RESEARCH ARTICLE OPEN ACCESS

Social Recommendation with Cross-Domain Transferable Knowledge

Yadala Sucharitha*, P Chandra Shaker Reddy**
*Assistant Professor, Dept. of CSE, CMR Institute of Technology, Hyderabad, India
Email: suchi.yadala@gmail.com
**Assistant Professor, Dept. of CSE, CMR College of Engineering & Technology, Hyderabad, TS, India.
Email: chandu.pundru@gmail.com

Abstract:

Recommender systems can suffer from data sparsity and cold start issues. However, social networks, which enable users to build relationships and create different types of items, present an unprecedented opportunity to alleviate these issues. In this paper, we represent a social network as a star-structured hybrid graph centered on a social domain, which connects with other item domains. With this innovative representation, useful knowledge from an auxiliary domain can be transferred through the social domain to a target domain. Various factors of item transferability, including popularity and behavioral consistency, are determined. We propose a novel Hybrid Random Walk (HRW) method, which incorporates such factors, to select transferable items in auxiliary domains, bridge cross-domain knowledge with the social domain, and accurately predict user-item links in a target domain. Extensive experiments on a real social dataset demonstrate that HRW significantly outperforms existing approaches.

Keywords — **Social recommendation, transferability, cross-domain, star-structured graph .**

I. INTRODUCTION

A social networking service is a platform on which users can create and adopt different types of items such as web posts (e.g., articles and tweets), user labels, images, and videos. The huge volume of items generates a problem of information overload. Traditional web post recommendation approaches suffer from data sparsity (i.e., limited interaction between users and web posts) and the issue of cold start (i.e., giving recommendations to new users who have not yet created any web posts). The social connections and multiple item domains found in social networks provide an unprecedented opportunity to alleviate these issues in real applications. One common type of approach to recommendations, known as collaborative filtering (CF) techniques, characterizes users' latent features independently with user-item interactions in a single item domain [1]. Similarly, the type of approach provided in [2] does not consider the question of multiple domains. However, users' characteristics relate both to social connections and to different user-item interactions. For example, users read web posts created by their community and may adopt similar user labels to their friends. Therefore, an effective social recommendation approach should acknowledge (1) social tie strength (henceforth, tie strength) between users and (2) different user-item interactions. . The problem of how to incorporate a social domain and auxiliary item domains (e.g., user labels and images) into a unified framework remains open. Frameworks exist that connect directly-related item domains, such as a music album and tags on that album [3], or web pages and queries to them [4]. However, these cannot be applied to indirectly-related item domains in social networks, such as tweets and user labels. The multiple item domains reflect users' intrinsic preferences and tend to be tightly connected among a massive number of users. In this paper, we reconsider the representation of social networks and propose a star-structured graph, where the social domain is at the center and is connected to the surrounding item domains. The value of the cross-domain link1 weight represents how often a given user adopts a given item, while the value of the within-domain link2 weight in the social domain represents the tie strength between users. Tie strength can refer to homophily [5], circle-based influence [6], [7], [8], or social trust. Users are more likely to have stronger ties if they share similar characteristics. Cross-domain links reflect users' characteristics in different ways. For example, a cross-domain link from a user to a web post about iPhones shows his/her short-term interest in iPhones, and a crossdomain link from him/her to a label "iPhone Fan" implies his/her long-term interest in iPhones. A basic assumption is that the more auxiliary knowledge we have, the more we know about the users, thereby enabling more accurate estimates of tie strength. When a user and his/her friend have many common user labels, we assume greater tie strength and expect them to be more similar in terms of their web post adoption behaviors [9].

A. Background

The domains are relational. Social network data provide social connections between users, semantic similarity between two items of the same type, and item adoptions by users. The issue of how to represent the user-user links, item-item links, and user-item links poses a challenge to method capability. The domains are heterogeneous. Heterogeneity is a challenging issue in social recommendation. Within domain links can be directed ("following" links in the social domain) or undirected (semantic similarity links in the item domains). Cross-domain links can be signed (indicating a positive or negative connotation, such as web-post adoptions and rejections) or unsigned (user-label adoptions) [10, 11]. The issue of how to transfer knowledge across heterogeneous domains poses a challenge to method comprehensibility. The domains are variously sparse. This data sparsity is essentially caused by the large amounts of users and items as well as the time and attention scarcity of these users. It is challenging to try to use relatively dense auxiliary information to help predict sparse links in the target domain. Items in the domains have varying transferability. Traditional literature often assumes that the most popular items have better transferability. However, later in this work, we will show that this assumption is incorrect. Therefore, transferable knowledge selection approaches for enhancing performance constitutes a literature gap. To address the above challenges, we propose an innovative hybrid random walk (HRW) method for transferring knowledge from auxiliary item domains according to a starstructured configuration to improve social recommendations in a target domain. HRW estimates weights for (1) links between user nodes within the social domain, and (2) links between user nodes in the social domain and item nodes in the item domain. The weights respectively represent (1) tie strength between users and (2) the probability of a user adopting or rejecting an item. Our proposedmethod integrates knowledge from multiple relational domains and alleviates sparsity and cold-startissues. We reconsider the representation of social networks and propose a star-structured graph, where the social domain is at the center and is connected to the surrounding item domains [12, 13].

B. Scope

The domains are relational. Social network data provide social connections between users, semantic similarity between two items of the same type, and item adoptions by users. The issue of how to represent the user-user links, item-item links, and user-item links poses a challenge to method capability. The domains are heterogeneous. Heterogeneity is a challenging issue in social recommendation. Within domain links can be directed ("following" links in the social domain) or undirected (semantic similarity links in the item

domains). Cross-domain links can be signed (indicating a positive or negative connotation, such as web-post adoptions and rejections) or unsigned (user-label adoptions). The issue of how to transfer knowledge across heterogeneous domains poses a challenge to method comprehensibility. The domains are variously sparse [14, 15]. This data sparsity is essentially caused by the large amounts of users and items as well as the time and attention scarcity of these users. It is challenging to try to use relatively dense auxiliary information to help predict sparse links in the target domain. We reconsider the representation of social networks and propose a star-structured graph, where the social domain is at the center and is connected to the surrounding item domains. Rest of the article organized as follows. In section related works are discussed, section 3 explained about methodology, section 4 represents about experimental results and conclusion is discussed in section 5.

II. RELATED WORKS

Ref. [16] proposed sample based algorithms that capture information in the neighbourhood of a user in dynamic social networks utilizing random walks. Ref. [17] studied the distribution of tags in the social bookmarking site del.icio.us and propose a generative model of collaborative tagging in order to evaluate the dynamics that lie beneath the act of collaborative recommendation. Their findings prove that the dataset collected follows a power-law distribution. Even though both studies examine social networks that are based on social tagging, they do not explore the dynamics of friendships among users. Taking into account the power of free-form tagging of items by users other than their authors/owners, researchers also focus on tag recommendation. Ref. [18] proposed a system that automatically recommends tags for blogs, using similarity ranking in a manner similar to collaborative filtering techniques. Ref. [19] studies a novel idea in tag recommendation, which bridges the gap between the keywords issued by a user in a query and the tags actually used by a social system. He argues that the tags used by a user when performing a query exhibit his or her intent, whereas the annotations of items describe content semantics. As a result, he proposes a new form of purpose tags, which extract the intent of the user and facilitate goal-oriented search in a social network [20]. Both studies underline the importance and discriminative power of social tagging, which is also validated by our work. Several studies exist in the field of applying Random Walks on bipartite graphs. Ref. [21] studied a click through data graph in order to perform item recommendation. Nevertheless, no social content is available between users. Ref. [22] proposed a novel recommendation algorithm which performs Random Walks on a graph that denotes similarity measures between items. They evaluate their system using data from MovieLens. Although, the use of the Random Walk model performs well in the context of recommendation, their use of an ItemItem similarity matrix raises some issues as to the ability of the system to extend when other similarities are introduced based on social tagging.

The usage of social network data has been found to improve the prediction accuracy of rating values, and various models for integrating these two data sources have been proposed, like Social Recommendation (SoRec), Social Trust Ensemble (STE) , Recommender Systems with Social Regularization [23], Adaptive social similarities for recommender systems, among which the SocialMF model was found to achieve a particularly low RMSE value, and is hence used as a baseline model in our experimental comparison study. The Social MF model was proposed in [24], and was found to outperform SoRec and STE with respect to RMSE. The social network information is represented by a matrix $S \in Ru0\times u0$, where u0 is the number of users. Recommender Systems (RS) have the goal of suggesting to every user the items that might be of interest for her. In particular, RSs based on Collaborative Filtering (CF) rely on the opinions expressed by the other users. In fact CF tries to automatically finds users similar to the active one and recommends to her the items liked by these similar users. This simple intuition is effective in generating recommendations and widely used [25].

A more detailed explanation of the architecture can be found in. We are going to present experiments we have run with different instantiations of the different modules. For the Trust Metric module we have tested a local and a global trust metric. As local trust metric we have chosen MoleTrust, a depth-first graph walking algorithm with a tunable trust propagation horizon that allows us to control the distance to which trust is propagated. As global trust metric we have chosen PageRank, probably the most used global trust metric. For the Similarity Metric module we have chosen the Pearson Correlation Coefficient since it is the one that is reported to be performed best [26]

III. PROPOSED METHODOLOGY

The value of the cross-domain link1 weight represents how often a given user adopts a given item, while the value of the within-domain link2 weight in the social domain represents the tie strength between users. Tie strength can refer to homophily, circle-based influence, or social trust. Users are more likely to have stronger ties if they share similar characteristics. Cross-domain links reflect users' characteristics in different ways. For example, a cross-domain link from a user to a web post about iPhones shows his/her short-term interest in iPhones, and a cross domain link from him/her to a label "iPhone Fan" implies his/her long-term interest in iPhones [27, 28]. A basic assumption is that the more auxiliary knowledge we have, the more we know about the users, thereby enabling more accurate estimates of tie strength. When a user and his/her friend have many common user labels, we assume greater tie strength and expect them to be more similar in terms of their web post adoption behaviors. Even if the web post domain is extremely sparse, we may still produce effective recommendations by transferring auxiliary knowledge from other item domains through the social domain and proposed model diagram is shown in Figure 1.

Thus, knowledge transfer procedures among multiple item domains in social networks should focus on updating tie strength in the social domain, but this is complicated by challenges associated with jointly modeling multiple relational domains, discovering transferable knowledge, and improving recommendations in the target domain [29].

A. *Preparation of Dataset*

The dataset for this research was crawled in January 2011 from Tencent Weibo (t.qq.com). We crawled data from users who own at least one user label. While the website allows users to have, at most, 10 user labels, the average number of user labels per user was 5.3. The average number of web posts per user was 12.8. We did not filter any social relationships. The average number of friends per user was 14.2. We used a 5-minute time window to derive negative links. That is, if a user had two adopting behaviors (sharing the web posts) in 5 minutes, we assumed that the user ignored the rest of posts that he/she received in the time window. Thus, besides the two positive user-post links, we noted several negative links. The data indicates that although both web-post and user-label domains are sparse, the latter is denser. (1) Distributions of user and post frequency, (2) Distributions of user and label frequency, and (3) Distributions of follower and followee frequency. We note that the data has smooth distributions, which look like power law relationships in log-log scale. Our dataset has no spiky outliers [30].

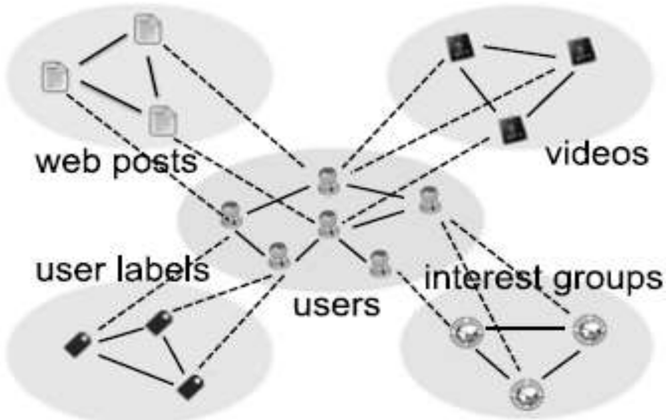

Fig.1. Proposed model Architecture

B. *Transferability of User Labels*

We conduct data analysis to demonstrate that (1) the auxiliary, user-label domain can be transferred to predict a target web-post domain; that is, user-label interactions are in some degree consistent with user-post interactions; (2) user-label interactions are also consistent with user-user interactions in the social domain; and (3) not every label can be transferred, and the most popular labels are not the most transferable.

C. Hybrid Random Walk Algorithm

We introduce our random walk-based method on social recommendation. Owing to data sparsity in the target domain, traditional bipartite random walk (BRW) algorithms cannot accurately derive user tie strength to predict user behaviors in the target domain. Fortunately, we have auxiliary domains in which user ties are formed for the same reason as in the target domain: homophily, trust, and influence. The key idea is to utilize rich knowledge from the auxiliary domains to better describe user tie strength and then more precisely predict user behaviors. Thus, we derive HRW algorithms on star-structured graphs.

We derive a random walk algorithm to predict missing links on G(UP) and G(UT), which includes both within-domain and cross-domain random walks. For G(U), G(P) and G(T), we derive steady-state distributions, indicating the intrinsic relevance among users, posts and labels. For a standard random walk model, a walker starts from the ith vertex and iteratively jumps to other vertices with transition probabilities pi ={pi1; . . . ; pin}. After reaching the steady state, the probability of the walker staying at the jth vertex corresponds to the relevance score of vertex j to i. We assumed that two types of item domains, web post and user label, are associated with each user. However, online social networks are an unprecedented comprehensive platform with a number of different types of user-generated content (UGC), e.g., posts, labels, music, and movies. Shows a typical example of a hybrid high-order star-structured graph with four different types of UGC. In this case, the second order graph is insufficient for describing all the UGC. Our random walk strategy can be easily extended to higher-order cases.

IV. EXPERIMENTAL RESULTS

In this section, we first compare the performance of our proposed HRW method and other comparable algorithms on predicting missing user-post links for social recommendation. Second, we discuss the transferability of different items in the auxiliary domain and the performance of our item selection method. Third, we discuss the transferability across domains, i.e., (1) how user tie strength works as a bridge, (2) how transferring auxiliary information performs, and (3) how positive/negative samples help. Finally, we show how our method solves the cold-start problem and sheds light on the usefulness of auxiliary information. Figure 2 shows MAP evaluations of our HRW-Borda and HRW-All methods and baselines for recommending the top K items. We observe that (1) our HRW methods produce higher MAP scores than the best BRW method (BRW-RU-P) and the matrix factorization method TLLSM, and (2) the HRW-Borda outperforms all the competitors. Figure 3 compares the RMSE of HRW-Borda method, with multiple settings of tie strength matrix W(U) initialization, including (1) Laplacian, i.e., the Laplacian matrix of degree matrix of the social graph. Further, in Figure 4, we plot precision-recall curves for all the algorithms and find that HRW-Borda reaches the best, almost perfect values. Figure 5 represents the recommendation analysis of different labels.

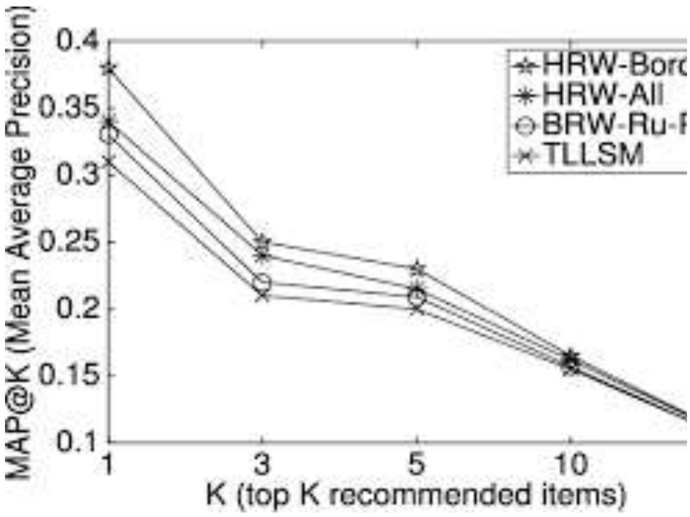

Fig. 2 . The results of MAP@K for different K for HRWBorda, HRW-all, BRW-RU-P and TLLSM.

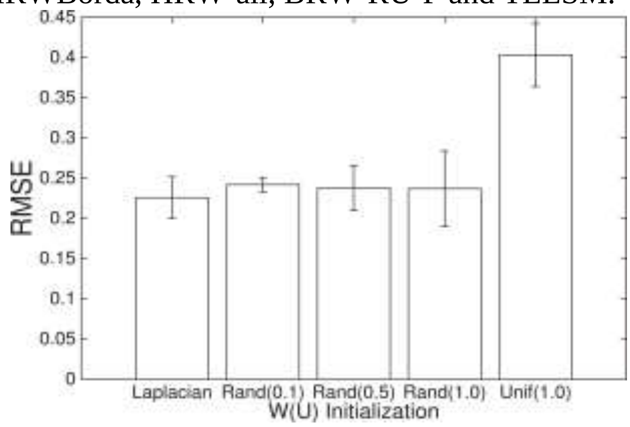

Fig. 3. Laplacian Rand(0.1) Rand(0.5) Rand(1.0) Unif(1.0) W(U) Initialization

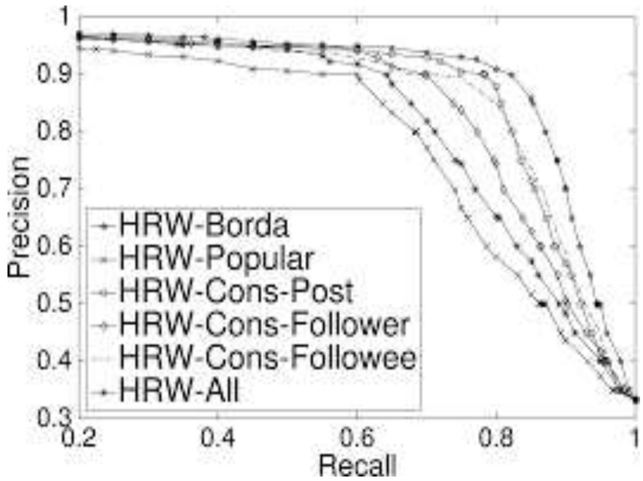

Fig. 4. Precision vs. recall for several HRW methods. HRW-Borda gives better precision-recall results than other HRW methods.

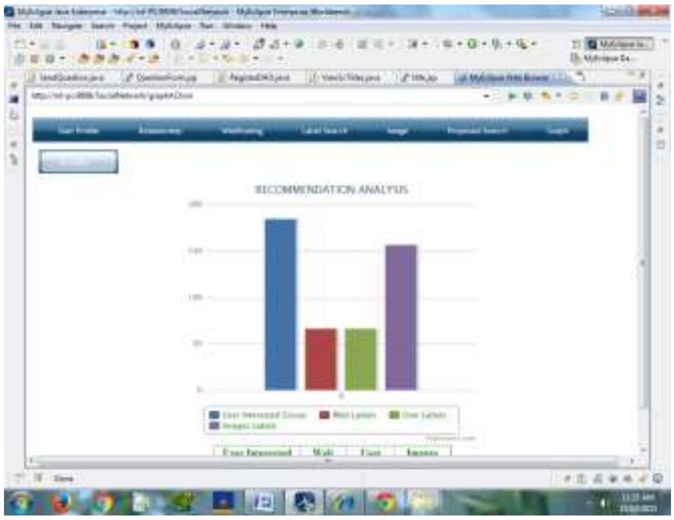

Fig. 5. Recommendation analysis

V. CONCLUSION

In this paper, we addressed the problems of data sparsity and cold start in social recommendation. We reconsidered the problem from the transfer learning perspective and alleviated the data sparsity problem in a target domain by transferring knowledge from other auxiliary social relational domains. By considering the special structures of multiple relational domains in social networks, we proposed an innovative HRW method on a star-structured graph, which is a general method to incorporate complex and heterogeneous link structures. We conducted extensive experiments on a large realworld social network dataset and showed that the proposed method greatly boosts the social recommendation performance. In particular, we gained improvement in web-post recommendation by transferring knowledge from the user-label domain for the user tie strength updating process, compared with the recommendation methods, which only use information from the web-post domain. In addition, we demonstrated that, by using only 27.6% of the available information in the target domain, our method achieves comparable performance with methods that use all available information in the target domain without transfer learning. The proposed method and insightful experiments indicate a promising and general way to solve the data sparsity problem.

REFERENCES

[1] Y. Koren, R. Bell, and C. Volinsky, ”Matrix factorization techniques for recommender systems,” in Computer, vol. 42, pp. 30–37, 2009.

[2] H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King, ”Recommender system with social regularization,” in WSDM, pp. 287–296, 2011.

[3] I. Konstas, V. Stathopoulos, and J. M. Jose, ”On social networks and collaborative recommendation,” in SIGIR, pp. 195–202, 2009.

[4] J. Leskovec, A. Singh, and J. Kleinberg, ”Patterns of influence in a recommendation network,” in LNCS, vol. 3918, pp. 380–389, 2006.

[5] X. Yang, H. Steck, and Y. Liu. ”Circle-based recommendation in online social networks.” in SIGKDD, pp. 1267–1275, 2012.

[8] X. Qian, H. Feng, G. Zhao, and T. Mei. ”Personalized Recommendation Combining User Interest and Social Circle.” in IEEE TKDE, vol. 26, no. 7, pp. 1763–1777, 2014.

[9] M. Jamali and M. Ester, ”TrustWalker: a random walk model for combining trust-based and item-based recommendation,” in SIGKDD, pp. 397–406, 2009.

[10] Sucharitha Y, Vijayalata Y, Prasad VK. Analysis of Early Detection of Emerging Patterns from Social Media Networks: A Data Mining Techniques Perspective. InSoft Computing and Signal Processing 2019 (pp. 15-25). Springer, Singapore.

[11] Sucharitha Y, Prasad VK, Vijayalatha Y. Emergent Events Identification in Micro-Blogging Networks Using Location Sensitivity. J. Adv. Res. Dynam. Cont. Syst. 2019;11:596-607.

[12] M. Harvey, M. J. Carman, I. Ruthven, and F. Crestanig, ”Bayesian latent variable models for collaborative item rating prediction,” in CIKM, pp. 699–708, 2011.

[13] J. Noel, S. Sanner, K. N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad, and E. D. Penna. ”New objective functions for social collaborative filtering.” in WWW, pp. 859–868, 2012.

[14] Reddy PC, Babu AS. Survey on weather prediction using big data analystics. In2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT) 2017 Feb 22 (pp. 1-6). IEEE.

[15] M. Jiang, P. Cui, R. Liu, F. Wang, W. Zhu, and S. Yang, ”Scalable Recommendation with Social Contextual Information,” in IEEE TKDE, vol. 26, no. 11, pp. 2789–2802, 2014.

[16] J. Tang, S. Wu, J. Sun, and H. Su. ”Cross-domain collaboration recommendation,” in SIGKDD, pp. 1285–1293, 2012.

[17] C. Li, and S. Lin. ”Matching users and items across domains to improve the recommendation quality,” in SIGKDD, pp. 801-810, 2014.

[18] Y. Shi, M. Larson, and A. Hanjalic. ”Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges,” in ACM Computing Surveys, 47(1):3, 2014.

[19] Reddy PC, Sureshbabu A. An applied time series forecasting model for yield prediction of agricultural crop. InInternational Conference on Soft Computing and Signal Processing 2019 Jun 21 (pp. 177-187). Springer, Singapore.

[20] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo. ”Crossdomain recommendation via cluster-level latent factor model,” in Machine Learning and Knowledge Discovery in Databases, pp.161-176, 2013.

[21] W. Chen, W. Hsu, and M. L. Lee. ”Making recommendations from multiple domains,” in SIGKDD, pp. 892-900, 2013.

[22] J. Tang, X. Hu, H. Gao, and H. Liu. ”Exploiting local and global social context for recommendation,” in IJCAI, pp. 2712–2718, 2013.

[23] J. Tang, X. Hu, and H. Liu. ”Social recommendation: a review,” in Social Network Analysis and Mining, vol. 3, no. 4, pp. 1113–1133, 2013.

[24] S. Sedhain, S. Sanner, L. Xie, R. Kidd, K. N. Tran, and P. Christen. ”Social affinity filtering: Recommendation through fine-grained analysis of user interactions and activities.” in COSN, pp. 51–62, 2013.
[25] Reddy PC, Sureshbabu A. An Adaptive Model for Forecasting Seasonal Rainfall Using Predictive Analytics. InInternational Journal of Intelligent Engineering and Systems 2019 (pp. 22-32).
[26] B. Shapira, L. Rokach, and S. Freilikhman. ”Facebook single and cross domain data for recommendation systems,” in User Modeling and User-Adapted Interaction, 23(2-3), pp. 211–247, 2013.
[27] Sucharitha Y, Vijayalata Y, Prasad VK. Predicting Election Results from Twitter Using Machine Learning Algorithms. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science). 2021 Jan 1;14(1):246-56.