A Secure Online Reputation Defense System from Unfair Ratings using Anomaly Detections

Asha baby
PG Scholar,Department of CSE
S.K.P Engineering College
Tiruvannamalai, India

A. Kumaresan
Professor, Department of CSE
S.K.P Engineering College
Tiruvannamalai, India

K. Vijayakumar
Professor, Department of CSE
S.K.P Engineering College
Tiruvannamalai, India

ABSTRACT
A reputation system collects feedbacks from users and aggregates these feedbacks as evidence and generates the aggregated results to the normal users. These aggregated results are called reputation scores. We can call this system as online feedback-based reputation system. To protect the reputation system many defense schemes have been developed. In this paper we propose a defense scheme; it is the combination of five modules. Evaluation based filtering, Time domain unfair rating detector, suspicious user correlation analysis, trust analysis based on Dempster-Shafer theory and malicious user identification and reputation recovery. This system identifies the items under attacks, the time when the attacks occur and unfair raters who insert unfair ratings. Compared with existing systems this system achieves detection of high unfair ratings and reduces the detection of false dishonest ratings.

General Terms
Unfair ratings, user correlation analysis, trust analysis, belief function, Euclidean distance

Keywords
CUSUM detector, Dempster-Shafer theory, K-mean algorithm

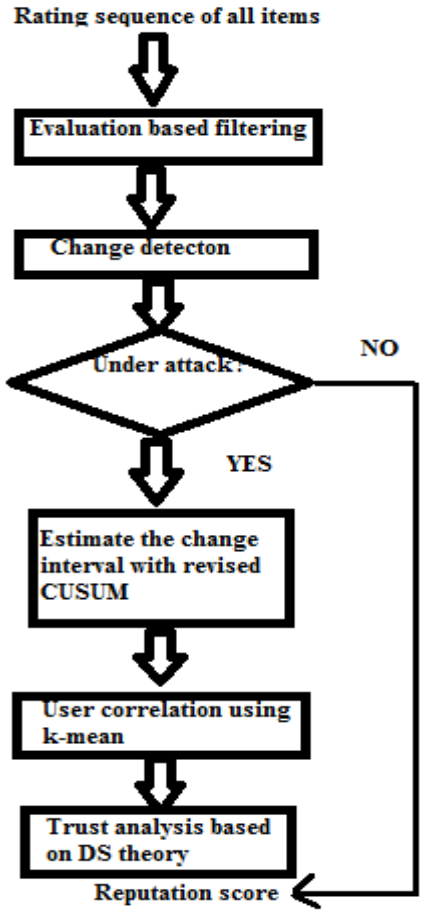
1. INTRODUCTION
Many of the people are using internet for their daily life such as entertainment, making personal relationship and business purposes. The internet has created large opportunities for online interactions. However the internet is more vulnerable to attacks, which makes online interactions risky. It will ask a number of questions in online interactions. For example Will a seller at an online shopping site provides the product in correct time? Is Amazon.com site will produce high quality and trustworthy product?[7] Is a video on YouTube really interesting? Here is one problem that satisfies how the online participants protect themselves by identifying the quality of strangers and unfamiliar items .

To solve this problem online reputation systems are introduced. A reputation system collects feedbacks as evidence, about individual items, including products, services, and digital contents, aggregate the evidence and disseminates the aggregated results into the normal users. These results are called reputation scores. The system which provides rating is referred to as online feedback-based reputation system [3].In order to protect reputation systems, many defense schemes have been developed. The efficiency of the defense scheme depends on the accuracy of the reputation system. Without the proper defense scheme items reputation score increase or decrease rapidly. It will reduce the quality of the reputation system.

In this paper we propose a defense scheme. It consists of 5 modules, Evaluation based filtering, A Time domain unfair rating detector[1], Suspicious user correlation analysis, Trust analysis based on Dempster-Shafer theory and finally Malicious user identification and reputation recovery. This system accepts time domain rating sequences as input, an unfair rating detector will identify whether changes occur due to rapidly or changes accumulated over time. Here revised CUSUM detector[9] is used. Which detects the change intervals, it is given to the correlation analysis module, and from suspicious user intervals this module forming the clusters containing suspicious users with a smaller distance using k-mean clustering algorithm. After forming the clusters, group rating will be calculated by taking average rating score of each users in this group. This module filtering out number of malicious users. Even though these clusters contain malicious users. In order to improve the detection rate again these clusters are given to the 3rd module trust analysis .It will calculate the trust value based on the Dempster-Shafer theory[9] and find out all malicious users based on their trust value. Here trust value is calculated by collecting history of individual users. Lower trust value indicates malicious user and finally accurate reputation score will be calculated by considering only honest raters.
The performance of this system is evaluated with existing systems such as [3], [4], [5], [6]. Our system provides better detection rate and low false alarm rate.
The remaining of the paper is arranged as follows section 2 describes the proposed mechanism, section 3 deals with the results and final section 4 gives the conclusion.

2. PROPOSED MECHANISM

2.1 System Architecture



The proposed scheme contains **4** modules. Time domain unfair rating detector, suspicious user correlation analysis, Trust analysis based on Dempster-Shafer theory and malicious user identification and Items reputation recovery

2.1.1 Evaluation based Filtering

When $D_r \leq 0$ or $D_r \geq T$, CUSUM detector detects the change and restart the detector by setting $D_r=0$, again starts another round of detection.

Disadvantages:

- Basic CUSUM only determines stopping time, at which the detector detects the change interval, but it is not the actual starting time of change.
- It doesn't determine the starting and ending time of change.
- It will't provide the exact change intervals.

In order to overcome these disadvantages, we introduce Revised CUSUM [9]. This detector is the advanced version of basic CUSUM, which detects the exact change starting and ending time.

2.1.2.1 Modified CUSUM

We analyse a new ranking method , a novel scoring scoring system that aggregates the evaluations of N agents over M objects by use of reputation and weighted averages. The method[3] can be implemented via an iterative algorithm, where the intrinsic bias of the estimators of the weights can be corrected.

2.1.2 Time Domain Unfair Rating Detector

In online reputation system, there is more chance of occurring unfair ratings. There exists number of change detectors like Shewhart, Finite weighted moving average and CUSUM[1] for different applications. In online reputation system, the rating doesn't follow specific distribution. The task of finding small shifts in rating is very complicated. So we introduce CUSUM detector[1] which detects small shifts in normal ratings.

2.1.2.1 Basic CUSUM

If there exist small changes in rating, basic CUSUM [9] determines the changes in parameter θ in a probability density function (PDF). Assume $P\theta_1$ and $P\theta_2$ be the PDF before and after change respectively. Let X_r denotes the r^{th} sample of rating sequence.

The basic CUSUM detection function is,

$$Dr = \max\left(Dr - 1 + \ln\left(\frac{P\theta_1(Xr)}{P\theta_2(Xr)}\right)\right) \quad \textbf{(1)}$$

$$Ts = \min(r) \quad \text{where} \quad Dr >= T \quad \textbf{(2)}$$

Table 1 specifies the variable and their indications in equations (1) and (2)

Table 1 variables and their indication

Variable	Indication
T	Change detection threshold
Ts	Stopping time
Dr	Detection rate of the r^{th} sample

In order to avoid the disadvantages of basic CUSUM, we introduce a modified CUSUM [9] which satisfies all of our requirements.

Change detection algorithm

1. Attack-list= [] //set contains objects under attacks, initially empty

2. For each object I do

3. Collect all ratings for 0i and arrange according to the time they provided.

4. Flag=0 //a flag ,which is 1 when D_r exceeds threshold

5. Compute $D_i=\{D_i(1),D_i(2),\ldots\ldots D_i(n)\}$

6. If (max (D_i)>T) then

7. Add Oi to attack-list[]

8. for each rating r do

9. If (flag==0) then

10. If($D_i(r)$>T)then

11. Estimate the change starting time T_{start}

12. Flag=1

13. End if

14. Else

15. If($D_i(r)$<T)then

16. Estimate change ending time T_{end}

17.Flag=0

18.End if

19. End if

20. End for

21. End if

Modification:

Here we introduce a new modification of the existing change detection algorithm. During the change interval we stop all other ratings for that particular item. And making it for the accuracy of malicious users.

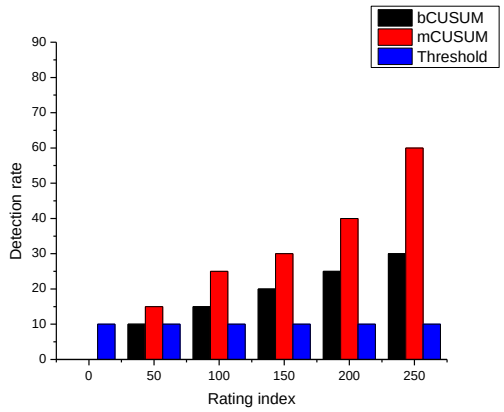


Fig1 Performance comparison of change detector

Fig 1 suggests that performance (in terms of detection rate) of basic CUSUM(bCUSUM) is lower than that of modified CUSUM (mCUSUM) and here threshold is taken as constant. The x axis is the rating index and y axis is the detection rate (Dr).

Advantages:

- Determine the changes occur on an item.
- Suspicious timing intervals are determined.

2.1.3 Suspicious User Correlation Analysis

Change detector determines the change intervals which are called suspicious user intervals which contains both normal and unfair ratings. Suspicious user provides the rating during suspicious intervals. Consider user A and B are the two users provides ratings during suspicious intervals. These users are referred to as suspicious users analyze all their ratings along with items under attacks. Assume here having total m number of items and their rating form objects are denoted as {x1, x2,……xm} and {y_1, y2…..ym} respectively.

In order to form the clusters perform the Euclidean distance function for these two users A and B.

$$d(A,B) = \sqrt{(x1 - y1)^2 + (x2 - y2)^2 + \cdots (xm - ym)^2} \quad \textbf{(3)}$$

This distance is calculated for each pair of suspicious users using k-mean [5] clustering algorithm. Suspicious users with smaller distances among them are clustered together. After forming the clusters group rating will be calculated by taking average rating score of each users in this group. Large group rating is marked as malicious user. This module filtering out number of malicious users. Even though these clusters again contain malicious users. In order to improve the detection rate again these clusters are given to the 3rd module trust analysis .

Advantages

- When variables are larger, k-means may be computationally faster than other clustering methods.
- K-mean may produce closed clusters than other clustering approaches, especially if the clusters are globular.

2.1.4 Trust analysis based on Dempster-Shafer theory

Suspicious users are derived from the users who rate during the detected change interval. But we can't consider all suspicious users as malicious users at time normal users tend to provide a biased rating because of personal reasons or human error. Hence we put forward trust analysis so as to differentiate normal users and malicious users [8]. As a preface for the proposed trust model here we introduce some major concept used in our model. Behavior, In this we define a user's behavior value of a particular item as binary value which will indicate the user's rating behavior as good or bad.

2.1.4.1 Dempster–Shafer theory (DST)

We propose the system with trust analysis. Trust analysis is based on Dempster-Shafer theory [13]. Using belief function it will combine evidences from different sources and achieve at a high degree of trust.

2.1.4.2 Trust Model based on DST

We proposed a trust model[9] for calculating the trust value of each user. Consider a feedback-based reputation system, here exists number of users who provide ratings to number of items.

Assume a user U_k has rated total N items such as I_1,I_2,I_3,……I_N. Our aim is to calculate the trust value of user U_k on item I_i.

Procedure for Calculating User Trust Value

Aim:

To calculate the trust value of user U_k to I_i

Methods:

- By considering user U_k ratings to other N-1 items
- By considering user U_k ratings to I_i items.

Steps:

1.Calculate the combined behavior value of user U_k on N-1 items.

$$C.Beh(\mathrm{Uj})(\mathrm{i}) = \frac{x}{x+y+2} \quad \textbf{(4)}$$

2. We calculate the behavior uncertainty of user U_j on item I_i.

$$C.Beh(\mathrm{Uj})(\mathrm{i}) = \frac{2}{x+y+2} \quad \textbf{(5)}$$

3.Calculate the trust value of user U_j based on the rating to N-1 items.

$$Com.T(Uj)(\mathrm{i}) = C.Beh(Uj)(\mathrm{i}) * (1 - Beh.Uncer(Uj)(\mathrm{i})) \quad \textbf{(6)}$$

4.Calculate the trust value of user U_j to I_i items.

$$ITEMiT(Uj)(\mathrm{i}) = Beh(Uj)(i) * Beh.Uncer(Uj)(\mathrm{i}) \quad \textbf{(7)}$$

5.Calculate total trust value on item Ii

$$TUj(i) = Com.T(Uj)(\mathrm{i}) + ITEMiT(Uj)(\mathrm{i}) \quad \textbf{(8)}$$

Table 2 variables and their description

Variable	Description
X	Number of items with U_j behavior value 1
Y	Number of items with Uj behavior value 0
C.Beh(Uj)(i)	Combined Behavior value of user U_j on item I_i
Beh.Uncer(Uj)(i)	Behavior uncertainity
Com.T(Uj)(i)	Trust value of user $U_j$on N-1 items.
ITEMi.T(Uj)(i)	Trust value of user U_j on I_i
TUj(i)	Total trust value

Table 2 describes the variable and their description used in the equations (4),(5),(6),(7),(8).

Based on the trust value feedback-based reputation system can analyze the rating behavior of users.

2.1.5 Malicious User Identification And Items Reputation Recovery

After the trust analysis, we have to examine the trust value of each user. Based on the Dempster-Shafer theory we evaluate the trust value, users with low trust value is considered as malicious user, for that we have to specify the trust threshold. Malicious users provide normal and unfair ratings. Instead of removing all the ratings ,only remove the ratings with low trust value. After the rating removal we use averaging method to calculate the reputation scores.

3. RESULT

In order to implement our system, we have simply developed a Online Shopping Website. Which contains following features.

- Admin can add product details(product name, price, validity etc…) and should maintain the product details.
- The user enters their credit card details such as credit card number, card holder name, DOB, credit card provider name and the credit card is validated.
- The user can select purchasing products displayed in the home page or search the product using keyword or based on category.

After developing online shopping site next step is to develop online feedback-based reputation system. Perform the following steps to detect the malicious users in online shopping site.

Step 1:

A novel scoring system that aggregates the evaluations of N agents over M objects by use of reputation and weighted averages.

Step 2:

Change interval estimation, using revised CUSUM detector identify the change intervals with starting and ending time. These intervals may or may not contain malicious users.

Step 3:

These change intervals contain suspicious users. Using k-mean clustering algorithm forming clusters with small distance function. After forming the clusters group rating will be calculated by taking average rating score of each users in the group. Large group rating is marked as malicious user. Filtering out these clusters. Again which contain less amount of malicious user's .Eliminate that users using trust analysis.

Step 4:

Calculate the trust value based on Dempster-Shafer theory. It is based on behavior uncertainty. When user having low trust value than the threshold it is considered as malicious users and eliminate that user. Through these steps our system increases the detection rate in a better level.

Step 5:

Identifies the malicious user and calculate the reputation score based on averaging method.

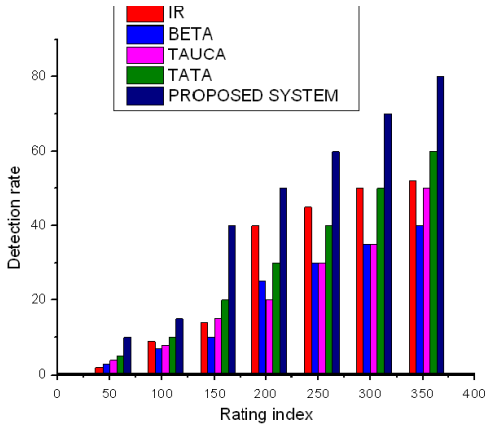


Fig 2 Performance Comparison

In this graph x axis denotes rating index and y axis denotes the malicious user number. This graph shows that our proposed system shows better detection rate. When compared with existing systems such as TATA[6], TAUCA[5], BETA[3], IR[4] our system increase the detection rate and improving the accuracy of defense scheme.

4. CONCLUSION

In this paper, we proposed a novel anomaly detection scheme for protecting online feedback-based reputation system. It consist of 5 modules ,evaluation based filtering, time domain unfair rating detector, suspicious user correlation analysis, trust analysis based on Dempster-Shafer theory and malicious user identification and items reputation recovery. When compared with existing systems such as TATA, TAUCA, BETA, IR our system increase the detection rate and improving the accuracy of defense scheme.

5. REFERENCES

[1] E. S. Page, "Continuous inspection schemes," *Biometrika, vol. 41, no. 1/2, pp. 100–115, Jun. 1954.*

[2] T. K. Philips, Monitoring Active Portfolios: *The CUSUM Approach.*

[3] A. Whitby, A. Jøsang, and J. Indulska, *"Filtering out unfair ratings in Bayesian reputation systems," Icfain J. Manage. Res., Vol. 4, no. 2, pp. 48–64, Feb. 2005.*

[4] P. Laureti, L. Moret, Y.-C.Zhang, and Y.-K. Yu, *"Information filtering via iterative refinement,"* Europhys.Lett.,Vol. 75, no. 6, pp. 1006–1012, 2006.

[5] Y. Liu and Y. Sun,"Anomaly detection in feedback-based reputation systems through temporal and correlation analysis," in Proc. *2nd IEEEInt.Conf. Social Computing, Aug. 2010, pp. 65–72.*

[6] Y. Liu and Y. Sun, "Securing Online Reputation System through Trust Modeling and *Temporal Analysis".*

[7] Y. Yang, Q. Feng, Y. Sun, and Y. Dai, *"Reputation trap: A powerful attack on reputation system of file sharing p2p environment,"* in Proc.4th Int. Conf. Security and Privacy in Communication Networks, Istanbul, Turkey, Sep. 2008.

[8] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, *"Sybilguard: Defending against Sybil attacks via social networks,"* in Proc. 2006 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communications, 2006, pp. 267–278.

[9] Y. Liu, Y. Sun and Alex C. Kot "*Securing Online Reputation System through Dempster-Shafer Theory based Trust Model "*

[10]