IMPROVING FRIENDS RECOMMENDATION USING FP-GROWTH ALGORITHM IN SOCIAL TAGGING SYSTEM

1Kabir Umar & 2Usman Bukar Usman

1Department of software engineering, Faculty of Computer Science and Information Technology, Bayero University Kano, Nigeria.
2Department of Computer Science, Faculty of Science, Mai Idriss Alooma Polytechnic Geidam, Yobe State, Nigeria.

*Corresponding Author’s Email: ubu1700016.msc@buk.edu.ng

ABSTRACT

Social tagging system are web-based sites that store user’s keywords called tags, continue to receive significant consideration in academic environment it became an interesting research topic, give good support for users to tag resources, communicate with friends as well as friend recommendation, due to increasingly acceptance of Web2.0. There is a constant growth in the number of users using social tagging. Friend recommendation is one of the most important aspects for overcoming information overloading problem and helps users to make choice. In this work we proposed improving friend recommendation using, to address the difficulty of tag ambiguities. The technique is consisted of the following stages: Apply FP-Growth algorithm to discovery the frequent events sets among the users, construct the combine trust graph, and then apply ant colony optimization Algorithm to joint trust graph to compute the optimal friend recommended through repetition. The results of our experiments on a Delicious dataset, using our model, precision improved in the range 0.1 % to 1.1 %, recall increased greatly in the range 21.85 % to 50.84 % and $F_2$ increased in the range 31.31 % to 50.85 % approximately. When compared to other methods, the outcome demonstrates a significant improvement. For friend commendation in a social tagging system, the future approach is to apply various data mining methods, large scale datasets, and community detection.

Keywords: Social Tagging, FP-Growth, Friends Recommendation, Ant Colony Optimization

INTRODUCTION

According to a report by the International Data Corporation (IDC, 2012), by 2020, the total amount of global data would be 22 times that in 2011. In the last few years’ people assisted with the continuous growth of online social networking sites, both in terms of services and size. With the rapid growth of the social tagging (STS) site, the social interaction is becoming a hot topic. The social tagging system (STS) has grown in both size and functionality. STS enables users to define keywords (tags) to characterize resources of interest to them, assisting in the organization and sharing of these resources with other network users (Farooq et al., 2007). Online social tagging platforms such as Amazon, Facebook, Twitter, WeChat, Flickr, delicious.com are a web-based site that store users’ keywords called tags, give good support for individuals to tag resources, such as videos and images as well as to communicate with friends. Carullo et al. (2014) stated that Social Networks Services (SNSs) are making a great effort to enlarge their importance and popularity Even if the strategy is based on as-yet-undisclosed algorithms, some of them provide utility services to promote friends. However, the human elements that influence how a user communicates with others are complicated. The overall amount of information in the network environment is expanding quickly as a result of considerable advancements in information technology, which has given rise to Internet resources and e-commerce. However, despite enjoying the abundance of information resources, internet users are confronted with the dimension disaster dilemma. Because of the duplicated and chaotic nature of enormous data, it has become increasingly difficult for humans to extract usable content from it quickly and efficiently (Peng et al., 2018). There is a constant increase in the number of users using social sites. Friend Recommendation is one of the most important aspects of Social networking sites. Recomender system has become an effective method to solve information overloading and helps users to make decisions (Ren et al., 2018). The purpose of this research is to overcome the challenges of information overloading on social networking sites as researchers gave their various opinions. The performance of a...
friend’s recommendation system in the social networking system in the past received little attention as a subject of research, whereby most researches focus on the uses of a single Algorithm. Personalized online social friendship is different from traditional offline friendships. Traditional friends, mostly have physical interactions frequently because they live near each other, or work or play in the same area while the online friend recommendation should be based on individuals’ interests where individuals have a similar interest in the same or different area based on designed Algorithm on social networking platform. Therefore, the genesis of this research focuses on combining two different Algorithms which will drive by observations in the disparity of the friend recommendation system in the social tagging system as a result of non-implementation based on FP-Growth Algorithm and ant colony Algorithm. Similarly, there is a need to strive towards achieving optimal solutions in a friend recommendation system in a social tagging system. Social network sites, however, this cannot be attainable unless problems such as; information overloading and cold start are addressed. Hence, this research will seek to recommend ways to improve the quality of friends generated from our proposed method. While Peda and Sharma (2015) also proposed a technique for friend recommendation based on pheromone based tags recommendation (PbTR) and ant colony optimization. The PbTR Algorithm does not address the problem of tag ambiguities. This ambiguity problem affects the quality of truth graph produced by the PbTR Algorithms during the first phase of Peda and Sharma technique. By replacing PbTR algorithm with FP-Growth algorithm in the first phase to classify the user activities will resolve the tag ambiguity problem and also improve the quality of the resulting truth graph and performance (Hasan et al., 2015).

This work aims at improving friends’ recommendation in social tagging system by incorporating FP-Growth algorithm with Ant colony optimization.

**RECOMMENDER SYSTEM**

![Diagram](image)

**Fig. 1**

Recommendation system is algorithms which is basically used to make perfect predictions based off user preferences or needs (Fessahaye et al., 2019). Recommendation systems for social networking sites are a relatively new area of study, as social people are increasingly interested in online social networking sites like as Facebook, Twitter, Flickr, LinkedIn, and other similar sites (Hasan et al., 2015). The basic purpose of the recommender system is to filter out unwanted information based on the preferences of the users, allowing them to find products that they are interested in (Wang et al., 2019). The system master will suggest things that are similar to those that the user has previously desired. The recommendation system is critical in assisting individuals in extracting useful information from large amounts of data. The five ways to recommender systems are content-based filtering, collaborative filtering (CF), hybrid approach, and knowledge-based recommender systems (Dhruv et al., 2019). The collaborative filtering (CF) recommender system is a filtering technique that uses information filtering based on the user's past appraisals. It is anticipated that a user with similar beliefs or interests to other users will enjoy the things that these individuals prefer. Memory-based and model-based recommendations are the two types of CF-based recommendations (Linden et al., 2003). The memory-based recommendation assumes that similar users have similar ratings for the same item or product, and that the user rates similar items with similar ratings (Chai et al., 2019). The model-based recommendation focuses on the use of training data to build a model for scoring criteria; this model is based on data mining algorithms and user ratings of unrated products (Koren et al., 2009). Knowledge-based systems promote things based on domain knowledge about how various item qualities satisfy users' preferences and, ultimately, how significant that item is to the user (Ricci et al., 2011). The hybrid system, which combines CB and CF methodologies, aims to exploit CB's advantages to compensate for CF's shortcomings. CF techniques have an issue with recommending new items, which means they can't forecast items with no evaluations.

**FRIENDS RECOMMENDATION IN SOCIAL TAGGING**

People have contributed to the continued expansion of online social networking platforms in recent years. With the rapid growth of social networking sites, social relationships have become a hot topic. Users can generate keywords (tags) to describe resources that are of interest to them in social tagging systems, which helps to organize and share these resources with other users in the network (Farooq et al., 2007). Overloading issues have arisen as a result of the ever-increasing amount of content and people that occupy social tagging systems. Overwhelming social interaction is a concern. Overwhelming social contact (Guy et al. 2013; Simon 1971; Shehu 2017). Because of the social ramifications of “friending,” recommending friends on social tagging sites is worth investigating. It differs from typical recommendations of books, movies, restaurants, and so on. However, the human elements that influence how a user communicates with others are
complicated. Because of the rapid advancement of information technology, which has given rise to Internet resources and e-commerce, the total amount of data in the network environment is rapidly expanding. According to Wu et al., (2015), friend recommendation in social networking sites may be divided into two categories: predicting friends based on social networks, such as friends of friends; and suggesting based on users’ interests, which is generated using text information in recommendation systems. Following friends of friends on social media can lead to recommendations for people you may already know. As a result, friend recommendations based on interests have practical value because they can forecast users with significant interests in common as potential friends. In our thesis, we underlined the importance of predicting user interest on a social platform. The summary of related work on friend recommendation systems in both social tagging and social networking is shown in Table 1

Table 1 RELATED WORK

<table>
<thead>
<tr>
<th>SN</th>
<th>AUTHORS</th>
<th>YEAR</th>
<th>METHODOLOGY</th>
<th>WEEKNESS</th>
<th>STRENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lian-ju and Hai-yan</td>
<td>2014</td>
<td>SimRank and ant colony optimization</td>
<td>Sequential ordering problem</td>
<td>Optimal solution, improve the efficiency of decisions</td>
</tr>
<tr>
<td>2</td>
<td>Manca et al.</td>
<td>2014</td>
<td>analyzing frequency of each used tag (graph analysis)</td>
<td>The bookmarks that a user receives are serendipitous.</td>
<td>Solve problem of complexity and scalability</td>
</tr>
<tr>
<td>3</td>
<td>Peda &amp; Sharma</td>
<td>2015</td>
<td>ant-based friends recommendation (AbFR)</td>
<td>Tag ambiguity problem</td>
<td>Expand network of friends</td>
</tr>
<tr>
<td>4</td>
<td>Wu et al.</td>
<td>2015</td>
<td>User similarity Graph (FRUG)</td>
<td>Community detection</td>
<td>Resolve the issues of tag redundancy and data scarcity.</td>
</tr>
<tr>
<td>5</td>
<td>Hasan et al.</td>
<td>2015</td>
<td>FP-Growth algorithm</td>
<td>Low accuracy</td>
<td>User’s online behavior</td>
</tr>
<tr>
<td>6</td>
<td>Naruchitparames et al.</td>
<td>2011</td>
<td>Genetic Algorithms and Network Topology</td>
<td>The lack of completeness of data is largely to blame for lower effectiveness in social-based approaches.</td>
<td>Examine the difficulty of discovering how and why social networks create links.</td>
</tr>
</tbody>
</table>
Table 1 shows the topics that are most closely related to our work. Hasan et al., (2015) used the FP-Growth algorithm to present a novel buddy recommendation framework based on online user behavior. The framework has five steps: extracting sub-networks, determining activity frequency, identifying common behavior, identifying uncommon behavior within common behavior, and ultimately recommending friends. However, according to Narvekar and Syed (2015), the Frequent Pattern (FP) growth algorithm employs a divide-and-conquer strategy. The frequent item sets are calculated using FP's tree of frequent patterns. When FP growth is compared to an Apriori method, FP is significantly more efficient. However, the FP's drawback is that it generates a large number of conditional FP-Trees. FP-Growth also performs badly with lengthy pattern data sets and consumes more memory space, according to Suman et al., (2012). Peda and Sharma (2015) suggested a PbTR algorithm that uses tags assigned to a specific resource by the target user and other users as input. The target user for whom friends are suggested may be an existing user who has already contributed tags to the resource or a new user who has not assigned tags to the resource. The cumulative trust graph of all users is used as feedback by the friend recommendation algorithm in recommending friends. Last but not least, Lian-ju and Hai-yan (2014) introduced a novel hybrid technique that calculates the similarity between any two nodes using a simrank algorithm based on the relationship between users and uses the previous Simrank result as the input ACO to produce the optimal friend recommendation. However, according to Farsania et al., (2013), SimRank is unable to generate a collection if the number of linkages between resources is low.

PROPOSED MODEL

Fig 2: The proposed model of friends recommendation system

MODULSES DESCRIPTION

User Interest Module

The FP-Growth algorithm is used to classify the user actions in the first stage of our model, which is utilized to compute the target user and other user activities. We measure the similarities between this user and other users of his social tagging who use the tags occurring in the rule's consequence for each extracted association rule A B whose antecedent applies to an active user. To determine the degree of similarity between two users, u1 and u2, each are represented by a binary vector including all of their tags, and the angle cosines between the two vectors are computed as follows:

\[
sim(u1, u2) = \frac{u1 \cdot u2}{||u1|| \cdot ||u2||}
\]

eqn. 1

Where u is user

Because of its linear complexity, the cosine similarity provides good findings at a reasonable processing cost (Cattuto et al., 2008; Koerner et al., 2010; Kaveri & Maheswari 2016; Beldjoudi et al., 2017).

Truth Graph Module

The second stage of our proposed algorithm was used to construct the combined truth graph of users. The result obtained from the user interest module as an input to this stage. The directed trust graph TG = (V, E) is initially created with root node as the target user tu and another set of vertices V equivalent to recommenders AU = {au1, au2, ..., aun} whose activities of the target user. E is the set of edges connecting Activities of target user au.

Step 1: Evaporation takes place on all the edges of the trust graph TG which exists for target user au using the following equation:

\[
\tau_{au}(t) = -(1 - \rho) * \tau_{au}(t-1)
\]

where (\(\rho = 0.01\)) is a modest constant value indicating the trust-activities weight evaporation rate on the edge (au, I between time t–1 and t to avoid its infinite buildup.

Step 2: If vi does not exist, a new vertex is created to the graph for each recommender rui.

Step 3: If an edge between vertex (v1) and v1 exists, then

\[
\tau_{au,i}(t) = \tau_{au,i}(t - +\Delta Q)
\]

Otherwise an edge with trust-pheromone weight \(\tau_{au}(t) = \Delta Q\) is added between vertex \(v_{au}\) and \(v_{i}\) where \(\Delta Q\) is a small positive real number.

ANT COLONY OPTIMIZATION MODULE

Final stage of the Algorithm where used to apply ant colony optimization Algorithm to recommend friend to the target user.
In real life when two users know each other, they know how much to trust each other. However, for the users who do not know each other, trust may be calculated from the data contained in the path connecting them in the trust graph. Although single level trust graph is maintained for each target user, but as the time passes, trust graphs of different target users when combined, forms a complex trust graph. Friends can not only be chosen from single level trust graph, but also from combined trust graph by computing optimal trust path as defined in the following technique:

**Step 1:** Let $S$=target user node

**Step 2:** Make $m$ ants (where $m$ is the number of outgoing links from $S$)

**Step 3:** $m$ ants should be placed on $S$.

**Step 4:** Place the $k$th ant’s beginning node $S$ in $tabuk(S)$, where $k = 1..m$. The $tabuk$ list is a data structure related with the $k$th ant that maintains track of the nodes that ant has already visited and prevents it from accessing them again.

**Step 5:** For all ants, build tour $Tk(t)$ up to level $lev$ using the probability function specified in equation (6). At time $t$, the probability that ant $k$ at node $I$ will travel to node $j$ is given as:

$$p_{ij}^{k}(t) = \frac{\tau_{ij}(t)}{\sum_{l \in N_{i}^{k}} \tau_{il}(t)} + r$$

if $j \in N_{i}^{k}$ and $i \in N_{j}^{k}$

$$p_{ij}^{k}(t) = 0$$

otherwise

(6)

If $ij(t)$ is the amount of trust that node $I$ has in node $j$ at time $t$.

$$j \in N_{i}^{k} = \{N – tabus\}$$

The number of nodes is $N$.

$tabuk$ is a dynamically increasing vector that contains the nodes that ant $k$ has already visited.

A random function is denoted by the letter $r$.

$lev$ denotes the depth of the trust graph till which an ant has traverse

**Step 6:** Sort the tour lengths computed in Step 5 in descending order and extract the top $m$ optimal tours for each ant.

**Step 7:** Friends can be recommended. For example, the nodes at $lev = 1$ are recommended as direct friends and nodes at higher values of $lev$ are the inferred friends.

**EXPERIMENTAL PLANNING**

In our experimental planning we conducted empirical evaluation of our proposed approach which ware implemented using FP-Growth and ant colony optimization algorithm, we adopt experimental design from (Peda and Sharma 2015), and we used key activities that are evaluation and performed our experiments on delicious dataset, a social bookmarking website.

**A. DATASET**

On the Delicious datasets, a popular social tagging resource, we test our proposed recommendation system. hetrec2011-delicious-2k is a Delicious datasets that was released in the literature during the HetRec2011 workshop.

**B. EXPERIMENTAL DESIGN**

In our research, we planned to make an empirical evaluation, our experimental design intended to perform two different experiments which are normally divided into test and training set. We will compare our proposed approach with pheromone based tag recommendation proposed by (Peda and Sharma 2015), enables the comparison between the approach applied to the same delicious dataset. We have to evaluate the performance of a system to evaluate the recommendations, by measuring the precision, recall, and F-measure of the system. we will get the following main conclusion: The comparison of the value of P, R, and F1, the number of potential friends recommended to a user is 2,5 and 7.

**A. EVALUATION MATRICS**

Precision, recall, and f-measure are the most popular metrics for evaluating friend recommendation system. Precision is a measure of accuracy or correctness and recall is a measure of completeness. The precision score of 100% indicates that every recommendation retrieved was relevant. A recall score of 100% indicates that all relevant recommendations were retrieved. The formulas are described in section 2.4.2 above adopted from (Wu et al., 2015; Peda and Sharma 2015).

**B. EXPERIMENTAL DESIGN**

We aimed to conduct an empirical evaluation in our study, and our experimental design called for two separate trials, which are typically divided into a test and a training set. We will compare our proposed strategy to (Peda and Sharma 2015)'s pheromone-based tag recommendation, which allows us to compare the two approaches using the same yummy datasets. We must measure the precision, recall, and F-measure of a system to evaluate its performance in order to evaluate its recommendations. The following is the primary conclusion: When the values of P, R, and F1 are compared, the number of possible buddies suggested to a user is 2,5 and 7.

**EVALUATION STRATEGY**

Friends can be recommended. For example, the nodes at $lev = 1$ are recommended as direct friends and nodes at higher values of $lev$ are the inferred friends.
divided its associated collection of resources into three parts and then chose one portion to be eliminated from each evaluation as a test set. We performed this method three times, each time selecting a different test set from the separated halves.

RESULT AND DISCUSSION
We compare our work (IFRAST) with that of Peda and Sharma (2015), Ant-based friends' recommendation in social tagging systems (PbTR), and discussion in this part. Table 2: the results

A. EXPERIMENTAL RESULT
Experimental results are shown in Table 4.1 The performance measured in terms of precision, recall, and F-Measure. The total number of prospective friends recommended to a user is 2 and 5 and 7 individually.

B. RESULT OBTAINED
In our experimental result shows significant improvement compared with the result of Peda and Sharma (2015), the below graphs are data representation of our proposed model contain their precision, recall and F-Measure analysis.

Tables 2
Note: Here, P stands for Precision in %, R stands for Recall in % and F represents F2 metric.

Results in Tables 2 show that recommendations generated through IFRAST have significant enhancement as compared to PbTR values of N, where N is the number of potential friends recommended to one user. The following observations were made when the two approaches were tested for different values of N ranging from 1 to 10 as represented in Table 4.1: IFRAST vs. PbTR: Using IFRAST, precision improved in the range 0.1 % to 1.1 %, recall increased in the range 21.85 % to 50.84 % and F2 increased in the range 31.31 % to 50.85 % approximately.

CONCLUSION
We present a unique friend recommendation system called improving friend recommendation using FP-growth Algorithm in social tagging system (IFRAST) that may be applied on social tagging sites in this paper. Based on the user's online actions and friend optimization in a social tagging system, the algorithm gives a novel friend recommended solution for virtual community operators. The precision improved in the range of 0.1 percent to 1.1 percent, recall increased in the range of 21.85 percent to 50.84 percent, and F2 increased in the range of 31.31 percent to 50.85 percent using our model, according to the results of our studies on a Delicious dataset. When compared to
the PbTR and the ant colony optimization algorithm, the experimental findings suggest that our proposed strategy is significantly better. We've described a new recommendation method based on the FP-Growth Algorithm, which yielded excellent results and a strong recommendation. In the future, more advanced research into the ways to be used to rate the recommendations in social tagging is required for the arranging buddy recommendation. Other features, such as location, user blood group, and so on, should be used in future studies. Future study will focus on other data mining methods, big scale datasets such as Facebook and Twitter, and community detection for buddy referral in social tagging systems.

REFERENCES


