2022 International Conference on Inventive Computation Technologies (ICICT) | 978-1-6654-0837-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICT54344.2022.9850852

Hierarchical K-Means Clustering Method for Friend Recommendation System

Anju Taiwade
M. Tech Scholar
Department of Computer Science and Engineering
Technocrats Institute of Technology & Science
Anjutaiwade11@gmail.com

Prof. Nitish Gupta
Associate Professor
Department of Computer Science and Engineering
Technocrats Institute of Technology & Science
Gupta_neetesh81@yahoo.com

Prof. Rakesh Tiwari
HOD
Department of Computer Science and Engineering
Technocrats Institute of Technology & Science
rakeshktiwari80@gmail.com

Dr. Shashi Kumar Jain
Director
Department of Computer Science and Engineering
Technocrats Institute of Technology & Science
shashi.k.jain@gmail.com

Upendra Singh
Assistant Professor
Department of Information Technology
Shri Govindram Seksaria Institute of Technology and Science, Indore
Upendrasingh49@gmail.com

Abstract: Friend suggestion is one of the most popular social networking sites since it helps users connect with others who are similar or known to them. The notion of friend recommendation originated on social media platforms such as Twitter and Facebook, suggesting individuals use the friends-of-friends mechanism. It may be claimed that users do not form friends randomly but rather with their friends' friends. Existing techniques have a limited suggestion scope and are inefficient. Therefore, a novel buddy recommendation model that overcomes the shortcomings of the present system has been proposed. To create a more accurate buddy recommendation system, collaborative filtering is utilized to analyze users' similar and dissimilar data, and a recommendation system that provides user-to-user recommendations based on their comparable choices, activities, and preferences is created. Location-based buddy recommendation systems are gaining popularity because they connect the actual world to the digital realm and provide a complete picture of users' preferences or interests. This recommendation method will broaden the reach of recommendations from one user to another who has similar interests and is located in the same area.

Keywords: Friend recommendation, collaborative filtering, social network, Recommendation system.

I. INTRODUCTION

Friend suggestion is one of the most frequently used and basic services on the LSBN platform since it connects familiar or interested users. Around 71% of internet users were online social network users, and this number will continue to expand in the foreseeable future. Social networking is one of the most popular online hobbies, owing to its high user involvement rate and developing mobile capabilities. The fast proliferation of smartphones and other mobile devices has created new areas of mobile social networks with enhanced functionalities. With more than a billion monthly active users on social media. Facebook is now the industry leader regarding reach and breadth of user engagement [1].

Location-based services have recently gained popularity. Thanks to recent advancements in location-based technology, users may now share their locations and location-related data with their friends and family. The term for these social networks is "location-based social networks" (LBSNs). Users can make new friends by using the buddy recommendation tool on LBSN based on their previous location information.

Friend recommender engines, which are becoming more popular, pair users with possible friends based on their profiles, social structure, and connections with other people, among other factors. You may be able to offer more effective recommendations if you are aware of where your thoughts originate. User location histories reveal preferences, according to the

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

fundamental premise of the research; users with similar location histories have similar interests and are more likely to be friends, the theory states [2].

For typical social networks, a friend referral service is employed. However, very few systems use LBSN data for a recommendation. Earlier approaches relied heavily on GPS data to determine user similarity. Compared to GPS data, check-in data provides additional context-dependent information [3]. Additionally, the majority of LBSNs gather check-in data in addition to GPS trajectory data. The purpose of our suggested recommendation systems is to include user profiles, interests, and location histories (check-in data) and to use collaborative filtering techniques for user-to-user suggestions to broaden the scope and improve recommendation efficiency [13-17].

A location-based social network (LBSN) is a novel social structure of persons linked by their actual locations and location-tagged material such as text, pictures, and video [4]. Physical location does not just refer to an individual's instantaneous point position at a specific timestamp; it also refers to their location history over a defined period. Additionally, knowledge, shared interests, and preferred activities are extracted from an individual's location data, and location-related material affects social interactions in LBSN [5].

LBSN comprises a G U, C> and a G U, E > social network. In G U, E>, U denotes the set of users, and E is the edges that link or show a social relationship between various users in LBSN. G U, C> Check-in 'c' is associated with set C and indicates that user 'u' is associated with set U and has a check-in activity at location l at time t [2].

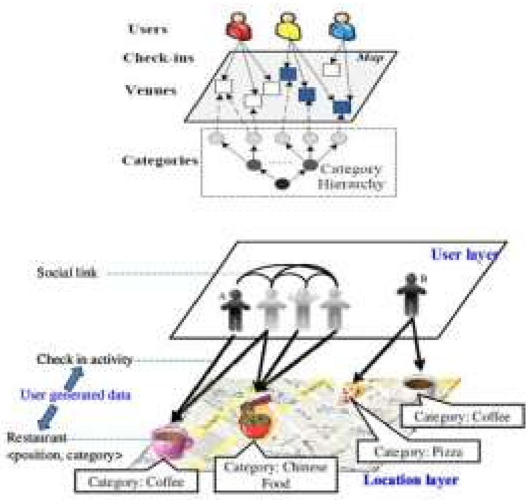

Figure 1 Location-Based Social Network [2]

II. RELATED WORK

Three primary characteristics of any LBSN are temporal, geographical, and social correlation. However, earlier algorithms are incapable of solving situations including these three characteristics. No solution fully leverages all available data. A novel technique for friend suggestion is offered to recommend friends that share LBSN users' location preferences. This strategy begins by calculating the user's social network relationship similarity using the local random walk method based on Markov chains. Second, it compares the user's location preferences in the actual world to those in the virtual world using check-in data and ultimately recommends friends to users using a mixed user preferences model [6].

The author proposes a novel friend recommendation model (FE-ELM) in which friend recommendation is seen as a binary classification issue. In this model, features are extracted using a variety of methodologies, and then ELM is used as the classifier to learn spatial-temporal, social, and textual features. Finally, trials are conducted on datasets to improve efficiency and accuracy [7].

The location brings unique traits and issues for LBSN recommender systems. Before that, the author grouped recommender systems into four categories: locations, people, activities and social media. Second, they categorize them into content-based, link-based, and collaborative filtering. In addition to user profiles, online and location histories may be used to categorize systems. Each system's goals and contributions are summarised, showing the usual research effort. It covers the basics, issues, evaluation

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

methods, and future Work in LBSN recommender systems [8].

We offer a hierarchical-graph-based similarity measurement framework (HGSM) to model people's location histories and determine user similarity. Three aspects are covered in this framework: the sequence of users' journeys, the hierarchy of physical places, and the popularity of various locales. In addition to using HGSM to assess user similarity, our approach employs a collaborative filtering-based technique to determine an individual's interest in previously unvisited geographical places [9].

The Random walk-based context-aware buddy recommendation system is presented (RWCFR). This model describes users, locations, and connections using an undirected unweighted graph. RWCFR creates a subgraph based on the user's current surroundings. This sub-graph includes popular users and well-known locations in the area. After generating the sub-graph is sent to the algorithm, which estimates the likelihood of users being suggested as possible friends. A list of possible buddies is constructed based on the random walk algorithm's output [10].

While recommendation systems use user profiles, friend descriptions, and prior behaviour, little attention has been paid to explicit social network customization. The author included the social graph between users, songs, and tags from the last—FM social network. We conducted a series of studies comparing the Random Walk with Restarts model against a collaborative filtering approach based on user input. The findings demonstrate that the graph model benefits from including social knowledge information [11].

The essay discusses the collaborative filtering algorithm's key issues and provides several remedies. To remedy the new user's cold start issue, we might refill the user's profile in several ways. The basic method would demand the user to supply their profile upon signing into their social account. At the same time, we could combine collaborative filtering and a content-based recommender algorithm for the new friend. The sparsity problem has a limited number of remedies. The first one utilizes filling or reducing the dimension to lessen the matrix's sparsity. Another way boosts algorithm efficiency without increasing the sparsity of the matrix [12].

Propose a new LBSN recommendation model that considers social, geographical, and textual factors. To increase buddy recommendation accuracy, the features are obtained based on criteria. We offer a novel component for extracting textual features and enhancing social and geographical elements. Two Twitter datasets were used to test the recommendation model. The model beat another friend recommendation model [13], with the results of proposing top-k friends being 97.88 percent correct.

Feed dynamic behaviour traits into a CNN to obtain an implicit vector representation of two users' interactions and conditions. A better multi-attention approach enhances the deep vector representation of attribute information. The LINE-based network embeddings approach is then used. A fully connected neural network (FCNN) constructs a deep feature representation that captures the possibility of two users becoming friends using the attribute attention vector and network embeddings. The possibility of two users becoming friends is calculated using FCNN. On a real-world Weibo dataset, DFRec++ outperforms various existing techniques [14].

Create an end-to-end framework that includes a Multi-Social Graph Convolutional Network buddy recommendation model and numerous social networks to analyze potential user attributes (MSGCN). Using an improved graph convolution neural network augments the target user's representation with higher-order neighbours from numerous social networks. Graph fusion algorithms that modify and fuse the network's Laplace matrix are used to add social relationships. Finally, Bayesian theory transforms buddy suggestions into a sorting problem. In experiments, the proposed model outperforms existing strategies [15].

Propose an approach that learns users' implicit interest topics and topic weights, then uses a symmetric Jensen-Shannon divergence to calculate user interest topic similarity. The weighted link relationship similarity between users is then determined using this technique, which is used to investigate implicit link connections between users in local topic cliques. Combining user interest topic similarity and weighted link connection similarity yields a latent buddy recommendation list. The suggested method outperforms state-of-the-art latent buddy recommendation algorithms on real-world datasets in four independent evaluation criteria [16].

TextCNN (Text Convolution Neural Network) is used to create a text classification model in this approach. This article uses the probability distribution of users' text categories to discover Top-N pals. The F1 of the KBERT-CNN text classification model is 92.26 percent, which is much

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

higher than that of other text classification models. Text-based friend recommendation algorithms are more precise than content-based recommendation systems [17].

Growing co-friend connections and building a relationship chain via friend interaction are examples of this. This study suggests a method for future users to assess co-affinity with friends with shared tastes. The friend-suggested weighting approach is created in both virtual and real-world groups. Furthermore, the system is designed and implemented to evaluate a way of locating potential mates among users [19].

We combine past user input and social information (e.g., friends' reactions to the same items) into a covariance matrix that learns the complicated links between users, interacting objects, and friends' perspectives using a social-aware covariance function. A stochastic gradient descent strategy is used to fit the model. The suggested technique outperforms existing social recommendation algorithms and Gaussian process-based approaches on three real-world datasets [20].

In the above literature work, we have to find limitation

1. In existing system has not applied dimension reduction.
2. Author finds a solution using a hierarchical-graph-based similarity measurement framework, which gets less accuracy.
3. Existing systems have applied Text Convolution Neural Network.
4. In existing system has got less accuracy with data is not proper.

III. PROBLEM DOMAIN

The traditional collaborative filtering recommendation algorithm lacks accuracy and efficiency as this uses a formal method of filtering, which makes it inefficient to use alone. In terms of recommendations made by the collaborative filtering algorithm, it may be concluded that the algorithm needs many more improvements.

By implementing the traditional collaborative filtering recommendation algorithm, we get less accuracy, making it typical to use and inefficient to apply on huge datasets, i.e., Big Data. Dealing with big data, the less accuracy makes it inappropriate and less accurate. When applying this algorithm to the massive amount of data in real-world applications, the less accuracy will not be efficient for making recommendations to users. The number of available attributes is considered for extracting information to recommend friends to users, making the collaborative filtering recommendation algorithm inefficient. Also, the higher the number of attributes used to make recommendations, the higher the computing time and the higher the number of comparisons to be made. The overall dimensions for making a recommendation should be removed as per the requirement.

The k-means clustering applied previously with the collaborative filtering algorithm can be replaced by a different clustering technique. Some drawbacks can be seen in the k-means clustering technique, which may be overcome by replacing this clustering technique with the newer one. In the k-means clustering, the numbers of the clusters that should be made need to be defined at the start of the algorithm, which makes it inefficient to use if the numbers of the clusters are not properly defined.

One more thing to note is that the dimensionality of the given dataset should be less in number to lower the comparisons that will be made at the time of execution. The more the number of dimensions to evaluate the results, makes the accuracy lesser and requires the more time to make recommendations to the user. Hence, reducing the number of attributes or the dataset's dimensionality is a major task.

IV. PROPOSED METHOD

The problem observed in the previous algorithm can be removed by replacing the existing techniques with newer techniques. As during the last, the algorithm combines the K-means clustering technique with the PCA as a dimensionality reduction technique. Combining this technique in the collaborative filtering algorithm was a solution proposed earlier by the authors.

Here we have proposed a better clustering technique than the k-means clustering while keeping the PCA as earlier it was used. The hierarchical clustering can replace the k-means clustering as it is a better clustering technique to work on. The PCA will be used as the dimensionality reduction technique to decrease the dimensionality of the data.

Hierarchical clustering will provide better results than k-means clustering; as stated that in hierarchical clustering, there is no need to define the number of clusters at the beginning of the clustering. Defining the required number of clusters after applying the hierarchical clustering will make it feasible to break the clusters as per the dataset. But before applying the clustering technique to the dataset, the dataset should be improved. If the Input to the algorithm is

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

accurate, then the obtained output will be more efficient. So, to improve the input dataset, dimensionality reduction should be made, and to do this, the PCA have to be applied to the dataset.

In final words, we will apply the PCA to the dataset before giving it Input. After getting the principal components, these are given as Input to the hierarchical clustering. The collaborative filtering algorithm will first perform the PCA. After that, the hierarchical clustering is applied, and the final recommendations are made. Hence, the collaborative filtering algorithm can be improved, and the requests can be made accurate.

4.1 Algorithm For Proposed Approach: The proposed algorithm uses both techniques. The first one is the PCA, which will help reduce the dimensions of the given dataset. The second one is the clustering technique, which is hierarchical clustering. Here in our algorithm, we are applying the PCA at first because it will reduce data dimensions. After that, the hierarchical clustering will be performed on the obtained principal components. The working algorithm is as follows:

Step 1: Data collection - collect the friend related data like name, rating etc., in the form of a CSV file.
Step 2: Data pre-processing - perform manual data analysis and eliminate the less correlated feature to another feature.
Step 3: Perform PCA (principal component analysis) on the data and save the data into a CSV file.
Step 4: Define hierarchical clustering (agglomerative) model.
Step5: Train the hierarchical clustering (agglomerative) model on the data.
Step 6: Take the one user input and apply PCA on that.
Step 7: Perform the prediction in the Input. It gives the cluster id.
Step 8: Fetch all the friend details that belong to this cluster id and make a list. (This list is recommended friend list)

4.2 Flowchart of the Proposed Approach: Below, we have given the flowchart for the proposed approach, which will help in understanding the flow of the steps performed:

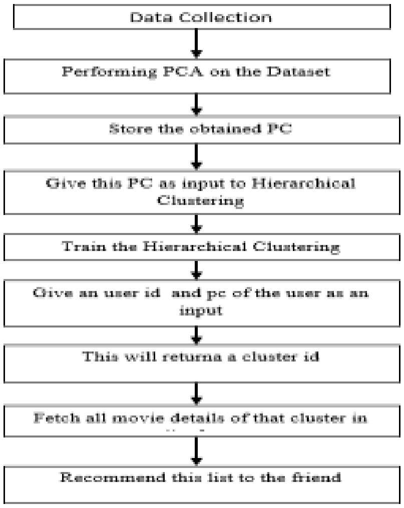
Data Collection
Performing PCA on the Dataset
Store the obtained PC
Give this PC as input to Hierarchical Clustering
Train the Hierarchical Clustering
Give an user id and pc of the user as an input
This will returna a cluster id
Fetch all movie details of that cluster in
Recommend this list to the friend

Fig.2 Flowchart for the proposed system

V. EXPERIMENTAL RESULTS AND EVALUATION

For analyzing the proposed approach, we have used the Kaggle dataset. The data about friends is taken from the Kaggle dataset, and the friend_likes pattern and user details are combined from the Kaggle dataset. The experiment was carried out to evaluate the accuracy of the recommendations produced by the algorithm proposed in our paper. The accuracy term is calculated in this experiment by which the comparison between the proposed and the existing algorithm can be made.

In the experimental section, we have used python language with the related library. We have used the Windows 10 operating system, 8GB RAM, and Intel Core i5 processor in the hardware section.

We are applying this data to the previous collaborative algorithm with PCA and k-means, and the results are obtained, so the accuracy of the previous algorithm is calculated.

Accuracy = ({Relevant Document} intersection {Retrieved Document} / {Relevant Document}) *100
Now the proposed algorithm with hierarchical clustering is taken for analysis. The collaborative filtering algorithm and pca and hierarchical clustering are analyzed over the same data. This algorithm's accuracy is compared with the existing algorithm.

The experiment increases the accuracy of the recommendations made by our proposed algorithm. The results are compared between both the algorithms using k-means clustering with PCA and

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

hierarchical clustering with PCA in terms of accuracy are shown in the following graph:

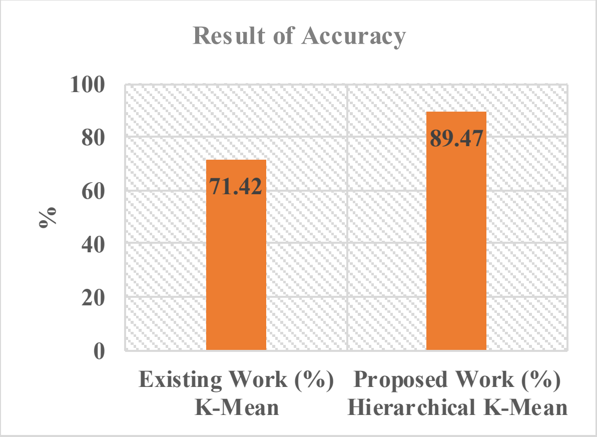

Fig. 3. Accuracy results for both the algorithms

Fig.3 concludes that the proposed hierarchical clustering works better than the previously used k-means clustering. The results in terms of the accuracy of the proposed algorithm are higher than the earlier clustering technique. So it is better to use the Hierarchical clustering with PCA on the collaborative filtering algorithm than the earlier one.

Table 1 Result between Existing and Proposed Work

	Existing Work (%) K-Mean	Proposed Work (%) Hierarchical K-Mean
Recall	71.47	79.166
Accuracy	78.25	81.25
Precision	67.5	90.9
F1_Score	84.38	86.95

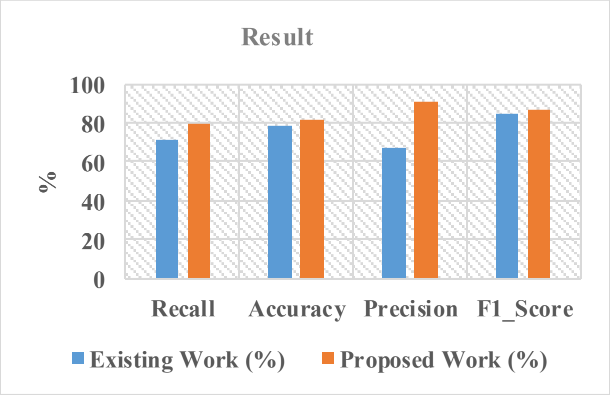

Figure 4 Results between Existing and Proposed Work

Table 1 and figure 4 examine the aftereffects of the trials performed; we can reason that our methodology is proficient in lessening the running time without giving up the proposal quality. This builds up that our technique is adaptable and may be utilized to manage significantly greater datasets.

VI. CONCLUSION AND FUTURE WORK

The proposed research work observes the recommendations made by the system to the user. The hierarchical clustering technique does the entire Work along with the PCA, by which the system's accuracy is evaluated. The system's accuracy is evaluated by the intersection of the recommended friends with the friend_likes made by the user for the earlier friends. The experiment shows better results than the earlier algorithms.

We can use other datasets to experiment in the future. The other parameters, apart from the accuracy, can be tested. Different clustering techniques may be applied to improve the algorithm. This Work required multidimensional data, and it's a limitation of our Work.

Reference

[1] Haruna K, Ismail MA, Damiasih D, Sutopo J, Herawan T. A collaborative approach for research paper recommender system. Plos One. 2017, 12(10): e0184516. https://doi.org/10.1371/journal.pone. 0184516 PMID: 28981512

[2] Rojas G, Garrido I. Toward a rapid development of social network-based recommender systems. IEEE Latin America Transactions. 2017, 15(4):753–759.

[3] Huang S, Zhang J, Wang L, Hua XS. Social Friend Recommendation Based on Multiple Network Correlation. IEEE Transactions on Multimedia. 2016, 18(2):287–299.

[4] Corbellini A, Mateos C, Godoy D, Zunino A, Schiaffino S. An architecture and platform for developing distributed recommendation algorithms on large-scale social networks. Journal of Information Science. 2015, 41(5):686–704.

[5] Fields B, Jacobson K, Rhodes C, Inverno M, Sanler M, Casey M. Analysis and Exploitation of Musician Social Networks for Recommendation and Discovery. IEEE Transactions on Multimedia. 2011, 13 (4):674–686

[6] Chamoso P, Rivas A, Rodrı' guez Sara, Bajo J. Relationship recommender system in a business and employment-oriented social network. Information Sciences. 2018, s 433–434:204–220.

[7] Guo G, Zhang J, Zhu F, Wang X. Factored similarity models with social trust for top-N friend recommendation. Knowledge-Based Systems. 2017, 122:17–25.

[8] Zhang Z, Liu H. Social recommendation model combining trust propagation and sequential behaviors. Applied Intelligence. 2015, 43(3):695–706.

[9] Maier C, Laumer S, Eckhardt A, Weitzel T. Giving too much social support: social overload on social networking sites. European Journal of Information

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.

Systems. 2015, 24(5):447–464.

[10] Lee S, Koubek RJ. The effects of usability and web design attributes on user preference for e-commerce web sites. Computers in Industry. 2010, 61(4):329–341.

[11] AndreasenT, Jensen PA, Nilsson JF, Paggio P, Pedersen BS, Thomsen HE. Content-based text querying with ontological descriptors. Data & Knowledge Engineering. 2004, 48(2):199–219

[12] Huang S, Zhang J, Wang L, Hua XS. Social Friend Recommendation Based on Multiple Network Correlation. IEEE Transactions on Multimedia. 2016, 18(2):287–299.

[13] B. Samir and N. El-Tazi, "Enhancing Multi-factor Friend Recommendation in Location-based Social Networks," 2020 International Conference on Data Mining Workshops (ICDMW), 2020, pp. 198-205, doi: 10.1109/ICDMW51313.2020.00036.

[14] J. Gong et al., "Hybrid Deep Neural Networks for Friend Recommendations in Edge Computing Environment," in IEEE Access, vol. 8, pp. 10693-10706, 2020, doi: 10.1109/ACCESS.2019.2958599.

[15] L. Chen, Y. Xie, Z. Zheng, H. Zheng and J. Xie, "Friend Recommendation Based on Multi-Social Graph Convolutional Network," in IEEE Access, vol. 8, pp. 43618-43629, 2020, doi: 10.1109/ACCESS.2020.2977407.

[16] L. Cui, J. Wu, D. Pi, P. Zhang and P. Kennedy, "Dual Implicit Mining-Based Latent Friend Recommendation," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 5, pp. 1663-1678, May 2020, doi: 10.1109/TSMC.2017.2777889.

[17] N. Pan, W. Yao and X. Li, "Friends Recommendation Based on KBERT-CNN Text Classification Model," 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1-6, doi: 10.1109/IJCNN52387.2021.9533618.

[18] T. -S. Chen and S. -W. Syu, "Friend Recommendation Based on Mobile Crowdsensing in Social Networks," 2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS), 2020, pp. 191-196, doi: 10.23919/APNOMS50412.2020.9236783.

[19] T. -S. Chen and S. -W. Syu, "Friend Recommendation Based on Mobile Crowdsensing in Social Networks," 2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS), 2020, pp. 191-196, doi: 10.23919/APNOMS50412.2020.9236783.

[20] D. Cao, X. He, L. Miao, G. Xiao, H. Chen and J. Xu, "Social-Enhanced Attentive Group Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 3, pp. 1195-1209, 1 March 2021, doi: 10.1109/TKDE.2019.2936475.

Authorized licensed use limited to: MIT-World Peace University. Downloaded on March 30,2023 at 05:02:16 UTC from IEEE Xplore. Restrictions apply.