International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 5, Sep - Oct 2016

RESEARCH ARTICLE OPEN ACCESS

Modified Fuzzy C Means Clustering To Study the Willingness Maximization Using Mean Weighted Artificial Bee Colony Algorithm (MWABC) For Social Activity Planning

Alagulaskhmisubha. A [1], Mohanapriya. C [2]
Research scholar [1], Assistant Professor [2]
Department of Computer Science
Tirupur Kumaran College for Women, Tirupur
Tamil Nadu – India

ABSTRACT

Studies show with intention of a person is willing toward join a social group movement if the activity is interesting, and if some close friends moreover join the activity as companions. In the recent work demonstrates that the interests of a person and the social tightness between friends have been able to successfully resultant and mined from social networking websites. On the other hand, even by means of the above mentioned information extensively available, social group activities at rest need to be matched physically, and the procedure is difficult and time consumption task for social users, particularly designed for a huge social group activity, appropriate toward difficulty of social connectivity and the variety of possible interests between friends. In order solve all this problem in this paper, new Modified Fuzzy C Means (MFCM) clustering algorithm is proposed to similar users activity designed for large social group activity. MFCM clustering is proposed which attendees of a social group activity, which becomes very helpful for social networking websites. Before that initially first formulate a problem, called Willingness mAximization for Social grOup (WASO). This work WASO problem is solved by using Mean Weighted Artificial Bee Colony (MWABC) algorithm. In MWABC algorithm the weight values is assigned to each users based on the behavior of bees with hive. ABC algorithm the colony of artificial bees consists of three groups of bees: employed bees, onlookers and scouts. For MWABC Optimization Algorithm every food source, there is only one employed bee to assign weight values to users. Given the presented computational budgets, the proposed MWABC Optimization Algorithm is being able to optimally solve WASO problem and find a WASO solution with an approximation ratio. In this research work implement the proposed MWABC algorithm in Facebook, and the user study demonstrates that social groups obtained by the proposed MWABC algorithm significantly outperform the solutions manually configured by users.

Keywords:- Social network, query processing, optimization, social media; data mining; social data; social media mining;

I. INTRODUCTION

Studies demonstrated with the intention of two major important criteria's are generally used for decision making of a person joining a group activity at her obtainable time. Initially, the person is concerned in the inherent characteristics of the activity, which might be in line through her desired hobby or based on their exercise. Subsequent, other people who are significant toward the person, such as her close friends, motivation join the action as friend. At the present time, several people are familiar toward sharing information by means of their friends on social networking websites such as Facebook, Myspace, Meetup, MyYearbook, and LikeALittle, and a recent studies [1-2] introduces a many algorithms towards measure the interests of a person related to the interest features in her personal profile and the contextual information in her communication by means of friends.

Furthermore, social connectivity methods have been extensively introduced and developed [3] designed for evaluating the rigidity among two friends in the above websites. However, even by means of the above information obtainable, to date there has been neither published work nor a real system search how toward influence the above two crucial factors designed for regular planning and recommending of a group activity, which is potentially very helpful designed for social networking websites as a value-added service. Furthermore, social connectivity models have been extensively developed [3] designed for evaluating the inflexibility with two friends in the above websites. However, even by means of the above information obtainable, to date there has been either published or non published work search how toward influence the above two crucial factors designed for regular planning and recommending of a group activity, which is very helpful designed for social networking websites as a value-added service. At present, many social networking websites only act as a platform for information sharing and exchange in activity planning. The attendees of a group movement still required toward be chooses physically, and such physical coordination is frequently difficult and time-consuming, mainly designed for a huge social activity, specified the complex social link structure and the various interests. To conquer these problem this research work uses an initial effort toward include the interests of people and their social rigidity as two important factors towards discover a group of attendees designed for automatic planning and recommendation. It is popular towards select more attendees who like and benefit from the activity by means of the shared concentration in the activity as friends. Actually, Psychology [4] and studies related to social networks [5-6] have been developed using willingness toward concentrate an activity as the sum of the interest of every attendee on the activity and the social rigidity among friends with the purpose is to potential to join it. It is predicted with the purpose of the chosen attendees are more disposed toward connect the activity if the motivation of the group increases.

Data mining which consists of several numbers of techniques is association rule mining, anomaly detection, feature selection, instance selection, and visual analytics. These methods have been implemented in Han et al. [7], Tan et al. [8], Witten et al. [9], Zhao and Liu [10], and Liu and Motoda [11]. Mining social media is a growing multidisciplinary area where investigators of varied backgrounds is be able to be make significant contributions with the purpose of substance designed for social media research and improvement .In data mining, methods have been generally categorized into three categories supervised , unsupervised and semi supervised learning algorithms. Among these methods unsupervised learning algorithms is applied to clustering problems. Used for a given task, unsupervised learning algorithms are performed based on the similarity or dissimilarity among data objects. Similarity or dissimilarity among the data objects might be determined or calculated by using distance measures such as Euclidean, Minkowski, and Mahalanobis. Some more distance or measures also used for measuring similarity and dissimilarity between data objects those are matching coefficient, Jaccard coefficient, cosine similarity, and Pearson’s correlation. These methods have been applied to traditional clustering methods such as K-means, hierarchical clustering and density-based clustering. But these traditional clustering methods have their own advantages and disadvantages which are solved by using semisupervised learning .Since the semisupervised learning requires only less number of labeled dataset samples and large amounts of unlabeled data.

From this motivation this research work uses a data mining methods that solves the Willingness mAximization for Social grOup (WASO) optimization problem. New MWABC is proposed that sequentially chooses an attendee that leads to the largest increment in the willingness at each iteration. Before that initially first formulate a problem, called Willingness mAximization for Social grOup (WASO). This work WASO problem is solved by using Mean Weighted Artificial Bee Colony (MWABC) algorithm. In MWABC algorithm the weight values is assigned to each users based on the behavior of bees with hive. ABC algorithm the colony of artificial bees consists of three groups of bees: employed bees, onlookers and scouts. For MWABC Optimization Algorithm every food source, there is only one employed bee to assign weight values to users. Modified Fuzzy C Means (MFCM) is proposed for the grouping of similar users particularly for large social group activity. MFCM clustering method suggests prospective attendees of a social group activity, which might be extremely helpful, designed for social networking websites as a value-added service. Note the MWABC, though simple, tends to be trapped in a local optimal solution, since it facilitates the selection of nodes only suitable at the corresponding iterations.

II. RELATED WORK

By means of considering randomization, the above specified methods are capable towards successfully and easily trapped into local optimal solution. In general methods have two major drawbacks. The first disadvantage of these systems, a start node with the purpose of has the probable toward create final solutions by means of high willingness is not provide by means of further computational budgets designed for randomization in the subsequent iterations. Each initial node in the randomization methods is enlarged towards simply one final solution. Therefore, initial nodes, which have the likely to develop and develop into the solution by means of high willingness, it might be unsuccessful towards find a final solution towards high willingness since only one solution is randomly constructed and expanded from the start node. The second drawback is with the purpose of the expansion of the partial solution should not differentiate the chosen of the neighboring nodes. Each neighboring node is pleasured regularly and preferred regularly at random designed for each iteration. In difference, a straightforward method toward solve this problem is towards assign the probability towards each neighboring node related to its interest score and social tightness of incident edges. On the other hand, this task is equivalent to the greedy algorithm in with the purpose of it restricts the scope to the local information related to each node and is not predictable towards create a solution by means of high willingness.

Expert team development in social networks has concerned widespread research interests. The difficulty of creating an expert team is to discover a group of social network user owning the particular skills; at the same time as the communications cost between the selected friends is minimized to assure the rapport between the team members designed a proficient operation. Two communications costs, diameter and minimum spanning tree, were evaluated. Many methods have been introduced and studied to solve these problems. For instance, each skill i required to consists at least k_i people in order to create a strong team [12], at the same time as all-pair shortest paths are integrated to illustrate the communications costs more accurately [13]. Furthermore, a skill leader is chosen for each skill by means of the objective to reduce the social distance from the skill members towards each skill leader [13], while the density of a team is also measured [14].

In addition to expert team formation, community detection as well as graph clustering and graph partitioning have been explored to find groups of nodes mostly based on the graph structure [15]. The quality of an obtained community is usually measured according to the structure inside the community, together with the connectivity within the community and between the rest of the nodes in the graph, such as the density of local edges, deviance from a random null model, and conductance [14]. Sozio et al. [16], for example, detected community by minimizing the total degree of a community with specified nodes. However, the objective function of WASO is different from community detection. Each node and each edge in WASO are associated with an interest score and social tightness score in the problem studied in this paper, in order to maximize the willingness of the attendees with a specified group size, which can be very useful for social networking websites as a value-added service.

From above mentioned methods some other methods are required to solve WASO problem so in the recent work two randomized algorithms such as Computational Budget Allocation for Start nodes (CBAS) and Computation Budget Allocation for Start nodes with Neighbor Differentiation (CBAS-ND) is proposed to solve the issues of crucial factors such as selection of initial nodes and enlarging the partial solutions, correspondingly. This work also uses a Optimal Computing Budget Allocation (OCBA) [17] ,it is performed in random manner with high computational cost in the initial nodes by the prospective to create the solutions by means of high willingness. CBAS initially chooses m and start with m number of nodes , it then randomly adds new nodes to enlarge the partial solution step-by- step, until k nodes are considered as a final solution. Each initial node in CBAS is enlarged towards finds a many final solutions. To properly maintain computational cost, CBAS at each stage finds the start nodes worth more computational cost related to sampled results of the earlier stages. Operational by means of considering computational resources, CBAS is improved to CBAS-ND by calculating probability value to each neighboring node at some stage in the expansion of the partial solution related to the cross entropy method.

Shuai et al [18] initial create a new problem, called Willingness mAximization for Social grOup (WASO). They solve this problem by proposing greedy algorithms which successfully and capably solve the WASO problem. Particular the obtainable computational time, the WASO algorithm is capable towards assign the resources and discovers a best solution via approximation ratio. This WASO algorithm is implemented to facebook social network site, and the user results show that social groups achieved by the proposed WASO algorithm extensively outperform the solutions physically configured by means of users.

III . PROPOSED WILLINGNESS MAXIMIZATION FOR SOCIAL GROUP (WASO) AND MEAN WEIGHTED ARTIFICIAL BEE COLONY (MWABC)

From this motivation this research work uses a data mining methods that solves the Willingness mAximization for Social grOup (WASO) optimization problem. New MWABC is proposed that sequentially chooses an attendee that leads to the largest increment in the willingness at each iteration. Before that initially first formulate a problem, called Willingness mAximization for Social grOup (WASO). The WASO problem is formulated via the use of graph G, where each node is denoted as candidate person and is associated by the use of approximation score of the person designed for each social user activity, and each edge has a social tightness score towards specify the mutual familiarity among the two persons. From the social network user activities not including an a priori fixed size, it is practical for a social network user towards specify a correct range for the group size, and the algorithm finds an optimal WASO solution for each k social network. This work WASO problem is solved by using Mean Weighted Artificial Bee Colony (MWABC) algorithm. In MWABC algorithm the weight values is assigned to each users based on the behavior of bees with hive. ABC algorithm the colony of artificial bees consists of three groups of bees: employed bees, onlookers and scouts. For MWABC Optimization Algorithm every food source, there is only one employed bee to assign weight values to users. Modified Fuzzy C Means (MFCM) is proposed for the grouping of similar users particularly for large social group activity. MFCM clustering method suggests prospective attendees of a social group activity, which might be extremely helpful, designed for social networking websites as a value-added service. Note the MWABC, though simple, tends to be trapped in a local optimal solution, since it facilitates the selection of nodes only suitable at the corresponding iterations.

a. Problem Definition

Let us consider graph model $G = (V; E)$ as social network model, where each node $v_i \in V$and each edge $e_{i,j} \in E$ are related with an interest score η_i and a social tightness score $t_{i,j}$ calculated from the literature [22] respectively .This research paper solves the optimization problem WASO to discovering a optimal set F of vertices inside size k to maximize the willingness W(F), i.e.,

$$\max_F W(F) = \max_F \sum_{v_i \in F} \eta_i + \sum_{v_j \in F: e_{i,j} \in E}^{1} \tau_{i,j} \tag{1}$$

where F is denoted as subgraph in G to support each attendee to be familiar by means of another attendee related to a social path in F. However the social tightness among v_i and v_j is not necessarily symmetric with the purpose of $t_{i;j}$ might be diverse with $t_{j;i}$. So, the willingness in Eq. (1) considers both $t_{i;j}$and $t_{j;i}$. These social network have interest score h as 0 or τ as 0 is and topic interest are intrinsic criteria involved in the decision of a person to join a group activity.

3.2. Modified Fuzzy c Means (MFCM) clustering

The objective of a MFCM algorithm is attempts to partition a finite collection of n elements X=$\{x_1, \ldots, x_n\}$into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centres C=$\{c_1, \ldots, c_n\}$ and a partition matrix W=$w_{i,j} \in [0,1]$; i=1,…,n ,j=1.,,,c , where each element $w_{i,j}$ tells the degree to which element x_i belongs to cluster c_j . Like the k-means algorithm, the FCM aims to minimize an objective function. Fuzzy c-means (FCM) is a method of clustering [20-22] which allows one piece of data to belong to two or more clusters. Here, this method is used in clustering for social network users Data. Generally fuzzy c means algorithm is defined in the following way : It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, 1 \le m < \infty \tag{2}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{3}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{4}$$

This iteration will stop when $max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$. Where ε, a termination criterion between 0 and 1 and k is are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. The FCM algorithm is composed of the following steps:

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$

2. At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $U^{(k)}$ $c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m}$

3. Update $U^{(k)}$, $U^{(k+1)}$ follows
$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. If $\left\|U^{(k)} - U^{(k+1)}\right\| < \varepsilon$ then STOP; otherwise return to step 2.

This algorithm groups the nodes into various clusters and includes the study of the data generation rate as well as the similarity between data series in the sink. The cluster-heads are chosen based on the unused energy, quality of link, and coverage of node in each cluster. The separation of nodes$N = (n_1, \ldots n_k)$ into c cluster is done by the standard FCM objective function which is given as

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} \mu_{ik}^p \|n_k - v_i\|^2 \tag{5}$$

Where $\{v_i\}_{i=1}^{c}$ are the prototypes of the clusters and the array $[\mu_{ik}^p] = U$ correspond to the partition matrix $U \in u$. The parameter $p$is each fuzzy membership weighting exponent and determines the amount of clusteringfuzziness. The minimization of FCM objective function is done if the nodes which are nearer to the centroid of its particular cluster head are assigned by high membership values, and the nodes which are far away from the centroid of its cluster head are assigned by low membership values, modification is proposed to (5) by using a term α that allows the influence of labeling of a cluster nodes to the labels in its immediate neighborhood nodes. The modified objective function is given by

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{p}\left\|n_k - v_i\right\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{p}\left(\sum_{n_r\in\mathcal{N}_k}\left\|n_k - v_i\right\|^2\right)$$

Where $\mathcal{N}_k$ is the set of neighbors nodes of similarity function from equation (1) and N_R is cardinality of the $\mathcal{N}_k$ .The parameter α is controlled by the effect of the neighborhood node term . This MFCM algorithm involves the similarity between data series in the sink as well as the study of the data generation rate.

DGreedy abusively chooses v1 as the initial node and discover basically a solitary series of nodes in the solution space. To increase the search space randomized algorithms such as Computational Budget Allocation for Start nodes (CBAS) and Computation Budget Allocation for Start nodes with Neighbor Differentiation (CBAS-ND) is proposed to solve the issues of crucial factors such as selection of initial nodes and enlarging the partial solutions, correspondingly. This work also uses a Optimal Computing Budget Allocation (OCBA) [22] ,it is performed in random manner with high computational cost in the initial nodes by the prospective to create the solutions by means of high willingness. CBAS initially chooses m and start with m number of nodes , it then randomly adds new nodes to enlarge the partial solution step-by- step, until k nodes are considered as a final solution. Each initial node in CBAS is enlarged towards finds a many final solutions. To properly maintain computational cost, CBAS at each stage finds the start nodes worth more computational cost related to sampled results of the earlier stages. Operational by means of considering computational resources, CBAS is improved to CBAS-ND by calculating probability value to each neighboring node at some stage in the expansion of the partial solution related to the cross entropy method. Give explanation the IP formulation is designed for WASO. Binary variable x_i denotes if node v_i is preferred in the solution F, and binary variable $y_{i;j}$ denotes if two neighboring nodes v_i and v_j are uniformly chosen in F. The objective function is

$$\max_F W(F) = \max_F \sum_{v_i\in F}\eta_i\, x_i + \sum_{v_j\in F: e_{i,j}\in E}^{1}\tau_{i,j}\, y_{i,j} \qquad (6)$$

where the first term is the total interest score, and the second term is the sum social rigidity score of the selected nodes.

3.3.Mean Weighted Artificial Bee Colony (MWABC) algorithm

The Artificial Bee Colony (ABC) algorithm is used for real-time optimization which suggests the foraging behaviour of a bee colony [22]. The ABC algorithm is performed based on the behavior of three kinds of bees: employed, onlooker and scout. Half of the colony includes of employed bees, and the other half consists of onlooker bees. In this paper work , employed bees related to nodes in the graph can be represented as $Y = (y_{i,j})_{m\times n}$, binary variable $y_{i;j}$ represent if two neighboring nodes v_i and v_j. Employed bees are responsible for selection of nearest social network users neigbours in the nodes that nectar sources explored before and giving information to the onlooker bees. Scouts bees randomly search the environment in order to find a new selected nearest neigbours in the nodes depending on fitness function in equation (9) can be summarized as follows:

1. At the initial phase of the foraging process, the bees start to search the randomly in order to find best WASO problem for social network users.

2. After finding a highest objective value the two neighboring nodes becomes an employed forager and starts to make use of the exposed source. After receiving the nectar, go back to find out preferred two neighboring nodes directly by means of performing a dance on the dance area. If it reaches maximum iterations preferred social network users is tired, it becomes a scout and starts toward randomly search designed for a new neighboring nodes.

3. Onlooker bees waiting in the hive watch the dances advertising the profitable neighboring nodes and choose a two neighboring nodes depending on the objective function of a dance proportional to the quality of the dataset samples$Y = (y_{i,j})_{m\times n}$.

In the ABC algorithm, the location of neighboring social network user nodes represents a possible social network activity results to the quality assurance, and the nectar amount of a food source relates to the fitness of the associated solution. If the search space of quality assurance is measured toward be the current environment of the hive with the purpose of consists many highest scores in the search space. Initial two neighboring nodes matrix are produced randomly within the range of the boundaries of the parameters.

$$z_{ij} = z_j^{min} + rand(0,1)(z_j^{max} - z_j^{min}) \qquad (7)$$

where $i = 1\ldots SN, j = 1\ldots D.SN$ is the number of social network users and D is the number of optimization parameters. After initialization, the population of the two neighboring social network users is subjected to repeat cycles of the search processes of the employed bees, the onlooker bees and the scout bees. Termination criteria for the ABC algorithm should be reaching a Maximum Cycle Number (MCN). As mentioned earlier, each employed bee is associated with only neighboring social network users nodes. Hence, the number of neighboring social network users is equal to the number of employed bees. An employed bee produces a variation on the location of the two social network users neighboring nodes in her memory depending on fitness function and finds neighboring social network users. In ABC, finding neighboring social network users is defined by

$$v_{ij} = z_{ij} + \phi_{ij}(z_{ij} - z_{kj}) \tag{8}$$

Within the neighboring social network users represented by z_i, a food source v_{ij} is determined by changing one parameter of z_i. In Eq. (8), j is a random integer in the range [1,D] and $k \in \{1, 2, \ldots SN\}$ is a randomly chosen index. ϕ_{ij} is a uniformly distributed real random number in the range [-1, 1]. As can be seen from Eq. (3.9), as the difference between the parameters of the $z_{i,j}$ and $z_{k,j}$ decreases. After producing v_{ij} within the cycle (MCN), a fitness value for a WASO problem can be assigned to the solution v_{ij} by (3.10).

$$fitness_i = \begin{cases} \dfrac{1}{(1 + f_i . w_i)} & if\ f_i . w_i \geq 0 \\ \dfrac{1}{(abs(f_i . w_i))} & if\ f_i . w_i < 0 \end{cases} \tag{9}$$

where f_i is the quality assurance value from objective function. For maximization problems, the cost function is determined directly from fitness function. Related to ABC declaration, assigned a weight $W(a_i)$ to each attribute a_i. The value of weight $W(a_i)$ for each a_i, which is set to zero initially, is calculated sequentially throughout the whole matrix using the mean value of the attribute and update using the following formula when a new entry a_i is met in the discernibility matrix:

$$w_i = w(a_i).\mu(a_i) \tag{10}$$

After all employed bees complete their searches, and then onlooker bee evaluates the nectar information that chooses a best neighboring social network users with a highest probability in the two neighboring social network users is employed (11):

$$p_i = \frac{fitness_i}{\sum_{i=1}^{SN} fitness_i} \tag{11}$$

In the ABC algorithm, a random real number within the range [0,1] is generated for each source. If the probability value (p_i in Eq. (11)) associated with that source is greater than this random number then the onlooker bee produces a modification on the position of this selected neighboring social network users by using Eq. (9) as in the case of the employed bee. After the highest quality assurance is evaluated, greedy selection is applied and the onlooker bee either memorizes the new neighboring social network users position by forgetting the old one or keeps the old one. If the selected neighboring social network users z_i cannot be improved, its counter holding trials is incremented by 1, otherwise, the counter is reset to 0. If the value of the counter is greater than the control parameter of the ABC algorithm, the current neighboring social network users is assumed to be exhausted and is abandoned. Assume that the abandoned source is z_i, then the scout randomly discovers a new neighboring social network users food source to be replaced with z_i.

$$v_{ij} = \begin{cases} z_{ij} + \mu_{ij}(z_{ij} - z_{kj})\ if\ R_{ij} < \mu_{ij} \\ z_{ij}\ otherwise \end{cases} \tag{12}$$

However in order to enhance the convergence rate, ABC algorithm is enhanced by using mean value of the attribure (μ) number, (0 <R_{ij} < 1), then the parameter v_{ij} is modified as in the Eq. (12).

Algorithm 1: Mean Weight Artificial Bee Colony (MWABC) algorithm

1. Initialize the population of solutions $z_{i,j}, i = 1 \ldots SN, j = 1 \ldots D$, trial$_i$ = 0 trial$_i$ is the non-improvement number of the solution z_i, weight values W(a_i)=0, i=1,···,n.used for abandonment
2. Evaluate the population
3. cycle = 1
4. repeat {——Produce a new food source population for Employed bee——}
5. for i = 1 to SN do
6. Produce a new feature selection solution v_i for the employed bee by using (3.13) and evaluate its fitness function
7. Apply a greedy selection process between v_i and z_i from this select the better features
8. If solution z_i does not improve $\text{trial}_i = \text{trial}_i + 1$, otherwise trial_i = 0
9. end for
10. Calculate the probability values p_i by (11) for the solutions using fitness values {——Produce a new food source population for onlookers———}
11. t = 0,i = 1
12. repeat
13. if random < p_i then
14. Produce a new v_{ij} food source by (12) for the

onlooker bee

15. Apply a greedy selection process between v_i and z_i from select the better features

16. If solution z_i does not improve $trial_i = trial_i + 1$, otherwise $trial_i = 0$

17. t = t + 1

18. end if

19. until (t = SN){——Determine Scout——}

20. if max($trial_i$) > limit then

21. Replace z_i with a new randomly produced solution by (7)

22. end if

23. Memorize the best classification accuracy achieved so far

24. Compute the lower and upper approximate classification accuracy

25. Form a approximation matrix merge all the same entries in the selected feature matrix, record their frequencies and sort all entries in the matrix according to their length in descending order; if two entries have the same length, the entry with more frequency is preferred

26. cycle = cycle+1

27. until (cycle = Maximum Cycle Number)

IV.PERFORMANCE ANALYSIS

In this section measure evaluate the effectiveness and efficiency of three optimization methods such as to DGreedy, CBAS-ND and proposed MWABC .These methods have been experimented to facebook social network which consists of information about 137 peoples with different organization such as schools, government, technology companies, and businesses. Additionally, each user in the social network is asked to preparation 10 social activities with social graphs extracted from their particular social network that is Facebook. These ten social network activities make use of differed network sizes and different numbers of attendees. In this section measure the performance results of the various optimizations methods such as DGreedy, CBAS-ND and proposed MWABC using the metrics like precision, recall, F-measure and accuracy.

Precision and recall

Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

$$P = \frac{TP}{(TP + FP)} \quad (13)$$

$$R = \frac{TP}{(TP + FN)} \quad (14)$$

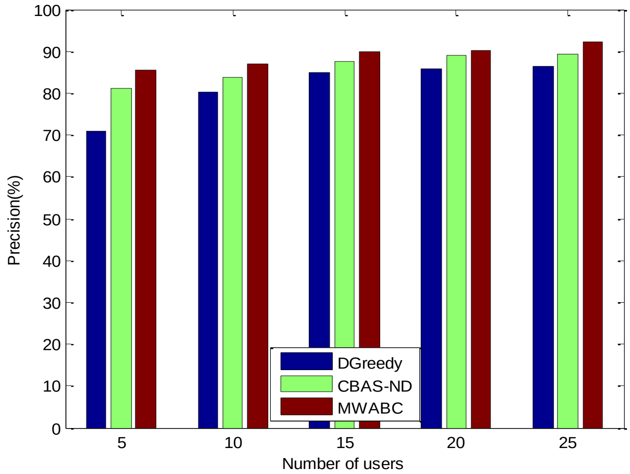

Figure 1. The chart of precision results comparison for number of users

Figure 1 shows the precision results comparison of the various methods DGreedy, CBAS-ND and proposed MWABC, the values are tabulated in Table .1. From the experimental results it is inferred that for the facebook dataset the proposed MWABC algorithm performs 2.81 % better than the CBAS-ND algorithm, 7.324% better than the DGreedy algorithm is illustrated in Figure 1. Proposed MWABC algorithm produces 88.992%, CBAS-ND algorithm produces 86.182% and Dgreedy produces 81.668% is illustrated in Figure 1.

Table.1. Precision results comparison

No.of users	Precision (%)		
	DGreedy	**CBAS-ND**	**MWABC**
5	70.97	81.28	85.63
10	80.27	83.81	86.97
15	85	87.63	89.89
20	85.71	88.91	90.12
25	86.39	89.28	92.35

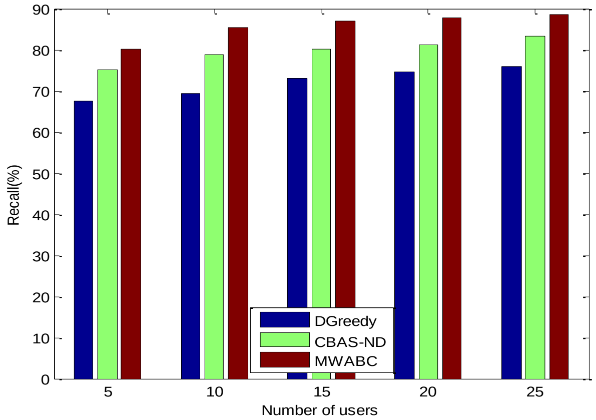

Figure 2. The chart of recall results comparison for number of users

Figure 2 shows the recall results comparison of the various methods DGreedy, CBAS-ND and proposed MWABC, the values are tabulated in Table 2. From the experimental results it is inferred that for the facebook dataset the proposed MWABC algorithm performs 6.0625 % better than the CBAS-ND algorithm, 13.7013% better than the DGreedy algorithm is illustrated in Fig 5.2. Proposed MWABC algorithm produces 72.048%, CBAS-ND algorithm produces 79.687% and Dgreedy produces 85.74% is illustrated in Figure 2.

Table.2. Recall results comparison

No.of users	Recall (%)		
	DGreedy	CBAS-ND	MWABC
5	67.3969	75.1237	80.1914
10	69.326	78.69	85.28
15	72.98	80.198	86.91
20	74.58	81.24	87.78
25	75.96	83.185	88.588

F-measure

Accuracy and F-measure are based on a combinatorial approach which considers each possible pair of objects. Each pair can fall into one of four groups: if both objects belong to the same class and same cluster then the pair is a True Positive (TP); if objects belong to the same cluster but different classes the pair is a False Positive (FP); if objects belong to the same class but different clusters the pair is a False Negative (FN); otherwise the objects belong to different classes and different clusters, and the pair is a True Negative (TN) . F-measure is described as equation (15)

$$\mathrm{F-measure} = \frac{2PR}{(P+R)}, \quad (15)$$

where $P = \frac{TP}{(TP+FP)}$, and $R = \frac{TP}{(TP+FN)}$

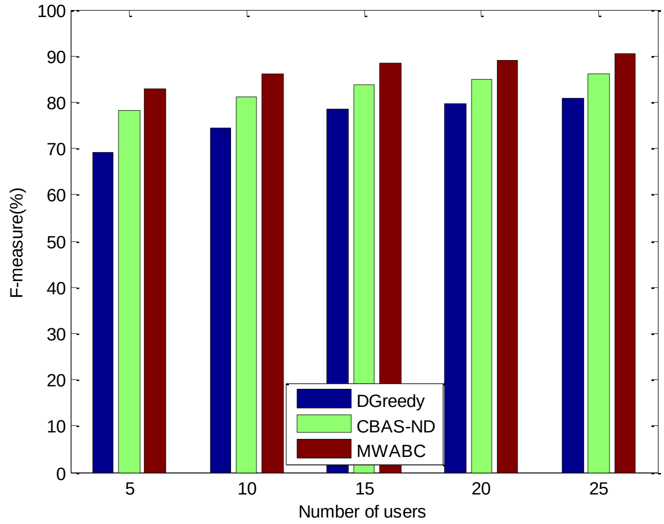

Figure 3. The chart of F-measure results comparison for number of users

Figure 3 shows the F-measure results comparison of the various methods DGreedy, CBAS-ND and proposed MWABC, the values are tabulated in Table 3. From the experimental results it is inferred that for the facebook dataset the proposed MWABC algorithm performs 4.53 % better than the CBAS-ND algorithm, 10.80% better than the DGreedy algorithm is illustrated in Figure 3. Proposed MWABC algorithm produces 76.5332%, CBAS-ND algorithm produces 82.80528% and Dgreedy produces 87.3355% is illustrated in Figure 3.

Table.3. F-Measure results comparison

No.of users	F-measure (%)		
	DGreedy	CBAS-ND	MWABC
5	69.13732	78.08069	82.82151
10	74.39768	81.16934	86.11671
15	78.53273	83.74944	88.37489
20	79.75859	84.90213	88.93461
25	80.83997	86.1248	90.42989

Accuracy: Usually, the accuracy rate in Eq. (16) is the most frequently used measure in assessment metrics. But in the framework of the leukemia datasets, the accuracy is a proper measure, because it distinguishes between the numbers of correctly classified examples of different classes,

$$\text{Accuracy (Acc)} = TP+TN/(TP+FN+FP+FN) \quad (16)$$

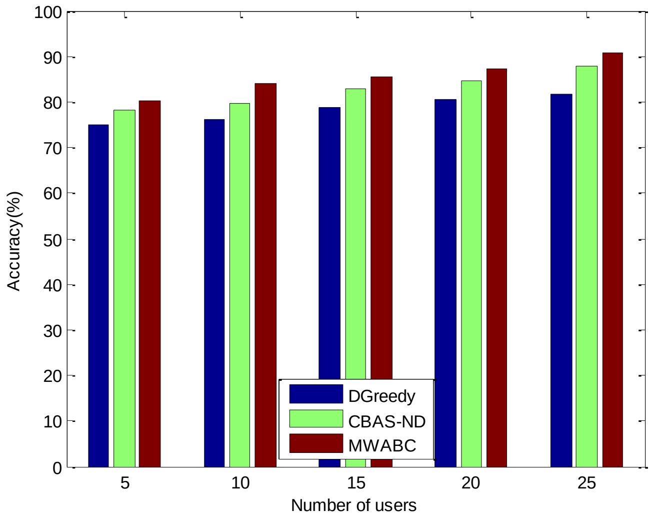

Figure 4. The chart of Accuracy results comparison for number of users

Figure 4 shows the accuracy results comparison of the various methods DGreedy, CBAS-ND and proposed MWABC, the values are tabulated in Table 4. From the experimental results it is inferred that for the facebook dataset the proposed MWABC algorithm performs 2.88% better than the CBAS-ND algorithm, 7.044% better than the DGreedy algorithm is illustrated in Figure 4. Proposed MWABC algorithm produces 78.536% , CBAS-ND algorithm produces 82.65% and Dgreedy produces 85.58% is illustrated in Figure 4.

Table.4. Accuracy results comparison

No.of users	Accuracy (%)		
	DGreedy	**CBAS-ND**	**MWABC**
5	75.12	78.08	80.15
10	76.25	79.81	83.98
15	78.81	82.97	85.63
20	80.63	84.57	87.29
25	81.87	87.83	90.85

Timing Analysis

Figure 5 shows the execution time results comparison of the various methods DGreedy, CBAS-ND and proposed MWABC, the values are tabulated in Table 5. From the experimental results it is inferred that for the facebook dataset the proposed MWABC algorithm performs 0.0354 seconds lesser than the CBAS-ND algorithm, 0.071 seconds lesser than the DGreedy algorithm is illustrated in Figure 5. Proposed MWABC algorithm performs 0.3126 seconds , CBAS-ND algorithm performs 0.348 seconds and Dgreedy performs 0.3836 seconds is illustrated in Figure 5.

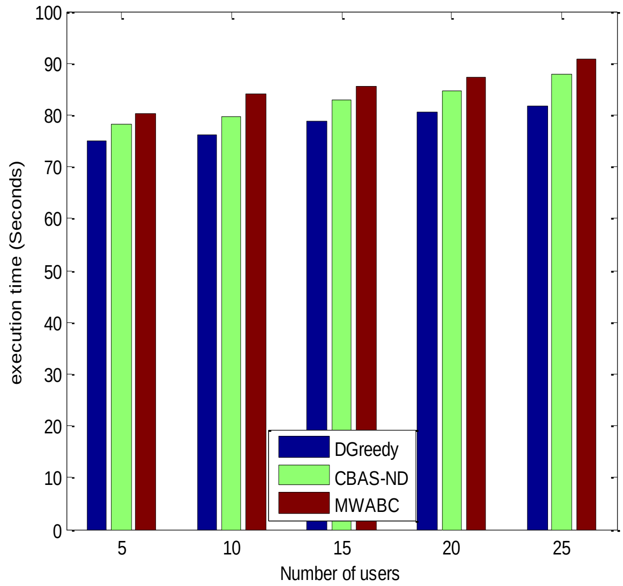

Figure 5. The chart of execution time results comparison for number of users

Table.5. Execution time results comparison

No.of users	execution time (Seconds)		
	DGreedy	**CBAS-ND**	**MWABC**
5	75.12	78.080689	80.15
10	76.25	79.81	83.98
15	78.81	82.97	85.63
20	80.63	84.57	87.29
25	81.87	87.83	90.85

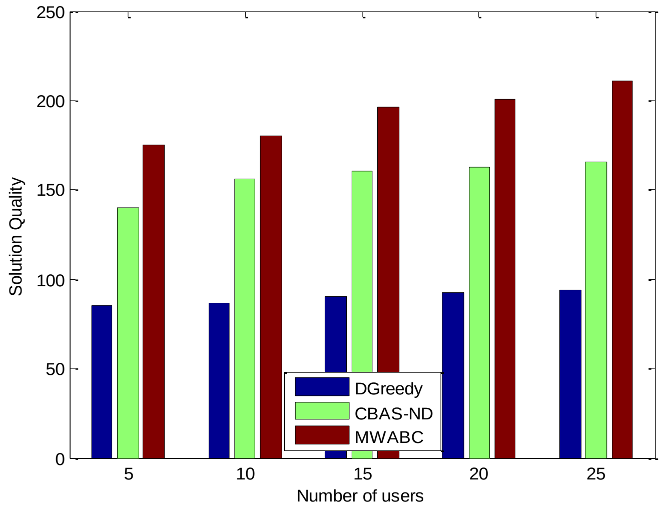

Figure 6. Results of user study

Figure 6 shows the performance comparison results of various methods with their solution quality and running time with varied network sizes, where the expected number of attendees k is 7. The experimentation results demonstrate that the proposed MWABC provides best optimal results for solving the WASO problem when compared to existing DGreedy and CBAS-ND method .Since the existing methods are hard for manual coordination, even when the network size is high.

V.CONCLUSION AND FUTURE WORK

In this paper describe a new problem description depending on Willingness mAximization for Social grOup (WASO) towards find a set of attendees and formulate use of the willingness. Demonstrated with the purpose of WASO is hard problem and proposed a new Mean Weighted Artificial Bee Colony (MWABC) algorithm via approximation ratio. Logically, moreover incrementally create the group, a deterministic MWABC sequentially choose an attendee that direct to the important addition in the willingness at every iteration. In MWABC algorithm the weight values is assigned to each users based on the behavior of bees with hive. MWABC algorithm is performed based on the colony of three groups of bees: employed, onlookers and scouts. Modified Fuzzy C Means (MFCM) is proposed in this paper for grouping of similar users mainly designed for large social group activity. MFCM clustering proposes possible attendees of a social group activity, which capability is mainly helpful, designed for social networking websites. Note the MWABC, although easy, tends to be attentive in a local optimal solution, because it facilitates the choice of nodes simply appropriate at the equivalent iterations. There are several interesting directions to be investigated. The user learning resulted in practical directions toward improve WASO designed for future research. Some users suggested with the purpose of incorporate the proposed MWABC willingness optimization system by means of regular obtainable time extraction toward filters busy users, such as by integrating the proposed MWABC system with Google Calendar. Since candidate attendees are related via multiple features in social networks those attributes are location and gender, these attributes have been specified as input parameters towards further filter out inappropriate candidate attendees. Last but not the least, some users pointed out that this work could be extended to allow users to specify some attendees that must be included in a certain group activity.

REFERENCES

[1] Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. In SIAM, 51(4):661–703, 2009.

[2] Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: Inferring user profiles in online social networks. In WSDM, pages 251–260, 2010.

[3] V. Chaoji, S. Ranu, R. Rastogi, and R. Bhatt. Recommendations to boost content spread in social networks. In WWW, pages 529–538, 2012

[4] M. F. Kaplan and C. E. Miller. Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. Journal of Personality and Social Psychology, 53(2):306–313, 1987.

[5] D. N. Yang, W. C. Lee, N. H. Chia, M. Ye, and H. J. Hung. Bundle configuration for spread maximization in viral marketing via social networks. In CIKM, pp. 2234–2238, 2012.

[6] M. Ye, X. Liu, and W. C. Lee. Exploring social influence for recommendation - a probabilistic generative model approach. In SIGIR, pp. 671–680, 2012.

[7] J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, 2011

[8] P.-N. Tan, M. Steinbach, and V. Kumar. Introduction to Data Mining. Pearson AddisonWesley, Boston, 2006.

[9] H. Witten, E. Frank, and M. A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2011.

[10] Z. A. Zhao and H. Liu. Spectral Feature Selection for Data Mining. Chapman & Hall/CRC Press, Virginia Beach, VA, 2012.

[11] H. Liu and H. Motoda. Computational Methods of Feature Selection. Chapman & Hall, Boca Raton, FL, 2008.
[12] C. Li and M. Shan. Team formation for generalized tasks in expertise social networks. In Proc. of IEEE International Conference on Social Computing, 2010.
[13] M. Kargar and A. An. Discovering top-k teams of experts with/without a leader in social networks. In Proc. CIKM, 2011.
[14] A. Gajewar and A. D. Sarma. Multi-skill collaborative teams based on densest subgraphs. In Proc. SDM, 2012
[15] D. F. Gleich and C. Seshadhri. Vertex neighborhoods, low conductance cuts, and good seeds for local community methods. In Proc. KDD, pages 597–605, 2012.
[16] U. Brandes, D. Delling, M. Gaertler, R. Goerke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. IEEE Transactions on Knowledge and Data Engineering, 20:172–188, 2008
[17] M. Sozio and A. Gionis. The community-search problem and how to plan a successful cocktail party. In Proc. KDD, pages 939–948, 2010.
[18] C. H. Chen, E. Yucesan, L. Dai, and H. C. Chen. Efficient computation of optimal budget allocation for discrete event simulation experiment. IIE Transactions, 42(1):60–70, 2010.
[19] Christophe Gauge & Sreela Sasi , "Modified Fuzzy C-means Clustering Algorithm with Spatial Distance to Cluster Center of Gravity", IEEE International Symposium on Multimedia (ISM), 2010 , pp. 308 - 313
[20] Wen-Liang Hung, and Yen-Chang Chang, "A Modified Fuzzy C-Means Algorithm for Differentiation in MRI of Ophthalmology", Modeling Decisions for Artificial Intelligence, Volume 3885 ,pp 340-350,2006.
[21] László Szilágyi, , Sándor M. Szilágyi and Zoltán Benyó, "A Modified Fuzzy C-Means Algorithm for MR Brain Image Segmentation", Image Analysis and Recognition of the series in Computer Science , vol. 4633 ,pp 866-877,2007.
[22] Karaboga D., An Idea Based On Honey Bee Swarm for Numerical Optimization, Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.