ISSN:0975-9646

Securing Online Reputation Systems through Trust Modelling and Temporal Analysis

Aanum Shaikh

MCT's Rajiv Gandhi Institute of Technology
Department of Computer Engineering
Andheri(West),Mumbai-53,India

Abstract— **Online Reputation systems are playing increasingly important roles in influencing people's online purchasing/downloading decisions. With the rapid development of reputation systems in different online social networks, manipulations against such systems are evolving quickly. This paper proposes a scheme TATA, the abbreviation of joint Temporal and Trust Analysis, which protects reputation systems from a new angle: the combination of time domain anomaly detection and Dempster–Shafer theory-based trust computation. It also demonstrates a great potential to effectively remove dishonest ratings and keep the online reputation system a secure and fair marketplace.**

Keywords—: **Information filtering, social network, information security.**

INTRODUCTION

Nowadays most of the people all around the world use internet for entertainment and other purposes, building personal relationships, and conducting businesses, the Internet has created vast opportunities for online interactions. However, due to the anonymity of the Internet, it is very difficult for normal users to evaluate a stranger's trustworthiness and quality, which makes online interactions risky. Does a product at Amazon.com have high quality as described? Is a video on YouTube really interesting or informative? In most cases, the questions can hardly be answered before the interactions are committed. The problem is how the online participants protect themselves by judging the quality of strangers or unfamiliar items beforehand. To address this problem, online reputation systems have been built up. The goal is to create large-scale virtual word-of-mouth networks where individuals share opinions and experiences, in terms of reviews and ratings, on various items, including products, services, digital contents and even other people. These opinions and experiences, which are called users feedback, are collected as evidence, and are analyzed, aggregated, and disseminated to general users. The disseminated results are called reputation score. Such systems are also referred to as feedback based reputation systems. Online reputation systems are increasingly influencing people's online purchasing/downloading decisions. For example, according to comScore Inc., products or services with a 5-star rating could earn 20% more than products or services with a 4-star rating could. More and more people refer to Yelp rating system before selecting hotels and restaurants; to Amazon product ratings before purchasing products online; to YouTube video ratings before viewing a video clip; and etc. Furthermore, a recent survey indicates that around 26% of adult Internet users in the U.S. have rated at least one item through online reputation systems. Meanwhile, driven by the huge profits of online markets, diverse manipulations against online reputation systems are evolving rapidly. Many sophisticated programs are developed to automatically insert feedback. Furthermore, some reputation management companies even control large affiliate networks of real user IDs to provide rating services for their customers. For about $750, a company named VideoViralViews.com can provide 100 real user ratings to a piece of music on iTunes. For just $9.99, a video on YouTube could receive 30 "I like" ratings or 30 real user comments provided by Increase YouTube Views.com. Without proper defense schemes, attacks against reputation systems can overly inflate or deflate the item reputation scores, crash users confidence in online reputation systems, eventually undermine reputation-centric online businesses and lead to economic loss. Securing online reputation systems is urgent. The concept of user behaviour uncertainty from the Dempster–Shafer theory can be used to model users behaviour patterns, and evaluate whether a user's rating value to each item is reliable or not.

System Model*:* We model the feedback-based reputation systems as the system in which users provide ratings to items .This model can describe many practical systems. For example, buyers provide ratings to products on Amazon.com, and reader's rate social news on Reddit.com. The items in above systems are products and social news, respectively. We consider that each user will provide rating to one item at most once, and the rating values are integer values ranging from 1 to 5. In practice, reputation systems often allow users to provide reviews as well. These reviews can also be untruthful. In this paper, we focus on the detection of dishonest ratings. The analysis of untruthful reviews is beyond the scope of this paper, whereas the dishonest rating detection and untruthful review detection complement each other. Attack Model*:* An attacker can control one or multiple user IDs and each of these user IDs is referred to as a malicious user. Malicious users provide ratings to manipulate the reputation score of items. The item whose reputation score is manipulated by malicious users is called a target item. The ratings provided by malicious users to target items are considered as dishonest ratings. An attack profile describes the behavior of all malicious users controlled by the attacker. Assumptions*:* In this work, we assume that items have intrinsic quality, which does not change rapidly. The rating values to a given item depend on the users' personal preference as well as the item quality. In some applications, such as ratings for movies or books, the item quality judgement is very subjective and users' personal preference plays a more important role, whereas in some other applications, such as Amazon product ratings, the item quality plays a more important role. In this work, we focus on the product-rating type applications, where the rating distribution of an item is relatively stable. Therefore, if rapid changes in the rating distribution occur, it is possible that anomaly happens. Furthermore, we notice that due to personal preference, normal users sometimes may also provide "biased ratings" that are far away from the real quality of the items. Meanwhile, to avoid being detected by reputation defense schemes, malicious users may imitate normal users' behaviors by providing honest ratings to the items that they do not care. We call these ratings as "spare ratings". We assume that most of the ratings from normal users can reflect the real quality of the items, Whereas malicious users who have limited rating resources would mainly focus on rating target items and can provide "spare ratings" to few other items. This is also observed in the attack data in the cyber competition. The trust module of the proposed scheme TATA is built up based on this assumption and may not well differentiate normal users from malicious users.

I. JOINT TEMPORAL AND TRUST ANALYSIS(TATA)

The Temporal and trust analysis(TATA) scheme contains the two main aspects: (a) a time domain anomaly detector and (b) a trust model based on the Dempster–Shafer theory. In TATA, the purpose is to detect anomaly from a new angle: analyzing time domain information. Specifically, ratings are organised to a given item as a sequence in the descending order according to the time when they are provided. This sequence actually reflects the rating trend to the given item. In practice, many items have intrinsic and stable quality, which should be reflected in the distribution of normal ratings. If there are rapid changes in the rating values, such changes can serve as indicators of anomaly. Therefore, a change detector in TATA is used as the anomaly detector, which takes the rating sequences as inputs and detects changes occurring in the rating sequences. The change detector will detect not only sudden rapid changes but also small changes accumulated over time. In this way, even if malicious users insert dishonest ratings with small shifts to gradually mislead items reputation scores, such type of changes will still be accumulated and finally be detected by the change detector. If the change detector is triggered by an item, the time intervals in which the changes occur are called change intervals. However, the change intervals may still contain normal ratings. Therefore, the trust analysis module comes into picture.

- Instead of assigning a user with an overall trust value, the trust model evaluates each user's reliability on different items separately. It can reduce the damage from the malicious users who aim to accumulate high trust values by providing spare ratings to uninterested items.
- Furthermore, based on the Dempster–Shafer theory, the trust model introduces user behaviour uncertainty. In this way, a user could yield high trust values only if the user's behaviour yields a sufficient amount of good observations.

Finally, the users with low trust values will be identified as malicious users and their ratings to the detected target items will be removed. The remaining ratings are used to calculate the item reputation.

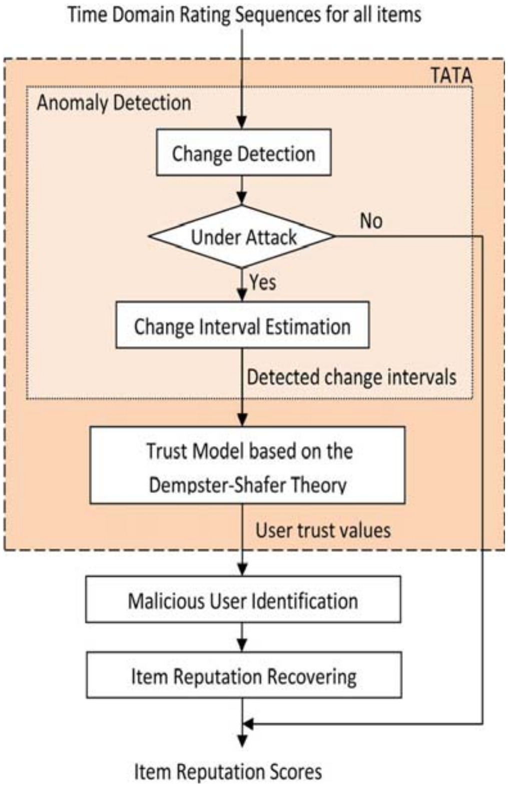
Time Domain Rating Sequences for all items
TATA
Anomaly Detection
Change Detection
Under Attack
No
Yes
Change Interval Estimation
Detected change intervals
Trust Model based on the Dempster-Shafer Theory
User trust values
Malicious User Identification
Item Reputation Recovering
Item Reputation Scores

Fig. 1 System architecture

II. TEMPORAL ANALYSIS-CHANGE DETECTOR

Almost all the change detectors have been developed for different application scenarios. In online reputation systems, since normal ratings do not necessarily follow a specific distribution and attackers may insert dishonest ratings with small bias, it is mandatory to choose a change detector that is insensitive to the probability distribution of data and is able to reliably detect small shifts. Therefore, the CUSUM detector which fulfils these requirements acts as the base to build the change detector.

Revised CUSUM: The basic CUSUM detector could sensitively detect changes occurring in the time domain. However, it cannot be directly applied to the problem for two reasons. First, basic CUSUM does not provide a way to estimate the starting time of a change. In the basic CUSUM, once the detection function exceeds the threshold, an alarm is issued and the CUSUM detector restarts. The stopping time denotes the time when the change has been detected but not the actual change starting time. We need to trace back on the change detection function to estimate the change starting time. Second, the change may last for some time. Restarting the detector will make it impossible to trace the change ending time.

Procedure 1: Change detection procedure

```
//the set containing items under attack
for each item do
    collect all ratings for and order them according to the
    time that they are provided
    //a flag, which is 1 when exceeds threshold
    Compute
    //Under attack or not
    if then
        add to
        //Get change starting time and ending time
        for each rating do
            if then
                if then
                    Estimate change starting time
                end if
            else
                if then
                    Estimate change ending time
                end if
            end if
        end for
    end if
end for
```

III. Trust Model Based On Dempster Shafer Theory

The Dempster Shafer Theory is a framework for combining evidence from different sources to achieve a degree of belief. It has introduced the concept of "uncertainty" by allowing the representation of ignorance. The users who provide ratings during the detected change intervals are termed as suspicious users. Not all suspicious users are malicious users because normal users may occasionally provide biased ratings due to personal reasons or even human errors. Therefore, we propose to further differentiate normal users from malicious users by trust analysis. In most trust models, users trust values are determined only by their good and bad behaviors. However, it is not sufficient. Consider two trust calculation scenarios. First, user has conducted 5 good behaviors and 5 bad behaviors. Second, user is a new coming user and has no behavior history. In several trust models, both of their trust values will be calculated as 0.5, although there is much confidence in user's trust value. To differentiate these two cases, the concept of behavior uncertainty is introduced by the Dempster–Shafer theory, to represent the degree of the ignorance of behavior history. In this work, we adopt the behavior uncertainty by proposing a trust model based on the Dempster–Shafer theory.

Behavior Value: We define a user's behavior *value* on a single item as a binary value to indicate whether his/her rating behavior is good or bad.

Combined Behavior Value: To evaluate users' behaviours on multiple items, we introduce the combined behaviour value,

Behavior Uncertainty: Similarly, we define a user's behaviour uncertainty using the Dempster–Shafer theory.

IV. Malicious User Identification And User Aggregation

Here, the trust values are examined of each user. The detection of the users with low trust values on items as malicious users determined. Instead of removing all the ratings provided by the malicious users, only their ratings that yield low trust values are removed. Specifically, for user a if Ta(i)<Tb, user a's rating to item Ii is removed and a is marked as malicious user. Here, Tb is called the trust threshold, which could be adjusted according to different application scenarios. In the testing data that is used, most normal users provide more than 10 normal ratings, while malicious users provide less normal ratings. Based on the trust model, if a user has provided 1 suspicious rating to item, and 10 normal ratings to other items, his/her trust value on item is calculated as 0*(2/12)+(10/12)*(10/12)=0.694. Therefore, we choose the Tb as 0.69 in the experiments below, so that a malicious user has to provide at least 10 other normal ratings to cover his/her dishonest rating to the target item. After the rating removal, our method is compatible with any existing reputation schemes that calculate the item quality reputation. Without loss of generality, we use simple averaging in this paper for computing item quality reputation.

V. Related Work

The Dempster Shafer Many people are involved in doing a lot of manipulations in the online reputation systems, defense schemes protecting reputation systems are also evolving accordingly. It is divided into four categories. In the first category, the defense approaches limit the maximum number of ratings each user could provide within certain time duration. Such type of approaches actually restricts the rating power of each user ID. This can prevent the attackers from inserting a large amount of dishonest ratings through a few user IDs within a short time. In the second category, the defense schemes aim to increase the cost of launching an attack. Some reputation systems in practice, such as Amazon, assign higher weights to users who commit real transactions. This method can effectively increase the cost to manipulate competitor's item reputation. However, it has little impact on attacks in which attackers buy their own products for reputation boosting. Some other schemes increase the costs of acquiring multiple user IDs by binding identities with IP addresses or using network coordinates to detect Sybil attacks. Such schemes will greatly increase the attack costs, but cannot defeat the attackers with plenty of resources. For example, some reputation boosting companies often acquire a large affiliate network of user IDs. In the third category, the defense approaches investigate rating statistics. They consider ratings as random variables and assume dishonest ratings have statistical distributions different from normal ratings. Representative schemes are as follows. A Beta-function based approach assumes that the underlying ratings follow Beta distribution and considers the ratings outside (lower) and (upper) quantile of the majority's opinions as dishonest ratings. An entropy based approach identifies the ratings that bring a significant change in the uncertainty of the rating distribution as dishonest ratings. In, dishonest rating analysis is conducted based on Bayesian model. Controlled anonymity and cluster filtering are used to eliminate dishonest ratings in. The defense approaches in the fourth category investigate users rating behaviors. Assuming that users with bad rating history tend to provide dishonest ratings, such approaches determine the weight of a rating based on the reputation of the user who provides this rating. Such reputation is also referred to as trust or reliability. Several representative schemes are as follows. Iteration refinement approach proposed in assigns weights to a user's ratings according to the inverse of this user's rating variance. A personalized trust structure is introduced so that different users may assign different trust values to the same user. A user's trust is obtained by accumulating neighbours' beliefs through belief theory. REGRET reputation system calculates user reputation based on fuzzy logic. Flow models, such as Eigen Trust and Google PageRank, compute trust or reputation by transitive iteration through looped or arbitrarily long chains.

VI. Conclusion

This paper analyses a comprehensive anomaly detection scheme, TATA, which is designed for protecting feedback-based online reputation systems. In order to analyze the time-domain information; a revised-CUSUM detector is used to detect change intervals. So as to further reduce false alarms, a trust model based on the Dempster–Shafer theory is introduced. When the number of malicious users is not very large, examining individual user's behavior (such as through a well designed trust model) is a very effective defense approach. When the number of malicious users is very large, investigating user behavior similarity (such as in the TAUCA scheme) becomes a promising method. In the future, one possibility is to jointly consider trust evaluation and user correlation. This future approach can be used to reduce the time consumption and increase the efficiency.

References

[1] J. Weng, C. Miao, and A. Goh, "An entropy-based approach to protecting rating systems from unfair testimonies," IEICE Trans. Inf. Syst., vol. E89-D, no. 9, pp. 2502–2511, Sep. 2006.

[2] M. Abadi, M. Burrows, B. Lampson, and G. Plotkin, "A calculus for access control in distributed systems," ACM Trans. Program. Lang. Syst., vol. 15, no. 4, pp. 706–734, 1993.

[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 IEEE Int. Conf. Social Computing, Aug. 2010, pp. 65–72.

[6] 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.

[7] Q. Zhang and T. Yu, "On the modeling of honest players in reputation systems," in Proceedings of IEEE ICDCS Workshop on Trust and Reputation Management, 2008.

[8] Y. Yang, Y. Sun, J. Ren, and Q. Yang, "Building trust in online rating systems through signal modeling," in Proceedings of IEEE ICDCS Workshop on Trust and Reputation Management, 2007.