

A Trust and Reputation Model for Quality Assessment of Online Content

Yousef Elsheikh

Department of Computer Science
Applied Science Private University
Amman 11931, Jordan, PoBox 166

Abstract—In recent years, online transactions have become more prevalent than it was. This means that the number of online users to perform such transactions keeps growing, causing an increase in the level of expectations for them. One of those expectations is to enable them to get a better understanding of such transactions before going ahead with it. Consequently, trust and reputation models represent an important milestone to support those users to make their own decisions to facilitate online transactions. Many of the common trust and reputation models used primitive methods to calculate the reputation of online content. These methods are usually inaccurate when there is a divergence in rating. In addition, the lack of predictability through the latter ratings in emerging trends. Others use a probabilistic model or the so-called weighted average, which usually focusing on a single dimension for online user ratings. Even those models that combine multiple dimensions of user ratings are usually not representative on the one hand, and on the other hand are with heterogeneous weights. This paper fills this gap by proposing a model to assess the trust and reputation of online content, relying on three factors namely user behavior, user reliability, and user tendency with homogeneous weights of interest to the user on the Internet. These homogenous weights will be used to measure the reputation of any online content. The proposed model has been validated and compared with some other well-known models, and showed a significant improvement in terms of the Mean Absolute Error (MAE). The proposed model is also good with sparse and dense datasets.

Keywords—Online content; Quality assessment; Trust and reputation model; User behaviour; User reliability; user tendency

I. INTRODUCTION

In recent years, online transactions have become a phenomenon [8]. However, many users are still reluctant to make such online transactions and the reason for this, the inability to assess online content, where the disparity in opinions and feedbacks provided by other users about that online content. Feedbacks are usually offered either in the form of ratings or reviews, as these are among the primary sources to assess the quality of online content [1]. Online content quality is usually assessed through a model of trust and reputation, which in turn collects processes and aggregates user ratings on a particular online content. The main component in such models is how to aggregate ratings. The results of such models is to assess the quality of online content in the form of either a numeric value or stars as is the case for some common models [1].

Recently, many of the trust and reputation models were used to evaluate the different forms of online content. Most of these models are used on a large scale, easy to obtain, for free at most and help in decision-making by the user [4]. However, the accuracy of trust and reputation models are always a source of interest to many users over the Internet because they are often reflect public opinion on a specific online content. The main challenge in the trust and reputation models is how to aggregate the user ratings to assess the quality of online content. The easiest solution is to use primitive methods to achieve a score of trust and reputation regarding specific online content. Despite the simplicity of these methods, they are not effective enough since they do not take into account the quality and reliability of the user or even the popularity of online content to be evaluated. It also cannot predict the trends emerging from recent ratings of the user [8]. Other methods were more mature, such as those used probabilistic and fuzzy logic models. These methods have achieved more accurate results than its predecessors, but it depends largely on the threshold points, which are set by the experts. However, some of the models used the weighted average in order to calculate the score of the reputation of any online content, since the weights may be the reputation of the user, user reliability [2], User leniency [7], rating time [4], or the difference between the current reputation score and the new rating score [11]. This method requires evaluating user rating to assess the quality and trustworthiness of online content, and thus reflect that weight.

This paper focuses mainly on aggregating the ratings using the weighted average method. Most current weighted average models of trust and reputation mostly focus on a single dimension of the weights for the user, for example, in the Lenient-Quality reputation model; the weight of the user is evaluated on the basis of being lenient or not in providing rates [7]. Even models that combine multiple user dimensions they often combine them through a discount function, and significantly associated with thresholds set by the experts [9]. Moreover, some of the weighted average methods do not take into account changes in the weights of the user in terms of reliability of the user, the time variation between the first rating and the latest rating of the same online content, and finally the user experience compared with the experience of other users about specific online content. Also, current methods of weighted average do not take into account user tendencies (positive or negative) during the process of rating online content. Therefore, this paper proposes a new model of trust

and reputation to assess the quality of online content through a combination of three factors and of great importance to the user on the Internet. These factors have been nominated to reflect the user's reputation as a value for the weight. These factors are: 1) user behaviour, 2) user reliability, and 3) user tendency.

II. RELATED WORK

In a review of the literature, trust and reputation models fall into three basic categories namely the weighted average models, fuzzy logic models, and probabilistic models. The Weighted average models are the most commonly used where weights are calculated based on time or data relating to the user. In a study of [5], they used non-linear function of aging, depending on the time-based approach in calculating the weights. In an example of the same approach, the number of previous transactions has been addressed while the times of those previous transactions have not been addressed [10]. All common models mentioned above are not able to adapt to the data when it grows in size. In a study of [8] & [13], they addressed the variation in the user's rating, which was reflected on the user's decision on that item. Moreover, these models often do not take into account the credibility dimension in the user's rating, which reflects negatively on the accuracy of those ratings that are placed.

The other approach of the weighted average models is based on data relating to the user, where the weights are calculated by the reliability of the user, the user's credibility and trustworthiness to the user. In a study of [11], they proposed a model to measure the reliability of the user through the same user ratings. Higher weights were given to the users with ratings that are closer somehow to the average ratings made by the user for a specific online content. Another study of [7] & [12], they addressed and proposed a model that is based on rating user behaviour and tendency in order to measure the leniency of the user. Rating user behaviour or tendency is a value that reflects the extent of user's behaviour or tendency to provide ratings higher or lower than other users. Also, a study of [6] proposed a model of trust and reputation takes user ratings in the account through what is known as a polynomial probabilistic Bayesian probability distribution and Dirichlet distribution. In a study of [2], they proposed a model of trust and reputation for aggregating ratings through mixing in use between the weighted average method and fuzzy logic. User reputation depends on the accuracy of prediction compared with the ratings of other users for various elements in any online content. In a study of [9], they recently addressed the problem of unfair ratings provided by some users by proposing a model of reputation and trust uses fuzzy logic to address that problem.

However, all the above models lacked during the evaluation of the user's reputation to a comprehensive approach that combines three factors making up the model of this paper. In addition, the proposed model in this paper is also good with sparse and dense data sets as opposed to the rest of the models listed above.

III. PROPOSED MODEL

The proposed model uses three input factors that have great impact on the user's trust and reputation. These factors are 1)

user behaviour, 2) user reliability and 3) user tendency, which are then fused together through the Arithmetic Mean to reflect the user's reputation of any online content as a value for the weight.

The first factor measures the user's behaviour in making the rating. In other words, it is measured when the user usually makes his/her rating. Is he/she the first who assessed online content? Or did he/she make his/her rating after many previous transactions? Rating time is important as it reflects the impact of past transactions on the user's decision. Therefore, time difference between the user rating and first rating received is measured in terms of day unite. Then, these are discounted using age decay function as shown in Equation 1.

$$y_1 = \frac{1}{m} \sum_j^m \mu^{T_{Cj} - T_{1j}} \quad (1)$$

Where m is the number of online items rated by a user. μ is the discounting variable (in this case we use $\mu = 0.95$). T_{Cj} is the timestamp when the user rated the online item j . T_{1j} is the timestamp of the first rating received for online item j .

The second factor measures the user's reliability. This factor assesses the accuracy of user in providing rating that is very close to the average of rating for the online item under assessment. Equation 2 shows how the distance between the user rating and average of ratings are discounted using discounting function μ^x .

$$y_2 = \frac{1}{m} \sum_j^m \mu^{|r_j - avg_j|} \quad (2)$$

Where r_j is the rating given by a user for online item j . avg_j is the rating average of online item j .

The third factor assesses the user's tendency in providing positive, neutral or negative ratings. Positive ratings are those that are larger than mid of the rating level, and negative ratings are the opposite while neutral ratings are those with the mid-range value. Rating level is the scale used to score the item, for example in most application the rating scale ranges from 1 to 5. The basic idea of this variable is to find the ratio between errors of either user positive ratings or negative ratings, and total ratings errors. The error is computed by finding the difference between a rating and average of the online item rating. Equation 3 shows how the user tendency variable is calculated. In this first step we classify user ratings into three sets, positive, neutral and negative set. To decide which set should be used, one should assess the rating of the online item under assessment, if the rating belongs to the positive set then we use positive error otherwise we use either negative or neutral set.

$$y_3 = 1 - \frac{\sum_{k=1}^L |r_k - avg_k|}{\sum_{j=1}^m |r_j - avg_j|} \quad (3)$$

Where L is the number of ratings in the targeted set, avg_k is the average of online item k in the targeted set.

The three factors y_1 , y_2 and y_3 are fused together using the Arithmetic Mean as shown in Equation 4. The final online item score is computed as shown in Equation 5.

$$w_i = \frac{y_1 + y_2 + y_3}{3} \tag{4}$$

$$score_j = \sum_{i=1}^n w_i \times r_i \tag{5}$$

Where n is the number of user ratings for online item j . w_i is the normalized weight for user i and r_i is the rating provided by user i .

IV. EXPERIMENT DESIGN

To validate the proposed trust and reputation model we applied 5-Fold cross validation, which divides the dataset into 5 sets of training and testing data. In each run, 80% of the users are used as training data to build trust and reputation model and 20% are used as testing data to validate the data against the generated score. The errors of validation at each run are recorded using the Mean Absolute Error (MAE) as shown in Equation 6. The MAE (MAE) assesses, for each online item, the closeness of the predicted items scores from the training dataset to the actual ratings in the testing dataset.

$$MAE = \frac{1}{m} \sum_{j=1}^m \frac{\sum_{i=1}^n (r_i - \bar{r}_j)}{n} \tag{6}$$

Where \bar{r}_j is the predicted score for online item j . m is the number of online items in the testing data. n is the number of ratings for j^{th} online item in the testing data.

To investigate the performance of our proposed model we employ two common stable datasets that are publically available on the internet. These two datasets are 100K and 1M which are taken from large benchmark data repository called Movielens. Both datasets contain ratings for movies. The first dataset (100K) consists of 943 users and 1682 movies, whereas the second dataset consists of 6040 users and 3706 movies as shown in Table 1. Each user in both datasets has rated at least one item online and each item has been rated by at least one user. Both datasets have been widely used in validating trust and reputation models.

TABLE I. DATASETS CHARACTERISTICS

Dataset	No. Of Users	No. Of Movies	No. Of ratings
100K	943	1682	100,000
1M	6040	3706	1,000,209
10M	72,000	10,000	10,000,000

V. EXPERIMENT RESULTS

To show the significant improvement, which we obtained in the results, we compared the proposed model with a range of well-known models for trust and reputation in the literature. Comparisons were made between the proposed model with six other models of trust and reputation namely the Average Based Reputation Model, Beta Distribution Based Reputation Model [1], Bayesian Reputation Model [6], Dirichlet Based

Reputation Model [5], Fuzzy Logic Rating Based Reputation Model [2] and Lenient-Quality Reputation Model [7]. Table 2 shows for each model, the value for Mean Absolute Error (MAE) obtained on all databases [14].

TABLE II. MEAN ABSOLUTE ERROR RESULTS

Dataset	Proposed Model	Average	BetaDR	Bayesian	Dirichlet	Fuzzy Logic Rating	Lenient-Quality
100K	0.826	0.905	0.893	0.911	0.898	0.916	1.024
1M	0.780	0.841	0.833	0.844	0.841	0.848	0.962
10M	0.749	0.791	0.812	0.790	0.776	0.795	0.917

It is noted above, the results of the proposed model provides better accuracy compared with other models. This shows, through the MAE value, that the proposed model has a lower MAE value than the rest of the models used, which means that it is better in terms of accuracy. Moreover, the above results confirmed that the proposed model can work well over the sparse and dense dataset alike. Sparse dataset is often a problem for many trust and reputation models as they do not work well when dealing with a small number of ratings. Although the Dirichlet model can work with uncertain, small number of ratings, the above results show that the proposed model was significant, compared with Dirichlet model, particularly in the dataset (100K). This is the biggest proof that the proposed model can deal with uncertain ratings in sparse datasets. Unusually, the proposed model shows significant improvement for dense datasets (1M and 10M). This means that the proposed model would generate accurate ratings for dense datasets. This is another proof of the efficiency of the proposed model regarding the dense datasets. The results showed that the proposed model is more reliable than the rest of the models over both spare and dense datasets.

VI. CONCLUSIONS

This paper presents a new model of trust and reputation, which uses three factors fused together through the Arithmetic Mean to reflect the user's reputation of any online content as a value for the weight. In other words, this value can be used to calculate the reputation or the quality of any online content alike. The proposed model showed good accuracy in terms of the Mean Absolute Errors (MAE) and addressed several problems for the rest of the models. All this was through sparse and dense datasets. The main limitation in this paper is that it focuses primarily on the aggregation of assessments using the weighted average method, while there are many aggregating methods that can be examined to demonstrate the accuracy of the proposed model.

ACKNOWLEDGMENT

The authors are grateful to the Applied Science Private University, Amman, Jordan, for the financial support granted to cover the publication fee of this research article.

REFERENCES

- [1] A. Abdel-Hafez, X. Yue "An accurate rating aggregation method for generating item reputation." Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on. IEEE, 2015.
- [2] K. Bharadwaj, M. Y. H. Al-Shamri. "Fuzzy computational models for trust and reputation systems." Electronic commerce research and applications 8.1: 37-47, 2009.
- [3] H. F. Maxwell, and J. A. Konstan. "The movielens datasets: History and context." ACM Transactions on Interactive Intelligent Systems (TiiS) 5.4: 19, 2016.
- [4] A. Jøsang, I. Roslan, B. Colin. "A survey of trust and reputation systems for online service provision." Decision support systems 43.2: 618-644, 2007.
- [5] A. Josang, H. Jochen. "Dirichlet reputation systems." Availability, Reliability and Security. ARES. The Second International Conference on. IEEE, 2007.
- [6] W. Andrew, A. Jøsang, J. Indulska. "Filtering out unfair ratings in bayesian reputation systems." Proc. 7th Int. Workshop on Trust in Agent Societies. Vol. 6. 2004.
- [7] L. Hady W. Ee-Peng Lim, K. Wang. "Quality and leniency in online collaborative rating systems." ACM Transactions on the Web (TWEB) 6.1: 4, 2012.
- [8] L. Christopher, S. Soumya, M. Chiang. "On the volatility of online ratings: An empirical study." Workshop on E-Business. Springer Berlin Heidelberg, 2011.
- [9] W.L. Teacy, J. Patel, N. R. Jennings, M. Luck. "Travos: Trust and reputation in the context of inaccurate information sources." Autonomous Agents and Multi-Agent Systems 12.2: 183-198, 2006.
- [10] M. Zaki, A. Bouguettaya. "Rateweb: Reputation assessment for trust establishment among web services." The VLDB Journal—The International Journal on Very Large Data Bases 18.4: 885-911, 2009.
- [11] R. Tracy, and R. Wilensky. "An algorithm for automated rating of reviewers." Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. ACM, 2001.
- [12] F.L. Garcin, L.I. Xia, B. Faltings. "How aggregators influence human rater behavior." Proc. Workshop@ 14th ACM Conference on Electronic Commerce (EC-13). 2013.
- [13] M. Zaki, and A. Bouguettaya. "Rater credibility assessment in web services interactions." World Wide Web 12.1: 3-25, 2009.
- [14] F. Maxwell Harper, Joseph A. Konstan. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems 2015, 5, 4, Article 19.