

Mitigating Power User Attacks on a User-Based Collaborative Recommender System

David C. Wilson and Carlos E. Seminario

University of North Carolina at Charlotte
Charlotte, North Carolina, USA
davils@uncc.edu cseminar@uncc.edu

Abstract

Collaborative Filtering (CF) Recommender Systems (RSs) ease the burden of information overload faced by online users who browse, search, or shop for products and services. Influential users, known as “power users” are able to exert substantial influence over recommendations made to other users, and RS operators encourage the existence of power user communities to help fellow users make informed purchase decisions. However, the influence power users wield can be used for both positive (addressing the “new item” problem) or negative (attack) purposes. Attacks on RSs tend to bias recommendations by introducing fake reviews or ratings and remain a key problem area for system operators. In prior work, we have shown that attackers emulating power users are effective against user-based, item-based, and SVD-based CF RSs. Previous research has shown that, in general, attacks on RSs can be mitigated by detecting the attackers and either removing them from the dataset or ignoring them during the prediction calculations. In live RS environments, however, these approaches impact legitimate users detected as attackers (false positives) and can lead to reduced coverage for those legitimate power users as well as reduced accuracy for other users that depend on the ratings of those legitimate power users. Our research is investigating alternative mitigation approaches to address these issues for power user attacks. We focus on techniques that remove or reduce the influence of power users and determine their impact on RS accuracy and robustness using established metrics. We introduce a new metric used to assess the trade-off between accuracy and robustness when our mitigation approaches are applied. And our results show that, for user-based systems, reducing power user influence is more effective than removing power users from the dataset.

1 Introduction

Recommender Systems help users decide what products and services to buy from online providers. Influential “power” users, in the RS context, are those that are able to influence the largest group of RS users, i.e., power users have the ability to positively or negatively impact RS predictions for many other users. For white-hat purposes (e.g., addressing the “new item” problem), online systems encourage the formation of power user communities, e.g.,

Amazon VineTM invites trusted reviewers to post opinions about new and pre-release items.¹ In previous work (Wilson and Seminario 2014; Seminario and Wilson 2014), we have shown that attackers emulating power users are effective against user-based, item-based, and SVD-based recommenders. In the literature, mitigating RS attacks usually consists of detecting the attackers and either removing them from the dataset or ignoring them during the prediction calculations (Chirita, Nejdl, and Zamfir 2005; Mehta and Nejdl 2009). While removing attack user profiles from recommendation calculations is a straightforward approach to eliminating the attacker’s influence in a laboratory environment, in live RS environments this approach could also have unwanted side effects (Mehta and Nejdl 2009). For instance, in cases where a legitimate power user is mistakenly identified as an attacker (false positive) and is removed, two issues could occur: (1) the removed legitimate power user would no longer receive recommendations, and (2) the users that rely on that legitimate power user’s neighborhood influence may be impacted. These approaches also assume that all attackers will be detected, i.e., no provision is provided for attackers that are not detected (false negatives).

This study investigates the potential for more effective mitigation approaches against Power User Attacks (PUAs), as compared to 100% removal of identified power users. PUA mitigation seeks to balance the trade-offs between accuracy (too many power user profiles are removed) and robustness (too few power user profiles are removed) impacts. *Our hypothesis is that reducing the influence of power users is a more effective and less impactful mitigation strategy than removing the profiles of identified power users.*

The following research questions are used as a guide:

RQ1: What happens to RS accuracy/robustness when power user profiles are removed from recommendation calculations to mitigate the power user attack impacts?

RQ2: What happens to RS accuracy/robustness when power user influence is reduced during similarity calculations?

RQ3: What are the trade-offs between accuracy and robustness when power user attacks are mitigated?

¹<http://www.amazon.com/gp/vine/help>

2 Related Work

Power users are of particular interest to RS operators and their client companies when launching a new item, because a positive endorsement (high rating) can result in making item recommendations to many other users. This “market-based” use of RS has been previously promoted as a solution to the “cold-start” or “new item” problem (Anand and Griffiths 2011). A viral marketing perspective to exploit the network value of customers was studied in (Domingos and Richardson 2001).

Attacks on RSs by providing false ratings are known as “shilling attacks” (Lam and Riedl 2004), or “profile injection attacks” (Mobasher et al. 2007b; O’Mahony, Hurley, and Silvestre 2005). Research in attacks on recommender systems began in 2002 (O’Mahony, Hurley, and Silvestre 2002); a recent summary in (Burke, O’Mahony, and Hurley 2011) describes RS attack models, attack detection, and algorithm robustness. In (Wilson and Seminario 2013), we defined a novel *Power User Attack* (PUA) model as a set of power user profiles with biased ratings that influence the results presented to other users. The PUA relies critically on the method of power user identification/selection, so we also developed and evaluated a novel use of degree centrality concepts from social network analysis for identifying influential RS power users for attack purposes (Wilson and Seminario 2013). In (Wilson and Seminario 2014; Seminario and Wilson 2014), we have shown that attackers emulating power users are effective against user-based, item-based, and SVD-based recommenders.

Previous research has shown that, in general, attacks on RSs can be mitigated by detecting the attackers (Mobasher et al. 2007a; Williams, Mobasher, and Burke 2007; Sandvig, Mobasher, and Burke 2008; Burke, O’Mahony, and Hurley 2011) and either removing them from the dataset or ignoring them during the prediction calculations (Chirita, Nejd, and Zamfir 2005; Mehta and Nejd 2009). However, removing attack user profiles from recommendation calculations can also have unwanted side effects (Mehta and Nejd 2009) that result in reduced accuracy and coverage.

Therefore, a gap in the research is that attack mitigation strategies that preserve adequate accuracy, coverage, and robustness have largely been ignored. And this remains an open question in RS attack research that we continue to explore in this study.

3 Power User Attack Background

In order to study RS attacks based explicitly on measures of influence, we previously defined a *Power User Attack* model as a set of power user profiles with biased ratings that influence the results presented to other users (Wilson and Seminario 2013). The PUA consists of one or more user profiles containing item ratings (called attack user profiles) that “push” (promote) or “nuke” (disparage) a specific item. The PUA relies critically on the method of power user identification/selection, so we implemented a set of heuristic approaches for comparative purposes, as follows:

In-Degree Centrality (InDegree): Our approach based on in-degree centrality where power users are those that partic-

ipate in the largest number of neighborhoods (Wasserman and Faust 1994; Lathia, Hailes, and Capra 2008). In our implementation, for each user i compute similarity with every other user j applying significance weighting $n_{cij}/50$, where n_{cij} is the number of co-rated items and 50 items was determined empirically by (Herlocker et al. 1999) to optimize RS accuracy. Next, discard all but the top- k neighbors for each user i . Count the number of similarity scores for each user j (# neighborhoods user j is in), and select the top- k user j ’s.

Aggregated Similarity (AggSim): The Most Central heuristic from (Goyal and Lakshmanan 2012) selects top- k users with the highest aggregate similarity scores as the selected set of power users. This method requires at least 5 co-rated items between user i and user j and does not use significance weighting.²

Number of Ratings (NumRatings): This method is based on (Herlocker et al. 2004) where “power user” refers to users with the highest number of ratings; it also is called the Most Active heuristic in (Goyal and Lakshmanan 2012). We selected the top- k users based on the total number of ratings they have in their user profile.

4 Mitigation Strategies

Removing 100% of the power user attackers as a mitigation strategy could result in (1) reduced coverage for the “removed” users including legitimate users (false positives), (2) reduced accuracy for users whose similarity neighborhoods no longer benefit from the influence of the “removed” users including legitimate users (false positives), and (3) no provision for attackers that are not detected (false negatives) and assumes that all (true) power user attackers will be detected. To address these issues, the following mitigation strategies were initially evaluated in this study:

- Remove attackers incrementally from 0% to 100%.
- Reduce the similarity weighting factor of all attackers incrementally from 1.0 to 0.0.
- Combine removal and influence reduction.

Our analysis of these initial mitigation strategies determined the following:

When removing power user attackers incrementally from the dataset, removal sequence matters. From the attacker’s standpoint, it would be better to remove starting from least influential to most influential; while from the system operator’s standpoint, removing starting from most influential to least influential would be better. And we also analyzed the impacts when removing power user attackers randomly. Since this is a mitigation study, we decided to use a removal sequence that favored system operators. Our analysis indicated that removing power users starting from most influential to least influential improves robustness at a faster rate than the other two methods.

When mitigating the PUA, the type of target item matters. We used “New” target items (those with one rating) and “New and Established” target items (those with one or more ratings). From previous research (Seminario and Wilson 2014), we knew that New target items are more vulnera-

²Based on personal communication with the authors.

ble to attack than New and Established targets. We analyzed the impacts of the PUA and found that, for New targets, robustness metrics were relatively high until the weighting was set to zero (ignore power user influence). For New and Established targets we found that robustness measures were significantly lower between weightings from 1.0 to 0.1, indicating that the PUA was not as effective with these target items.

Combining power user attacker removal and influence reduction resulted in outcomes similar to the removal approach when the similarity weighting was greater than zero. Robustness was at a minimum only when the similarity weighting was zero. So, this approach did not provide additional information regarding power user attack mitigation.

In our initial approach, all injected attackers were also considered to be power users (even when they were not) so that removing and/or reducing the influence of power users assumed a perfect power user attacker detection method. This was not a very realistic assumption so we decided to use our power user selection methods (§ 3) to allow for a mix of real and synthetic power user attackers. Consequently, the final mitigation strategies (MS) for this study are:

MS1: Remove selected power users incrementally from 0% to 100%, starting from most influential to least influential.

MS2: Reduce the similarity weighting factor of all selected power users incrementally from 1.0 to 0.0.

MS3: Reduce the number of power users that influence predictions. The percentage of power users selected is reduced incrementally from 100% to 0% and the similarity weighting is set to one if selected, zero otherwise.

To implement these mitigation strategies, the following methodology was used:

1. Generate power user lists from selected datasets using power user selection techniques from Section 3: InDegree, NumRatings, and AggSim. This generates a list of real power users (RPU).
2. Generate synthetic power user (SPU) attack profiles based on power user statistical characteristics (Wilson and Seminario 2014) and insert them into the dataset. Select power users from the updated dataset using power user selection techniques described in Section 3. The top- k list of power users is expected to be a combination of RPUs and SPUs.
3. Select target items from a given dataset: New (items with one rating), New and Established (randomly-selected items with a range of popularity and likeability values).
4. Create incremental datasets with most-to-least-influential power users removed from the top of the top- k list.
5. Execute attacks for each mitigation strategy for the selected target items and calculate averaged metrics over all target items. Only SPUs will be used for attack purposes, leaving RPUs to provide their influence but not be part of the attack. Note that some SPU attack profiles will remain in the dataset after the top- k power users are removed during the experiments described in Section 5.
6. Compare accuracy, coverage, and robustness metrics for variations of the mitigation strategy to determine impacts of removing and reducing influence of power users.

Other parameters such as recommender algorithms, datasets, and metrics are also specified in Section 5.

5 Experimental Design

To address our research questions and hypothesis, we conducted three main experiments to correspond with the three final mitigation strategies (MS1-MS3) described in § 4:

- E1: Power User Removal
- E2: Power User Influence Reduction: All power users
- E3: Power User Influence Reduction: Selected power users

Evaluation Metrics: Evaluations were performed before and after the attacks. We use Mean Absolute Error (MAE) for accuracy and prediction coverage (Herlocker et al. 2004; Shani and Gunawardana 2011) using a holdout-partitioned 70/30 train/test dataset. To compare MAE before and after attacks, we use $\delta MAE = MAE_{after} - MAE_{before}$. For robustness metrics (Mobasher et al. 2007b; Burke, O’Mahony, and Hurley 2011), we use Hit Ratio (HR), Average HR (\overline{HR}), Prediction Shift (PS), Average PS (\overline{PS}), Rank (R), and Average R (\overline{R}), where a high Hit Ratio and a low Rank after the attack indicate that the attack was successful (from the attacker’s standpoint) assuming that the target item had a lower Hit Ratio and higher Rank before the attack. The top- N list of recommendations for Hit Ratio calculations is $N=40$. Also, when the PUA being evaluated uses “new” target items (items with 1 rating), the Prediction Shift is expected to be close to the maximum rating as defined by the RS. For “new and established” target items, the Prediction Shift may also be high because some of the SPUs may fall below the threshold of power users to be removed or have their influence reduced; the SPUs that remain after removal or influence reduction are still used in the attack and may contribute to the high Prediction Shift. Finally, to assist in the assessment of the effectiveness of the mitigation strategies and the trade-offs between accuracy and robustness, we developed the new Accuracy/Robustness/Mitigation measure (ARM),

$ARM = (2 * \frac{MAE_{after} * \overline{HR}}{MAE_{after} + \overline{HR}}) * (1 - \rho)$, where ρ is the percentage of power users or influence being evaluated. ARM varies between 0 and 1 and a higher ARM indicates a more effective mitigation for a given experiment. The major motivation behind the ARM metric is to find a measure that determines the level of removal or influence reduction that is best for mitigating the PUA. The ARM metric combines MAE and Hit Ratio in such a way that it balances the increase in MAE with the reduction in Hit Ratio as power user influence is removed or reduced.

Datasets and Algorithms: We used MovieLens³ ML100K⁴ and ML1M⁵ datasets. The CF user-based weighted algorithm (UBW) (Desrosiers and Karypis 2011) uses Pearson similarity with a threshold of 0.0 (positive correlation), neighborhood size of 50, and significance

³<http://www.grouplens.org>

⁴nominal 100,000 ratings, 1,682 movies, and 943 users.

⁵nominal 1,000,209 ratings, 3,883 movies, 6,040 users.

weighting of $n/50$ where n is the number of co-rated items (Herlocker et al. 1999). We used UBW from Apache Mahout⁶ and added functionality to implement the MS2 and MS3 strategies (§ 4).

Power User Selection: The InDegree (ID), NumRatings (NR), and AggSim(AS) methods described in § 3 were used.

Target Item Selection: Fifty target items with no more than one rating, regardless of their rating value, were selected randomly as “New” (New) target items. We also used 50 “New and Established” (New/Est) target items, i.e., target items were selected randomly and had the following average number of ratings, average rating, and average rating entropy, respectively: ML100K (73.780, 3.133, 1.769), ML1M (280.399, 3.296, 1.883).

Attack Parameter Selection: The Attack Intent is Push, i.e., target item rating is set to max (= 5). The Attack Size or number of SPUs in each attack varied by dataset: up to 50 for ML100K and up to 300 for ML1M. Attack sizes, also expressed as $(\frac{\#attackers}{\#users} * 100)\%$, were selected based on previous research (Mobasher et al. 2007b; Burke, O’Mahony, and Hurley 2011), where a 5-10% attack size was shown to be effective; we use a 5% attack size for each dataset. Power user attack profiles were generated as described in (Wilson and Seminario 2014) and target item ratings were injected at run time.

Test Variations: We used 3 experiments to evaluate the final set of mitigation strategies using one prediction algorithm, 2 datasets, 3 power user selection methods, 2 target item types, and 8 attack sizes.

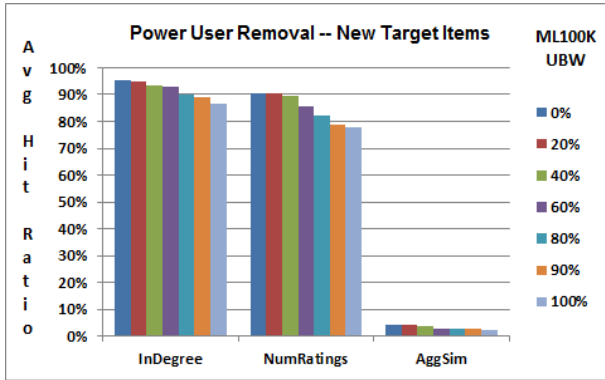


Figure 1: E1 – Hit Ratio and MAE as 0% to 100% of Power Users (Real & Synthetic) are removed using ML100K

6 Results and Discussion

(E1) Power User Removal: Consisted of removing power users from the dataset (incrementally from 0% to 100%) prior to recommendation calculations (similarity and prediction). We conducted a series of PUAs against the user-based CF algorithm. Each PUA in this experiment uses a dataset with a specified number of SPU attackers (§ 5). The SPU profiles are injected with New or New/Est target items at

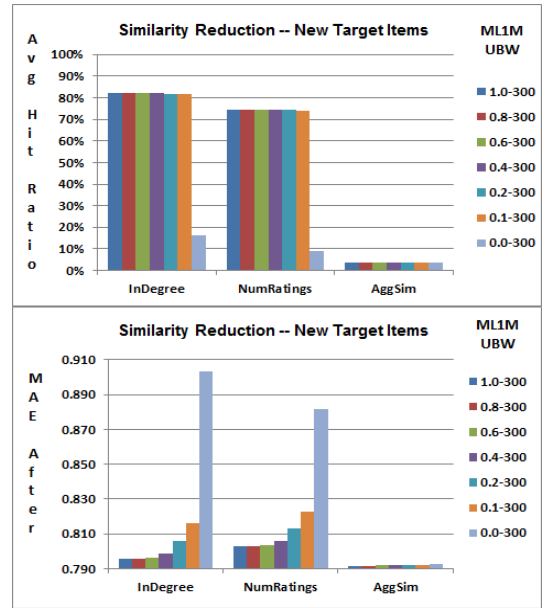


Figure 2: E2 – Hit Ratio and MAE as Power Users’ (Real & Synthetic) Influence reduced from 1.0 to 0.0 using ML1M

runtime; the 50 targets are evaluated one at a time for each PUA, then the HR/Rank/PS metrics are averaged across all 50 target items. We use the most-to-least-influential SPU removal approach since that better mitigates the attack effectiveness, from a system operator’s perspective, i.e., \overline{HR} drops off faster as the most influential power users are removed first.

\overline{HR} before the attack is 0% for New target items and 2% for New/Est target items for ML100K across all power user removal levels; these values serve as the \overline{HR} baseline and indicate that without attackers, the target items appear in few, if any, top- N lists of recommendations. Figure 1 shows the results for ML100K as the percentage of power users removed increases from 0% to 100% (0-50 power users). The drop in \overline{HR} is modest as power users are removed because some SPU attackers remain in the dataset, i.e., they were below the specified power user selection threshold and contributed to increasing the \overline{HR} . New/Est target items (not shown) are more difficult to attack compared to New targets and have a lower \overline{HR} starting at 50% for ID and NR, 4% for AS. Removing 100% of the power users still leaves SPU attackers in the dataset, hence \overline{HR} remains high; the PUA with AS is not effective at any level of removal. And as power users are removed, MAE varies between 0.80 to 0.83 for New and New/Est target items across all power user selection methods and all removal levels. \overline{PS} for ML100K and New target items averaged 4.9 for all power user selection methods across all removal levels; for New/Est target items \overline{PS} was 4.5 for InDegree and NumRatings, across all removal levels; 3.0 for AggSim. \overline{R} for ML100K and New target items varied between 3 and 4 for all power user selection methods across all removal levels; for New/Est target items \overline{R} varied between 13 and 15 across all selection meth-

⁶<http://mahout.apache.org>

ods and removal levels, indicating a less effective attack.

We observed similar results with ML1M except that for ID, \overline{HR} increased slightly as power users were removed, most likely due to the influence characteristics of the SPUs. \overline{R} for New target items varied between 7 and 9, and for New/Est target items varied between 16 and 21 for all power user selection methods across all removal levels. The ARM measure indicated that 100% removal is the best mitigation for all attacks in E1.

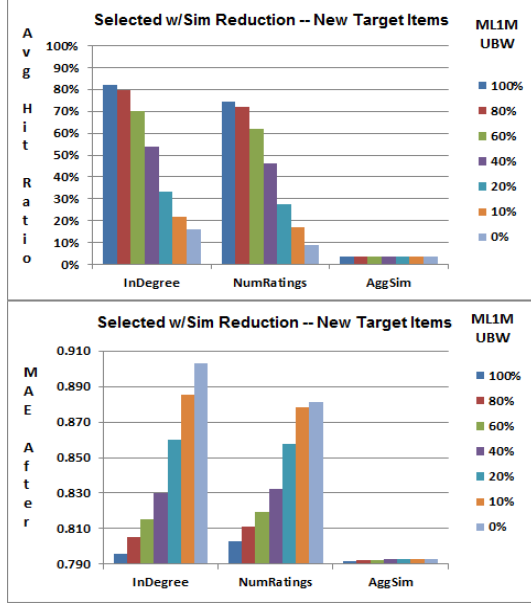


Figure 3: E3 – Hit Ratio and MAE as 100% to 0% of Power Users’ (Real and Synthetic) Influence is applied using ML1M

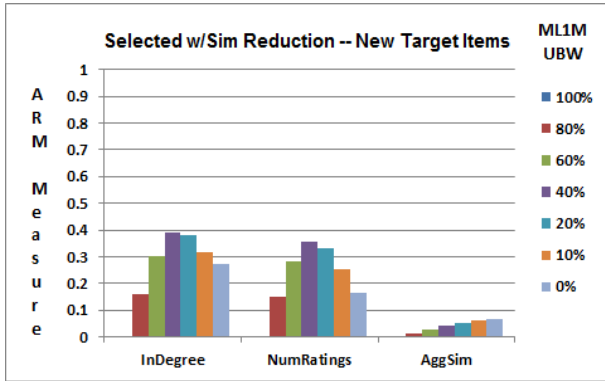


Figure 4: E3 – ARM Measure as 100% to 0% of Power Users’ (Real and Synthetic) Influence is applied using ML1M

(E2) Influence Reduction, all Power Users: Consisted of varying the similarity weighting (incrementally from 0.0 to 1.0) applied to all power users (selected RPUs and SPUs) who are nearest neighbors during the prediction calculation. Each PUA in this experiment uses a dataset with the

same number of power users, i.e., there is no removal of power users in this experiment. \overline{HR} before the attack is 0% for New target items and 1% for New/Est target items for ML1M across all power user influence reduction levels. New target item results in Figure 2 for ML1M indicate that as similarity weighting is reduced from 1.0 to 0.1, \overline{HR} remains flat for ID (81%), NR (74%), and AS (4%). When similarity weighting is set to zero, \overline{HR} drops significantly for ID (to 16%) and for NR (to 9%), while remaining flat for AS (4%). \overline{HR} is flat for ID and NR as similarity weighting is reduced from 1.0 to 0.1 and can be attributed to the fact that the SPUs are, in most cases, the only users that have rated the New target items; therefore, the SPUs dominate the influence within the neighborhoods keeping \overline{HR} high and \overline{R} low. The influence of power users (a mix of RPUs and some SPUs) can be observed in the significantly higher MAE results (lower accuracy) when similarity weighting is set to zero, i.e., without the power user influence, accuracy becomes much worse. The non-zero \overline{HR} when similarity weighting is set to zero indicates that not all attackers (SPUs) have been removed from the prediction calculations. \overline{PS} for ML1M and New target items was >4 and \overline{R} ranged between 9 and 14 for all power user selection methods across all reduction levels.

For New/Est targets (not shown), \overline{HR} begins at a much lower level (40% for ID and NR, 4% for AS) and remains flat until similarity weighting drops to 0.0, mainly because SPU influence is not very dominant within the neighborhoods as many other users have rated established items. As similarity weighting is reduced from 1.0 to 0.1, MAE increases for ID and NR, and remains flat for AS; at 0.0 similarity weighting MAE also rises significantly. \overline{PS} was >3.8 for New/Est target items and \overline{R} ranged between 14 and 22 over all similarity weightings and selection methods.

We observed similar results with ML100K except that \overline{HR} starts slightly higher (90%) and \overline{R} ranges lower (4-6 for New, 15-17 for New/Est targets) for ID and NR. For all attacks in E2, the ARM measure indicated that a similarity weighting reduction setting of 0.1 is the best mitigation for ID and NR, avoiding the spike in MAE_{after} albeit with high \overline{HR} , and 0.0 for AS.

(E3) Influence Reduction, selected Power Users: There is no power user removal in this experiment and only a percentage (varied from 0% to 100%) of power users are involved in the prediction calculation. The power users are selected randomly and have a similarity weighting of 1.0 if selected and 0.0 if not selected, during the prediction calculation. \overline{HR} before the attack (not shown) is 0% for New target items and 1% for New/Est target items for ML1M across all power user influence reduction levels. For ML1M with New target items shown in Figure 3, as the percentage of power users is reduced from 100% to 10%, \overline{HR} decreases from 80% to 20% for both ID and NR; for AS, \overline{HR} remains flat at 4%. When the percentage of power users is zero, \overline{HR} goes to 15% and 10% for ID and NR, and no change for AS. As the percentage of power users is reduced from 100% to 10%, MAE increases for ID and NR, remains flat (0.79) for AS. \overline{PS} for ML1M and New target items was >4 for all power user selection methods and power user percentages. \overline{R} ranges between 9 and 17 over all power user percentages.

and selection methods. For ML1M with New/Est targets (not shown), we observed a similar set of results except that \overline{HR} begins at a lower level (40% for ID and NR, 3% for AS), \overline{PS} is >3.8 , and \overline{R} ranges between 13 and 25 for all power user percentages and selection methods.

We observed very similar results with ML100K except that \overline{HR} starts slightly higher (90%) and \overline{R} ranges lower (3-6 for New, 13-18 for New/Est targets) for ID and NR. For New target items, the ARM measure (see Figure 4) indicated that a percentage of power user reduction of 20-40% is the best mitigation for ID and NR (avoiding the larger values of MAE_{after}) and 0.0 for AS. ARM results were the same for New/Est target items (not shown).

Based on our results and using the ARM metric, the mitigation strategy that best balances accuracy and robustness for ID and NR PUAs is MS3; the AS PUA was not an effective attack in this study and did not require mitigation. For MS1, the ARM metric indicates 100% removal which leaves a very high Hit Ratio. And MS2 is marginally better than MS1, with the ARM metric indicating a similarity weighting of 0.1. Our hypothesis is accepted for MS3, partially accepted for MS2, and rejected for MS1.

7 Conclusion

This paper evaluated power user attack mitigation approaches to address issues encountered when legitimate influential users (false positives) are removed along with attackers. We have shown that reducing similarity weighting during prediction calculation is an improvement over removal. We showed that there is a trade-off between accuracy (MAE) and robustness (Hit Ratio) when implementing mitigation strategies and have developed a metric to assist in evaluating this trade-off. Consistent with our previous work using user-based recommenders, we also showed that reducing the influence of power users contributes to a reduction in recommender system accuracy indicated by an increase in MAE; this shows how power users can impact recommendations. Our future work in this area will examine other recommender system algorithms and datasets.

References

- Anand, S. S., and Griffiths, N. 2011. A market-based approach to address the new item problem. In *Proceedings of the ACM Recommender Systems Conference*.
- Burke, R.; O'Mahony, M. P.; and Hurley, N. J. 2011. Robust collaborative recommendation. In Ricci, F., et al., eds., *Recommender Systems Handbook*. Springer.
- Chirita, P.-A.; Nejdl, W.; and Zamfir, C. 2005. Preventing shilling attacks in online recommender systems. In *WIDM '05: Proceedings of the 7th annual ACM international workshop on Web information and data management*. ACM.
- Desrosiers, C., and Karypis, G. 2011. A comprehensive survey of neighborhood-based recommendations methods. In Ricci, F.; Rokach, L.; Shapira, B.; and Kantor, P. B., eds., *Recommender Systems Handbook*. Springer.
- Domingos, P., and Richardson, M. 2001. Mining the network value of customers. In *Proceedings of KDD'01*. ACM.
- Goyal, A., and Lakshmanan, L. V. S. 2012. Recmax: Exploiting recommender systems for fun and profit. In *Proceedings KDD '12*. ACM.
- Herlocker, J. L.; Konstan, J. A.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proc of the ACM SIGIR Conf.*
- Herlocker, J. L.; Konstan, J. A.; Terveen, L. G.; and Riedl, J. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*.
- Lam, S. K., and Riedl, J. 2004. Shilling recommender systems for fun and profit. In *Proceedings of the 13th international conference on World Wide Web*. ACM.
- Lathia, N.; Hailes, S.; and Capra, L. 2008. knn cf: A temporal social network. In *Proceedings of the 2nd ACM Recommender Systems Conference (RecSys '08)*.
- Mehta, B., and Nejdl, W. 2009. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User-Adapted Interaction* 19(1-2).
- Mobasher, B.; Burke, R.; Bhaumik, R.; and Sandvig, J. 2007a. Attacks and remedies in collaborative recommendation. *Intelligent Systems, IEEE* 22(3):56–63.
- Mobasher, B.; Burke, R.; Bhaumik, R.; and Williams, C. 2007b. Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. *ACM Trans. Internet Technol.*
- O'Mahony, M. P.; Hurley, N.; and Silvestre, G. C. M. 2002. Promoting recommendations: An attack on collaborative filtering. In *Proceedings of DEXA '02*.
- O'Mahony, M. P.; Hurley, N.; and Silvestre, G. C. M. 2005. Recommender systems: Attack types and strategies. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI-05)*.
- Sandvig, J.; Mobasher, B.; and Burke, R. 2008. A survey of collaborative recommendation and the robustness of model-based algorithms. *IEEE Data Engineering Bulletin*.
- Seminario, C. E., and Wilson, D. C. 2014. Attacking item-based recommender systems with power items. In *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14*. ACM.
- Shani, G., and Gunawardana, A. 2011. Evaluating recommendation systems. In Ricci, F., et al., eds., *Recommender Systems Handbook*. Springer.
- Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.
- Williams, C.; Mobasher, B.; and Burke, R. D. 2007. Defending recommender systems: detection of profile injection attacks. *Service Oriented Computing and Applications*.
- Wilson, D. C., and Seminario, C. E. 2013. When power users attack: assessing impacts in collaborative recommender systems. In *Proceedings of the 7th ACM conference on Recommender Systems, RecSys '13*. ACM.
- Wilson, D. C., and Seminario, C. E. 2014. Evil twins: Modeling power users in attacks on recommender systems. In *Proceedings of the 22nd Conference on User Modelling, Adaptation and Personalization*.