму, можна помітити невеликий, але помітний зниження ефективності атаки.

Зміщення користувача є тактикою затуманювання, при якій змінюють рейтинги частини оцінених предметів для кожного впровадженого профілю. Рейтинги цієї частини предметів можуть бути збільшені або зменшені, щоб зменшити схожість між атакуючими профілями. Для різних груп впроваджених профілів можна модифікувати рейтинги різних частин оцінених предметів.

Target Shifting – це зсув рейтингу цільового елементу на один рівень нижче максимально можливого в push-атаках. У нюк-атаках рейтинг цілевого елементу зсувається на один рівень вище найменш можливого рейтингу. Ця стратегія особливо корисна для ухилення від методів виявлення, які покарають користувачів, які встановлюють екстремальний рейтинг для елементів. Якщо цільовий елемент вже популярний, його важче просунути, використовуючи зсув цілей. У таких випадках слід використовувати інші методи затемнення.

Average Over Popular – це техніка, яку використовують для затемнення атак типу Average. Тут заповнювачі вибираються з топ-X% найбільш популярних елементів з рівною ймовірністю. Цей метод гораздо ефективніший, ніж випадковий вибір з усієї колекції елементів. Вибір X впливає на виявленість атаки.

Mixed Attack – це використання в рівних пропорціях випадкових, середніх, bandwagon та сегментованих атак одночасно. Техніка виявлення повинна мати здатність виявляти всі стандартні атаки, щоб бути успішною. Різні методи атак використовуються для того, щоб просунути/знищити той самий цільовий елемент. Це допомагає ухилятися від кількох методів виявлення.

Атака Power Item використовує сильні елементи, які вибираються на основі трьох методів. Сильні елементи визначаються як набір елементів, які можуть впливати на найбільшу групу елементів. Ці елементи ефективно змінюють рекомендації, які надаються іншим користувачам. У PIA-AS, сильні елементи вибираються з верхніх N елементів з найвищою загальною

схожістю. Така схожість можлива лише в тому випадку, якщо значна кількість користувачів оцінила однакові два елементи. У PIA-ID критерієм вибору сильних елементів є ступінь входження в граф, а схожість кожної пари елементів розраховується з використанням вагового значення, і вибирається топ-N для кожного елементу. У PIA-NR вибираються елементи з найбільшою кількістю користувачів, які оцінили їх [14].

Power User Attack, подібно до PIA, вибирає набір користувачів, які мають максимальний вплив на найширшу групу користувачів. У PUA-AS, верхніх X користувачів з найвищою загальною схожістю обираються як сильні користувачі. У PUA-ID обираються користувачі, які беруть участь в найбільшій кількості околиць на основі концепції входження в граф. У PUA-NR, сильні користувачі – це користувачі з найбільшою кількістю оцінок у своєму профілі.

SAShA – це стратегія атак, яка використовує семантичні функції, отримані з графа знань, щоб поліпшити роботу стандартних моделей атак на фільтрацію на основі співробітництва. Граф знань – це структурований репозиторій фактичної, категоріальної та онтологічної інформації. Ця атака працює, обчислюючи семантичну подібність між особливостями графа знань цільового елемента та всіма іншими елементами в системі. Цю інформацію використовують для генерації найбільш ефективного набору елементів-заповнювачів.

У Чен та ін. описують метод, який використовує як кореляцію між елементами, так і популярність елементів, щоб генерувати зловмисних користувачів зі сильним потенціалом атаки та схожістю з реальними користувачами. У їхньому підході кожен зловмисний профіль користувача генерується окремо. Оцінені елементи профілю вибираються на основі матриці профілів реальних користувачів.

Як тільки була виявлена вразливість фільтрації на основі співробітництва до шилінг-атак, були створені різноманітні методи виявлення. Можна широко класифікувати ці методи на навчані та ненавчані

методи виявлення. В літературі існує маса ознак виявлення, які керують цими методами.

1.3 Дослідження методів виявлення шилінг-атак

Шилінг-атаки – це спроби штучного збільшення рейтингу продукту, послуги, контенту або користувача в мережі Інтернет. Ці атаки часто використовуються в онлайн-магазинах, соціальних мережах, рейтингових системах та інших веб-платформах з метою збільшення популярності та покращення рейтингу.

Для виявлення цих атак можна використовувати різні методики. Деякі з найпоширеніших методів виявлення включають наступне:

Використання статистичних методів: ці методи використовують статистичні моделі, щоб виявляти аномальні патерни у поведінці користувачів та їх взаємодії з продуктом або платформою. Наприклад, можна аналізувати рівень активності користувачів, їх інтереси, та інші фактори, щоб виявити незвичайні та неадекватні патерни поведінки.

Використання машинного навчання: ці методи використовують алгоритми машинного навчання для виявлення шаблонів та аномалій у поведінці користувачів. Наприклад, можна використовувати алгоритми класифікації та кластеризації для виявлення підозрілих аккаунтів або груп користувачів.

Використання графових методів: ці методи використовують топологію графу, що відображає взаємодії між користувачами та продуктом, щоб виявляти шаблони та аномалії. Наприклад, можна використовувати алгоритми виявлення спільнот та аналізу центральності графу для виявлення підозрілих користувачів або груп користувачів, які мають надмірну кількість взаємодій з продуктом.

Використання методів аналізу контенту: ці методи аналізують зміст контенту, що створюється користувачами, щоб виявити підозрілі патерни.

Наприклад, можна використовувати методи аналізу тональності та використання певних слів або фраз для виявлення підозрілих відгуків або коментарів.

Використання методів аналізу взаємодії: ці методи вивчають спосіб, як користувачі взаємодіють з продуктом або платформою, щоб виявити незвичайні патерни та аномалії. Наприклад, можна вивчити часову лінію взаємодії користувача з продуктом або аналізувати патерни дій користувачів, щоб виявити підозрілі взаємодії.

Використання соціальної мережевої аналітики: цей метод вивчає соціальну мережу користувачів, щоб виявити незвичайні патерни взаємодії між ними.

Використання блокчейн технологій: цей метод може бути використаний для створення системи, яка забезпечує надійну інформацію про взаємодію користувачів з продуктом або контентом. Блокчейн може зберігати історію взаємодії користувачів з продуктом та використовувати її для виявлення підозрілих патернів [15].

Описані методи можуть бути використані окремо або в поєднанні між собою для виявлення Шилінг-атаки. Кожен з цих методів може бути використаний для виявлення Шилінг-атаки залежно від контексту та характеру платформи. Результати аналізу можуть бути використані для виявлення підозрілих аккаунтів та взаємодій, які можуть бути підозрілі на шилінг-атаки (таблиця 1.3 та 1.4).

Таблиця 1.3 – Порівняння методів виявлення шилінг-атак

Метод	Методи аналізу взаємодії	Соціально-мережева аналітика	Блокчейн технології
1	2	3	4
Використовувані техніки	Методи аналізу взаємодії	Аналіз графу взаємодії між користувачами	Використання децентралізованих блокчейн мереж

Продовження таблиці 1.3

1	2	3	4
Показники	Незвичайні патерни та аномалії взаємодії користувачів з продуктом	Групи користувачів з надмірною взаємодією з продуктом або певним контентом	Безпека та прозорість даних, протидія шахрайству, захист від зловживань
Недоліки	Не може виявити підозрілі дії користувачів, що не залежать від продукту	Не може виявити дії користувачів, що не залежать від соціальної мережі	Можуть бути вкрадені, якщо більшість вузлів мережі підконтрольовані одній організації або особі. У такому випадку, може бути здійснено шилінг атака, яка може зламати систему.

Таблиця 1.4 – Порівняння методів виявлення шилінг-атак

Метод	Статистичні методи	Машинне навчання	Графові методи
Використовувані техніки	Статистичні моделі	Алгоритми класифікації та кластеризації	Алгоритми виявлення спільнот та аналізу центральності графу
Показники	Рівень активності користувачів, інтереси, незвичайні та неадекватні патерни поведінки	Шаблони та аномалії у поведінці користувачів, підозрілі аккаунти або групи користувачів	Підозрілі користувачі або групи користувачів з надмірною кількістю взаємодій з продуктом
Недоліки	Помилкові результати при використанні нерепрезентативних даних	Вимагає значної кількості даних для навчання моделі	Складно побудувати граф для продукту зі значною кількістю користувачів

Атрибути, які відрізняють підроблені профілі від справжніх, вважаються атрибутами виявлення. Атрибути виявлення, які призначені для роботи незалежно від типу моделі атаки, відомі як загальні атрибути.

Ці атрибути не адаптовані під конкретні моделі атак і належать до загальних атрибутів. Ефективність цих атрибутів залежить від застосованих моделей атак.

Rating Deviation from Mean Agreement (RDMA) – це міра відхилення рейтингів користувача на наборі цільових елементів порівняно з іншими користувачами, поєднана з оберненою частотою рейтингів цих елементів.

Weighted Deviation from Mean Agreement (WDMA) ґрунтується на атрибуті RDMA. Відмінність полягає в тому, що важливість відхилень оцінок від середньої збільшується для рідкісних предметів. Експериментально було встановлено, що WDMA дає більший приріст інформації.

Weighted Degree of Agreement (WDA) визначає кумулятивну різницю між рейтингом користувача для певного товару та середнім рейтингом цього товару, поділену на кількість рейтингів для даного товару. WDA в емпіричному відношенні збігається з чисельником RDMA.

Показник Length Variance (LengthVar) вимірює відмінності у довжині профілю користувача від середньої довжини профілю. Тут довжина означає кількість оцінених елементів в конкретному профілі користувача. Деякі атакуючі профілі можуть містити надмірну кількість оцінених елементів, значно відрізняючись від середньої довжини профілю [16].

Проблема використання лише загальних атрибутів полягає в тому, що вони іноді не можуть відрізнити зловмисні профілі від аутентичних користувачів, особливо коли аутентичний користувач виявляє незвичну поведінку. Для подолання цих недоліків було розроблено атрибути, специфічні для певних типів атак. Ці атрибути виявляють розділи в профілях користувачів, щоб їхня поведінка виявляла схожість з одним конкретним типом атак. Атрибут середньої дисперсії (MeanVar)

використовується для виявлення середніх атак. Він розділяє атакуючі профілі на три частини: елементи з екстремальними оцінками (цільові елементи), всі інші оцінені елементи в профілях (заповнювачі елементи) та непрооцінені елементи. Цей атрибут працює, обчислюючи середньо-варіаційне відхилення між всіма заповнювачами елементів та загальним середнім значенням. Низька дисперсія свідчить про можливість середньої атаки [17].

Різниця середніх між вставними елементами та цільовою моделлю (FMTD) націлена на сегментовану модель атаки. Цей атрибут базується на різниці між рейтингами елементів у цільовому розділі та елементів у вставному розділі.

Filler Average Correlation (FAC) фокусується на виявленні випадкових атак. При випадковій атакі оцінки, які надаються позначаються випадково. Ця атрибут обчислює кореляцію між оцінками в профілі та середніми оцінками предметів. Для випадкових атак очікується низький рівень кореляції.

Атрибут Filler Mean Difference (FMD) використовує той факт, що середні оцінки елементів filler-групи випадкової атаки схожі на загальний середній рівень системи. Якщо середні оцінки схожі, то користувацький профіль може бути потенційним профілем випадкової атаки.

Алгоритми виявлення можуть бути загалом класифіковані на дві категорії: навчання з вчителем (Supervised detection methods) та навчання без вчителя (Unsupervised detection methods). При навчанні з вчителем дані повинні бути позначені мітками під час тренування, тоді як при навчанні без вчителя цього не потрібно. У наборах даних систем рекомендацій наявність позначеного зразка з обмеженою.

Проблему атак з підробленими профілями розглянуто як задачу класифікації в роботі [24], використовуючи RDMA та DegSim як метрики ознак для виявлення зловживань. Метод був розроблений для виявлення випадкових та приєднаних атак. Пізніше, Burke et al. додали ще дві загальні

метрики, а саме WDMA та WDA, щоб покращити продуктивність класифікатора. SVM, kNN та C4.5 були найбільш поширеними класифікаторами для виявлення фальшивих профілів. Проблема використання загальних атрибутів полягала в тому, що багато аутентичних користувачів з екстремальними поведінковими властивостями помилково класифікувалися як підроблені профілі. Щоб подолати цю проблему та покращити точність класифікації, були сформульовані атак-специфічні атрибути. Для середніх, випадкових, сегментованих та приєднаних атак були створені різні атак-специфічні атрибути [18].

У роботі [25] використовували три стратегії для підвищення точності виявлення у навчаному підході: подібність до обернено-інженерних атак, концентрація на цілях та виявлення аномалій в рейтингах. Ця методика виявлення ефективна завдяки додатковій стійкості системи, але дуже залежить від вибору класифікатора. У їхньому дослідженні показано, що поєднання різних атрибутів покращує роботу класифікатора, особливо у випадку методу опорних векторів, і значно зменшує вплив найбільш потужних моделей атак. Використовувані в методі атрибути – RDMA, WDMA, DegSim, LengthVar, MeanVar, FMD, FAC та FMTD. Для покращення точності виявлення було введено використання мета-навчання.

У роботі було запропоновано використання мета-навчання для покращення точності виявлення атак. Цей алгоритм можна розглядати як двоетапний процес, де на першому етапі проводиться навчання на атакуючих профілях та наявних рейтингах. Другий етап полягає у поєднанні вихідних даних на першому етапі з мета-рівнем для кінцевого виявлення атак. Цей алгоритм мав вищу точність, ніж попередні методи. Різноманітність класифікаторів зменшує кореляцію помилкових класифікацій, позитивно впливаючи на мета-рівень передбачення. Вони перевірили свій підхід порівняно зі звичайним SVM та голосуванням SVM та експериментально довели його більшу ефективність. Атрибути, які

використовуються в їхньому методі, це WDMA, RDMA, WDA, LengthVar, DegSim, MeanVar, FMD та FAC.

SVM-TIA мав навчання з вчителем, навчання без вчителя та напівнаглядний підхід виявлення атак. Недоліком використання навчання з вчителем було те, що потрібна збалансована даних, тобто повинна бути рівна кількість автентичних профілів та профілів атак. Точність підходу з навчанням з вчителем була нижчою, ніж у підходу без нагляду, який включав кластеризацію та статистичні методи. Це двофазний процес, де перша фаза полягає у отриманні грубих результатів виявлення шляхом зменшення невідповідності класів. У другій фазі потенційні профілі атак аналізуються для виявлення цільових профілів. У цьому методі використовуються специфічні для моделі атрибути, такі як FMTD, MeanVar, FAC та FMD.

Як вже зазначалося раніше, незбалансованість доступних даних спотворювала результати класифікаторів навчання з учителем. В був використаний AdaBoost, щоб зменшити вплив незбалансованості. Автори спочатку спрощували задачу важкої класифікації, використовуючи добре розроблені ознаки для профілів користувачів. Цього досягнуто шляхом застосування ваг до різних спостережень для підкреслення погано модельованих зразків. Цей процес повторюється, щоб зміцнити корекцію помилкової класифікації. Використовуються ознаки RDMA, WDMA, WDA, LengthVar, MeanVar, FMTD та FAC. Крім того, вони також використовують ознаки, які виявляють розмір заповнювача з малопопулярними елементами.

У роботі [26] використали метод енсемблювання для виявлення з функцій, отриманих з рейтингів, популярності товарів та графів користувачів. Вилучення функцій здійснювалося за допомогою методу стекованих автокодерів з підтримкою зведення та головних компонент. Це дозволяє автоматично вилучати функції користувачів з різними рівнями пошкодження. Застосовувалася триетапний процес, що включав попередню обробку даних, вилучення функцій та виявлення за допомогою слабких

класифікаторів. Новизна товарів – ступінь відмінності між різними товарами – також використовувалася як функція.

Початковий метод ненавченого виявлення, запропонований [27], використовував аналіз головних компонент для проблеми виявлення профілю. Чотири фактори сприяли використанню PCA для цієї проблеми: спам-користувачі мають високу кореляцію, низьке відхилення від середньої значення рейтингу, висока схожість з великою кількістю користувачів та припущення, що спам-користувачі працюють разом. Усі профілі користувачів у системі рекомендацій проектувалися на гіперплощину, що утворена з матриці користувач-предмет. Профілі користувачів, які були скластеризовані ближче до початку гіперплощини, були профілями атаки. Рідкість матриці користувач-предмет робить ці передбачення менш надійними. RDMA та WDMA також використовуються як атрибути виявлення.

У роботі [28] розробили загальний атрибут, який допомагає виявляти профілі атак у надзвичайному режимі. Їхній підхід вирішує проблему виявлення профілів атак як проблему виявлення аномальних структур. Метрика, що використовується, є варіацією метрики Hv-score, яка спочатку використовувалася при аналізі генних даних для виявлення бікластерів. Цей алгоритм, названий UnRAP, здатний виявляти як стандартні, так і затуманені атаки. Їхній підхід має кращі шанси виявляти нові стратегії атак, які можуть уникнути виявлення з використанням навчання з учителем.

За припущенням, що атакуючі профілі менші в кількості та мають високу схожість, у застосували метод кластеризації k-середніх на основі атрибутів. Користувачі були розділені на дві кластери, менший з яких було визначено як атакуючі профілі. Цей метод продемонстрував вищу точність та меншу помилкову класифікацію справжніх користувачів. Незалежно від використаної стратегії атаки, ця робота стверджує, що має менше помилкової класифікації справжніх користувачів, ніж попередні методи. Серед використовуваних атрибутів – RDMA, WDMA, WDA, LengthVar, a

також метрика Hv-score, що використовувалась в . Chung et al. застосували алгоритм Бета-розподілу для виявлення атак. Цей метод виявляв якомога більше атак, не позбавляючи справжніх користувачів. Більшість проблем, пов'язаних з цим методом, успадковані від самого Бета-розподілу. Переваги використання цього методу – низька швидкість помилок та висока швидкість виявлення. Цей метод стверджує, що працює з розрідженими даними та незбалансованим співвідношенням атакуючих та нормальних профілів. Цей підхід виявляє високу продуктивність навіть з невеликим розміром атаки та має низький рівень помилкових спрацювань.

Ще один підхід, який базується на подібності профілів атаки, використовував k-середніх кластеризацію, щоб перемістити фальшиві профілі до листкових вузлів бінарного дерева. За допомогою матриці користувач-товар та оптимального числа сусідів N, рекурсивно застосовується k-середніх кластеризація для кластеризації користувачів на дві відмінні групи. Індексні центри кластерів та внутрішня кореляція бінарного дерева використовуються для виявлення профілю атаки. Цей підхід має особливо високу успішність у моделях середнього, сегментаційного та бандвагонного типів атак. У роботі [29] розробили алгоритм, який фокусується на аналізі цільових користувачів та товарів. Це двохфазний метод. Спочатку до набору даних застосовується метод кластеризації на основі щеки на основі деяких вибраних ознак для виявлення зловмисних користувачів. DBSCAN використовується для визначення підозрілих користувачів на основі їх ознак. Друга фаза допомагає детальніше дослідити користувачів з першої фази, виявляючи підозрілі товари на основі адаптивного навчання структури на вибраних ознаках.

У роботі [30] розробили метод кластеризації на основі прихованої моделі Маркова (HMM) та ієрархічної кластеризації для виявлення профілів-шахраїв [19]. Поведінка користувачів, що стосується рейтингів, моделюється за допомогою HMM. На основі послідовності вподобань

користувачів та модельованої поведінки їх рейтингів, обчислюється ступінь підозрілості кожного користувача. Потім використовується метод ієрархічної кластеризації, щоб згрупувати цих користувачів на основі їх ступеня підозрілості на подвійні кластери: для чесних користувачів та атакуючих користувачів. Автори також застосували свій метод на вибірці даних з Amazon, щоб продемонструвати його ефективність. У роботі [31] запропонували метод для поліпшення підходу PCA у виявленні профілів-шахраїв. Спочатку PCA використовується для розділення профілів на дві категорії: позитивні мітки для виявлених і негативні мітки для всіх інших користувачів. Потім використовуються ознаки виявлення – RDMA, WDMA, WDA та LenVar – як складність даних для обчислення CCMeasure набору даних. CCMeasure – це кількісна оцінка класифікаційної складності, що показує, наскільки складно класифікувати набір даних. Якщо ця міра висока, це свідчить про те, що значна кількість чесних користувачів помилково класифікували, тому мітки перевертають, щоб зменшити складність даних.

Після обговорення технік виявлення атак, також досліджуються інші ризики приватності, пов'язані з методами виявлення атак. Луо та Лянг [32] обговорюють вплив внутрішньої атаки на виявлення шахрайських атак у системах рекомендацій. Вони розглядають можливий сценарій, де нападник виступає як екзаменатор, якого утримують від окремих рейтингових профілів захищеними обчисленнями. Їхня модель атаки може визначити цільовий рейтинговий профіль з малою кількістю попереднього знання та виходу захищених обчислень. Така внутрішня атака становить серйозну загрозу приватності користувачів.

Паралельно з роботами, які фокусуються на виявленні атаки, існує лінія дослідження, спрямована на створення стійких алгоритмів, які стійкі до атак. Ці алгоритми не мають механізму для знаходження та видалення фальшивих профілів, але можуть знизити ефективність атаки. Ми коротко обговоримо деякі з останніх стійких алгоритмів у цьому підрозділі.

У роботі [33] поєднали алгоритм м'якого СК з методом схильності користувачів, щоб покращити стійкість системи рекомендацій та виявляти шилінг-атаки. Вони використовують Байєсівське СК, алгоритм м'якого СК, який дозволяє змішане членство рядків та стовпців, що дуже підходить для реальних даних. Ця модель поєднує RDMA з м'яким СК, щоб зменшити вплив шилінг-атак. Усі профілі атаки кластеризуються в один кластер, обмежуючи вплив атаки серед профілів атаки.

Turk та Bilge розробили стійкий алгоритм багатокритеріальної КФ. Багатокритеріальна CF має кілька категорій, в яких користувач може оцінити кожен елемент. MCCF допомагає краще розуміти уподобання та неподобання клієнта. Стійкість їх методу досягається шляхом виключення підозрілих оцінок на основі ступеня невизначеності. Користувачі також категоризуються на різні групи за схожістю уподобань, щоб обмежити автентичних користувачів від змішування з профілями атак. У роботі [34] інтегрували шкалювання ентропії в процес КФ для зменшення впливу дуже позитивних та негативних користувачів. Вони також використовують мінімальний поріг, щоб ще більше ускладнити ентропію та запобігти випадковим атакам.

У роботі [35] розрахували значення надійності для кожного прогнозування користувача щодо товару. Якщо спостерігається незвичайна зміна значення надійності прогнозування товару, це свідчить про можливу атаку залучення підроблених відгуків. Вони використовують метод матричного розкладання, щоб нейтралізувати вплив такої атаки. Відхилення від таких підроблених прогнозів може бути уникнуто, щоб зменшити масштаб атаки та знейтралізувати наявність профілів залучення підроблених відгуків. Швидкодія цього методу знижується при зменшенні розміру атаки, але стверджується, що такий невеликий розмір атаки має знехтувальний вплив [20].

Поточні тенденції у дослідженні шилінг атак: показано кількість публікацій, які виходили щороку, пов'язаних із атаками. Ця фігура

представляє як однокритеріальні, так і багатокритеріальні системи оцінювання, з провідних конференцій та журналів, що охоплюють як навчання з вчителем, так і без вчителя. На початкових етапах дослідження атак фокус був спрямований на створення нових моделей атак для оцінки впливу різних атак на систему рекомендацій. Стандартні атаки були створені на початку 2000-х років, проте зі зростанням методів виявлення на цих початкових етапах дослідження більше фокусувалися на оманливих атаках.

Дослідження виявлення шилінг-атак дозволяє виявити нерегулярність у взаємодії користувачів з системою рекомендацій, такі як створення фальшивих профілів або масове накручування рейтингів. Шляхи виявлення включають аналіз статистичних показників, залучення алгоритмів машинного навчання та спеціально розроблених моделей для виявлення аномалій. Результати досліджень методів виявлення шилінг-атак в рекомендаційних системах допомагають покращити якість рекомендацій та забезпечити довіру користувачів до системи. Це важливий крок у забезпеченні справедливості, точності та надійності рекомендаційних систем. Однак, виявлення шилінг-атак залишається складним завданням через постійне змінювання та вдосконалення методів атак. Тому дослідники та розробники повинні продовжувати працювати над розробкою нових та ефективних методів виявлення шилінг-атак, щоб забезпечити стабільність та безпеку рекомендаційних систем у майбутньому.

1.4 Постановка задачі та дослідження

Актуальність дослідження шилінг-атак визначається тим, що існуючі методи виявлення шилінг-атак не враховують короткострокові зміни вподобань користувачів. Тому такі зміни можуть розглядатись як нові атаки. Для вирішення цієї проблеми необхідно врахувати темпоральну складову цих атак.

Розробка нових методів, які використовують адаптивне комбіноване навчання, може забезпечити більш ефективне виявлення шилінг-атак на основі обробки коротких вибірок даних.

Об’єктом дослідження кваліфікаційної роботи є процес побудови рекомендацій.

Предметом дослідження є методи виявлення шилінг-атак в рекомендаційних системах.

Метою даної роботи є розроблення ефективного алгоритму виявлення шилінг-атаки та визначення їх впливу на якість рекомендаційної системи.

2 МЕТОД ВИЯВЛЕННЯ ШИЛІНГ-АТАК З УРАХУВАННЯМ НЕЯВНОГО ЗВОРОТНОГО ЗВ’ЯЗКУ ВІД КОРИСТУВАЧІВ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ

2.1 Підхід до виявлення шилінг-атак на основі порівняння явного і неявного зворотного зв’язку

Зворотний зв'язок в РС може бути явним або неявним. Явний зворотний зв'язок відбувається тоді, коли користувач прямо вказує свої вподобання стосовно певних товарів або послуг, наприклад, залишає відгук про товар чи послугу, ставить оцінку, додає товар до кошика, купує його, тощо.

У контексті рекомендаційних систем, неявний зворотній зв'язок означає невиразне вираження відгуку користувача на якість товару або послуги. Наприклад, користувач може просто переглядати товари, додавати їх до кошика, зберігати в обране, але не залишати коментарів або оцінок. Такі дії можуть бути інтерпретовані як неявний зворотній зв'язок, який може вказувати на інтерес користувача до товару або послуги. Неявний зворотній зв'язок може бути корисним для рекомендаційних систем, оскільки він дозволяє враховувати велику кількість даних про поведінку користувачів, що не дається виразним відгуком або оцінкою. Наприклад, якщо багато користувачів додають товар до кошика, але не купують його, це може свідчити про те, що ціна товару занадто висока або що товар не відповідає їхнім очікуванням. Рекомендаційна система може врахувати такий неявний зворотній зв'язок і пропонувати користувачам інші товари, які можуть відповідати їхнім потребам краще.

Проте, неявний зворотній зв'язок може також бути використаний для шилінг атак. Наприклад, продавець може створити багато штучних облікових записів, які будуть додавати його товар до кошика або зберігати його в обраних, щоб збільшити його рейтинг [21]. Це може призвести до

неправильних рекомендацій користувачам, оскільки система буде вважати, що багато людей зацікавлені цим товаром.

Явний зворотній зв'язок в рекомендаційних системах описується в явному способі відгуку користувача, який включає безпосередній вибір товару та поставлення оцінки. Користувачі можуть висловлювати свої побажання щодо товарів та послуг, які вони б хотіли придбати, або оцінювати товари, які вони вже придбали, за допомогою рейтингів, відгуків та коментарів.

Шилінг-атаки можуть використовувати явний зворотній зв'язок для накручування рейтингів та відгуків на товари та послуги. Наприклад, відомі методи шилінг-атак включають створення псевдо-користувачів, які можуть створювати фальшиві відгуки та рейтинги для певних товарів або послуг.

Для виявлення шилінг-атак в рекомендаційних системах, можна порівнювати результати явного та неявного зворотного зв'язку користувачів. Якщо відгуки та рейтинги не відповідають дійсному вибору користувачів, можливо, що вони стали жертвами шилінг-атаки. Додатково можна використовувати методи аналізу часових залежностей, щоб виявити незвичайні зміни в звичайному користувацькому поведінці, що можуть свідчити про накрутку рейтингів.

Зміни відвідувачів системи рекомендацій щодо конкретного товару відображаються у процесах його купівлі та встановлення оцінок. Для опису порядку цих процесів запропоновані адаптовані правила двох типів: «Подія в наступному» та «Подія в майбутньому». Кожне з цих правил встановлює порядок в часі для пари фактів ρ_m та ρ_s, які відображають вибір (купівлю) продукту або встановлення його оцінки. Факт стає істинним, коли відбуваються певні події, такі як вибір певного об'єкта в певний час τ; вибір декількох екземплярів товару на певному підмножині покупок.

Кожне правило $\eta_{m,s}^{(j)}$ встановлює відносний порядок типу раніше-пізніше. Правило $\eta_{m,s}^{(j)}$ встановлює, що після факту ρ_m покупки товару i_j на

інтервалі $\Delta\tau_m$, факт ρ_s покупки товару i_j на інтервалі $\Delta\tau_s$ буде істинним. Такі правила можуть встановлювати відносини між інтервалами або точками в часі та між підмножинами фактів, впорядкованими в часі. Правило типу «Наступне» використовує оператор X, який пов'язує дві послідовні події вибору/оцінки. Коли це правило виконується, між фактами ρ_m та ρ_s не можуть існувати інші істинні факти. Правило типу «Майбутнє» використовує оператор F, який пов'язує дві несусідні події. Між фактами ρ_m та ρ_s у композиції правила F повинен бути щонайменше один проміжний факт. Узагальнена форма правила має наступний вигляд:

$$\eta_{m,s}^{(j)} = \rho_m(X \vee F)\rho_s. \qquad (2.1)$$

Правила P_R описують $\eta_{m,s}^{(j)}$ послідовність купівель (оцінок) об'єкта i_j за період T:

$$P_R = (\eta_{1,2}^{(j)}, \eta_{1,3}^{(j)}, \dots, \eta_{1,s}^{(j)}, \dots, \eta_{s+1,s}^{(j)}, \dots, \eta_{s-1,s}^{(j)} : \forall s \Delta\tau_s \in T). \qquad (2.2)$$

Вираз правила $\eta_{1,2}^{(j)}$ (2) містить оператор часу X, оскільки він зв'язує факти ρ_1 та ρ_2 покупок (або призначення оцінок) на двох прилеглих інтервалах $\Delta\tau_1$ та $\Delta\tau_2$. Залежність $\eta_{1,3}^{(j)}$ є прикладом правила з оператором часу F, який зв'язує факти ρ_1 та ρ_3.

Окрім переформування фактів, адаптація правил полягає в налаштуванні їх ваг, з урахуванням динаміки інтересів користувача. Вага $\upsilon_{m,s}^{(j)}$ правила $\eta_{m,s}^{(j)}$ встановлюється через нормалізовану різницю між кількістю покупок або середньою оцінкою рейтингу товару i_j на інтервалах $\Delta\tau_m$ та $\Delta\tau_s$.

Послідовність вибору товарів або встановлення їх оцінок представлена упорядкованим набором нормалізованих ваг P_υ правил :

$$\mathrm{P}_{\upsilon} = (\upsilon_{1,2}, \dots, \upsilon_{1,s}, \dots, \upsilon_{m,m+1}, \dots, \upsilon_{s,s-1} : \forall s \Delta\tau_s \in T). \qquad (2.3)$$

На основі послідовності ваг (3) для будь-якого інтервалу часу $\Delta\tau_s$ можливо оцінити $\upsilon_s^{(j)}$, зміну інтересу користувача до предмета i_j з часом. Ця оцінка поєднує зміну інтересів користувачів до вибраного об'єкта з першого інтервалу $\Delta\tau_1$ до поточного інтервалу $\Delta\tau_s$:

$$\upsilon_s^{(j)} = \frac{\sum_{m=1}^{s-1} \upsilon_{m,s}}{max_s(\sum_{m=1}^{s-1} \upsilon_{m,s})}. \qquad (2.4)$$

Приклад набору правил $\{\eta_{m,s}^{(j)}\}$, які використовуються для розрахунку оцінки $\upsilon_4^{(j)}$, зображено на рисунку 2.1.

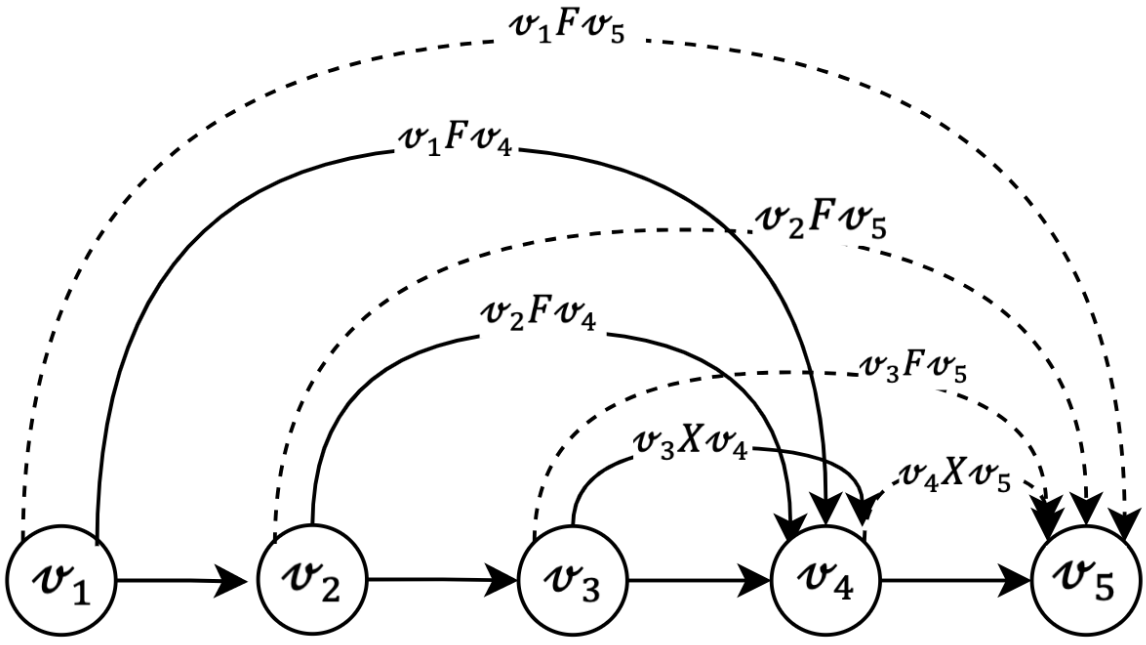

Рисунок 2.1 – Алгоритм реалізації

Цей приклад представляє послідовність фактів, упорядкованих у $\langle \rho_1, \rho_2, \rho_3, \rho_4, \rho_5 \rangle$. Такі факти описують процес вибору певного товару (або

встановлення його рейтингу) різними користувачами на постійній послідовності інтервалів.

$$T = \langle \Delta\tau_1, \Delta\tau_2, \Delta\tau_3, \Delta\tau_4, \Delta\tau_5 \rangle. \quad (2.5)$$

У прикладі для спрощення не вказано індекс j товару. Оцінка формується для поточного факту ρ_4. Цей факт пов'язаний з попередніми фактами за допомогою F-правил $\rho_1 F \rho_4$ та $\rho_2 F \rho_4$, а також X-правила $\rho_3 X \rho_4$. Ваги цих правил $\upsilon_{1,4}^{(j)}$, $\upsilon_{2,4}^{(j)}$ та $\upsilon_{3,4}^{(j)}$ використовуються для розрахунку оцінки $\upsilon_4^{(j)}$. Якщо на поточному інтервалі немає значень рейтингів, то використовуються попередні правила діапазону для побудови оцінки $\upsilon_s^{(j)}$. Наприклад, якщо на інтервалі $\Delta\tau_5$ для того, що показано на рис. 1 прикладу, немає нових значень рейтингів, то його середнє значення залишається таким самим, як на інтервалі $\Delta\tau_4$. Тоді ваги $\rho_1 F \rho_5$, $\rho_2 F \rho_5$ та $\rho_3 F \rho_5$ будуть дорівнювати відповідно $\upsilon_{1,4}^{(j)}$, $\upsilon_{2,4}^{(j)}$ та $\upsilon_{3,4}^{(j)}$. Вага правила $\rho_4 F \rho_5$ дорівнює нулю, оскільки середній рейтинг на інтервалі $\Delta\tau_5$ не змінився порівняно з діапазоном $\Delta\tau_4$ через відсутність нових рейтингів.

Таким чином, оцінка (4) на кожному поточному інтервалі $\Delta\tau_s$ показує, як змінилися уподобання користувачів до продукту i_j порівняно з усіма попередніми інтервалами часу.

Тоді модель $M^{(j)}$ процесу зміни уподобань користувача до предмета i_j є послідовністю оцінок $\upsilon_s^{(j)}$, впорядкованих за інтервалами $\Delta\tau_s$.

$$M^{(j)} = \langle \upsilon_2^{(j)}, \upsilon_3^{(j)}, \dots, \upsilon_s^{(j)} \rangle. \quad (2.6)$$

Ця модель описує зміну продажів або рейтингів для кожного інтервалу $\Delta\tau_s$, що дозволяє виявляти фальсифікацію рейтингів шляхом порівняння відповідних оцінок $\upsilon_s^{(j)}$ поетапно.

2.2 Метод виявлення шилінг-атак

Цей метод, при виявленні шилінг-атак, порівнює моделі процесів зміни вподобань користувачів (6), отриманих як результат неявної (продажі) та явної (рейтинги) зворотного зв'язку. Метод формує кількісну оцінку розбіжностей між цими процесами. Початкові дані методу: список продажів L; список рейтингів Q; період аналізу T; об'єкт можливої атаки i_j; підмножина користувачів – потенційних нападників $U = \{u_k\}$; рівень деталізації часу (година, день, тиждень, місяць), представлений довжиною інтервалу $\Delta\tau_s$. Початкові списки продажів та рейтингів містять наступні елементи: ідентифікатор користувача u_k; момент вибору/встановлення рейтингу τ_s; кількість продуктів i_j, проданих користувачем n_k; рейтинг p_k продукту i_j, встановлений користувачем n_k.

Метод включає наступні етапи.

Етап 1. Попередня обробка вхідного набору даних. На цьому етапі формуються набори даних для конструювання фактів продажу, а також фактів призначення рейтингів на інтервалах $\Delta\tau_s$. Результатом цього етапу є набори фактів покупок товарів $\{\eta_s^{j,item}\}$ та рейтингів $\{\eta_s^{j,\ rating}\}$. Ці факти містять інформацію про кількість куплених товарів $\eta_s^{(j)}$ та середній рейтинг цих товарів $r_s^{(j)}$ на інтервалах $\Delta\tau_s$ для користувачів u_k.

Етап 2. Конструювання відповідно до (2) наборів правил для вибору користувачів $P_R^{j,item}$ товарів та призначення рейтингів $P_R^{j,rating}$. На цьому етапі утворюються Наступні-правила та Правила-майбутнього у вигляді (1) з парами фактів $(\eta_m^{j,item}, \eta_s^{j,items})$ та $(\eta_m^{j,rating}, \eta_s^{j,rating})$.

Етап 3. Формування згідно з виразом (3) наборів вагових коефіцієнтів правил для покупок $P_v^{j,item}$ та $P_v^{j,rating}$ рейтингів.

Етап 3.1. Формування набору $P_{\vartheta}^{j,item}$ виконується шляхом нормалізації різниці у кількості покупок $n_s - n_m$ для всіх правил з набору $P_R^{j,item}$.

Етап 3.2. Формування набору $P_{\vartheta}^{j,rating}$ виконується шляхом нормалізації різниці у рейтингах для елементів набору $P_R^{j,rating}$.

Етап 4. Побудова моделей процесів зміни користувацьких уподобань $M^{j,invoice}$ для продажів та $M^{j,rating}$ для встановлення рейтингів.

Етап 5. Виявлення інтервалів можливого підроблення. На цьому етапі беруться до уваги якісні та кількісні різниці між відповідними елементами послідовностей $M^{j,invoice}$ та $M^{j,rating}$.

Етап 5.1. Формування набору кількісних розбіжностей $D^{(j)}$ між процесами покупок та рейтингами:

$$D^{(j)} = \{d_s^{(j)}\}, \tag{2.7}$$

$$d_s^{(j)} = \begin{cases} |\eta_s^{j,item}| + |\eta_s^{j,rating}| \; if \; (\eta_s^{j,item} \geq 0 \wedge \eta_s^{j,rating} < 0) \vee \quad (\eta_s^{j,item} \leq 0 \wedge \eta_s^{j,rating} > 0). \\ 0 \; otherwise \end{cases} \tag{2.8}$$

На цьому етапі, оцінки $\eta_s^{j,item}$ та $\eta_s^{j,rating}$ сумуються по модулю у випадку протилежно напрямлених змін попиту та рейтингу, оскільки така багатонапрямленість може свідчити про можливу атаку.

Крок 5.2. Формування множини ознак $B^{(j)}$ можливої атаки фальшивих користувачів на основі невідповідностей $D^{(j)}$.

Умова рівності рейтингів $\eta_s^{j,\ rating} = \eta_{s-1}^{j,\ rating}$, блокує формування ознаки атаки, коли рейтинг залишається незмінним на парі підряд інтервалів $\Delta\tau_{s-1}$ та $\Delta\tau_s$. Ця умова виконується у відсутності інформації про рейтинги на інтервалі $\Delta\tau_s$. Негативні значення $b_s^{(j)}$ вказують на можливу атаку з метою зниження рейтингу конкуруючих продуктів, тоді як позитивні значення вказують на атаку з метою збільшення рейтингу цільового товару.

Цей метод підходить для виявлення шилінг-атак, які відбуваються в рекомендаційних системах з використанням явних (рейтинги) та неявних (продажі) зворотних зв'язків. Він враховує короткострокові зміни вподобань користувачів та дозволяє виявляти атаки на основі обробки коротких вибірок даних. Цей метод може бути застосований для різних типів шилінг-атак, які мають вплив на рекомендаційну систему. Наприклад, це може бути створення фальшивих профілів користувачів, що впливають на рекомендації, або надмірне купування товарів з метою збільшення їх рейтингу.

3 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА ОТРИМАНИХ РЕЗУЛЬТАТІВ

3.1 Перевірка результатів

Метою дослідження було перевірити, наскільки ефективно метод виявлення підривних атак працює на відсортованому за часом наборі даних про оцінки та продажі, де не відомі абсолютні значення часу. Базуючись на використанні відносної шкали часу та F-правилах, припускали, що можна відрізнити інтервали за кількістю покупок товарів чи послуг. Дослідження довело, що використання опису подій за шкалою «раніше-пізніше» та їх атрибутів підтверджує це припущення.

Експеримент складався з двох етапів: на першому етапі було виявлено підривні атаки та проаналізовано способи їх приховування, на другому етапі ефективність запропонованого методу порівнювали з іншими методами, які також розбивали початковий набір даних на інтервали часу [22].

Під час експерименту було використано набір даних з інформацією про читання та оцінки декількох мільйонів книг. Записи про читання та оцінки впорядковані за часом, але в оригінальних даних немає абсолютних відміток часу. Факти η_m^j та η_s^j були сформовані на підмножинах вхідних даних з фіксованою кількістю записів.

Перша фаза спрямована на виявлення атак за довгими періодами, що представлені великою кількістю покупок. Під час формування фактів використовувалися підмножини з 100 000 послідовних елементів (читання, оцінки). Такі підмножини відповідають оригінальним інтервалам методу і позначаються через Δ_s. Результати ключових кроків методу для об’єкта i_{10000} (книги з ідентифікатором $id = 10000$) представлені в таблиці 3.1.

Таблиця 3.1 – Порівняння методів виявлення шилінг-атак

Шаг	Результат	Components of the result by subsets Δ_s								
		Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9	Δ_{10}
4. 1	$M^{10000,ite}$	0.19	0.19	-0.41	0.04	-0.63	-0.52	0.07	-1.00	-0.59
4. 2	$M^{10000,rat}$	-0.10	-0.10	-0.30	-0.30	-0.30	-0.03	-0.03	-0.03	1.00
5. 1	$D^{(10000)}$	0.28	0.28	0.00	0.33	0.00	0.00	0.11	0.00	1.59
5. 2	$B^{(10000)}$	-1	-1	0	0	0	0	-1	0	1

Результати кроків 4.1 та 4.2 в таблиці 3.1 містять опис процесу вибору об'єкта i_{10000} користувачами, а також процесу формування рейтингу для цього об'єкта. Наприклад, згідно з результатами кроку 4.1, для другого підрядка Δ_2, дійсним є лише одне правило $\rho_{1,2}^{(10000)} = \eta_1 X \eta_2$, що пов'язує його з попереднім підрядком. Значення ваги правила $W_2^{10000,item} \in M^{10000,item}$ становить 0,19 і є нормалізованою вагою цього правила, що показує збільшення продажів за Δ_2 порівняно з Δ_1. Показник $W_4^{10000,item} = -0.41$ відображає загальний спад продажів за Δ_4 порівняно з Δ_1, Δ_2 та Δ_3. Знак $a_2^{10000} \in A^{(10000)}$ має від'ємне значення і, отже, вказує на можливу атаку з метою зниження продажів. Цей знак a_{10}^{10000} з свого боку, вказує на можливу атаку з метою збільшення рейтингу. Показники динаміки рейтингу для Δ_5 та Δ_6 були сформовані з використанням правил для Δ_4, оскільки рейтинг для об'єкта i_{10000} не був представлений в вказаних підрядках.

Результати експерименту з вибраним об'єктом дозволяють проаналізувати підходи до маскування атак. Розкриття методу приховування можливої атаки здійснюється на основі порівняння показників атаки: $d_s^{(10000)}$ і $a_s^{(10000)}$, як показано на рисунку 3.1. На Δ_2 та Δ_3 спостерігається розбіжність між збільшенням продажів та одночасним зменшенням рейтингу порівняно з Δ_1. Оскільки величина розбіжності не

змінилася: $d_2^{(10000)} = d_3^{(10000)} = 0.28$, то ймовірна атака сталася в інтервалі Δ_2. Розбіжність менше 50% максимально можливого значення вказує на можливість приховання атаки, враховуючи середній рейтинг наявних елементів. Значення $a_2^{(10000)} = -1$ вказує на можливу атаку типу з метою зниження рейтингу. Також існує відхилення $d_5^{(10000)}$ на інтервалі Δ_5. Це відхилення не є ознакою атаки $\left(a_5^{(10000)} = 0\right)$ оскільки оцінки $W_4^{10000,item}$ і $W_5^{10000,item}$ збігаються. Збіг рейтингів означає, що користувачі не ставили оцінки в наступному інтервалі Δ_5.

Розбіжність $d_8^{(10000)}$ описує динаміку вибору користувача в інтервалах від Δ_2 до Δ_8 включно. Це відхилення свідчить про можливий фальсифікацію рейтингу на одному з цих інтервалів.

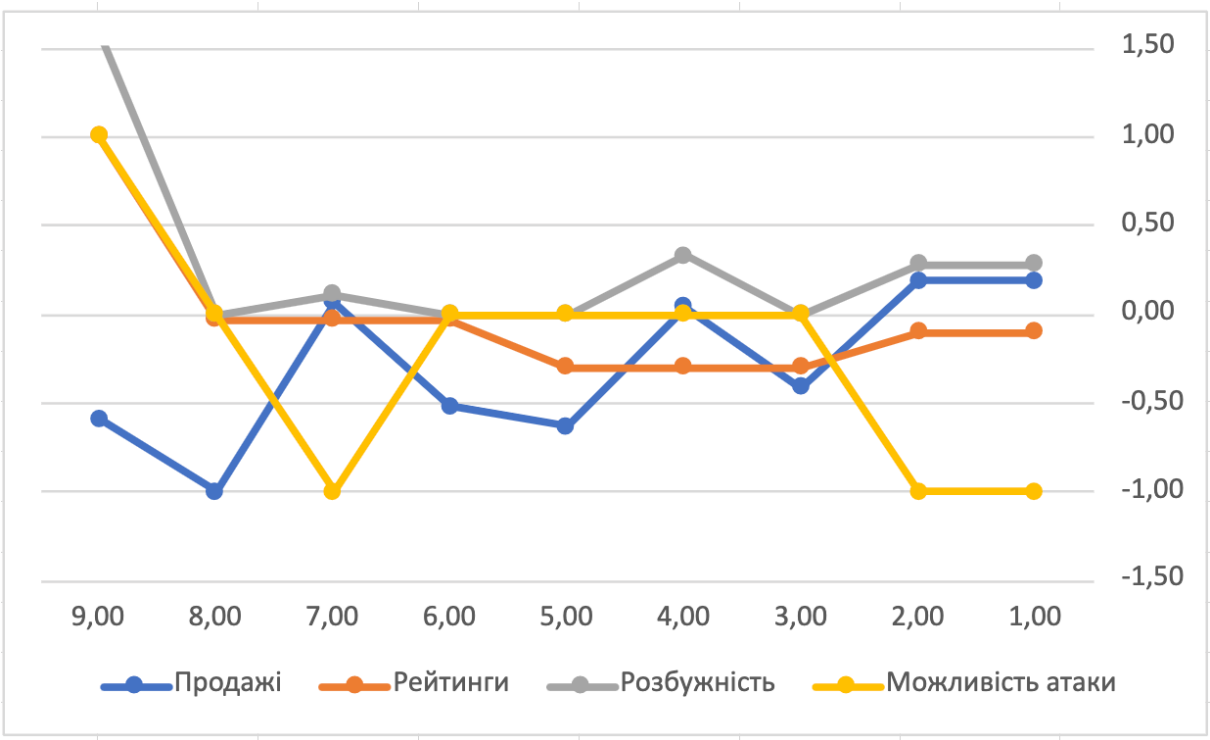

Рисунок 3.1 – Вразливості методів побудови РС

Проте, оскільки значення $d_8^{(10000)}$ враховує всі попередні відхилення та має менше значення, ніж $d_2^{(10000)}$ ймовірна атака сталася в інтервалі Δ_2. Розбіжність $d_2^{(10000)} = 1.59$ свідчить про можливу атаку на інтервалах від Δ_2 до Δ_{10}. Оскільки це значення значно вище, ніж $d_2^{(10000)}$d та $d_8^{(10000)}$ ймовірна атака сталася на інтервалі Δ_{10}. Значення $a_{10}^{(10000)} = 1$ свідчить про можливу атаку для збільшення рейтингу. Розбіжність $d_{10}^{(10000)}$, яка перевищує 50% максимально можливого значення, свідчить про можливе приховування атаки за допомогою популярних елементів. Була проведена перевірка на популярних властивостях (з великою кількістю покупок), щоб підтвердити це. Наприклад, популярний елемент i_{10} був вибраний понад 91 000 разів. Рейтинг цього об'єкта збільшується на множинах Δ_9 та Δ_{10}, що підтверджує гіпотезу про маскування атаки за допомогою рейтингу популярних елементів. Таким чином, цей метод дозволяє ідентифікувати способи маскування атаки в умовах неповних даних про рейтинг за допомогою даних, впорядкованих за часовою шкалою «раніше-пізніше» [23].

Друга фаза експерименту присвячена порівнянню ефективності запропонованого методу з методами. Ці методи розбивають інтервали за умови різкого змінювання значень рейтингу протягом обмеженого періоду. Для порівняння методів була використана оцінка точності.

Оцінка точності визначається як кількість виявлених атак до загальної кількості атак. Атаки були згенеровані для зниження та підвищення рейтингу трьох цільових товарів. У першому випадку рейтинг було встановлено рівним нулю, а в другому - рівним 5.

Ці атаки були включені до множини Q рейтингів. Раніше, згідно з алгоритмом на рисунку 2, проводилось розбиття на інтервали за допомогою оцінки змін в виборі користувачів для трьох книжок. Інтервал з максимальним значенням ваги, з урахуванням округлення, становив 9000 покупок.

Існує метод виявлення аномалій на основі динамічного розбиття для часових рядів та метод виявлення аномальних елементів на основі часових інтервалів. Розроблений метод, показав підвищену точність порівняно з методами ДР та ЧІ на початковій стадії фальсифікації рейтингу, коли формується до 10 атак для підвищення рейтингу, збільшивши точність від 8% до 23% та понад 30% відповідно. Збільшення точності на початковій стадії становило від 5% до 23% порівняно з методом ДР і понад 30% для методу ЧІ. Однак, при збільшенні кількості атак, методи ДР та ЧІ показують схожу або вищу точність. Такі характеристики визначають область застосування розробленого методу на початковій стадії дій атакуючого користувача.

3.2 Алгоритм реалізованого методу

Алгоритм, який реалізує цей метод, показаний на рисунку 3.2. Алгоритм включає наступні кроки.

Крок 1. Введення початкових даних. На цьому етапі необхідно ввести дані про продажі L, рейтинг Q, період T, суб'єкт i_j; користувачів U та рівні деталізації часу. Рівень деталізації часу залежить від формату відміток часу в журналах продажів та рейтингування.

Крок 2. Розбиття періоду T на інтервали часу $\Delta\tau_s$. На цьому етапі вибирається рівень деталізації часу, що забезпечує найвище значення ваг правил процесу продажів

$$\eta^{j,item} = \sum_s \eta_s^{j,item}. \quad (3.1)$$

для всіх інтервалів періоду T. Результатом цього кроку є набір інтервалів $\Delta\tau_s$ для періоду T.

Крок 3. Реалізація розробленого методу для виявлення атак із підробкою голосів.

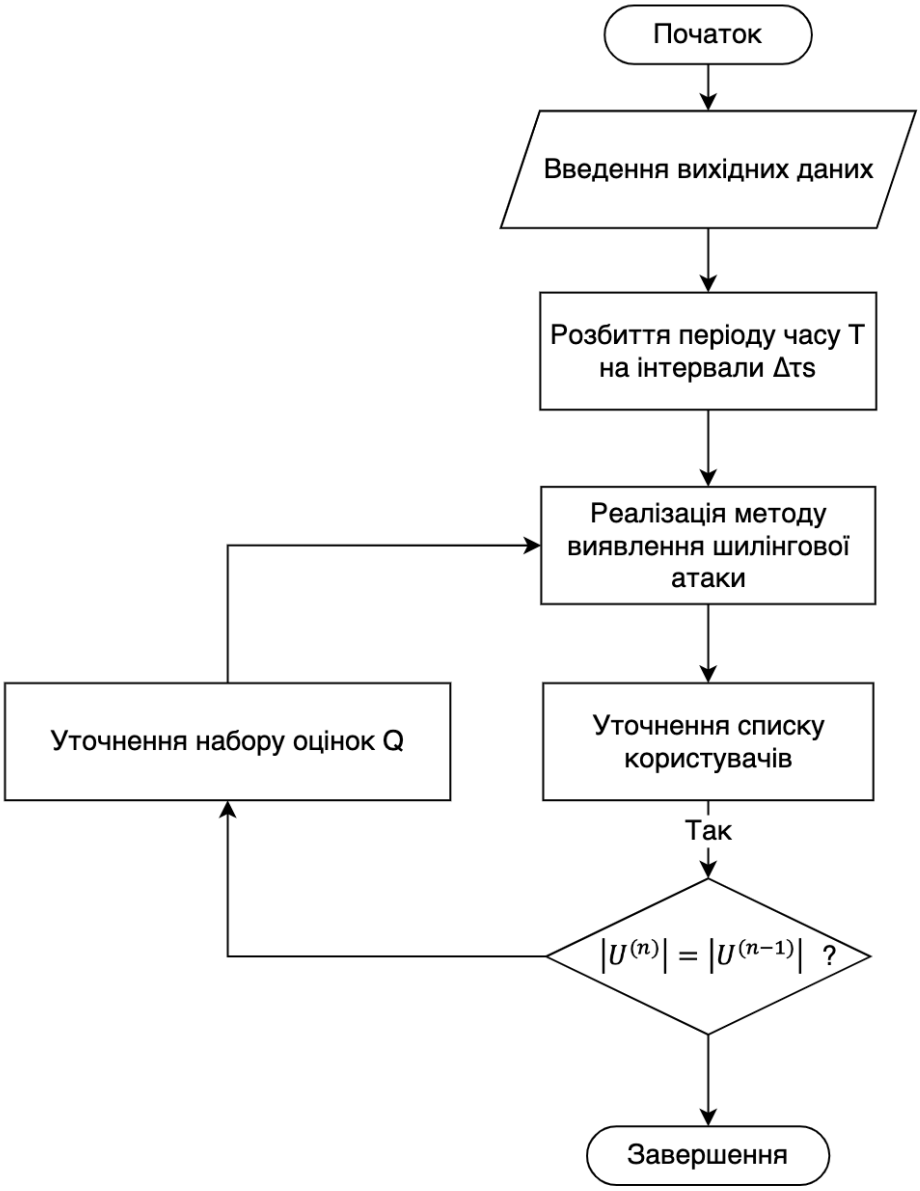

Рисунок 3.2 – Алгоритм реалізації

Крок 4. Уточнення списку користувачів. З множини U на n ітераціях видаляються користувачі $u_k^{(n)}$, які не встановили рейтинги на інтервалах $\Delta\tau_s$ з позначкою $b_s^{(j)} \neq 0$, оскільки ці користувачі не брали участь у спотворенні рейтингів. Результатом цього кроку є підмножина користувачів $U^{(n)}$, що містить потенційних атакувальників.

Крок 5. Перевірка умови завершення алгоритму $\left|U^{(n)}\right| = \left|U^{(n-1)}\right|$, згідно з якою кількість користувачів на поточній n ітерації не змінилася порівняно з ітерацією n-1. Якщо ця умова виконується, робота алгоритму завершується. В іншому випадку виконується крок 6.

Крок 6. Вдосконалення множини рейтингів Q. Рейтинги користувачів видаляються з цієї множини, оскільки ці користувачі $u_k^{(n)}$ не є атакувальниками. Далі виконується перехід до виконання методу на кроці 3 алгоритму.

Результати виконання алгоритму можуть бути використані для додаткового аналізу методу приховування атаки маніпулювання рейтингами. Цей аналіз здійснюється шляхом порівняння відхилень $d_s^{(j)}$ та характеристик $b_s^{(j)}$.

ВИСНОВКИ

Розглядалась проблема виявлення шиллінг-атак з урахуванням як явного, так і неявного зворотного зв’язку. Метод виявлення шиллінг-атак був удосконалений шляхом порівняння змін у рейтингах та часу користування на послідовних інтервалах часу. Була обґрунтована сфера застосування методу для виявлення атак такого типу. Було проведено експериментальну перевірку методу та практичну сферу застосування удосконаленого методу для виявлення шиллінг-атак у рекомендаційних системах.

Актуальність дослідження шилінг-атак визначається тим, що існуючі методи виявлення шилінг-атак не враховують короткострокові зміни вподобань користувачів. Тому такі зміни можуть розглядатись як нові атаки. Для вирішення цієї проблеми необхідно врахувати темпоральну складову цих атак.

Розробка нових методів, які використовують адаптивне комбіноване навчання, може забезпечити більш ефективне виявлення шилінг-атак на основі обробки коротких вибірок даних.

Мета даної роботи досягнута, було розроблено ефективний алгоритм виявлення шилінг-атаки та визначений їх вплив на якість рекомендаційної системи.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Riedl, J., Jameson, A., & Konstan, J. (2004). "AI Techniques for Personalized Recommendation." Proposal.
2. Aggarwal, C. (2016). Recommender Systems. Springer, 498. doi: https://doi.org/10.1007/978-3-319-29659-3 (дата звернення 15.04.2023).
3. Melville, P., Sindhwani, V, "Recommender Systems." IBM T.J. Watson Research Center, Yorktown Heights, NY 10598: 21 p.
4. Hallinan, B., Striphas, T. (2014). Recommended for you: The Netf lix Prize and the production of algorithmic culture. New Media & Society, 18 p, 117–137. doi: https://doi.org/10.1177/1461444814538646 (дата звернення 15.04.2023).
5. D. Goldberg, D. Nichols, B. M. Oki, and D. Terry.(1992). "Using collaborative filtering to weave an information tapestry". Communications of the ACM, 10 p.
6. Hanani.U, Shapira.B, Shoval.P, (2001), "Information Filtering: Overview of Issues, Research and Systems". User Modeling and User-Adapted Interaction, 57 p.
7. Shardanand.U, Maes.P. (1995). "Social Information Filtering: Algorithms for Automating “Word of Mouth". In CHI ’95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 8 p.
8. Aha.W, Kibler.D, Albert.M. (1991). "Instance-Based Learning Algorithms". Machine Learning, 30 p.
9. Sarwar.B, Karypis.G, Konstan.J, and Riedl.J. (2001). "Item-Based Collaborative Filtering Recommendation Algorithms". In WWW ’01: Proceedings of the 10th International Conference on World Wide Web, New York, NY, USA, 11 p.
10. Karypis.G. (2001). "Evaluation of Item-Based Top-N Recommendation Algorithms". In CIKM ’01: Proceedings of the Tenth

International Conference on Information and Knowledge Management, New York, NY, USA. 8 p.

11. Herlocker.J, Konstan.J, Al Borchers, and Riedl.J. (1999). "An Algorithmic Framework for Performing Collaborative Filtering". In SIGIR ’99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 8 p.

12. Breese. J. S, Heckerman. D, Kadie. C. (1998). "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." Microsoft Research: 19 p.

13. Chickering, D., Heckerman, D., and Meek, C. (1997). "A Bayesian approach to learning Bayesian networks with local structure".

14. Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments.” User Modeling and User- Adapted Interaction: 40 p.

15. Andrew I. Schein, A. P., Lyle H. Ungar, and David M. Pennock (2002). "Methods and Metrics for Cold-start Recommendations." In SIGIR ’02: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,New York: 8 p.

16. Mark Claypool, A. G., Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin. (1999). "Combining Content-Based and Collaborative Filters in an Online Newspaper." In Proceedings of ACM SIGIR Workshop on Recommender Systems.

17. Burke, R. (1999). "Integrating Knowledge-Based and Collaborative-Filtering Recommender Systems." In Proceedings of the AAAI Workshop on AI in Electronic Commerce: 4 p.

18. Lang, K. (1995). "NewsWeeder: Learning to Filter Netnews." In ICML ’95: Proceedings of the 12th International Conference on Machine Learning,San Mateo, CA, USA,: 9 p.

19. Roy, R. J. M. a. L. (2000). "Content-Based Book Recommending Using Learning for Text Categorization." In DL ’00: Proceedings of the Fifth ACM Conference on Digital Libraries, New York, NY: 10 p.

20. Rashid.A.M, Albert.I,Cosley.D, K.Lam.S., McNee.S, A.

Konstan.J,and Riedl.J.(2002). "Getting to Know You: Learning New User Preferences in Recommender Systems." In IUI ’02: Proceedings of the 7th International Conference on Intelligent User Interfaces,New York, NY: 8 p.

21. McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S., Rashid, A., Konstan, J., and Riedl, J. (2002). "On the recommending of citations for research papers". In Proceedings of the 2002 ACM Conference on Computer-Supported Cooperative Work. ACM Press, New Orleans, LA, USA, 10 p.

22. Torres, R., McNee, S., Abel, M., Konstan, J., and Riedl, J. (2004). "Enhancing digital libraries with TechLens". In Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries. ACM Press, Tuscon, AZ, USA, 9 p.

23. Nichols, D. M. "Implicit Rating and Filtering." 6 p.

24. C.-Y. Chung, P.-Y. Hsu, and S.-H. Huang, ‘‘βP: A novel approach to filter out malicious rating profiles from recommender systems,’’ Decis. Support Syst., vol. 55, no. 1, pp. 314–325p, Apr. 2013.

25. A. Bilge, Z. Ozdemir, and H. Polat, ‘‘A novel shilling attack detection method,’’ Procedia Comput. Sci., vol. 31, pp. 165–174p, Jan. 2014.

26. Z. Yang, Z. Cai, and Y. Yang, ‘‘Spotting anomalous ratings for rating systems by analyzing target users and items,’’ Neurocomputing, vol. 240, pp. 25–46p, May 2017.

27. F. Zhang, Z. Zhang, P. Zhang, and S. Wang, ‘‘UD-HMM: An unsupervised method for shilling attack detection based on hidden Markov model and hierarchical clustering,’’ Knowl.-Based Syst., vol. 148, pp. 146–166p, May 2018.

28. F. Zhang, Z.-J. Deng, Z.-M. He, X.-C. Lin, and L.-L. Sun, Detection of shilling attack in collaborative filtering recommender system by PCA and data complexity, in Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC), vol. 2, Jul. 2018, pp. 673–678p.

29. Z. Luo and C. Liang, ‘‘An insider attack on shilling attack detection for recommendation systems,’’ in Proc. 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS), Aug. 2016, pp. 277–280p.

30. L. Yang, W. Huang, and X. Niu, ‘‘Defending shilling attacks in recommender systems using soft co-clustering,’’ IET Inf. Secur., vol. 11, no. 6, pp. 319–325p, Nov. 2017.

31. A. M. Turk and A. Bilge, ‘‘A robust multi-criteria collaborative filtering algorithm,’’ in Proc. Innov. Intell. Syst. Appl. (INISTA), Jul. 2018, pp. 1–6p.

32. D. Deng, J. J. Mai, C. K. Leung, and A. Cuzzocrea, ‘‘Cognitive-based hybrid collaborative filtering with rating scaling on entropy to defend shilling influence,’’ in Proc. 8th Int. Conf. Netw., Commun. Comput., Dec. 2019, pp. 176–185p.

33. S. Alonso, J. Bobadilla, F. Ortega, and R. Moya, ‘‘Robust modelbased reliability approach to tackle shilling attacks in collaborative filtering recommender systems,’’ IEEE Access, vol. 7, p. 41782–41798, 2019.

34. G. Adomavicius and A. Tuzhilin, ‘‘Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,’’ IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749p, Jun. 2005.

35. D. W. Oard and J. Kim, ‘‘Implicit feedback for recommender systems,’’ in Proc. AAAI Workshop recommender Syst., Wollongong, NSW, Australia, vol. 83, 1998, pp. 81–83p.

ДОДАТОК А

Відомість кваліфікаційної роботи

Позначення	Найменування	Дод. відомості
	Текстові документи	
1.	Пояснювальна записка	61 с.
	Інші документи	
2.	Презентаційні матеріали	17 плакатів

.	.	Прізвище та ініціали.	Підп.	Дата	
Розробив		Іщенко А.І.			Дослідження та розробка методів виявлення атак користувачів на рейтинги в рекомендаційних системах
Перевірив		Чалий С.Ф.			
Н.контр.		Малєєва І.А.			
Затв.		Філатов В.О.			

Шифр групи	Код напр./спец.
СШІм-21-2	122
ХНУРЕ кафедра ШІ	