Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 10, Number 2, Summer 2021, pp. 53-67
Received: 06/02/2017, Revised: 11/07/2020, Accepted: 06/07/2021

Correlations Farm, An Effective Shilling Attack Model to Item-based Graph Recommender Systems

Sima Iranmanesh[1], Mohammad-Reza Pajoohan[2*]
1- School of Computer Engineering, Yazd University, Yazd, Iran.
2*- School of Computer Engineering, Yazd University, Yazd, Iran.
[1] siranmanesh@stu.yazd.ac.ir, [2*] pajoohan@yazd.ac.ir

*Corresponding author's address: Mohammad-Reza Pajoohan, School of Computer Engineering, Yazd University, Yazd, Iran.

Abstract- Despite their great success in e-commerce sites, recommender systems are vulnerable to fake profiles created by malicious users. There are numerous considerable studies in the area of recommender systems robustness in the face of profile injected attacks. However, as far as we know, there is limited research on the robustness of graph-based recommendation algorithms. There are a group of recommendation algorithms that use the PageRank idea for ranking the items of a system and generate recommendation lists for users. In this paper, we study the vulnerability of these algorithms against different attack models. We proposed a new attack model by taking advantage of link spamming attacks on the web. Our experimental results show that the proposed attack model is more effective against these algorithms comparing old attack models.

Keywords- Recommender Systems, Graph-based, Correlation graph, Shilling attacks, Link farm.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۰، شماره ۲، تابستان ۱۴۰۰، صفحه ۵۲-۶۷
دریافت: ۱۳۹۶/۰۳/۱۰۱۲، بازنگری: ۱۳۹۹/۰۸/۱۷، پذیرش: ۱۴۰۰/۰۳/۱۷

مزرعه ارتباط، روشی مؤثر برای حمله به الگوریتم PageRank در سیستم‌های توصیه‌گر مبتنی برگراف آیتم‌ها

سیما ایرانمنش[۱]، محمدرضا پژوهان*[۲]
۱- دانشکده مهندسی کامپیوتر ، دانشگاه یزد، یزد، ایران.
*۲- دانشکده مهندسی کامپیوتر ، دانشگاه یزد، یزد، ایران.
[1] siranmanesh@stu.yazd.ac.ir, [2*] pajoohan@yazd.ac.ir

* نشانی نویسنده مسئول: محمدرضا پژوهان، یزد، دانشگاه یزد، دانشکده مهندسی کامپیوتر.

چکیده- امروزه سیستم‌های توصیه‌گر به جزء جداناپذیری از وب‌سایت‌های تجارت الکترونیک تبدیل‌شده‌اند. بااین‌حال، عمومی و قابل‌دسترس بودن این سیستم‌ها موجب آسیب‌پذیری آن‌ها در مقابل حمله کاربرهای سودجو گشته است. مطالعات بسیاری آسیب‌پذیری الگوریتم‌های مختلف توصیه‌گر را در مقابل حمله‌هایی که با ایجاد پروفایل‌های جعلی صورت می‌گیرند، مورد بررسی قرار داده اند، که تمرکز بسیاری از آن‌ها بر روش‌های قدیمی از جمله الگوریتم‌های پالایش گروهی بوده است. گروهی از الگوریتم‌های توصیه‌گر که مورد توجه سرویس های مختلف اینترنتی قرار گرفته‌اند، برای ارائه پیشنهاد به کاربر از روش‌های مختلف تحلیل گراف از جمله قدم‌زنی تصادفی بهره می‌برند. مطالعات محدودی در زمینه بررسی آسیب‌پذیری الگوریتم‌های توصیه‌گر مبتنی بر گراف صورت گرفته که بر انواع خاصی از این روش‌ها تمرکز دارند. از این رو در این مقاله، آسیب‌پذیری گروهی از الگوریتم‌های مبتنی بر گراف آیتم‌ها که از ایده الگوریتم رتبه‌بندی PageRank در وب برای امتیازبندی آیتم‌ها و تولید پیشنهادهایشان استفاده می‌کنند، مورد بررسی قرارگرفت. بدین منظور با بهره‌گیری از حمله‌های هرزه‌نگاری اعمال‌شده به الگوریتم رتبه‌بندی PageRank ، مدل حمله جدیدی به نام مزرعه ارتباط، پیشنهاد می‌شود. نتایج به‌دست‌آمده از اعمال حمله‌های مختلف به این تکنیک‌ها نشان داده‌اند که مدل حمله ارائه شده، بر این دسته از الگوریتم‌های توصیه‌گر مبتنی بر گراف آیتم‌ها تأثیرگذار است.

واژه‌های کلیدی: سیستم توصیه‌گر، رویکرد مبتنی بر گراف، گراف ارتباط، حمله شیلینگ، مزرعه پیوند، مزرعه ارتباط

۱- مقدمه

تصمیم‌گیری بخشی از زندگی روزمره افراد است؛ اما گاهی اوقات تصمیم‌گیری به دلایلی مانند عدم تجربه یا وجود انتخاب‌های بسیار، مشکل خواهد بود. از لحاظ تاریخی افراد در تصمیم‌گیری‌هایشان به توصیه متخصصان توجه زیادی می‌کنند. اما در دنیای امروزی با حجم وسیع انتخاب‌ها که در پی گسترش اینترنت به وجود آمده‌اند، استفاده از روش‌های قدیمی دریافت پیشنهاد، غیرعملی است. بدین منظور مطالعات گسترده‌ای بر روی چگونگی ارائه پیشنهاد خودکار به افراد صورت گرفته است و سیستم‌هایی تحت عنوان سیستم‌های توصیه‌گر پیشنهاد شده‌اند [۱]، [۲]. امروزه سیستم‌های توصیه‌گر به صورت گسترده‌ای در انواع سرویس‌های اینترنتی از جمله آمازون، ای‌بی و سرویس‌های مختلف شرکت گوگل مورد استفاده قرار گرفته است.

انواع مختلفی از سیستم‌های توصیه‌گر وجود دارند که از رویکردهای متعددی برای تولید پیشنهادها استفاده می‌کنند. رویکردهای مبتنی بر محتوا [۳] و پالایش گروهی [۱]، از قدیمی‌ترین رویکردهای توصیه‌گر هستند. در رویکردهای مبتنی بر محتوا آیتم‌هایی برای توصیه انتخاب می‌شوند که از لحاظ محتوایی مشابه آیتم‌هایی باشند که قبلاً توسط کاربر مشاهده یا خریداری شده باشند.

روش‌های پالایش گروهی محبوب‌ترین روش‌ها در حوزه سیستم‌های توصیه‌گر هستند. ایده اصلی این روش‌ها، پیش‌بینی آیتم‌های مورد علاقه یک کاربر با تحلیل رفتار رتبه‌دهی وی در گذشته است [۱]. مزیت اصلی روش‌های پالایش گروهی نسبت به روش‌های مبتنی بر محتوا این است که این روش‌ها تنها متکی به داده‌های رتبه‌دهی هستند. درحالی‌که روش‌های مبتنی بر محتوا نیازمند اطلاعات دقیقی از آیتم‌ها و کاربرها در سیستم هستند [۳].

الگوریتم‌های مبتنی بر گراف، گروه جدیدتری از الگوریتم‌های توصیه‌گر هستند. این دسته از الگوریتم‌های توصیه‌گر به نوعی زیرمجموعه رویکردهای پالایش گروهی هستند و به منظور رفع برخی از کاستی‌های این روش‌ها ارائه شده‌اند [۴]. این دسته از الگوریتم‌های توصیه‌گر مورد توجه بسیاری از پژوهشگرهای این حوزه قرار گرفته‌اند [۵]، [۶]، [۷]. دسته‌ای از سیستم‌های توصیه‌گر مبتنی بر گراف وجود دارند که از ایده الگوریتم رتبه‌بندی PageRank(PR) [۸] در حوزه بازیابی اطلاعات[۱]، به‌منظور رتبه‌بندی محصولات و تولید پیشنهادها استفاده می‌کنند [۶]، [۹].

اخیراً با افزایش محبوبیت و موفقیت تکنیک‌های یادگیری عمیق[۲] و تعبیه[۳] در حوزه‌های مختلفی از جمله پردازش تصویر و پردازش متن، استفاده از این روش‌ها برای تولید پیشنهادهای موثر در سیستم‌های توصیه‌گر موردتوجه محققان قرارگرفته است. سرویس محبوب یوتیوب [۱۰] و شبکه اجتماعی پینترست [۱۱] از جمله این سیستم‌ها هستند.

باوجود رویکردهای مختلفی که در زمینه سیستم‌های توصیه‌گر ارائه شده است، تمامی این رویکردها ویژگی مشخصی دارند؛ به منظور تولید توصیه‌های شخصی‌سازی‌شده، این سیستم‌ها به اطلاعاتی در مورد سلیقه و احتیاج کاربرهایشان نیاز دارند. معمولاً هر چه سیستم‌های توصیه‌گر اطلاعات دقیق‌تری از کاربرهایشان داشته باشند، توصیه‌های مناسب‌تری ارائه می‌دهند. اما این انگیزه برای تولیدکنندگان وجود دارد که بخواهند احتمال پیشنهاد شدن محصولاتشان بیشتر از محصولات رقیبانشان در یک سیستم توصیه‌گر باشد. ازاین‌رو، افراد سودجو ممکن است با ایجاد پروفایل‌های جعلی و وارد کردن اطلاعات غیرواقعی، نحوه عملکرد این سیستم‌ها را به نفع خودشان تحت تأثیر قرار دهند. چنین مسائلی ممکن است منجر به افت کیفیت و دقت عمل این سیستم‌ها شود. چنین اقدامات سودجویانه‌ای در سیستم‌های توصیه‌گر، حمله شیلینگ[۴] یا سمی نامیده شده‌اند [۱۲]، [۱۳].

با توجه به الگوریتم‌های مختلفی که توسط سیستم‌های توصیه‌گر به منظور تولید پیشنهادها به کار می‌رود، تکنیک‌های مختلفی نیز به منظور حمله به این سیستم‌ها نیاز است. بر این اساس پژوهش‌های قابل‌توجهی در زمینه بررسی آسیب‌پذیری الگوریتم‌های مختلف توصیه‌گر در مقابل حمله‌های شیلینگ صورت گرفته است [۱۳]-[۱۷]. اما بر اساس مطالعه‌ها و بررسی‌هایی که بر روی روش‌های توصیه‌گر مبتنی بر گراف انجام شد، پژوهش‌های محدودی در زمینه آسیب‌پذیری روش‌های مبتنی بر گراف وجود دارد [۱۸] که البته تمامی تکنیک‌های مبتنی بر گراف را شامل نمی‌شوند.

این مقاله میزان آسیب‌پذیری گروهی از روش‌های مبتنی بر گراف در مقابل حمله‌هایی که با ورود اطلاعات غیرواقعی به سیستم صورت می‌گیرند، مورد بررسی قرار داده است. بدین منظور گروهی از رویکردهای توصیه‌گر مبتنی بر گراف که با بهره‌گیری از ایده الگوریتم رتبه‌بندی PR بر روی گرافی ایجاد شده از ارتباطات آیتم‌ها عمل می‌کنند، انتخاب و مورد بررسی قرار گرفتند. با توجه به شباهت رویکردهای توصیه‌گر مذکور با الگوریتم رتبه‌بندی PR در حوزه بازیابی اطلاعات، به نظر می‌رسد این الگوریتم‌ها دارای نقاط ضعف مشترکی باشند. بنابراین می‌توان احتمال داد که حمله مزرعه پیوند [۱۹] در سیستم‌های بازیابی اطلاعات احتمالاً بر سیستم‌های توصیه‌گر مبتنی بر گراف نیز اثرگذار هستند. به این منظور در این مقاله، تکنیک‌هایی مشابه با حمله مزرعه پیوند[۵] برای اعمال حمله شیلینگ به سیستم‌های توصیه‌گر منتخب ارائه شد. همچنین میزان آسیب‌پذیری این روش‌ها در مقابل مدل‌های حمله مطرح در سیستم‌های توصیه‌گر، مورد بررسی قرار گرفت.

بخش‌بندی مقاله در ادامه به شرح زیر خواهد بود. در بخش دوم، مروری بر سیستم‌های توصیه‌گر مبتنی بر گراف ارائه می‌شود. در بخش سوم، مفهوم حمله در سیستم‌های توصیه‌گر، به همراه پژوهش‌های صورت گرفته در این زمینه مورد بررسی قرار خواهد گرفت و سپس در بخش چهارم روش حمله پیشنهادی در این مقاله شرح داده خواهد شد. در بخش پنجم روش حمله پیشنهادی مورد ارزیابی قرار گرفته و درنهایت در بخش ششم، به جمع‌بندی و

پیشنهاد کارهای آینده پرداخته شده است.

۲- سیستم‌های توصیه‌گر مبتنی بر گراف

این بخش ابتدا مفاهیم پایه در سیستمهای توصیه‌گر را معرفی می‌کند و سپس مروری بر سیستم‌های توصیه‌گر مبتنی بر گراف خواهد داشت.

۲-۱- مفاهیم پایه

با توجه به کاربردهای مختلف سیستم‌های توصیه‌گر، محصولات مختلفی وجود دارند که در پژوهش‌ها آیتم خطاب می‌شوند. بنابراین، هر سیستم توصیه‌گر شامل مجموعه‌ای از آیتم‌ها و کاربرها است. مجموعه آیتم‌ها با $I = \{i_1, i_2, \dots, i_{|I|}\}$ و مجموعه کاربرها با $U = \{u_1, u_2, \dots, u_{|u|}\}$ نمایش داده می‌شوند. $|I|$ و $|U|$ به ترتیب برابر با تعداد آیتم‌ها و کاربرها در سیستم است. نظر یک کاربر نسبت به یک آیتم به صورت یک رتبه ارائه می‌شود. رتبه‌های ارائه‌شده توسط کاربرها، اعداد صحیحی هستند که در یک مقیاس مشخص قرار می‌گیرند که با $\{r_{min}, \dots, r_{max}\}$ نمایش داده می‌شوند. رتبه r_{min}، حداقل رتبه و به معنای عدم علاقه کاربر است و رتبه r_{max} بالاترین رتبه در سیستم است. رتبه کاربر u به آیتم i را معمولاً به صورت سه‌تایی (u, i, r) یا $r_{u,i}$ نمایش می‌دهند. داده‌های رتبه‌دهی در یک سیستم توصیه‌گر در قالب یک ماتریس رتبه‌دهی[۶] (RM) در نظر گرفته می‌شوند. هر درایه غیر صفر $RM(u, i) \neq 0$ از این ماتریس بیانگر رتبه کاربر u به آیتم i است. همچنین I_u و U_i به ترتیب مجموعه آیتم‌هایی که توسط کاربر u رتبه‌دهی شده‌اند و مجموعه کاربرهایی که به آیتم i رتبه داده‌اند می‌باشند.

۲-۲- الگوریتم‌های مبتنی بر گراف

الگوهای موجود در داده‌های رتبه‌دهی یک سیستم توصیه‌گر را می‌توان به صورت یک گراف مدلسازی و اطلاعات موجود را در قالب گره‌ها و یال‌ها ارائه کرد. در چنین گرافی گره‌ها بیانگر موجودیت‌هایی مثل کاربرها و یا آیتم‌ها هستند و انواع ارتباط بین موجودیت‌ها در قالب یال‌ها ارائه می‌شوند. پژوهش‌های ارائه شده از سیستم‌های توصیه‌گر مبتنی بر گراف، ساختارهای مختلفی را برای ایجاد گراف در نظر گرفته‌اند. بر این اساس، گراف ایجاد شده در یک سیستم توصیه‌گر می‌تواند گرافی ناهمگون[۷] [۷] از کاربرها، آیتم‌ها و ویژگی‌هایشان باشد و یا گرافی همگون [۹] ، [۲۰] ازآیتم‌ها یا کاربرها باشد.

تکنیک‌های توصیه‌گر مبتنی بر گراف متعددی ارائه شده‌اند که از الگوریتم‌های مختلف جستجو و تحلیل گراف، به منظور تولید پیشنهادها استفاده می‌کنند. اما در این مقاله تمرکز ما بر الگوریتم‌های مبتنی بر گرافی است که از گراف آیتم‌ها و ایده رتبه‌بندی PR برای تولید پیشنهادهایشان استفاده می‌کنند.

گوری و پوچی در [۹] یک الگوریتم امتیازدهی به نام امتیاز آیتم[۸] ارائه دادند. روش ارائه‌شده آیتم‌ها را بر اساس سلیقه هر کاربر امتیازبندی می‌کند. این الگوریتم با بهره‌گیری از ایده الگوریتم رتبه‌صفحه به تولید پیشنهادها در یک سیستم توصیه‌گر می‌پردازد. در این الگوریتم ابتدا داده‌های رتبه‌دهی به صورت یک گراف مدل‌سازی می‌شوند. در گراف ایجادشده که گراف ارتباط‌ها[۹] نام دارد، گره‌ها نشان‌دهنده آیتم‌ها و یال‌ها نشان‌دهنده ارتباط بین آیتم‌ها هستند. در صورتی بین دو آیتم در گراف یال برقرار است که کاربر یا کاربرهای یکسانی وجود داشته باشند که به هر دو آیتم رتبه داده باشند.

گراف ارتباط‌ها در قالب یک ماتریس ایجاد و به کار گرفته می‌شود که ماتریس ارتباط‌ها نام دارد. اگر U_{i_a,i_b} مجموعه کاربرهایی باشند که هم به آیتم i_a و هم به آیتم i_b رتبه داده باشند، آنگاه هر درایه از ماتریس ارتباط‌ها طبق معادله (۱) محاسبه می‌شود.

$$\tilde{C}_{i_a,i_b} = |U_{i_a,i_b}| \qquad (۱)$$

در معادله بالا، $|.|$ به معنای تعداد اعضای مجموعه U_{i_a,i_b} است. ماتریس ارتباط ایجادشده یک ماتریس متقارن در ابعاد $|I| \times |I|$ است. پس از ایجاد ماتریس ارتباط‌ها این ماتریس نرمال‌سازی می‌شود. نرمال‌سازی هر درایه از این ماتریس بر طبق معادله (۲) صورت می‌گیرد.

$$C_{i_a,i_b} = \frac{\tilde{C}_{i_a,i_b}}{\sum_{i_k \in I} \tilde{C}_{i_a,i_b}} \qquad (۲)$$

ماتریس ارتباط نرمال‌سازی شده، یک ماتریس نامتقارن است که هر درایه آن نشان‌دهنده میزان ارتباط بین دو آیتم است. در ماتریس نرمال‌سازی شده هر درایه $C_{i_a,i_b} \neq 0$ به معنی وجود یک یال مابین دو آیتم i_a و i_b در گراف ارتباط‌ها است. همچنین مقدار هر درایه از این ماتریس برابر با وزن یال متناظر در گراف ارتباط‌ها است.

اگر گراف $G = \{E, V\}$ را با مجموعه گره‌های V و مجموعه یال‌های E در نظر گرفته‌شود، آنگاه الگوریتم رتبه‌بندی PR، میزان اهمیت هر گره در گراف را محاسبه می‌کند. این الگوریتم اتصالات یک گراف را به منظور محاسبه میزان اهمیت هر گره در نظر می‌گیرد؛ یعنی یک گره در الگوریتم رتبه‌بندی PR اهمیت بالایی

دارد اگر تعداد اتصالات ورودی بالایی از گره‌های مهم داشته باشد. معادله (۳) نحوه عملکرد این الگوریتم را به منظور محاسبه امتیاز هر صفحه نمایش می‌دهد.

$$PR(n) = \alpha . \sum_{q:(q,n)\in E} \frac{PR(q)}{w_q} + (1-\alpha).\frac{1}{|V|} \qquad (۳)$$

که در آن w_q درجه خروجی گره q است و α برابر است با ضریب میرایی[۱۰] که معمولاً مقدار 0.85 می‌گیرد.

نحوه عملکرد الگوریتم امتیاز آیتم در معادله (۴) نمایش داده شده است. در این معادله از ماتریس ارتباط نرمال‌سازی شده استفاده می‌شود و به ازای هر کاربر مقادیر امتیاز آیتم متفاوتی برای آیتم‌ها محاسبه می‌شود.

$$IR_{u_a} = \alpha . C . IR_{u_a} + (1-\alpha).d_{u_a} \qquad (۴)$$

که در آن C برابر است با ماتریس ارتباط نرمال‌سازی شده و بردار d_{u_a} بردار ترجیحات کاربر u_a است که بر اساس رتبه‌دهی‌های کاربر u_a به آیتم‌ها محاسبه می‌شود. بردار d_{u_a}، بردار نرمال‌سازی شده بردار $\tilde{d}_{u_a}$ است که هر درایه آن طبق معادله (۵) محاسبه می‌شود.

$$\tilde{d}^b_{u_a} = \begin{cases} 0, \ if \ t_{a,b} \notin L_{u_a} \\ r_{a,b}, \ if \ t_{a,b} \in L_{u_a} \ \wedge \ t_{a,b} = (u_a, i_b, r_{a,b}) \end{cases} \qquad (۵)$$

که در آن L_{u_a} مجموعه آیتم‌هایی است که توسط کاربر u_a در مجموعه یادگیری رتبه‌دهی شده‌اند. هر رتبه در مجموعه یادگیری به صورت یک سه‌تایی $t_{a,b} = (u_a, i_b, r_{a,b})$ نمایش داده می‌شود که بیانگر رتبه $r_{a,b}$، توسط کاربر u_a به آیتم i_b است. هر درایه از بردار $\tilde{d}^b_{u_a}$ به صورت $d_u = \tilde{d}^b_{u_a}/sum(\tilde{d}^b_{u_a})$ نرمال‌سازی می‌شود و نهایتاً نحوه محاسبه امتیاز آیتم در معادله (۶) نمایش داده شده است.

$$\begin{cases} IR_{u_a}(0) = \frac{1}{M}.1_{|M|} \\ IR_{u_a}(t+1) = \alpha.C.IR_{u_a}(t) + (1-\alpha).d_{u_a} \end{cases} \qquad (۶)$$

معادله ارائه‌شده به ازای هر کاربر جداگانه اجرا می‌شود. بردار $IR_{u_a}(0)$ بیانگر مقدار اولیه امتیاز هر آیتم قبل از اجرای الگوریتم است که هر درایه آن برابر است با $1/|I|$. نهایتاً بعد از محاسبه بردار امتیاز آیتم‌ها به ازای هر کاربر، این بردار به منظور ارائه پیشنهادها استفاده می‌شود. هر چه امتیاز یک آیتم بیشتر باشد، کاربر با احتمال بیشتری به آن آیتم علاقه خواهد داشت. بنابراین از بین آیتم‌هایی که توسط کاربر رتبه‌ای دریافت نکرده‌اند، آیتم‌هایی که بیشترین مقدار امتیاز آیتم را دریافت کرده‌اند به کاربر پیشنهاد می‌شوند [۹].

۳- معرفی حمله در سیستم‌های توصیه‌گر

به منظور حمله به سیستم‌های توصیه‌گر، حمله‌کنندگان بسته به رویکرد مورد استفاده سیستم توصیه‌گر هدفشان، تکنیک‌های متفاوتی را به منظور تولید پروفایل‌های جعلی‌شان به کار می‌گیرند. این بخش به معرفی حمله در سیستم‌های توصیه‌گر می‌پردازد.

۳-۱- تعریف یک حمله

نتایج یک سیستم توصیه‌گر، به‌خصوص سیستم‌هایی که از بازخورد کاربرهایشان برای تولید پیشنهادها استفاده می‌کنند، شدیداً به صداقت کاربرهایشان در ارائه بازخوردها وابسته است. توصیه‌هایی که با استفاده از بازخوردهای حقیقی محاسبه می‌شوند، کیفیت بالایی دارند. در سیستم‌های توصیه‌گر، امکان ایجاد پروفایل‌های جعلی که نسبت به کاربرهای واقعی قابل‌تشخیص نیستند، وجود دارد. در این حالت، یک فرد خرابکار می‌تواند با ایجاد پروفایل‌های جعلی و ارائه بازخوردهای خاص، تولید پیشنهادها در سیستم توصیه‌گر را طبق اهداف خودش تغییر دهد. همان‌طور که پیش‌تر نیز بیان شد، این نوع آسیب‌پذیری سیستم‌های توصیه‌گر، حمله شیلینگ نامیده می‌شوند [۱۲].

۳-۲- ابعاد حمله

حمله‌های شیلینگ بر اساس پنج پارامتر میزان دانش موردنیاز، قصد حمله، مدل حمله، اندازه پروفایل و اندازه حمله بررسی می‌شوند [۲۱].

- دانش موردنیاز: میزان دانش حمله‌کننده به یک سیستم توصیه‌گر از دوجنبه میزان دسترسی به داده‌های سیستم و درجه شناخت الگوریتم توصیه‌گر و پارامترهایش سنجیده می‌شود.
- قصد حمله: بسته به هدف حمله‌گر، حمله‌ها به دو دسته تقسیم می‌شوند. دسته اول حمله‌هایی هستند که با هدف افزایش احتمال توصیه یک یا گروهی از آیتم‌ها در سیستم صورت می‌گیرند. این نوع حمله‌ها، اصطلاحاً حمله‌های افزایشی[۱۱] نام دارند. دسته دوم حمله‌هایی هستند که هدفشان کاهش احتمال توصیه یک یا گروهی از آیتم‌ها است. چنین حمله‌هایی اصطلاحاً حمله‌های کاهشی[۱۲] نامیده می‌شوند.
- مدل حمله: نحوه انتخاب آیتم‌ها و الگو و سیاست رتبه‌دهی در پروفایل‌های جعلی را مدل حمله می‌نامند.

- اندازه پروفایل: تعداد آیتم‌هایی است که توسط هر پروفایل جعلی رتبه‌دهی می‌شوند.
- اندازه حمله: بیانگر تعداد پروفایل‌های جعلی‌ای است که در ارتباط با یک حمله به سیستم اضافه می‌شوند.

۳-۲-۱- انواع مدل‌های حمله

برای خوانایی بهتر مقاله باید سعی شود تا حد امکان علامت‌گذاری متن مقاله به‌درستی انجام شود. دقت کنید تمام علامت‌ها مثل نقطه، ویرگول، نقطه ویرگول، دونقطه و علامت سؤال باید به کلمه قبل از خود چسبیده باشند و از کلمه بعدی تنها به اندازه یک فضای خالی فاصله داشته باشند. علامت خط تیره باید به اندازه یک فضای خالی از کلمه قبل و بعد از خود فاصله داشته باشد؛ مگر این که کلمه قبلی یا بعدی یک عدد باشد که در این صورت باید به آن بچسبد. بین کلماتی که جدا هستند باید یک فضای خالی فاصله باشد. از قرار دادن فاصله در دو طرف نوشته داخل هلالین خودداری کنید.

بسته به الگوریتم مورد استفاده در یک سیستم توصیه‌گر، تکنیک حمله متفاوتی موردنیاز است. برای مثال، یک تکنیک حمله موثر به سیستم‌های توصیه‌گری که بر پایه مشاهدات مشترک[۱۳] [۱۶] عمل می‌کنند، در سیستم‌هایی که بر پایه تحلیل داده‌های رتبه‌دهی هستند تاثیرگذار نخواهد بود. از این رو، برای الگوریتم‌های مختلف توصیه‌گر، تکنیک‌های حمله متفاوتی ارائه شده است.

ساختار کلی یک پروفایل حمله که توسط مباشر در [۲۱] ارائه شد، آیتم‌های موجود در یک پروفایل حمله را به چهار دسته تقسیم می‌کند.

- مجموعه آیتم‌های انتخابی[۱۴] (I_S): این مجموعه آیتم توسط حمله‌گر انتخاب می‌شوند تا ویژگی‌های مدل حمله را مشخص کنند.
- مجموعه آیتم‌های تکمیلی[۱۵] (I_F): این مجموعه از آیتم‌ها به صورت تصادفی انتخاب می‌شوند تا مانع تشخیص حمله شوند.
- مجموعه آیتم‌های رتبه‌دهی نشده[۱۶] (I_N): این مجموعه در پروفایل‌های حمله رتبه‌ای دریافت نمی‌کنند.
- آیتم هدف[۱۷] (i_t): این آیتم توسط حمله‌گر انتخاب می‌شود و بسته به‌قصد حمله کمترین یا بیشترین رتبه را دریافت می‌کند.

نحوه انتخاب و رتبه‌دهی به آیتم‌های هر مجموعه آیتم در پروفایل‌های جعلی، بستگی به مدل حمله دارد. تاکنون مدل‌های مختلفی از حمله به سیستم‌های توصیه‌گر تعریف‌شده‌اند که در ادامه به معرفی برخی از آن‌ها پرداخته خواهد شد.

حمله تصادفی توسط لم و رایدل در [۱۲] ارائه شد. در این مدل مجموعه آیتم تکمیلی به صورت تصادفی و با توزیع نرمال از میانگین و انحراف معیار کل رتبه‌های موجود در سیستم، رتبه دریافت می‌کنند. برای یک حمله افزایشی رتبه آیتم هدف برابر با بیشترین رتبه (r_{max}) و برای یک حمله کاهشی رتبه آیتم هدف برابر با کمترین رتبه (r_{max}) در سیستم خواهد بود. مجموعه آیتم انتخابی در این نوع حمله، وجود ندارد. پیاده‌سازی این حمله آسان است اما بروک و همکارانش در [۲۲] نشان دادند که این حمله چندان تأثیرگذار نیست.

حمله میانگین نیز اولین بار توسط لم و رایدل در [۱۲] ارائه شد. این حمله به‌نوعی مشابه با حمله تصادفی عمل می‌کند با این تفاوت که به‌جای استفاده از میانگین رتبه‌های کل سیستم، از میانگین رتبه‌های هر آیتم به صورت جداگانه استفاده می‌کند. این مدل حمله بر روش پالایش گروهی مبتنی بر آیتم تأثیرگذار نیست.

حمله باندواگن به‌نوعی گسترش‌یافته مدل حمله تصادفی است که در آن مجموعه I_S نیز به پروفایل‌های جعلی اضافه‌شده است. مجموعه I_S مجموعه‌ای از آیتم‌های محبوبی هستند که به صورت مکرر رتبه دریافت می‌کنند (برای مثال، باکس آفیس در حوزه فیلم). مجموعه I_S در مدل باندواگن بالاترین رتبه را می‌گیرند. مدل باندواگن تنها برای حمله‌های افزایشی به کار می‌رود [۲۱].

مشابه مدل حمله باندواگن، برای اهداف کاهشی نیز وجود دارد که حمله باندواگن معکوس[۱۸] نامیده می‌شود. در این مدل آیتم‌های مجموعه I_S از بین آیتم‌هایی انتخاب می‌شوند که محبوبیت پایینی دارند و کمترین رتبه را دریافت می‌کنند. مدل‌های باندواگن و باندواگن معکوس نیازمند دانش خاصی از سیستم نیستند، زیرا آیتم‌های محبوب و غیر محبوب در هر حوزه به‌راحتی قابل‌تشخیص‌اند [۲۱].

ماهونی و همکارانش در [۲۳]، مدل حمله کاوشگر را ارائه داده‌اند. در این مدل حمله‌گر ابتدا با تعداد محدودی آیتم، یک پروفایل می‌سازد و به‌تدریج با آیتم‌هایی که سیستم به او پیشنهاد می‌دهد پروفایل‌های بعدی را ایجاد می‌کند. در این حالت شباهت بین پروفایل‌های حمله و پروفایل‌های کاربرهای معمولی بالا خواهد بود و درنتیجه احتمال شناسایی پروفایل‌های جعلی بسیار کاهش خواهد یافت. این مدل نیاز به دانش خاصی از سیستم ندارد.

مدل‌های حمله قبلی بر روی روش‌های پالایش گروهی مبتنی بر

آیتم تأثیر چندانی نداشتند، ازاین‌رو بروک و همکارانش در [۲۴] مدل حمله سگمنت را ارائه دادند. این مدل حمله هدف گروهی از کاربرهای خاص که با احتمال بیشتری به آیتم هدف رتبه می‌دهند را هدف قرار می‌دهد. در این مدل مجموعه I_S از بین آیتم‌هایی انتخاب می‌شوند که موردعلاقه کاربرهای این گروه‌اند. همچنین مجموعه I_F پایین‌ترین رتبه را دریافت می‌کنند. اگرچه در مدل سگمنت تمامی کاربران از حمله تاثیر نمی‌گیرند اما این مدل تأثیر بیشتری بر روش مبتنی بر آیتم می‌گذارد.

هو و همکارانش در [۱۷] مدل بهینه‌ای برای حمله به سیستم‌های توصیه‌گر اجتماعی ارائه کردند که از روش‌های عامل بندی ماتریس[۱۹] استفاده می‌کنند. مدل ارائه شده مشابه با حمله سگمنت کاربر یا گروهی از کاربران را در یک شبکه اجتماعی هدف قرار می‌دهد.

مدل حمله ترکیبی توسط ژانگ به منظور حمله به یک سیستم مبتنی بر اعتماد ارائه شد. این مدل حمله، مدل‌های میانگین و باندواگن را به صورت ترکیبی استفاده می‌کند[۲۵].

استراتژی‌های مختلف دیگری برای حمله به سیستم‌های توصیه‌گر وجود دارند که به منظور کاهش احتمال کشف شدنشان، به‌نوعی با پیچیده کردن مدل‌های حمله پیشین، از کشف آن‌ها جلوگیری می‌کنند. این مدل حمله‌ها توسط ویلیامز و همکارانش در [۲۶] معرفی و حمله مبهم نام‌گذاری شده‌اند.

آسیب پذیری یک الگوریتم ترکیبی مبتنی بر گراف کاربران و پالایش گروهی در [۲۷] مورد بررسی قرار گرفته است. این الگوریتم از گراف کاربرها به منظور بهبود عملکرد پیش‌بینی در روش پالایش گروهی مبتنی بر آیتم استفاده می‌کند. در حالیکه در مقاله حاضر، الگوریتم هدف حمله، تنها از یک گراف آیتم‌ها برای رتبه‌دهی آیتم‌ها به ازاء هر کاربر استفاده می‌نماید. بنابراین دستاورد اصلی مقاله حاضر برای حمله به الگوریتم‌هایی است که از گراف آیتم‌ها جهت تولید پیشنهادها استفاده می‌نمایند.

اخیراً فنگ و همکارانش در [۱۸] آسیب‌پذیری گروهی از سیستم‌های توصیه‌گر مبتنی بر گراف را با ارائه یک روش حمله افزایشی موثر مورد بررسی قرار دادند. الگوریتم توصیه‌گر هدف [۱۸] در این مقاله با اعمال قدم‌زنی تصادفی بر گراف دوبخشی[۲۰] ناهمگون از دادگان رتبه‌دهی سیستم به رتبه‌بندی آیتم‌ها و ارائه پیشنهادها می‌پردازد. تکنیک حمله ارائه شده یک مسئله بهینه‌سازی برای انتخاب بهینه آیتم‌های کمکی در پروفایل‌های جعلی ارائه کرده است.

همچنین ژانگ و همکارانش در [۱۵] آسیب‌پذیری تکنیک‌های تعبیه را در کاربردهای مختلفی از جمله سیستم‌های توصیه‌گر و پرسش و پاسخ مورد بررسی قرار داده‌اند.

۴- حمله به سیستم‌های توصیه‌گر مبتنی بر گراف

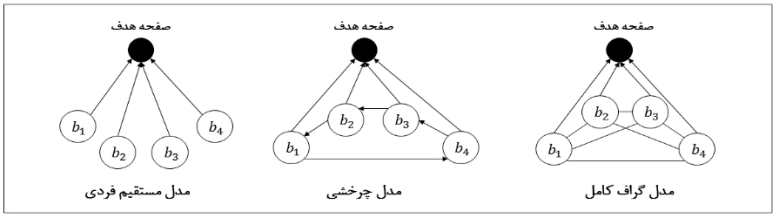
همان‌طور که پیش‌تر نیز بیان شد، دسته‌ای از روش‌های توصیه‌گر مبتنی بر گراف، از الگوریتم رتبه‌بندی PR الهام گرفته و به تولید پیشنهادها می‌پردازند که این الگوریتم در وب، در مقابل حمله‌های هرزه‌نگاری آسیب‌پذیر است. بنابراین با توجه به شباهت عملکرد الگوریتم‌های توصیه‌گر مبتنی بر گراف با الگوریتم رتبه‌بندی PR، احتمال وجود نقاط ضعف مشابهی بین این الگوریتم‌ها وجود دارد. درنتیجه در این مقاله با الهام از حمله‌های هرزه‌نگاری، روش جدیدی برای حمله به الگوریتم‌های توصیه‌گر مبتنی بر گراف آیتم‌ها ارائه شده است. در این بخش ابتدا حمله مزرعه پیوند معرفی می‌شود و سپس روش حمله پیشنهادی شرح داده خواهد شد.

۴-۱- حمله مزرعه پیوند

حمله هرزه‌نگاری مزرعه پیوند، الگوریتم‌هایی را هدف قرار می‌دهد که میزان اهمیت یک صفحه را با توجه به ساختار پیوندهای بین صفحه‌های وب به دست می‌آورند. به‌منظور افزایش امتیاز یک صفحه خاص در نتایج موتورهای جستجو، هرزه‌نگارها گروهی از صفحه‌های جعلی را ایجاد می‌کنند که با ساختار خاصی به هم متصل‌اند. چنین صفحه‌هایی را که توسط هرزه‌نگارها ایجاد و کنترل می‌شوند، مزرعه پیوند [۱۹] می‌نامند. یک مزرعه پیوند در ساده‌ترین مدل شامل ویژگی‌های زیر است [۱۹]:

- هر مزرعه پیوند شامل یک صفحه هدف است که هرزه نگار قصد دارد امتیاز آن را در نتایج موتورهای جست‌وجو بالا ببرد.
- هر مزرعه پیوند شامل مجموعه‌ای از صفحه‌های کمکی (تقویتی) [۲۱] است که به افزایش امتیاز صفحه هدف کمک می‌کنند.
- هر مزرعه پیوند می‌تواند تعدادی پیوند از صفحه‌های خارج از مزرعه نیز دریافت کند. این نوع صفحه‌ها و پیوندها را به ترتیب صفحه‌های ربوده‌شده[۲۲] و پیوندهای ربوده‌شده[۲۳] می‌نامند.

آدلی در [۲۸] مدل‌های مختلف سازمان‌دهی صفحه‌های موجود در مزرعه پیوند را مورد بررسی قرار داده است. برخی از این مدل‌ها در شکل ۱ نمایش داده شده و در ادامه معرفی می‌شوند.



شکل ۱: مدل‌های مختلف سازمان‌دهی صفحه‌ها در مزرعه پیوند [۲۸]

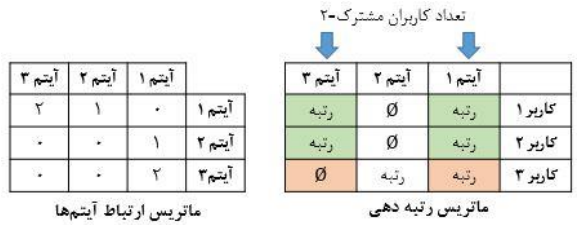
- مدل فردی مستقیم[۲۴]: در این مدل تمامی صفحه‌های کمکی فقط به صفحه هدف پیوند می‌دهند.
- مدل چرخشی[۲۵]: در این مدل صفحه‌های کمکی علاوه بر صفحه هدف، به صورت چرخشی به یکدیگر پیوند داده‌اند.
- مدل کامل[۲۶]: در این مدل صفحه‌های کمکی علاوه بر صفحه هدف، دوبه‌دو به یکدیگر پیوند داده‌اند.

۴-۲- روش حمله پیشنهادی

همان‌طور که پیش‌تر بیان شد، برای اعمال حمله مزرعه پیوند به الگوریتم رتبه‌بندی PR، هرزه‌نگارها صفحه‌هایی با ساختار پیونددهی خاص ایجاد می‌کنند. چنین صفحه‌هایی باعث تغییر در ساختار گراف وب شده و روند انتشار امتیاز را در گراف تحت تأثیر قرار می‌دهند. نتیجه تولید چنین صفحه‌هایی، انتشار امتیاز قابل‌توجهی به سمت صفحه هدف است که باعث افزایش امتیاز آن صفحه در نتایج موتور جست‌وجو می‌شود. برای شرح حمله پیشنهادی در قدم اول، به تطابق مفاهیم بین دو حوزه وب و سیستم‌های توصیه‌گر پرداخته می‌شود.

در گراف وب گره‌ها و یال‌ها به ترتیب نشان‌دهنده صفحه‌ها و پیوندهایشان در وب هستند. در مقابل، گراف ارتباط‌ها در یک سیستم توصیه‌گر مشخص‌کننده میزان ارتباط بین موجودیت‌ها در سیستم است. برخی از الگوریتم‌ها تنها از کاربرها [۲۰] یا آیتم‌ها [۹]، [۲۹] برای ساخت گراف و تولید پیشنهادها استفاده می‌کنند. برخی دیگر ممکن است گرافی ناهمگون از آیتم‌ها و کاربرها ایجاد کنند [۳۰] و یا ویژگی‌های [۳۱] آیتم‌ها یا کاربران را نیز برای ساخت گراف به‌کارگیرند. تمرکز این مقاله بر الگوریتم‌های مبتنی بر گراف آیتم‌ها [۹] است که از ایده رتبه‌بندی PR استفاده می‌کنند.

در روش‌های مبتنی بر گراف آیتم‌ها، گراف ارتباط‌ها بر اساس میزان رتبه‌های مشترک بین آیتم‌ها ایجاد می‌شود. برای مثال، در گراف آیتم‌ها در صورتی بین دو آیتم یال وجود دارد که حداقل یک کاربر به هر دو آیتم رتبه داده باشد و هر چه تعداد این کاربرها بیشتر باشد ارتباط نزدیک‌تری بین دو آیتم وجود دارد. ماتریس ارتباط آیتم‌ها، یک ماتریس $|I| \times |I|$ است که هر درایه آن برابر با تعداد کاربرهای مشترک بین دو آیتم است. مثال ارائه‌شده در شکل ۲ چگونگی ایجاد ماتریس ارتباط آیتم‌ها را در یک سیستم توصیه‌گر نمونه نمایش می‌دهد.



شکل ۲: مثالی از نحوه ایجاد ماتریس ارتباطات در یک سیستم توصیه‌گر. هر درایه در ماتریس ارتباط آیتم‌ها مشخص کننده تعداد کاربرهایی است که به هر دو آیتم رتبه داده‌اند.

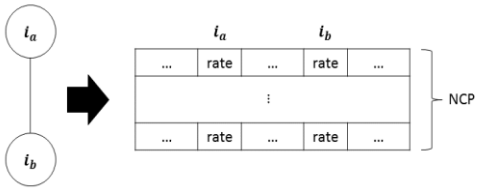
یک هرزه نگار برای ایجاد مزرعه پیوند در گراف وب، صفحه‌های کمکی را به‌گونه‌ای ایجاد می‌کند تا گراف وب را تحت تأثیر قرار دهد؛ اما در سیستم‌های توصیه‌گر یک کاربر و به عبارتی یک حمله‌گر تنها قادر به ایجاد پروفایل است. در این حالت اعمال یک ساختار خاص در گراف آیتم‌ها ساده نخواهد بود. از آنجا که در گراف آیتم‌ها برای ایجاد یک حمله با ساختار مشابه با مزرعه پیوند از ارتباط بین آیتم‌ها استفاده می‌شود، ساختار ایجادشده **مزرعه ارتباط**[۲۷] نامیده شد. در ادامه نحوه ایجاد پروفایل‌ها به منظور تولید مزرعه ارتباط و حمله به گراف آیتم‌ها به‌تفصیل شرح داده خواهد شد.

در سیستم‌های توصیه‌گر معمول مانند آمازون کاربرها نقشی در افزایش آیتم‌های سیستم ندارند. بنابراین برای حمله به گراف آیتم‌ها با ساخت پروفایل جعلی گره جدیدی به گراف اضافه نخواهد شد. تنها تغییری که می‌توان در گراف آیتم‌ها ایجاد کرد اضافه کردن یال بین گره‌های گراف است که این امر تنها با رتبه‌دهی خاص در پروفایل‌های جعلی صورت می‌گیرد.

بنابراین برای حمله به الگوریتم‌هایی که از گراف آیتم‌ها استفاده می‌کنند، ساختار مزرعه ارتباط تنها با ایجاد یال بین گره‌های موجود ایجاد می‌شود. برای ایجاد یک یال بین دو آیتم در گراف، کافی است حداقل در یک پروفایل به هر دو آیتم رتبه داده شود.

لازم به ذکر است که برخلاف مدل‌های قدیمی حمله، در حمله به الگوریتم‌هایی که از گراف آیتم‌ها استفاده می‌کنند، مقدار رتبه‌ها در پروفایل‌های جعلی مهم نیست، زیرا در این روش‌ها پیشنهادها تنها با استفاده از ارتباط بین آیتم‌ها در گراف تولید می‌شوند و به عبارت دیگر مقدار رتبه داده شده به آیتم‌ها تاثیری بر روند تولید پیشنهاد ندارند[۹]. هر چه تعداد پروفایل‌های مشترک بین دو آیتم بیشتر باشد ارتباط نزدیک‌تری بین دو آیتم برقرار است و همچنین وزن یال بین دو گره در گراف بیشتر خواهد بود. تعداد کاربرهای مشترک بین دو آیتم NCP[۲۸] نامیده می‌شود. شکل ۳ نحوه ایجاد یال بین دو آیتم در گراف آیتم‌ها را نمایش می‌دهد.

برای ایجاد یک مزرعه ارتباط در گراف آیتم‌ها، سه دسته از آیتم‌های موجود در سیستم انتخاب می‌شوند. آیتم هدف، آیتم‌های کمکی و آیتم‌های ربوده‌شده.



شکل ۳: نحوه ایجاد یال بین دو آیتم در گراف آیتم‌ها

- آیتم هدف (i_t): هدف از حمله، افزایش امتیاز آیتم هدف است. ساختار آیتم‌های مزرعه ارتباط باید به‌گونه‌ای ایجاد شود که باعث افزایش امتیاز آیتم هدف شوند.
- آیتم‌های کمکی (I_b): از آیتم‌های کمکی به منظور افزایش امتیاز آیتم هدف استفاده می‌شود. ساختار مجموعه آیتم‌های کمکی را مشابه با مدل‌های ذکرشده در شکل ۱ به سه مدل می‌توان ایجاد کرد. در مدل فردی، بین آیتم هدف و هر آیتم کمکی در گراف یک یال ایجاد می‌شود. در مدل چرخشی علاوه بر ایجاد یال بین هر آیتم کمکی با آیتم هدف، بین آیتم‌های کمکی نیز به صورت چرخشی یال ایجاد می‌شود. در مدل گراف کامل نیز علاوه بر ایجاد یال بین هر آیتم کمکی با آیتم هدف، بین تمامی آیتم‌های کمکی دوبه‌دو یال ایجاد می‌شود. تعداد آیتم‌های کمکی با نماد NB[۲۹] نمایش داده می‌شوند.
- آیتم‌های برتر (I_h): آیتم‌های برتر نقشی مشابه با صفحه‌های ربوده‌شده در مزرعه پیوند دارند. این آیتم‌ها از میان آیتم‌هایی انتخاب می‌شوند که بیشترین رتبه را در سیستم دریافت کرده‌اند. یافتن این نوع آیتم‌ها به‌راحتی امکان‌پذیر است زیرا اغلب سیستم‌های توصیه‌گر میانگین رتبه هر آیتم را با تعداد

Algorithm 1. Creating an Edge between two items in the Items Graph *(createEdge)*

Input:
v_{src},v_{tar} *// source and target items for edge*
NCP *// number of common profiles*
Output:
FakeProfiles *// fake profiles for created edge*
Procedure*:*
1. FakeProfiles=Ø
// add fake profiles p one by one
2. **For** p = (p_1 , p_2 , …,p_{NCP}) **do**
3. r (p,v_{src}) = r_{random} *// assign random rate*
4. r (p,v_{tar}) = r_{random} *// assign random rate*
// append profile (p) to set of fake profiles
5. FakeProfiles⟵ FakeProfiles ∪ p
6. **Endfor**

الگوریتم ۱. نحوه تولید یک یال بین دو آیتم در گراف آیتم‌ها

Algorithm 2. LinkFarm Attack (Cycle type)

Input:
RM *// rating matrix of system*
NCP,NH,NB,t *//attack parameters*
Output:
AttRM *// attacked rating matrix*
Procedure*:*
1. fp= Ø; *//fake profiles*
//select random boosting and popular hijacked Items
2. Select B , H
3. **For** b_i = (b_1 , b_2 , …,b_{NB-1})∈B **do**
// add edges inside boosting items
4. fp ⟵ fp ∪ *createEdge* (b_i , b_{i+1}, NCP)
5. **Endfor**
6. fp ⟵ fp ∪ *createEdge* (b_{NB} , b_1) *// add Cycle*
7. **For** b_i = (b_1 , b_2 , …,b_{NB})∈B **do**
// add edges from boosting set to target item
8. fp ⟵ fp ∪ *createEdge* (b_i ,t , NCP)
// add edges to from boosting set to Hijacked items
9. fp ⟵ fp ∪ *createEdge* (b_i ,h_i , NCP)
10. **Endfor**
11. AttRM⟵RM ∪ fp

الگوریتم ۲. حمله مزرعه ارتباط، مدل چرخشی

رتبه دهندگان به آن آیتم مشخص می‌کنند. همچنین اغلب آیتم‌های محبوب معمولاً تعداد رتبه‌دهی بالایی نیز دارند. بنابراین حمله‌گر برای تولید حمله مزرعه ارتباط در الگوریتم‌هایی که از گراف آیتم‌ها استفاده می‌کنند، نیاز به دانش خاصی ندارد. تعداد آیتم‌های برتر با نماد NH[۳۰] نمایش

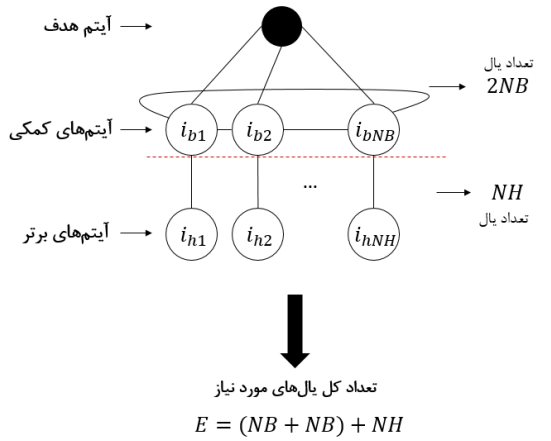
داده می‌شوند. در مزرعه ارتباط بین این آیتم‌ها و آیتم‌های کمکی یال برقرار می‌شود.

پس از تعیین انواع آیتم‌ها، نوبت به تولید پروفایل‌های جعلی بر اساس ساختار مدل موردنظر می‌رسد. الگوریتم ۱ شبه کد نحوه ایجاد پروفایل‌های جعلی را برای ایجاد یک یال در ساختار مزرعه ارتباط نمایش می‌دهد. برای ایجاد هر یال در مدل‌های مختلف مستقیم فردی، چرخشی و گراف کامل، این الگوریتم اجرا می‌شود و نهایتاً پروفایل‌های جعلی حاصل به ماتریس رتبه‌دهی کل سیستم اضافه می‌شوند. برای نمونه، از میان سه مدل معرفی شده، شبه کد مدل حمله چرخشی در الگوریتم ۲ نمایش داده شده است.

نحوه محاسبه تعداد یال‌های موردنیاز (E) به منظور ایجاد مدل‌های مختلف مزرعه ارتباط در جدول ۱ نمایش داده شده است. برای نمونه شکل ۴ نحوه محاسبه تعداد یال‌های مورد نیاز در ساختار چرخشی را نمایش می‌دهد. با توجه به تعداد یال‌ها در مزرعه ارتباط تعداد پروفایل‌های جعلی موردنیاز برای اعمال حمله مزرعه پیوند برابر با $E \times NCP$ است.

جدول ۱: نحوه محاسبه تعداد یال‌های موردنیاز به منظور ایجاد مزرعه ارتباطات در مدل‌های مختلف

مدل مزرعه ارتباطات	تعداد یال‌های موردنیاز (E)
مستقیم فردی	$E = NH + NB$
چرخشی	$E = NH + 2NB$
گراف کامل	$E = NH + NB + \frac{NB(NB-1)}{2}$

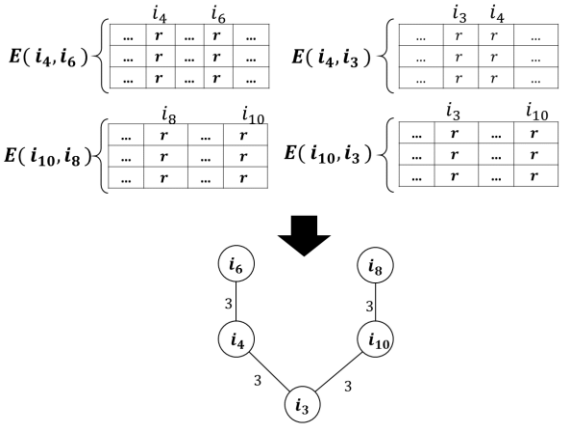


شکل ۴: نحوه محاسبه تعداد یال‌های مورد نیاز در ساختار چرخشی

مثال ۱. به منظور ایجاد یک حمله مزرعه ارتباط با مدل مستقیم فردی در سیستمی با پارامترهای

$$\begin{cases} NCP = 3, NB = 2, NH = 2 \\ I = [i_1, \dots, i_{10}], I_h = [i_6, i_8], I_b = [i_4, i_{10}], i_t = i_3 \end{cases}$$

که در آن NCP، NB و NH به ترتیب برابراند با تعداد کاربرهای مشترک بین دو آیتم، تعداد آیتم‌های کمکی و تعداد آیتم‌های برتر. همچنین مجموعه‌های I_h و I_b به ترتیب بیانگر مجموعه آیتم‌های برتر و آیتم‌های کمکی و در نهایت i_t آیتم هدف حمله است. ساخت پروفایل‌های جعلی بر طبق شکل ۵ صورت می‌گیرد. به ازای هر یال در چنین ساختاری ۳ پروفایل جعلی ایجاد می‌شود و طبق جدول ۱، جمعاً ۱۲ پروفایل جعلی نیاز است. در واقع برای ایجاد چنین ساختاری، الگوریتم ۱، ۱۲ مرتبه فراخوانی شده و درنهایت پروفایل‌های جعلی ایجاد شده به ازای هر ۱۲ اجرا، به عنوان حمله به سیستم اضافه می‌شوند (E×NCP=(NH+NB)×NCP=(2+2)×3=12).



شکل ۵: مثالی از نحوه ایجاد ساختار مزرعه ارتباطات در گراف

۵- ارزیابی روش پیشنهادی

در این بخش دادگان و معیارهای استفاده‌شده به منظور ارزیابی روش حمله پیشنهادی معرفی می‌شوند. الگوریتم امتیاز آیتم (IR) [۹]، الگوریتم پایه در زمینه تکنیک‌های مبتنی بر گرافی است که از گراف آیتم‌ها استفاده می‌کنند. بدین منظور این الگوریتم از میان الگوریتم‌های مبتنی بر گراف انتخاب و پیاده‌سازی شده است. لازم به ذکر است که کلیه پیاده‌سازی‌های مربوط به الگوریتم پایه و روش‌های پیشنهادی، در نرم‌افزار متلب صورت گرفته است.

۵-۱- نحوه ارجاع به منبع‌های مورد استفاده

به‌منظور ارزیابی الگوریتم ارائه شده و مقایسه با روش‌های دیگر، مجموعه داده‌های MovieLens [۳۲] و FilmTrust [۳۳] مورد

استفاده قرار گرفت. مجموعه داده MovieLens از پرکاربردترین مجموعه داده‌های موجود در حوزه سیستم‌های توصیه‌گر است. این مجموعه شامل ۱۰۰،۰۰۰ رتبه است که توسط ۹۴۳ کاربر به ۱۶۸۲ فیلم داده‌شده‌اند. مجوعه داده FilmTrust شامل ۱،۵۰۸ کاربر، ۲،۰۷۱ فیلم و ۳۵،۴۹۷ رتبه است.

۵-۲- معیار ارزیابی

هدف ارزیابی امنیت سیستم‌های توصیه‌گر، محاسبه کارایی خام نیست، بلکه محاسبه میزان تغییراتی است که درنتیجه یک حمله به کارایی تحمیل شده است. نتیجه مطلوب یک حمله این است که آیتم هدف با احتمال بیشتری توسط سیستم توصیه شود. یک روش برای اندازه‌گیری این احتمال، محاسبه میزان تغییر در رتبه پیش‌بینی شده است. اما در تکنیک‌های مبتنی بر گراف پیش‌بینی مقدار رتبه صورت نمی‌گیرد. برای مثال، در روش IR به ازای هر کاربر، لیست امتیازبندی شده‌ای از تمامی آیتم‌ها ایجاد می‌کند و سپس از میان آیتم‌هایی که توسط کاربر مشاهده نشده‌اند، آیتم‌هایی با بالاترین امتیازها را برای توصیه به کاربر انتخاب می‌کند.

معیارهای مختلفی ازجمله تای کندال[۳۱] و رُو اسپیرمن[۳۲] [۳۴]، برای ارزیابی لیست‌های امتیازبندی شده وجود دارند که تفاوت بین دو لیست امتیازبندی شده را محاسبه می‌کنند. اما در ارزیابی یک حمله، هدف محاسبه میزان تغییر به ازای یک آیتم خاص است نه تغییر کل لیست امتیازبندی شده؛ بنابراین استفاده از چنین معیارهایی راه‌حل مناسبی برای ارزیابی تأثیر حمله نیست. ازاین‌رو، در این پژوهش معیار **تغییر امتیاز**[۳۳] ارائه شد.

میزان تغییر امتیاز، برابر است با تغییر مکان آیتم هدف در لیست امتیازبندی شده، قبل و بعد از حمله. اگر I_T و U_T به ترتیب مجموعه آیتم‌ها و کاربرهای هدف باشند، آنگاه برای هر جفت کاربر و آیتم (u,i) تغییر امتیاز طبق معادله (۷) محاسبه می‌شود.

$$\Delta_{u,i} = R_{u,i} - \acute{R}_{u,i} \qquad (۷)$$

R و $\acute{R}$ به ترتیب بیانگر مکان آیتم هدف در لیست، قبل و بعد از حمله می‌باشند. اگر مقدار به‌دست‌آمده از تغییر امتیاز مثبت باشد، به این معنی است که آیتم هدف بعد از حمله به ابتدای لیست امتیازبندی شده تغییر مکان داشته است و درنتیجه حمله بر آیتم هدف موفقیت‌آمیز بوده است. میانگین تغییر امتیاز آیتم i برای تمام کاربرها طبق معادله (۸) محاسبه می‌شود. به‌طور مشابه، محاسبه میانگین تغییر امتیاز برای تمام آیتم‌هایی که مورد حمله قرارگرفته‌اند، طبق معادله (۹) محاسبه می‌شود.

$$\Delta_i = \frac{\sum_{u \in U_T} \Delta_{u,i}}{|U_T|} \qquad (۸)$$

$$\overline{\Delta} = \frac{\sum_{i \in I_T} \Delta_i}{|I_T|} \qquad (۹)$$

معیار ارائه‌شده میزان جابه‌جایی یک آیتم را پس از حمله در لیست امتیازبندی شده مشخص می‌کند. ممکن است این سؤال مطرح شود که چرا معیار تغییر امتیاز، تفاوت امتیاز یک آیتم را قبل و بعد از حمله محاسبه نمی‌کند و تنها به میزان تغییر مکان آیتم توجه می‌کند؟ پاسخ این سؤال را می‌توان این‌گونه بیان کرد که در روش‌هایی که لیست امتیازبندی شده‌ای از آیتم‌ها را با قدم زنی تصادفی در گراف ارائه می‌دهند، ازآنجایی‌که امتیاز آیتم‌ها با انتشار امتیاز در گراف محاسبه می‌شوند، تفاوت امتیاز بین برخی آیتم‌ها ممکن است بسیار ناچیز باشد و با محاسبه تفاوت امتیاز یک آیتم قبل و بعد از حمله میزان اثرگذاری حمله مشخص نشود. در چنین روش‌هایی میزان اثرگذاری یک حمله با تغییر مکان یک آیتم در لیست امتیازبندی شده آیتم‌ها نمایان‌تر است.

۵-۳- جزییات تست‌ها و حمله‌ها

برای انجام آزمایش‌ها ابتدا ۵۰ کاربر و ۲۵ آیتم که در دادگان انتخابی از نوع فیلم هستند، به عنوان هدف آزمایش‌ها انتخاب‌شدند. ۲۵ فیلم هدف به صورت تصادفی و از میان فیلم‌هایی انتخاب شدند که میانگین رتبه‌دهی آن‌ها کمتر از میانگین رتبه‌دهی کل آیتم‌ها در سیستم است. هدف از حمله، افزایش احتمال پیشنهاد شدن هر یک از این آیتم‌ها است. در اغلب الگوریتم‌های توصیه‌گر که از داده‌های رتبه‌دهی استفاده می‌کنند، رفتار رتبه‌دهی کاربرها در تولید توصیه‌های ارائه‌شده برای آن‌ها تأثیرگذار است. اما کاربرهایی وجود دارند که رفتار رتبه‌دهی عادی ندارند. ازاین‌رو در برخی پژوهش‌های [۲۱]، [۲۲]، [۲۴] مربوط به حمله در سیستم‌های توصیه‌گر، میزان اثرگذاری حمله بر روی کاربرهایی بررسی شده‌است که رفتار رتبه‌دهی نرمالی در سیستم دارند، یعنی میانگین رتبه‌دهی آن‌ها در توزیع نرمال میانگین رتبه‌دهی کل کاربرهای سیستم وجود داشته باشد. بنابراین ۵۰ کاربر تست به صورت تصادفی و از بین کاربرهایی انتخاب‌شده‌اند که اولاً به هیچ یک از ۲۵ فیلم هدف رتبه نداده باشند، ثانیاً میانگین رتبه‌دهی آن‌ها در توزیع نرمال میانگین رتبه‌دهی کل کاربرهای سیستم باشد.

پس از انتخاب کاربرهای تست و فیلم‌های هدف، هر فیلم به صورت جداگانه مورد حمله قرار می‌گیرد. مسئله ارزیابی یک حمله به این صورت است که میزان تغییر امتیاز برای هر کاربر چه مقدار بوده است. بنابراین به ازای حمله به هر آیتم هدف، میزان این

تغییرات برای تمامی کاربرهای تست محاسبه می‌شود. نتایج ارائه‌شده میانگین تغییرات به ازای حمله به تمامی آیتم‌های هدف بوده است.

علاوه بر مدل حمله پیشنهادی، از میان مدل‌های حمله اولیه که پیش‌تر معرفی شدند، سه مدل حمله تصادفی، میانگین و باندواگن که به منظور مقایسه میزان تأثیرگذاری مدل‌های مختلف حمله در اکثر پژوهش‌ها استفاده می‌شوند نیز به الگوریتم IR اعمال‌شده‌اند. نحوه رتبه‌دهی در پروفایل‌های جعلی به ازای هر مدل حمله اولیه، مشابه توضیحات ارائه‌شده در بخش ۳-۱-۲ صورت گرفته است. برای انتخاب مجموعه آیتم انتخابی در مدل حمله باندواگن ۵ فیلم از میان فیلم‌هایی که بیشترین میانگین رتبه‌دهی را داشته‌اند انتخاب شدند. همچنین این حمله‌ها در ۵ اندازه پروفایل ۱، ۳، ۵، ۱۰ و ۲۰ درصد و ۳ اندازه حمله ۵، ۱۰ و ۱۵ درصد اعمال‌شده‌اند. انتخاب این مقادیر برای اندازه پروفایل و اندازه حمله بر اساس مقادیر انتخابی در اکثر پژوهش‌های موجود در زمینه حمله به سیستم‌های توصیه‌گر صورت گرفته است. در ادامه این بخش نحوه اعمال حمله‌ها و نتایج به‌دست‌آمده شرح داده می‌شود.

۵-۴- نتایج حمله‌های قدیمی

جداول ۲، ۳ و ۴ به ترتیب مقادیر به‌دست‌آمده از معیار تغییر امتیاز برای حمله‌های تصادفی، میانگین و باندواگن به این الگوریتم را نمایش می‌دهند. همان‌طور که مشاهده می‌شود، نتایج به‌دست‌آمده از حمله‌های تصادفی و میانگین کاملاً برخلاف هدف حمله است. یعنی میزان تغییر امتیاز مقدار منفی داشته است و این به معنای کاهش جایگاه آیتم هدف در لیست امتیازبندی شده کاربرهای تست است.

اما چنین نتیجه‌ای با توجه به ماهیت الگوریتم IR و همچنین مدل‌های حمله میانگین و تصادفی دور از انتظار نیست. چراکه در الگوریتم IR مقدار رتبه‌ها تأثیر چندانی در تولید پیشنهادها ندارد. از طرف دیگر در مدل‌های حمله تصادفی و میانگین، آیتم‌های متفاوتی به صورت تصادفی انتخاب و به همراه آیتم هدف در پروفایل‌های جعلی رتبه‌دهی می‌شوند، که نتیجه آن ایجاد یال‌های متعدد و ناهماهنگ در گراف آیتم‌ها است.

در این میان ازآنجایی‌که در تمامی پروفایل‌های جعلی ایجادشده به آیتم هدف رتبه‌دهی شده است، تعداد زیادی یال با وزن کم برای آیتم هدف در گراف ایجاد می‌شود و همین امر باعث کاهش امتیاز آیتم هدف در الگوریتم IR می‌شود.

اما همان‌طور که مشاهده می‌شود مدل حمله باندواگن نسبت به مدل‌های میانگین و تصادفی چندان بی‌تأثیر نبوده است. زیرا در این مدل حمله، در هر پروفایل جعلی علاوه بر آیتم هدف به گروهی از آیتم‌های محبوب نیز رتبه‌دهی می‌شود. ازآنجایی‌که آیتم‌های محبوب معمولاً تعداد رتبه بالایی در سیستم دریافت می‌کنند، بنابراین در گراف آیتم‌ها نیز اعتبار بالایی دریافت می‌کنند. درنتیجه این مدل حمله، باعث می‌شود که بین آیتم هدف و آیتم‌های محبوب یک یال با وزن بالا (تعداد پروفایل‌های جعلی) در گراف آیتم‌ها ایجاد شود و باعث افزایش امتیاز آیتم هدف در الگوریتم IR شود.

اما یک نکته وجود دارد، همان‌طور که در نتایج مشاهده می‌شود این حمله زمانی تأثیرگذار است که اندازه پروفایل‌های جعلی پایین است. زیرا با افزایش اندازه پروفایل یا به عبارتی افزایش تعداد آیتم‌های رتبه‌دهی شده در پروفایل‌های جعلی، تعداد یال‌های ناهماهنگ با وزن‌های کم نیز در گراف افزایش می‌یابد. همچنین مشابه با مدل‌های میانگین و تصادفی تعداد زیادی یال با وزن کم برای آیتم هدف در گراف ایجاد می‌شود و همین امر باعث کاهش امتیاز آیتم هدف در الگوریتم IR می‌شود.

جدول ۲: نتایج معیار تغییر امتیاز از اعمال حمله تصادفی به الگوریتم IR در دادگان MovieLens

اندازه حمله	اندازه پروفایل				
	۱٪	۳٪	۵٪	۱۰٪	۲۰٪
٪۵	-۲۵	-۶۰	-۷۰	-۸۱	-۵۴
٪۱۰	-۴۵	-۸۸	-۱۰۴	-۱۳۰	-۱۱۳
۱۵٪	-۶۲	-۱۴۱	-۱۸۹	-۲۳۹	-۱۶۹

جدول ۳: نتایج معیار تغییر امتیاز از اعمال حمله میانگین به الگوریتم IR در دادگان MovieLens

اندازه حمله	اندازه پروفایل				
	۱٪	۳٪	۵٪	۱۰٪	۲۰٪
٪۵	-۲۷	-۶۳	-۷۲	-۷۴	-۵۳
٪۱۰	-۳۹	-۸۲	-۱۰۳	-۱۳۲	-۱۰۶
۱۵٪	-۶۸	-۱۳۳	-۱۷۹	-۲۴۱	-۱۷۲

جدول ۴: نتایج معیار تغییر امتیاز از اعمال حمله باندواگن به الگوریتم IR در دادگان MovieLens

اندازه حمله	اندازه پروفایل				
	۱٪	۳٪	۵٪	۱۰٪	۲۰٪
٪۵	۶۷	۶	-۲۷	-۴۵	-۵
٪۱۰	۹۱	۱۲	-۳۷	-۶۰	-۵۶
۱۵٪	۱۳۸	۱۵	-۴۹	-۱۰۱	-۱۱۷

۵-۵- تنظیم پارامترهای مدل حمله پیشنهادی

برای اعمال حمله مزرعه ارتباط، تولید پروفایل‌های جعلی مشابه با توضیحات ارائه‌شده در بخش روش حمله پیشنهادی، در سه مدل مزرعه ارتباط چرخشی، کامل و فردی انجام شد. برای اعمال هر یک از این سه مدل، چهار نوع آیتم کمکی انتخاب شدند.

در نوع اول، آیتم‌های کمکی از بین آیتم‌های با تعداد رتبه‌دهی بالا انتخاب‌شده‌اند، یعنی آیتم‌هایی که بیشترین تعداد رتبه را در سیستم دریافت کرده‌اند.

در نوع دوم، آیتم‌های کمکی از بین آیتم‌های با تعداد رتبه‌دهی پایین انتخاب‌شده‌اند، یعنی آیتم‌هایی که کمترین تعداد رتبه را در سیستم دریافت کرده‌اند.

در نوع سوم، آیتم‌های کمکی از میان آیتم‌های انتخاب‌شده‌اند که تاکنون در سیستم رتبه‌ای دریافت نکرده‌اند و درنتیجه هیچ

جدول ۵: پارامترهای استفاده شده در حمله مزرعه ارتباطات

آیتم‌های			پارامترها		
کد	کمکی	ساختار حمله	NCP	NB	NH
{T1-Ind-B5H0}		فردی			
{T1-Cyc-B5H0}	نوع۱	چرخشی	۱۰	۵	۰
{T1-Com-B5H0}		کامل			
{T2-Ind-B5H5}		فردی			
{T2-Cyc-B5H5}	نوع۲	چرخشی	۱۰	۵	۵
{T2-Com-B5H5}		کامل			
{T3-Ind-B5H5}		فردی			
{T3-Cyc-B5H5}	نوع۳	چرخشی	۱۰	۵	۵
{T3-Com-B5H5}		کامل			

ارتباطی در گراف آیتم‌ها ندارند. ازآنجایی‌که در دادگان انتخابی هر فیلم حداقل از ۱ کاربر رتبه دریافت کرده است، بنابراین برای اجرای چنین حالتی، رتبه‌های آیتم‌های انتخاب‌شده در حالت دوم را حذف کرده و از همان آیتم‌ها برای حالت سوم استفاده شد.

جدول ۵ مشخصات و مقداردهی به پارامترهای تعداد پروفایل‌های مشترک (NCP)، تعداد آیتم‌های کمکی(NB) و تعداد آیتم‌های ربوده‌شده (NH) را به ازای حمله‌های مزرعه ارتباط که به الگوریتم IR اعمال شده است، نمایش می‌دهد.

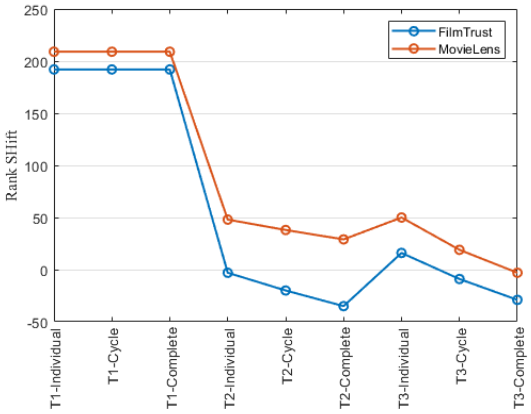
به هر حمله یک کد اختصاص داده شده که در جداول از آن‌ها استفاده می‌شود. همان‌طور که در جدول ۵ مشاهده می‌شود، در حالتی که آیتم‌های کمکی جزو آیتم‌های نوع ۱هستند تعداد آیتم‌های ربوده‌شده صفر است. دلیل این انتخاب این است که آیتم‌های کمکی نوع ۲ و نوع ۳ تعداد رتبه‌دهی بالایی در سیستم ندارند، بنابراین برای دریافت میزان اعتماد بالا نیاز به ایجاد ارتباط با یک آیتم قوی دارند. اما ازآنجایی‌که آیتم‌های کمکی نوع ۱ بیشترین تعداد رتبه‌دهی را در سیستم دارند بنابراین ارتباط‌های زیادی در گراف دارند و درواقع نیاز به آیتم‌های ربوده‌شده ندارند.

همچنین حداکثر مقدار پارامتر NCP برابر با ۱۰ در نظر گرفته شد. دلیل این انتخاب این بوده است که برای مثال در یک حمله با مدل گراف کامل که تعداد آیتم‌های کمکی و آیتم‌های ربوده‌شده برابر با ۵ بوده است، بر طبق جدول ۱ جمعاً ۲۰ یال نیاز است. اگر مقدار NCP برابر با ۱۰ در نظر گرفته شود، یعنی به ازای هر یال ۱۰ پروفایل ایجاد می‌شود، درنتیجه برای چنین حمله‌ای ۲۰۰=۲۰×۱۰ پروفایل نیاز است! برای مثال در مجموعه داده مووی لنز که شامل ۹۴۳ کاربر است، ایجاد ۲۰۰ پروفایل جعلی یعنی حمله‌ای با اندازه ۲۱ درصد اعمال‌شده است. در این حالت افزایش مقدار پارامتر NCP اندازه حمله را افزایش می‌دهد که هزینه و تلاش زیادی نیاز خواهد داشت، از طرف دیگر ازآنجایی‌که این پارامتر مشخص‌کننده وزن یال بین دو گره از گراف است، هر چه مقدار کمتری داشته باشد میزان اعتبار کمتری به سمت آیتم هدف منتشر خواهد شد و درنتیجه کاهش مقدار پارامتر NCP اثر حمله را نیز کاهش می‌دهد. همچنین تعداد آیتم‌های کمکی و به عبارتی تعداد آیتم‌های برتر ۵ در نظر گرفته شد تا به‌نوعی مشابه با ۵ آیتم انتخابی در مدل باندواگن و قابل‌مقایسه باشد. جدول ۶، اندازه حمله به دو مجموعه داده انتخاب شده را به ازای هر کد حمله نمایش می‌دهد.

۵-۶- نتایج روش حمله پیشنهادی

هر یک از حملات با پارامترهای مشخص شده در جدول ۵ صورت جداگانه به آیتم‌های هدف در هر دادگان اعمال شده‌اند. نتایج معیار تغییر امتیاز از اعمال حمله مزرعه ارتباط به الگوریتم IR در جدول ۶ و شکل ۵ نمایش داده شده است. همان‌طور که مشاهده می‌شود انتخاب آیتم‌های کمکی از میان آیتم‌های نوع ۱ بیشترین تأثیرگذاری را بر الگوریتم IR داشته است؛ زیرا در این حالت بین آیتم هدف و آیتم‌های برتر مستقیماً یال برقرار می‌شود و به‌نوعی اعتبار آیتم‌های برتر بدون واسطه به سمت آیتم هدف منتشر می‌شود. اما در دیگر حالت‌ها که آیتم‌های برتر به‌عنوان آیتم‌های ربوده‌شده و با واسطه آیتم‌های کمکی به آیتم هدف متصل می‌شوند، میزان اعتبار انتشاریافته کمتر است و درنتیجه تأثیر حمله نیز قابل‌توجه نیست. بنابراین می‌توان نتیجه گرفت که استفاده از آیتم‌های کمکی ضعیف (نوع ۲ و نوع ۳)، به‌منظور ایجاد یک ساختار مشابه با صفحه‌های کمکی در مزرعه پیوند، برای حمله مزرعه ارتباط مناسب نیست.

از طرفی همان‌طور که در شکل ۵ مشاهده می‌شود، ساختار مستقیم فردی در مقایسه با دیگر مدل‌ها در هر دو مجموعه داده عملکرد بهتری داشته است زیرا در این حالت علاوه بر ایجاد پروفایل‌های جعلی کمتر، ارتباط‌های بین آیتم‌های کمکی نیز کم‌تر است و به‌نوعی میزان اعتبار مستقیماً به سمت آیتم هدف انتشار می‌یابد. به عبارتی با توجه به عملکرد الگوریتم رتبه آیتم که در بخش ۳ ارائه شد، در ساختارهای چرخشی و گراف کامل، میزان امتیاز انتشار یافته از سمت آیتم‌های ربوده شده به سمت آیتم‌های کمکی (به دلیل وجود یال‌های متعدد مابین آیتم‌های کمکی) تقسیم شده و بین یال‌ها منتشر می‌شود و درنهایت امتیاز منتشر شده به سمت آیتم هدف تاثیر کمتری می‌گذارد. در حالی که در ساختار مستقیم فردی از آنجایی که بین آیتم‌های کمکی یال ایجاد نمی‌شود، در نتیجه امتیاز انتشار یافته از آیتم‌های ربوده شده به سمت آیتم هدف در میان راه بین یال‌های متعدد تقسیم



شکل ۵: میزان تغییر امتیاز در الگوریتم IR پس از اعمال حمله مزرعه ارتباط با پارامترهای تعیین شده

نمی‌شود. به همین دلیل در ساختار مستقیم فردی میزان امتیاز آیتم هدف به صورت قابل‌توجهی بیشتر می‌شود.

درنتیجه طبق نتایج به‌دست‌آمده از حمله مزرعه ارتباط با ساختارهای مختلف، می‌توان دریافت که مدل فردی مستقیم روشی مناسب برای حمله به الگوریتم‌های مبتنی بر گراف آیتم‌ها است. همان‌طور که در شکل ۵ مشاهده می‌شود، مدل حمله ارائه شده بر دو مجموعه داده با تعداد آیتم و کاربر و پراکندگی مختلف تاثیر گذار بوده است. همچنین لازم به ذکر است که این مدل حمله اثرگذاری قابل توجهی حتی در اندازه‌های پایین حمله دارد (۳٪ و ۵٪) که افزایش پارامترهای NCP و NB و افزایش اندازه حمله تأثیرگذاری بیشتری نیز به همراه خواهد داشت.

این باور وجود دارد که در یک حمله با افزایش اندازه حمله، میزان اثرگذاری نیز بیشتر می‌شود، درنتیجه ممکن است این سؤال مطرح شود که چرا در مدل مزرعه ارتباط با توجه به افزایش اندازه حمله در هر دو مجموعه داده میزان اثرگذاری کاهش‌یافته است؟ پاسخ این سؤال را می‌توان این‌گونه بیان کرد که در الگوریتم IR میزان اثرگذاری حمله‌ها بستگی به ارتباط بین آیتم‌ها دارد و هر چه ساختار ارتباط بین آیتم هدف با آیتم‌های برتر پیچیده‌تر باشد، اثر حمله نیز کاهش می‌یابد.

۵-۷- مقایسه نتایج حمله‌های قدیمی با حمله پیشنهادی

به‌منظور مقایسه بهتر مدل مزرعه ارتباط با مدل‌های حمله قدیمی، نتایج به‌دست‌آمده از حمله‌های قدیمی با اندازه حمله ٪۵ و اندازه پروفایل ٪۱، به همراه نتایج به‌دست‌آمده از حمله مزرعه ارتباطات بهینه با کد T1-Ind-B5H0 و اندازه حمله ٪۵ در جدول ۷ نمایش داده‌شده‌اند. همان‌طور که مشاهده می‌شود، حمله مزرعه ارتباط نسبت به حمله‌های قدیمی نتایج قابل‌توجهی دارد. لازم به ذکر است که با توجه به جدول ۷، حمله باندواگن حتی با اندازه حمله ٪۱۵ نیز نسبت به مدل حمله مزرعه ارتباط با اندازه حمله ٪۵ ضعیف‌تر عمل کرده است.

در نهایت می‌توان نتیجه گرفت که الگوریتم IR در مقابل

جدول ۶: نتایج معیار تغییر امتیاز از حمله مزرعه ارتباطات با پارامترهای مختلف به الگوریتم IR

کد حمله	MovieLens		FilmTrust	
	اندازه حمله	تغییر امتیاز	اندازه حمله	تغییر امتیاز
{T1-Ind-B5H0}	5%	209	3%	192
{T1-Cyc-B5H0}	10%	209	7%	192
{T1-Com-B5H0}	15%	209	10%	192
{T2-Ind-B5H5}	10%	48	7%	-3
{T2-Cyc-B5H5}	15%	38	10%	-20
{T2-Com-B5H5}	21%	29	13%	-35
{T3-Ind-B5H5}	10%	50	7%	16
{T3-Cyc-B5H5}	15%	19	10%	-9
{T3-Com-B5H5}	21%	-3	13%	-29

حمله‌هایی که گراف ارتباطات را هدف قرار می‌دهند آسیب‌پذیر است. این مسأله با توجه به نتایج به دست‌آمده از اعمال حمله مزرعه ارتباط ارائه شده در این مقاله، قابل مشاهده است.

جدول ۷: مقایسه نتایج به‌دست‌آمده از مدل‌های حمله قدیمی و مدل حمله مزرعه ارتباطات به الگوریتم IR

مدل حمله	مزرعه ارتباطات	باندواگن	میانگین	تصادفی
میانگین تغییر امتیاز	۲۰۹	۶۷	-۲۷	-۲۵

۶- نتیجه‌گیری و پژوهش‌های آینده

علی‌رغم پژوهش‌های گسترده‌ای که درزمینه حمله در سیستم‌های توصیه‌گر صورت گرفته است، تعداد محدودی آسیب‌پذیری الگوریتم‌های توصیه‌گر مبتنی بر گراف را در مقابل حمله‌های مختلف مورد بررسی قرار داده‌اند. ازاین‌رو، در این مقاله روش جدیدی برای حمله به ساختار گراف آیتم‌ها در یک سیستم توصیه‌گر ارائه شد. گروهی از روش‌های توصیه‌گر مبتنی بر گراف وجود دارند که از ایده الگوریتم محبوب رتبه‌بندی PR بهره گرفته و به تولید پیشنهادها می‌پردازند. برای حمله به این مدل الگوریتم‌ها از ایده حمله‌های هرزه‌نگاری در وب بهره برده و مدل حمله مزرعه ارتباط ارائه شد. درنهایت مدل حمله ارائه‌شده به همراه تعدادی از مدل‌های حمله قدیمی به منظور بررسی میزان آسیب‌پذیری سیستم‌های توصیه‌گری که با تحلیل گراف آیتم‌ها به تولید لیست پیشنهادهایشان می‌پردازند، مورد بررسی قرار گرفت.

نتایج به‌دست‌آمده از اعمال مدل‌های حمله قدیمی و حمله مزرعه ارتباط به الگوریتم IR، نشان دادند که این الگوریتم در مقابل حمله‌های بررسی شده در این مقاله آسیب‌پذیر است. به‌خصوص، حمله مزرعه ارتباط که روند تولید پیشنهادها در این الگوریتم را به صورت قابل‌توجهی نسبت به مدل‌های حمله قدیمی، تحت تأثیر قرار می‌دهد. به‌طورکلی روش‌هایی که در آن‌ها رتبه‌دهی هر کاربر مستقیماً بر توصیه‌های ارائه‌شده برای آن کاربر تأثیرگذار است، در مقابل حمله مقاوم‌تر خواهند بود. زیرا در این روش‌ها هرچقدر هم که حمله‌گر از یک استراتژی قوی استفاده کند، نهایتاً نمی‌تواند بر رتبه‌های پیشین ارائه‌شده توسط کاربرها تأثیری بگذارد.

مدل حمله ارائه‌شده از مجموعه حمله‌هایی است که هدفشان افزایش مجبوبیت یک آیتم خاص است. بنابراین در ادامه می‌توان بر روی ایجاد مدل‌های حمله‌ای تمرکز کرد که هدفشان کاهش محبوبیت یک آیتم خاص باشد. مدل حمله پیشنهادشده تنها در الگوریتم‌هایی کاربرد دارد که از ایده رتبه‌بندی PR بر گراف آیتم‌ها استفاده می‌کنند. درنتیجه بررسی میزان آسیب‌پذیری دیگر الگوریتم‌های مبتنی بر گراف آیتم‌ها که بر مبنای ایده رتبه‌بندی PR نیستند، می‌تواند پیشنهادی برای پژوهش‌های بیشتر در این زمینه باشد. با توجه به آسیب‌پذیری الگوریتم مبتنی بر گراف IR، تمرکز پژوهش‌های آینده می‌تواند بر نحوه کشف انواع حمله در این الگوریتم‌ها باشد. همچنین از آنجا که امروزه ظهور داده‌های کلان چالش مهمی در اکثر سرویس‌های اینترنتی است، سنجش عملکرد مدل‌های حمله مختلف در سیستم‌های توصیه‌گر مبتنی بر گراف که با داده‌های عظیم روبه‌رو هستند [۳۵]، می‌تواند در پژوهش‌های آینده مورد بررسی قرارگیرد.

مراجع

[1] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, ‘Collaborative filtering recommender systems’, Found. Trends® Human–Computer Interact., vol. 4, no. 2, pp. 81–173, 2011.

[2] 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. 734–749, 2005.

[3] P. Lops, M. De Gemmis, and G. Semeraro, ‘Content-based recommender systems: State of the art and trends’, in Recommender systems handbook, Springer, 2011, pp. 73–105.

[4] Z. Huang, H. Chen, and D. Zeng, ‘Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering’, ACM Trans. Inf. Syst., vol. 22, no. 1, pp. 116–142, 2004.

[5] X. Ma and R. Wang, ‘Personalized Scientific Paper Recommendation Based on Heterogeneous Graph Representation’, IEEE Access, vol. 7, pp. 79887–79894, 2019.

[6] L. Zhang, J. Xu, and C. Li, ‘A random-walk based recommendation algorithm considering item categories’, Neurocomputing, vol. 120, pp. 391–396, 2013.

[7] Z. Jiang, H. Liu, B. Fu, Z. Wu, and T. Zhang, ‘Recommendation in heterogeneous information networks based on generalized random walk model and bayesian personalized ranking’, in Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 288–296.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, ‘The pagerank citation ranking: Bringing order to the web’, Stanford InfoLab, 1999.

[9] M. Gori, A. Pucci, V. Roma, and I. Siena, ‘Itemrank: A random-walk based scoring algorithm for recommender engines’, in IJCAI, 2007, vol. 7, pp. 2766–2771.

[10] P. Covington, J. Adams, and E. Sargin, ‘Deep neural networks for youtube recommendations’, in Proceedings of the 10th ACM conference on recommender systems, 2016, pp. 191–198.

[11] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, ‘Graph convolutional neural networks for web-scale recommender systems’, in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 974–983.

[12] S. K. Lam and J. Riedl, ‘Shilling recommender systems for fun and profit’, in Proceedings of the 13th international conference on World Wide Web, 2004, pp. 393–402.

[13] L. Chen, Y. Xu, F. Xie, M. Huang, and Z. Zheng, ‘Data poisoning attacks on neighborhood-based recommender systems’, Trans. Emerg. Telecommun. Technol., p. e3872, 2019.

[14] I. Gunes, C. Kaleli, A. Bilge, and H. Polat, 'Shilling attacks against recommender systems: a comprehensive survey', Artif. Intell. Rev., vol. 42, no. 4, pp. 767–799, 2014.

[15] H. Zhang et al., 'Data poisoning attack against knowledge graph embedding', in Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 4853–4859.

[16] G. Yang, N. Z. Gong, and Y. Cai, 'Fake Co-visitation Injection Attacks to Recommender Systems', in NDSS, 2017.

[17] R. Hu, Y. Guo, M. Pan, and Y. Gong, 'Targeted poisoning attacks on social recommender systems', in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1–6.

[18] M. Fang, G. Yang, N. Z. Gong, and J. Liu, 'Poisoning attacks to graph-based recommender systems', in Proceedings of the 34th Annual Computer Security Applications Conference, 2018, pp. 381–392.

[19] Z. Gyöngyi and H. Garcia-Molina, 'Link spam alliances', in Proceedings of the 31st international conference on Very large data bases, 2005, pp. 517–528.

[20] M. Gao, Z. Wu, and F. Jiang, 'Userrank for item-based collaborative filtering recommendation', Inf. Process. Lett., vol. 111, no. 9, pp. 440–446, 2011.

[21] B. Mobasher, R. Burke, R. Bhaumik, and C. Williams, 'Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness', ACM Trans. Internet Technol., vol. 7, no. 4, pp. 23-es, 2007.

[22] R. Burke, B. Mobasher, and R. Bhaumik, 'Limited knowledge shilling attacks in collaborative filtering systems', in Proceedings of 3rd International Workshop on Intelligent Techniques for Web Personalization (ITWP 2005), 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), 2005, pp. 17–24.

[23] M. P. O'Mahony, N. J. Hurley, and G. C. M. Silvestre, 'Recommender systems: Attack types and strategies', in AAAI, 2005, pp. 334–339.

[24] R. Burke, B. Mobasher, R. Bhaumik, and C. Williams, 'Segment-based injection attacks against collaborative filtering recommender systems', in Fifth IEEE International Conference on Data Mining (ICDM'05), 2005, p. 4 pp.

[25] F. Zhang, 'Analysis of bandwagon and average hybrid attack model against trust-based recommender systems', in 2011 Fifth International Conference on Management of e-Commerce and e-Government, 2011, pp. 269–273.

[26] C. Williams, B. Mobasher, R. Burke, J. Sandvig, and R. Bhaumik, 'Detection of obfuscated attacks in collaborative recommender systems', in Proceedings of the ECAI'06 Workshop on Recommender Systems, 2006, vol. 94.

[۲۷] سیما ایرانمنش، محمدرضا زارع میرک‌آباد، فاطمه کاوه یزدی، "بررسی آسیب‌پذیری روش‌های مبتنی بر گراف در سیستم‌های توصیه‌گر با استفاده از روش‌های مزرعه پیوند"، نخستین کنفرانس ملی محاسبات نرم، دانشگاه گیلان، ۱۸ و ۱۹ آبان ۱۳۹۴.

[28] S. Adalı, T. Liu, and M. Magdon-Ismail, 'An analysis of optimal link bombs', Theor. Comput. Sci., vol. 437, pp. 1–20, 2012.

[29] A. N. Nikolakopoulos, M. A. Kouneli, and J. D. Garofalakis, 'Hierarchical itemspace rank: Exploiting hierarchy to alleviate sparsity in ranking-based recommendation', Neurocomputing, vol. 163, pp. 126–136, 2015.

[30] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, 'Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation', IEEE Trans. Knowl. Data Eng., vol. 19, no. 3, pp. 355–369, 2007.

[31] D. Yang, J. He, H. Qin, Y. Xiao, and W. Wang, 'A graph-based recommendation across heterogeneous domains', in proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, pp. 463–472.

[32] 'MovieLens DataSet', GroupLens Research. https://grouplens.org/datasets/movielens/.

[33] G. Guo, J. Zhang, and N. Yorke-Smith, 'A Novel Bayesian Similarity Measure for Recommender Systems', in Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 2619–2625.

[34] G. Shani and A. Gunawardana, 'Evaluating recommendation systems', in Recommender systems handbook, Springer, 2011, pp. 257–297.

[35] M. Eirinaki, J. Gao, I. Varlamis, and K. Tserpes, 'Recommender systems for large-scale social networks: A review of challenges and solutions'. Elsevier, 2018.

پاورقی‌ها:

[1] Information retrieval
[2] Deep learning
[3] Embedding
[4] Shilling attacks
[5] Link farm
[6] Rating matrix
[7] Heterogeneous graph
[8] Item-Rank
[9] Correlations graph
[10] Damping factor
[11] Push attacks
[12] Nuke attacks
[13] Co-visitation
[14] Selected items
[15] Filler items
[16] Unrated items
[17] Target item
[18] Reverse bandwagon attack
[19] Matrix factorization
[20] Bipartite
[21] Boosting pages
[22] Hijacked pages
[23] Hijacked links
[24] Direct individual
[25] Cycle
[26] Complete
[27] Correlations Farm
[28] Number of Common Profiles
[29] Number of Boosting items
[30] Number of Hijacked items
[31] Kendall's tau
[32] Spearman's rho
[33] Rank shift