International Journal of Security and Its Applications
Vol.8, No.6 (2014), pp.115-126
http://dx.doi.org/10.14257/ijsia.2014.8.6.11

Shilling Attack Detection Algorithm based on Non-random-missing Mechanism

Man Li

Shandong Huayu University of Technology,
Dezhou 253034, china,
xdclm@126.com

Abstract

Besides unsupervised feature, universality serves as another important factor determining the practical value of attack detection technology. Considering the difficulty of possessing both features for the existing attack detection techniques, this paper reveals the latent factors invoking missing ratings under the non-random-missing mechanism and further combines these latent factors with Dirichlet process in the framework of probabilistic generative model, thus proposes the Latent Factor Analysis for Missing Ratings(LFAMR)model. Based on performing user clustering with this model, this paper achieves the goal of attack detection by presenting the method for identifying attack cluster in ideal situation. Experimental results show that comparing with the existing detection techniques, LFAMR is more universal and unsupervised, and it can effectively detect shilling attacks of typical types and their derivatives even in lack of the apriori inputs such as user cluster numbers.

Keywords: *shilling attack, shilling attack detection, not missing at random, robust recommendation*

1. Introduction

In this paper, the data missing mechanism is as the foundation, fully tap the potential information loss event score implied, committed to develop shilling attack detection technology both unsupervised and strong universal, which has higher practical value.

At present, shilling attack diversity and hidden are to so that the detection technology of existing faces two limitations:

Unsupervised degree is low, the key parameters is to be unknown input, such as attack strength or user category number, PCA VarSelect, PLSA and EMSVD algorithm.

Universal is poor, for some shilling attack are effective, the scope of application is limited

These limitations restrict practical degree of shilling attack detection technology. The most effective way is to overcome this limitation to fully dig two of all known in recommender systems, and even the value of implicit information. Detection technology of existing ignores potentially valuable information (the sparsity of the rating matrix) [1-2]. Sparsity refers to the scoring matrix contains a large number of loss score. The formation of this characteristic is because users usually only in the evaluation system of a few items of interest. Obviously, the lack of scoring and user preferences and other factors are a specific association. The mining potential information will loss score behind contribute to a more efficient attack detection [3-4].

Therefore, this paper proposes Latent Factor Analysis for Missing Ratings (LFAMR) model. The main idea includes:

(1) Based on non-random missing data mechanism to resolve the underlying factors

ISSN: 1738-9976 IJSIA
Copyright ⓒ 2014 SERSC

leading to the missing score, established Logistic regression model and the corresponding lack of events between the potential factors

(2) Based on (1) regression model and Dirichlet process, established user generated clustering model, and using the variational expectation maximization (Variational EM) algorithm to learn the model parameters

To reveal the attacker enrichment class in the ideal case identification characteristics theoretically, achieved to shilling attack detection.

2. Related Concepts

2.1 The Data not Missing Mode

The missing data mode can be divided into three categories [5-6]: Missing completely at random (MCAR), missing at random (MAR), non-missing at random (NMAR). Formally, consider random vector R and the indicator vector M. Where, $R = R^0 \cup R^m$, R^0 and R^m are respectively the observational data and missing data. The elements of the M is $M_i \in \{0,1\}$, When M_i=0, $R_i \in R^m$, otherwise, $R_i \in R^0$. Let R, M and T implicit variables distribution are:

$$p(R, T, M) = p(R, T)p(M|R, T)$$

In recommendation system, users are generally not evaluation do not like, which indicates that the score of random events missing are not independent, but depends on the user preferences and other potential factors. Therefore, the recommendation system score loss model should be NMAR. Intuitively, such as the effect of MovieLens100K and MovieLens1M data centralized user score distribution [7]. From Figure 1 to show its distribution is uneven, but toward the high evaluation of the regional, apparently user preferences and score deletion event have correlation.

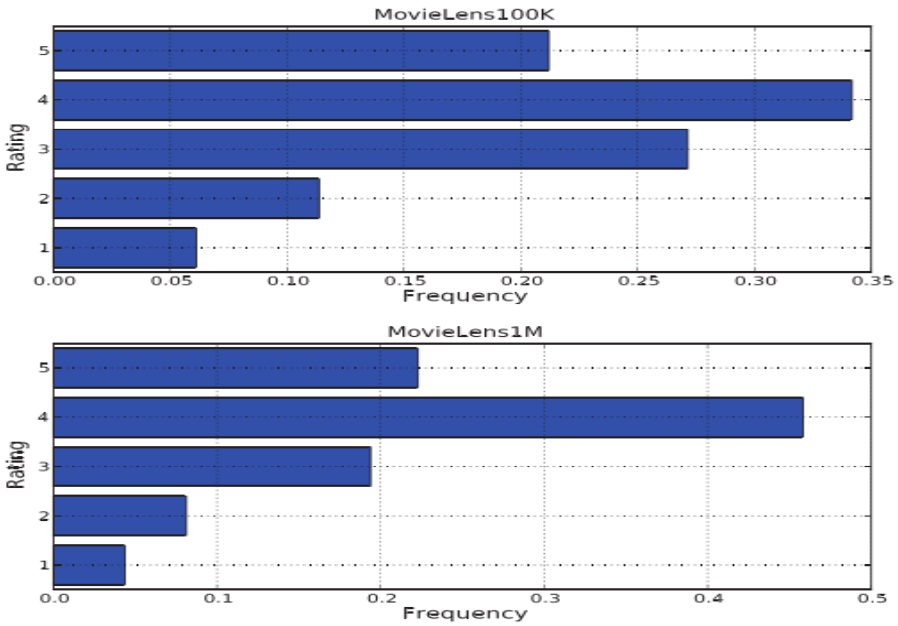

Figure 1. The Score Distribution

Copyright ⓒ 2014 SERSC

2.2. The Dirichlet Process

The core idea of this paper is in accordance with the specific similarity clustering of users, and identifying the attacker enrichment classes. LFAMR assume the user clustering function, to avoid the error fitting set in advance the number of unknown cluster may cause, LFAMR uses a non parametric Bayesian method - Dirichlet process. On one hand, this method allows the number of clusters increased or decreased as required, the model is given full scalability. On the other hand, with the aid of Bayesian method for internal "Occam razor" effect [8], determine the reasonable clustering number.

In essence, the Dirichlet process is a kind of distribution, means that the probability distribution of each sample is itself. The Dirichlet process can be written as $GP(\alpha, G_0)$, α and G_0 are respectively for the scaling factor and the base distribution. Typical application of Dirichlet process is as shown in Figure 2. Consists of two sampling levels:

$$G|\alpha, G_0 \sim GP(G|\alpha, G_0)$$

$$\theta_k \sim G$$

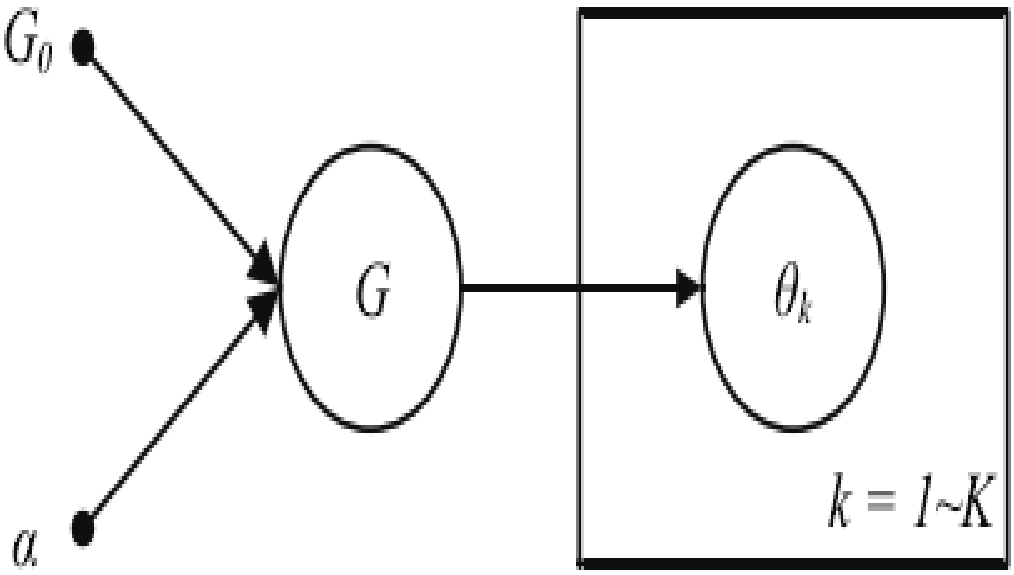

Figure 2. The Sampling Process of the Hierarchical Dirichlet

3. Analysis of Potential Factors Score Model

To reflect the score not missing at random, without introducing too many complicated

factors. LFAMR uses the form of $p(M \mid R, T) = p(M \mid T)$ choice model, namely M depends only on the hidden content T. Ignore the impact score. LFAMR thinks the hidden content on behalf of users interested in the item value. Because "interested" in general is the most important factor in triggering scoring event, scoring just the user after the interested item evaluation. Potential factors are as loss score value of interest, which is the key to construct the LFAMR model.

3.1 Lack of Scoring Potential Factors

Set $user_i$ interest value for $item_j$ is T_{ij}, it can be regarded as the superposition of three factors:

Users Factors U_i, $user_i$ may be abnormal score will be given from the average score of behavior, this personalized by the user factors explain;

Items Factors H_j, If $item_j$ is very popular, we obtain a large number of scores, then factors will enhance the user of this interest. On the contrary, it will inhibit the user of this interest;

Users -items factors UH_{ij}. $user_i$ is pure interest value for $item_j$, determined entirely by user preference and fit the intrinsic nature of the degree of item attributes, not mixed with any external factors. At this point, you can create T_{ij} Logistic regression model for M_{ij}

$$T_{ij} = U_i + H_j + UH_{ij}$$

$$p(M_{ij}) = Bern(M_{ij} \mid \sigma(T_{ij}))$$

3.2. The Formal Description of LFAMR

LFAMR is a model of a probabilistic fusion of the Logistic regression model and DPM. Figure 3 is the probability graph model of LFAMR.

 Copyright ⓒ 2014 SERSC

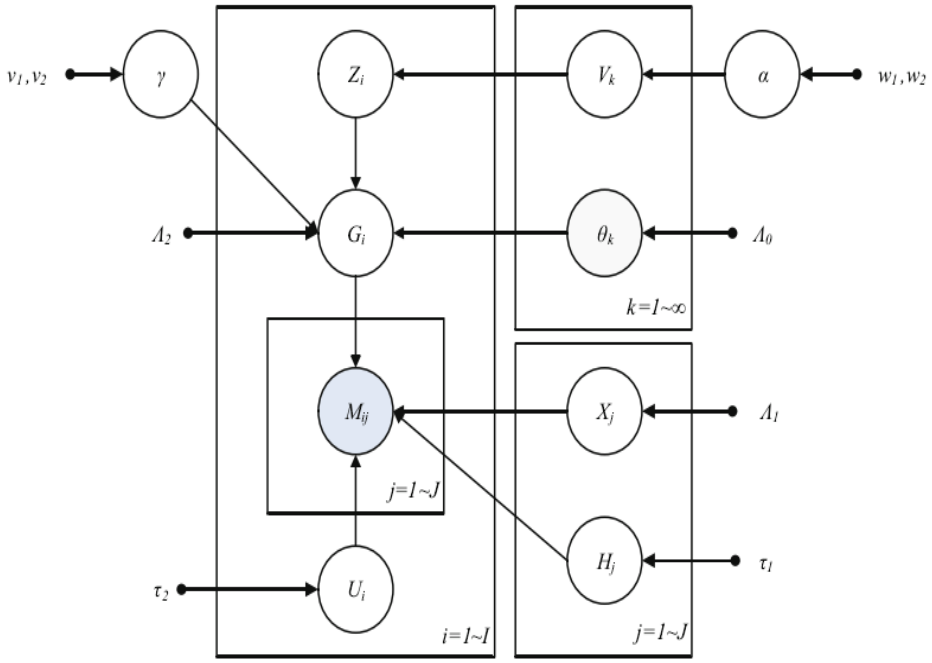

Figure 3. Lack Score Graph Model Potential Factor Analysis

Random variables meet the following distribution:

$$p(\alpha \mid w_1, w_2) = Gam(\alpha \mid w_1, w_2)$$
$$p(\gamma \mid v_1, v_2) = Gam(\gamma \mid v_1, v_2)$$
$$p(V_k \mid \alpha) = \text{Beta}(V_k \mid 1, \alpha)$$
$$p(\theta_k \mid \wedge_0) = N(\theta_k \mid 0, \wedge_0^{-1})$$
$$p(H_j \mid \tau_1) = N(H_j \mid 0, \tau_1^2)$$
$$p(X_j \mid \wedge_1) = N(X_j \mid 0, \wedge_1^{-1})$$
$$p(Z_i \mid \{V_K\}) = \text{Mult}(Z_i \mid \{X_k\})$$

4. Experiment Design and Discussion

4.1. Data Sets and Experimental Setup

The experimental data sets are MovieLens100K and MovieLens1M. For the MovieLens100K data set, the experimental uses all users and items. For the MovieLens1M data set, the experimental uses all items, but chosen randomly 1/4 users. Set priori model parameters of LFAMR, $w_1 = w_2 = v_1 = 1$, $v_2 = 10^{-3}$, $\wedge_0 = \wedge_1 = \wedge_2 = I$, $\tau_1^2 = \tau_2^2 = 0.1$ And the element number is $d = 5$, the cutoff value of is $T = 20$, threshold epsilon is $\varepsilon = 2$. In order to detect the ability to validate LFAMR, assume that the data from original users for real users. In different attack strength of p^{att} and filling rate of p^{fill}

The data set into four shilling attack: random attack, mean attack, bandwagon attack and segments attack users in attack class. N^a is the number of attackers in attack class. , N^t is the total number of attackers in the system, then:

Copyright ⓒ 2014 SERSC

$$\begin{cases} f_{pre} = \dfrac{N^{a}}{N} \\ f_{rec} = \dfrac{N^{a}}{N^{t}} \\ F = \dfrac{2 f_{pre} f_{rec}}{f_{pre} + f_{rec}} \end{cases}$$

4.2. Shilling Attack Detection Example

The following examples demonstrate the user clustering process of LFAMR and identify the attack class. Now Movie-Lens100K data set injected parameters $p^{att} = 10\%$ $p^{fill} = 20\%$. Fisher [9] discriminant method will use the user's expectations preference characteristics $(E[G_i])$ projected into two dimensional spaces. The top half of Figure.5 shows iteration number is 0, 3, 7, 23, lower shows mixing coefficient in the corresponding number of iterations of each user class. Dirichlet process is by adjusting the mixing coefficient of user class, keep user class of the data fitting and generalization performance. Because of using the low dimensional projection technology, so most of the user class overlapped together in the graph, and boundary is blurred. However, it can be observed that the attacker, gradually separated from real users in the background, to the near origin. And in the first 23 iterations are highly enriched in cls_{10} class, which is to identify the attack class. The type of cls_{10} center is closest to the origin, confirms the effectiveness of attack recognition method. In particular, in the Figure4, cls_{10} is the closest to the plane coordinate system origin.

4.3. The Detection Results Analysis

This paper tests the comprehensive LFAMR attacks on MovieLens100K data detection capability. Selected PCA VarSelect, PLSA, EMSVD and UnRAP algorithm are as the performance of the LFAMR reference. At present, PCA VarSelect in this data set with the detection performance of the best. The experiment adopted $4 \times 4 \times 6$ design patterns, attack model (random attack, average attack, bandwagon attack and segments attack), The attack strength of p^{att} (5%, 7%, 10%, 12%) and filling rate of p^{fill} (3%, 6%, 9%, 12%, 15%, 20%) corresponds to a different combination of a group of experimental configuration. Each configuration of the experimental results obtained from ten independent experiments mean.

Outside the parentheses of table 1-4 shows the LFAMR data detection results.

 Copyright ⓒ 2014 SERSC

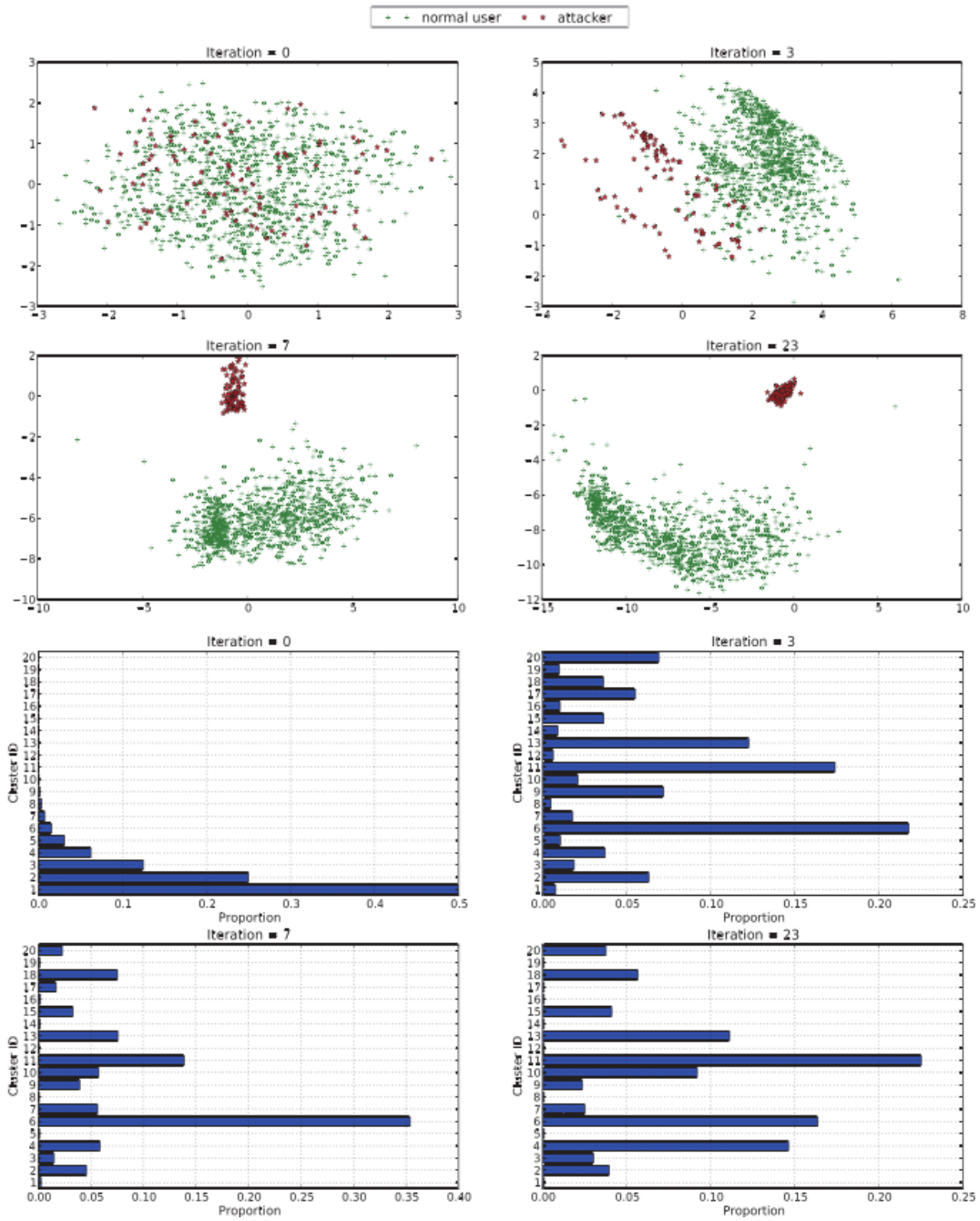

Figure 4. Application Examples of LFAMR

Table 1. The F Value of Random Attack Detection (MovieLens100K)

p^{fill} p^{att}	3%	6%	9%	12%	15%	20%
5%	0.69(0.98)	0.98(0.99)	0.99(0.98)	0.99(0.99)	0.98(0.99)	0.99(0.99)
7%	0.87(0.98)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)
10%	0.80(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)
12%	0.62(0.99)	1.00(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)

Copyright ⓒ 2014 SERSC

Table 2. The F Value of Average Attack Detection (MovieLens100K)

p^{fill} p^{att}	3%	6%	9%	12%	15%	20%
5%	0.75(0.98)	0.98(0.98)	0.99(0.97)	0.99(0.97)	0.99(0.97)	0.98(0.96)
7%	0.88(0.97)	1.00(0.98)	0.99(0.98)	0.99(0.98)	0.99(0.97)	0.99(0.97)
10%	0.79(0.98)	0.98(0.98)	1.00(0.98)	1.00(0.98)	0.99(0.98)	0.99(0.97)
12%	0.71(0.98)	0.98(0.98)	1.00(0.98)	1.00(0.98)	0.99(0.98)	0.99(0.97)

Table 3. The F Value of Bandwagon Attack Detection (MovieLens100K)

p^{fill} p^{att}	3%	6%	9%	12%	15%	20%
5%	0.60(0.96)	0.99(0.97)	0.99(0.98)	0.99(0.98)	0.99(0.98)	0.99(0.99)
7%	0.61(0.97)	0.99(0.98)	0.99(0.98)	0.99(0.99)	0.99(0.99)	0.99(0.99)
10%	0.70(0.96)	0.99(0.98)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)
12%	0.71(0.97)	0.99(0.98)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)

Table4 The F Value of Segment Attack Detection (MovieLens100K)

p^{fill} p^{att}	3%	6%	9%	12%	15%	20%
5%	0.51(0.00)	0.98(0.00)	1.00(0.28)	0.99(0.58)	0.99(0.68)	0.98(0.78)
7%	0.54(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.29)	0.99(0.64)
10%	0.63(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.00)
12%	0.63(0.00)	0.99(0.00)	1.00(0.00)	0.99(0.00)	0.99(0.00)	0.99(0.00)

Inside the parentheses of Table 1-4 is PCA VarSelect. To achieve the best performance the value of F in the corresponding experimental configuration. In most cases, the detection ability of LFAMR is better than that of PCAVarSelect. In particular, PCA VarSelect faced with segment attack to be almost lapsed, and LFAMR effective detection range covers the four attack model. PCA VarSelect achieves optimal detection performance that the attack must be informed of the exact strength. Similarly, PLSA, EMSVD and UnRAP also have different degrees of universality or non-supervisory limitations.

Figure 5 shows the four attack models. $p^{att} = 10\%$, $p^{fill} = 6\%$, LFAMR, PLSA, EMSVD and UnRAP algorithm are comparison of detection performance. PLSA and EMSVD require the user to the number of categories as input parameters, and the parameters are generally only by the test method. In a number of different categories of users, the detection performance of PLSA is very sensitive to the change of the input. Users cannot guarantee the same number of classes in all the attacks have made the best detection results. Detection performance of EMSVD sensitivity to input parameters is weaker than PLSA. But it cannot detect the average attack, while facing the attack is complete failure. In contrast, LFAMR can detect four shilling attack, and there is no need to change according to attack situation adjustment input parameters, degree of unsupervised is very high. The unsupervised degree of UnRAP and LFAMR are same. But the universal is weak, unable to detect attacks. However, in the remaining three attack model, it is ability to detect and LFAMR are almost same.

 Copyright ⓒ 2014 SERSC

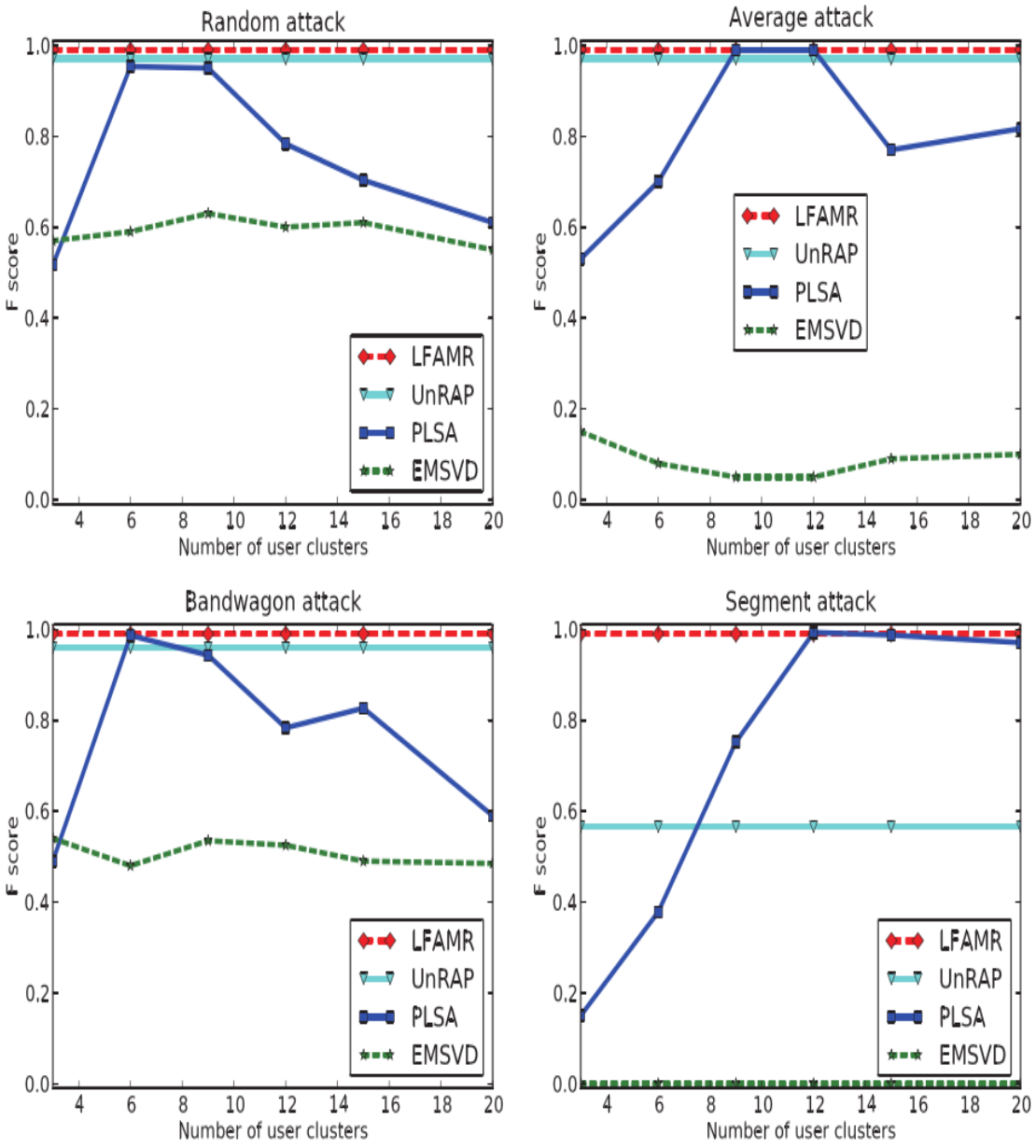

Figure 5. Detection Performance Depends on the Input Parameter

In order to show the differences of LFAMR and UnRAP, we use the attack model of nuclear attack.

Next, on the MovieLens1M data set is to evaluate LFAMR detection ability of shilling attack. The experiment adopted $4 \times 4 \times 5$ design patterns, attack model (random attack, average attack, bandwagon attack and segments attack), The attack strength of p^{att} (5%, 7%, 10%, 12%) and filling rate of p^{fill} (6%, 9%, 12%, 15%, 20%) corresponds to a different combination of a group of experimental configuration. Each configuration of the experimental results obtained from ten independent experiments mean.

Copyright ⓒ 2014 SERSC

Table 5-8 shows the effect of detection LFAMR, The same reasons mentioned above based on LFAMR, the test cannot detect the filling rate was 3% shilling attacks, it omitted the corresponding experimental data. From the Table 5-8 we can see that, in all other attack scenario, LFAMR shows good detection capability, F value is in more than 98%, it can accurately identify the attacker

Table 5. The F Value of Random Attack Detection (MovieLens1M)

p^{fill} p^{att}	6%	9%	12%	15%	20%
5%	0.99	0.99	0.99(0.99)	1.00	0.99
7%	0.99	0.99	0.99(0.99)	0.99	0.99
10%	0.99	1.00	0.99(0.99)	0.99	0.99
12%	0.99	0.99	0.99	0.99	0.99

Table 6. The F Value of Average Attack Detection (MovieLens1M)

p^{fill} p^{att}	6%	9%	12%	15%	20%
5%	0.98	0.99	0.99	0.99	0.99
7%	0.99	0.98	1.00	1.00	0.99
10%	0.99	0.99	1.00	0.99	0.99
12%	0.99	0.99	0.99	0.99	0.99

Table 7. The F Value of Bandwagon Attack Detection (MovieLens1M)

p^{fill} p^{att}	6%	9%	12%	15%	20%
5%	0.99	0.99	0.99	0.99	0.99
7%	0.99	0.99	0.99	0.99	0.99
10%	0.99	0.99	0.99	0.99	0.99
12%	0.99	0.99	1.00	0.99	0.99

Table 8. The F Value of Segment Attack Detection (MovieLens1M)

p^{fill} p^{att}	6%	9%	12%	15%	20%
5%	0.98	0.99	0.99	0.99	0.99
7%	0.99	0.99	0.99	0.99	0.99
10%	0.99	0.99	0.99	0.99	0.99
12%	0.99	0.99	0.99	0.99	1.00

Conclusion

In this paper, the existing shilling attack detection algorithm in the unsupervised and universality are limitations, the data missing mechanism is as the basis, for the potential cause score missing carries on the analysis. And the probability of generating these potential factors and Dirichlet process are integrating in the framework of the model, proposed Latent Factor Analysis for Missing Ratings (LFAMR) model to use shilling attack detection. LFAMR makes use of the clustering effect of Dirichlet process, through user identification features clustering and reveal attack class to achieve the purpose of shilling attack detection.

 Copyright ⓒ 2014 SERSC

References

[1] Z. Huang, H. Chen and D. D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering", ACM Transactions on Information Systems, vol. 22, no. 1, **(2004),** pp. 116–142.
[2] B. M. Marlin, R. S. Zemel, S. Roweis, *et al.*, "Collaborative Filtering and the Missing at Random Assumption", In Proceedings of the 23rd conference on Uncertainty in artificial intelligence, **(2007)**, pp. 267–275.
[3] B. M. Marlin and R. S. Zemel, "Collaborative Prediction and Ranking with Non-random Missing Data", Bergman L D, Tuzhilin A, Burke R D, et al. In Proceedings of the 3rd ACM conference on Recommender systems, New York, New York, USA, **(2009),** pp. 5–12.
[4] B. Marlin, S. Roweis and R. Zemel, "Unsupervised Learning with Non-ignorable Missing Data", In Proceedings of the 10th international workshop on Artificial intelligence and statistics, **(2005)**, pp. 222–229.
[5] R. J. A. Little and D. B. Rubin, "Statistical Analysis with Missing Data (2nd edition)", John Wiley, **(2002)**.
[6] B. M. Marlin. "Missing Data Problems in Machine Learning", Canada: University of Toronto, **(2008).**
[7] T. S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems", The Annals of Statistics, vol. 1, no. 2, **(1973)**, pp. 209–230.
[8] D. J. C. MacKay, "Bayesian Interpolation", Neural Computation, vol. 4, no. 3, **(1992),** pp. 415–447.
[9] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, vol. 7, no. 2, **(1936),** pp. 179–188.

Author

Man Li, she is an Associate Professor in Shandong Huayu University of Technology. She is in the research of computer application technology

Copyright ⓒ 2014 SERSC

Copyright ⓒ 2014 SERSC