КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№1 (5), 2019
ISSN 2663 - 4023
DOI: 10.28925/2663-4023.2019.5.95104
УДК 004.9

Мелешко Єлизавета Владиславівна
кандидат технічних наук, доцент, докторант кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна
OrcID: 0000-0001-8791-0063
elismeleshko@gmail.com

Хох Віталій Дмитрович
аспірант кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна
OrcID: 0000-0002-5608-4632
vd.khokh@gmail.com

Улічев Олександр Сергійович
аспірант кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна
OrcID: 0000-0003-3736-9613
askin79@gmail.com

ДОСЛІДЖЕННЯ РОБАСТНОСТІ РЕКОМЕНДАЦІЙНИХ СИСТЕМ З КОЛАБОРАТИВНОЮ ФІЛЬТРАЦІЄЮ ДО ІНФОРМАЦІЙНИХ АТАК

Анотація. У даній статті здійснено дослідження робастності рекомендаційних систем з колаборативною фільтрацією до інформаційних атак, метою яких є накручування рейтингів деяких об'єктів системи. Досліджено вразливості методів колаборативної фільтрації до інформаційних атак, а також розглянуто основний вид атак на рекомендаційні системи – атаку ін'єкцією профілів. Розглянуто способи оцінки робастності рекомендаційних систем до атак ін'єкцією профілів за допомогою таких показників як середній зсув прогнозування оцінок та коефіцієнт звернень користувачів до рекомендацій. Описано загальний спосіб тестування робастності рекомендаційних систем. Наведено класифікацію методів колаборативної фільтрації та здійснено порівняння їх робастності до інформаційних атак. Виявлено, що методи колаборативної фільтрації засновані на моделі більш робастні, ніж методи засновані на пам'яті, а методи на основі коефіцієнтів подоби об'єктів, більш стійкі до атак, ніж методи засновані на коефіцієнтах подоби користувачів. Досліджено методи виявлення інформаційних атак на рекомендаційні системи на основі класифікації профілів користувачів. Розглянуто показники, на основі яких можна виявити як окремі профілі ботів у системі, так і групи ботів. Наведено способи оцінки якості роботи класифікаторів профілів користувачів, зокрема, обчислення таких показників як точність, повнота, точність негативного прогнозу та специфічність. Розглянуто спосіб підвищення робастності рекомендаційних систем за допомогою введення параметра репутація користувачів, а також методів одержання числового значення параметру репутації користувачів. Результати даних досліджень у подальшому будуть спрямовані на розробку програмної моделі рекомендаційної системи для тестування робастності різних алгоритмів колаборативної фільтрації до відомих інформаційних атак.

Ключові слова: рекомендаційні системи; колаборативна фільтрація; інформаційна безпека; інформаційна атака; робастність; ідентифікація атаки

1. ВСТУП

Рекомендаційні системи використовуються на веб-сайтах та у програмних додатках для створення списків рекомендацій користувачам на основі їх попередніх дій,

наприклад, переглянутого контенту, виставлених оцінок, списків друзів, демографічних даних, тощо.

Алгоритми роботи рекомендаційних систем можуть використовувати різні методи фільтрації даних, зокрема, контентну, колаборативну, соціальну фільтрацію, тощо [1]. Кожен з видів фільтрації має багато різних способів реалізації та модифікацій. Сучасні рекомендаційні системи – це складні гібриди різних методів фільтрації даних [2]. Переважна більшість гібридних рекомендаційних систем використовують у своєму складі ті чи інші методи колаборативної фільтрації, так як цей вид фільтрації даних дозволяє створювати досить точні рекомендації користувачам, які активно взаємодіють з контентом системи, здійснюють перегляди, виставляють оцінки, тобто дають якісний зворотній зв'язок системі. Колаборативна фільтрація на основі дій користувача дозволяє знаходити схожих на нього користувачів («сусідів») та створювати йому список рекомендацій на основі відомих вподобань «найближчих» до нього «сусідів», що складається з контенту, який він ще не переглядав.

Одним з суттєвих недоліків колаборативної фільтрації є її вразливість до інформаційних атак ін'єкцією профілів [1, 3-8]. Даний тип атак полягає у створенні деякої кількості профілів ботів, які узгоджено будуть виставляти потрібні оцінки об'єктам системи [1, 3, 4, 7, 8]. Дані атаки будуть спрямовані на цільові об'єкти, яким зловмисник хоче підвищити або понизити рейтинги у системі, щоб змінити частоту їх потрапляння у списки рекомендацій справжнім користувачам. Цільовим об'єктам боти будуть виставляти цільові оцінки (максимальні, або мінімальні в залежності від мети зловмисника). Також будуть виставлятися оцінки деяким нецільовим об'єктам для наповнення профілів ботів історією дій та імітації поведінки справжніх користувачів. Спосіб вибору об'єктів для наповнення профілю та оцінки для них залежать від моделі атаки, яку обере зловмисник.

Постановка проблеми. Оскільки рекомендаційні системи використовуються все частіше на різних веб-ресурсах та фактично стають доповненням, а на деяких сайтах навіть альтернативою, до пошукових систем і дозволяють не тільки створювати маркетингові пропозиції, але і формувати стрічки новин та покращувати пошукові видачі, то впливати та результати роботи рекомендаційних систем стає все привабливіше для зловмисників. Впливаючи на вміст інформації, яку переглядають користувачі, зловмисники можуть впливати на їх думки та судження, а таким чином і на їх рішення та дії. Тому актуальною є задача дослідження та підвищення робастності рекомендаційних систем до інформаційних атак.

Аналіз останніх досліджень і публікацій. У роботах [1, 3, 4, 7, 8] розглянуті відомі атаки на рекомендаційні системи з колаборативною фільтрацією, основним типом таких атак є атаки ін'єкцією профілів, до відомих моделей атак відносяться: випадкова атака, середня атака, атака приєднання до більшості, популярна атака, атака любов/ненависть, тощо. У роботах [1, 5, 6, 7] розглянуті методи підвищення робастності рекомендаційних систем, для підвищення надійності рекомендаційних систем доцільно використовувати підсистему виявлення та нейтралізації профілів ботів, а також додавати до системи параметр репутація користувачів. У [1, 3, 5] здійснено порівняння відомих алгоритмів колаборативної фільтрації з погляду їх стійкості до атак ін'єкцією профілів, зокрема, існують дослідження, які показують, що методи засновані на моделі більш стійкі до атак, ніж методи засновані на пам'яті.

Метою статті є дослідження робастності рекомендаційних систем з колаборативною фільтрацією до інформаційних атак ін'єкцією профілів та методів захисту від даних атак.

ISSN 2663 - 4023

2. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Робастність (надійність) рекомендаційної системи – здатність її адаптуватися до потреб звичайних користувачів, ігноруючи інформаційні впливи користувачів-зловмисників [1]. Надійна рекомендаційна система повинна мати захист від інформаційних атак, щоб надавати якісні і достовірні рекомендації своїм користувачам.

Підвищувати робастність рекомендаційної системи можна наступними способами [1, 3, 10]:

1. Заміняти одні методи колаборативної фільтрації всередині рекомендаційної системи на інші більш робастні аналоги.

2. Використовувати методи виявлення профілів ботів та вилучати їх дані з розрахунків для формування рекомендацій користувачам системи.

3. Впровадити та враховувати у рекомендаційній системі параметр репутація для користувачів таким чином, щоб користувачі з низькою репутацією, визначеною на основі їх попередніх дій, слабо впливали або не впливали на формування списків рекомендацій.

Перед тим як розглянути вищезгадані способи підвищення надійності рекомендаційних систем, слід дослідити способи визначення рівня робастності різних методів колаборативної фільтрації.

2.1 Оцінювання робастності рекомендаційних систем

Оскільки мета інформаційних атак на рекомендаційні системи – це зміна рейтингу цільового об’єкту, то для вимірювання робастності рекомендаційної системи, потрібно оцінити, наскільки змінилися рейтинги об’єктів системи після атаки.

Також показники робастності повинні фіксувати відмінності в рекомендованому статусі цільового об’єкту до та після атаки, тобто визначати наскільки змінилася частота потрапляння цільового об’єкту у списки рекомендацій користувачам та чи покращилися його позиції у даних списках. Ці дані важливіші за дані про зміни у рейтингах, оскільки показують ефективність атаки. Але показники, спрямовані на вявлення змін у рекомендованому статусі об’єкту, можуть не зафіксувати невдалу атаку на відміну від показників орієнтованих на виявлення змін у рейтингах, так як незначна зміна рейтингу об’єкту системи може не вплинути на його рекомендований статус, але може бути зафіксована.

Багато дослідників використовують [1, 3] середній зсув прогнозування для оцінювання змін у прогнозованих рейтингах.

Нехай U_T та I_T – це набори користувачів та об’єктів системи. Для кожної пари user-item (u, i) зсув прогнозування може бути виміряний як:

$$\Delta_{u,i} = p'_{u,i} - p_{u,i}, \quad (1)$$

де p і p' є прогнози до- та після атаки відповідно.

Позитивне значення $\Delta_{u,i}$ означає, що атака зуміла збільшити прогнозовані рейтинги об’єкту, а негативне – зменшити їх. Середній зсув прогнозування рейтингів об’єкта i для всіх користувачів можна обчислити наступним чином:

$$\Delta_i = \sum_{u \in U_T} \frac{\Delta_{i,u}}{|U_T|}, \quad (2)$$

де $|U_T|$ – кількість елементів у наборі користувачів U_T.

Аналогічно середній зсув прогнозування рейтингів для всіх об’єктів у тестовій виборці може бути обчислений як:

ISSN 2663 - 4023

$$\overline{\Delta} = \sum_{i \in I_T} \frac{\Delta_i}{|I_T|}, \quad (3)$$

де $|I_T|$ – кількість елементів у наборі об'єктів I_T.

Зсув прогнозування дозволяє дослідити як атаки впливають на рейтинги цільових об'єктів. Однак навіть дуже сильні зміни у рейтингу об'єкту можуть не змінити його рекомендований статус. Така ситуація може виникнути, напр., якщо його початкова середня оцінка дуже низька, що навіть сильний її приріст недостатній для потрапляння у списки рекомендацій.

Для оцінювання впливу атаки на списки рекомендацій існує наступний показник [1, 3, 10] – коефіцієнт звернень. Нехай R_u – це набір найпопулярніших N рекомендацій для користувача u. Якщо цільовий об'єкт потрапляє в R_u, для користувача u, функція оцінювання результату атаки $H_{u,i}$ має значення 1; інакше – 0. Коефіцієнт звернень для елемента i визначається як:

$$HitRatio_i = \sum_{u \in U_T} \frac{H_{i,u}}{|U_T|}. \quad (4)$$

Середнє значення коефіцієнту звернень може бути обчислене як:

$$\overline{HitRatio} = \sum_{i \in I_T} \frac{HitRatio_i}{|I_T|}. \quad (5)$$

Для оцінювання робастності різних методів колаборативної фільтрації формується два набори тестових даних, одні без додавання профілів, які моделюють атаку, інші з додаванням таких профілів. Для кожного набору даних створюються списки рекомендацій та обчислюються вищерозглянуті показники робастності. Потім результати для тестових наборів порівнюються.

Існують різні моделі атак ін'єкцією профілів на рекомендаційні системи [1, 3-10], зокрема, випадкова атака, середня атака, атака приєднанням до більшості, популярна атака, атака любов/ненависть, тощо. Різні методи колаборативної фільтрації мають різний рівень робастності до різних атак.

2.2 Робастність різних методів колаборативної фільтрації до відомих моделей інформаційних атак на рекомендаційні системи

Спочатку розглянемо загальну класифікацію методів колаборативної фільтрації (рис. 1).

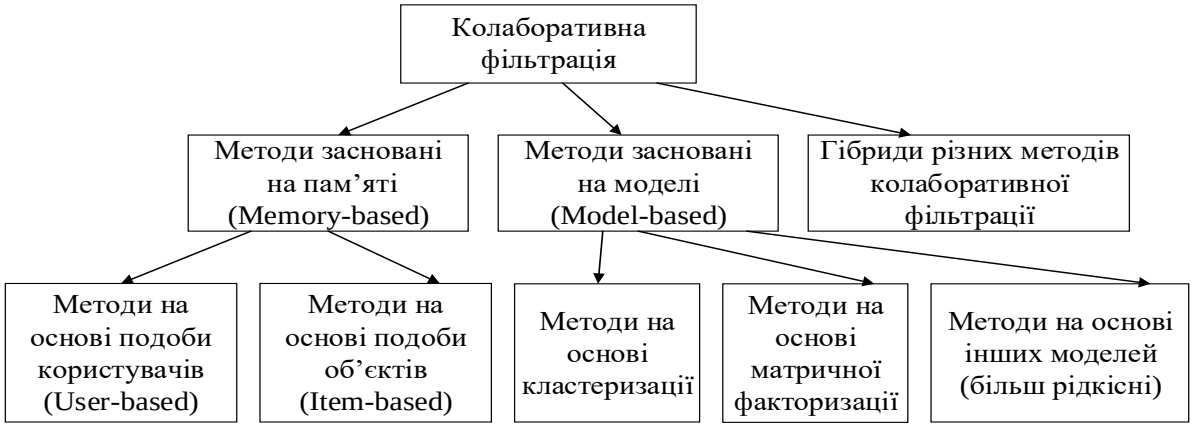

Рис. 1. Загальна класифікація методів колаборативної фільтрації

Як видно з рисунку, методи колаборативної фільтрації поділяються на дві основні групи – методи засновані на пам’яті та методи засновані на моделі [1, 2].

У методах заснованих на пам’яті рекомендації користувачам створюються на основі коефіцієнтів подоби або користувачів, або об’єктів. У методах заснованих на моделі рекомендації визначаються на основі прихованих факторів, які виявляє обрана модель. Найпоширенішими методами заснованими на моделі є методи на основі матричної факторизації (найвідоміші – svd та svd++) та методи на основі кластеризації. Більш рідкісні методи засновані на моделі можуть використовувати нейронні мережі, асоціативні правила, байєсівські мережі тощо.

Алгоритми на основі пам'яті використовують усі наявні дані з системної бази даних для обчислення прогнозів та рекомендацій. А алгоритми на основі моделей спочатку виводять модель із наявних даних, і ця модель згодом використовується в процесі створення рекомендацій.

У [11] було показано, що методи колаборативної фільтрації, засновані на моделі, забезпечують більш високу робастність рекомендаційних систем до атак, ніж методи засновані на пам’яті. Крім того, ця надійність не призводить до значного зниження точності рекомендацій.

Тож у системах з високими вимогами до робастності слід обирати алгоритми колаборативної фільтрації засновані на моделі.

Також у [1, 3] наводяться експерименти, які показують, що Item-based методи більш стійкі до багатьох інформаційних атак, ніж User-based.

Таким чином, якщо все ж таки у системі, що може зазнавати інформаційних атак, необхідно застосовувати колаборативну фільтрацію засновану на пам’яті слід надавати перевагу методам на основі подоби об’єктів.

Тим не менш обрання більш робастних методів та алгоритмів недостатнє для повного захисту рекомендаційної системи. Більш надійний захист від інформаційних атак надає впровадження підсистеми виявлення та нейтралізації профілів ботів.

2.3 Методи виявлення атак на рекомендаційну систему та профілів ботів

Виявлення атак на рекомендаційні системи базується на виявленні групи профілів користувачів, що здійснюють атаку. Після виявлення профілів ботів, можна вилучити їх інформацію з бази даних, що використовується для формування списків рекомендацій. Таким чином поставлені ними оцінки та виконані дії (перегляди, коментарі, тощо) не будуть впливати на рейтинги об’єктів системи.

Виявлення атаки на рекомендаційну систему можна розглядати як задачу бінарної класифікації профілів системи [1, 3] з двома можливими результатами для кожного профілю, а саме:

– Профіль справжнього користувача (Authentic);

– Профіль бота, створеного для атаки на систему (Attack).

Для створення такого класифікатору можна використовувати різні методи машинного навчання, які навчати на навчальній виборці профілів, що містять як справжні профілі, так і профілі ботів.

Якщо класифікатор використовує алгоритм навчання з учителем, то на етапі навчання використовується анотований набір даних профілів, тобто набір профілів, позначених як Authentic або Attack. Оскільки більшість моделей атак використовують групи профілів ботів, які працюють узгоджено, є корисним розглядати саме групи профілів разом під час класифікації.

Для запобігання атаці необхідно застосовувати навчений на навчальній вибірці класифікатор, який буде класифікувати профілі користувачів на нормальні (Authentic) та

зловмисні (Attack). Дані профілів визначених як Attack повинні вилучатися з обчислень прогнозування рейтингів та створення рекомендацій.

Якість роботи такого класифікатору можна оцінювати за допомогою стандартних метрик, таких як точність позитивного прогнозу (6) та повнота (7):

$$precision = \frac{tp}{tp + fp}, \tag{6}$$

де *tp* – правильна класифікація профілю як Attack; *fn* – неправильна класифікація профілю як Attack.

$$recall = \frac{tp}{tp + fn}, \tag{7}$$

де *tp* – правильна класифікація профілю як Attack; *fn* – неправильна класифікація профілю як Authentic.

Також корисними будуть метрики точність негативного прогнозу (8) та специфічність (9):

$$NPV = \frac{tn}{tn + fn}, \tag{8}$$

де *tp* – правильна класифікація профілю як Authentic; *fn* – неправильна класифікація профілю як Authentic.

$$specificity = \frac{tn}{tn + fp}, \tag{9}$$

де *tn* – правильна класифікація профілю як Authentic; *fp* – неправильна класифікація профілю як Attack.

Неправильна класифікація справжніх профілів призводить до вилучення хороших даних із бази даних рекомендаційної системи, що може негативно вплинути на загальну якість роботи рекомендаційної системи. Один із способів оцінити цей вплив – обчислити MAE системи до та після виявлення атаки та фільтрації профілів.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y(t) - \hat{y}(t)\right|. \tag{10}$$

Ефективність нейтралізації атаки можна оцінити за допомогою зсуву прогнозування рейтингів цільового об’єкту до та після виявлення атаки та вилучення профілів ботів з процесу обчислень рекомендаційної системи за формулою (1).

Виявлення профілю бота

Розподіл оцінок у профілі бота з високою ймовірністю буде відрізнятися від розподілу оцінок у профілях справжніх користувачів. Незважаючи на те, що зловмиснику вигідно створювати профілі ботів якомога схожими на профілі звичайних користувачів системи, у нього ніколи не може бути достатньо інформації та ресурсів для повного усунення відмінностей між ботами та звичайними користувачами.

Ознаками профілю бота можуть бути наступні статистичні особливості [1, 10]: відхилення від середнього значення оцінок є більшим, ніж зазвичай, деяка група профілів, має вищу, ніж зазвичай, подобу до профілю, який перевіряється.

Відхилення від середнього значення оцінок можна оцінити за допомогою наступного показника – відхилення оцінок від середньої угоди (RDMA):

$$RDMA_u = \sum_{i=0}^{n_u}\frac{\left|\frac{r_{u,i} - \bar{r}_i}{l_i}\right|}{n_u}, \tag{11}$$

де n_u – кількість об'єктів, які оцінив користувач u; r_u – оцінка, яку поставив користувач u елементу i; l_i – кількість оцінок, виставлених об'єкту i всіма користувачами; $\bar{r}_i$ – середнє значення усіх оцінок об'єкту i.

Ступінь подібності з топ-сусідами, дозволяє виявляти цілі групи ботів, оскільки профіль бота буде сильніше схожий на профілі найбільш схожих на нього користувачів, ніж це відбувається з профілями справжніх користувачів:

$$DegSim_u = \sum_{v=1}^{k} \frac{sim_{u,v}}{k}, \quad (12)$$

де $sim_{u,v}$ – коефіцієнт подоби між користувачами u та v.

Виявлення групи профілів ботів

Як правило, ці алгоритми використовують кластеризацію профілів системи та намагаються відрізнити кластери профілів атаки від кластерів автентичних профілів.

Основні методи виявлення груп ботів передбачають розділення профілів користувачів на два кластери на основі їх коефіцієнтів подоби [1, 12]. Аналізуючи статистику кластерів, приймається рішення про наявність нападу, і якщо так, то який кластер містить профілі атаки. Усі профілі кластера атаки ігноруються у процесі формування рекомендацій. Вважається, що напад відбувся, якщо різниця в статистичних особливостях профілів для двох кластерів досить велика. Кластер з меншим стандартним відхиленням визначається як кластер ботів. Однак, особливо для невеликих розмірів атак, значна частина кластера ідентифікованого як кластер ботів може містити справжніх користувачів.

2.4 Використання репутації користувачів для підвищення робастності рекомендаційних систем

Для боротьби з інформаційними атаками на рекомендаціні системи може бути корисним додавання параметру репутація для користувачів системи [1, 12-14].

При наявності параметру репутація, можна використовувати коефіцієнти репутації та зважувати вклад кожного користувача в прогноз рейтингів та формування рекомендацій.

Найпростіший спосіб визначення репутації користувача – зробити на веб-ресурсі можливість оцінювати його. Таким шляхом ідуть Інтернет-магазини, дошки об'яв, ресурси з оренди житла, де можна виставляти оцінки та переглядати загальні рейтинги продавців та покупців. Визначена таким чином репутація може використовуватися рекомендаційною системою, але вона сама по собі вразлива до інформаційних атак, таку репутацію можна накрутити діями ботів. А також вона лише опосередковано може вказувати на нечесність користувача у виставлені оцінок, адже така репутація є індикатором чесності/нечесності поведінки в інших діях.

Також цей параметр може визначатися за допомогою різних алгоритмів на основі статистики дій користувачів, напр.: значення репутації підвищується, коли користувач оцінює об'єкт більш правдоподібно (для кластеру, у якому знаходиться) та зменшується, коли користувач неправдиво оцінює об'єкт (його оцінка протилежна до оцінок користувачів з його кластеру) [1]. В такій системі для користувача, що прагне максимізувати свій вплив на рекомендації іншим користувачам, доцільно чесно оцінювати об'єкти.

5. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

ISSN 2663 - 4023

Проведене дослідження показало, що існує декілька стратегій підвищення робастності рекомендаційних систем до інформаційних атак, зокрема: вибір більш стійких до атак алгоритмів колаборативної фільтрації, створення підсистеми ідентифікації та нейтралізації профілів ботів, додавання параметру репутація користувача у рекомендаційні системи.

Оскільки існують різні моделі атак на методи колаборативної фільтрації, при розробці робастних рекомендаційних систем слід використовувати різні методи захисту та проводити їх тестування на стійкість до основних атак.

Для визначення робастності рекомендаційної системи до інформаційних атак необхідно моделювати атаки на неї та вимірювати зсув рейтингів об'єктів і коефіцієнт звернень до та після атаки.

Подальші дослідження будуть спрямовані на розробку програмної моделі рекомендаційної системи для тестування робастності різних алгоритмів колаборативної фільтрації до відомих інформаційних атак.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Recommender Systems Handbook. 1st edition. (2010) Editors Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. *New York, NY, USA: Springer-Verlag New York, Inc.*, 842 c.
[2] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, vol. 12, DOI: https://doi.org/10.1023/A:1021240730564
[3] O'Mahony, M.P., Hurley, N.J., Silvestre, G.C.M. (2002) Promoting recommendations: An attack on collaborative filtering. *Lecture Notes in Computer Science*, vol. 2453, pp. 494–503.
[4] Lam, S.K., Riedl, J. (2004) Shilling recommender systems for fun and profit. *Proceedings of the 13th International World Wide Web Conference*, pp. 393–402.
[5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl J. (1994) Grouplens: An open architecture for collaborative filtering of netnews. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 175–186.
[6] Sarwar, B., Karypis, G., Konstan, J., Riedl, J. (2001) Item-based collaborative filtering recommendation algorithms. *Proceedings of the Tenth International World Wide Web Conference*, pp. 285–295.
[7] Williams, C.A., Mobasher, B., Burke, R. (2007) Defending recommender systems: detection of profile injection attacks. *Service Oriented Computing and Applications*, pp. 157–170.
[8] Chirita, P.A., Nejdl, W., Zamfir, C. (2005) Preventing shilling attacks in online recommender systems. *Proceedings of the ACM Workshop on Web Information and Data Management*, pp. 67–74.
[9] Zhou, W., Wen, J., Qu, Q., Zeng, J., Cheng, T. (2018) Shilling attack detection for recommender systems based on credibility of group users and rating time series. *PLoS ONE 13(5): e0196533.* DOI: https://doi.org/10.1371/journal.pone.0196533
[10] Kumari, T., Punam, B. (2017) A Comprehensive Study of Shilling Attacks in Recommender Systems. *IJCSI International Journal of Computer Science Issues*, Vol. 14, Issue 4, URL: https://www.ijcsi.org/papers/IJCSI-14-4-44-50.pdf
[11] Mobasher, B., Burke, R.D., Sandvig, J.J. (2006) Model-based collaborative filtering as a defense against profile injection attacks. *AAAI. AAAI Press*, pp. 1388-1393.
[12] Dellarocas, C. (2000) Immunizing on-line reputation reporting systems against unfair ratings and discriminatory behavior. *Proceedings of the 2nd ACM Conference on Electronic Commerce*, pp. 150-157.

[13] Ozsoy, M.G., Polat, F. (2013) Trust based recommendation systems. *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1267-1274.

[14] Mohammadi, V., Rahmani, A.M., Darwesh, A.M., Sahafi, A. (2019) Trust-based recommendation systems in Internet of Things: a systematic literature review. *Human-centric Computing and Information Sciences*, DOI: https://doi.org/10.1186/s13673-019-0183-8

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (5), 2019
ISSN 2663 - 4023

Yelyzaveta Meleshko
Candidate of Technical Sciences, Associate Professor, Doctoral Student of Cybersecurity and Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine
OrcID: 0000-0001-8791-0063
elismeleshko@gmail.com

Vitaliy Khokh
graduate student of Cybersecurity and Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine
OrcID: 0000-0002-5608-4632
vd.khokh@gmail.com

Oleksandr Ulichev
graduate student of Cybersecurity and Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine
OrcID: 0000-0003-3736-9613
askin79@gmail.com

THE RESEARCH TO THE ROBUSTNESS OF RECOMMENDATION SYSTEMS WITH COLLABORATIVE FILTERING TO INFORMATION ATTACKS

Abstract. In this article research to the robustness of recommendation systems with collaborative filtering to information attacks, which are aimed at raising or lowering the ratings of target objects in a system. The vulnerabilities of collaborative filtering methods to information attacks, as well as the main types of attacks on recommendation systems - profile-injection attacks are explored. Ways to evaluate the robustness of recommendation systems to profile-injection attacks using metrics such as rating deviation from mean agreement and hit ratio are researched. The general method of testing the robustness of recommendation systems is described. The classification of collaborative filtration methods and comparisons of their robustness to information attacks are presented. Collaborative filtering model-based methods have been found to be more robust than memory-based methods, and item-based methods more resistant to attack than user-based methods. Methods of identifying information attacks on recommendation systems based on the classification of user-profiles are explored. Metrics for identify both individual bot profiles in a system and a group of bots are researched. Ways to evaluate the quality of user profile classifiers, including calculating metrics such as precision, recall, negative predictive value, and specificity are described. The method of increasing the robustness of recommendation systems by entering the user reputation parameter as well as methods for obtaining the numerical value of the user reputation parameter is considered. The results of these researches will in the future be directed to the development of a program model of a recommendation system for testing the robustness of various algorithms for collaborative filtering to known information attacks.

Keywords: recommendation systems; collaborative filtering; information security; information attack; robustness; attack identification

REFERENCES

[1] Recommender Systems Handbook. 1st edition. (2010) Editors Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. *New York, NY, USA: Springer-Verlag New York, Inc.*, 842 c.
[2] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, vol. 12, DOI: https://doi.org/10.1023/A:1021240730564
[3] O'Mahony, M.P., Hurley, N.J., Silvestre, G.C.M. (2002) Promoting recommendations: An attack on collaborative filtering. *Lecture Notes in Computer Science*, vol. 2453, pp. 494–503.