Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Representation and querying of unfair evaluations in social rating systems

Mohammad Allahbakhsh [a,*], *Aleksandar Ignjatovic* [a], *Boualem Benatallah* [a], *Seyed-Mehdi-Reza Beheshti* [a], *Norman Foo* [a], *Elisa Bertino* [b]

[a] *University of New South Wales, Sydney 2052, Australia*
[b] *Purdue University, West Lafayette, IN, USA*

ARTICLE INFO

Article history:
Received 23 May 2013
Received in revised form
26 August 2013
Accepted 30 September 2013

Keywords:
Unfair evaluation
Reputation
Degree of fairness
Collusion
Biclique
Rating system

ABSTRACT

Social rating systems are subject to unfair evaluations. Users may try to individually or collaboratively promote or demote a product. Detecting unfair evaluations, mainly massive collusive attacks as well as honest looking intelligent attacks, is still a real challenge for collusion detection systems. In this paper, we study the impact of unfair evaluations in online rating systems. First, we study the individual unfair evaluations and their impact on the reputation of people calculated by social rating systems. We then propose a method for detecting collaborative unfair evaluations, also known as collusion. The proposed model uses frequent itemset mining technique to detect the candidate collusion groups and subgroups. We use several indicators to identify collusion groups and to estimate how destructive such colluding groups can be. The approaches presented in this paper have been implemented in prototype tools, and experimentally validated on synthetic and real-world datasets.

1. Introduction

Due to huge number of people and contents involved in Web 2.0 enabled systems, assessing quality of people or contents is crucial to the success of these systems. For example, in an e-commerce system, trustworthiness of the parties plays a key role when people choose to sell or purchase goods. Also, the quality of products, items and services provided on the web is the key factor when a person intends to hire someone, buy a product or use an online service.

Relying on evaluations cast by others is a very common approach to assess quality when dealing with people or contents in the web (Allahbakhsh et al., 2013a; Feng et al., 2012; Doan et al., 2011; Quinn and Bederson, 2011). People can easily share their opinions or experience of using a service, purchasing a product or hiring a person with other community members, in order to help them judge an item or a person who have no direct experience with. *Social Rating Systems* collect and aggregate such opinions to build a rating score or level of trustworthiness for items and people (Feng et al., 2012). These rating scores reflect the overall quality of the person or item from community's point of view. Reputation systems and product rating systems are two major types of social rating systems.

* *Corresponding author*. Tel.: +61 423 688 680.
E-mail addresses: mallahbakhsh@cse.unsw.edu.au (M. Allahbakhsh), ignjat@cse.unsw.edu.au (A. Ignjatovic), boualem@cse.unsw.edu.au (B. Benatallah), sbeheshti@cse.unsw.edu.au (S.-M.-R. Beheshti), norman@cse.unsw.edu.au (N. Foo), bertino@cs.purdue.edu (E. Bertino).
0167-4048/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cose.2013.09.008

Reputation systems are cornerstones of human centric systems such as crowdsourcing systems on the web. Crowdsourcing involves receiving, incorporating and consolidating contributions from a large crowd with varied levels of expertise. The people who own crowdsourcing task are called *requesters* and the people who do the tasks are called *workers* (Allahbakhsh et al., 2013a). Due to lack of enough information, lack of expertise, dishonesty of workers or evaluators, bias in user interests and many more reasons (Allahbakhsh et al., 2013a; Ipeirotis et al., 2010; Agichteinet al, 2008), quality of contributions in crowdsourcing tasks is always under question. The overall quality of outcome of a crowdsourced task depends on the quality of the workers, the processes which govern the task creation, selection of workers, coordination of sub-tasks including reviewing intermediary outcomes, aggregation of individual contributions, etc. Using reputation as an indicator of community-wide judgment on workers' trustworthiness is a popular method for evaluation of the quality of workers in existing crowdsourcing platforms (Allahbakhsh et al., 2013a; De Alfaro et al., 2011).

The key role of the quality metrics calculated by social rating systems motivates people to manipulate such scores by posting *unfair evaluations* (Brown, 2006; Harmon et al., 2004). Unfair evaluations are evaluations which are cast regardless of the quality of a product or a person and usually are given based on personal vested interests of the users. For example, providers may try to submit supporting feedback to increase the rating of their product in order to increase their income (Harmon et al., 2004). The providers also may attack their competitors by giving low scores on their products. Also, sometimes sellers in eBay boost their reputations unfairly by buying or selling feedback (Brown, 2006). Unfair evaluations may be given individually or collaboratively (Swamynathanet al, 2010). Collaborative unfair evaluations are also called *collusion* (Sun and Liu, 2012; Swamynathanet al, 2010).

1.1. Problem statement

Although both individual and collaborative unfair evaluations has been well studied in the literature (Ciccarelli and Cigno, 2011; Kamvar et al., 2003; Lee et al., 2010; Mukherjee et al., 2012; Lim et al, 2010; Kerr, 2010; Yang et al., 2008), there are still some issues which need more research and investigations. We can simply highlight some of the existing issues as follows:

- Most of the existing models, while aggregating individual evaluations, usually ignore some important aspects of evaluation. The first aspect is the time on which the evaluation has been cast. The quality of people or items may change over time, hence the evaluations they receive in time may be different correspondingly. The more recent the evaluations, the more credible they should be while are taken into account. However, in existing models, time is not considered. The other important aspect is the credit of transaction corresponding to that particular evaluation. The feedback received for doing well in a high credit (e.g., monetary credit) task should have higher impact on the final quality metric (i.e., reputation score of the person) than evaluations received for low credit transactions. This aspect is also missing in existing work.
- Sometimes colluding evaluators try to completely take control of a product. In such cases, the number of unfair reviewers is significantly higher than the number of honest users. Existing models cannot detect such groups. Also, the existing models do not perform well against intelligent attacks, in which group members try to give an appearance of honest users. For example, typically they will not deviate from the majority's ranking on most of the cast feedback and target only a small fraction of the products. Such attacks are hard to identify using the existing methods (Yang et al., 2008).

1.2. Contributions and outline

To address the above mentioned problems, we first propose a method for detecting individual unfair evaluation. We study this case in the context of reputation management in crowdsourcing systems. We present a method in which, in addition to the evaluations, the trustworthiness of the evaluators, the time in which each evaluation has been posted and the credit paid for the task corresponding to each evaluation are also considered when calculating workers' reputation scores. To detect dishonest evaluators (e.g., outliers, self-promoters, etc.) we analyze the behavior of evaluators in three different levels: short time intervals, pairwise relations and community-wide. This contribution of the paper has been the main idea of one of our recent work published in Allahbakhsh et al. (2012). In the current paper, we extend this work and added additional people evaluation parameters and an algorithm mainly for calculating overall degree of fairness of evaluators.

As the second significant contribution of this research, we study impact of collusion in online rating systems and assess their susceptibility to collusion attacks. We propose a model which uses Frequent Itemset Mining (FIM) technique (Agrawal and Srikant, 1994) to detect candidate collusion groups. Then, several indicators are used for identifying collusion groups and to estimate how damaging such colluding groups might be. This part of the work partially has presented in our previous work (Allahbakhsh et al., 2013b). Here, in current research, we extended this work by changing the main collusion detection algorithm, extended our query language and performed new performance evaluations.

In summary, the unique contributions of the paper are as follows:

1. We propose a graph data model to facilitate the analysis of unfair evaluations in social rating systems. This model allows representing evaluators, the items which are being evaluated (workers or products), evaluations cast on the quality of these items, pairwise trust, and the degree of fairness. We also propose a new notion and representation for collusion groups called *biclique*. A biclique is a group of users and a group of products such that every reviewer in such a group has rated every product in the corresponding group of products. We use bicliques to detect collusion.
2. We propose a new measure for assessing fairness of evaluators when evaluating workers' contributions and call it *degree of fairness*. This metric is calculated according to