

**О МЕТОДАХ ПОВЫШЕНИЯ НАДЕЖНОСТИ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ
С ИСПОЛЬЗОВАНИЕМ КЛАСТЕРИЗАЦИИ ПОЛЬЗОВАТЕЛЕЙ****Евгений Юрьевич ЩЕТИНИН**

доктор физико-математических наук, доцент, профессор департамента анализа данных,
принятия решений и финансовых технологий,
Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация
riviera-molto@mail.ru
https://orcid.org/0000-0003-3651-7629
SPIN-код: 8132-6640

История статьи:

Получена 07.03.2019
Получена в доработанном
виде 02.04.2019
Одобрена 14.05.2019
Доступна онлайн
30.07.2019

УДК 519.7, 338.67**JEL:** C38, C55, L81, L86**Аннотация**

Предмет. Растущая популярность систем электронной коммерции все чаще привлекает внимание их пользователей к услугам систем рекомендаций. Коллаборативная фильтрация является одним из хорошо известных методов рекомендаций, который помогает клиентам выбирать возможные продукты, представляющие для них интерес. Однако рекомендательные системы открыты для атак злоумышленников в целях искусственного продвижения или, наоборот, дискредитации определенных продуктов. Для этого ими создаются и внедряются в системы рекомендаций поддельные профили пользователей, так называемые шиллинг-атаки, вызывающие существенное изменение рейтингов продуктов в социальной сети, нанося ей материальный и моральный ущерб.

Цели. Повышение надежности систем онлайн-коммерции к манипуляциям ими со стороны злоумышленников и недобросовестных пользователей. Разработка компьютерных алгоритмов предотвращения внедрения в систему поддельных рекомендаций.

Методология. В процессе исследования использовались методы интеллектуального анализа данных и машинного обучения.

Результаты. Рассмотрены характеристики шиллинг-атак, описаны основные модели и их параметры, а также предложены эффективные методы обнаружения поддельных профилей. Разработан компьютерный алгоритм рекомендаций на основе кластеризации как эффективный метод противодействия атакам шиллинга в социальных сетях. Исследована его устойчивость к атакам шиллинга, наиболее популярным у злоумышленников, а также проанализировано влияние различных параметров атаки.

Выводы. Компьютерное моделирование вторжений и анализ их влияния на прогноз сдвига рекомендаций показали, что алгоритм устойчив к атакам шиллинга без значительного влияния внедряемых вредоносных профилей на работу рекомендательной системы.

Ключевые слова:

рекомендательная система,
шиллинг-атака,
кластеризация

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2019

Для цитирования: Щетинин Е.Ю. О методах повышения надежности рекомендательных систем с использованием кластеризации пользователей // *Экономический анализ: теория и практика*. – 2019. – Т. 18, № 7. – С. 1348 – 1361.

<https://doi.org/10.24891/ea.18.7.1348>

Введение

С увеличением объема информации, доступной в повседневной жизни благодаря широкому использованию Интернета, рекомендательные системы (РС) стали наиболее востребованным инструментом для сбора и анализа необходимой информации.

Алгоритмы РС эффективны в автоматизации привычек людей, они собирают информацию об их предпочтениях в продуктах, товарах и услугах, таких как фильмы, музыка, книги и т.д. Как правило, системы РС представляют собой матрицу «пользователь – продукт», содержащую рейтинги продуктов,

составленные пользователями в результате либо покупки либо рекомендаций, полученных от других пользователей [1]. Всякий раз, когда пользователь запрашивает прогноз по интересующему его продукту, система строит его в виде средневзвешенного значения рейтингов других пользователей этого продукта.

Современные рекомендательные системы пока еще не в состоянии надежно различать профили подлинных пользователей и вредоносных, что делает их уязвимыми для манипуляций со стороны недобросовестных пользователей. Злонамеренные пользователи или конкурирующие компании могут вторгаться в базы данных, чтобы искусственно увеличить или снизить популярность определенного продукта.

Впервые этот процесс был описан¹ и терминологически определен как шиллинг-атака (shilling attack). Определение поддельных (фейковых) профилей и устойчивость к ним имеют решающее значение для успешности работы рекомендательных систем. Все это сделало необходимым разработку эффективного алгоритма рекомендательных систем для создания персонализированных прогнозов рекомендаций с высокой точностью в условиях неизбежных атак по внедрению поддельных профилей в рекомендательную систему.

С ростом популярности рекомендательных систем возникло несколько моделей атакующих механизмов для совершения манипуляций над рекомендациями в пользу конкретных продуктов. Шиллинг-атаки на систему генерируются путем внедрения поддельных профилей в базы данных пользователей.

В литературе атаки шиллинга классифицируются по двум направлениям: атаки, направленные на повышение рейтинга продукта, выбранного злоумышленниками (push attack) в качестве объекта атаки, и атаки, направленные на понижение его рейтинга

(nuke attack)². Чтобы повысить прогноз, то есть повысить популярность некоторого продукта, атакующие формируют новый профиль участника, который присваивает ему своим голосованием в сети высокий рейтинг. Для снижения популярности целевого товара ему присваивается низкий рейтинг. Злоумышленники генерируют фиктивные профили, назначают своим целевым продуктам максимальные или минимальные значения рейтинга в соответствии с намерениями и вносят их в базы данных, таким образом, манипулируя рекомендациями в системе в свою пользу.

Исследованиям шиллинг-атак и разработке устойчивых алгоритмов их противодействию посвящено значительное количество работ и оно продолжает расти. Публикации посвящены разработке методов отражения шиллинг-атак в системах онлайн-отчетности³, обсуждению уязвимости процесса автоматизированного обнаружения манипуляций в системе, в частности описанию объема информации, необходимого для реализации эффективных шиллинг-атак [2]. В ряде работ проанализирована структура атак и показано существование связи между конфиденциальностью информации о пользователях РС и стоимости эффективных шиллинг-атак. Ряд работ посвящен описанию моделей шиллинг-атак, таких как случайные, средние, групповые, сегментные атаки и др.⁴ [2–9].

Многие работы посвящены использованию методов кластеризации для решения поставленной задачи⁵. Успешное применение

² Burke R., Mobasher B., Williams C., Bhaumik R. Classification Features for Attack Detection in Collaborative Recommender Systems. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia, PA, USA, August 2006, pp. 542–547.

³ Cheng Z., Hurley N.J. Robust Collaborative Recommendation by Least Trimmed Squares Matrix Factorization. In: 22nd IEEE International Conference on Tools with Artificial Intelligence. 2010, pp. 105–112.

⁴ Davoudi A., Chatterjee M. Product Rating Prediction Using Trust Relationships in Social Networks. In: 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016, pp. 115–118.

⁵ См., напр.: Mehta B. Unsupervised Shilling Detection for Collaborative Filtering. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, AAAI, Menlo Park (July 2007), pp. 1402–1407.

¹ Bhaumik R., Mobasher B., Burke R.D. A Clustering Approach to Unsupervised Attack Detection in Collaborative Recommender Systems. Proceedings of the 2011 International Conference on Data Mining (DMIN'11). CSREA Press, 2011, pp. 181–187.

методов кластеризации в алгоритмах обнаружения атак шиллинга вдохновили нас на гипотезу о том, что в качестве надежного алгоритма прогнозирования может быть предложен метод кластеризации k -средних. Подобное утверждение связано прежде всего с тем, что в задаче обнаружения шиллинга невозможно применение методов машинного обучения с учителем из-за отсутствия обучающей выборки. Следовательно, остается относительно небольшой выбор методов машинного обучения без учителя, каким и является кластеризация. В дополнение к обнаружению вредоносных профилей мы предполагаем, что метод кластеризации может использоваться для разработки алгоритмов построения надежного прогноза рейтинга рекомендуемых продуктов. Кроме того, многократное применение алгоритма кластеризации должно устранить все профили шиллинга после некоторого уровня построенного бинарного дерева.

Нами предложен алгоритм прогнозирования рейтинга рекомендуемых продуктов на основе метода кластерного анализа k -средних⁶, обеспечивающий их высокую точность. Мы также проводим сравнительный анализ между предложенным методом и ранее известными подходами на основе полученных оценок сдвигов прогноза.

Модели и свойства шиллинг-атак

Наиболее известная стратегия злоумышленника по внедрению поддельных профилей в группу пользователей системы состоит в следующем. Прежде всего этот пользователь стремится к более частому упоминанию некоторого продукта в системе. Очевидно, что для этого необходимо изменить прогнозируемое значение рейтинга этого продукта для как можно большего количества пользователей. Эффективная шиллинг-атака делает это значение максимально высоким. Мерой эффективности атаки может служить величина смещения прогноза рейтинга целевого

продукта. Она определяется как разница в прогнозируемом значении рейтинга до и после нападения:

$$P = \sum_u \tilde{r}_{u_i,y} - r_{u_i,y} = \sum_u P_u, \quad (1)$$

где $\tilde{r}_{u_i,y}$ – прогноз рейтинга продукта u для пользователя u_i после атаки;

P_u – прогноз сдвига рейтинга продукта u по пользователям.

Таким образом, задача атаки состоит в максимизации P . Сам прогноз $\tilde{r}_{u_i,y}$ может быть вычислен следующим образом:

$$\tilde{r}_{u_i,y} = \bar{r}_{u_i} + \frac{\sum_{u_j \in N} (r_{u_j,y} - \bar{r}_{u_j}) C_{u_j,u_i}}{0 \sum_{u_j \in N} C_{u_j,u_i}}, \quad (2)$$

где $r_{u_j,y}$ – рейтинг продукта u для пользователя u_j ;

$\bar{r}_{u_i}, \bar{r}_{u_j}$ – средние рейтингов пользователей u_i, u_j соответственно;

N – множество ближайших соседей пользователя u_i ;

C_{u_j,u_i} – мера подобия между пользователями u_i, u_j , в качестве которой выбирается корреляция Пирсона:

$$C_{u_j,u_i} = \frac{\sum_y (r_{u_i,y} - \bar{r}_{u_i})(r_{u_j,y} - \bar{r}_{u_j})}{\sqrt{\sum_y (r_{u_i,y} - \bar{r}_{u_i})^2} \sqrt{\sum_y (r_{u_j,y} - \bar{r}_{u_j})^2}}. \quad (3)$$

Заметим, что корреляция (3) вычисляется только для тех продуктов u , на которые имеются рейтинги от обоих пользователей u_i, u_j . Также нужно отметить и другие меры подобия, например, косинусную⁷. На рис. 1 представлена общая структура шиллинг-атаки на рекомендательную систему.

Состав элементов шиллинг-атаки:

⁶ Cheng Z., Hurley N.J. Robust Collaborative Recommendation by Least Trimmed Squares Matrix Factorization. In: 22nd IEEE International Conference on Tools with Artificial Intelligence, 2010, pp. 105–112; Shchetinin E. Cluster-Based Energy Consumption Forecasting in Smart Grids. In: 21st International Conference, DCCN, 2018.

⁷ Davoudi A., Chatterjee M. Product Rating Prediction Using Trust Relationships in Social Networks. In: 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016, pp. 115–118.

- I_T – набор целевых продуктов, состоящий из одного или нескольких наименований, называемый одиночной или множественной атакой. Рейтинг $\gamma(i_j^T)$, обычно определяется как максимальное (минимальное) значение профилей целевых продуктов в сети в зависимости от типа атаки (push/nuke);
- I_S – набор выбранных элементов с рейтингом, определяемым функцией $\sigma(i_k^S)$;
- I_F – набор элементов наполнения, полученный случайным образом из выбранных продуктов со случайным присвоением рейтинга $\rho(i_k^F)$;
- I_N – набор наименований без рейтинга.

Далее рассмотрим основные модели шиллинг-атак и опишем их свойства.

Случайная атака (RN). Случайная атака может рассматриваться как базовая модель push-атаки, то есть атаки, направленной на занижение рейтинга целевого продукта. Выбранный набор целевых продуктов пусть и произвольно выбранный набор элементов заполнителя содержит случайные значения, полученные из нормального распределения со средним значением и стандартным отклонением по системе. Целевому элементу присваивается максимальный рейтинг, доступный в рекомендательной системе:

$$I_S = \emptyset, \rho(i) \sim N(\bar{r}_i, \bar{\sigma}_i^2).$$

Групповая атака (bandwagon). Как модель шиллинг-атаки типа push атака bandwagon нацелена на продукты, привлекающие особое внимание многих потребителей для манипуляций с людьми, которые склонны к тому, чтобы закупить такие продукты. Выбранный набор продуктов состоит из популярных и хорошо оцененных предметов, имеющих высокие средние значения. Для атак на рекомендательные системы таким выбранным элементам присваивается максимальный доступный рейтинг, элементам наполнения присваиваются случайные значения, а целевому элементу – наивысший

рейтинг. Набор I_S содержит популярные наименования:

$$\sigma(i) = r_{\max} / r_{\min}(\text{push / nuke}), \rho(i) \sim N(\bar{r}_i, \bar{\sigma}_i^2).$$

Сегментная атака (SG). Сегментная атака разработана как модель push-атак, направленных на группу пользователей, которые более склонны к приобретению определенных видов продуктов, чем остальные пользователи сети. Состав группы шиллинг-атаки I_S выбирается из элементов с высоким средним рейтингом с определенным свойством (например, фильмы ужасов или джазовая музыка). Таким выбранным элементам присваивается максимальное значение рейтинга, элементам наполнителя присваивается минимальное значение рейтинга, а целевому продукту назначается самый высокий голос, чтобы повысить его популярность. Набор I_S содержит набор сегментированных продуктов:

$$\sigma(i) = r_{\max} / r_{\min}(\text{push / nuke}),$$

$$\rho(i) = r_{\min} / r_{\max}(\text{push / nuke}).$$

Сгруппированная атака (RBW). Обратная модель атаки bandwagon nuke является зеркальным отражением модели атаки push bandwagon. Элементы выбираются среди непопулярных (имеющих низкие средние) элементов, оцененных многими пользователями. Для атак на рекомендательные системы таким выбранным элементам присваивается минимальный доступный рейтинг, элементам-наполнителям назначаются случайные значения, а целевому элементу – самый низкий рейтинг. Набор I_S содержит продукты с низким рейтингом:

$$\sigma(i) = r_{\min} / r_{\text{vfx}}(\text{push / nuke}), \rho(i) \sim N(\bar{r}, \bar{\sigma}^2).$$

Атака любви/ненависти (L/H). Простая модель атаки nuke, которая не требует никаких знаний о системе. Набор I_S выбранных элементов пусть и произвольно выбранным элементам наполнителя присваиваются самые высокие доступные значения рейтинга, в то время как целевой элемент получает минимальное количество голосов:

$$I_S = \emptyset, \rho(i) = r_{\max} / r_{\min} (\text{push / nuke}).$$

Основные приведенные модели шиллинг-атак и их характеристики позволяют сформулировать некоторые метрики, позволяющие различать поддельные профили от обычных пользователей системы.

Метрики подобия и различия поддельных пользователей

Рассмотрим методы обнаружения поддельных профилей в структуре рекомендательных систем с использованием различных метрик.

Rating Deviation from Mean Agreement (RDMA).

Измеряет расхождение оценки пользователя u с другими подлинными пользователями в системе, обратно пропорциональное количеству продуктов, который оценил пользователь u :

$$RDMA_u = \frac{\sum_{i=0}^{N_u} |r_{u,i} - Avg_i|}{NR_i}, \quad (4)$$

где Avg_i – средний рейтинг продукта i ;

N_u – количество продуктов, которые оценил пользователь u .

Weighted Deviation from Mean Agreement (WDMA) основана на RDMA, однако она придает больший вес отклонениям рейтинга для разреженных элементов⁸. Определяется как

$$WDMA_u = \frac{\sum_{i=0}^{N_u} |r_{u,i} - Avg_i|}{NR_i^2}, \quad (5)$$

где $r_{u,i}$ – рейтинг продукта i , присвоенный пользователем u ;

NR_i – общий диапазон рейтингов в системе, присвоенных пункту i .

Degree of similarity (DegSim). Основана на гипотезе, что профили злоумышленников

очень похожи друг на друга из-за своих характеристик и генерируются одним и тем же образом:

$$DigSim_u = \frac{\sum_{v \in neighbors(u)} W_{u,v}}{k}, \quad (6)$$

где $W_{u,v}$ – мера подобия пользователя u и k его ближайших соседей w , в качестве которой нами выбран коэффициент корреляции (3):

$$W_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2 (r_{v,i} - \bar{r}_v)^2}}.$$

Сходство связей, взаимоотношений между пользователями: подобие между двумя пользователями может также быть измерено их общими связями с другими участниками или похожими рекомендациями (тем, что принято называть общими интересами). Большее количество взаимных связей означало бы большее сходство между ними. Для фейковых связей ожидается, что количество взаимных связей мало, следовательно, сходство между пользователями невелико. Мы определяем сходство таких связей как

$$C(u, v) = \frac{|N(u) \cap N(v)|}{|N(u)|}. \quad (7)$$

Как и подобие рейтинга пользователя u , мы определяем сходство для пользователя u :

$$C(u) = \frac{\sum_{v \in N(u)} C(u, v)}{|N(u)|}. \quad (8)$$

Теперь рассмотрим фейковых пользователей.

Показатель экстремального поведения рейтингов: пользователи назначают продуктам только или высокие оценки (например, 5) или низкие оценки (например, 1). Мы характеризуем такие рейтинги пользователя u как

$$S(u) = 1 - \frac{\sigma(R_u)}{R_h - R_l}, \quad (9)$$

⁸ Davoudi A., Chatterjee M. Product Rating Prediction Using Trust Relationships in Social Networks. In: 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016, pp. 115–118.

где $\sigma(R_u)$ – стандартное отклонение всех рейтингов пользователя u ;

R_h, R_l – высший и низший рейтинги в системе.

При экстремальном поведении пользователя значение $E(u)$ будет близко к 1.

Также экстремальное поведение пользователя можно охарактеризовать с помощью показателя вариативности его рейтингов:

$$B_j(u) = \frac{\sum_{v \in N(u)} |r_{u,j} - r_{v,j}|}{|N(u)|}. \quad (10)$$

Алгоритм обнаружения шиллинг-атак с использованием кластеризации пользователей

На первом этапе алгоритм формирует матрицу «пользователь – продукт» $U_{n,m}$, где n и m представляют количество пользователей и продуктов соответственно. Обычно матрица U представляет собой разреженную матрицу, состоящую из рейтингов (лайков, фактов покупки и т.п.), которые пользователи (строки матрицы) присвоили продуктам (столбцы матрицы). Наша задача – построить прогноз рейтинга $\tilde{r}_{u_i,y}$ для продукта y , который, возможно, будет ему присвоен покупателем u_i в случае его приобретения. При этом известны некоторые уже расставленные в матрице оценки r_{ui} . Известны два подхода к созданию матрицы: искать рекомендации, основанные на пользователях (user-based collaborative filtering), либо искать похожие продукты – рекомендации, основанные на продуктах (item-based collaborative filtering). Найти, насколько другие пользователи (продукты) в базе данных похожи на данного пользователя (продукт). По оценкам других пользователей (продуктов) предсказать, какую оценку даст данный пользователь данному продукту, учитывая с большим весом тех пользователей (продукты), которые больше похожи на данный. Веса определяются из (3), а прогноз, который даст новый пользователь на целевой продукт, определяется как (2). Однако вряд ли рационально для каждой рекомендации суммировать данные от миллионов пользователей. Поэтому вместо того, чтобы

учитывать всех пользователей, мы отбираем только N ближайших соседей – N пользователей, максимально похожих на данного пользователя u_i и уже оценивших этот продукт. Критерии, по которому будут формироваться эти N пользователей мы сформулировали ранее в виде показателей (4)–(9).

Далее строим бинарное дерево решений, используя алгоритм кластеризации N -средних для разделения матрицы U на два кластера, где центры кластеров нумеруются для дальнейшего использования. Если число пользователей в каком-либо кластере превышает допустимое число ближайших соседей к центру кластера N , то такие кластеры продолжают делиться на подмножества с помощью алгоритма кластеризации N -средних. В итоге получим бинарное дерево с размеченными центрами кластеров в качестве узлов ветвей и сгруппированными соседними пользователями на конечных узлах. Таким образом, получим непрерывный процесс обновления дерева решений, которое сохраняет алгоритм для классификации нового пользователя и его возможного включения в систему.

Когда активный пользователь (a) запрашивает прогноз рейтинга на какой-либо продукт, вместо вычисления сходства со всеми пользователями алгоритм только перенаправляет активного пользователя в соответствии с его сходством в два центра кластера на каждом уровне. В процессе обхода дерева на каждом уровне выполняются два вычисления подобия, где более высокое сходство определяет следующий прогноз. Хотя глубина бинарного дерева d зависит от N , интуитивно она намного меньше в больших рекомендательных системах с высокой масштабируемостью. Поэтому после формирования дерева выполняется не более $2 \times (d - 1) + N$ вычислений подобия вместо n для формирования ближайших соседей. Затем определяется конечный узел, к которому принадлежит новый пользователь, и все пользователи в этом узле рассматриваются как ближайшие соседи. Затем прогноз рассчитывается как средневзвешенное

значение оценок соседей по целевому элементу, приведенное в (1), и возвращается к (а) как прогноз (2).

Мы полагаем, что предложенный алгоритм последовательного разделения пользователей на две группы в итоге приведет к тому, что все поддельные профили, внедренные различными шиллинг-атаками, начиная с определенного уровня дерева, будут перемещены в один кластер. Чтобы подтвердить или опровергнуть эту гипотезу, были проведены численные эксперименты на реальных базах данных.

Компьютерное моделирование шиллинг-атак и анализ надежности рекомендательных систем

Для тестирования построенного алгоритма использовался набор данных с отзывами о фильмах MovieLens, созданный компанией GroupLens. Он содержит оценки от 943 пользователей на 1 682 фильма, и диапазон оценок составляет от 0 до 5 ед. Мы также провели компьютерный анализ результатов атак по двум параметрам: размер наполнения I_F (Filler size) и размер атаки I_T (Size attack). Параметр Filler size определяет процент ячеек, которые будут заполнены поддельными рейтингами при создании профилей атаки. Размер атаки Size attack можно описать как количество поддельных профилей, пропорциональное числу пользователей в базе данных. Для оценки изменений прогноза, вызванных эффектами атак шиллинга, мы использовали метрику сдвига прогноза (1). Для проведения компьютерных экспериментов, описанных далее, нами был разработан комплекс компьютерных алгоритмов, реализованных на языке Python, и все представленные расчеты, графики получены с их помощью.

Шиллинг-атаки были внедрены в два отдельных набора из 100 фильмов для push- и nuke-атак. Эти наборы для push- и nuke-атак были построены произвольным образом из разных диапазонов рейтингов для представления характеристик исходного набора данных. Ввиду того, что неразумно искусственно повышать популярность фильма

с высокими оценками или аналогично снижать рейтинг непопулярного фильма, мы в основном выбрали фильмы с низкими средними значениями для push-атак и высокие средние для nuke-атак. Далее были построены бинарные деревья решений путем внедрения поддельных профилей в систему описанным ранее способом. Затем построены оценки прогнозов на основе полученных результатов бинарной классификации, а также вычислены оценки сдвига прогноза (1) для различных атак шиллинга. Критерий остановки для построения бинарных деревьев решений был установлен 30. Полученные результаты для push- и nuke-атак представлены далее.

Анализ надежности алгоритма k -средних в отношении шиллинг-атак

Исследование влияния параметра шиллинг-атаки Filler size (FS). Мы провели компьютерные эксперименты, чтобы показать, как величина параметра FS влияет на надежность прогнозирования рейтинга продукта с использованием кластеризации по отношению к четырем моделям push и трем nuke шиллинг-атак. Напомним, что параметр FS определяет количество поддельных рекомендаций, добавленных для заполнения области атаки I_F . Параметр атаки Size attack установлен равным 15%, а диапазон изменений FS составил от 3 до 30%. Средние значения прогноза сдвига для push- и nuke-атак представлены на рис. 2, 3.

Анализ результатов манипуляций рейтингом с помощью указанных атак (рис. 2) показал, что ни одна из четырех использованных моделей push-атаки не привела к значительным изменениям прогноза рейтинга для рассмотренного диапазона изменений параметра FS . Максимальный сдвиг прогноза 5% наблюдается для средней атаки при $FS = 20\%$. По сравнению со случайными и групповыми атаками, средние и сегментные атаки являются более эффективными. Считаем, что их влияние на устойчивость системы по-прежнему незначительно, поскольку максимальный сдвиг прогноза составляет всего около 5%. При увеличении значений FS с 3 до 20% наблюдается заметный рост значений сдвига прогноза для средней

атаки. Таким образом, можно заключить, что предложенный алгоритм прогнозирования рейтинга продукта на основе метода кластеризации N -средних устойчив к push-атакам на рекомендательные системы.

Результаты расчетов, представленные на рис. 3, показали, что модели nuke-атак не эффективны для атак рекомендательных систем в указанных пределах вариации параметра FS 0–25%. Изменения в значениях сдвига прогноза P при атаках L/H с увеличением размера наполнителя незначительны. Таким образом, атака любви/ненависти совершенно неэффективна. В отличие от атаки L/H , RW -атака наносит некоторый ущерб, следовательно, более эффективна, чем атака L/H . Однако максимальное значение сдвига прогноза (около 15%) доставляет средняя атака. Тем не менее такие изменения можно считать незначительными относительно значений рейтингов целевых продуктов.

Оценка влияния размера атаки (Attack Size). Далее мы провели компьютерные эксперименты по оцениванию характера влияния параметра AS на устойчивость алгоритма для четырех моделей атаки push и двух моделей атак nuke. Чтобы оценить устойчивость нашего алгоритма к различным значениям размера атаки, мы приняли параметр filler size равным 15%, а размер атаки изменяется от 1 до 15%. Мы оценили значения сдвига прогноза и отображали их общие средние значения для моделей атаки push и nuke на рис. 4, 5. Как видно из обоих рисунков, параметр AS является более эффективным при атаках, чем параметр FS . Результаты на рис. 4 показывают, что наиболее эффективными являются средняя и случайная атаки.

По сравнению со средними и случайными атаками, сегментные и групповые атаки оказались неэффективными. Сегментные и групповые атаки вызывают стабильные изменения прогнозов с увеличением размера атаки. Почти все значения размера сдвига прогноза для таких атак близки к 1%, что ничтожно мало. Таким образом, можно сделать вывод, что алгоритм устойчив к ним,

не вызывает значительных сдвигов прогноза P . Хотя средние и случайные атаки вызывают манипуляции, максимальный сдвиг при этом не превышает 15%, когда параметр AS составляет 3%. При увеличении размера атаки с 6 до 15% значения сдвига прогноза для средних и случайных атак уменьшаются, что можно прокомментировать как повышение точности работы алгоритма. Результаты в целом демонстрируют, что предложенный алгоритм устойчив к push-атакам для рассмотренных значений размера атаки AS .

Как видно, RW -атака оказалась наиболее эффективной атакой шиллинга (рис. 5). Когда величина параметра AS равна 6%, значение сдвига прогноза для нее достигает максимального значения, которое составляет около 7%. В отличие от RW -атаки L/H -атака практически неэффективна, так как сдвиг рейтинга от ее воздействия стабилен и составляет около 3%. Таким образом, можно заключить, что алгоритм также устойчив и к атакам L/H .

Выводы

Таким образом, на основании проведенных исследований можно утверждать, что предложенный алгоритм в основном устойчив к шиллинг-атакам. Мы проанализировали его устойчивость к шести известным атакам шиллинга (включая модели push-атак и nuke-атак). Большинство исследованных атак оказалось неэффективным, так как полученные оценки сдвига рейтинга продукта, на который были направлены эти атаки, как правило, не превышают 5%. В большинстве случаев атака любви/ненависти вызывает почти нулевые сдвиги. В некоторых случаях значения сдвига прогноза достигают 16%, но по-прежнему это является приемлемым в сравнении с диапазоном значений рейтингов. Средняя push-атака оказалась наиболее эффективной атакой против нашего алгоритма.

Нами исследованы популярные методы манипуляций поведением пользователей в рекомендательных системах, так называемые шиллинг-атаки, а также предложен эффективный компьютерный алгоритм, позволяющий снизить их влияние на рейтинг

атакуемых продуктов системы рекомендаций. В основу этого алгоритма положен метод кластеризации k -средних с модификацией выбора центров кластеров. Для тестирования предложенного метода в работе проведены компьютерное моделирование шиллинг-атак и анализ их воздействия на оценки прогноза рейтингов рекомендуемых кинофильмов из базы данных MovieLens.

Результаты компьютерных экспериментов продемонстрировали, что исследованные модели шиллинг-атак оказались не способными существенно изменить прогнозы рейтинга

рекомендаций при использовании предложенного алгоритма кластеризации фейковых пользователей рекомендательной системы.

В заключение необходимо отметить, что нами рассмотрен только один из возможных подходов к решению задачи обнаружения шиллинг-атак. В последнее время интерес к рассмотренной проблеме значительно возрос и количество работ, методов, посвященных ее решению, сложно перечислить в одной статье. Мы надеемся, что их исследование и сравнительный анализ послужат материалами для дальнейшей работы.

Рисунок 1

Общая структура шиллинг-атаки на рекомендательную систему

Figure 1

General structure of shilling attack on the recommender system

I_T			I_S			I_F			I_N		
i_1^T	...	i_j^T	i_1^S	...	i_k^S	i_1^F	...	i_l^F	i_1^N	...	i_v^N
$\gamma(i_1^T)$	...	$\gamma(i_j^T)$	$\sigma(i_1^S)$	...	$\sigma(i_k^S)$	$\rho(i_1^F)$	...	$\rho(i_l^F)$	null	...	null

Источник: авторская разработка

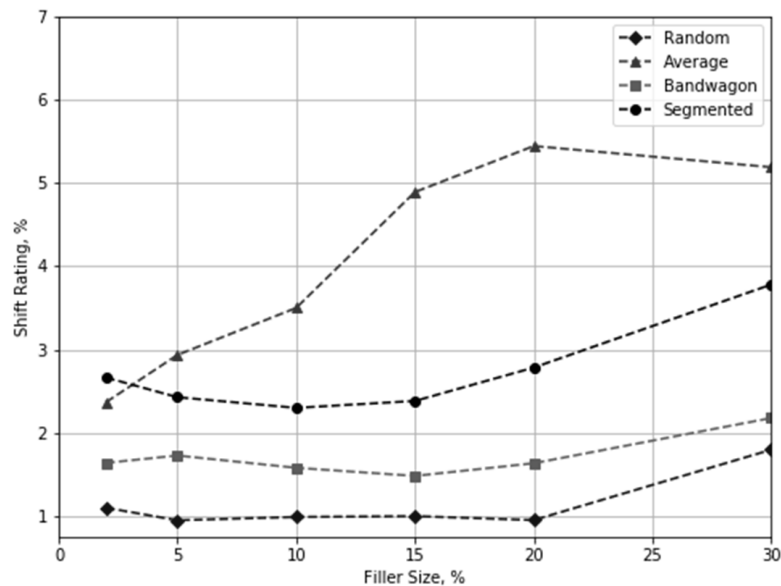
Source: Authoring

Рисунок 2

Графики смещения прогноза рейтинга целевого продукта (1) в зависимости от параметра Filler Size для различных моделей push-атаки: 1 – RN, 2 – AV, 3 – RW, 4 – SG. SA = 15%

Figure 2

Charts of bias in the forecast of target product rating (1) depending on the Filler Size parameter for different push-attack models: 1 – RN, 2 – AV, 3 – RW, 4 – SG. SA = 15%



Источник: авторская разработка

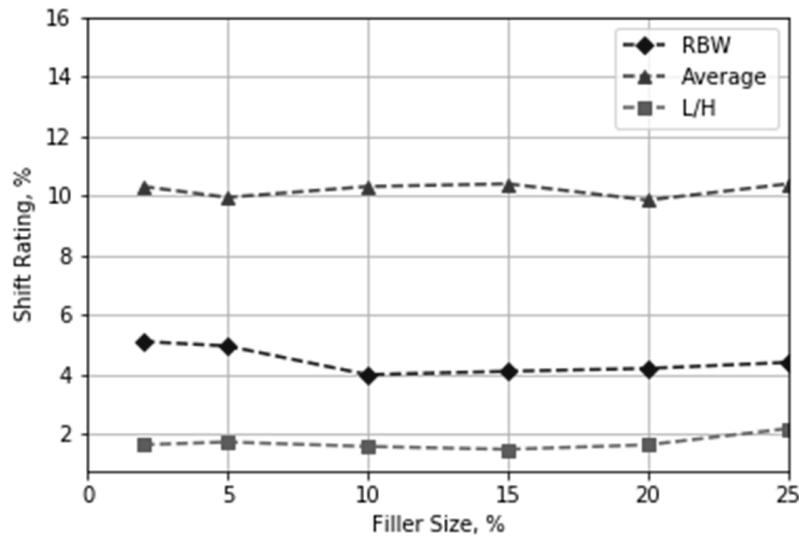
Source: Authoring

Рисунок 3

Графики смещения прогноза рейтинга целевого продукта (1) в зависимости от параметра Filler Size для различных моделей nuke-атаки: 1 – RBW, 2 – L/H, 3 – AV

Figure 3

Charts of bias in the forecast of target product rating (1) depending on the Filler Size parameter for different nuke-attack models: 1 – RBW, 2 – L/H, 3 – AV



Источник: авторская разработка

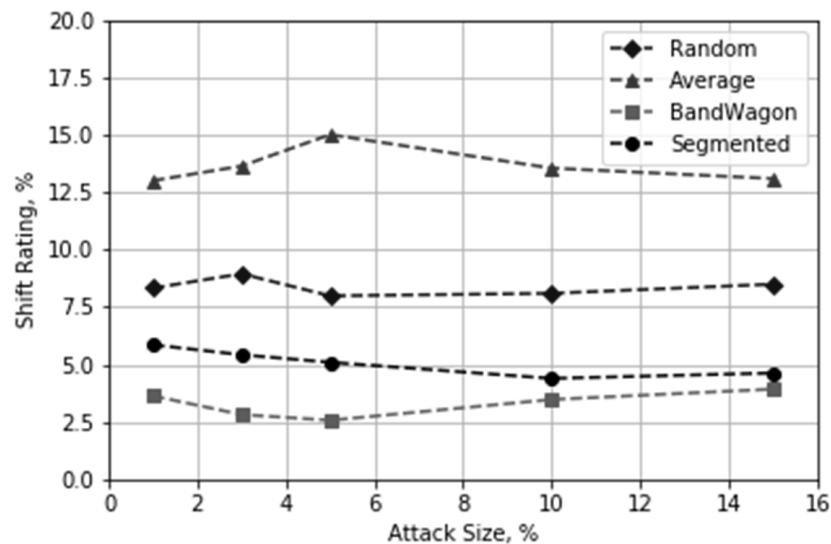
Source: Authoring

Рисунок 4

Графики смещения прогноза рейтинга атакованного продукта в зависимости от параметра Attack Size для четырех различных моделей push-атаки: 1 – RN, 2 – AV, 3 – RW, 4 – SG

Figure 4

Charts of bias in the forecast of attacked product rating depending on the Attack Size parameter for four different push-attack models: 1 – RN, 2 – AV, 3 – RW, 4 – SG



Источник: авторская разработка

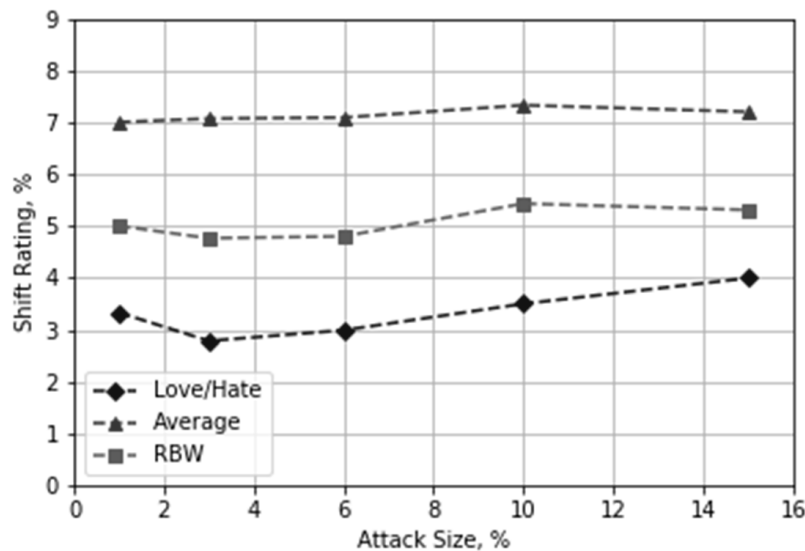
Source: Authoring

Рисунок 5

Графики смещения прогноза рейтинга атакованного продукта в зависимости от параметра Attack Size для различных моделей нuke-атаки: 1 – L/H, 2 – RBW, 3 – AV

Figure 5

Charts of bias in the forecast of attacked product rating depending on Attack Size parameter for different nuke-attack models: 1 – L/H, 2 – RBW, 3 – AV



Источник: авторская разработка

Source: Authoring

Список литературы

1. Agarwal Deepak K., Chen Bee-Chung. Statistical Methods for Recommender Systems, Cambridge University Press, 2016, 293 p.
2. Chung C.Y., Hsu P.Y., Huang S.H. χ^2 P: A Novel Approach to Filter out Malicious Rating Profiles from Recommender Systems. *Decision Support Systems*, 2013, vol. 55, iss. 1, pp. 314–325. URL: <https://doi.org/10.1016/j.dss.2013.01.020>
3. Zhou W., Wen J., Koh Y.S. et al. Shilling Attacks Detection in Recommender Systems Based on Target Item Analysis. *PLoS ONE*, 2015, no. 10(7). URL: <https://doi.org/10.1371/journal.pone.0130968>
4. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems: An Introduction. Cambridge University Press, 2010, 352 p.
5. Thorat P.B., Goudar R.M., Barve S. Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System. *International Journal of Computer Applications*, 2015, vol. 110, no. 4, pp. 31–36. URL: <https://pdfs.semanticscholar.org/51f4/78339cc86f20fff222baf17be5bab7f29bc3.pdf>
6. Azadjalal M.M., Moradi P., Abdollahpouri A., Jalili M. A Trust-Aware Recommendation Method Based on Pareto Dominance and Confidence Concepts. *Knowledge-Based Systems*, 2017, vol. 116, pp. 130–143. URL: <https://doi.org/10.1016/j.knosys.2016.10.025>
7. Deng S., Huang L., Xu G. Social Network-Based Service Recommendation with Trust Enhancement. *Expert Systems with Applications*, 2014, vol. 41, iss. 18, pp. 8075–8084. URL: <https://doi.org/10.1016/j.eswa.2014.07.012>

8. Kalai A., Zayani C.A. et al. Social Collaborative Service Recommendation Approach Based on User's Trust and Domain-Specific Expertise. *Future Generation of Computer System*, 2018, vol. 80, pp. 355–367. URL: <https://doi.org/10.1016/j.future.2017.05.036>
9. Li J., Chen C., Chen H., Tong C. Towards Context-Aware Social Recommendation via Individual Trust. *Knowledge-Based Systems*, 2017, vol. 127, pp. 58–66. URL: <https://doi.org/10.1016/j.knosys.2017.02.032>

Информация о конфликте интересов

Я, автор данной статьи, со всей ответственностью заявляю о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке данных, написанию и подготовке статьи, принятию решения о публикации рукописи.

**ON METHODS TO IMPROVE THE RELIABILITY OF RECOMMENDER SYSTEMS
BASED ON USER CLUSTERING****Evgenii Yu. SHCHETININ**Financial University under Government of Russian Federation, Moscow, Russian Federation
riviera-molto@mail.ru
<https://orcid.org/0000-0003-3651-7629>**Article history:**Received 7 March 2019
Received in revised form
2 April 2019
Accepted 14 May 2019
Available online
30 July 2019**JEL classification:** C38, C55,
L81, L86**Keywords:** recommender
system, collaborative
filtering, shilling attack,
clustering**Abstract****Subject** The growing popularity of e-commerce services is increasingly attracting the attention of their users to the services of recommendation engines. Collaborative filtering is one of the well-known recommendation methods that help customers select products of their interest. However, recommender systems are open to malicious attacks to promote or discredit certain products. For instance, the so-called shilling attacks cause a significant change in the rating of products in the social network, thereby causing significant material and moral damage.**Objectives** The purpose of the study is to increase the robustness of e-commerce systems and protect them against any manipulation by attackers and unscrupulous users; to develop computer algorithms to prevent the intrusion of fake recommendations into the recommendation system.**Methods** The study employs data mining techniques and machine learning methods.**Results** The paper considers characteristics of shilling attacks, describes their main models and parameters, and offers effective methods to detect fake profiles. I developed a computer algorithm of recommendations based on clustering as a powerful method to counter shilling attacks in social networks. I investigated its resistance to shilling attacks being the most popular among intruder, and analyzed the influence of various attack parameters on the results of its work.**Conclusions** Computer simulation of shilling attacks and analysis of their impact on the forecast of shift in recommendations showed that the algorithm is resistant to shilling attacks without significant impact of introduced malicious profiles on the work of the recommender system.

© Publishing house FINANCE and CREDIT, 2019

Please cite this article as: Shchetinin E.Yu. On Methods to Improve the Reliability of Recommender Systems Based on User Clustering. *Economic Analysis: Theory and Practice*, 2019, vol. 18, iss. 7, pp. 1348–1361.
<https://doi.org/10.24891/ea.18.7.1348>**References**

1. Agarwal Deepak K., Chen Bee-Chung. Statistical Methods for Recommender Systems. Cambridge University Press, 2016, 293 p.
2. Chung C.-Y., Hsu P.-Y., Huang S.H. χ^2 P: A Novel Approach to Filter out Malicious Rating Profiles from Recommender Systems. *Decision Support Systems*, 2013, vol. 55, iss. 1, pp. 314–325. URL: <https://doi.org/10.1016/j.dss.2013.01.020>
3. Zhou W., Wen J., Koh Y.S. et al. Shilling Attacks Detection in Recommender Systems Based on Target Item Analysis. *PLoS ONE*, 2015, no. 10(7). URL: <https://doi.org/10.1371/journal.pone.0130968>
4. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems: An Introduction. Cambridge University Press, 2010, 352 p.

5. Thorat P.B., Goudar R.M., Barve S. Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System. *International Journal of Computer Applications*, 2015, vol. 110, no. 4, pp. 31–36.
URL: <https://pdfs.semanticscholar.org/51f4/78339cc86f20fff222baf17be5bab7f29bc3.pdf>
6. Azadjalal M.M., Moradi P., Abdollahpouri A., Jalili M. A Trust-Aware Recommendation Method Based on Pareto Dominance and Confidence Concepts. *Knowledge-Based Systems*, 2017, vol. 116, pp. 130–143. URL: <https://doi.org/10.1016/j.knosys.2016.10.025>
7. Deng S., Huang L., Xu G. Social Network-Based Service Recommendation with Trust Enhancement. *Expert Systems with Applications*, 2014, vol. 41, iss. 18, pp. 8075–8084.
URL: <https://doi.org/10.1016/j.eswa.2014.07.012>
8. Kalai A., Zayani C.A. et al. Social Collaborative Service Recommendation Approach Based on User's Trust and Domain-Specific Expertise. *Future Generation of Computer System*, 2018, vol. 80, pp. 355–367. URL: <https://doi.org/10.1016/j.future.2017.05.036>
9. Li J., Chen C., Chen H., Tong C. Towards Context-Aware Social Recommendation via Individual Trust. *Knowledge-Based Systems*, 2017, vol. 127, pp. 58–66.
URL: <https://doi.org/10.1016/j.knosys.2017.02.032>

Conflict-of-interest notification

I, the author of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.