Field-aware User Influence Recommendation Model Based on Trust Relationship

Bai Yun*, Cai Wandong
School of Computer, Northwestern Polytechnical University, Xi'an, ShaanXi, 710072, China
*Corresponding author’s e-mail: 395943619@qq.com

Abstract. In current determining of user influence based on network structure, the overall importance of user node in network usually attracts great attention. However, the significance of the users in specific field has not been fully studied, which result in low accuracy and reliability in user influence measurement. In order to resolve these problems, this research proposes a field-aware user influence model, which constructs user influence for specific fields, and analyses the global influence of users in the whole network on a network structure basis. The model, assisted by historical behaviour data, not only considers the spread of user influence, but also takes the relationship between user influence and fields into account. Referring to the global influence, the model is able to optimize the composition of user influence, and further improve the accuracy of recommendation. The experiments based on real data sets conducted in this research un exceptionally proved that the field-aware user influence model proposed in this paper, compared with other methods, can effectively improve the recommendation accuracy.

1. Introduction
Social influence is widespread in social networks. And research on it has been widely used in recommendation systems, such as advertising, innovation, public opinion monitoring and guidance. Many researchers have confirmed that user influence is directly related to recommendation accuracy. According to life experience, people are more inclined to listen to expert opinions when making decisions, that is, opinions of people with certain influence in a certain field. This shows that user influence is related to the field. The influencing factors of the field must be considered in user influence research. In addition, the research on user influence is inseparable from network structure[1]. The structure of social network is based on the social relationship of users. At present, the study on the structure of social network composed by trust relationship is the most common. And the recommendation algorithm based on trust relationship mostly assumes that the trust relationship is generalized and there is no field distinction. In fact, trust is inherently different and multifaceted[2-3]. User A trusts User B can only indicate that User A may agree with User B’s point of view in one or some fields, and does not mean that User A trusts B in any aspect (i.e. any field). Therefore, the network structure formed by trust relationships in different fields is different. When studying user influence, it is necessary to consider the greatness of influence in a specific field and the composition of the network structure in a specific field. Therefore, from the perspective of field, this research, by combining user historical behaviour data with trust relationship and field attribute, measures user influence in specific fields, and finally proposes a field-aware user influence recommendation model.
2. Literature review
User influence in a social network can be understood as the capability to drive other users to agree with a point of view or to make an action[4]. More and more researches currently are focusing on the application of user influence to recommendation systems[5]. The network topology analysis is the most direct method in measuring user influence. It measures user influence from following three aspects: the importance of node location in social network structure, network structure similarity, and network reach capability. The user influence measurement based on the importance of nodes mainly considers the out-degree, in-degree or node degree centrality[6], the between centrality[7], the tightness centrality[8] as well as the eigenvector centrality[9] to show the magnitude of the influence of the node. Wu et al.[10] measured the average influence of a node on its neighbours by the degree centrality of nodes. Newman[11] mainly studied the role of indirect influence in calculating user influence. His study used tightness centrality to measure the indirect influence of nodes on others. Sporns[12] investigated the influence of users in information dissemination, the measurement method of which is median centrality. Chen et al.[13] constructed a semi-local centrality method based on node degree centrality and between centrality to measure node influence. Some scholars measure user influence by studying the similarity of network structure. For instance, SimRank algorithm[14] measured structural similarity according to the relationship between the node and other nodes, and analysed social influence from the perspective of similarity. Tang et al.[15], based on the similarity analysis of network structure, proposed a local affinity propagation algorithm to measure the social influence according to the similarity of information content. Besides, there are some studies that measure user influence on the basis of the accessibility of the network structure. The basic principle is that the influence of the node is closely related to the influence of the nodes around it. In such researches, typical methods such as eigenvector centrality[16], Katz centrality[17], and Page Rank[18] metrics were applied. The influence measurement simply with traditional network structure analysis methods neglected the influence of various factors such as relationship strength between users, field relevance of social relations, and information content characteristics on node behaviour decision-making in social networks. Therefore, some scholars try to use the user's historical interaction behaviour record to make up for the deficiencies of the network structure analysis method, and measure the social influence of the user by modelling the frequency of interaction between users or activities. Romero et al.[19] proposed the Influence-Passivity algorithm based on the PageRank algorithm by referring to the forwarding rate between users. Tang[20] et al. studied the topic-based influence issue among users by defining a topic factor graph TFG model. Hwak et al.[21] calculated PageRank values through the network topology of social user followers, which was considered as the user influence. However, the method neglected the relevance between trust and fields.

The above researches all failed to measure user influence in a specific field, so there is a phenomenon that the global influence is large but specific field influence is small or the influence is large in one field but is small in another field. When the user influence is applied to the recommendation system, the recommended effect will be reduced. Currently, the recommendation algorithm based on trust relationship have attracted most attention from scholars. Since trust relationship is related to fields, in different fields, the user sets that trusted are accordingly different. And thus the network topology formed in different fields is also varied. In this sense, filed influence should be considered in measuring user influence through the network topology. This research investigates the network structure differences in different fields. In view of the network structure in specific fields, it studies the field influence of users in specific field, and then combines the global influence to propose a user influence recommendation model based on trust relationship.

3. Field-aware user influence model
When making product recommendations, users are more likely to accept opinions from users with higher influence in the field to which the product belongs. This paper, considering the role of user influence in specific field, proposes a recommendation model of field-aware user influence, and divides user influence into filed influence and global influence.
3.1. Field influence
Field influence reveals the importance of users in social networks in specific field. Tang et al.[2-3] found that users have different preferences in different fields, and they may also have different activities in different fields. These characteristics lead to differentiated importance of users in different fields. In addition, users’ field transmission capability is also an indicator to measure the importance of users in the field. The strong transmission capability can exert greater influence on more users. Therefore, users’ activity characteristics and social relationships in a certain field is able to reflect their influence greatness in the field.

Figure 1 is a partial social network diagram of the user $u_i$. Users’ field influence will be explained in the following sections on this basis.

![Figure 1. Partial social network diagram of user $u_i$.](image)

As can be seen from Figure 1, the influence of the target user $u_i$ in field $f$ is measured from the following two aspects: one is the activity characteristics of the target user $u_i$ in the field $f$ and its potential influence (users whose activities in the field may be directly affected), which is called inherent influence in field $f$; the second is to consider other users’ transmission on social network of the target user $u_i$, which is called the transmission of user field influence.

### 3.1.1. Users’ inherent field influence
According to the definition of users’ inherent field influence and the above-mentioned the research theory of Tang et al.[2-3], the user's activity characteristics in field $f$ are mainly affected by the user's activity in field $f$ and the preference for field $f$.

The user prefers to post comments in a certain field, indicating that the user has a strong interest in the field and has certain opinions, and then the user has a greater influence on the preferred field than other fields. Therefore, the user's field preference can be used as a factor of measuring user's field influence. The preference impact factor of user $u_i$ in field $f$ can be calculated with the following formula:

$$ prf(u_i, f) = \frac{N_i^f}{N_i} $$

(1)

In this formula, $N_i^f$ refers to the rating number of user $u_i$ in field $f$ and $N_i$ is the total rating number of user $u_i$.

The more frequently that the user activates in a certain field, the more comments the user may post and the greater influence will be exerted on the field. Therefore, the user's field preference can be used
to measure the user’s field influence. The calculation formula of field active impact factor of user $u_i$ in field $f$ is demonstrated as follows:

$$act(u_i, f) = \frac{N_f^i}{N_f}$$  \hspace{1cm} (2)

In the formula, $N_f^i$ is the total rating number in field $f$.

The activities of user $u_i$ have potential influence on social network in field $f$, the affected object is the link-in users of user $u_i$. The link-in users trust user $u_i$, so user $u_i$ will exert influence on these link-in users. The larger the number of these link-in users, the greater the influence will be. Therefore, the user’s inherent field influence can be determined by:

$$inherent\_inf\_ulence(u_i, f) = indg(u_i) \times prf(u_i, f) \times act(u_i, f)$$  \hspace{1cm} (3)

In this formula, $indg(u_i)$ is the in-degree of user $u_i$.

3.1.2. Transmission capability of user field influence. Information is transitive, so the capability of users to transmit information in social networks determines the depth of the user influence. Most studies have found that the information provided by nodes with a depth of less than 2 hops is most valuable to users[22]. And hence, this paper mainly considers the contribution of nodes within 2 hops to field transmission when measuring the user field influence. In this sense, the field influence transmission capability of user $u_j$ in the figure is related to the contribution of directly link-in user $u_j$ and indirectly link-in user $v_k$. The larger the contribution of user $u_j$ and user $v_k$, the more profound field influence that user $u_i$ may have. There is the same principle in the contribution of user $u_j$ and $v_k$ in field $f$. Take user $u_j$ as an example:

In Figure 1, user $u_j$ trusts user $u_i$, the larger proportion that the common rating items of the two users take in the total rating items of user $u_j$, the greater that the influence of user $u_i$ may have on user $u_j$. In addition, the greater the intrinsic influence of user $u_j$ in field $f$, the greater its potential influence will be. Therefore, the field transmission contribution of user $u_j$ in field $f$ to user $u_i$ is determined by rating data and trust relationship in the field. The formula can be demonstrated as:

$$Contribution(u_j, u_i, f) = \frac{com^j_i}{N_j^f} \times inherent\_inf\_ulence(u_j, f)$$  \hspace{1cm} (4)

In this equation, $com^j_i$ is the commonly rated items of user $u_j$ and $u_i$ in field $f$, while $N_j^f$ is the number of the items that user $u_j$ reviewed in field $f$.

Similarly, in field $f$, the field transmission contribution of user $v_k$ to user $u_j$ is:

$$Contribution(v_k, u_j, f) = \frac{com^k}{N_k^f} \times inherent\_inf\_ulence(v_k, f)$$  \hspace{1cm} (5)

The field influence transmission capability of user $u_i$ is:

propagation($u_i, f$) = $contribution(u_j, u_i, f) + Contribution(v_k, u_j, f)$  \hspace{1cm} (6)

Hence, the influence capability of user $u_i$ is:

field\_inf\_ulence($u_i, f$) = inherent\_inf\_ulence($u_i, f$) + propagation($u_i, f$)  \hspace{1cm} (7)
3.2. global influence

Global influence refers to the influence of users in the entire social network. Among all the sorting algorithms, the PageRank algorithm is the most well-known and widely used one. The algorithm evaluates the importance of nodes in the whole network, but the average distribution method used in the iteration process would assign the PageRank value of the leading nodes to the post-positioned node set. However, this average allocation method ignores the importance of the post-positioned nodes, hence it is not applicable to evaluating the influence of social network nodes. In addition, life experience tells people that the users accept trust would exert greater influence on the target user. Based on the above two points, this paper proposes the TrustRank algorithm that evaluates the influence of users from the perspective of trust relationship among users, and distributes the influence of leading nodes according to the influence weight.

The following concepts are involved in TrustRank algorithm:

- **Intuitive trust**: it refers to the trust degree that is directly observed from the number of users trusted and the number of highly-recognized comments posted by users. If the number of trusted users is larger and the number of the comments with high-recognition that the user posted is larger, the user's intuitive trust rank should be higher.

- **Initial influence**: this concept refers to the initial influence obtained without considering trust link relationship but only through the user's intuitive trust. Its formula can be expressed as

  \[ I = T \times \log_{10} C. \]

  In the formula, \( C \) is the number of highly approved comments posted by the user, \( T \) is the number of users who trust the user. Since in the sample, some users have posted a large number of reviews, which are much more than those of common users, logarithms are adopted to narrow the difference. This method, therefore, will increase the weight of the influence of trust.

- **Influence weight**: The degree of trust between users in a trust network is different. The more greatly the user trusts another user, the more greatly he will be influenced. If user \( u_j \) trusts user \( u_i \), then \( u_j \) is the pre-node while \( u_i \) is the post-node. For the pre-node, the greater the influence of the post-node, the more susceptible it is to the post-node, so the greater the distributed influence. Therefore, the formula for calculating the influence weight is

  \[ w_{ji} = \frac{I(u_j)}{\sum_{v \in Y(u_i)} I(v)}, \]

  in which \( w_{ji} \) is the influence weight in the process that user \( u_j \) trusts user \( u_i \). \( I(u_j) \) is the initial influence of user \( u_j \), and \( Y(u_j) \) is the user set that user \( u_j \) trusts, \( u_i \in Y(u_j) \).

The TrustRank algorithm evaluates the global influence, and its formula is:

\[ ginfluence(u_i) = \frac{1-p}{n} + p \sum_{u_j \in X(u_i)} ginfluence(u_j) \times w_{ji}. \tag{8} \]

In this process, \( p \) is the damping coefficient, indicating the random probability that the user affects another user. Its value range is \([0, 1]\), and its usually accepted value is 0.85. \( n \) is the total number of users, \( X(u_i) \) is the user set pointing to user \( u_i \), that is, the set of users trusting user \( u_i \).

In a trust network, there is a case where the node has a degree of 0, that is, a user does not trust any user, and such users are generally referred to as hanging nodes. A hanging node does not trust any user, indicating that the hanging node has no trust tendency to other users in the trust network, then the probability that the hanging node may trust a certain node is the same. The probability that the hanging node is affected by other nodes is the same. This is called influence probability. The formula for calculating the probability of influence is:

\[ \eta = \frac{I(u_0)}{n!}. \tag{9} \]
I(u₀) is the initial influence of the node with an out-degree of 0, and n' is the total number of nodes in the trust network.

The way to handle the situation when encountering a hanging node in the TrustRank iteration is to jump to any node on the trust network with an influence probability of \( \eta \).

### 3.3. Recommendation model of user influence

When recommending an item for a certain user, the authority of users trusted by the user in the field will exert influence on the user. The more authoritative they are, the more susceptible that the target user will be. At the same time, the higher the overall authority of the users trusted by the target user, the greater influence they will exert on the user. Therefore, the greater the influence on the target user, the easier the target user will be recommended. And the better the recommendation would become.

The influence of user \( u_j \) in field \( f \) is mainly determined by the influence of the user set of user \( u_i \) in the same field, and the global influence. It can be calculated by:

\[
user\_inf\_luence(u_i) = \sum_{u_j \in Trust(u_i)} \left( w \times field\_inf\_luence_{norm}(u_j, f) + (1 - w) \times ginfluence_{norm}(u_j) \right)
\]

In this formula, \( Trust(u_i) \) is the \( n \)th layers of user sets of user \( u_i \), \( w_1 \) and \( w_2 \) are the weighting coefficients of the user’s field influence and global influence, and \( w_1 + w_2 = 1 \). \( field\_inf\_luence_{norm}(u_j, f) \) and \( ginfluence_{norm}(u_j) \) respectively refers to the field influence and global influence after normalization, the calculation formula is as follows:

\[
field\_inf\_luence_{norm}(u_j, f) = \frac{\sum_{u_j \in f} field\_inf\_luence(u_j, f)}{\sum_{u_j \in u} field\_inf\_luence(u_j, f)}
\]

\[
ginfluence_{norm}(u_j) = \frac{\sum_{u_j \in u} ginfluence(u_j)}{\sum_{u_j \in u} ginfluence(u_j)}
\]

### 4. Experimental analysis

#### 4.1. Experimental data set

To measure the effect of the algorithm, this research adopts the Epinions data set[23] for the experiments. The data set contains 22,166 users, 296,277 items, 355,813 trust relationships, 922,267 user ratings for items, 922,267 user feedback ratings for the quality of ratings, and classification of 27 classifications of the items. The scoring mechanism of 1~5 points is adopted to represent the degree of preference degree in ascending order.

#### 4.2. Measurable indicators

Five cross-validations have been conducted in this research. In each cross-validation, 80% of the data is used as the training set, with 20% being the test set. The most commonly used recommended accuracy index in the recommended system literature are used as the standard to measure the pros and cons of the algorithm: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). They are defined as follows:

\[
RMSE = \sqrt{\frac{\sum_{(u_i) \in test} (r_{ui} - \bar{r}_{ui})^2}{|test|}}
\]
In these two equations, $r_{ui}$ refers to the actual rating of user $u$ on item $i$. And $\hat{r}_{ui}$ is the rating of user $u$ on item $i$ predicted by the system. $r_{test}$ is a test sample. The value of MAE and RMSE are both as small as better, among which RMSE increases the penalty for inaccurate scoring prediction and it is relatively stricter in evaluating recommended algorithms.

4.3. Recommended accuracy analysis

To analyse the role of user influence on recommendation, this paper proposes a field-aware user influence model (Hereinafter referred to as FieldUI) and comparatively studies it with project-based collaborative filtering algorithm (i.e. Item-based CF), trust-dependent random walk algorithm (i.e. Trust Walk) and typical LeaderRank on Epinions data base[13]. Figure 2 shows a comparison of the MAE and RSME values of all the four algorithms.

![Figure 2. Comparison of evaluation indexes of different algorithms.](image)

The two indicators MAE and RMSE in the figure reveal that the Item-based CF algorithm has poorer recommendation effect due to its heavy dependence on data. TrustWalk, however, has been improved in performance after it introduces trust relationship and thus effectively alleviates the problem of data sparsity. LeaderRank measures the importance of nodes in the network from the perspective of network structure. The node importance indexes are added during recommendation, which further enhances the recommendation effect. FieldUI proposed in this paper not only considers the importance of the node from an overall perspective, that is, the global influence, but also considers the field influence of the node in a certain field. It also corrects the calculation method of the user influence, and further improves the accuracy of recommendation in a single field. Ultimately, the overall recommendation is improved.

4.4. Choosing parameter $w_1$ and $w_2$

Parameter $w_1$ and $w_2$, as two weighting factors of field influence and global influence of users, are used to measure the impact of field influence and global influence on recommendation. Figure 3 shows the impact of the user field influence weight $w_1$ on RMSE and MAE. Since it needs to satisfy the equation $w_1 + w_2 = 1$, only the influence of the parameter $w_1$ change on the evaluation index is examined. Nine proportional allocations is selected in the [0, 1] interval. It can be seen from the figure that when considering only the single influence of the field influence or global influence on the recommendation effect (i.e. $w_1=1$, $w_2=0$ or $w_1=0$, $w_2=1$), the model only considering that field influence is obviously superior to the model simply considering global influence. This indicates that when an item is recommended, the expert opinion in the field to which the item belongs is more
informative. This conclusion is consistent with common sense of life. When $w_1=0.7$, $w_2=0.3$ and $w_1=0.6$, $w_2=0.4$, the recommendation precision MAEs are equivalent in effect. However, from the more stringent RMSE index of the recommendation algorithm evaluation, when $w_1=0.7$, $w_2=0.3$, the recommendation is the most effective. Therefore, when measuring user influence, users should take both field influence and global influence into consideration. To improve the recommendation effect, the user influence in a single field accounts for a large proportion, and meanwhile, the influence of the user in the overall situation cannot be ignored. This is also consistent with the situation in daily life. People are more willing to listen to the opinions of authoritative people in a certain field, but may also think about the overall cognitive level of the authority.

5. Conclusion
The research proposed a field-aware user influence model to resolve above mentioned problems. Firstly, based on the trust network in the specific field and the historical behaviour data of users, it fully took the influence transmission factors into account and constructed the user field influence model. And then, from the perspective of the overall importance of users in the network, TrustRank algorithm was suggested to be adopted to measure the global influence. The model could optimize user influence from both field influence and global influence. Experiments on the public dataset[2-3] show that the field-aware user influence model proposed in this paper is superior in accuracy to the comparative collaborative filtering recommendation algorithm, TrustWalk algorithm and LeaderRank algorithm. The experimental results unexceptionally revealed that measuring user influence in specific field has a significant positive effect on improving the recommendation effect. And meanwhile, it should be compatible with considering the global influence of users in the overall network. Therefore, in the trust network, the user influence is analysed in field awareness, and the comprehensive evaluation of the user influence from the specific field and the whole perspective plays an important role in improving the recommendation effect of social recommendation algorithm.

Acknowledgments
This paper supported by Science and Technology Planning Project of Yulin City (2016-24-4), Special Scientific Research Project of Shaanxi Provincial Education Department. Shaanxi Provincial Department of Education special scientific research plan (19JK0526).

References
[1] Sun, J., & Tang, J. (2011). A survey of models and algorithms for social influence analysis. In Social network data analytics. Springer, Boston, MA. pp. 177-214.
[2] Tang, J., Gao, H., & Liu, H. (2012). mTrust: discerning multi-faceted trust in a connected world. In: the fifth ACM international conference on web search and data mining. pp.93-102

[3] Tang, J., Gao, H., Liu, H., & Sarma, A. D. (2012). eTrust: understanding trust evolution in an online world. In: the 18th ACM SIGKDD international conference on knowledge discovery and data mining. pp. 253-261.

[4] Aggarwal, C. C. (2011). An introduction to social network data analytics. In Social network data analytics. Springer, Boston, MA. pp. 1-15.

[5] Wang, Q., Liu, X., Zhang, S., Jiang, Y., Du, F., Yue, Y., & Liang, Y. (2015). A Novel APP Recommendation Method Based on SVD and Social Influence. In: international conference on algorithms and architectures for parallel processing pp. 269-281

[6] Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology, 2:113-120.

[7] Newman, M. E. (2005). A measure of betweenness centrality based on random walks. Social Networks, 27: 39-54.

[8] Sabidussi, G. (1966). The centrality index of a graph. Psychometrika, 31:581-603.

[9] Newman, M. E. (2016). Mathematics of networks. The new Palgrave dictionary of economics, 1-8.

[10] Wu, X. D., Li, Y., & Li, L. (2014). Influence analysis of online social networks. Chinese Journal of Computers, 37:355-752.

[11] Newman, M. E. (2005). A measure of betweenness centrality based on random walks. Social Networks, 27: 39-54.

[12] Sporns, O. (2013). Structure and function of complex brain networks. Dialogues in clinical neuroscience, 15: 247-262.

[13] Chen, D., Lu, L., Shang, M., Zhang, Y., & Zhou, T. (2012). Identifying influential nodes in complex networks. Physica A-statistical Mechanics and Its Applications, 391:1777-1787.

[14] Jeh, G., & Widom, J., (2002). SimRank: a measure of structural-context similarity. In: the eighth ACM SIGKDD international conference on knowledge discovery and data mining. pp.538-543

[15] Tang, J., Sun, J., Wang, C., & Yang, Z., (2009). Social influence analysis in large-scale networks. In: the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 807-816

[16] Bonacich, P. (2007). Some unique properties of eigenvector centrality. Social Networks, 29:555-564.

[17] Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18:39-43.

[18] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

[19] Romero, D. M., Galuba, W., Asur, S., & Huberman, B. A., (2011). Influence and passivity in social media. In: european conference on machine learning. pp.18-33.

[20] Tang, J., Sun, J., Wang, C., & Yang, Z., (2009). Social influence analysis in large-scale networks. In: the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 807-816

[21] Kwak, H., Lee, C., Park, H., & Moon, S. B., (2010). What is Twitter, a social network or a news media?. In: the web conference pp. 591-600.

[22] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P., (2010). Measuring user influence in Twitter : the million follower fallacy. In: international conference on weblogs and social media. pp. 10-17.

[23] Zhang, B., Huang, Z., Yu, J., & Xiang, Y. (2014). Trust computation for multiple routes recommendation in social network sites. Security and Communication Networks, 7(12), 2258-2276.