ПОВЫШЕНИЕ УСТОЙЧИВОСТИ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ ЕЕ ПОЛЬЗОВАТЕЛЕЙ

Щетинин Е.Ю., Рассахан Н.Д.

Финансовый Университет при Правительстве РФ,
Россия, Москва, Ленинградский проспект, 49
riviera-molto@mail.ru, rassahan@gmail.com

Аннотация: Рекомендательные системы открыты для атак злоумышленников с целью искусственного продвижения или дискредитации определенных продуктов. Ими создаются поддельные профили пользователей, вызывающие изменение рейтингов продуктов. Нами предложен алгоритм рекомендаций на основе кластеризации k-средних в качестве устойчивого метода противодействия атакам шиллинга.

Ключевые слова: рекомендательные системы, коллаборативная фильтрация, шиллинг-атака, кластеризация, k-средних.

Введение

С увеличением объема информации, доступной в повседневной жизни благодаря широкому использованию Интернета, рекомендательные системы (РС) стали одним из наиболее востребованных инструментов для сбора и анализа необходимой информации. Алгоритмы РС эффективны в автоматизации привычек людей, собирая информацию о предпочтениях в продуктах, таких как фильмы, музыкальные компакт-диски, книги и т.д. Как правило, РС представляют собой матрицу пользователь-продукт (user-item), содержащую рейтинги продуктов, составленные пользователями в результате либо покупки, либо рекомендаций (лайки, отзывы фактов покупки и т.п.), полученных от других пользователей [1]. Обычно она является сильно разреженной в силу того, что далеко не все пользователи оставляют отзывы по результатам покупки. а кроме того не все продукты, предлагаемые в системе, приобретаются. Известны два подхода к заполнению матрицы: строить рекомендации, основанные на пользователях (user-based collaborative filtering), либо искать схожие продукты – рекомендации, основанные на продуктах (item-based collaborative filtering). Задача РС - построить прогноз рейтинга на интересующий пользователя продукт при условии, что известны рейтинги либо похожих продуктов, либо они установлены другими пользователями. Всякий раз, когда пользователь запрашивает прогноз по интересующему его продукту, система строит его в виде средневзвешенного значения рейтингов других пользователей этого продукта.

1 Шиллинг-атаки на рекомендательные системы

Современные системы РС пока еще не в состоянии надежно отличать профили подлинных пользователей от 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 соответственно снизить популярность определенного продукта. Впервые этот процесс был описан в работе [2] 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 разработать эффективный алгоритм РС для создания персонализированных прогнозов рекомендаций с высокой точностью в условиях неизбежных атак по внедрению поддельных профилей в рекомендательную систему. Кроме того, из-за постоянно увеличивающихся размеров матрицы пользователь-продукт такие алгоритмы должны быть устойчивы к проблемам масштабируемости. Известны различные методы повышения качества получаемых прогнозов путем модификации методов расчета подобия [3,4] и обработки разреженных профилей пользователей [5,6]. Некоторые исследователи предложили алгоритмы для преодоления

проблем масштабируемости с использованием матричной факторизации [7], снижения размерности и методов кластеризации [8].

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, избранного злоумышленниками (push attack) и атаки, направленные на понижение рейтинга целевого продукта (nuke attack) [3,4]. Чтобы повысить популярность некоторого продукта, атакующие формируют профиль нового пользователя РС, который присваивает ему в сети высокий рейтинг, для снижения популярности целевого товара ему присваивается низкий рейтинг. Злоумышленники генерируют фиктивные профили, назначают своим целевым продуктам максимальный или минимальный рейтинг в соответствии с намерениями и внедряют их в РС. Таким образом, они манипулируют рекомендациями в свою пользу.

Решению задачи обнаружения и устранения шиллинг-атак уделяется значительное внимание и посвящено немало публикаций [1, 3, 5]. В настоящей работе для обнаружения атак шиллинга нами предложен метод кластеризации k-средних с адаптивным выбором центров [2]. В дополнение к обнаружению вредоносных профилей мы показали, что предложенный метод может использоваться для построения надежного прогноза рейтинга продуктов.

Наиболее известная стратегия в 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 больше 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 элемента до и после нападения

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

где $\tilde{r}_{u_i,y}$ обозначает прогноз рейтинга продукта y для пользователя ui после атаки, Pu – прогноз сдвига рейтинга продукта у по пользователям. Таким образом, задача атакующего состоит в максимизации Р. Сам прогноз $\tilde{r}_{u_i,y}$ может быть вычислен следующим образом:

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

где $r_{u_j,y}$ - рейтинг продукта у для пользователя uj, $\bar{r}_{u_i}, \bar{r}_{u_j}$ – средние рейтингов пользователей u_i, u_j N – множество ближайших соседей пользователя ui, C_{u_j,u_i} - мера подобия между пользователями ui,uj, в качестве которой выбирается корреляция Пирсона. Заметим, что корреляция (3) вычисляется только для тех продуктов у, на которые имеются рейтинги от обоих пользователей ui, uj.

2 Алгоритм кластеризации пользователей с целью выделения фальшивых профилей

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 сильно разреженную матрицу, состоящая из рейтингов (лайков, фактов покупки и т.п.), которые пользователи (строки матрицы) присвоили продуктам (столбцы матрицы). Наша задача – предсказывать оценки $\tilde{r}_{u_i,y}$, зная некоторые уже расставленные в матрице оценки $r_{u_i,y}$. 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 корреляции Пирсона, 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 ui и уже оценивших этот продукт.

Затем алгоритм строит бинарное дерево, 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, сервер только перенаправляет активного пользователя в соответствии с его сходством в два центра кластера на каждом уровне. Во время обхода 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. Наконец, определяется конечный узел, к которому принадлежит новый пользователь, и все пользователи в этом соответствующем узле рассматриваются как ближайшие соседи. Затем прогноз рассчитывается как средневзвешенное значение оценок соседей по целевому элементу, приведенное в (1), и возвращается к (a) 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 с достаточно высокой точностью. Чтобы подтвердить или опровергнуть эту гипотезу в работе были проведены численные эксперименты на реальных базах данных.

Заголовок раздела «Литература» набирается шрифтом Arial размером 11 пт., жирным, выравнивание – по левому краю, интервал с отступом до абзаца 6 пт., после абзаца – 3 пт.

Литература

1. Agarwal Deepak K., Chen Bee-Chung, Statistical Methods for Recommender Systems, Cambridge University Press, 2016.

2. R. Bhaumik, B. Mobasher, R. D. Burke, A clustering approach to unsupervised attack detection in collaborative recommender systems, IEEE international conference on data mining, pp. 181–187, 2011.

3. R. Burke, B. Mobasher, C. Williams, and R. Bhaumik, Classification features for attack detection in collaborative recommender systems, ACM SIGKDD, pp 542–547, 2006.

4. Z. Cheng, N. J. Hurley, Robust collaborative recommendation by least trimmed squares matrix factorization, IEEE conference on tools with artificial intelligence, pp. 105–112, 2010.

5. C. Y. Chung, P. Y. Hsu and S. H. Huang, A novel approach to filter out malicious rating profiles from recommender systems, Decision Support Systems, vol. 55, No. 1, pp. 314–325, 2013

6. A. Davoudi, M. Chatterjee, Product rating prediction using trust relationships in social networks, IEEE Annual Consumer Communications Networking Conference (CCNC), pp. 115–118, 2016.

7. W. Zhou, J. Wen , Y. S. Koh, Q. Xiong, M. Gao, G. Dobbie, and Sh. Alam, Shilling Attacks Detection in Recommender Systems Based on Target Item Analysis, PLOS ONE journal, 2015.

8. B. Mehta. Unsupervised shilling detection for collaborative filtering. In AAAI, p. 1402 1407, 2007.