Received: 22 June 2023 | Revised: 14 Aug 2023 | Accepted: 24 Sept 2023 | Published: 15 Nov 2023
Volume 1(2) Pages 33-51 DOI: https://doi.org/10.56279/jicts.v1i2.40

JICTS
Journal of ICT Systems

Impact of Trust Adjustment Factor on Artificial Intelligence Driven Collaborative Filtering Recommendation Algorithm

Edwin Ngwawe[1], Elisha Abade, Stephen Mburu

Department of Computer Science and Informatics, University of Nairobi, Nairobi, Kenya

[1]Corresponding author
Email: edwin.ngwawe@tukenya.ac.ke

Abstract

Trust on artificial intelligence (AI) is a major concern in the contemporary computing paradigms. Studies show that AI systems may outsmart humans, leading to an ultimate extinction of mankind. Therefore, the behavior of these systems must be controlled to avert potential use by bad actors. Recommender systems, which are variant of AI products, learn shoppers past data and predict items that shoppers may prefer. This helps in identifying items that may be recommended to the active user. Studies indicate that classical recommender systems allow untrustworthy data, tempting unscrupulous dealers to misdirect the learning process. This action potentially defrauds buyers. Our study introduces trust adjustment factor into the AI learning pipeline. We conducted experiments to test the difference in robustness of the trust-enhanced collaborative filtering recommendation algorithm against the classical counterpart. Prediction shift and hit ratios for the two sets of algorithms were measured when subjected to various forms of profile injection attacks. We found that the trust-enhanced variant of the algorithm significantly outperforms classical collaborative filtering recommendation in terms of robustness by up to 52% when measured by prediction shift and by up to 18% when measured by hit ratio. Confirmed by t-test, results suggest that embedding trust adjustment factor into recommender systems improves its robustness.

Keywords

Artificial Intelligence
Decision Support
Ethics in e-commerce
E-commerce
Recommender Systems

1. Introduction

The shopping trend around the world are geared towards e-commerce [1], and was partly ignited by the COVID-19 pandemic [2]. Shoppers were encouraged to purchase online products as much as possible from home to avoid potential exposure to the pandemic. This has led to information-intensive

environments in the e-commerce space as buyers get exposed to too much alternatives from many e-commerce platforms. It has become too difficult to make informed and timely decisions about correct items, thereby presenting a problem of information overload. This situation invites a need for decision support tools, such as recommender systems [3]. These are tools which suggest suitable items to users amidst a myriad of alternatives [4].

One of the most widely used recommender system algorithms is the Collaborative Filtering Recommendation Algorithm (CFRA). CFRA works by profiling users according to their previous purchase history, and then tries to estimate or predict active users' preferences. The aim of this profiling is to provide an active user with recommendations about purchase of the next item. This involves learning the preferences from previous data. Therefore, an application of machine learning is part of Artificial Intelligence (AI) [4,5].

As part of AI, one of the greatest concerns in research and implementation of recommender systems is the control of the output. This is necessary to prevent abuse of the mathematical properties by the malicious actors.

Control of AI is a major concern in the contemporary world [6]. Indeed, as of this writing, there is an open letter to tech giants to pause AI experiments for at least six months. This is expected to allow time for regulation to catch up [7,8]. If this control is not addressed, then, for recommender systems, it will provide a room for malicious actors to manipulate the output. This can be done by manipulating the input data in a manner which detriments the active user. This can be achieved by inserting fake item rating data into the recommendation system database to mislead the learning process, hence misleading the prediction process. This misled prediction process may be designed to result into more qualified items being suppressed, a process known as product nuking. It may also be designed to result into less qualified items being promoted for recommendation, a process known as product promotion. Both of the above possibilities are forms of profile injection attacks [9].

The threat is even more amplified as it manifests in e-commerce which is now becoming the new normal way of shopping [10]. This means that more people are likely to be affected if the concern is not addressed. To emphasize this, in the year 2023, Walmart, the leading American retail store, is set to close 20 stores due to underperformance [11]. Even though they do not cite specific reasons for the closure, observers and analysts connect it partly to online giants, such as Amazon. These online giants are taking too much of the retail market space [10]. This also leads to development of dark stores, types of stores which do not involve walk-in customers but only employees. These employees work to deliver orders placed online or through mails. Such stores provide more room for employees to deliver orders faster and further takes up the brick and mortar market shares. Indeed, in the contemporary world, for the brick and mortar stores, the greatest competitor is not the store across the street but the online giants. This indicates the trends on the transition of shopping styles across the world.

The contemporary e-commerce is geared towards full automation, which involves assessing customer needs from the trend of consumption, recommending a list of shopping items to the customer, letting the customer approve the recommendation as a way of placing order (and have funds deducted from their accounts for the order in the process of approving the recommendation, or at the time of items delivery), and then having a robot package the order and finally, an artificial intelligence or self-driven cars deliver the order to the customer premise. The work of Bogue [12] indicates that Amazon was the

pioneer in the use of robots in order fulfilment centres. As it can be seen from the foregoing discussion, this fully bypasses active human intervention and fully relies on machines which in turn rely on the data that have been provided to them to learn from. If the data is not trustworthy, then there is likelihood that unscrupulous vendors may have a room to manipulate the system in a manner which misdirects the whole process. This does not only disadvantage the end buyer, but also it stands on the way of this interesting progress. Inserting fake customer items profiles makes the recommendation engine recommend to the buyers' items that are not of value to them, albeit fraudulently.

Burke et al. [9] carried out a research on robust collaborative recommenders and demonstrated the effects of profile injection on classical recommender systems. Classical recommender system is the recommender system in its regular occurrence and when trust is not added to the pipeline. The authors used various attack vectors or forms of profile injection attacks. From their work, it is clear how the common filtering recommendation algorithm is open to manipulations which are risky to the business process.

In 2017, Yin et al. [13] worked on improving the recommendation algorithm using trust in sociology. They carried out an experiment and demonstrated that trust based in sociology improves the prediction accuracy of collaborative filtering recommendation algorithm. The challenge of this work is that the authors used data which was actively collected from users. This means of estimating trust for computational purpose is cumbersome, and is subject to the user's discretion to provide opinion. Also, the dataset [14], which was collected and used by then, is now old and can no longer be relied upon to predict current user preferences. The collection of these data stopped long time ago and user preferences change over time.

The work of Yasmin et al. [15] illustrates the power of digital marketing tools. Without trust, these tools can also be misused to misdirect business actors unfairly.

Ziheng et al. [16] demonstrated poisoning of recommender systems with counterfactual examples using an attack method known as Horn-Clause Attacks to Recommender Systems (H-CARS). They carried out an experiment on two distinct datasets by using well-known counter factual generation methods. The finding was that H-CARS yields significant and successful attack on performance. This work also supports the desire to incorporate trust mechanism into the recommendation process.

In 2018, Mingdan and Qingshan reviewed shilling or profile injection attacks against Collaborative Filtering algorithms [17]. They analyzed the shortcomings of existing detection schemes and gave some proposals to improve shilling attack detection rates and robustness of collaborative recommendation. The recommendations included use of crossing media data to raise the robustness of Collaborative Filtering Recommender Systems. This involves taking advantage of users' trust relationships and distrust information to strengthen the discrimination of both genuine and fake users. This proposal is yet to be implemented, so not much can be said about it. But the paper emphasizes a need to curb profile injection.

In 2019, Shuai and Yao surveyed recommender systems based on deep learning [18]. They provided a comprehensive review of recent research efforts on deep learning-based recommender systems. The key challenge they found is that big and complex neural models are just fitting the data without any true understanding. This

still leaves room for malicious data to find their way into the recommendation process. It therefore still creates a need to find a mechanism to filter out such data [18].

In 2021, Fei et al. [19] used supervised learning to identify online article reviewers as spammers or not. The aim was to improve robustness of recommender systems. They tried to prevent the recommender system from ingesting input from spammers who try to promote items unfairly. Their results were consistent with human judgment [19]. This method can, however, still be evaded if the attack was statistically calculated. This way, the item is promoted moderately to avoid a burst which will trigger suspicion.

In 2022, Manqing et al. [20] did a literature survey for trust-aware recommender systems. On robustness, they stated that recommender systems face a series of intended or unintended noises that challenge the robustness of the recommender systems. One of their suggestions was to filter out noisy or malicious feedback from the data before executing the recommendation algorithm. Our work aims to address the suggestion.

Search advertising, also known as sponsored search, ranks search results to imply ordered superiority of the output. This is done by putting paid for items in areas which capture the active user's attention. Sometimes this implied superiority does not regard whether the paid for item is the best match for the search. This is another example of an unfair indirect recommendation of a product to the buyer. Studies on search advertising have mostly focused on maximizing the advertisers' profit [21,22,23,24,25]. They, however, disregard ethics or trustworthiness of the service the advertiser offers.

In 2022, Ngwawe et al. [26] came up with factors that buyers consider to classify an e-commerce platform as trustworthy or not. These factors can be considered and be automated to be used computationally and automatically. The study started with an exploratory stage to discover the items of concern, and later extended to confirmatory stage, yielding a model as a trust adjustment factor (Figure 1). Here, four components, namely security, privacy, deception and reliability were generated as the confirmed components of trust, alongside their indicators [27].

In 2023, Ngwawe et al. [28] embedded the generated trust model into a recommendation system pipeline for empirical evaluation in an e-commerce platform to test its impact. The finding was that the trust model improved the accuracy of a recommender system significantly. The authors proposed further research to evaluate the impact of this trust parameter on other properties of recommenders, such as robustness and serendipity, among others.

This paper addresses robustness, a measure of the resilience of recommender system when confronted with various forms of attacks. Attempt is made towards mitigating the possibility of recommender system taking up malicious data as input and potentially derailing the learning process. An experiment is carried out to test if the model helps in filtering out untrustworthy data. This approach is expected to improve the robustness of the recommender system.

2. Methodology

2.1 Hypothesis

The ongoing concern involves worries that recommendation process can potentially be manipulated by insertion of fake data. The property of a recommender system which measures how a recommendation process is resilient to malicious input data is referred to as the robustness of the recommender system. Robustness also means how the output remains stable even when an attacker tries to manipulate it using malicious or

untrustworthy input data. This means that using trust to filter out untrustworthy data is expected to improve the recommender system's robustness.

The robustness of a collaborative recommender system is measured using two metrics, namely prediction shift and hit ratio. The prediction shift is the difference between the rating before and after attack. The lower the prediction shift for an attack, the more robust the algorithm is against that attack. The hit ratio is the average likelihood that a top-n recommender will recommend the pushed item. The lower the hit ratio for an attack, the more robust the algorithm is against that attack.

It is therefore, hypothesized that a trust model, called trust adjustment factor, can help improve the robustness of a collaborative filtering recommendation algorithm. The hypothesis is expressed as follows:

H_o: A trust adjustment factor in the form of a trust model has no significant effect on the robustness of a collaborative recommendation algorithm.

H_1: A trust adjustment factor in the form of a trust model has positive significant effect on the robustness of a collaborative recommendation algorithm.

As described earlier, robustness can be measured using both prediction shift and hit ratio. Therefore, the hypothesis should be expressed in both terms as follows:

$$H_o: \mu_{predo} = \mu_{pred} \text{ or } H_o: \mu_{predo} - \mu_{pred} = 0 \quad (1)$$

Where:

μ_{predo} is the prediction shift before embedding trust

μ_{pred} is the prediction shift after embedding trust

$$H_o: \mu_{hit\_ro} = \mu_{hit\_r} \text{ or } H_o: \mu_{hit\_ro} - \mu_{hit\_r} = 0 \quad (2)$$

Where:

μ_{hit_ro} is the hit ratio before embedding trust

μ_{hit_r} is the hit ratio after embedding trust.

It is important to note that the existing collaborative filtering recommendation algorithm has two forms, user-based and item-based. The categorization is based on the where data for prediction of user preferences is drawn from. Data for prediction can be drawn from on items and their similarities or from the users and their similarities [9].

Since both forms of collaborative recommender algorithm can each be attacked by several forms of profile injection attacks [9], and these several forms of attacks affect both the prediction shift and the hit ratio, it was imperative to further break down the two sub hypotheses into further sub hypotheses to illustrate how each attack is mitigated for both the prediction shift and the hit ratio and for both forms of the collaborative filtering algorithms (Table 1).

The steps of hypothesis testing described by Shafer and Zhang [29] were relied upon to test the hypothesis after carrying out the experiment. To identify the relevant test, the concept of central limit theorem [30] was considered. Since all of the observations were less than 30, t-statistic was used. Confidence level of 95% was chosen since this is the most widely used threshold, changed only with a strong reason that could not be in our study; so, σ or significant level is 0.05.

We used one tailed paired two sample for means t-test from the Data Analysis ToolPak Add-in in Microsoft Excel to compute the p-values, σ. The results of this procedure can be interpreted by considering recommendations from Dorfman et al. [31]. To make a decision from the p-value, it is compared with the significance level α. The general rule is that when the p value is less than the significance level, we reject the null hypothesis.

Table 1. Robustness Hypothesis Testing Results.

S/N	Description	Sub Hypothesis	P - value (σ)	T-stat	t-critical one tail	Number of observations (n)	Reject Null?
1	Prediction Shift for product push attack on user based collaborative filtering algorithm	Average	5.91139E-11	12.56767938	1.729132792	20	YES
		Bandwagon	9.44649E-11	12.22601834	1.729132792	20	YES
		Random	4.94171E-10	11.07731306	1.729132792	20	YES
2	Hit Ratio for product push attack on user-based collaborative filtering algorithm	Average	3.73515E-07	12.04270923	1.833112923	10	YES
		Bandwagon	7.29474E-07	11.1291124	1.833112923	10	YES
		Random	7.29474E-07	11.1291124	1.833112923	10	YES
		Baseline	0.018393749	2.449489743	1.833112923	10	YES
3	Prediction Shift for product push attack on item-based collaborative filtering algorithm.	All Users	3.12212E-16	24.78574859	1.729132792	20	YES
		In Segment	6.38908E-11	12.51052343	1.729132792	20	YES
4	Hit Ratio for product push attack on item-based collaborative filtering algorithm	All User	4.64392E-06	8.907784453	1.833112923	10	YES
		In Segment	1.11802E-08	18.05320007	1.833112923	10	YES
		Baseline	0.011449747	2.738612788	1.833112923	10	YES

S/N	Description	Sub Hypothesis	P - value (σ)	T-stat	t-critical one tail	Number of observations (n)	Reject Null?
5	Prediction shifts achieved by nuke attacks against the user-based algorithm	Average	2.08517E-13	-17.34469513	1.729132792	20	YES
		Bandwagon	6.09932E-15	-21.08166868	1.729132792	20	YES
		Random	4.05566E-14	-18.99524052	1.729132792	20	YES
		Love/Hate	6.33388E-14	-18.53242445	1.729132792	20	YES
		Reverse Band Wagon	1.80508E-14	-19.86250197	1.729132792	20	YES
6	Prediction shifts achieved by nuke attacks against the item-based algorithm	Average	2.98625E-18	-31.83765479		20	YES
		Bandwagon	4.49569E-09	9.671342365	1.729132792	20	YES
		Random	0.005577841	2.810891842	1.729132792	20	YES
		Love/Hate	7.08453E-08	8.090729559	1.729132792	20	YES
		Reverse Band Wagon	2.39677E-17	-28.46979978	1.729132792	20	YES
7	Hit ratios achieved by the popular, probe and average push attacks against the user-based algorithm.	Popular	0.00019816	5.467934261	1.833112923	10	YES
		Probe	8.33625E-07	10.95445115	1.833112923	10	YES
		Average	7.78135E-06	8.358885556	1.833112923	10	YES

2.2 Setup of the Experiment

It was sought to augment the trust model generated as per the steps described by Ngwawe et al. [27] and as shown in Figure 1 as a new parameter, called trust adjustment factor, into the pipeline of existing collaborative recommendation

algorithm. This is because the existing collaborative recommendation algorithm is well-known and is also known to be susceptible to profile injection attacks [9]. It was imperative to assess if the trust model improves the collaborative recommendation algorithm against such attacks.

The steps of existing collaborative recommendation algorithm are well known and have been tried and tested. These steps are described in the work of Yin et al. [13].

The known steps were extended by adding the trust parameter, called the trust adjustment factor. The extension was done using the procedure described by Ngwawe et al. [28]. The authors describe the algorithmic steps to derive the trust parameter, choose the threshold for minimum trust score, set up the model for empirical tests, gather experimental data, and provide scientific tools and procedure for estimation of user preferences using artificial intelligence. Figure 2 shows a sample program of the model as a function.

2.3 Mounting Attacks

Several profile injection attacks were mounted, one at a time. This was done using various statistical methods for profile injection to evade obvious flagging by naïve anomaly detection systems. These methods have been tried and tested by Burke et al. [9].

2.3.1 Random Attack (Basic Attack)

To mount the Random Attack, which is considered as a basic attack because of its simplicity, the following steps were followed:

i. Assign random ratings distributed around the overall mean assigned to the filler items; and
ii. Assign a pre-specified rating assigned to the target item, r_{max} (maximum rating), for push, r_{min} (minimum rating) for nuke.

2.3.2 Average Attack (Basic Attack)

To mount the Average Attack, which is also considered as a basic attack because of its simplicity, the following steps were followed:

i. For each filler item, assign a rating that corresponds to (either exactly or approximately) to the mean rating for that item, across the users in the database who have rated it; and
ii. Assign a pre-specified rating assigned to the target item, r_{max} for push, r_{min} for nuke.

2.3.3 Bandwagon Attack (Low-knowledge attacks)

To mount the Bandwagon Attack, the following steps were followed:

i. Associate the attacked item with a small number of frequently rated items; and
ii. Assign a pre-specified rating assigned to the target item, r_{max} for push, effective for user-based, not item-based algorithm.

2.3.4 Segment Attack (Low-knowledge attacks)

To mount the Segment Attack, the following steps were followed:

i. Find a targeted group of users with known or easily predicted preferences; and
ii. Assign a pre-specified rating assigned to the target item, r_{max} for push, r_{min} for nuke.

2.3.5 Love/Hate Attack - Nuke Attack

To mount the love/hate attack, the following steps were followed:

i. Assign r_{min} to the target item; and
ii. Assign r_{max} to all other filler items.

2.3.6 Reverse Bandwagon Attack - Nuke Attack

To mount the reverse bandwagon nuke attack, the following steps were followed:

i. Identify items that tend to be rated poorly by many users; and
ii. Assign these items low ratings together with the target item.

2.3.7 Popular Attack (Informed)

To mount popular attack, which is considered as an informed attack because the attacker needs some prior information, the following steps were followed:

i. Get the average rating for the target item;
ii. Rate the filler items either r_{min} +1 and r_{min}, according to whether the average rating for the item is higher or lower; and
iii. For negative prediction shifts, assign the target item a rating of r_{min}, and ratings of r_{max} and r_{max}−1 to the filler items.

2.3.8 Probe Attack Strategy

To mount the probe attack, the following steps were followed:

i. Create a seed profile then use the seed profile to generate recommendations from the system (will be well-correlated with real users' opinions);
ii. Then learn the system with these recommendations; and
iii. Finally, use the knowledge to perform an attack – To mount a segment attack, probe narrowly and to mount an average, probe widely.

2.4 Procedure for the Experiment

Two sets of experiments were then run on the same dataset, generated as described by Ngwawe et al. [28] and then subjected to profile injection attacks.

The first set of experiments involved generating recommendations using the classical common filtering recommendation algorithm, which is the control or ablation experiment. The other set involved generating recommendations using the trust enhanced variant of the algorithm, which has the trust model embedded into the recommendation pipeline as a trust adjustment factor so as to filter out the products which do not meet the trust threshold.

To test the robustness of the trust enhanced algorithm, which is the key for our problem statement, we test prediction shift and hit ratio. The procedure for evaluating prediction shift and hit ratio is described by Burke et al. [9]. Every item is attacked individually by inserting fake user profiles which nuke or promote the item as suitable for the stage of the research.

2.5 Assumptions and limitations

There is an assumption that users were utmost genuine when providing ratings for the data [28], which was considered benign before the attacks were mounted into the data.

The key limitation of this work, and also with collaborative recommendation in general, is the cold boot problem. This means that a new item cannot be recommended to users until a few users have provided some ratings to it, even when the new item is superior. This can affect the hit ratios metrics [32].

2.6 Data Analysis

Data analysis involved a comparative analysis. This was done by comparing the performance of the trust enhanced algorithm against the performance

of the regular algorithm in terms of their robustness.

3. Results

Table 2 shows maximum effects of incorporating the trust adjustment factor for different forms of profile injection attacks. The Table shows values of hit ratios and prediction shifts before and after trust adjustment factor are incorporated into the pipeline. It also shows the change in the values as well as the percentage changes in the face of various forms of attacks.

Figures 3, 5, 7 and 8 show graphs which depict the prediction shift against percentage attack size. Percentage attack size is the ratio of maliciously injected profiles to the total profiles under consideration. For Figures 4, 6 and 9, the graphs depict percentage hit ratios against the number of recommendations. For every Figure, the graphs capture forms of attack when trust adjustment factor is incorporated into the pipeline and when it is not. The graphs where the trust adjustment factor is incorporated are suffixed with the word trust, after the name of the form of attack.

Table 1 shows t-test results for every sub-hypothesis. These sub hypotheses are based on forms of profile injection attacks. They are categorized by a combination of parameter to measure intention of attack and the type of algorithm. It indicates the t-test results, remark (in the 8th column) which is the consideration as to whether to reject the sub hypothesis or not and finally the decision to reject the sub hypothesis in the 9th column.

4. Discussion

The effect of the trust model, called a trust adjustment factor, was tested on the robustness of collaborative filtering recommendation system. The results indicate that the trust enhanced collaborative filtering recommendation system outperforms the classical collaborative filtering recommendation system.

The classical system refers to the recommender system in its regular occurrence and when trust is not added to the pipeline. The measurements are in terms of units of robustness. The performance improves by up to 52% when measured by prediction shift and up to 18% when measured by hit ratio.

Table 2 shows summary of cases of profile injection attacks where maximum effect of adding trust adjustment factor to the pipeline are realized. Table 1 shows t-test results for every sub hypothesis. The results from both Tables confirm that these effects are not by chance, but have statistical significance and, therefore, the null hypothesis must be rejected.

Figures 3 to 9 show the general trend which involve several forms of attack for the two forms collaborative filtering recommender algorithm. In each of the Figures, the graphs illustrate a case where trust is incorporated into the pipeline and one where trust is not incorporated.

The Figures are categorized according to intention of attack, type of algorithm and the parameter being measured. This means product push or nuke, item-based or user-based algorithms and measurements by either prediction shift or hit ratio. This type of categorization of results is based on literature [9].

In general, for product push (product promotion) attacks, both hit ratio and prediction shifts are lower when trust is incorporated into the pipeline. The prediction shift is also lower in the case of product nuke when trust is incorporated into the pipeline. Hit ratio is still high for product nuke attack for the case where trust is incorporated into the pipeline. These mean good results because, in all the cases, the attackers do not succeed in meeting their objectives.

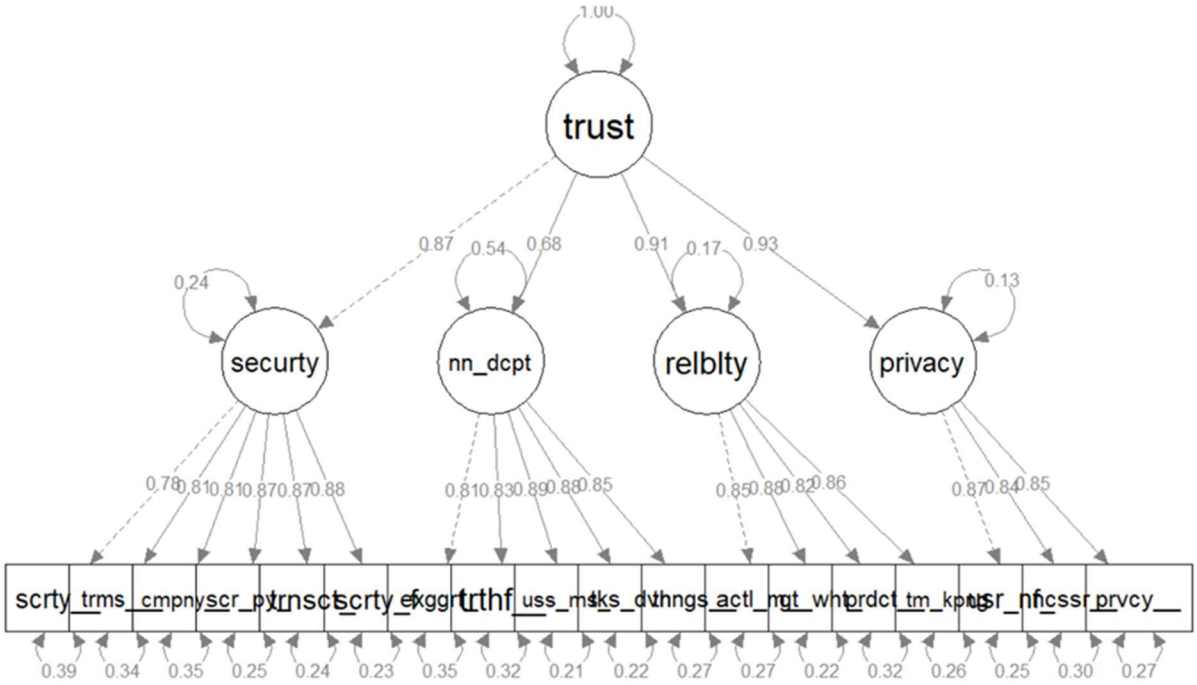

Figure 1. Trust model with security, non-deception, reliability and privacy constructs.

```
function compute_trust_score(){
    $query = "select * from wp_affiliates";
    $result = $this->select($query);
    while($row=mysqli_fetch_array($result)){
        $id = $row['id'];
        $y_1 = $row['y_1'];
        $y_2 = $row['y_2'];
        $y_3 = $row['y_3'];
        $y_4 = $row['y_4'];
        $y_5 = $row['y_5'];
        $y_6 = $row['y_6'];
        $security = (($y_1*$Cy_1/$sum_security)+($y_2*$Cy_2/$sum_security)+($y_3*$Cy_3/$sum_security)+($y_4*$Cy_4/$sum_security)+($y_5*$Cy_5/$sum_security)
              +($y_6*$Cy_6/$sum_security));
        $y_7 = $row['y_7'];
        $y_8 = $row['y_8'];
        $y_9 = $row['y_9'];
        $privacy = (($y_7*$Cy_7/$sum_privacy)+($y_8*$Cy_8/$sum_privacy)+($y_9*$Cy_9/$sum_privacy));
        $y_10 = $row['y_10'];
        $y_11 = $row['y_11'];
        $y_12 = $row['y_12'];
        $y_13 = $row['y_13'];
        $reliability = (($y_10*$Cy_10/$sum_reliability)+($y_11*$Cy_11/$sum_reliability)+($y_12*$Cy_12/$sum_reliability)+($y_13*$Cy_13/$sum_reliability));
        $y_14 = $row['y_14'];
        $y_15 = $row['y_15'];
        $y_16 = $row['y_16'];
        $y_17 = $row['y_17'];
        $y_18 = $row['y_18'];
        $non_deception = (($y_14*$Cy_14/$sum_non_deception)+($y_15*$Cy_15/$sum_non_deception)+($y_16*$Cy_16/$sum_non_deception)+
        ($y_17*$Cy_17/$sum_non_deception) +($y_18*$Cy_18/$sum_non_deception));

        $trust = ($security*$CSec/$sum_second)  - ($non_deception*$Cnd/$sum_second) +($reliability*$Crel/$sum_second) +($privacy*$Cpr/$sum_second);
        $this->write($this->log, "SECURITY: $security | PRIVACY: $privacy |  RELIABILITY: $reliability | NON DECEPTION: $non_deception | TRUST: $trust");

        $query = "update wp_affiliates set trust_score = '$trust' where id = $id limit 1";
        $this->write($this->log, "About to update trust score");
        $this->update($query);
```

Figure 2. Computation of trust value.

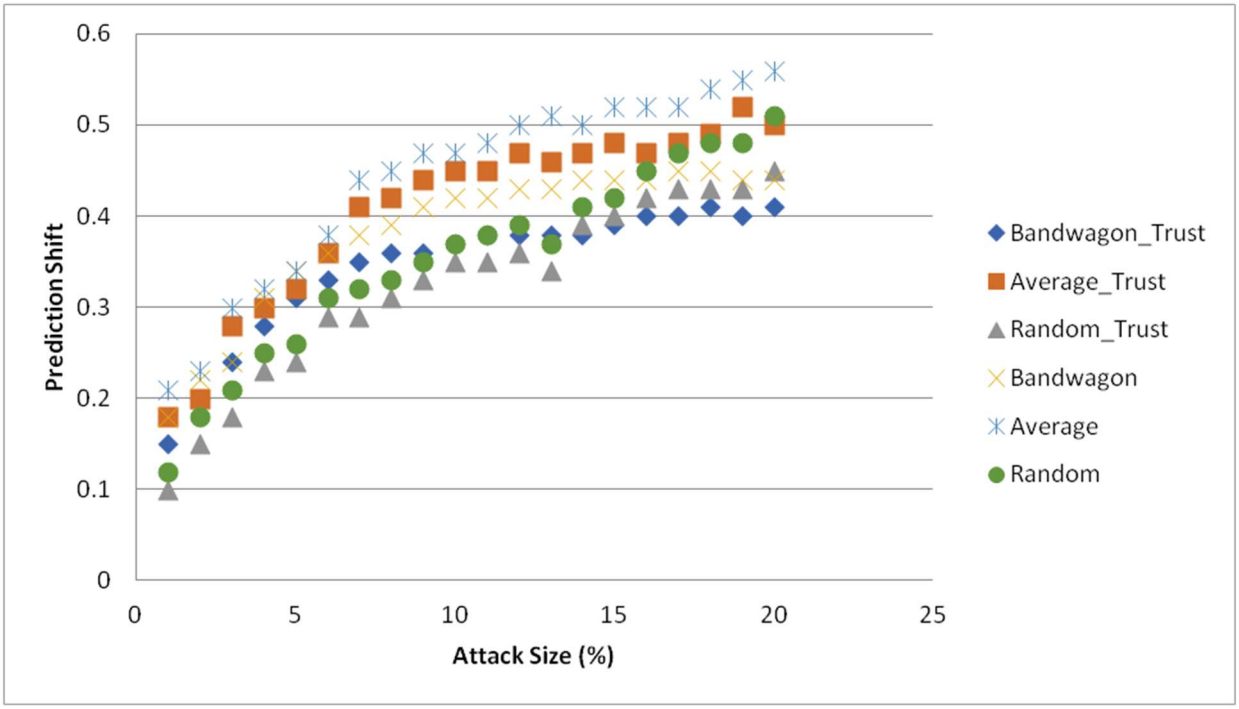

Figure 3. Prediction Shift for product push attack on user based collaborative filtering algorithm.

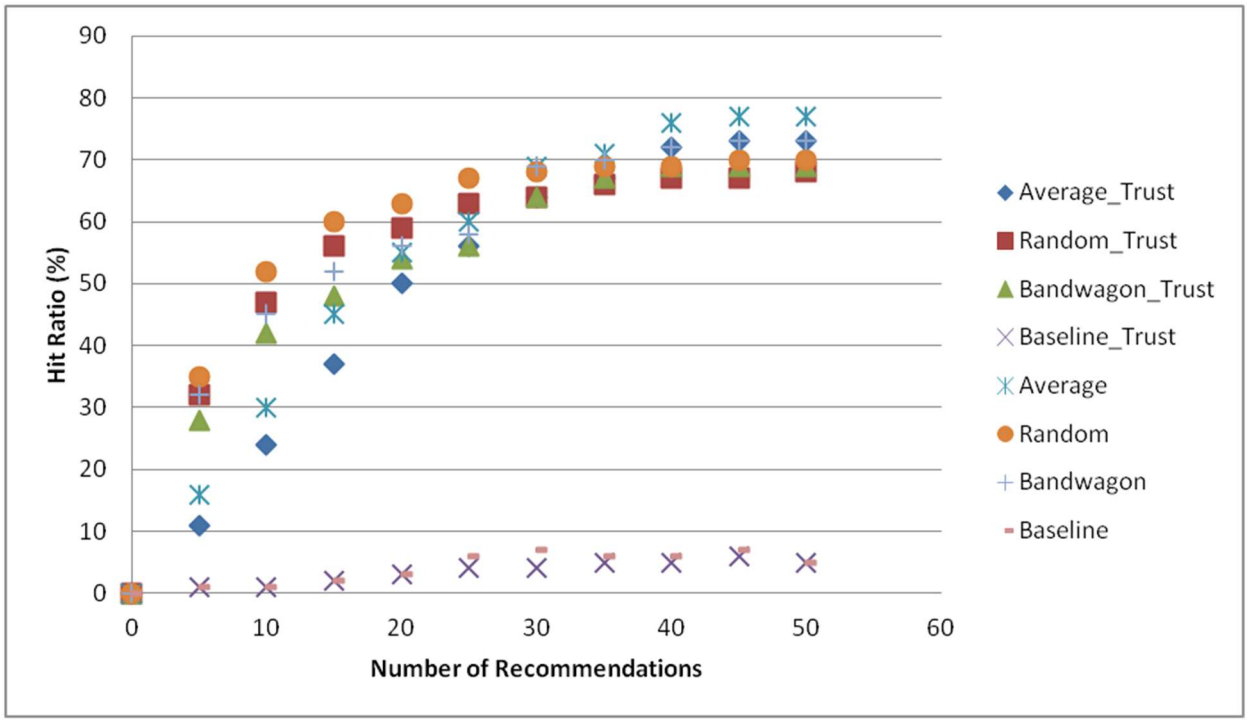

Figure 4. Hit Ratio for product push attack on user-based collaborative filtering algorithm.

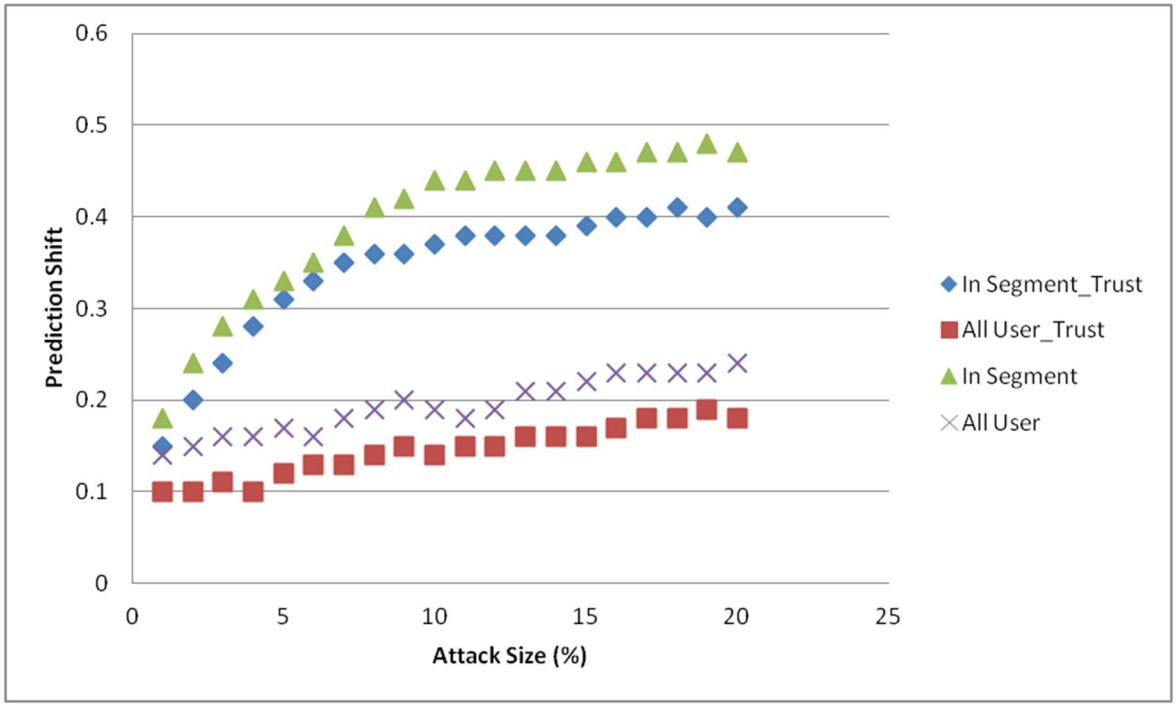

Figure 5. Prediction Shift for product push attack on item-based collaborative filtering algorithm.

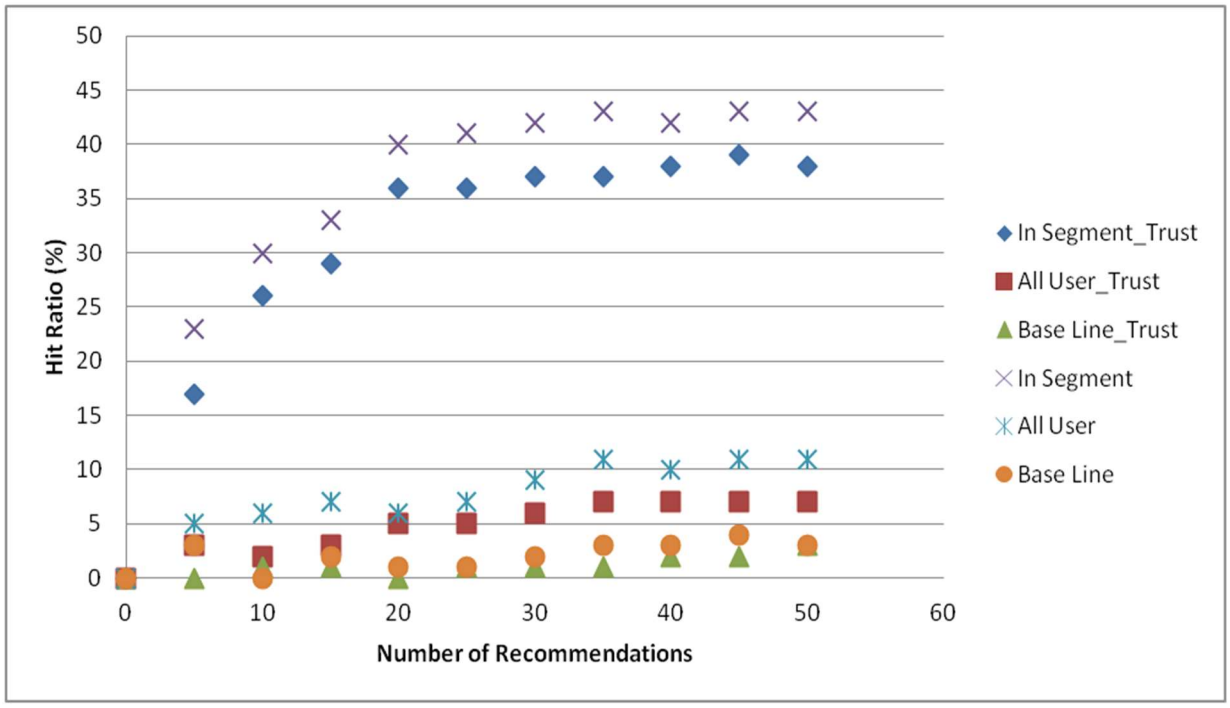

Figure 6. Hit Ratio for product push attack on item-based collaborative filtering algorithm.

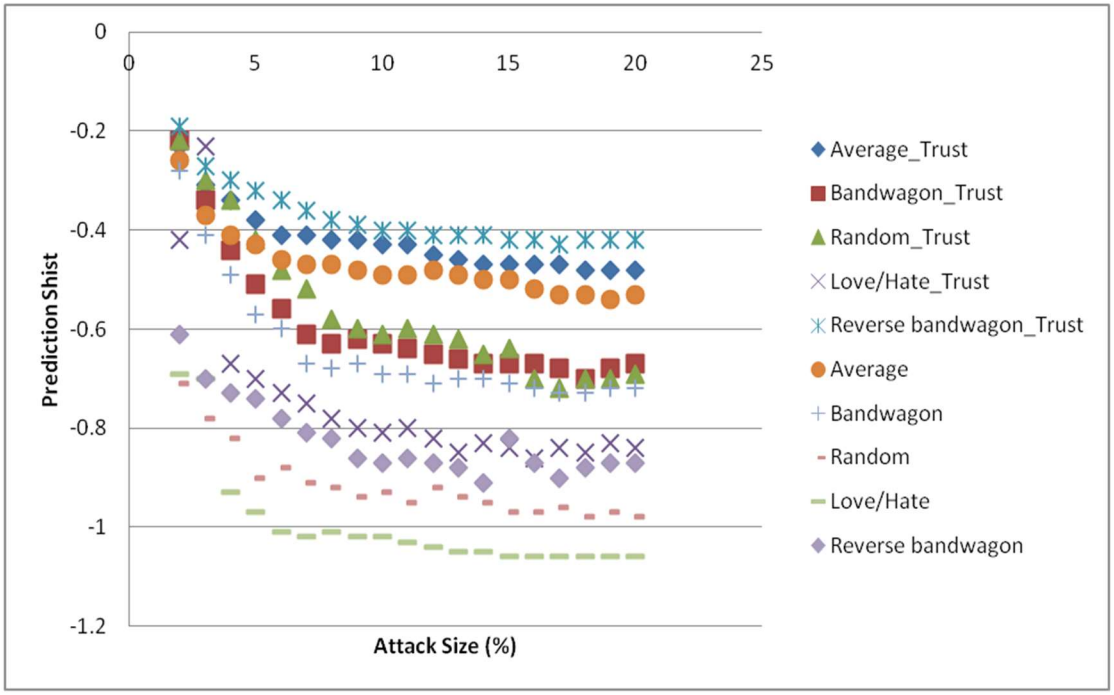

Figure 7. Prediction shifts achieved by nuke attacks against the user-based algorithm.

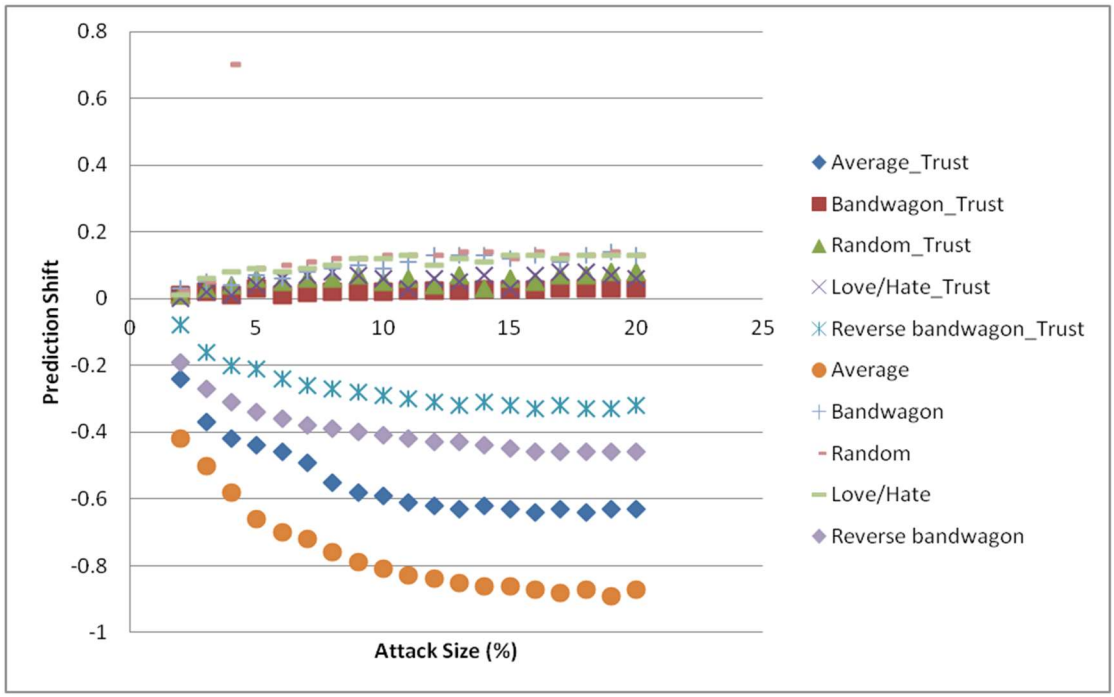

Figure 8. Prediction shifts achieved by nuke attacks against the item-based algorithm.

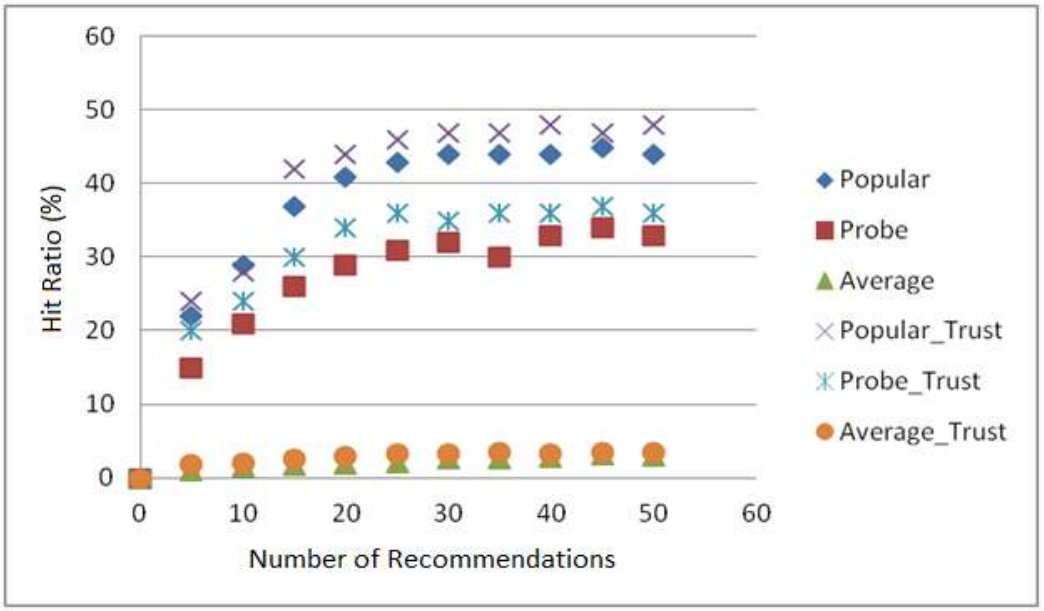

Figure 9. Hit ratios achieved by the popular, probe and average nuke attacks against the user-based algorithm.

In Figure 3, prediction shift is observed against the percentage attack size for bandwagon, average and random attacks. Here, product push attacks are carried out against a user-based algorithm. There is generally a lower prediction shift when a trust-enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 4, hit ratio is observed against the number of recommendations for average, random, bandwagon and baseline attacks. Here, product push attacks are carried out on user-based algorithm. There is generally a lower hit ratio when a trust enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 5, prediction shift is observed against the percentage attack size for in segment and all user attacks. Here, product push attacks are carried out against item-based algorithm. There is generally a lower prediction shift when a trust enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 6, hit ratio is observed against the number of recommendations for in segment, all user and baseline attacks. Here, product push attacks are carried out against item-based algorithm. There is generally a lower hit ratio when a trust enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 7, prediction shift is observed against the percentage attack size for average, bandwagon, random, love/hate and reverse bandwagon attacks. Here, product nuke attacks are carried out against user-based algorithm. There is generally a lower prediction shift when a trust-enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 8, prediction shift is observed against the percentage attack size for average, bandwagon, random, love/hate and reverse bandwagon attacks.

Here, product nuke attacks are carried out against item-based algorithm. There is generally a lower prediction shift when a trust enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

In Figure 9, hit ratio is observed against the number of recommendations for popular, probe and average attacks. Here, product nuke attacks are carried out against user-based algorithm. There is still generally a higher hit ratio when a trust enhanced version of the system is used as opposed to its classical counterpart. This means that the trust enhanced version outperforms the classical one for this case.

The importance of these findings is confirmation of value in incorporating the trust adjustment factor into the pipeline of the recommender system. It improves robustness and therefore helps in mitigating potential abuse.

It is, therefore, recommended that the trust adjustment factor be incorporated into the pipeline of artificial intelligence driven recommender system. This will aid in averting potential fraud. The fraud might lower the user experience of the e-commerce shopping process, when unworthy items are recommender to the user. It may also make the user to either spend more time looking for the right item or just abandon online shopping. Some user may be forced to meet locomotion costs of going to buy the item in brick and mortar shops, which is an additional expense. Some users may also just settle for inferior items because the suitable ones have been hidden from them. This prevents such users from getting best value for their money.

These results are relevant because they aid to conclude different pieces of related work which have tried to bridge the gap mentioned herein. Burke et al. [9] identified the room for potential abuse of recommender systems and proposed further research on this area. Ngwawe et al. [26, 27] then worked on finding out factors that predict trustworthiness of e-commerce platform and engineered a trust model. They then incorporated this trust model into the recommender system pipeline and confirmed its positive impact on recommender system accuracy [28]. They were also cognizant of the fact that some properties of recommender systems do trade with each other.

For example, the goal to improve recommender system's accuracy property may lower its serendipity and robustness properties. Ngwawe et al. [26, 27] recommended further study to empirically evaluate the impact of this new trust model on both the serendipity and robustness. Our work fills in the gap as far as robustness is concerned. Further research is proposed to investigate the impact of this new trust model, called trust adjustment factor on the serendipity and other properties. Other properties of recommender systems, including serendipity, are described by Shani and Gunawardana [33].

The key limitation of using this trust adjustment factor is that it increases the required computing power. This is because it adds an additional computing process into the pipeline. This will be a great deal if recommendations were to be computed for each user in an e-commerce platform serving millions of users. This extra cost will, however, be taken care of by the improved user experience in the shopping process, which leads to customer satisfaction. It is expected that customer satisfaction will lead to customer retention which naturally has a positive impact in revenue.

5. Conclusion

It has been demonstrated that a trust enhanced collaborative filtering recommendation system outperforms the classical system in terms of robustness. Trust adjustment factor was incorporated into the pipeline of artificial intelligence driven recommender system. Robustness was found to improve by up to 52%

when measured by prediction shift and by up to 18% when measured by hit ratio

This trust adjustment factor can help to filter out inferior items from the output of the recommendation process in a product promotion attack. It may also help to reduce chances of an attacker succeeding to demote a superior product maliciously in a product nuke attack.

Further work is proposed to investigate the impact of trust adjustment factor on other properties of recommender systems, such as the serendipity property.

Table 2 Summary of Maximum Percentage Changes After Adding Trust Adjustment Factor.

Intention of attack	Parameter to Measure	Type of Collaborative Filtering Algorithm	Type of Attack that is most mitigated	Original Value	Value After Adding Trust Parameter	Change	Percentage Change
Product Promotion	Prediction Shift	User-based	Average	0.56	0.5	0.06	10.71429
Product Promotion	Prediction Shift	Item- based	In segment	0.47	0.41	0.06	12.76596
Product Promotion	Hit Ratio	User- based	Average	77	73	4	5.194805
Product Promotion	Hit Ratio	Item- based	In segment	43	35	8	18.60465
Product Demotion	Prediction Shift	User- based	Reverse Bandwagon	-0.88	-0.42	-0.46	52.27273
Product Demotion	Prediction Shift	Item- based	Average	-0.87	-0.63	-0.24	27.58621
Product Demotion	Hit Ratio	User- based	Popular	44	48	4	8.333333

CONTRIBUTIONS OF CO-AUTHORS

Edwin Ngwawe	ORCID: 0000-0002-1324-9628	Conceived the idea, conducted experiments and analyzed results and wrote the paper
Elisha Abade	ORCID: 0000-0002-8978-4213	Overall supervision and guidance in the research process
Stephen Mburu	ORCID: 0000-0002-7500-9312	Overall supervision and guidance in the research process

REFERENCES

[1] Stuart J. B., *Information management research and practice in the post-COVID-19 world*, International Journal of Information Management, **55**, 2020.

[2] United Nations Conference on Trade and Development, *How COVID-19 triggered the digital and e-commerce turning point*, 15 March 2021. [Online]. Available: https://unctad.org/news/how-covid-19-triggered-digital-and-e-commerce-turning-point. [Accessed 02 May 2023].

[3] Aljukhadar, M., Senecal, S., Daoust, C.,*Using Recommendation Agents to Cope with Information Overload*, International Journal of Electronic Commerce, **17**(2): p. 41-70, 2014.

[4] Ricci, F.,. Rokach, L., hapira , B., *Recommender Systems Handbook*, New York: Springer, 2011.

[5] Jordan, M. I., Mitchele, T. M., *Machine learning: Trends, perspectives, and prospects*, Sicence, **349**: p. 225-260, 2015.

[6] Etzioni, A., Etzioni, O., *Incorporating Ethics into Artificial Intelligence*, The Journal of Ethics, **21**: p. 403-418, 2017.

[7] Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, 22 March 2023. [Online]. Available: https://futureoflife.org/open-letter/pause-giant-ai-experiments/. [Accessed 01 April 2023].

[8] Yudkowsky, E., *Pausing AI Developments Isn't Enough. We Need to Shut it All Down*, 29 March 2023. [Online]. Available: https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/. [Accessed 19 April 2023].

[9] Burke, R,. O'Mahony, M. P., Hurley, N. J., *Robust Collaborative Recommendation*, in Recommender Systems Handbook,, F. Ricci, L. Rokach and S. Bracha, Eds., New York, Springer, p. 961-995, 2015.

[10] DW Documentary, *Amazon, Jeff Bezos and collecting data | DW Documentary*, DW Documentary, 10 May 2019. [Online]. Available: https://www.youtube.com/watch?v=O90PShJVu58. [Accessed 15 April 2023].

[11] Ben, T., Dominick, R., *Walmart is closing a batch of stores in 2023 — here's the full list*, 11 April 2023. [Online]. Available: https://sports.yahoo.com/walmart-closing-batch-stores-2023-161826876.html. [Accessed 02 May 2023].

[12] Bogue, R., *Growth in E-commerce Boosts Innovation in the Warehouse Robot Market*, Industrial Robot: An International Journal,**43**(6): p. 583-587, 2016.

[13] Yin, C., Wang, J., Park, J. H., *An Improved Recommendation Algorithm for Big data Cloud Service based on the Trust in Sociology*, Neurocomputing, **256**: p.49-55, 2017.

[14] Leskovec, J., *Epinions social network*, 1 January 2003. [Online]. Available: https://snap.stanford.edu/data/soc-Epinions1.html.

[15] Yasmin, A., Tasneem, S., Fatema, K.,*Effectiveness of Digital Marketing in the Challenging Age: An Empirical Study*, International Journal of Management Science and Business Administration, **1**(5): p. 69-80, 2015.

[16] Ziheng, C., Fabrizio, S., Jia, W., Yongfeng , Z., Gabriele, T., *The Dark Side of Explanations: Poisoning Recommender Systems with Counterfactual Examples*, in 46th International ACM SIGIR Conference on Research and Development in Information Retrieval *(SIGIR '23)*, Taipei, p. 2426–2430, 2023.

[17] Mingdan, S., Qingshan, L., *Shilling attacks against collaborative recommender systems: a review,* Artificial Intelligence Review, **53**: p. 291–319, 2018.

[18] Shuai, Z., Lina, Y., Aixin, S., Yi, T., *Deep Learning Based Recommender System: A Survey and New Perspectives*, ACM Comput. Surv., **52**(1): p. 1-38, 2019.

[19] Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos,, M., Ghosh, R.., *Exploiting Burstiness in Reviews for Review Spammer Detection,* in Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 7(1): p. 175-184, 2021.

[20] Manqing, D., Feng, Y., Lina, Y., Xianzhi, W., Xiwei, X., iming, Z., *A survey for trust-aware recommender systems: A deep learning perspective*, Knowledge-Based Systems, **249**, 2022.

[21] Cornière, A., *Search Advertising*, American Economic Journal: Microeconomics, **8**(3): p. 156-88, 2016.

[22] Athey, S., Nekipelov, D., *A Structural Model of Sponsored Search Advertising Auctions*, in Sixth ad auctions workshop, New Haven, 2010.

[23] Narayanan, S., Kalyanam, K., *Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach*, Marketing Science,**34**(3): p. 309-472, 2015.

[24] Aggarwal, G., ., Muthukrishnan, S.,, Pál, D., Pál M., *General auction mechanism for search advertising*, in WWW '09: Proceedings of the 18th international conference on World wide web, Madrid, p. 241–250, 2009.

[25] Ghose, A., Yang, S., *An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets*, Marketing Science, **55**(10): p. 1605-1622, 2009.

[26] Ngwawe, E. O., Abade, E. O., Mburu, S. N,. *Predicting Trustworthiness of an E-commerce Platform from the Consumer Perspective*, Indian Journal of Computer Science, **7**(4): p. 16-30, 2022.

[27] Ngwawe, E. O., Abade, E. O., Mburu, S. N,. *Estimating Perceived Risk in E-Commerce Platforms from the Consumer Perspective*, World Journal of Engineering Research and Technology, p. 39-53, 2022.

[28] Ngwawe, E., Abade, E., Mburu S., *Trust Enhanced Collaborative Filtering Recommendation Algorithm*, International Research Journal for Computer Science, p. 88-96, 2023.

[29] Shafer, D. S., Zhang, Z., Introductory Statistics, Saylor Foundation, 2012.

[30] Kwak, S. G., Kim, J. H., *Central limit theorem: the cornerstone of modern statistic*s, Korean Journal of Anesthesiology, **70**(2): p. 144-156, 2017.

[31] Dorfman, K., *Comparing Means: The t-Test,* UMass biology Department, 2019.

[32] Narges, H., Parham, M., Abbas, K., *An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems*, Knowledge-Based Systems, **256**, 2022.

[33] Shani, G., Gunawardana, A., *Evaluating Recommendation Systems*, in Ricci F., Rokach L., Shapira B., Kantor P. B., (eds.) Recommender Systems, Springer, Boston, 2011.