Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/44/20230247

Enhancing recommendation with causal embedding: Considering social network influence

Zhuyiheng Chu[1,6,11], Pengxiang Zhang[2,7], Yibo Zhou[3,8], Shihan Huang[4,9], Yanze Guo[5,10]

[1]School of Mathematical Sciences, Zhejiang University, Hangzhou, 310058, China
[2]School of Mathematical Sciences, Xiamen University, Xiamen, 361005, China
[3]School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou, 310018, China
[4]Shijiazhuang NO.24 high school, Shijiazhuang, 050051, China
[5]School of Physical Science, University of California, Irvine, Irvine, California, 92612, United States

[6]3200104144@zju.edu.cn
[7]pxzhang2@gmail.com
[8]BruceZhou0201@163.com
[9]shihanhuang911@gmail.com
[10]Tedguo0511@gmail.com
[11]corresponding author

Abstract. In the realm of recommendation models, we consistently rely on observational interaction data. This data encompasses a variety of aspects, such as user conformity and genuine user interests. The key challenge for recommender systems is to extract a user's authentic interests from this interaction data in order to provide accurate recommendations. The current method, DICE, attempts to separate conformity and interest by assigning distinct embeddings for each to users and items. The method ensures that each embedding captures only one causal factor through training with specific causal data. In our research, we've enhanced this existing method by incorporating social networks into the disentanglement of conformity and real interest from observational interaction data. The results from our proposed method surpass those of the prevailing baseline, demonstrating significant improvements across various backbone models using a real dataset. Furthermore, we conducted a sensitivity analysis and provided recommendations for scenarios in which our new model would be most effective.

Keywords: Recommender systems, conformity bias, social networks, causal embedding, DICE.

1. Introduction

The recommendation system has developed very rapidly in recent years. More and more methods are proposed to disentangle users' real interests from other causes. DICE [1] is one of the state-of-the-art frameworks that consider to disentangle users' conformity and interest from observational interaction data. However, DICE does not consider social networks while disentangling conformity and interest, which will greatly affect the conformity of users. A user's clicking behavior on an item may be

DOI: 10.54254/2755-2721/44/20230247

influenced by their friends' previous clicks. In fact, these clicks are more likely to occur than clicks generated by other conformity behaviors because users will subconsciously trust his friends and have an urge to click. For example, when choosing a pencil case, a student purchases a same pencil case when he sees his friend has bought a certain pencil case. With the existing method, it will be difficult to distinguish whether the user buys the pencil case out of love or because his friends have bought the pencil case. Therefore, it is crucial to disentangle users' pure interests from conformity in order to enhance the quality of recommendations.

In this work, we use the Epinions [2], a dataset that containing information about social networks to calculate the effect of social networks on user conformity. There are three tables in the Epinions. The user_rating table contains social network information between users, i.e. users they trust or do not trust. The rating table contains the user's rating of different items. Based on above two tables, we propose the personal popularity which includes information of social networks, and replace the popularity previously adopted by DICE with personal popularity. And we propose to use the Epinions to generate datasets for subsequent training and testing. Thus, we can judge the impact of social networks on users' conformity. Specifically, we face four key challenges. First, a method to calculate personal popularity that includes users' social network information. Second, a way to process and generate datasets containing information about social networks needs to be worked out. Third, the Epinions is too large to be fully trained, so rules need to be determined for cleaning useless data in the dataset. Fourth, we need to modify the DICE framework so that it can train datasets containing social network information.

We evaluated the performance of the DICE trained on the dataset with social network information added. We also controlled the proportion of social networks used in the dataset and observed the influence of different proportions of social networks on the training effect of DICE. Experimental results show that DICE trained on the new dataset outperforms about 10 percent better than DICE trained on the previous datasets when the test index remains unchanged, in terms of Recall and Hit Ratio. And the test results are also better than another state-of-the-art baseline model.

In summary, the main contribution of this paper are as follows:

●We propose a method that takes social networks into account when disentangling users' conformity and interest.

●Extensive experiments were carried out on a real-world recommendation system's large-scale dataset, and the findings demonstrate that the DICE model, trained on the new dataset, performs better than the state-of-the-art baseline model.

●Sensitivity analysis on the impact of our proposed social network utilization on dice model results

The remaining parts of this paper are as follows. In Section 2, we introduce the motivation and outline the problem. Then, in Section 3, we provide a detailed explanation of the proposed personal popularity approach. In Section 4, we present the experimental setup and results. Finally, in Section 5, we conclude the paper.

2. Motivation and Problem Formulation

Motivation Disentangling the influence of user interest and conformity towards an item provides challenges within the framework of independent and identically distributed reality. In other words, in the recommendation model, a user's click can be attributed to either their personal interest or conformity influenced by other users. According to the original paper, "Disentangling User Interest and Conformity for Recommendation with Causal Embedding" (Yu Zheng, Chen Gao, Xiang Li, etc. 2021,) the DICE model initially concludes the relationship based on the popularity of the items by using the colliding effect. The user shows interest when he clicks on an unpopular item instead of a popular item, and he shows both interest and conformity when he clicks on a popular item instead of an unpopular item. By splitting the data into interest and conformity embeddings, the Dice model integrates these causes using multitask curriculum learning, resulting in a more accurate result compared to other state-of-the-art casual recommendation models.

However, this model does not take into account the impact of social networks which is the relationship between different users. Since the recommendation system without social networks will prioritize recommending items that are based on the user's historical data, this type of system will gradually become more accurate with increasing user interactions. Simultaneously, the drawbacks of this recommendation system are also apparent. Suppose a new user registers and visits the system for the first time, the traditional recommendation system lacks historical data for analysis, resulting in a low accuracy result. In addition, there is a risk of generating induced data, which ultimately produces results that do not meet expectations. On the other hand, considering social networks by analyzing the relationship between users can enhance the robustness of the recommendation system.

Social networks facilitate more interactions among users, enabling the exchange of opinions. To be specific, Users' interests are often influenced by their peers, as they tend to trust and value the recommendation from friends. For example, when Andy posts a photo of his favorite bicycle on a social media platform, his friends who have seen the post are more inclined with Andy's recommended products when they need to purchase a bicycle in the future. Additionally, social networks influence the prospective connections between users. Through interactions and establishing connections, users and their friends are more likely to form social groups with similar interests, which allows for a closer analysis of the reasons behind user clicks that align with real-world scenarios. Therefore, by considering the influence of social networks, we can enhance the robustness and interpretability of recommendation systems, thereby improving the overall user experience. To reach this approach, we introduce personal popularity in this paper, separating the relationship into trust [3] and distrust between different users.

Problem Formulation The problem can be formulate as disentangling users conformity and interest considering with social networks. It involves processing and manipulating the original data set, selecting high rating data as the training set and test sets, introducing the concept of personal popularity (pp_i) based on the existing DICE model. Additionally, the relationship between users are indicated by tU_i and dU_i, representing the set of users that user u trust and distrust. Our objective is to maximize performance in terms of the ratio, NDCG, and hit ratio.

Input: 60 percent intervened record as training set and 40 percent intervened record as test set, with trust and distrust personal popularity tU_i and dU_i.

Output: A model that can predict users' interactions based on the social networks.

3. Popularity

Recommendation systems usually use observation data for model learning, and observation data is often affected by the popularity of the item, resulting in Popularity Bias (users do not necessarily like an item but click on the item because they follow the trend). The existing methods use the method of weighting the training samples or the method of using a small amount of completely random data for correction. However, these methods do not consider that different users are affected by popularity bias differently, and do not really distinguish between the reasons for an interaction (whether it is because it is following the trend or because it is liked), but all the interaction reasons are expressed in a unified vector.

3.1. Refer to the Specific Method of the Model in the Paper

The user's consumption behavior of materials can be derived from both user interest and user herdness, and the author splits the user's feedback score S_{ui} on an item into the sum of two parts, user interest and user affected by popularity:

$$S_{ui} = S_{ui}^{interest} + S_{ui}^{conformity} \tag{1}$$

DICE uses two sets of embeddings to separately calculate the interest matching score $S_{ui}^{conformity}$ and the conformity matching score $S_{ui}^{conformity}$:

$$S_{ui}^{int} =< u^{int}, i^{int} >, S_{ui}^{con} =< u^{com}, i^{con} > \tag{2}$$

As can be seen from the following cause-and-effect diagram, the user's clicking behavior depends on both user interest and user conformity (can also be called consistency), interest and herd are the cause, and the user's clicking behavior is the effect. Intuitively, interest embedding and herd embedding need their own specific samples to train, but in fact, it is impossible to accurately determine the cause behind user behavior, so it is impossible to accurately segment the sample. However, in the causal inference, the three nodes of user interest, herdness and user click behavior form a collision structure, and the conditional correlation theory in the collision structure can be used to roughly speculate the cause behind user behavior.

3.2. Refer to the Mthodology for Calculating Popularity in the Papar

When "user behavior" is determined, the reason behind it, "user interest", is correlated with herdness". For example, when a user clicks and the popularity of the material is low, it can be inferred that the click behavior is triggered by user interest. In order to separate, the paper splits the data set into O1 and O2 according to the popularity of positive and negative samples, before introducing the following symbols, let M^I represent the interest matching matrix between users and items, and M^C represent the popularity (consistency) matching matrix between users and items.

- O1: When the user clicks product A but does not click on product B, and the popularity of positive sample A is greater than that of negative sample product B, it can be seen from the popularity size that the herd matching score of positive sample A is greater than that of negative sample B, and the total matching score of positive sample A is greater than that of negative sample B from the relationship between positive and negative samples:

$$M_{ua}^{C} > M_{ub}^{C}$$

$$\mathrm{M}_{\mathrm{ua}}^{\mathrm{C}} + \mathrm{M}_{\mathrm{ua}}^{\mathrm{I}} > \mathrm{M}_{\mathrm{ub}}^{\mathrm{C}} + \mathrm{M}_{\mathrm{ub}}^{\mathrm{I}}$$

- O2: When the user clicks on product A and does not click on product B, and the popularity of negative sample product B is greater than that of positive sample product A, it can be seen from the popularity size that the positive sample herd matching score is less than the negative sample, and the positive sample total matching score can be obtained from the positive and negative sample relationship is greater than the negative sample, and it can be seen from the collision structure condition that the click behavior generated in low popularity may be caused by user interest, so the positive sample interest matching score is greater than the negative sample:

$$M_{ua}^{C} < M_{ub}^{C}$$

$$\mathrm{M}_{\mathrm{ua}}^{\mathrm{I}} > \mathrm{M}_{\mathrm{ub}}^{\mathrm{I}}$$

$$\mathrm{M}_{\mathrm{ua}}^{\mathrm{C}} + \mathrm{M}_{\mathrm{ua}}^{\mathrm{I}} > \mathrm{M}_{\mathrm{ub}}^{\mathrm{C}} + \mathrm{M}_{\mathrm{ub}}^{\mathrm{I}}$$

In the discussion of the above inequality relationship, the motivation behind the user's behavior depends on the comparison of the popularity of positive and negative samples, and the negative sample is randomly sampled, as if the motivation for its behavior is also random, which is counterintuitive. The user's motivation for behavior solidifies at the beginning of the behavior, and no longer depends on how it is subsequently derived, and from this point of view, the above inequality is difficult to believe. In fact, the derivation of this inequality is rather crude and only holds true if the positive sample popularity is large enough or small enough. The popularity of negatively sampled materials is limited in the paper to enhance confidence in the inequality. That is, when the positive sample popularity is p, the negative sample is randomly sampled from materials with popularity greater than $p + m_{up}$ or less than $\mathrm{p} - \mathrm{m}_{\mathrm{down}}$. Let the sample space be O, the part of the positive sample popularity greater than the negative sample head O_1 and the positive sample popularity less than the negative sample part O_2. The loss of the model is the pairwise loss of the relative size of the learning matching score. From inequality 3 and inequality 4, it can be seen that sample O_1 can be used to train herd embedding, and sample O_2 can be used to train both interest embedding and herd embedding. The paper uses Equation 1 to additionally add the click main task to prevent the model from being deviated by the rough inequality.

3.3. A Method of Calculating Popularity after Joining Social Networks

But we don't think this makes sense, we need to consider the factors affected by popularity, there are many big and small factors, and this article does not consider the impact of the most important social networks on popularity.

To take an extreme hypothesis, if the users in our dataset come from two completely disconnected societies, such as two small islands on the ocean with no communication. Then intuitively, when we train the relevant embedding of the user of Kojima A, we should not consider the user data in Kojima B. However, if we use the algorithm of the original text, an item that is only frequently interacted with by the user of Island A, when training the user in Island B, we will still think that it is a very popular item, but in fact, the user in Island B does not interact with this item. Because when two individuals have no relationship, we should not use our inherent thinking to look at the same situation, which may be another result, which is the great role of social networks.

When calculatting popularity in the article, P_i is the popularity calculated using the method in the original article, the calculation method used in the original model was to count the total number of times an item was interacted with by all people:

$$P_i = \Sigma_{u \in U, i \in I}[(u, i) \in \text{TrainRecord}] \tag{3}$$

That dosen’t take into account the impact of social networking. So we consider using social networks to improve the way popularity is calculated. The user_rating data present in the epinions dataset was used and run, documenting trust and distrust relationships between users.

In a dataset where some users may trust and distrust a very small number of users, the popularity of many different items that Equation 2 gets will be the same. It is not convenient to determine whether the sample pair (u, i, i) belongs to O_1 or O_2.

Since we want to determine whether the sampled positive and negative sample pairs (u, i, j) belong to O_1 or O_2, then we need to know how popular items i and j are among users related to user u. Suppose the set of users that user u trusts is tU_i; The set of users that user u does not trust is dU_i, pp_i is the popularity of personal calculated after joining social networks. From this, we get a new popularity calculation formula (Equation 2), which indicates personal popularity:

$$PP_i = \Sigma_{u \in tU_i, i \in I}[(u, i) \in \text{TrainRecord}] - \Sigma_{u \in dU_i, i \in I}[(u, i) \in \text{TrainRecord}] \tag{4}$$

However, because in the data set, some users trust and do not trust users may be particularly few, the popularity of many different items obtained in Formula 2 will be the same, and it is not convenient to determine whether the sample pair (u, i, j) belongs to O_1 or O_2, p_i' is a result used to determine whether a sample pair (u,i,j) belongs to O_1 or O_2 by considering the weighted average of old popularity and new popularity, that is, (Equation 3):

$$P_i' = (|U| - |tU_i| - |dU_i|)P_i + (|tU_i| + |dU_i|)PP_i \tag{5}$$

4. Experiments

This section presents a series of experiments aimed at demonstrating the effectiveness of the proposed approach, with a particular focus on addressing the two following research questions.

(1) What impact does the social network information incorporation have on the interpretability and robustness of DICE framework? Does it contribute to an enhanced recommendation performance?

(2) How does modulating the proportion of social network information combined into the dataset influence the training efficacy of DICE?

4.1. Experimental Settings

Datasets The experiments utilize a dataset sourced from real-world applications, Epinions, comprising approximately fourteen million records. Table 1 provides a list of the files included in the dataset as well as the related column details, and Table 2 outlines the statistical details.

DOI: 10.54254/2755-2721/44/20230247

Table 1. Files included in the Epinions dataset

rating.txt	user_rating.txt	mc.txt
OBJECT_ID	MY_ID	CONTENT_ID
MEMBER_ID	OTHER_ID	AUTHOR_ID
RATING	VALUE	SUBJECT_ID
STATUS	CREATION	NA
CREATION	NA	NA
LAST_MODIFIED	NA	NA
TYPE	NA	NA
VERTICAL_ID	NA	NA

Table 2. Statistics of the Epinions dataset

User	Item	Rating	Trust Relation	Distrust Relation
120492	755760	13668320	717667	123705

Data Preprocessing Our work primarily focuses on the following tasks.

(1) Cleaning the data.

In the Epinions dataset, a binarization process[4] is applied where ratings of five stars are assigned a value of one, while any other ratings are assigned a value of zero. As a result, we retain only the relevant data with a rating of five. Additionally, to reduce the dataset size, we remove values associated with user-item interactions that occur fewer than one hundred times.

(2) Generating trust and distrust numbers.

In light of the aforementioned objective, this approach endeavours to augment the concreteness of social networks of users. Specifically, we focus on classifying values to trust and distrust, with a value of one denoting trust and a value of negative one representing distrust.

(3) Determining personal popularity.

In addition to considering item popularity, we incorporate a user perspective into the popularity calculation, resulting in a measure of the level of user interest in different items. By obtaining the item lists associated with social networks of a user, we update the popularity of these items in the personal popularity matrix at their respective positions. This updating method takes into account the influence of trust and distrust relationships among users, allowing the personal popularity of items for a particular user to be adjusted based on the rating behaviours in the social networks.

(4) Intervening test sets.

To assess the performance of causal learning in non-IID (non-independent and identically distributed) [5] scenarios, it is necessary to use intervened test sets.

In order to create the training and testing datasets, we employ a random sampling approach, selecting 60% of the records based on items with equal probability. These selected records are then designated as the training data. The remaining 40% of the records are kept as the initial test data. Consequently, a test set will consist of 20% of the records with interventions applied. In our research, two intervention methods on the initial test set are predominantly employed.

●Intervening solely on items.

With the aim of mitigating conformity bias, we implement a method that considered item popularity. The level of popularity is calculated using the frequency of occurrence associated with each item in the

initial test set. Probabilities are then assigned to the items, with higher-popularity items receiving lower probabilities. Thus, the likelihood of selecting highly popular items can be reduced.
Additionally, we apply a cap on the probabilities to limit the inclusion of items in the initial test set that did not appear in the training set. This adjustment ensures that even relatively unpopular items have a chance to be selected during the intervention process.

●Intervening on both users and items simultaneously.

Building upon the previous method, we incorporate user preferences into the selection process.

To begin with, ten thousand items are randomly chosen based on the probabilities, and the list of users who interacted with them can be obtained. Afterwards, we match the users to the items based on their relationship weight, where the relationship weight between a user and an item is determined by considering the personal popularity, as well as trust and distrust numbers.
As a consequence, there might not be a substantial difference in the conformity of test data compared to that of the training data. This is because users can encounter more popular items in both the training and test sets, resulting in a certain level of overlap and similarity between the two datasets.

Recommendation Models In the field of recommendation systems, causal methods are often utilized as supplementary techniques to improve the performance of core recommendation models, but it's still an integral part [6]. In our research, we test the results of the cause model and the new Dice model with different social network utilization rates.

Experiment Setups During our experiments, we kept the embedding size consistent at 256 for both CausE [7] and DICE. Additionally, we assign fixed values of 0.1 to α and 0.01 to β. We utilize the Bayesian Personalized Ranking [8] loss function across all baselines and optimized our models using the Adam [9] optimizer. The optimal values for other hyperparameters of our method and the baselines are identified through a grid search.

4.2. Performance Comparison (RQ1)

We evaluate our method against two state-of-the-art causal recommendation techniques: CausE and DICE. We measure the performance of all approaches using three commonly used metrics: Recall, Hit Ratio, and NDCG. The results of our experiments are presented in Table 3.

Table 3. Test results on Epinions dataset

Method	**TopK=20**			**TopK=50**		
	Recall	**Hit Ratio**	**NDCG**	**Recall**	**Hit Ratio**	**NDCG**
Cause	0.02018	0.56924	0.12705	0.03769	0.68372	0.11485
DICE	0.01932	0.51521	0.10031	0.03394	0.64742	0.09491
New DICE	0.02592	0.57156	0.12282	0.04528	0.69968	0.11527

Our findings indicate that, based on the test set we designed, the performance of the DICE recommendation model was not satisfactory and even inferior to that of the causal models. However, the incorporation of social network information in the new DICE model resulted in superior performance compared to the causal models.

4.3. Sensitivity Analysis (RQ2)

To investigate the impact of social network utilization on the DICE model, we conducted experiments with the same parameters as the previous section but with varying levels of social network utilization[10] (social_usage). Social network utilization is defined as the proportion of users who use new Linear interpolation in training. For example, a model with social_usage=0 is equivalent to the original DICE model, while a model with social_usage=1 is equivalent to the new DICE model. The results of these experiments are shown in Figure 1.

It is evident from the above graph that with the increase in social usage, the performance metrics of the Dice model also improve. When the social usage reaches 75%, the new Dice model's performance is comparable to that of the Cause model. Furthermore, when social usage reaches 100%, the Dice model's performance is significantly better than that of the original Dice model and slightly better than Cause model.

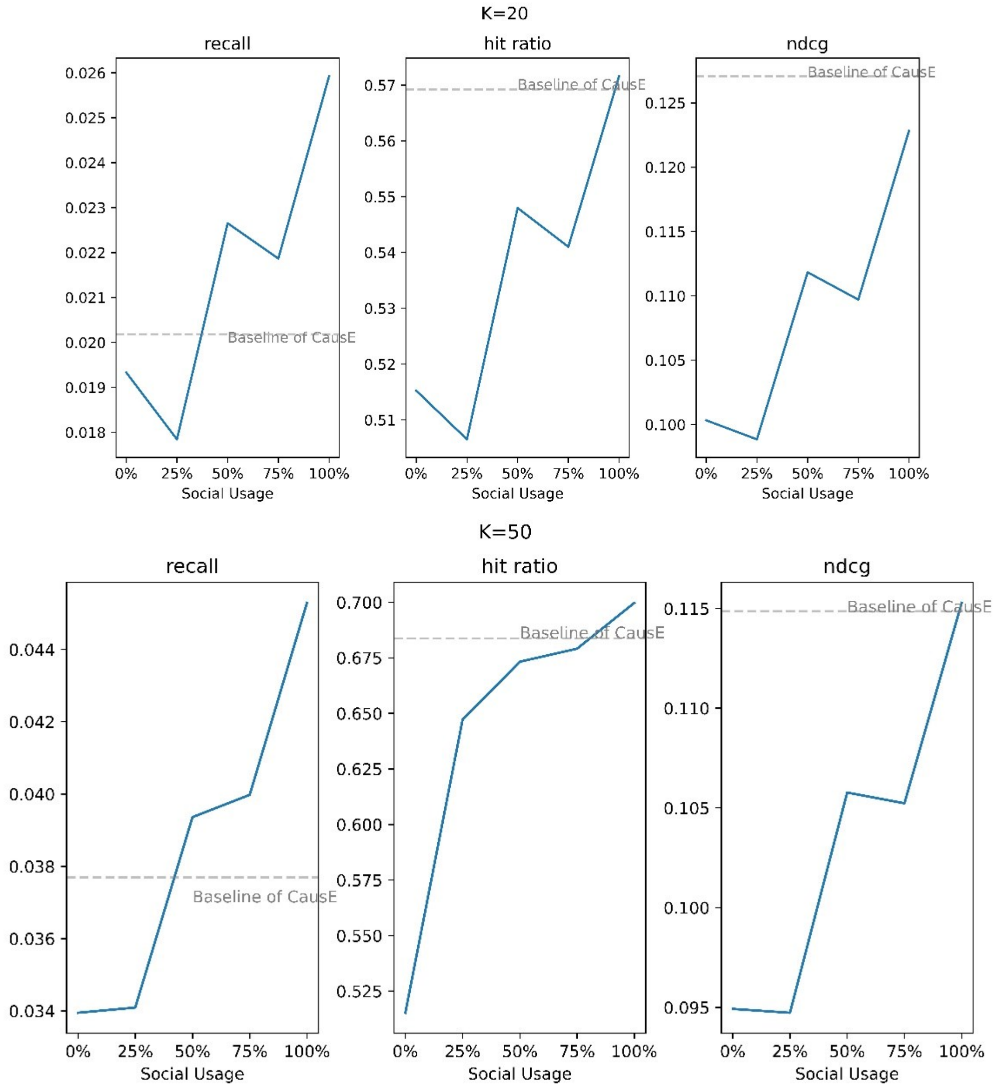

Figure 1. Test results under different social_usage

5. Conclusion

In this paper, we address the issue of inadequate consideration of social network information in the previous DICE model by proposing an optimization approach. Specifically, we generate a new popularity measure by combining the commonly used item popularity and the personal popularity derived from social network data. This popularity measure is then utilized in the training of the DICE model to enable the model to better incorporate social network information and achieve improved performance.

Experimental results demonstrate that our new DICE model outperforms the original DICE model on the Epinions dataset and achieves slightly better performance than the commonly used CausE model. Sensitivity analysis further reveals that the improvement in performance is positively correlated with the increase in social network utilization. In practical applications, we suggest that the new DICE model should be considered when the social network utilization rate is above 75%. Otherwise, the simple CausE model may yield better results.

References

[1] Zheng, Y., Gao, C., Li, X., He, X., Jin, D., & Li, Y. (2021). Disentangling User Interest and Conformity for Recommendation with Causal Embedding. Proceedings of the 2021 World Wide Web Conference. DOI: 10.1145/3442381.3449788
[2] Masoud Reyhani Hamedani, Irfan Ali, Jiwon Hong, and Sang-Wook Kim, "TrustRec: An Effective Approach to Exploit Implicit Trust and Distrust Relationships along with Explicit ones for Accurate Recommendations," Computer Science and Information Systems, Vol. 18, No. 1, Jan. 2020.
[3] Frank E. Walter, Stefano Battiston, and Frank Schweitzer. 2008. "A Model of a Trust-based Recommendation System on a Social Network." Autonomous Agents and Multi-Agent System 16, 57-74, 2008.
[4] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S. (2017). Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web. Perth. pp. 173–182.
[5] Jonas, P., Dominik, J., Bernhard, S. (2017). Elements of causal inference: Foundations and learning algorithms. MIT press, Cambridge.
[6] Chen, G., Yu, Z., Wang, W., Feng, F., He, X., Li, Y. (2022) Causal Inference in Recommender Systems: A Survey and Future Directions. arXiv:2208.12397.
[7] Bonner, S., Vasile, F. (2018). Causal embeddings for recommendation. In: Proceedings of the 12th ACM Conference on Recommender Systems. Vancouver. pp. 104–112.
[8] Rendle, S., Freudenthaler, C., Gantner, Z. Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. In: The 25th Conference on Uncertainty in Artificial Intelligence. Montreal. pp. 452-461.
[9] Diederik P Kingma and Jimmy Ba. (2014). Adam: A method for stochastic optimization. In: The 3rd International Conference for Learning Representations. San Diego.
[10] Xu, R., Frank, K. (2021). Sensitivity analysis for network observations with applications to inferences of social influence effects. Network Science, 9(1), pp. 73-98. doi:10.1017/nws.2020.36