Enhancing Mole Trust Algorithm Based Analysis User Profile

Saba M. Hussain 1, Ghaidaa A. Al-Sultany r 2
1, 2 Department of Information Network, College of Information Technology, University of Babylon, Babil, Iraq.
E-mail: saba.alshebeeb@uobabylon.edu.iq 1 , Ghaidaa.almulla@itnet.uobabylon.edu.iq 2

Abstract. Social networking is exploited for Internet-based marketers’ purposes in which, advertising companies are seeking to engage customers for their products. The research presented to improve the work of trust-based algorithms by analysing the user's behaviour of the check and this analysis was done by taking a set of features and parameters that can be granted knowledge of this behaviour if it is doubtful or normal. Parameters are (fairness, similarity and number of previews per products) generated for each user to get user weight score. User weight score is using to enhancing the weight of mole trust algorithm. Mole Trust algorithm used to build social trust network between users based on the rating given by the users on the products. whilst the proposed model used user weight score to enhancing the algorithm, therefore, it gives markable and effective result comparing with traditional models. The results were very satisfactory compared to previous work to classify users and build networks using traditional social network technologies. Furthermore, proposed model performance has provided better accuracy (72%) than traditional mole algorithm (35%).

Keywords: Social Trust Network, Mole algorithm, Number of Reviews per Product (NRP), Fairness and user weight score.

1. Introduction

Trust social networks are built on trust between users according to their shared products in terms of user-given rating. Social network algorithms rely on ratings to calculate the trust between users [1]. Most of the goods are promoted online, so there is a need for a network to support trust between users through shared comments on the products and the rating of the products given [2]. Most social algorithms depend on the rating given by the user for the product, which suggests the possibility that this rating is not right or fair. Therefore, it has been excluded from these algorithms that take into account other features of trust among users, such as transactions to deal with it [3]. Thus, to build a trusted network, one needs to distinguish between users. Although most existing detection approaches depend on the modelling of suspicious spammer behaviours and/or reviews analysis abstracted and derived from observations, some other metrics have not been used in an integrated manner with the reviewers and reviews data. In this paper suggests a trusted network between users built using the characteristics of social networking algorithms, and it determines the amount of trust between users by analysing the user properties such as reviews, rates, and several reviews. Here, enhancing the performance of social algorithms by using behaviour analysis of user instead of depending on one parameter.

Performance measurement is the act that measures how efficient, effective, and capable a procedure or system is against particular criteria or targets [4]. Accuracy and precision have always been a priority, and despite that these terms seem to be interchangeable in casual talks, they do carry different meanings.
Accuracy can be seen as how proximate measurement results are to the true value, meanwhile, precision is merely how repeatable or reproducible the measurement is [5].

The rest of this paper is organized as follow: the related work are presented in section 2, section 3 is explaining social network and mole trust algorithm, section 4: Illustrate the methodology that used in the proposed system, The experiments results and conclusion dissection in section 5, and respectively.

2. Related works

Some of the previous works in the same field have been reviewed within this section, such as:

Munmun and Nashreen, presented their work in an attempt to create better approaches to use the construction of the social graph as well as the trust connections shared. It deducts the extent to which two persons may trust each other even when there is no direct connection between them. The research presented an algorithm to interpret the amount of trust propagation shared by two individuals through an indirect connection using opinionated trust ratings along the path that appears to be the shortest and most trustworthy. The authors used DIJK star to define the SPA (Shortest Path Algorithm). This paper used four smaller datasets of just about 20 to 45 users in the analysing process, applying statistical techniques [6].

Hao Tian et. al., suggested an enhanced approach of recommending with a basis of trust connections within the social network to enhance their recommending performance. They particularly presented a definition of trust relationships, considering a number of representation factors when formalizing it. Furthermore, they presented an improved recommendation algorithm named IRATR (Improved Recommendations Based on Trust Relationships in Social Networks). The results of the experiment showed that the suggested method could enhance how accurate recommendations are, more than the methods that already exist. In addition to, it also indicated that the accuracy in predictions increased significantly [7].

Chuanmin Mi et.al, proposed a model of probability matrix factorization based on the integration of social trust as well as user interest. The model defined certain trust connections among users and possible labels of interest in light of user rating, and it has been put into use to perform matrix decomposition of user's rating and interest label data, as well as their trust connections. IT also identified the user's features to ensure that the data sparseness is carried out easily. The concept of the PMF model is mainly for predicting the user rating for a certain item in terms of probability [8].

M. Naderan et. al., proposed a work to classify trust between users, as the feature vector is calculated for each pair of social network users. Fuzzy logic is incorporated to rank the members of trust to a specific class, according to two-, three- and five-classes classification. Then, to classify the trust values of users, three machine learning techniques are used, instead of traditional weighted sum methods, to express the trust between any two users in the presence of a special pattern. These techniques are Support Vector Machine (SVM), Decision Tree (DT), and k-Nearest Neighbors (kNN) [9].

R. Logesh, et. al, suggested a POI (Point of Internet) recommendation method using trust improvement in social networks, so-called social pertinent trust walker (SPTW). First, a calculation of the trust level shared by users within the same social network is done, using a matrix factorization technique. Next, the SPTW with an algorithm of high probability location categorization is used in generating POIs in the form of recommendation lists. The evaluation of the suggested algorithm in terms of its accuracy is reached through experimenting with real-world databases. In conclusion, the results indicated that these recommendations are of better quality, and the suggested algorithm turns out to be relatively more effective than previous ones [10].

Faezeh Sadat et. al., proposed a novel trust-based method, known as Semantic-enhanced Trust based Ant Recommender System (STARS). This system performed a depth-first search to find the best trust paths within the trust networking and selected the most suitable neighbours for an active user for the sake of providing an enhanced recommendation. After performing experiments on real-world datasets, the results show that STARS outweighed other systems in light of accurate predictions,
recommendations of higher quality and the extent to which it overcame the problems mentioned earlier [11].

3. Social Network based Algorithms

One of the significant issues to which a large number of studies is devoted is trust within online social networks. Many algorithms have been widely developed by researchers to calculate the confidence in light of particular theories [12]. Several algorithms concern propagating trust between the source and the sink within a network to infer trust depending on the perspective of the user himself. Among the probably most commonly works referred to in this field are Tidal trust, WalkerTrust, and MoleTrust. A trait commonly shared by these three is the weighted average strategies used in computing a trust value for the sink [1].

There are many algorithms used to build social trust network such as Mole, tidal and walk random algorithms. The MoleTrust algorithm can calculate the trust that exists between the initial or first source and any other nodes covered by the maximum threshold span of two stages in advance [13]. Since there is no source specification for this algorithm, another standard has been applied for terminating shorter-path calculations, especially with regards to the maximum path length (δ). Besides, the threshold of user-defined confidence δ has been applied to all source pairs instead of calculating it for every source pair individually [1]. The higher efficiency of MoleTrust over TidalTrust is owed to the fact that it requires a single application for each source made, instead of each source-sink pair.

Having clarified how trust can be calculated between users within an online social network using a single algorithm of trust, as it utilizes the dataset part of the application [14]. Besides the principle algorithm execution that is the center of exploration, a direct trust algorithm depends on the activeness of companions towards the targeted user, as well as the TidalTrust and MoleTrust calculation [15]. This part of use gives additional usage of the techniques for adjusting loads indirect algorithms. This part of use includes different elements [16]:

- **Count:** This script presents user data such as a count number of likes, reviews, and tags. When gathered, this data is counted only once, and then stored in the form of a dataset to be used later on in computing direct algorithms.

- **Algorithms:** It is coordination that supplies functions for algorithms: direct, Mole and Tidal.

- **Rating:** It is the implementation of weight calibration for direct algorithms.

4. Methodology

This section clarifies how the proposed system designed and implemented to build a social trust network among users. It includes a description of the dataset that has been used in testing the system. As for the system itself, is has three stages, namely pre-processing, building main parameters and building a social network. Constructing relationships among users within social networks mainly depends on the trust value, which represents the amount of trust shared between users as well as how much of this is trust is determined by the rate value. The product trust network, for instance, is technically built on a rate for sharing products. Such a relationship is considered to have a higher level of comprehension whenever it is determined by several parameters rather than one. Hence, the suggested system makes use of a number of parameters in the calculation of the trust value shared among users. Figure (1) represents the main framework of the system for the build of social network trust. The novelty implied in this system is the use of MoleTrust algorithm, as it essentially depends on the ratings for purchased items. It eventually presents the relationship weight between users who are rating the same item. In this thesis, the analysis of user behaviour did not only include their ratings, but also the extraction of a collection of features for building trust relationships for each user, among which are Review Similarity, Number of Reviews per Product, User Fairness and Product Goodness. The data resulting from this process have been used as a weight for the MoleTrust algorithm to create a social trust network.
4.1. Preprocessing Stage

The Amazon dataset is originally available in JSON format. Since the data cannot be comprehended in this form, it is a prose piece of information, it is therefore converted into tabulation data such as CSV. All missing values and incomplete information from the user are then removed. The essential features of the data set used in this system are (User Name) and (product ID).

4.2. Generated Main Parameters

In this section, a set of features have been extracted to be applied. Therefore, more than one parameter is to be used in calculating the weight among users, also known as user weighted score. Parameters are:

- Number of Review per Product (NRP)

One of the ways used by spammers to degrade a product is to post a number of comments on that given item. Restricting this form of behaviour, the number of reviews per product (NRP) is calculated through the suggested system. This variable represents the ratio of the number of reviews left by a user on a certain item to the total number of reviews made on that product. A user posting several reviews repeatedly on a certain item is considered to be suspicious, as calculated in Equation (1).

\[ NRP_i = \frac{Hist_{i,j}}{N_j} \]  

(1)

Where, Hist i,j is a number of reviews from user i for product j and Nj is the number of all reviews on product j.

Equation (2) calculates the number of user's reviews.

\[ Hist_{i,j} = \sum_{i=1}^{n} \text{no. user review}_{i,j} \]  

(2)

The nj calculating by equation (3)

\[ n_j = \sum_{i=1}^{n} \text{no. review}_{j} \]  

(3)
4.2.1. User Fairness
Fairness is demonstrating how the system can be depending on a Fairness illustrates how the work of a system could be determined by a certain user. What distinguishes reasonable users is that they provide product ratings without an inclination. That is to say, higher scores are given to good quality products, and lower scores to lesser ones. Whenever users behave differently from what has been described above, this would soon be regarded as an indication of suspicious, "out of line" users. The rating that a user gives to a product reflects its quality according to the user's opinion. The rating values are span with the interval \[1-5\], where each product has different percentages of positive \((\geq 3)\) and negative \((<3)\) rates. Users who frequently give positive or negative rates are found to be in a circle of doubt, as it seems rather far from logical for all products to meet the user's satisfaction. Equation (4) presents a calculation of fairness of user.

\[
\text{Fairness}_i = \left[ \frac{\sum_{j=1}^{n} \hat{R}_j - \sum_{j=1}^{m} \bar{R}_j}{\sum_{j=1}^{n} \bar{R}_j} \right] \quad \ldots \ldots (4)
\]

Where, \(R=\text{total rating}\), \(\bar{R}=\text{number of rates } \geq 3\), \(\bar{R}=\text{number of rates } < 3\)

4.2.2. Reviews Similarity
Spammers often repeat similar sentences in their reviews, so, to manage this spam strategy, the proposed system identifies spammers based on the calculation of the similarity rate of their reviews using TF.IDF. Equation (5) calculates this measure similarity, as follows:

\[
\text{Sim}_i = TF_{ij} \times d \log \left( \frac{N}{df_i} \right), \text{ where } \ldots \ldots \quad (5)
\]

\[
tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad \text{Calculates the IDF: } idf(w) = \log \left( \frac{N}{df_i} \right)
\]

Where \(N\) documents. Define \(f_{ij}\) to be the frequency (number of occurrences) of the term (word) the \(i\) the \(n\) document\(j\). Then, define the term frequency \(TF_{ij}\) to suppose term \(i\) appears in \(n\) of the \(N\) documents in the set. Spammers frequently use the same phrases when reviewing different items. Thus, this sense of similarity could also be employed as one of the factors in detecting spammers.

4.2.3. User Weighted Score
The User Weighted Score creates a user ranking to distinguish between them. Some of the features, including NRP, Fairness and SIM, are used in the calculation of User Weighted Score, having a similar influence on the user weight score equation [17]. The calculation of the User Weighted Score is done by summing up the aforementioned characteristics. Equation (6) calculates the spam score:

\[
\text{User weighted Score}_i = \frac{(\text{Fairness}_i + \text{NRP}_i \times \text{Sim}_i)}{3} \quad \ldots \ldots (6)
\]

4.3. Social Network
For every user or customer, there is a trusted network built. Each user represents one node in-network, and the weighted connection shared by them is of use in the calculation of their trust. Below will be explained two trust networks by using Mole Trust algorithms before and after enhancing.

4.3.1. Mole trust algorithm
Mole Trust Network algorithm could be suitable for building social networks among users because of the items that are shared among them, as it technically derives trust among users without a direct connection. A disadvantage of applying this algorithm is the fact that it only uses the product rate in computing users' trust, which eventually does not provide the relation accurately and comprehensively. This leads to the use of traditional Mole algorithm in this work, with the modification of considering the spam score result to be a weight between users. Formula (7) extracts the trust between users:

\[
\text{trust}(u) = \frac{\sum_{i \in \text{predecessor}(u)} \text{trust}(i) \times \text{trust}_\text{edge}(i, u)}{\sum_{i \in \text{predecessor}(\text{trust}(i))}} \quad \ldots \ldots \quad (7)
\]
4.3.2 Update Mole Trust Algorithm
This section discusses the adjusting of weight used in Mole Trust algorithms, as it is replaced by the user weight score. The user weight score results into two categories considered in the identification of how authentic users are. Creating social network trust requires the extraction of new weights between users (user weight score) from their activity.

5. Experiments and Results

Initially, the original data set undergoes a converting process from JASON to a Comma-Separated Value (CSV) format, as the latter is considered to be clearer and easier to understand. Book dataset that is used in this paper has nine attributes used in describing user behavior, four factors have been elected as part of the suggested system, namely User Name, Product ID, Reviews, and Rating, from which the essential parameters will be extracted for supporting this research. The results of the extraction main parameters can be discussion here for each parameter.

Table 1 presents an explanation of the NRP sample results for the data set used. The first column shows the user name unique to each customer, while the second and third show the product name and value of NRP, which should be a real number between zero and one.

| SQ. | User name                               | ASIN         | NRP         |
|-----|-----------------------------------------|--------------|-------------|
| 1   | MR. KNOW IT ALL ☺DR SHOCK               | B000LVIIUC   | 0.5         |
| 2   | Capn Crusty                             | B002QS5OQ4   | 0.090909    |
| 3   | Frans Fanciful Folly                    | B000VU4GW2   | 0.076923    |
| 4   | Movie 2sday                             | B00BCWNXG8   | 0.2         |
| 5   | Amazon Customer Author                  | B004MSP04C   | 0.2         |
| 6   | Amazon Customer chemming                | B000OGTRC2   | 0.1         |
| 7   | Andrew                                  | B00CBNOBYU   | 0.058824    |
| 8   | Anne Nguyen                             | B00C5K6Z48   | 0.5         |
| 9   | April Vawter Wilderness Photographer    | B0012QRPU4   | 0.1         |
| 10  | Rosanne Dutzer                          | B001V58C4O   | 0.25        |

Figure (2) explain the NRP for two users the first normal behaviour user and second up normal behaviour user.

![Figure 2: NRP for each User](image-url)
The proposed approach is determined by the calculation of user rates on all products which will technically prove the user's authenticity. Table (2) below illustrates the results of the user authenticity, where the first column shows the unique user name, while the second and third show the product ID and value of user fairness respectively.

| SQ. | User name                                      | ASIN        | Fairness      |
|-----|-----------------------------------------------|-------------|---------------|
| 1   | "MR. KNOW IT ALL ☺ DR SHOCK                   | B000LVIIUC  | 0.692308      |
| 2   | Capn Crusty                                   | B002QS50Q4  | 0             |
| 3   | Frans Fanciful Folly                          | B000VU4GW2  | 1             |
| 4   | Movie 2sday                                   | B00BCWNXG8  | 0.333333      |
| 5   | Amazon Customer Author                        | B004MSP04C  | 0.285714      |
| 6   | Amazon Customer chemming                      | B000OGTRC2  | 1             |
| 7   | Andrew                                        | B00CBNOBYU  | 0             |
| 8   | Anne Nguyen                                   | B003N1KND2  | 0.677778      |
| 9   | April Vawter Wilderness Photographer           | B0012QRPU4  | 1             |
| 10  | Rosanne Dutzer                                 | B001V58C4O  | 1             |

The result of the second parameter of fairness, demands all users to rate the products over ten times. The ratio that is closest to 0 is considered to be better, as it indicates a good value of user authenticity. On the other hand, values closer to 1 are worse because they indicate the user's untrustworthiness. Figure (3) illustrate fairness for users.

![Figure 3: The Fairness of user](image)

Table (3) explains the review similarity, where the first column shows the unique user name, while the second shows the value of review similarity, which is a real number between zero and one.

| SQ. | User name                                      | ASIN        | Similarity     |
|-----|-----------------------------------------------|-------------|---------------|
| 1   | "MR. KNOW IT ALL ☺ DR SHOCK                   | B000LVIIUC  | 0.086417      |
| 2   | Capn Crusty                                   | B002QS50Q4  | 0.026438      |
| 3   | Frans Fanciful Folly                          | B000VU4GW2  | 0.132889      |
| 4   | Movie 2sday                                   | B00BCWNXG8  | 0.023648      |
| 5   | Amazon Customer Author                        | B004MSP04C  | 0.074673      |
| 6   | Amazon Customer chemming                      | B000OGTRC2  | 0.033145      |
The value of review content similarity ranges between (0 and 1), representing the likeness that exists between reviews placed by a customer. Figure (4) explains the distribution of user according to the similarity ratio.

![Figure 4: Density of user content similarity](image)

From all the above parameters get user weight score as a result to use it as a weight to build the social network trust between users. The suggested system is easily affected by the activity of the user with regards to the item, in light of the way users are classified, as shown in Figure (5).

![Figure 5: User Weight Score Density](image)

Table (4) explain the collection result of spam score for users depending on all above parameters for identifying users.

| SQ. | User Name                     | Product ID     | user weight score |
|-----|-------------------------------|----------------|-------------------|
| 1   | ! MR. KNOW IT ALL ☺DR SHOCK   | B000LVIIUC     | 0.426241477       |
| 2   | Capn Crusty                   | B002QS5OQ4     | 0.039115533       |
| 3   | Frans Fanciful Folly          | B000VU4GW2     | 0.403270841       |
| 4   | Movie 2sday                   | B00BCWGXG8     | 0.185660333       |

Table 4: The result of all the features
The initial stage of creating a social network of online customers depends on products that are shared among them. Figure (6) presents a sample of thirty-seven users of a social network in the Instant Video dataset.

Next, the network is divided into a subnet of users who share similar behaviour, on which the strength of the trust bond among users in one subnet depends. The system determines the user's authenticity based on analyzing their profile and behaviour in the data centre of a host online commercial webpage. Users can have normal activity on a certain product and irregular behaviour with another. Mole Trust algorithm creates networks between users based on the ratings given. These rates illustrate the trust weight among users who share the same item. Among the thirty-seven users, the mole figures were obtained from two subnets. The networks generated are: “{[3,10], [27,36]}”. Figure (7) illustrates the Mole Trust algorithm.

With the use of user weight score as trust weight among users who share products, five groups of subnets are determined from the same sample used in the Mole. The trust subnets generated are “{[1,14,22,23], [7,24,32], [0,2,11,12,18,21,27,29,30,31], [13,19], [16,17]}”. Figure (8) clarifies the sub-networks that have been obtained from the Updated Mole trust algorithm.
Mole algorithm aims at having a network with maximum size and reliability between users. The application of both algorithms noted resulted in the following: The Mole Trust algorithm found an average rate of 35% in Book dataset. The Updated Mole Trust algorithm, on the other hand, resulted in 72% for the same data sets.

5. Conclusions

The most important characteristics that have been obtained from the results of the proposed system are explained in this section. Most of the previous research that concerned the classification of the user relied on one parameter in the classification for either the rating or reviews given by the user on the products. In our proposed system, we worked on generating than one parameter and the results were more efficient than others in describing users. Rating only doesn’t give a sufficient impression of the person as a weight between users. Therefore, in this work applied traditional mole algorithm and will be modifying it by using the result of spam score as a weight between users. In this work, we used several algorithms in this research such as tidal trust algorithm and walk random trust network, each algorithm has advantages and disadvantages. The result was complexity in time and analysis because the two ways had a special rule to do. The results of enhancing Mole trust algorithm that more efficient than traditional algorithms. we suggest for future works using the sentiment analysis to develop this build social trust networks.

References
[1] Charu C. Aggarwal, 2016, Recommender Systems: The Textbook, Book, ISBN 978-3-319-29659-3 (eBook).
[2] Dematis Ioannis, 2017, Fake Review Detection via exploitation of Spam Indicators and Author Behavior Characteristics.
[3] Hao Tian , and Peifeng Liang, 2017, Improved Recommendations Based on Trust Relationships in Social Networks, Future Internet , 9; doi:10.3390/fi9010009.
[4] Deborah Nightingale, 2005, Metrics and Performance Measurement System for the Lean Enterprise, ESD.61J / 16.852J: Integrating the Lean Enterprise.

[5] Paolo Avesani, Paolo Massa and Roberto Tiella, 2005, a Trust-aware Recommender System for Ski Mountaineering. ISBN:1-58113-964-0 C ACM.

[6] Munnun Bhattacharya, Nashreen Nesa, 2016, An Algorithm for Predicting Local Trust based on Trust Propagation in Online Social Networks, Volume 156 – No 7.

[7] Hao Tian, and Peifeng Liang, 2017, Improved Recommendations Based on Trust Relationships in Social Networks, Future Internet, 9; doi:10.3390/fi9010009.

[8] Chuanmin Mi, Peng Peng, and Rafał Mierzwiak, 2018, A Recommendation Algorithm Considering User Trust and Interest, Springer International Publishing AG, part of Springer Nature, CAISC 2018, LNAI 10842, pp. 408–422, 2018.

[9] M. Naderan(C.A.), E. Namjoo and S. Mohammadi, 2019, Trust Classification in Social Networks Using Combined Machine Learning Algorithms and Fuzzy Logic, Iranian Journal of Electrical and Electronic Engineering 03 - 294–309.

[10] R. Logesh1 • V. Subramaniyaswamy, 2017, A Reliable Point of Interest Recommendation based on Trust Relevancy between Users, _Springer Science+Business Media, 97:2751–2780.

[11] Faeezeh Sadat Gohari, Hassan Haghighi, Ferëidoon Shams Aliie, 2017, A semantic-enhanced trust based recommender system using ant colony optimization, Springer Science+Business Media New York, 46:328–364 DOI 10.1007/s10489-016-0830-y.

[12] Mirjam Štim, 2014, ANALYSIS OF LGORITHMS FOR DETERMINING TRUST AMONG FRIENDS ON SOCIAL NETWORKS, FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING, Vienna.

[13] RICI, F., ROKACH, L., SHAPIRA, B., KANTOR, P.B. Trust and Recommendations. In Recommender Systems Handbook (pp. 645-675), London, (2011).

[14] Sherchan, W., Nepal, S., and Paris, 2013, A Survey of Trust in Social Networks, ACM Comput. Surv. 45, 4, Article 47 (August 2013), 33 pages.

[15] Vedran Podobnik, Darko Striga, Ana Jandras, Ignac Lovrek, 2012, How to Calculate Trust between Social Network Users? Proceedings of the 20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2012).

[16] Saba M. Hussain, Ghaidaa A. Al-Sultany, 2019, Using Costumer Behavior to Build Social Network Based on Mole-Algorithm, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November.

[17] Ghaidaa A. Al-Sultany and Saba Mohammed Hussain, 2019,” Fake reviews detection through user’s behavior analysis”, Jour of Adv Research in Dynamical & Control Systems, Vol. 11, 10-Special Issue, 2019.