International Journal of Security and Its Applications
Vol.8, No.4 (2014), pp.273-286
http://dx.doi.org/10.14257/ijsia.2014.8.4.25

Shilling Attack Detection Algorithm based on Genetic Optimization

Tao Li

Information Engineering Department, Jilin Police College
Jilin 130032, China
Taolili2014@126.com

Abstract

Aiming at the low limitation of shilling attack detection technology unsupervised degree, this paper takes the group effect attack profile as the breakthrough point to construct the attack profile groups and the corresponding genetic optimization objective function of quantitative measure of the effects, and prove that the maximum value of the objective function in the ideal state marks the optimum detection effects in ideal situation. On this basis, the combination of genetic optimization process will be adaptive parameter posterior inference and objective function, and proposes the Iterative Bayesian Inference Genetic Detection Algorithm (IBIGDA).Experimental results show that IBIGDA can effectively detect shilling attacks of typical types even in lack of the system or attack-related prior parameters. IBIGDA algorithm can detect common shilling attack, unsupervised degree is high, with the actual application requirements.

Keywords: *shilling attack, IBIGDA algorithm, shilling attack detection*

1. Introduction

Considering the low degree of unsupervised feature that the existing shilling attack detection techniques suffer from, Committed to the development of unsupervised degree higher attack detection technology, in order to reduce the dependence degree detection performance of the prior knowledge of the system. Supervised detection technology only has theoretical value, because practice is difficult to tailor the training set to construct the classifier [1-2]. Unsupervised detection technology is both less prior demand and stronger generalization ability, with the actual situation of the shilling attack defense. However, Unsupervised is only a relative concept, it is essential for a small amount of prior input, and a priori knowledge accurate is usually determined key detection performance[3-4]. For example, EMSVD and PLSA detection algorithms need to enter the number of user classes. PCA Var Select algorithm need to know accurate attack strength and this significant impact on the detection performance parameters are usually unable to accurately obtain weakened unsupervised feature of these algorithms [5-6].

To further reduce dependence on an unsupervised algorithm to detect a priori input to improve the algorithm of unsupervised; this paper proposes the Iterative Bayesian Inference Genetic Detection Algorithm (IBIGDA)

a. According to real users and attackers on a large scale ratings rule, to examine the correlation between the user profile distribution characteristics, deriving the group effect measure quantitative based on the generalized variance attacker.

b. On the basis of 1, construct genetic optimization objective function, the maximum value of this function marks the optimal detection results under ideal conditions.

ISSN: 1738-9976 IJSIA

c. Parameter Bayesian inference ideas is thought into genetic optimization process, depending on the objective function is adaptive parameter, according to the results of each iteration to optimize test update, and as a priori parameters attend the next iteration, until achieving the best detecting effect.

The experimental results show that IBIGDA algorithm can achieve better detection level in a variety of common attack configuration, and without accurate a priori input.

2. Iterative Bayesian Inference Genetic Detection Algorithm (IBIGDA)

This section will introduce IBIGDA algorithm, including the introduction of a quantitative measure of the attack profile of the group effect. Genetic optimization objective function is established; the algorithm specific processes related other content. Before the discussion, any user profile u_i are subject to pretreatment. There are two steps pretreatment: First, use 0 instead of u_i missing values to get $u_i^{'}$ .secondly, the normalization $u_i^{'}$ to get $u_i^{''}$. $u_i^{''T} u_i^{''} = 1$, $1^T u_i^{''} = 0$. Convenience, it still u_i represents $u_i^{''}$.

2.1. The Statistical Characteristics of Exists Attack between the User Profile Attack

Now MovieLens100K data set is (the experimental part of this paper) Injection into the random attack. Parameter configuration $p^{att} = 10\%$ $p^{fill} = 8\%$, Figure 1 shows the distribution of the correlation coefficient between the user profiles. A, B, C respectively represent the correlation coefficient distribution among real profile, attack profile, real and attack profile. After pretreatment, easy to push any score matrix $R_{I \times J}$ two user profile u_i u_j the correlation coefficient equal to their Cosine in J dimensional space angle $\rho(u_i, u_j) = \cos(u_i, u_j)$. Figure 1 shows the distribution of A set in about 0.1, and the distribution of B, C concentrated in about 0. It shows that the angle between the real user profiles is an acute angle, and the attack is approximately perpendicular to each other. Further experiments showed that this overview of the angular relationship under the same configuration parameters in mean attack and popular attack scenario. Intuitively, if J dimensional space $[e_1, e_2, ..., e_J]$ represents the J point of unrelated interest. There is interest relationship between real users, and the attacker's interest is relatively decentralized. There is a reasonable explanation of this phenomenon [7-8]. Real user ratings object is generally more popular items, rather than the popular item was rated a little, so degree of coincidence between profiles are high. Fill attacker randomly selected items from each item to get equal probability score, so degree of coincidence is low.

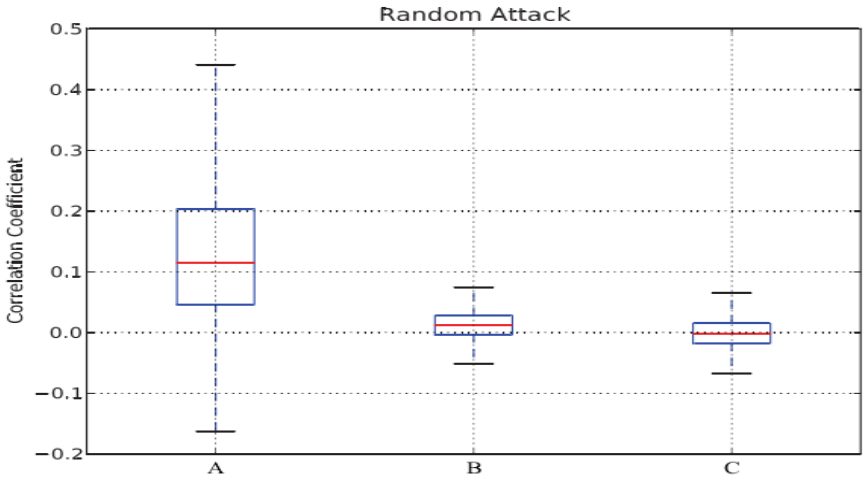

Figure 1. The Distribution of the Correlation Coefficient between the User Profiles Random Attack

To simplify the discussion of the problem, according to the statistical characteristics between user profiles, given the definition of the ideal situation:

Definition.1: set K^f K^t were pretreated attack profile and real user profile, ideally, there are:

$$\begin{cases} \cos(u,v) = 0 \quad u,v \in K^f, u \neq v \\ \cos(u,v) = 0 \quad u \in K^f, v \in K^t \\ \cos(u,v) = 0 \quad u,v \in K^f, u \neq v, 0 < a < 1 \end{cases} \tag{1}$$

2.2. Generalized Variance Induced Attack Profile Group Effect Metric

Attack model determines the dispersion of a single attack profile points of interest, this is an overview of the individual effects of the attack, as well as individual user profiles suspects index. On the one hand, using only the individual effects profile attack detection is not rigorous, because a small amount of interest in a wide range of real users also have this effect, and these user are conducive to promoting the new recommendation. On the other hand, if the individual effect of user profiles is more than a certain number, and coincidence degree between interests is the low, then the group effect is so that we have reason to suspect the shilling attacks appeared in the system. As this thought into detection means, need to make quantitative description of the group effect, generalized variance provides a good solution. Generalized variance is defined as the determinant of the covariance matrix. Let all users in the recommendation system are composed of a multivariate random variables, each item rating is the random variable observation value. After pretreatment score matrix and covariance matrix has the following relationship

$$S_{I\times I} = \frac{1}{J-1} R_{I\times J} R_{I\times J}^T \tag{2}$$

Generalized variance is $\det(S_{I\times I})$

Generalized variance has excellent geometric interpretation. Let I user profile in a J -dimensional hyper volume is V . $\det(S_{I\times I})$ and V satisfies the relationship

$$\det(S_{I\times I}) = (J-1)^{-I} V^2 \qquad (3)$$

Formula3 into the formula2 is equal to

$$\begin{aligned} &\det(\frac{1}{J-1} R_{I\times J} R_{I\times J}^T) = (J-1)^{-I} \\ &\Rightarrow (J-1)^{-I} \det(R_{I\times J} R_{I\times J}^T) = (J-1)^{-I} V^2 \\ &\Rightarrow V = \sqrt{\det(R_{I\times J} R_{I\times J}^T)} \\ &\Rightarrow V = \sqrt{\det(X_{I\times I})} \quad X_{I\times I} = R_{I\times J} R_{I\times J}^T \end{aligned} \qquad (4)$$

Easy to prove, after pretreatment, when the same ultra-flat all user profiles located, the generalized variance is 0, hyper-volume is minimum. When all user profile perpendicular to each other, the generalized variance is maximum, hyper-volume is maximum. Because the user profile length is 1, so the hyper-volume maximum value is 1.

For example:

$$R_{3\times 3} = [u_1, u_2, u_3]^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

As $u_1 = u_2 + u_3$, the u_1, u_2, u_3 coplanar, as shown in Figure 2, they surrounded hyper-volume:

$$V = \sqrt{\det(X_{I\times I})} = \sqrt{dept\left\{\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}\right\}} = 0$$

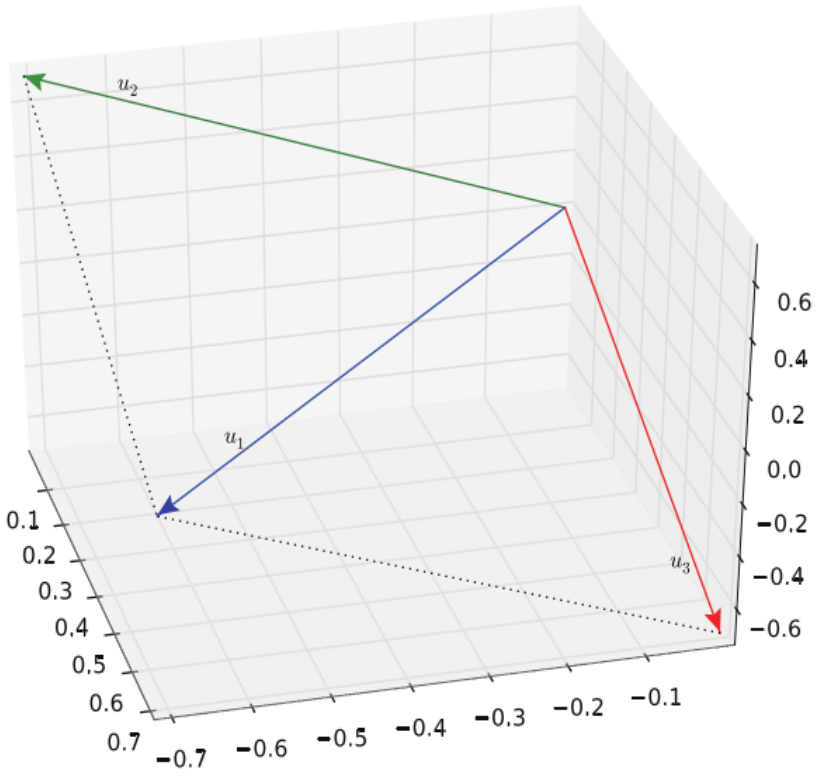

Figure 2. Coplanar Vectors u_1, u_2, u_3

Because the attack profile perpendicular to each other, they surrounded hyper-volume is $V = 1$. While the angle between real profiles is an acute angle, so they surrounded the hyper-volume $V \in [0,1]$. However, hyper-volume as a scalar with dimensions, hyper-volume of surrounded by different number of profiles is with different dimension level, there is no comparability. To solve this problem, define average contribution degree for hyper-volume ($advo$) function:

Definition.2: Let K be a subset of user profile in recommendation system, the enclosed volume is V and $|K| = I$ the average hyper-volume contribution function $advo(\cdot)$ is defined as: $advo(\cdot) = \sqrt[I]{V} = \sqrt[2I]{\det(X_{I \times I})}$

This paper will $advo$ as a quantitative measure of the attack profile group effect. From the above discussion, when $|K|$ is higher than a certain level, $Advo(K)$ value increases with the increasing ratio of the attack profile. In particular, when K is attack profile, $Advo(K)$ maximum value is 1. So $Advo(K)$ is a good index of concentration profile attack in K, it will serve as the basis for the construction of IBIGDA algorithm

2.3. Genetic Optimization Objective Function

Definition 3: set recommendation system has user I, Indication Vector $IV = [sign_1 \, sign_2, ..., sign_I]^T \quad \forall sign_i \in \{+, -\}$, if the IV in the j elements $IV[j] = +$, indicates that the user j is regarded as the attacker. Apparently the user j belongs to the true

or false positive. p_{IV}^{t+} is the proportion of real, p_{IV}^{f+} is the proportion of false. $p_{IV}^{+} = p_{IV}^{f+} + p_{IV}^{t+}$ .If IV overview of the entire positive user composed set K^{+} . $Advo(IV) = Advo(K^{+})$.

Using $Advo$ attack detection is a typical combinatorial optimization problem. The purpose of optimization is to maintain $Advo(IV) = 1$ .So that the relationship $p_{IV}^{+} = p_{IV}^{f+} + p_{IV}^{t+}$ is established. Which p^{ture} as an attacker to account for all the proportion of users. This paper chooses genetic algorithm as optimization method, need to introduce IV as the objective function of variables $g^{obj}(\cdot)$, so that when $g^{obj}(\cdot)$ is the maximum to achieve the best detection results.

Function of $g^{obj}(\cdot)$ should have two capabilities, First, as far as possible to maintain $Advo(IV)$ is equal to the maximum value of 1. Second, limit p_{IV}^{+} excessive deviation of p^{ture} . The median of p^{ture} Cauchy distribution density function is as the penalty factor.

$$\mathrm{Penalty}(p_{IV}^{+} \mid p^{ture}, \sigma) = \frac{1}{\sigma\pi\left(1 + \left(\frac{p_{IV}^{+} - p^{ture}}{\sigma}\right)^{2}\right)}$$

Where, σ is the expansion factor of Cauchy distribution, the Cauchy distribution instead of normal distribution density function is because Cauchy distribution density function in the form of more simple, to reduce the amount of calculation.

In order to explore the specific expression of $g^{obj}(IV)$, suppose IV is a random variable, from the view of probability, $p(IV \mid X) \propto p(IV \mid X)p(IV)$. If $Advo(IV)$ and $\mathrm{Penalty}(p_{IV}^{+} \mid p^{ture}, \sigma)$ are respectively corresponding to $p(X \mid IV)$ and Priori $p(IV)$. $g^{obj}(IV)$ is a posteriori $p(IV \mid X)$. The optimization objective is equivalent to find IV the maximum a posteriori estimation of IV^{MAP} . From the above, so

$$g^{obj}(IV \mid p^{ture}, \sigma) = Advo(IV) \cdot \mathrm{Penalty}(p_{IV}^{+} \mid p^{ture}, \sigma)$$

As $p_{IV}^{+} = p_{IV}^{t+} = p^{ture}$, detection results achieve the best. According to the formula 2 and the penalty factor expression. $Advo(IV)$ and $\mathrm{Penalty}(p_{IV}^{+} \mid p^{ture}, \sigma)$ respectively to get maximum value. Thus $g^{obj}(IV \mid p^{ture}, \sigma)$ reached the maximum.

2.3. Algorithm Description and Interpretation

The following is a description of the algorithm IBIGDA

Input. Genetic optimization objective function parameters $p_{(0)}^{'}$ and $\sigma_{(0)}$, priori proportions of attack profile and expansion factor of penalty function

Output. The best individual IV^{best}

Step1. Pretreatment evaluation matrix

Step2 $n \leftarrow 0$, initial population $IV_{(0)}^{[1]} \square IV_{(0)}^{[k]}$, k is the population base.

Step3 Start n-th iteration.

Step3.1. Taking $IV_{(0)}^{[1]} \square IV_{(0)}^{[k]}$ as the initial population, for $g^{obj}(IV \mid p^{ture}, \sigma_{(n)})$ carried on genetic optimization, until convergence, get populations $IV_{(n+1)}^{[1]} \square IV_{(n+1)}^{[k]}$.

Step3.2 Take the best individual IV^{best}, Posteriori update $p_{(n+1)}^{'} \leftarrow p_{IVbest}^{+}, \sigma_{(n+1)} \leftarrow \bar{\sigma}(IV^{best})$

Step3.3 If $| p_{(n+1)}^{'} - p_{(n)}^{'} | < \varepsilon$, return IV^{best}, the end of execution. Else $n \leftarrow n+1$, go to setp3.

In IBIGDA algorithm, because p^{ture} is generally small, the $p_{(0)}^{'}$ takes a larger value, without any prior knowledge. The $p_{(0)}^{'}$ value is close to p^{ture}, the number of iterations is less. It should be emphasized that $p^{'}$ and σ are updated after each iteration of the algorithm. This is the Countermeasures for the non-ideal conditions.

Ideally, Equation 2 determines IBIGDA algorithm to get the best detection results when the value of the fixed σ. Meanwhile, the size of the σ values will have an impact on the test values of $p^{'}$. σ is smaller, the deviation $p^{'}$ punishment is bigger. The posterior value is close to $p^{'}$. Whereas it is close to p^{ture}.So choose the larger σ can speed up the convergence of the algorithm.

However, the real system is running in non-ideal circumstances, although the differences are not significant in both cases. But this time the two vertical attack profiles is subset of all attack profile. So it may be in a certain iterations, appears $p_{IV}^{+} < p^{true}$ cross-border cases, influence of effect of detection algorithm. In this regard, IBIGDA algorithm will also take σ regarded as an adaptive parameters, in the p_{IV}^{+} far from the p^{true}, using larger σ value, Speed up the p_{IV}^{+} to the approximation rate of p^{true}. While p_{IV}^{+} near p^{true}, in order to control the cross-border degree. We must reduce σ value gradually. Approximation degree can $Advo(IV)$ measured, $Advo(IV)$ is bigger, show that p_{IV}^{+} is closer to p^{true}, so σ is $Advo(IV)$ function. Under non ideal conditions, $Advo$ value attack profile is less than 1, so $Advo(IV)$ at close to 1, σ to speed reduced to try to control the degree of cross-border. In IBIGDA algorithm $\bar{\sigma}(IV) = (1 - Advo(IV))^{\frac{1}{2}}$.

3. Experiment Design and Discussion

4.1. Data Sets and Experimental Setup

The experiment adopts Group Lens research group Minnesota University published and widely used two data sets, MovieLens100K and MovieLens1M. They are the two most commonly used benchmark data sets in the field, as shown in Table 1. For the MovieLens100K data set, the experimental adopts all users and items. For the MovieLens1M data set, the experimental adopts all items. But random is chosen 1/4 (about 1510) user. In the IBIGDA algorithm the relevant parameters value is $p_{(0)}^{'} = 0.3 \quad \sigma_{(0)} = 0.12 \quad \varepsilon = 0.004$

Table 1. Dataset Overview

Dataset	Number of users	Number of item (film)	Final score	Score range
MovieLens100K	943	1682	100000	1-5
MovieLens1M	6040	3900	1000209	1-5

In order to detect the ability to validate IBIGDA algorithm, assumes that the original user is real users in the dataset. In the different strength attack p^{att} and filling fraction p^{fill}, respectively to the dataset into three attacks, Random attack, mean attack and popular attack. Among them, the popular attack filling selects single items of popular highest degree.

Evaluation of shilling attack detection effect adopts the correct rate f^{pre}, recall rate f^{rec}, the comprehensive index F value. Let IBIGDA method returns IV^{best}, then

$$\begin{cases} f^{pre} = \dfrac{p_{IV^{best}}^{t+}}{p_{IV^{best}}^{+}} \\ f^{rec} = \dfrac{p_{IV^{best}}^{t+}}{p^{true}} \\ F = \dfrac{2 f^{pre} f^{rec}}{f^{pre} + f^{rec}} \end{cases}$$

4.2. The Example Analysis of Shilling Attack Detection Process

Figure 2 shows typical attack detection process of IBIGDA algorithm on the dataset of MovieLens100K. This data set was injected into $p^{att} = 12\%$, $p^{fill} = 6\%$ random attack. The dotted line corresponds to the time to update the adaptive parameters; the dashed line is the optimization process for the updated target genetic posterior function. The algorithm termination condition is reached in the twelfth iteration, return results. Further experiments showed that different attack models and parameters configuration, by monitoring the changes of p_{IV}^{+} and p_{IV}^{t+}, IBIGDA algorithm to get the attack detection process is similar to Figure 3.

In Figure 3, although in a non-ideal situation, but the approximation shows the detection process ideal. As can be seen, p_{IV}^{+} in σ domination, slowing down the rate of approximation p^{true}, when the program exits this rate falls below a certain extent, in order to limit cross-border.

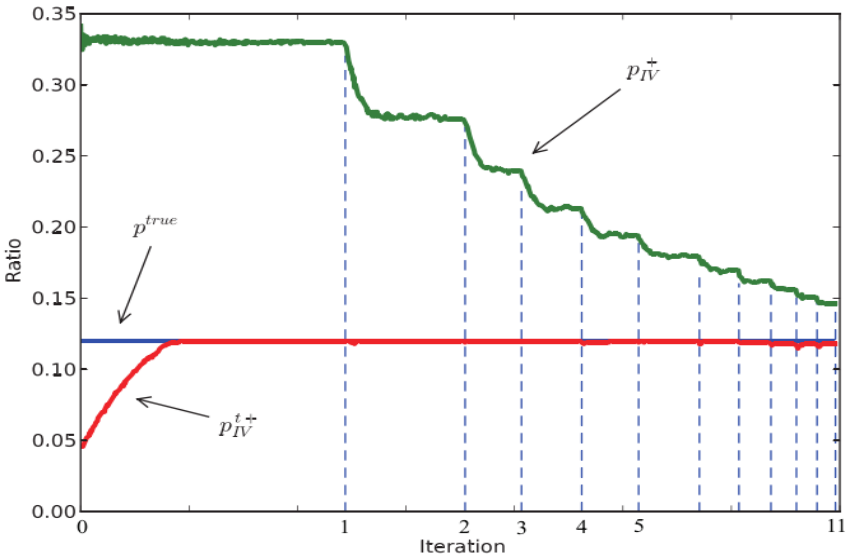

Figure 3. Detection Process of Typical Attack

Intuitively, IBIGDA algorithm consists of two processes of interaction, one is to absorb attacks overview process that aims to achieve and maintain $p_{IV}^{t+} \approx p^{true}$. The other is the real user profiles emptying process that aims to reduce the false positive rate $p_{IV}^{f+} = (p_{IV}^{+} - p_{IV}^{t+})$. The two process runs through the beginning of algorithm execution. In Figure 3, the initial iteration can be seen obvious absorption process, by continuously into the attack profile, p_{IV}^{t+} in the initial iteration of the interim had been close to or equal to p^{true}. And after each update adaptive parameter, p_{IV}^{+} decreased rapidly, to move closer to the p^{true}, display the emptying process significantly. When the last few iterations, p_{IV}^{t+} has a slight downward trend, this is due to the presence of a small amount of interest specific real user in the real system, their ratings have more aggressive behavior, causing the emptying process wrongly ruled out some weak offensive attack profile. Thus, the IBIGDA algorithm is chose to withdraw when p_{IV}^{+} rate below a certain degree. Not only limits the cross-border, also happens to stop the decline of p_{IV}^{t+}

4.3. Detection Effect of Shilling Attack

First, this paper IBIGDA algorithm of comprehensive evaluation attacks detection ability on the MovieLens100K dataset, and the detection performance compared with PCA Var Select, PLSA, UnRAP and EMSVD algorithm. At present, the PCA Var Select algorithm has the best performance in the detection of Movie Lens datasets. The experiment adopted $3 \times 4 \times 5$ design patterns (random attack, mean attack, popular attack); the different combination of attack strength $p^{att}(5\%, 7\%, 10\%, 12\%, 15\%)$ and filling of $p^{fill}(3\%, 6\%, 9\%, 12\%, 15\%, 20\%)$ corresponds to a group of experimental configuration.

Each configuration of the experimental results obtained from ten independent experiments mean.

Table 2-4 shows the effect of detection of IBIGDA algorithm. Only in the face of filling rate of popular attack 3%, algorithm of detection ability is limited. Because the choice of filling is a large proportion, so the attack profile is similar to actual situation, the attack characteristics is not obvious. In other cases, the algorithm has better detection performance, especially the recall rate is more than 90%, shows that the IBIGDA algorithm can detect most attacks. With the attack strength and fill rate increases, the suspects of the attack profile are even more significant, thus the detection performance of the algorithm has enhanced the trend.

In addition, it was found that when no injection attacks, IV^{best} has only 25 users. Since the detection accuracy of the algorithm is less than 1, so the real user necessarily is more less, algorithm can show attack of not exist in a certain degree.

In general, in order to achieve higher recall rate, need to sacrifice accuracy for a certain price. Due to a large user base recommendation system, some false positive error shielding the user does not have a significant impact on the results recommended in recommendation process.

Table 2. Accuracy Rate and Recall Rate of Random Attack Detection (MovieLens100K)

p^{fill}	3%		6%		9%		12%		15%		20%	
p^{att}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}
5%	0.58	0.91	0.62	1.00	0.60	1.00	0.64	1.00	0.62	1.00	0.64	1.00
7%	0.70	1.00	0.67	1.00	0.67	1.00	0.65	1.00	0.71	1.00	0.71	1.00
10%	0.75	0.97	0.78	1.00	0.78	1.00	0.74	1.00	0.74	1.00	0.76	1.00
12%	0.80	0.95	0.82	0.95	0.81	0.99	0.78	1.00	0.81	0.99	0.80	1.00
15%	0.84	0.90	0.85	0.95	0.83	0.99	0.86	0.99	0.85	0.99	0.86	0.99

Table 3. Accuracy Rate and Recall Rate of Mean Attack Detection (MovieLens100K)

p^{fill}	3%		6%		9%		12%		15%		20%	
p^{att}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}
5%	0.63	0.85	0.65	0.98	0.65	1.00	0.64	1.00	0.69	1.00	0.65	1.00
7%	0.69	0.86	0.68	0.98	0.73	0.98	0.70	1.00	0.71	0.97	0.74	1.00
10%	0.78	0.85	0.91	0.91	0.80	0.97	0.81	0.98	0.82	0.99	0.81	0.97
12%	0.79	0.82	0.90	0.90	0.85	0.96	0.85	0.98	0.85	0.97	0.84	0.97
15%	0.80	0.69	0.84	0.84	0.86	0.89	0.89	0.94	0.86	0.89	0.88	0.96

Table 4. Accuracy Rate and Recall Rate of Popular Attack Detection (MovieLens100K)

p^{fill}	3%		6%		9%		12%		15%		20%	
p^{att}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}	f_{pre}	f_{rec}
5%	0.57	0.68	0.59	1.00	0.59	1.00	0.63	1.00	0.59	1.00	0.62	1.00
7%	0.64	0.77	0.70	0.98	0.72	1.00	0.70	1.00	0.68	1.00	0.69	1.00
10%	0.68	0.61	0.79	0.95	0.78	0.99	0.99	1.00	0.78	1.00	0.78	1.00
12%	0.67	0.57	0.81	0.94	0.82	0.99	0.99	0.99	0.82	1.00	0.81	1.00
15%	0.73	0.59	0.83	0.87	0.84	0.97	0.97	0.96	0.84	0.97	0.97	0.99

PCA Var Select detection algorithm to achieve the best performance on the premise that the attack must be informed of the exact strength; otherwise it will seriously reduce the accuracy or recall [9]. Figure 4 shows in three attack model of the $p^{att} = 10\%$ $p^{fill} = 6\%$, comparison of detection performance of PLSA, EMSVD, IBIGDA and UnRAP algorithm. In the number of different user categories PLSA algorithm performance changes significantly, intermediate high on both sides of the lower trend, showing strong sensitivity to prior knowledge. Performance EMSVD algorithm also varies with the number of categories of users showed some fluctuations. The sensitivity is not significant, but the detection ability is not very ideal. Especially when is in the face of the mean attack, the algorithm is failure. In fact, the EMSVD algorithm is only effective for random attack high filling rate. Limitations are larger, because the attack profile adopts rarely high filling rate. This will increase the attack cost, and may reduce the effect of the attack. In practical application, regardless of whether the attack strength or the user number of categories is not obtained in advance. The IBIGDA algorithm has been used the parameters of the same group input in the experiment. Without accurate a priori knowledge, without the supervision of strong. The UnRAP algorithm and IBIGDA algorithm in unsupervised equal, its detection performance is better than that of the latter. However, in the Un-coordinated attack case, this advantage does not exist. Un-coordinated attack is variety of shilling attacks exist at the same time, and each target item attacks generally is different from each other. Shilling attack in real environment is a significant part of the Un-coordinated attack. The limitations of UnRAP algorithm in each detection can only focus on a single target, thus unable to effectively deal with the situation of multi target. While the IBIGDA algorithm is based on the group effect, without considering the target item.

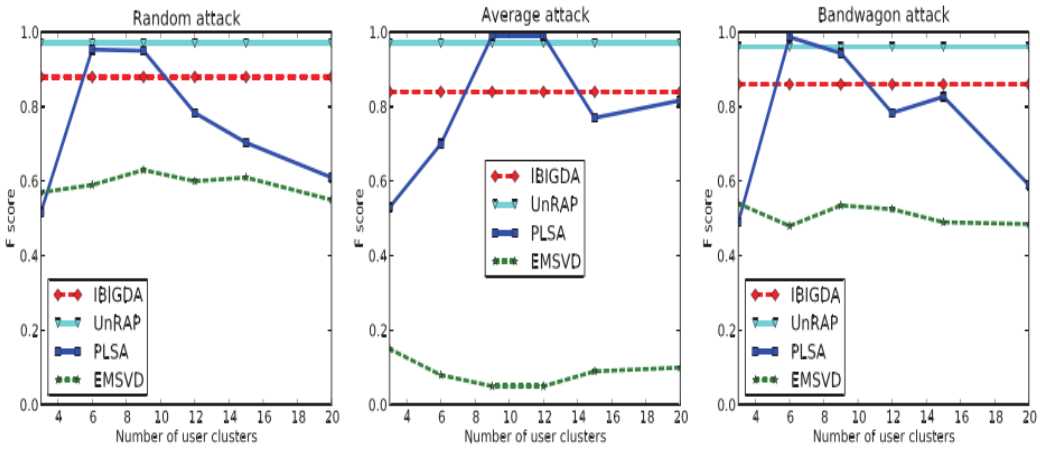

Figure 4. Detection Performance is Dependent on Input Parameters

Figure 5 shows the two algorithms for the Un-coordinated attack detection capability, configuration parameters for each attack are $p^{att} = 5\%$ $p^{fill} = 6\%$. It is easy to see, in the entire attack scenario, the IBIGDA algorithm shows balance. The F value is in more than 86%, the detection performance is much better than UnRAP algorithm. Especially in the three attack superposition, there are three different target items; IBIGDA algorithm has more significant advantages. In fact, the number of attack superposition is more the performance of UnRAP is low. And the IBIGDA algorithm can effectively detect the Un-coordinated attack.

The detection performance in single objective and multi-objective case has reached high level, practical value is better than UnRAP algorithm.

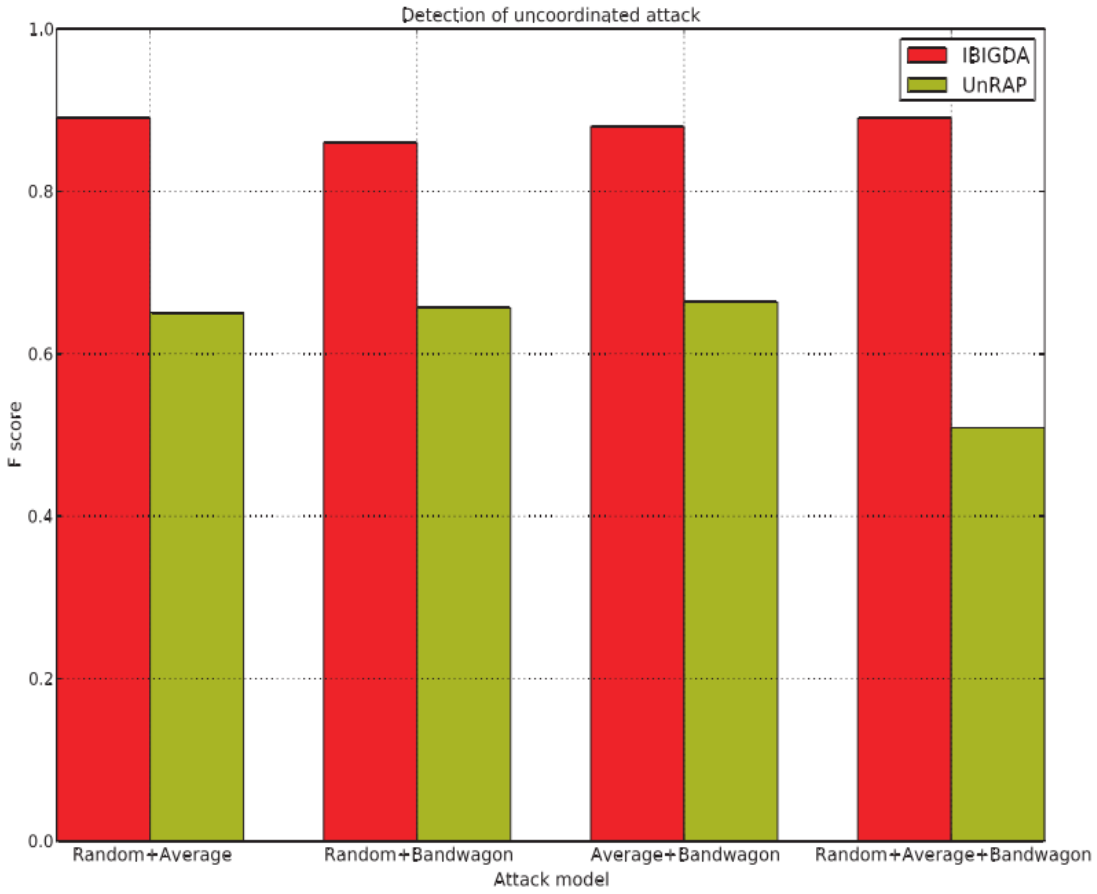

Figure 5. Comparison of IBIGDA and UnRAP in the Detection Performance of the Un-Coordinated Attack

5. Conclusion

Considering the low limitation of shilling attack detection technology unsupervised degree, introducing a quantitative measure of group effect of attack profile, and based on genetic optimization objective function, the adaptive parameter posterior process of inference and attack detection is fusing, and proposed an iterative Bayesian inference of genetic detection algorithm (IBIGDA). IBIGDA algorithm has unsupervised higher degree. Even in the absence of the attack strength, user categories prior knowledge of the number of cases. Still on the random attack, attack and popular support mean attack is reliable and accurate detecting. Recommended system administrator is provided a more practical means of detecting attacks and new research ideas.

References

[1] L. Cong, L. Zhigang and S. Jinlong, "An unsupervised algorithm for detecting shilling attacks on recommender systems", Automation Journal, vol. 02, **(2011)**, pp. 160-167.
[2] Z. Guangqun, "The high Kai... Research on DDoS attack detection system of DNS server", Computer engineering and applications, vol. 33, **(2011)**, pp. 94-97.

[3] Z. Wu, and Z. Yi, "Wang is entitled to, Cao Jie", Recommendation system feature selection support attack detection algorithm. Journal of Electronic, vol. 08, **(2012)**, pp. 1687-1693.
[4] W. Bo, "Research on detection algorithm against profile injection recommendation system", Yanshan University, **(2012).**
[5] X. Yuchen, L. Qiang and Z. Fuzhi, "A user profile attack detection algorithm based on target item identification. Electronic Computer Systems, vol. 07, **(2011),** pp. 1370-1374.
[6] G. Peng, "Luo source, Recommendation system based on unsupervised strategy supporting attack detection", Communication technology, vol. 04, **(2013),** pp. 5-12.
[7] B. Mobasher, R. Burke, R. Bhaumik, *et al.*, "Toward Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness", ACM Transactions on Internet Technology, vol. 7, no. 4, **(2007),** pp. 167-170.
[8] N. Hurley, Z. Cheng and M. Zhang, "Statistical Attack Detection", In Proceedings of the 3rd ACM conference on Recommender systems, New York, New York, USA, **(2009),** pp. 149–156.
[9] B. Mehta, T. Hofmann and P. Fankhauser, "Lies and Propaganda: Detecting Spam Users in Collaborative Filtering", In Proceedings of the 12th international conference on intelligent user interfaces. Honolulu, Hawaii, USA, **(2007),** pp. 14–21.

Authors

Tao Li, she is a lecturer at Information Engineering Department of Jilin Police College. She is in the research of Internet Monitoring.