Collaborative Item Recommendations Based on Friendship Strength in Social Network

N. Jamil, S. A. Mohd Noah, and M. Mohd

Abstract—Recommender Systems (RSs) are among the solutions in addressing the information overload problems. One of the RS main problem is cold start users where RSs do not have enough information to identify the user’s preferences and thus unable to recommend relevant items. One of the approaches to overcome such problem is through social relationship which can be extracted from social networks. Existing studies used friendship relation to find nearest neighbors combined with implicit data such as tweet content, posts, like or tags to identify user’s preference. This study proposed friendship strength through user’s interaction and item rating values to identify users’ preferences. User-based collaborative filtering methods are used in developing the system. Two datasets are synthetically obtained from MovieLens database for user’s rating information. Meanwhile Twitter users’ interactions data, was retrieved from Higgs Boson topic available from the Standford University. There are four phases of development namely user profile phase, friendship strength phase, user’s preference phase and items recommendation phase. The findings of the study show that there is an improvement in the proposed method compared to the baseline based on the recall (R), precision (P) and F score measures.

Index Terms—Collaborative filtering, recommender system, user interactions, social relationship, item recommendation.

I. INTRODUCTION

Social media has proven to have a huge impact on its users. Users can search for information without having to meet with others to get an idea of certain item. They can also share their opinions and views on an item with other users [1]. As a result, there will be a vast of online information that ultimately confuses users to find information that is accurate and relevant to them.

Recommender systems (RSs) is one of the solution in solving online information overload by providing appropriate information in the era of information explosion today [2], [3]. Amazon, Netflix, Movie Lens and Yahoo! Movies are among the e-commerce web sites that use recommender system [4]. Studies show that 20% of sales on Amazon came from a RSs that recommend products to their users based on their preferences [5]. Users need a good system in recommending items that are relevant to their interest [6].

In real world applications, RSs that rely on users’ ratings face the issues of cold start users. This is due to the fact that the rating matrices tend to be very sparse, as users typically provide ratings for only a small fraction of the items. Thus, the RSs unable to accurately model the users’ preferences. Previous works to overcome such a problem is by exploiting implicit rating based on user’s behavior such as browsing, clicking, opening or closing certain web sites, time spent on reading certain product information and even commenting. Although implicit ratings can be collected constantly, one cannot be sure whether the user behavior is correctly interpreted.

With the advent of social media, social recommender systems were introduced to solve the problem of cold start users [7], [8]. It aims to get relevant items to the users from their online social friends. However, not all social relationships within the social network can solve this problem because there are also relationships that are based on relatives and they might not represent the same interests [9], [10].

Research by Seo et al. [11] evaluate the strength of users’ friendship in social networks. The strength of friendships is generated through their interactions with friends. In their study, rating information was obtained from the content of the user’s tweets, posts and tags. Thus, the item’s rating value is based on the estimation and not based on real ratings given by users on items.

This study proposed a solution in the selection of nearest neighbors through similarity of friendship strength between users and friends on social networks. In addition, it also uses the items rating information given by the users. The proposed approach uses the collaborative filtering approach, which is the most popular and a proven approach in commercial recommender system [3], [4], [12]-[13]. This technique assumes that users with similar interests will like the same items without the need to know the content of the items.

The paper is organized into five sections. The first part introduces an overview of RSs. The second part discuss about related work done in RSs and some new approaches used to increase the system performance. The third part presents the proposed RSs based on friendship strength and the flow or this approach. The fourth part show the datasets, evaluation and result of the proposed approach compared to conventional approach. The last part is the conclusion.

II. RELATED WORKS

A. Recommender System

RSs help users make choices of the items based on their preferences. Each system has different analytical capabilities in identifying the nearest neighbors and user preferences.

RSs are usually classified into the following categories according to the approach of recommendations [14]:

Manuscript received October 11, 2019; revised January 19, 2020.
N. Jamil is with the Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia (e-mail: author@nrim.go.jp).
doi: 10.18178/ijmlc.2020.10.3.954 437
• Content-based recommendation: in which a user is recommended items that are similar to those that the user liked in the past [6].
• Collaborative recommendation: where a user is recommended items that other user with similar tastes liked in the past.
• Hybrid approaches: combination of content-based and collaborative methods [15].

Collaborative recommender systems have received greater attention and widespread industrial use today. The main idea of this approach is to exploit information about the past behavior or the opinions of an existing user community for predicting which items the current user of the system will most probably like or to be interested in.

In this study, we focus on user based collaborative recommender technique which has three main task. First task is to find nearest neighbors by measuring the similarity between users. Many correlation metrics can be used to compute similarity between users, for example, Pearson correlation, Euclidean distance, cosine similarity and others [18]. Nearest neighbors can be obtained by sorting the similarity between users through two approaches, K Nearest Neighbors (KNN) or by setting the threshold value. By finding the nearest neighbors with similarity preference, the system then will predict the unrated item for user as a second task. The list of recommended items based on Top-N items will be produce as a final task. Fig. 1 show the task of user collaborative technique in recommender system.

Fig. 1. Task of user collaborative technique in recommender system.

The major problems in recommender systems is that the number of initially available ratings is relatively small. In such cases, it becomes more difficult to apply traditional collaborative filtering models [17], [18]. Thus resulting the cold-start problem [6]. It affects the system’s performance because the users’ preferences are unknown and the system fails to recommend relevant items to them. One of the approaches to overcome this problem is by introducing a social recommender system that use online social relationships [19].

To address the problem of cold start users, the social recommender system was introduced in 1997 by Kautz et al. [19]. It has successfully increased the accuracy of recommendations by 20% [6] which is in line with the concept of homophily. This concept stated that individuals who have similar interests are more likely to develop friendships than those who do not have similarities [20]. Conventional recommender systems do not take into account social interactions in social networks whereas in reality, an individual will ask individuals with similar priorities to get an idea of an item they are interested in [21].

B. Social Recommender System

Most individuals have online social connections. Relationships between users and their friends are varied and they are usually related to each other [22]. Such social information is deem meaningful to be exploited in order to enhance recommendation. Thus, social recommender system was introduced in extracting information about user’s preference from their social relationship.

By definition, the social recommender system is any recommendation that uses online social relations as input [6] [23]. It assumes, each user is connected to each other when they have a social connection [24]. Based on this assumption, the user’s preferences may be similar to or influenced by their related peers. It was found that 66% of users would ask their social friends for decision-making while 88% of users would click on links provided by their friends and 78% of users would trust their friends’ opinions more than advertisement or people they did not recognize [6]. They also tend to recommend similar items. The social recommender system has two inputs which are the user’s social information and the rating information as shown in Fig. 2.

Sun et al. [23] proposed the Recommender System based on Social Network known as the RSboSN. This study clustered a group of users based on the same preferences. Item recommendations will be provided by the system to users based on items rating by friends in the same group only. The results show that this system is better than the existing system.

The IRN (Individual Relationship Network) was introduced by Ma et al. [22], combines matrix factoring methods with neighborhood models. This approach identifies the social network of each user or item by considering the complexity of relationships between users such as size, consistency and data imbalance. The results show that the recommendations are more accurate in most cases. The results also indicate that social relationships between users cannot be considered equal.

A new method for integrating social relationship information with social tags was studied by Ma et al. [13]. User-based collaborative filtering approach was used in this study to address the sparseness and cold start problem. The dataset was taken from Last.fm and MovieLens. The results show that by using social relationship and tags also increased the accuracy of the recommended items.
C. Friendship Strength Based on User’s Interactions

A friend relationship is a direct relationship between the user and the other users. It is widely found on social media like Facebook, Twitter, and Instagram. This relationship exists when a user follows other users (followers) or they are followed by other users (followers).

Friends on social networks have powerful influences on recommending items. This is proved by a study conducted by et al. [11], Ardissono et al. [25], Vosecky et al. [26], and Li & Xiong [27] that claimed system performance improved in recommending relevant items when using friendship relations. This is because individuals trust their friends more than strangers [28].

Studies on the strength of friendship have been carried out by Seo et al. [11] which proposed an approach for measuring the intimacy between users and friends in social networks. This measure is known as the friendship strength. User’s preferences were gathered from user’s generated content, user’s social relationship information and user’s interaction. The results of the study show that a combination of user’s interaction and user’s generated topics can help recommender system to identify user’s preferences.

Research by Ardissono et al. [25] find the impact of friend’s influence on determining user preferences. This study used Yelp dataset which provide data for friendship, rating, and social tag information. The result of the analysis show that the filtered friendship relationship can predicts the user’s preferences in recommending relevant items. Vosecky et al. [26] proposed the Collaborative Personalized Twitter Search (CPTS) to model user preferences accurately, alleviate data sparseness and address cold start issues. To achieve this goal, a collaborative user model was developed to exploit user’s relationships in social networks. User’s interactions are taken into account to gain the strength of each friend to the user.

The analysis by Li & Xiong [27] show that the number of mentions and comments were very effective in calculating influence between friends, while the number of followers contributes to the global influence of the entire social network. In this study, two definitions of influence in social networks are MISI (Microscopic Social Influence) and MASI (Macroscopic Social Influence) was. MISI is looking for a legitimate relationship by putting a lot of weight on the relationship, while MASI shows the strength of each user’s influence on social networks. The main goal of this study is to improve the accuracy of recommendations in media networks that involve direct user interaction such as Facebook, Twitter and LinkedIn.

Friendship between users must have a strong bond of trust. Based on a strong and long term friendship, an individual will share his or her views with their friends and they will be more trustworthy than those of strangers [9]. Granovetter [29] defines friendship strength as a combination of friendship, emotion (earnestness), intimacy (trustworthiness) and reciprocal relationships.

There are two types of friendship bonds, which are strong bonds and weak bonds. A strong bond of friendship involves the people user trust and their friends come from the same circle. While the bond of friendship is considered weak when information about friends is not known in the friendship circle. There are four characteristics of friendship strength, which are two ways communication, recent communication, frequency of interaction and having at least one mutual friend [30]. This has been proven in the study of Seo et al. [11] where they used friendship strength through interactions between users in recommending relevant items to users. The three interaction used are as follows:

1) Interaction frequency

Frequency of interaction between two users indicate the strength of the relationship between user and friends [31]. This is because user with similar interests or preferences will actively interact and share information [32]. The more user interacts with their friends, the more likely they have the same preferences as a discussion topic [33].

2) Last interaction

The last interaction is information about the recent interaction between user and friends in social network. This is to determine if the interaction is still running until recent day. Last interactions are important as they indicate that the relationships are still going on and they are still relevant to recommend items based on current topics [34].

3) Friendship longevity

Longevity is the period when a friendship develops between two users in a social network. Friendship periods are viewed to ensure that they have a mature relationship. Relationships that have been around for a long time indicate that they share the same interests and topics of discussion [35]. The duration of the friendship will be calculated starting when the user begins to interact with his or her friends in social networks to the present.

III. PROPOSED RECOMMENDER SYSTEM BASED ON FRIENDSHIP STRENGTH

The main purpose of the study is to develop a collaborative model of recommendation based on friendship strength through interaction among friends in enhancing the accuracy and effectiveness of the recommender system. The proposed approach is based on the user-based collaborative filtering approach. Social contact information and user’s interaction with friends on social networks are used to minimize the search for neighbors.

Three characteristics of interaction and friendship were defined. The characteristics are the frequency of interaction, the last interaction and the longevity of friendship that has been established between users on social networks. Fig. 3 shows the proposed approach.

Fig. 4 shows the concept of the proposed system, where the friend relationships between user \( u \) and user \( f \) are based on the interactions between them. The weight of the friendship (longevity) and interaction (frequency and last interaction) are calculated to produce the friendship strength. The values of friendship strength represent the similarity between users. This similarity is used in choosing the nearest neighbors of the target user in order to identify the preference of user \( u \) and user \( f \). To find user’s preference item, system will look at item rating value. Rating similarity weight \( \bar{r}_u \) is used to identify which items are preferred by the user. Then it will calculate rating prediction value for item \( i \) by user \( u \). Finally the
recommendations are generated based on this value. The list of item recommendation is based on top-k.

The computation of the last interaction weight and longevity of friendship between user $u$ and user $f$ is depends on concept of functions $L(u)$, $r(u, f)$ and $l(u, f)$ as shown in Fig. 5.

The last interaction of user $u$ with user $f$ is the interval time from last interaction to current time. The last interaction weight $W_l(u, f)$ is the ratio of the last interaction $r(u, f)$ to the period of time the user begins to register on social media $L(u)$. The calculation of the last interaction weights is as in equation (2) below.

$$W_l(u, f) = \frac{r(u, f)}{L(u)}$$  \hspace{1cm} (2)

The similarity of friendship strength is calculated based on the three main variables:

1) Interaction frequency ($W_q(u, f)$)

The frequency of interaction between user $u$ and user $f$ is measured as $W_q(u, f)$ and is calculated by normalizing the frequency of interaction $q(u, f)$ between the range 0 to 1. It uses the log function as shown in equation (1)

$$W_q(u, f) = \log \left(1 + q(u, f)\right) \text{if } q(u, f) < 10$$
$$W_q(u, f) = \frac{10}{\log_q(10)} \text{if } q(u, f) \geq 10$$  \hspace{1cm} (1)

2) Last interaction ($W_l(u, f)$)

3) Friendship longevity ($W_l(u, f)$)

The third phase is to identify user preference so that relevant items can be suggested to the user. User items rating are used to get their preferences. Average rating of user $u$ ($r_u$) and user $f$ ($r_f$) is calculated. The item recommendation phase is the final phase of the Friendship Strength RS. Once the average value of user $u$’s rating and their friends are calculated, the rating of item $i$ for user $u$ ($p_{ui}$) is can be predicted using equation (5).

$$p_{ui} = r_u + \sum_{f \in \mathcal{F}_u} \sum_{i \in \mathcal{I}_f} \frac{f \in \mathcal{F}_u}{\mathcal{I}_f} \alpha f \beta f \gamma f \\sim \text{Sim}_f(u, f)$$  \hspace{1cm} (5)

IV. RESULTS AND EVALUATION

A. Datasets Preparation

We used two types of data: interactions data between Twitter users and their movie rating data. The Twitter users’ interactions dataset was obtained through the Higgs Boson topic from Stanford University (https://snap.stanford.edu/data/higgs-twitter.html). The data
was retrieved between 1st till 7th July 2012. It contains information about functions reply, mention, retweet, relationship among Twitter users and timing of activity occurring in the form of timestamp during the discovery of Higgs Boson. The total interactions data is 563,069 and was divided into three functions, which are 354,930 retweets, 36,902 replies and 171,237 mentions. It also contained 1,048,576 friend relationships between users. In this experiment we consider 270 users with 400 relationships and 967 interactions between them.

The movie ratings dataset is based on the Movie Lens dataset (http://grouplens.org/datasets/movielens/20m/). It contains 27,278 movies from various genres such as comedy, documentary, action, and horror. In this study, the rating data of 270 users were synthetically generated and involved 39,655 rating values on the selected movies.

**B. Evaluation**

Evaluation is conducted through experiments. The data is divided into two parts: 80% for training set and 20% for testing. The proposed approach (Friendship Strength RS) is compared against the standard collaborative recommendation approach (Conventional RS). To obtain the nearest neighbors, three values of friendship strength similarity ($Sim_{f}(u, f)$) was set between 0.5 to 0.9. It is based on the study of Sun et al. [23]. $N$ number of neighbors were selected in identifying user’s preference. In order to test the effectiveness of the recommendation list, three values of $k$ recommended items are considered.

The relevance of the predicted rating value was compared to the actual rating value given by the user $u$. The predicted rating value was considered relevant if the value is higher than or equal to 3.5.

**C. Result**

The results of the experiment used recall, precision and F-score metric. Recall ($R$) is the probability that relevant item available in the list of recommended item. The measurement is based on equation (6).

$$R = \frac{\text{Total of relevant item recommended} (N_{rs})}{\text{Total of relevant item} (N_r)} \quad (6)$$

Precision ($P$) is the probability of recommended item is selected by user $u$ as relevant. Equation (7) shows the calculation of precision ($P$).

$$P = \frac{\text{Total of relevant item recommended} (N_{rs})}{\text{Total of item recommended} (N_s)} \quad (7)$$

The average balance between precision weight ($P$) and recall ($R$) is calculated using the Score $F$ metric as shown in equation (8).

$$F = 2 \times \frac{(PR)}{(P + R)} \quad (8)$$

The results from Table I and Fig. 6 shows that Friendship Strength RS has higher recall, precision and F score values than Conventional RS. In Friendship Strength RS, the highest recall value is 0.9842 when friendship strength similarity ($Sim_{f}(u, f)$) is at or greater than 0.83 and $k = 20$. This represents a 96% increase in the proposed approach compared to conventional approach. An increase of 34.6% also occurred in the precision values. The highest precision values for both approaches were when $Sim_{f}(u, f)$ values is at or greater than 0.83 and $k = 5$. However, the proposed approach has better precision value which is 0.5667 as compared to 0.2203 for the conventional approach. The F score results show that the approach to find the nearest neighbor through friendship strength similarity increased by 59.7%.

<table>
<thead>
<tr>
<th>$Sim_{f}(u, f)$</th>
<th>$k$</th>
<th>Friendship Strength RS</th>
<th>Conventional RS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>&gt;= 0.5</td>
<td>5</td>
<td>0.4643</td>
<td>0.3751</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.6644</td>
<td>0.4643</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.6314</td>
<td>0.3252</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7193</td>
<td>0.4268</td>
</tr>
<tr>
<td>&gt;= 0.66</td>
<td>10</td>
<td>0.8078</td>
<td>0.4771</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.7603</td>
<td>0.3557</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.6833</td>
<td>0.5667</td>
</tr>
<tr>
<td>&gt;= 0.83</td>
<td>10</td>
<td>0.6614</td>
<td>0.3833</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.9482</td>
<td>0.3917</td>
</tr>
</tbody>
</table>
V. CONCLUSION AND FUTURE WORKS

The finding of the study clearly shows that the proposed approach is better than the conventional approach in finding the nearest neighbor. The higher similarity of friendship strength between two users, the higher the relevance of the recommended items. This proves that friendship strength through interactions between friends have an influence on the users’ choice of an item. This approach can be improved in the future. Interaction between users can be broken down into positive and negative interactions. Besides that, friend relationship can take into account the reciprocal relations between users in determining the nearest neighbors.

Other potential future works include the addressing the elements of serendipity in recommendation. Exploiting friend’s relationship is expected to improve elements of serendipity whereby items suggested are considered novel, relevant and unexpected [36].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

N. Jamil, S. A. Mohd Noah, and M. Mohd conceived of the presented idea. N. Jamil; S. A. Mohd Noah, designed and performed the experiments, analysed data and co-wrote the paper. S. A. Mohd Noah, and M. Mohd involved in supervising the work and assist in the technical details.

ACKNOWLEDGMENT

The authors graciously acknowledge the Ministry of Education (MOE), Malaysia for awarding the scholarship under the Federal Training Awards (FTA) scheme to complete this work. This study was supported by the Universiti Kebangsaan Malaysia grant: DCP-2017-007/4.

REFERENCES


Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

**Normasлина Ямил** has received her M.IT in 2011 and B.IT in 2000 from Universiti Kebangsaan Malaysia. She is mathematics and computer lecturer at Politeknik Banting Selangor, Malaysia. Her current research work is focused on recommender systems. She is currently completing her PhD in computer science from Universiti Kebangsaan Malaysia.

**Shahrul Azman Mohd Noah** received his MSc and PhD degrees in information studies from Sheffield University. He is a professor in the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia and currently heads the Knowledge Technology Research Group. His current research work is focused on semantic computing with special emphasis on information retrieval, ontology and recommender systems.

**Masnizah Mohd** is an associate professor at the Universiti Kebangsaan Malaysia (UKM). She received her PhD in computer and information sciences from the University of Strathclyde in 2010, the M.IT in 2002 and B.IT in 1999 in information science from UKM. Her current research work is focused on Information retrieval and natural language processing.