An Evolutionary Algorithm for Homogeneous Grouping to Enhance Web-based Collaborative Learning

M. Mahdi Barati Jozan¹, Fattaneh Taghiyareh²

¹²School of Electrical and Computer Engineering
Collage of Engineering, University of Tehran
Tehran, IRAN
E-mail: {m_barati, ftaghiyar}@ut.ac.ir
Author Correspondence: School of Electrical and Computer Engineering, University of Tehran, North Kargar Ave., P.O.Box 14395/515, Tehran, IRAN, +98-21-88352148, ftaghiyar@ut.ac.ir

Abstract

Grouping of students is an important educational activity in traditional learning and e-learning environments and lots of research has been done in this area. In this paper the new algorithm is proposed for grouping of students with unlimited features. Our proposed algorithm considers the priority of features as well as their values. Priority of features is involved in the grouping by taking advantage of the Inversion concept. The results indicate that our algorithm is successful in both intra-fitness and inter-fitness grouping criteria. The discrepancy between the members of group that is called intra-group fitness and the similarity between heterogeneous formed groups that is called inter-fitness group.

Keywords: Group Learning, Inversion, Genetic Algorithm, Collaborative Learning.

1. Introduction

Collaborative learning in traditional learning and e-learning environments is very important and highly regarded. One of the major activities in collaborative learning is grouping of students that many researches has been performed in this activity. grouping methods include a wide range from simple random selection (Huxham & Land, 1955) to more sophisticated methods that used advanced algorithms (Graf & Bekele, 2006), (Gogoulou, 2007), (Wang et al. 2007), (Hwang et al. 2008), (Ho et al. 2009), (Ani et al. 2010), (Lin et al. 2010), (Moreno et al. 2012), (Abnar et al. 2012).

One of the disadvantages of simple methods is forming groups that may not have the same efficiency and homogeneity. This heterogeneous groups causes some of them to be unable to reach all predefined goals or even fail (Wang et al. 2007), (Lin et al. 2010), (Moreno et al. 2012).

If grouping of students is done efficiently, workload will be divided proportionally between members of groups, more diverse solutions to solve their problems will be innovated, more incentives will be created (Ani et al. 2010), students in a group can promote each other’s success through helping, sharing, assisting, explaining, and encouraging (Hwang et al. 2008) and communication and management skills and problem-solving ability in the members of group will be increased (Hwang et al. 2008), (Ani et al. 2010). Therefore, the selection of students for placement in groups is a particular important factor in collaborative learning tasks.

As mentioned above, forming efficient group needs effective and more sophisticated algorithms. Researchers have been done much effort in this area, some of them have been tried the data mining techniques such as clustering, Bayesian networks, classification algorithms and etc. In (Romero & Ventura, 2010) a survey of data mining techniques that have been used in grouping of students has been done. Besides the use of data
mining techniques, biologically inspired and artificial intelligence (AI) algorithms have also been used, and can form efficient groups.

In (Gogoulou 2007), (Wang et al. 2007), (Hwang et al. 2008), (Ani et al. 2010), (Moreno et al. 2012), (Abnar et al. 2012), the proposed algorithms use genetic approach for group formation. In (Graf & Bekele, 2006) the introduced algorithm uses Ant Colony Optimization (ACO) to forming groups. In (Ho et al. 2009), (Lin et a., 2010) the proposed algorithms use Particle Swarm Optimization (PSO) for composing group.

The relevant literature was reviewed for background study. After reviewing the literature, four important factors for grouping of students were extracted and for having a successful grouping must be considered. The first, the kinds and number of features for grouping of students should be specified (Bradley & Hebert, 1997) (Moreno et al. 2012), (Martín & Paredes, 2004). In (Bradley & Hebert, 1997), (Martín & Paredes, 2004) emphasize that features such as gender, ethnic background, motivations, attitudes, interests, and personality beside performance level should be considered in the forming of groups. Grouping methods should not have a restriction on the number of features. The second, the number of groups and number of students in each group must be specified (Moreno et al. 2012). The third factor is group type -heterogeneous, homogenous and mixed- (Wang et al. 2007) and final factor is the interaction between students in a group. In (Stahl et al. 2006) points out students learn not only from the contents of the course, but also learn from interactions with each other in a group.

In most studies in the grouping of students, only the values of features have been considered. In this paper, an algorithm is proposed that in addition to the values of features, also uses their priorities. In this study, priority as fifth factor that should be considered in grouping of student is introduced. The proposed algorithm also, in contrast to (Graf & Bekele, 2006), (Gogoulou et al. 2007), (Wang et al. 2007), (Ho et al. 2009), (Ani et al. 2010), (Lin et al. 2010), doesn’t limit the number of features of students for grouping. The experiments show that considering the priorities of the features will form groups that are better and more coherent.

2. The Proposed Method

Before introduce the proposed algorithm in this paper, it need to give definitions for some important concepts that will refer to them in the later sections.

2.1 Definitions

Feature: a variable that is used to describe one of the attributes of a student. Each feature has two parts 1) metadata: includes information about that feature such as name, description, value, range of value it can possess and etc.2) value: a specific amount in the legitimate range for that feature that is specified in the metadata. Feature is represented as F :< f-metadata, f-value>.

Student: an order vector that each member of it is a feature that is represented as ST :<F1, F2 ... Fk>.

Group: Group is set of students presented as G :<st1, st2 ... stn>. There are 3 different ways in the formation of groups. In the first way that is named as homogeneous, the criterion for the formation of group is that the members of a group must have common features between them and the number of these common features must be as much as possible. In other words, the higher the number of common features the better. In the second way that is named as heterogeneous, students with different features will be in a group. The rule that is governing the formation of the groups is to consider differences in features among students. In other words, unlike the first way, the students those have different values for their features will be in a group. In the third way that is named as mixed, features of members will be divided into two categories, a category that includes the features that members must have same values for them, and the second category of features that the group members must have different values for them.

Intra-group fitness: intra-group fitness is the criterion that is used for assessing the quality of the formed group from a set of specific students.

Inter-group fitness: inter-group fitness compares the competency of a formed group with others. This measure is important in the field of education. There are two types available 1) several groups are almost equally balanced in terms of the features of students who were in them and 2) one or more groups are excellent and other groups are very weak. The first type of grouping is more favourable (Dembo, 1994), (Moreno et al. 2012). For example, if we have a classroom of 16 students and their learning ability is labelled A, B, C, and D and in each group there are four students. Between the following two types 1) One member of each label is in each group 2) all student that have rate A form group 1, all student that have rate B form next group and etc., the first type can be considered a better grouping, because in the second type of formation of groups according to the members in each group, some groups have very good values for their fitness, but others possess very low
values for their quality. In other word, optimal grouping is, groups should have highest homogeneity between
themselves and students in each group should have highest heterogeneity between each other.

2.2 Explaining the proposed method

As presented in the previous section, finding the optimal groups is one of the most important activities in
learning. Finding the optimal solution for grouping of students is a NP-hard problem (Lin et al, 2010) and need
exponential time to solve. For example, if we have n students and k groups, in a way that n is divisible by k;
total number of formation of groups is equal to (1)

\[
\binom{n}{k} + \binom{n-k}{k} + \cdots + \binom{n}{k} = \sum_{i=0}^{k-1} \binom{n-(i+1)k}{k}
\]

For each case we must calculate the amount of intra-group and inter-group fitness and then choose the best
among all of them that it is very time consuming.

Since obtaining the optimal solution is very cumbersome or intractable in the general case, using an
approximation algorithm that does not find the optimal solution but its answer is close to the optimal solution
is an appropriate strategy. One of the most used and popular algorithm of this kind is the genetic algorithm
(Falkenauer, 1999). In this algorithm, first the students are placed randomly in groups (initial population), then
in each iteration, for each group, fitness function calculates the group fitness. Finally by using genetic
operators the algorithm try to move the members between groups in order to find better groups (groups formed
in each iteration are called a generation). Exchanging group members will continue and only will be aborted
when a predetermined terminating condition is reached and in that case the algorithm will finish its execution.

In almost all the previous algorithms, values of features are being considered to govern the process of
grouping of the students and genetic operators were defined only based on features values. We believe that in
addition to the values of features, the values that students have acquired in each feature, the priority of the
features to each other is important. For example, if three features 1) social characteristics, 2) IQ and 3) scores
of students are to be considered, then if the values of these features in order for students be as follows (Table
1):

<table>
<thead>
<tr>
<th>Students</th>
<th>Social characteristics</th>
<th>IQ</th>
<th>Scores of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student1</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Student2</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Student3</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Student4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2: Group1

<table>
<thead>
<tr>
<th>Feature</th>
<th>Student1</th>
<th>Student2</th>
</tr>
</thead>
<tbody>
<tr>
<td>social characteristics</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>IQ</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>scores of students</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3: Group 2

<table>
<thead>
<tr>
<th>Feature</th>
<th>Student3</th>
<th>Student4</th>
</tr>
</thead>
<tbody>
<tr>
<td>social characteristics</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>IQ</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>scores of students</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

The values of features have been normalized between 1 and 5 and placed in the above tables.

In the example above, if we consider only the values of features as the decision criterion, a noticeable
decision criterion implicit in the above data set has been neglected that is the priority of features against each
other. In the Table 2, the first group, it is clear that both students have the same value for their second feature
but student1 is a social person and good at scores, in contrast, second student’s scores are higher but is not a
social person. In the Table 3, the second group, both students are the same in first feature but student4 have
better IQ and good at scores, in contrast, third student’s scores are higher but is not clever and creative student.
This point -priority of features- has not been noticed in the past. For example, in (Gogoulou 2007) Group Quality of these groups is calculated as (3) and (4).

\[
\text{Quality of Group for each feature} = (\text{maximum value of feature } - 1) - \\
(\text{max(value of that feature in students of that group)} - \\
\text{min(value of that feature in student of that group)}) \tag{3}
\]

\[
\text{Quality Group} = \sum_{\text{feature}(e) \in \text{student's features}} \text{Quality of Group for feature}(e) \tag{4}
\]

Based on the above criteria, the qualities of groups in the previous example are: Quality Group (G1) = Quality Group (G2) = 8.

As mentioned, although the qualities of groups, G1 and G2, are the same but the features that students in these groups had different values for them, were not the same. This observation can be considered as a big deficiency in those algorithms which only consider the value of the features.

In this paper, before explaining the proposed algorithm used in this study, the concept of Inversion will be introduced, and then new Inversion-base algorithm will be described.

Inversion is defined as if \(A[1 \ldots n]\) be an order vector of \(n\) distinct values, if \(i < j\) and \(A[i] > A[j]\), then the pair \((i, j)\) is called an Inversion of \(A\). Several algorithms have been proposed to account Inversions in an order vector. The best algorithm for this propose has the \(O(n\log n)\) complexity (Kleinberg & Tardos, 2005). For a better intuitive understanding of the Inversion, suppose that we have a list of 5 movies and we want from 2 person to rate them based on their preferences between 1 and 10. Suppose the result is showed in the Table 4.

If movies are sorted into descending order based on their rating for all the voters, and then we draw a line between same movies in the two rating list that we are currently comparing them with each other, then the number of Inversions can simply be obtained (Figure 1).

<table>
<thead>
<tr>
<th>Films</th>
<th>Person1</th>
<th>Person2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film_1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Film_2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Film_3</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Film_4</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Film_5</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5 : Result of Rating

Figure 1 : Number of Inversions is equal to the number of intersection that lines have with each other

In remainder of this section, the new algorithm based on the concept of Inversion that is of the category of genetic algorithms will be introduced. Goal or problem definition, genetic representation of solutions, product the initial population, the fitness function, the crossover and mutation operators, and termination condition are expressed as main part of each genetic algorithm (Floreano & Mattiussi, 2008). We assume that all the features of the student’s vector are sorted base on their values.

Goal: there are two gaols. The first goal is to from groups that their members are as different as possible to each other. The second goal is that members distribute uniformly among the groups base on values and priorities of their features. In other words, the homogeneous groups will be formed.
Genetic representation of solutions: assume that we want to find $k$ groups, and the number of students is equal to $n$, without loss of generality, suppose that $n$ is divisible by $k$. Given that our goal is to form groups that have equal size, the number of students of each group will be equal to $\frac{n}{k}$ as presented in Figure 2.

Initial population: To generate initial population, each of the $n$ individual population randomly put in a $k$ group.

Fitness function: We consider two fitness functions base on Inversion to evaluate the formation groups.

Intra-group fitness: For each pair of students in a group, the number of their Inversions between their feature vectors is calculated and the total number of these Inversions is represented as intra-group fitness. As mentioned previously, because the goal is to create heterogeneous groups, with the help of genetic operators, algorithm tries to place students who have the most number of Inversions between their feature vectors in the same groups. Intra-group fitness is calculated as bellow:

$$\text{intra - group fitness } (g1 \in \text{ group}_1, ..., \text{ group}_k) = \sum_{m1=\forall \text{ Members in group}(g1)} \sum_{m2=\forall \text{ Members in group}(g1) \text{ and } m1\neq m2} \text{ inversion}(m1,m2) \tag{5}$$

$$\text{intra - group fitness (grouping algorithm)} = \sum_{g1=\forall \text{ groups in grouping}} \text{ intra - group fitness } (g1) \tag{6}$$

Inter-group fitness: We expect that groups to be close to each other as possible as it can be, and students are distributed uniformly between groups. So we use the standard deviation of intra-group fitness between formed groups. Intra-group fitness is calculated as follow:

$$\text{inter - group fitness} = \text{ standard - devisrion } (g \in \text{ group}_1, ..., \text{ group}_k) \tag{7}$$

Crossover: the algorithm uses a slightly modified one-point crossover. To determine how the new generation form, two numbers are randomly generated. At first, the proposed algorithm randomly determines the crossover point in the range of 1 to the number of members in each group. To do this it generates a random number that is called crossover-point. Next, it determines how the groups should be combined together and make up the new generation. To do this it generates another random number that is called rotation. These numbers define how each new solution in next generation is obtained from the current solutions. For example, suppose the grouping shown in Figure 3 has been created as a current generation. Numbers indicate students that are in each group, and random numbers generated for a crossover-point is equal to 2 and 3 for rotation. The new generation based on the previous generation and Crossover-point and iteration is shown in Figure 4.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>...</th>
<th>Group k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual 1</td>
<td>Individual 1</td>
<td>Individual 1</td>
<td>...</td>
<td>Individual 1</td>
</tr>
<tr>
<td>Individual 2</td>
<td>Individual 2</td>
<td>Individual 2</td>
<td>...</td>
<td>Individual 2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Individual n/k</td>
<td>Individual n/k</td>
<td>Individual n/k</td>
<td>...</td>
<td>Individual n/k</td>
</tr>
</tbody>
</table>

Figure 2: Genetic representation of solutions

In the end, to create more randomness, from each group randomly select two members and change their place. In the previous example, suppose in group 1 members 1 and 4, in group 2 members 2 and 4, in group 3 members 2 and 3, and finally in group 4 members 3 and 3 change their place. According to random mumers that produced for group 4, the change will not happen. The final grouping has been shown at Figure 5.
Mutation: in this operation, the algorithm randomly selects two groups and exchanges a member between them and random selection is used for choosing that member.

Before analyzing the algorithm, a brief discussion regarding the fitness functions will be introduced. One of the weaknesses in some of the previous introduced algorithms is that value of fitness function is not sensitive to the number of features of students or the number of students in a group. Another noteworthy point is that the growth of introduced fitness functions with respect to the number of features or students in the group is linearly. In this paper, new fitness functions that are introduced cover up those weaknesses. For first case, the maximum number of Inversion in an order vector is $\binom{n}{2}$ that $n$ is number of features (Kleinberg & Tardos, 2005), so it is sensitive to number of features of students. Also, the algorithm obtains intra-group fitness by calculating the number of Inversions between each pair of students in a group, so it is sensitive to the number of students in groups.

For the second weakness, if there are $n$ features for each student and in each group we have $m$ students, the growth of intra-fitness function will be $m\binom{n}{2}$ that is not linear. Because of this non-linearity any increment in the number of features or the number of students causes more variations in the intra-group and inter-group fitness functions value and can achieve better results than the linear mode.

3. Result

In this section we evaluate the proposed algorithm. Two criteria for the evaluation of the algorithm have considered; these criteria are intra-group fitness and inter-group fitness. The lesser the magnitude of the inter-group fitness the more similar the groups are to each other. The greater the value of the intra-group the more heterogeneous the members of the groups are to each other and the more robust group is formed. In the next sessions, Evaluation on synthetic data and real data has been performed.

3.1 Synthetic data

To evaluate the algorithms, 4 randomized dataset is created; their information is showed in the following table.

6. We generate a random value for each feature of each student between 0 and 9. Various aspects such as variation in the number of groups, members of each group, attributes, and students that should be grouped have been considered in producing randomized dataset process.

To evaluate the proposed algorithm, some famous clustering algorithms such as K-means, X-mans, Farthest-First, sIb, and Expectation Maximization algorithms have been modified in order to form equal size clusters. At first, these algorithms are run on the data, after forming clusters, so that each cluster will have a member in the final grouping. The proposed algorithm is performed for 10 times, and in each experiment 500 generations were produced by using proposed genetic operations and average of inter-fitness and intra-fitness group are considered as result.
Table 6: Data sets

<table>
<thead>
<tr>
<th>Name</th>
<th>#student</th>
<th>#Attribute</th>
<th>#Group</th>
<th>#member</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DS2</td>
<td>80</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>DS3</td>
<td>120</td>
<td>4</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>DS4</td>
<td>300</td>
<td>10</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

In Table 7, for groups that have been created by proposed algorithm and one of modified algorithms, intra-group and intra-group fitness is showed (result of DS1):

Table 7: Result of the Algorithm

<table>
<thead>
<tr>
<th></th>
<th>proposed Alg</th>
<th>modified Alg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inversions of Group 1</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>Inversions of Group 2</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td>Inversions of Group 3</td>
<td>38</td>
<td>44</td>
</tr>
<tr>
<td>Inversions of Group 4</td>
<td>40</td>
<td>34</td>
</tr>
<tr>
<td>Inter-group fitness</td>
<td>154</td>
<td>146</td>
</tr>
<tr>
<td>Intra-group fitness</td>
<td>0.87</td>
<td>9.53</td>
</tr>
</tbody>
</table>

In Table 7, can be seen the modified algorithm forms two groups that have a very good inter-group fitness (46 and 44), and this groups have highest number of Inversions but the other groups are less favourable. In addition intra-group fitness is not good. In the results obtained from proposed algorithm, not only the groups are similar to each other but also members within these groups have considerable discrepancy to each other than the results of the modified algorithm. Due to lack of space, in evaluation, only the inter-group and intra-group fitness is used for the comparison of the algorithms. The results are shown in Figures 6 to 13.

At first, we compare our algorithm with the random algorithm base on inter-group group.

**Figure 6**: DS1 - Inter-group fitness

**Figure 7**: DS2 - Inter-group fitness
As previously mentioned, the lesser the magnitude of the measure the more similar the groups are to each other and students are more evenly distributed in the groups. The proposed Algorithm has the lowest value at 4 synthetic databases. In result of DS1, It is remarkable to note that the proposed algorithm in 9 experiments finds optimal groups. In other words, the algorithm forms groups such that the standard deviation of number of Inversions in these groups was zero and it is significant improvement over the other algorithms.

In the following, we compare our algorithm with the random algorithm base on intra-fitness group.
The greater the value of the intra-group the more distinct the members of the groups are to each other. As shown in the above figures, results of proposed algorithm are not very good in this criterion. There are some significant notes in relation to the following observations:

1- In average case, results of our algorithm are comparable with other algorithm.
2- There is a trade-off between two mentioned criteria. Given the importance of the inter-group fitness to the intra-group fitness, the proposed algorithm is focused on that criterion.
3- According to significant improvement the proposed algorithm over the other mentioned algorithms in forming inter-heterogenous groups, this weakness is as minor as to be negligible.
As is clear from the results, with the increasing number of groups and individuals, the difference between inter-group fitness of the proposed algorithm and other algorithms is greater, and in intra-group fitness the proposed algorithm has considerable betterment over other algorithms, and also the value that reported for this criterion in our algorithm is close to the best result.

### 3.1 Real data

In this section, we evaluate our algorithm on real data.

Participants: The participants for this study were Master of Science students (N = 20) from the virtual faculty of Imam Reza (PBUH) International University, 55% were male and 45% female. The participants were all enrolled in the very large scale integrated circuits (VLSI) course.

Feature: Three categories features used for grouping of students, which include: demographic, learning style and presence information. Learning style is one of the most important factors in learning that affect all learning activities of students. In this study, the model that proposed by Felder & Silverman (1988) is used. To obtain learning style, students were asked to fill out a 44-item questionnaire. Another important factor in the learning environment is presence. Social presence is one of the aspects of presence that its effect on learning process has been studied by many researchers. Kim (2011) defines social presence as “the specific awareness of relations among the member s in a mediated communication environment and the degree of proximity and affiliation for med through it”. To extract information about social presence, 19-item questionnaire that is presented by Kim (2011) was used.

Grouping of students: The students were divided into 7 groups that each of them includes 2 or 3 students. 4 groups were formed by the method presented in this paper and 3 groups were formed by other methods.

Collaborative activity: Students were asked to analyse a circuit in their groups. Score of activity was divided into two parts. 40% score was assigned to analyse and answer the question and 60% score was assigned to interactions between students in the virtual learning environment in order to achieve the answer. All interactions must be carried out through the forums in virtual learning environment. A forum was designed so that member of each group could see the discussion and post their teammates, but the other groups were not able to view their forums. 4 days after putting question in e-learning environment are considered to groups submit their answer.

To evaluate the results of the two grouping method, the following criteria are compared.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Proposed method</th>
<th>Other methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of formed groups</td>
<td>4 groups</td>
<td>3 groups</td>
</tr>
<tr>
<td>Number of groups that find answer</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Number of created Discussion</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>Number of Added Post</td>
<td>122</td>
<td>12</td>
</tr>
<tr>
<td>Number of View Discussion</td>
<td>1601</td>
<td>510</td>
</tr>
<tr>
<td>Forum per Group</td>
<td>10</td>
<td>2.3</td>
</tr>
<tr>
<td>Post per Group</td>
<td>30.5</td>
<td>4</td>
</tr>
<tr>
<td>View Discussion per Group</td>
<td>400.2</td>
<td>170</td>
</tr>
<tr>
<td>Forum per Member</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>Post per Member</td>
<td>10.2</td>
<td>1.7</td>
</tr>
<tr>
<td>View Discussion per Member</td>
<td>133.4</td>
<td>72.9</td>
</tr>
</tbody>
</table>

In the table 8, the results are divided in to 3 different granularities. In the top view, groups that formed by the proposed algorithm have a significant improvement over other groups in all factors in this level - Number of View Discussion, Number of Added Post, and Number of created Discussion. In the second level comparison is done base upon activity of each groups. According to the criteria that are defined for this level - Discussion per Group, Post per Group, and View Discussion per Group - the superiority of the proposed algorithm is clear. In final granularity level, algorithms are compared on the level of individuals. The selected criteria for this level include Discussion per Member, Post per Member, and View Discussion per Member. The results show the significant advantage of the proposed algorithm to other algorithms.
4. Acknowledgment

The first author wishes to thank the virtual faculty of Imam Reza (PBUH) International University for providing necessary assistance to do this study, also likes to thank DR. Hoshmand, Learning Center Director and Mr. Nasiry, Online classes Director and Mr. akbarzadeh, teacher of VLSI course in Imam Reza (PBUH) International University.

5. Conclusion and Future Work

This study was to provide a new algorithm for grouping students. Criteria for evaluating the algorithm were inter-fitness and intra-fitness groups. In Section 3, the results were shown that the proposed algorithm has significant improvement over the other algorithms, especially in inter-group fitness.

The main features of the algorithm are:
- Considering the priority of features, unlike most previous algorithms that only the value of the features was considered.
- Lack of restrictions on the number of features for grouping of students.
- A significant improvement in forming inter-heterogenous and intra-homogenous groups.
- Use a non-linear fitness function with respect to the number of students and features of each student. If we have m students and each student has n features the fitness function growth is of the order m^n.
- The used fitness function is sensitive to number of students and number of features of each student.

As was shown in earlier sections, the performance of the proposed algorithm was much higher than the mentioned modified algorithms. This idea can also be used in other activities that need grouping. So the above algorithm can be considered as a general-purpose clustering algorithm.

Besides the positive features of the proposed algorithm, there are also weaknesses that include:
- Having problem with categorical data and their values that cannot be compared, for example colors or jobs.
- Having weakness in the Intra-group fitness measure.

As a future work, we will try to offer solutions to the above weaknesses. Another work is to apply our algorithm to other areas as a general-purpose clustering algorithm and compare it with other clustering algorithms.

References


A Brief Author Biography

M. Mahdi Barati Jozan is a Master of Science student at the Information Technology Department, School of Electrical and Computer Engineering, University of Tehran since 2010. He received his bachelor of science from the sadjad Institute for Educational Research in 2009. His research interests are in educational data mining, web-based collaborative learning, personalization in LMS and social presence in E-learning environment. Currently he is a member of laboratory of E-learning in university of Tehran, IRAN.

Fattaneh Taghiyareh is a faculty member at the Information Technology Department, School of Electrical and Computer Engineering, University of Tehran since 2001. She received her Ph.D. from the Tokyo Institute of Technology in 2000. Her research interests are in Technology Enhanced Learning, Multi-Agent Systems, and Human-Centered Computing Systems. Her current research involves the intersection of Web-Services, Personalization, and adaptive LMSs as well as applying ontology to semantic web. She was past-manager of the Department of Information Technology at the School of Electrical and Computer Engineering, and former Director of Technology Incubator, University of Tehran. She is a member of Editorial Board of International Journal of Information & Communication Technology. Recently she established eLearning laboratory with the mission of providing an standard view for eLearning materials and platforms. Currently she is supervising IT Foundation lab, eLearning lab, and Multi Agent Systems lab.

Copyright for articles published in this journal is retained by the authors, with first publication rights granted to the journal. By the appearance in this open access journal, articles are free to use with the required attribution. Users must contact the corresponding authors for any potential use of the article or its content that affects the authors’ copyright.