AN EXTENDED HYBRID RECOMMENDER SYSTEM BASED ON ASSOCIATION RULES MINING IN DISCUSSION GROUPS

Mahnaz Ebrahimi[1] and Mandana Goudarzian[2]

[1,2]Department of Computer Engineering and Information Technology,
Amirkabir University of Technology, Tehran, Iran
[1]m_ebrahimi@aut.ac.ir
[2]m.goudarzian@aut.ac.ir

ABSTRACT

Social groups in the form of different discussion forums are proliferating rapidly. Most of these forums have been created to exchange and share members' knowledge in various domains. Members in these groups may need to use and retrieve other members' knowledge. Therefore, recommender systems are one of the techniques which can be employed in order to extract knowledge based on the members' needs and favorites. It is noteworthy that not only the users' comments and posts can have valuable information, but also there are some other valuable information which can be obtained from social data; moreover, it could be extracted from relations and interactions among users. Hence, association rules mining techniques are one of the techniques which can be applied in order to extract more implicit data as input to the recommender system. Our objective in this study is to improve the performance of a hybrid recommender system by defining new hybrid rules. In this regard, for the first time, we have defined new hybrid rules by considering both users and posts' content data. Each of the defined rules has been examined on an asynchronous discussion group in this study. In addition, the impact of the defined rules on the precision and recall values of the recommender system has been examined. We found that according to this impact, a classification of the defined rules can be considered and a number of weights can be assigned to each rule based on their impact and usability in the specific domain or application. It is noteworthy that the results of the experiments have been promising.

KEYWORDS

Asynchronous discussion group, Association rules mining, Hybrid recommender system

1. INTRODUCTION

Nowadays, the importance of using social groups and sharing information in discussion groups, forums and other similar mediums is evident to everyone. However, there are still a number of issues in these domains which need to be studied further. Huge volume of information and extracting the desired knowledge are among those issues. Recommender systems are one of the techniques which can assist researcher to find and suggest the suitable information to users during the recommendation process. It should be mentioned that not only the users' comments and posts can have valuable information, but also there are some other valuable information which can be obtained from monitoring the relations and interactions among users as social data. Therefore,

association rules mining techniques are one of the techniques which can be applied in order to discover implicit relations between explicit data by defining new rules. Discovered data can be applied as an input for the recommender systems to empower them in order to produce recommendations. In fact, by increasing the input data, generated recommendations will be more accurate and reliable.

In this study, we have applied association rules mining technique inside a hybrid recommender system in an asynchronous discussion group. In fact, we have attempted to make new rules and also to make better recommendations in discussion groups as a digital social media. For the first time, we have considered both user and item data to define new association rules. In the other words, we have defined new hybrid rules in an asynchronous discussion group. It is noteworthy that using hybrid association rules instead of the rules which only consider the user or item data can improve the accuracy of the recommender system. Based on our knowledge, until now, there have been no such rules discovered by data mining techniques in discussion groups' domain.

In our previous study [1], a hybrid recommender system has been proposed which combined the collaborative and content based techniques. This system, in its collaborative section, applies association rule mining technique in order to find the similar users. Hence, a number of rules have been defined and evaluated. In this study, we have tried to extend capability of the proposed recommender system by applying association rules mining technique more effectively.

1.1. Motivation

The precision of recommender system is depended to its input data which can be gathered explicitly or implicitly. By increasing amount of input data and also accuracy of them, more valid and trustworthy recommendations will be generated.

Providing accurate input data is a main challenge in the recommender system domain. This challenge is a potent motivator to find suitable technique that is capable of discovering implicit information about users and contents inside discussion groups.

1.2. Contribution

In response to the aforementioned challenge, in this study, some new hybrid association rules are defined in order to extract implicit data about users and contents from available data in discussion groups. The contribution of this study is to improve accuracy of the hybrid recommender system by defining new and advanced association rules by considering both user and post's contents.

Overall, the main goal of this study is to improve the accuracy of the generated recommendations by adding a new part to the proposed hybrid recommender system. In this part, hybrid association rules will be generated. Therefore, a new architecture of hybrid recommender system has been proposed. To validate the proposed architecture, several experimental rules has been defined and tested on the Mata Filter dataset.

2. BACKGROUND THEORIES

In this section, the related subjects to this study will be explained. At first, a short introduction concerning the recommender systems is presented and then, the association rules mining technique is briefly explained.

2.1. Recommender Systems

Recommender systems are effective systems for guiding users through the large number of possible options to achieve their favorites so that the process is personalized for that specific user [1]. These systems have applications in different areas. One of the areas in which they can have remarkable impact on the process improvement is e-learning and especially collaborative learning and its application in asynchronous discussion groups as a developmental tool in this field of study.

Collaborative learning is actually the collaboration among the groups of users in order to solve problems or to exchange knowledge. This kind of learning has a high impact on the knowledge construction. In the learning process, it also combines the rules of social relations and information processing techniques. It is noteworthy that an interactive environment is required in this regard [2]. Asynchronous discussion groups are the mostly used tool in e-learning systems which support collaborative learning. These groups have an important role in developing scientific discussions and construction of human knowledge.

It should be noted that the data extracted from the posts in the discussion groups can have a large volume of knowledge [3]. Extracting this knowledge and using that as a new resource in learning process is an important and valuable issue in e-learning systems.

Discussion groups have a simple nature and that is why they could have been able to attract many users and application domains. Yet, there are still a number of limitations and challenges related to these groups which prevent them to become an effective and efficient tool in supporting collaborative learning. As a result of weak structure as well as huge volume of information in discussion groups, it is difficult and time consuming for users to find related and suitable information in such groups. As the search engines are based on the string matching patterns, they cannot be suitable tools for knowledge extraction in order to discover the semantically related information.

As a solution for the aforementioned problems, researchers are working on several techniques in order to provide personalized information for users. Personalization can be considered as the ability to provide suitable information and services for individuals based on the knowledge of their preferences and behaviors. Personalization process has three main steps: 1- To recognize the needs and favorites of the user and making his profile based on this knowledge, 2- To propose items and information based on user profile and post content and 3- To evaluate the quality and usefulness of the personalized information and recommendations based on the user feedbacks and other criteria [10].

In discussion groups, recommender systems search for the information related to the user query regarding his profile and content of posts. They carry out this process by identifying the users who are similar to the current user, or by identifying the contents which are similar to that user's query. It is noteworthy that the user feedbacks will result in quality improvement and recommendations preciseness.

Recommender systems are suitable options to find relevant and useful information in discussion groups. Hitherto, there have had been several researches regarding this issue so that the researchers have tried to find proper techniques in order to retrieve and recommend related items to the users. Collaborative filtering techniques are focused on the similarities among the users in the group such as [4] while content-based filtering methods have focused more on the similarities among the post contents [5].

It should be mentioned that collaborative filtering techniques are not sufficient to find relevant information in a group. These techniques only consider the priorities and features of the user and identify the similar users based on this criterion and not based on their posts. On the other hand, content-based filtering techniques concentrate on content information and do not consider the favorites and features of the users. Therefore, to solve the problems related to each of these techniques, a comprehensive method must be used so that it will be able to remove the weak points and improve the functionality of these systems. Hybrid recommender systems try to fulfil this goal by combing the existing filtering techniques.

2.2. Association Rules Mining Technique

The idea of mining association rules comes from the analysis of market-basket data which has rules such as "*If a customer buy item X, then he also may buy item Y*", or "*If a patient's disease is X, then he will have disease Y as well*".

Association rule is a phrase such as $X \rightarrow Y$ (if X then Y) that X and Y are the item sets in database D and $X \cap Y = \emptyset$. it is noteworthy that D is a database of transactions that each transaction $T \subseteq D$ is a set of items.

To extract rules from a data file, association rules mining technique primarily finds the frequent items, and then makes the association rules based on them. As the number of the defined rules may exceed, a kind of filtering must be applied in order to choose the most effective and useful rules in accordance with the criteria of evaluating rules. Hence, according to different types of measuring rules, the best rules will be selected finally.

A part of the transaction T that supports the item set X in database D, is called support (X) and defined as (1) and (2) [6]:

$$supp(X) = |\{ T \in D \mid X \subseteq T \} | \, / \, |D| \qquad (1)$$

$$supp(X \rightarrow Y) = supp(X \cup Y) \qquad (2)$$

Probability or confidence of a rule is defined as the number of transactions including Y in addition to X considering the total number of transactions including X. It means that the confidence of a rule can be defined as conditional probability $p(Y \subseteq T \mid X \subseteq T)$. Confidence of a rule is defined as (3) [6]:

$$conf(X \rightarrow Y) = supp(X \cup Y) \, / \, supp(X) \qquad (3)$$

The important issue in finding the association rules is to select rules which have the support and confidence values more than the minimum value (threshold) defined by a user.

3. RELATED WORKS

Hitherto, there have been a few studies in the area of recommender systems for asynchronous discussion groups. In this section, we will discuss about the most important ones. One of the main issues in the recommender systems is to specify authority of a member in the group based on his/her knowledge and social data so that it can help the recommender systems to generate more accurate recommendation. In the recommender system presented in [7], authority of each user in the group has been calculated based on specific and static data. This data includes: the total number of sent items by user, the number of other users' clicks on the user's sent messages, and the number of items sent by user which are considered as good messages. Users with more

authority score have been considered as reliable users so that their information can be used in the recommendation process for other group members. In this paper, the authority calculation formula for all the group members has been fixed and cannot be changed based on the accessible data or user's activity in the group.

Another main issue in the recommendation process is to find similar users based on their preferences and social relationships and activities in the group. In the recommender systems presented in [8] and [9], the users who are related and similar to an active user (i.e. the user for whom a recommendation is generated) have only been extracted based on their contribution in the common posts. New posts have been recommended based on these similarities, users' favorites and the related posts. It is noteworthy that in this method, the other useful implicit data which can be gained by a user contribution have not been considered.

The system presented in [4] is a recommender system which has identified similar users based on the limited number of rules. These rules calculate the similarities based on the limited number of implicit and explicit information. This information includes the sent posts or comments by the user and also the scores given by the user to the others' posts.

Due to the great number of posts and comments as well as large number of members and their social data, finding reliable users and required contents will be a complex and tedious task for members. The hybrid recommender system presented in [1] considers both the user and item information for generation recommendations. This system is a combination of collaborative filtering and content-based filtering methods. In the collaborative section, it uses the association rules mining technique in order to find the similar users while in the content-based section, it uses the WSD (Word Sense Disambiguation) method [1] in order to find the related and similar semantic posts based on Leacock–Chodorow algorithm [1]. By defining a number of rules, this system discovers implicit data in the group. It should be noted that as the number of these rules increases, precision and performance of the recommender system will be increased as well.

4. RESEARCH METHODOLOGY

In the previous study [1], a hybrid recommender system was presented. This system was based on these two techniques: collaborative filtering and content-based filtering. In the collaborative filtering, to find similar users, the association rules mining technique has been used based on different parameters. The most important feature of this system is that in order to recommend posts to the users, it considers the similarity between users as well as the similarity among the contents.

In the collaborative filtering part of the system, several simple and extended rules have been defined based on implicit users' data which is shown in tables 1 and 2. The confidence and support values for each rule have been specified for all the existing users in the training set. The users with the confidence and support values more than the threshold have been considered as the most similar users.

Table 1. Basic rules extracted from discussion groups [1]

Rule	Description
Rule 1	if u_i contributes to p_i $\rightarrow$ u_a will contribute to p_i
Rule 2	if u_i likes pi $\rightarrow$ u_a will contribute to p_i
Rule 3	if u_i contributes to p_i $\rightarrow$ ua will like p_i
Rule 4	if u_i likes p_i $\rightarrow$ u_a will like p_i

Table 2. Extended rules extracted from discussion groups [1]

Rule	Description
Rule 5	if u_a has the same rating style with u_i $\rightarrow$ u_a will like, or contribute to the posts which u_i contributes
Rule 6	if u_i and u_a add same tags for p_i $\rightarrow$ u_a will like, or contribute to the posts which u_i contributes
Rule 7	if u_a has tagged the posts which u_i contributes $\rightarrow$ u_a will like, or contribute to the posts which u_i contributes
Rule 8	if u_a has tagged the posts with the same subjects $\rightarrow$ u_a will contribute to the posts containing those subjects

In the system evaluation, only the first four rules have been considered and tested; however in fact, in order to improve the functionality of the recommender system, more rules must be considered. These new rules will increase the system performance and accuracy by identifying more similar users.

To improve the quality and accuracy of the generated recommendations of the proposed system, an extended architecture of pervious system is presented in Figure 1. In this architecture, a new part has been added to the recommender system. This part called "Hybrid Rule Engine" which generates hybrid rules as input of "Recommendation Generator" part. User data, posts' content data as well as the generated rules in "Association Rule Generator" part are input of rule engine. In the previous system, we have applied hybrid concept to combine output of collaborative and content-based filtering sections. However in the new proposed system, we have also applied hybrid idea in the first part of recommendation process i.e. manipulating user and content data. The main goal of defining hybrid rules is: 1- To extract similarity between users and contents, and 2- To discover knowledge regarding users and content such as finding reliable and expert users.
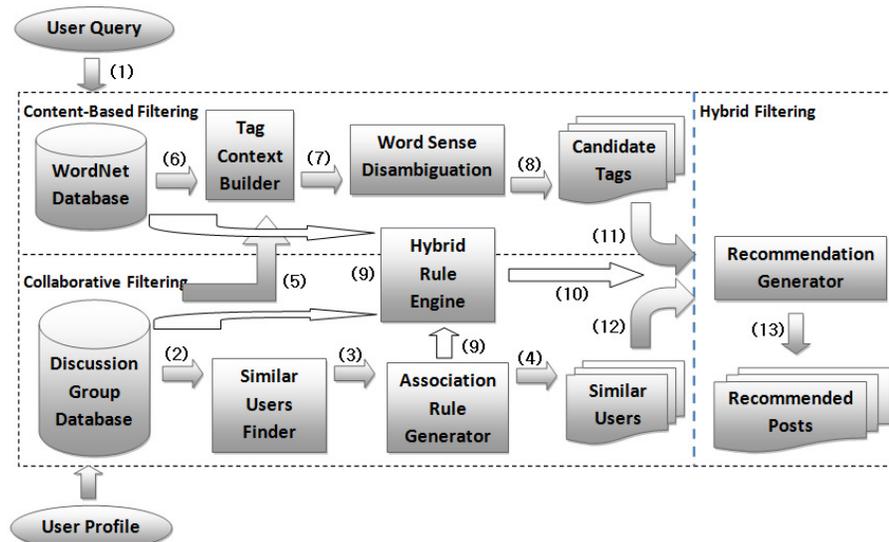

Figure 1. Extended architecture of proposed hybrid recommender system

To validate the proposed recommender system, we have experimentally defined several hybrid association rules in the discussion group.

In this study, to extract more rules, we have considered more implicit data in the group. These rules have been extracted according to the users' performance and activity in the group based on their contribution in the posts and comments. These rules are presented in Table 3.

A number of these rules have been defined in order to find expert users in the group while some of the other rules have considered the users' style of writing posts and comments. It is noteworthy that it can be beneficial to find expert users in the group so that we are able to use their information and experience. In the recommendation process, we can trust these expert users. As a case in point, we can suggest those posts which the expert users have sent or have made comments on.

In the previous study, the rules (basic and extended) have been only defined based on the user information, whereas in this study, new rules are defined based on both the user information and post contents. The important feature of this study is that the implicit information regarding the post content and its relation with the users can be discovered and used in the recommendation process. The rules #2, #3, and #4 are some samples of these rules which will be discussed further. In other words, as the recommended system as well as the defined rules is hybrid, the performance of the recommender system will be improved.

As it is shown in Figure 1, recommendation process will be started by entering the user query. First, the recommender system finds the users who are similar to the active user based on their common contributed posts and also their favorites ("Similar User Finder" unit). Then, system will find more similar users according to defined association rules in the "Association Rule Generator" unit. In the content-based part, system finds posts with similar tags matched with user query ("Tag Context Builder" unit), and then semantic similarity between contents will be calculated in the "Word Sense Disambiguation" unit. User data, post contents and also generated association rules in collaborative filtering part will be the input of the "Hybrid Rule Engine". New hybrid rules will be generated in this unit. Finally a list of similar users and contents will be sent to "Recommendation Generation" unit to produce the recommendations.

Table 3. New rules extracted from discussion groups

Rule	Description
Rule #1	if u_a is an expert user $\rightarrow$ u_a will properly be a reliable user
Rule #2, #3	if u_a posts short messages $\rightarrow$ u_a will properly be interested in short messages
Rule #4	$R\,\#2 \cup R\,\#3$
Rule #5	if u_a is a regular user $\rightarrow$ u_a will properly be a reliable user
Rule #6	if u_a doesn't start his/her activity immediately after join date $\rightarrow$ u_a will properly be a reliable user
Rule #7	$R\,\#1 \cap R\,\#5 - R\,\#6$

4.1. Extended Rules

4.1.1. Rule#1

"The user who has more knowledge or experience (the expert user) is probably more reliable."
In this rule, the user's knowledge or experience is defined based on the following criterion:

1) The number of comments made by the user as the answer to the questions or as an illustration of an idea.
   Based on this criterion, it will be determined if the user has sufficient knowledge or experience regarding the discussed issues or not.
2) The speed of answering or giving an idea compared with the other users. In this case, the quicker user is the one who makes the first or second comments.

Using this rule, users with more knowledge will be extracted. In the recommendation process, the posts and comments in which these users have made contribution in (with similar and related content) will be recommended to the active user.
To determine which user has been the high/low inquirer and also concerning which user has been the high/low respondent, the average of the questions count and answers count will be calculated, respectively. If the question/answer counts have been less than/equal the related average amount, it means the user has been the low respondent/inquirer, respectively and vice versa. The same method is applied to find the users which have been active in answering or not. In case the calculated amount for a user is more than the average of the comments in which he has been the first or second responder, he will be considered as the one who is active in answering and vice versa.

4.1.2. Rule#2 and Rule#3

"The user, who posts short messages, is probably interested in short messages."

In this rule, it is supposed that the user who usually posts short messages, most frequently is interested in answers which are also short in characters. In this method, the calculation of the post length is based on the following methods.

1) The total number of comments on a post (Rule #2)
2) The total number of the characters of a post's comments (Rule #3)

In Rule #2, for the user who is interested in short messages, we will choose the post which its number of comments is less than/equal the average value and vice versa.
In Rule #3, if the calculated total number of characters for comments of a post is less than/equal the average value, that post/comment will be considered as short and vice versa. After determining the length of the posts and comments of the user, the number of his short and long length comments and posts will be calculated. If the number of short messages is more, he will be considered as the one who is interested in short length posts and short length posts will be recommended to him.

4.1.3. Rule#4

This rule is actually the combination of rules #3 and #2. It means that in calculation of a post length, both the total number of comments and the sum of total length of the comments will be considered and the total length of a post will be determined based on that.

4.1.4. Rule#5

"The user, who has regularly been active in the group, is probably more reliable in recommendation process compared to the other users."

In the previous study [1], the users have been classified in five groups. This classification was defined by considering the user behavior and his activity type in the group. That classification is as follows:

1) Regular User: The user who regularly posts or makes comments in the group. These users are known as the knowledge creators.
2) Casual User: The user who is infrequently active in the group.
3) Regular favorite maker user: The user whose activity does not send posts or makes comments in the group, but he regularly expresses his opinion regarding the posts and comments of other users.
4) Casual favorite maker user: The user who infrequently expresses his opinion concerning the posts and comments of other users in the group.
5) Passive User: The user who has no special activity in the group.

We have used the aforementioned classification in order to extract the regular users. We assume that the user, who is considered as regular, has more knowledge and is more reliable to be recommended.

In the previous study, this classification was only employed to define some groups of users to show the hybrid recommender system capability in making recommendations to all kind of defined users by considering both the user and content data. In this study, the regular users have been considered as one of the reliable input data of the recommender system.

In this method, we have considered an 'Interval Indicator' which will obtain the number of '1' if the time interval between comments is less than or equal to 3 and in case that the interval between comments is more than or equal to 4 it will obtain the '0'. We have also considered a 'Number of Comments Indicator' that will be '0' if the number of comments is '0' and in case the number of comments is less than or equal to 10, it will be '1' and if it is more than or equal to 11, it will be set as '0'. It is noteworthy that these thresholds have been considered experimentally. Finally, we have considered to multiply the derived numbers for the 'Interval Indicator' and the 'Number of Comments Indicator' as the weight. To make decision concerning the user type of activity in the group, we calculate if the total number of '1's in the calculated weights are more/less than the calculated '0's. If the numbers of '1's have been equal or more than the number of '0's, the user will be considered as 'Regular'. If the number of '0's have been more than the number of '1's, then the user will be considered as 'Casual'. It is noteworthy that the whole period has been considered a month, so the number of the days in that month will be considered as the whole period number of days. In case the number of '0's are more than '1's and also the difference between the number of '1's and '0's is more than 2/3 of the whole period number, the user will be considered as 'Passive'.

4.1.5. Rule#6

"The user, whose join date in the group is far from the date of beginning his activity in the group, is probably a reliable user in the recommendation process."

This claim is based on the assumption that the user, who starts his activity more lately than his join date in the system, is more conservative. This kind of user probably first tries to increase his knowledge by studying the previous post and searching in the group, and then will start answering and sending comments.

To implement this rule, the minimum of the time interval between the users join date of the user and the time when he sends his first post/comment/favorite will be calculated. To decide regarding which user has been 'Quick' and which one has been 'Late', we calculated the average of the calculated minimum amounts and check if any of these calculated minimum numerical values are less/greater than that average. In case it has been less than/equal to the average, the user will be considered as 'Quick'. Similarly, if that numerical value is more than the average, the user will be considered as 'Late' in activity.

4.1.6. Rule#7

This rule will calculate the overall authority of a user in the group based on the rules #1, #5, and #6. This rule is defined as (4):

$$R\#7 = R\#1 \cap R\#5 - R\#6 \qquad\qquad (4)$$

It means the users who are in the list of extracted users by the use of rules #1 and #5, and are not in the list of the users regarding rule #6, will be considered as the reliable users.
Rules #1 and #5 considers users who are more active and regularly participate in discussions but rule #6 considers users who are less active by considering his activity start time comparing to the time of his join date. The reason why rule #6 has been omitted from rule #7 is that in spite of the rule proposed in rule #6, and also the results of the experiments, the user who starts his activity later than his join date, is probably less expert. Therefore, he cannot be a reliable user. The goal of this rule is to have a closer look at the rule #6.

5. SYSTEM EVALUATION

To test the defined rules, the experiments have been carried out in three time periods for similar users as it was done in [1] for MetaFilter dataset (http://stuff.metafilter.com/infodump).

 Each time period is one month so that it includes 20 days for training and 10 days for testing. In the experiments, 100 users extracted from the dataset and 20 of them selected as test users to generate recommendations for them. The selected users are the same as those in [1] who are selected based on the classification in section IV of that paper. In each time period, each user searches a special query in the group. By checking the user information and posts, the hybrid recommender system recommends the related posts to that user.

To conduct the experiments, each of the rules, which are defined in this study, has been tested separately for each of the test users and the results have been compared with the results of the experiments of previous study [1]. The results are presented in tables 4, 5, and 6.

Experiments in this study are same as previous study [1], but in this work new unit is added to the system called "Hybrid Rule Engine" to generate more association rules to empower collaborative filtering part of system. The goal of these experiments is to show effectiveness of association rules in finding more implicit data about users and posts and also improving accuracy of recommender system. As it is shown in Figure 1, experiment steps could be explained as follow:

1) User enters a query.
2) In the first part, "Similar Users Finder'' and "Association Rule Generator" units extract similar users according to basic association rules (based on users' common contributed posts and their favourites) as shown in Table 1. List of similar users is the output of this part.

3) In the second part, "Tag Context Builder" and "Word Sense Disambiguation" units extract senses of user's query tag by the Word Sense Disambiguation algorithm. Output of this part is list of similar and candidate tags of user's query.
4) In the third part, "Hybrid Rule Engine" new hybrid rule based on the user data, post data and defined basic association rules in the first part, will be generated. These rules are shown in Table 3. More similar users will be extracted in this part.
5) In the final part, "Recommendation Generation" unit combines output of previous parts to recommend related and useful posts.

The evaluation metrics adopted in this study are Precision and Recall, which are standard and common metrics in the evaluation of recommender systems. The Precision metric calculates the recommender system's ability to recommend only relevant items among a set of irrelevant and relevant items, while Recall metric shows the ability of the recommender system to recommend all useful and relevant items [1]. These metric are defined according to the confusion matrix shown in Table 7 by (5) and (6) [1]:

$$Precision = \frac{a}{a + b} \tag{5}$$

$$Recall = \frac{a}{a + c} \tag{6}$$

Where meaning of a, b, c values are shown in Table 7.

Table 4. Average of Precision, Recall and F values of proposed hybrid recommender system [1] and new extracted rules in first experiment for MetaFilter dataset

Technique	Experiment 1		
	Precision (%)	Recall (%)	F
Hybrid [1]	33.72	67.41	0.44
Rule #1	23	77.91	0.35
Rule #2	42.01	47.38	0.44
Rule #3	43.21	50.74	0.46
Rule #4	49.29	43.24	0.46
Rule #5	27.41	68.66	0.39
Rule #6	28.95	72.41	0.41
Rule #7	30.18	68.66	0.41

Table 5. Average of Precision, Recall and F values of proposed hybrid recommender system [1] and new extracted rules in second experiment for MetaFilter dataset

Technique	Experiment 2		
	Precision (%)	Recall (%)	F
Hybrid [1]	28.51	70.83	0.40
Rule #1	15.9	75.83	0.26
Rule #2	21.8	25	0.23
Rule #3	22.5	25	0.23
Rule #4	21.65	22.5	0.22
Rule #5	22.88	70.83	0.34

Technique	Experiment 2		
	Precision (%)	Recall (%)	F
Rule #6	19.84	70.83	0.3
Rule #7	23.82	70.83	0.35

Table 6. Average of Precision, Recall and F values of proposed hybrid recommender system [1] and new extracted rules in third experiment for MetaFilter dataset

Technique	Experiment 3		
	Precision (%)	Recall (%)	F
Hybrid [1]	21.42	71.66	0.32
Rule #1	17.58	78.33	0.28
Rule #2	14.81	31.66	0.2
Rule #3	9.18	21.66	0.12
Rule #4	9.71	21.66	0.13
Rule #5	18.38	74.16	0.29
Rule #6	18.1	75.83	0.29
Rule #7	21.89	74.16	0.33

Table 7. Confusion matrix

Predicted/Actual	Relevant	Irrelevant
Recommended	a	B
Not Recommended	c	D

6. RESULT ANALYSIS

In this section, the results of this study will be discussed and analyzed. It will also be stated how each of the defined rules have positive or negative impacts on the results comparing to the previous study.

In the tables 4, 5, and 6, the precision and recall values for each rule are compared with the similar values in the previous recommender system (Hybrid section). It can be seen that each of the new rules has influenced the precision, recall, or both of them. In some cases, the precision and recall values have increased while in others, they have decreased comparing with their values on the previous study. The impact of each rule on the precision and recall values can be predicted and justified based on the nature and type of the rule. As a case in point, rule#1 recommends the posts based on the expert user who has more knowledge (that is specified based on the number of sent comments and the speed of sending comments), and will have an impact on the recall value. The reason for the impact is the fact that this rule recommends the posts based on the user contribution which a number of these posts may be of those the active user (the user who is supposed to receive the recommendation) has contributed in. it is noteworthy that this rule can have a negative impact on the precision value as there may be some posts that have not been expected by the user. In this case, the precision value will be decreased. In other words, by applying this rule, recommendable posts will be identified based on trustable expert users rather than the content of the posts. So, probability of contribution of active user in recommended posts in future will be decreased. Rules #5, #6, and #7 have similar behavior, because in both of these rules, user's authority has been considered rather than the content of the recommended posts. In

Rule #5, the contribution type of user is a criterion for user expertise. So, the active users that contribute regularly in discussions are valid and trustworthy users. Rule #6 defines a relationship between user's join date and date of the beginning of his activity in the group. Based on this relationship, the user who starts his activity more lately than his join date in the group, must be somebody who first read others' posts and comments, gets enough information and finally begins to have activity, so he must be a more reliable person comparing somebody who begins to have activity without getting enough information of the previous activities and other issues. Rule #7 considers both rules #1 and #5 and ignores Rule #6.

Some of the other rules have positive impact on the precision value. For example, rule #3 recommends the posts based on the user's writing type of the post (the length of the sent comments, user's favorite to short/high length comments). In this rule, posts are not selected based on the contribution of other users and are just selected based on the post length; thus it has no considerable impact on the recall and it impacts on the precision value. Rules #2 and #4 have the same effect as Rule #3, because they consider content and features of the posts and comments rather than users contribution in finding recommendable posts.

By this means, a classification can be made for the rules based on their impact on the precision and recall values. To improve performance of the rules, we can combine those rules which have the similar impact on these values.
To have a dynamic use of the defined rules in different application domains, a number of conditions and limitations can be defined for the recommender system. For example, in the domains that the precision value of the recommender system is more important, the rules which lead to higher precision values can be applied. The same scenario can be applied when the precision value has a higher priority than proposing all the desired items (higher recall value). Conversely, in the system that proposing all or most of the items is needed, it is more appropriate to use those rules which have impact on the recall value. In other words, regarding the requirements for each application area, we can give weights to the rules. It means that we can give higher weight to the rules which are more significant or effective compared to the other rules.

As it is shown in the tables 4, 5, and 6, the precision and recall values for several of the rules in the second and third time periods have lesser values compared to the first time period. The reason of this issue could be the lesser contribution of the similar users in the second and third time periods compared to the first time. In other words, the number of posts and comments of the users have been minimal in these two periods; thus the available information and the input data for the rules are minimal. As a result, the implicit data that was expected to be extracted from the defined rules have been decreased.

The major advantage of the proposed system and defined association rules is the effectiveness of the rules in empowering collaborative filtering part of hybrid recommender system to discover more implicit data about user and also post's content to generate precise and useful recommendations. Another advantage of this work would be classification and usage of the defined rules based of their effectiveness on precision and recall of recommender system.

6. CONCLUSION AND FUTURE WORK

The main purpose of this study is to improve the accuracy of the generated recommendations of proposed hybrid recommender system in [1] by adding a hybrid rule engine part to the pervious system. Responsibility of this engine is to generate hybrid association rules from user and posts' content data to empower the input data of recommender system. The main goal of this engine is to find similarity among users and among contents and also discover more knowledge about user

and contents in the discussion group. To validate the new proposed hybrid recommender system, some hybrid rules have been defined experimentally. Despite of the previous study which has only considered the user information in defining rules, in this study, both the users and posts information has been considered in order to define new rules. In fact in this study, the defined rules for the hybrid system are hybrid as well. Additionally, the impact of the defined rules on the precision and recall values have been considered and the result revealed that some of these rules have positive impact on the precision while the others have only positive impact on the recall value. In this study, it can be concluded that we can have a dynamic use of these rules. It means in the domains, proposing all the desired items have more priority over the rules with more impact on the recall value which can be used or it can have a higher weight.

To present a more precise study and increase the system precision, we can increase the training time period to analyze the user behavior in the future studies. This cannot be accomplished until a robust state is achieved. On the other hand, to analyze the effectiveness of each rule, weights can be defined so that their values could be changed based on different parameters.

REFERENCES

[1]   Ahmad. A, Kardan & Mahnaz, Ebrahimi.,(2013) "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups", *Elsevier Information Sciences*, vol. 219, pp. 93-110.

[2]   Tammy, Schellens & Martin, Valcke, (2005) "Collaborative learning in asynchronous discussion groups: what about the impact on cognitive processing?", *Computers in Human Behavior*, vol. 21, pp. 957–975.

[3]   Yanyan, Li. Mingkai, Dong & Ronghuai, Huang, (2008) "Semantic organization of online discussion transcripts for active collaborative learning", *8th Int. Conf. IEEE Advanced Learning Technologies*, pp. 756–760.

[4]   Fabian, Abel. Ig Ibert, Bittencourt. Evandro, Costa. Nicola, Henze. Daniel, Krause & Julita, Vassileva, (2010) "Recommendations in online discussion forums for e-learning systems", *IEEE Transactions on Learning Technologies*, vol. 32, pp. 165–176.

[5]   Osmar.R. Zaı̈ane, (2002)"Building a recommender agent for e-learning systems", *Proc. Int. Conf. Computers in Education*, pp. 55–59.

[6]   Jochen, Hipp. Ulrich, Güntzer & Gholamreza, Nakhaeizadeh,(2000) "Algorithms for Association Rule Mining A General Survey and Comparison", *J. SIGKDD Explorations*, vol 12, pp. 58-64.

[7]   Yanyan, Li. Mingkai, Dong & Ronghuai, Huang, (2008) "Semantic organization of online discussion transcripts for active collaborative learning", *8th Int. Conf. IEEE Advanced Learning Technologies*, pp. 756–760.

[8]   Carlos, Castro-Herrera, (2010) "A hybrid recommender system for finding relevant users in open source forums", *3th Int. Managing Requirements Knowledge (MARK) Workshop*, pp. 41–50.

[9]   Carlos, Coastro-Herrera. Jane, Cleland-Huang & Bamshad, Mobasher, (2009) "A recommender system for dynamically evolving online forums", *Proc. 3th Conf. ACM Recommender Systems*, pp. 213–216.

[10]   Chenn-Jung, Huang. Hong-Xin, Chen & Chun-Hua, Chen,(2009) "Developing argumentation processing agents for computer supported collaborative learning", *J. Expert Systems with Applications,* 362: pp. 2615-2624.