

Automatic recommendation of trustworthy users in online product rating sites

Betsy Baby¹, Soumya Murali²

¹Sree Buddha college of engineering, Alappuzha, India

²Sree Buddha college of engineering, Alappuzha, India

Abstract - The online product rating sites provide recommendations for the users on different items. Some of the recommendations are valuable to the user but others might be misleading and harmful. It is because some recommenders might have malicious intentions or not the required competence. When a user add other users to his trust list some fake users with high rating on products also get into the list and corrupt the rating prediction for the user. The ratings of fake users will be very different from actual ratings. It will reduce the accuracy of rating prediction system. Thus, we have to find out and to decide carefully whom we can trust. Users can't always validate the trustworthiness of everyone providing recommendations.

Key Words: Recommender systems, user trust score, online social networks (OSN).

1. INTRODUCTION

The Recommender Systems are software tool and technique providing for interest items to a user. The suggestions provided are aimed at supporting their users in numerous decision-making processes, such as what items to buy, what music to listen, or what news to read. The recommender systems have proven to be valuable means for online users to deal with the information excess and have developed one of the most powerful and popular tools in electronic commerce. Numerous techniques for recommendation generation have been proposed and during the last period many of them have been successfully deployed in ecommerce environments. Trust relations between the users have a great impact on the recommendation task. The trusted users can influence the rating of a product. Many of the products rating sites like epinion, ciao etc allow users to create their own trust list and block list. The trust list contains the users trusted by the particular user and block list contains the users the particular user thinks to be as untrustworthy. The user cannot always able to validate the trustworthiness of every users providing recommendation

for them. While adding users to the trust list some fake or untrustworthy users also get into the trust list. The Feedbacks, ratings, scores, recommendations and any other information given by users are very important for the trust calculation. The users are interested in referring to the reviews of the users, in order to conceive their own trust and reputation on different products. By this reviews of the product it trusts the product and purchases it. However we must need to verify the reliability of reviews to ensure only genuine review is taken into consideration. So we need to identify misleading behaviours of dishonest or fake users. They will purposefully distract users into buying the products that are not qualified into an intended good category.

2. LITERATURE SURVEY

Reputation systems are commonly used in online transactions but they are suffering from "all good reputation" problem. The high reputation score of every seller makes it difficult for the buyer's to choose amongst them for the transaction. This paper [1] has proposed a reputation system combining the concepts of Natural Language Processing, Opinion Mining and Summarization. They have computed the reputation score and ranked the sellers. The reputation score is the weighted summation of criteria based reputation ratings for each seller. This combines the knowledge out of user ratings given in the form of five-star ratings and the opinions expressed in the form of reviews. This is able to solve "all good reputation" problem.

Reputation is an estimation of the trust the community has built in you. It is a complex and context-dependent opinion of the community about any entity in question. It is highly influential in ecommerce applications. A reputation system is a system that can assists people selecting whom to trust. The

“all good reputation” problem is common in the e-commerce domain, making it difficult for buyers to choose credible sellers. Observing high growth of online data in Hindi language, in this paper [2] they proposed a reputation system in this language. The functions of this system include 1) review mining for different criteria of online transactions 2) calculation of reputation rating and reputation weight for each criterion from user reviews and 3) ranking sellers based on computed reputation score.

As the rapid development of e-commerce, reputation model has been proposed to help customers make effective purchase decisions. However, most of reputation models focus only on the overall ratings of products without considering reviews which provided by customers. We believe that textual reviews provided by buyers can express their real opinions more honestly. As so, in this paper [3] proposed a Textual Reputation Model (TRM) to obtain useful information from reviews, and evaluate the trustworthiness of objective product. It is based on the word2vector model.

In this paper [4] they studied the influence of interaction attributes on trust and friendship probability in virtual communities and proposed a novel interaction ranking-based model (IntRank) for recommending trustworthy friends to community members. They also identified and validated four new interaction attributes: reply frequency, comment length, time difference, and domain similarity that influence trust and friendship probability in virtual communities by performing logistic regression analysis on real and large interaction and friendship data from Slashdot. This novel model has several advantages over many traditional friend recommendation mechanisms. Because IntRank provides a more comprehensive approach to recommend trustworthy friends with whom the user has intensively interacted before. It maintains the social network as a “manifestation of trust”.

Nowadays, the internet plays a significantly role in our daily lives. Online users usually select products by referring to their peer’s ratings on that product. Without a reputation system, there is a risk that user’s choices might be misled by some spamming ratings. The method used in this paper [5] is not only effective in estimating the true quality of the objects but also very robust to spamming ratings. This paper also introduces two penalty factors to the iterative ranking algorithm which can significantly improve its robustness against the malicious spamming behaviours. In this paper,

they proposes a robust iterative ranking algorithm. It is a reputation redistribution process, which can effectively enhance the weight of the highly reputed users and lower the weight of the users with low reputation in estimating the quality of objects.

On Amazon, many purchase reviews by the users are dishonest spam entries. It’s main aim is to skew product ratings. Though users have the opportunity to rate reviews as helpful or unhelpful, socio- logical factors and prior ratings influence users to rate these reviews for reasons other than the truth of their content. Many studies have evaluated the content of these user reviews to detect spam entries by mining and classifying the text entry. This paper [6] proposes a graph based algorithm to determine the honesty of reviews and trustworthiness of the reviewer for general product review data. They propose to apply this algorithm to Amazon review data and compare helpfulness data to the resulting quantitative assessments of honesty of reviews and trustworthiness of reviewers. They also analyzed the trust of the reviewers and their correlation with helpfulness in order to classify spam.

QoS identification for untrustworthy web services is critical in QoS management in the service computing. Because the performance of untrustworthy Web services may result in QoS downgrade. The key issue is to intelligently learn the characteristics of trustworthy Web services from different QoS levels, and then to identify the untrustworthy ones according to the characteristics of QoS metrics. In this paper [7] they proposed a two-phase neural network model to identify the untrustworthy Web services. In the first phase, the web services are collected from the published QoS dataset. Then designed a feedforward neural network model to build classifier for the web services with different QoS levels. In the second phase, they used a probabilistic neural network (PNN) model to identify the untrustworthy Web services from each classification.

In this paper [8] they have presented a new reputation model for products. This model used text reviews rather than user’s ratings for reputation calculation. They extracted opinions about different features and calculated the frequencies for positive and negative opinions. It is because frequent negative opinions about features and sub-features should get more weight in the reputation calculation, as they indicates a problem in a product a customer may face if they buy it. Additionally, they calculated the impact of features because certain features in the some products are more inspiring for

users than the others. Therefore they are more important in the reputation model. The model integrates strength of opinions and provides summary about user's opinions for customers rather than representing reputation as a number of stars.

Viral marketing is becoming very popular due to the large population base in OSNs. It opens the door for shill attacks. Shill attacks are the attacks where malicious users make wrong recommendations to distort the sales market. In this paper [9] they propose to defend against shill attack in OSNs based on the idea of shrinking suspicious set. They considered both the baseline shill attack and the intelligent shill attack in the work. They developed a set of distributed and randomized detection algorithms to identify dishonest users who give misleading recommendations in OSNs. The proposed system allows each honest user to independently perform the detection so as to discover her dishonest friends. This detection framework can be viewed as a valuable tool to maintain the viability of viral marketing.

Online product reviews have become an important source of user opinions. Due to profit, the fake reviewers have been writing deceptive or fake reviews to promote and/or to demote some target products. Such fake reviewers are called review spammers. In this work [10] they take a different approach, which exploits the burstiness nature of reviews to identify review spammers. Bursts of reviews can be either due to sudden popularity of products or spam attacks. Reviewers and reviews appearing in a burst are considered in the sense that spammers always tend to work with other spammers and genuine reviewers always tend to appear together with other genuine reviewers. This makes the sense to build a network of reviewers appearing in different bursts. They modeled reviewers and their co-occurrence in bursts as a Markov Random Field (MRF), and employed a Loopy Belief Propagation (LBP) method to infer whether a reviewer is a spammer or not in the graph. They also evaluated the detected spammers automatically using supervised classification of their reviews.

Many online social networks are using the past friend recommendation method. There is a huge requirement to implement privacy-preserving friend recommendation methods for social networks. It is because maintaining user privacy is a major issue. This paper [11] proposed a system to assist online social network (OSN) users to establish trust with strangers securely. It is done by multi-hop recommendation process. In the system any user can utilize their attributes to discover a new friend, and to set up social

connections with the outsiders by means of a multi hop trust chain. The proposed method allows OSN users to enhance social circles with maintaining their privacy.

On product review sites, people may write faked reviews to promote their products, or defame their competitor's products. It is important to identify and filter out these faked reviews or review spams. In this paper [12] they studied the review spam identification in product review mining system. For that they manually build a review spam collection based on the crawled reviews. Then they used supervised learning methods and analyzed the effect of different features in review spam identification. They mainly classified the features as review related and reviewer related features. Review related features include: content features, sentiment features, product features and meta data features. Reviewer related features include: profile features and behaviour features.

This work [13] proposes a novel angle to the problem by modeling spamicity as latent. An unsupervised model, called Author Spamicity Model (ASM) is proposed. It works in the Bayesian framework which facilitates modeling spamicity of authors and allows us to exploit various observed behavioural footprints of reviewers. The intuition is that opinion spammers have different behavioural distributions than non-spammers. This creates a distributional divergence between the population distributions of the two clusters: spammers and non-spammers. Model inference results in learning the population distributions of the two clusters.

Nowadays, the anonymous nature of the Internet has made it difficult to detect the manipulation of user reputations in the social media. Also it make difficult to ensure the qualities of users and their posts. To deal with this issue, this paper [14] designed and examined an automatic approach that adopts writing style features to estimate the user reputations in social media. Under the different ways for defining the Good and Bad classes of user reputations based on the collected data, this paper uses the four writing style features of the user posts. They are the lexical, syntactic, structural, and content-specific features. In this paper eight classification techniques, i.e. C4.5, Neural Network, Support Vector Machine, Naïve Bayes and four Random Subspace (RS) ensemble methods are used and the accuracy is measured.

All social media platforms allow users to create a group of friends with whom they shares their personal and social information. There are no tools to suggest a trust worthy

friend to a user. This lack of trust verification system would potentially lead to security issues like identity theft, cyber stalking etc. A method for calculating trust level of a user on social networking sites is needed This paper [15] proposes a system which calculate the trust score of a user based on the weighted average score of information related to him. The method enables a user to create web of trust on the web based social platform. The web of trust includes the most trusted friends, average trusted friends and least trusted friends.

Acknowledgements

We are thankful to our project guide Prof. Soumya Murali and PG Coordinator Prof. Minu Lalitha Madhavu for her remarks, suggestions and for providing all the vital facilities like providing the Internet access and important books, which were essential. We are also thankful to all the staff members of the Department of Computer Science & Engineering of Sree Buddha College of Engineering, Alappuzha.

REFERENCES

[1] Latent aspect rating analysis on review text data: A rating regression approach Hongning Wang, Yue lu, Chengxiang Zhai department of computer science university of Illinois .

[2] Multiple aspect ranking using the good grief algorithm in snyder and regina barzilay computer science and artificial intelligent{bsnyder,regina}@csail.mit.edu

[3] A joint model of text and aspect ratings for sentiment summarization ivan titov department of computer science university of illinois at urbana-champaign urbana, il 61801 titov@uiuc.edu

[4] Rated aspect summarization of short comments yue lu department of computer science university of Illinois at Urbana department of computer science university of Illinois at Urbana Champaign Urbana, il 61801 czhai@cs.uiuc.edu

[5] Latent aspect rating analysis without aspect keyword supervision Hongning Wang, Yue lug, Chengxiang Zhai department of computer science university of Illinois at

urbana-champaign Urbana il, 61801 usa {wang296, yuelu2, chai}@cs.uiuc.edu

[6] Modelling online reviews with multi-grain topic models Ivan titan department of computer science university of Illinois at urbana-champaign Urbana, il 61801 titov@uiuc.edu Ryan McDonald google inc. 76 ninth avenue new York, NY 10011 ryanmcd@google.com

[7] Jointly modelling aspects and opinions with a maxent-lda hybrid Wayne xin zhao†, jing jiang‡, hongfei yan†, xiaoming li† †school of electronics engineering and computer science, peking university, china

[8] Extracting and visualizing trust relationships from online auction feedback comments john o'donovan and barry smyth adaptive information cluster school of computer science and informatics university.

[9] Formalising trust as a computational concept stephen paul marsh department of computing science and mathematics university of stirling

[10] Inferring trust relationships in web-based social networks jennifer golbeck, james hendler university of maryland, college park computer science department college park, md 20742 {golbeck, handler}@cs.umd.edu

[11] Trust in multi-agent systems school of electronics and computer science, university of Southampton, Southamuk esdr01r@ecs.soton.ac.uk utdh02r@ecs.soton.ac.uk; nrj@ecs.soton.ac.uk

[12] The Eigen trust algorithm for reputation management in p2p networks spender d. kava Stanford university.

[13] Design of feedback analysis system for deciding trustworthy e-commerce organization sonali rakhi d.wajgi m. tech student, dept. of computer science & engg.

[14] Rlm: a general model for trust representation and aggregation ieee, ling liu, senior member, ieee and jinshu su, member, ieee

[15] Identifying honest recommenders in reputation systems farag azzedin king fahd university of petroleum and minerals, information and computer science department, dhahran, 31261, saudi arabia.