Group Learning, Student Clustering and Peer Mentorship

Doru Anastasiu Popescu[1], Ovidiu Domsa[2] and Nicolae Bold[3]

[1]University of Pitesti, Faculty of Mathematics and Computer Science, Romania, dopopan@gmail.com
[2]University "1 Decembrie 1918" University of Alba Iulia, Romania,
[3]University of Agronomic Sciences and Veterinary Medicine Bucharest, Faculty of Management, Economic Engineering in Agriculture and Rural Development, Slatina Branch, Romania, bold_nicolae@yahoo.com

Abstract: Adaptive learning is an important factor in the normal development of the education of the individual students. Taking into account the differences between the students in a classroom (or even employees in a company), things get complicated in the sense of the distribution of the information for the learners. Solutions to this issue are given to newer or upgraded methods of learning and assessment which are implemented nowadays and their results are quite encouraging. These methods form a system that can be very useful, a system which would comprise techniques such as student clustering, peer mentorship and group learning. In this matter, this paper presents a modality to gather these methods into a system that is applied in some courses in lower grades of high-school or faculty, with results regarding the quality of learning of the students.
Keywords: learning, cluster, group, mentoring.

1 Introduction

The learning process has multiple aspects and various situations appear within this process. One of the most problematic aspect of these is that, in the process of learning in clusters (or in groups), the levels of knowledge of different students may be distinct, which requires adapted methods of learning. Also, perspectives of teacher and students on the educational process are different. Comparative studies on this item can be found in papers [6] and [7].

Thus, frequent situations of this kind appear. For example, at the beginning of a course, the student may not have the minimal knowledge which will be used in the next period. In this matter, the teacher must make knowledge updates or introduce new concepts. These situations are frequent in cases of passing from a tuition cycle to a superior cycle or in cases of perfecting courses, where learners have different ages or come from different institutions.

For example, in the first year of high school, a class may be formed of 5 students from a school, 10 from another and so on. In the majority of the cases, the students that come from the same school have similar levels of knowledge, but it is a major possibility that students from different schools to be trained in a distinct way. The teacher may be constrained to reiterate some concepts. Due to this fact, some students can find the situation dull and unnecessary, being more interested in the applicative part of these concepts, while those who are unaware of these notions have to be trained differently. A useful solution to this problematic situation consists in the modality of work differentiated on clusters of students. This clusterization is made for the easiness of the educational process and does not imply any other aspect of any kind. Clusters are useful in situations as the one presented, as shown in papers [12] and [13].

To know exactly the situation at the start of the learning period, we use a system which provides as accurate as possible data related to levels of knowledge of students, in order to make the grouping on clusters of students. After this step, the customized data is taught to different clusters of students. This paper presents a model of creating clusters of students that can be later used in the process of learning, using different methods and techniques of education.

We can also say that a characteristic of this model is the adaptiveness, due to the fact that it follows the specific needs of the learner by reporting to the status of known notion (marked by 1) or unknown notion (marked by 0). More specific aspects of an adaptive system in education can be seen in [1].

Also, these kinds of methods of education can be considered as exploratory methods of education, whose characteristics are presented in paper [8]. We called them exploratory because their introduction in education is relatively novel (at least in Romania and surrounding regions) and they have to be studied harder in order for their full efficiency to be shown.

2 Student clustering: model description and algorithm

Based on some variables of the student, a general level of knowledge of a student can be established based on the school of origin and the results of an initial test made by the teacher.

In this matter, we will present a model which solves the problem presented in the introduction. In this model, we will use some notations. These are:

- N_{St}: the number of the students in the classroom;
- K: the number of institutions of origin of the students in the classroom;
- $SC_1, SC_2, \ldots, SC_K$: the institutions of origin of the students from the classroom;
- $ns_1, ns_2, \ldots, ns_K$: number of students that come from the institution $SC_1, SC_2, \ldots, SC_K$;
- $T_1, T_2, \ldots, T_K$: the set of unknown concepts of the students from the school $SC_1, SC_2, \ldots, SC_K$, found out after a set of questions answered at the start of the course; these variables can consist in two values, for a given notion W:

$$T_i = \begin{cases} 1, if\ W\ is\ known\ by\ the\ group\ ns_i \\ 0, otherwise \end{cases}, i = \overline{1, K}$$

- $W_1, W_2, \ldots, W_P$: the set of concepts that are contained in the curricula, where P is their number.

The sets ns and T are found in the relations presented in (1) and (2).

$$N_{St} = ns_1 + ns_2 + \cdots + ns_K \quad (1)$$

$$T = T_1 \cup T_2 \cup \ldots \cup T_K = \{0, 1\} \quad (2)$$

This level of knowledge (LK) for each student is measured as an average between the results on the test (also measured as a ratio and denoted by TR) and the degree of confidence (DC) of the institution of origin, as shown in (3).

$$LK = \frac{0.75 \times TR_i + 0.25 \times DC_j}{0.75 + 0.25}, i = \overline{1, N_{St}}, j = \overline{1, K} \quad (3)$$
$$TR_i = \frac{no\ of\ known\ concepts\ (W_z\ where\ T_y = 1)}{P}, z = \overline{1, P}, y = \overline{1, K}$$

TR is calculated as a ratio between the number of known concepts at the test and the total number of concepts. The number of known concepts is considered the number of concepts for which *T* for the student *i* which is part of the group ns_y is equal to 1.

The degree of confidence of the institution can be related in this way:

- for educational institutions (e.g. schools, faculties), the degree of confidence can be considered as a ratio between the number of graduates and the number of enrolled students;
- for companies, any indicator that shows the efficiency of the personnel (e.g., the proportion of submitted projects related to the total number of projects) can be taken into account.

In order to work in an easier and more attractive manner with the students, these will be grouped in two clusters M and P. The cluster M is formed of the students that must receive more detailed both practical and theoretical information related to certain notions established in the set T. The other cluster, P, will contain the students that have learned the theoretical aspects of the concept T_i from the set T and want to deepen and apply into practice this concept.

The reason for choosing two clusters instead of a larger number of clusters is given by the easier focus of the assessor on two clusters. Besides, related to a piece of information, the learner has two possible states: (a) know or (b) not know the information. Another reason is that the process of peer-mentoring would be clearer if two clusters are involved.

The established variables used in this model will have to be used in different methods in order to obtain the desired results.

Firstly, the students are tested for establishing the variables N_{St}, K, SC_1, SC_2, …, SC_K, ns_1, ns_2, …, ns_K, T_1, T_2, …, T_K. The test is built in order to ease the creation of the clusters. In this matter, the models of testing described in papers [2-4] or in paper [5] can be used. Based on certain keywords given at the beginning, the teacher can select some generated questions after certain restrictions are set. Then, LK for each student is calculated.

After this step, the clusters are formed based on LK. After certain set periods of time, the teacher can assess in a customized way the students using the same system of testing based on keywords. The final result would consist in reaching a balance between the two clusters, reducing the information gaps between those two clusters, assuring the performance of the cluster P and reaching a good level for the cluster M.

The clusters are formed dynamically. This means that:

- for every piece of information or notion W, the clusters can change;
- the whole group may be considered a specific type of group (e.g., a team that participate at a contest), then the clusters are formed analogous.

The graphical presentation of the system can be seen in Figure 1.

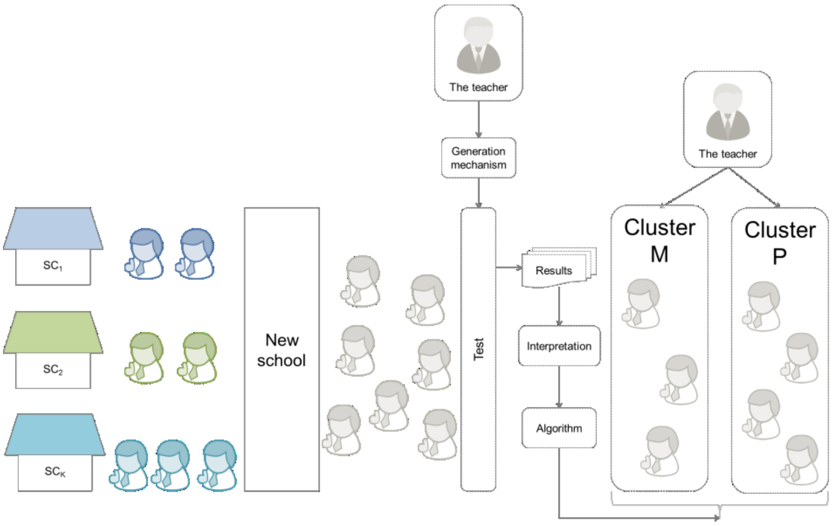

Fig. 1: The LMSC model proposed by the authors

In order to ease the work of the teacher, an algorithmic solution exists to solve this model. Practically, the model is transported in a mathematical language. The steps followed by this algorithm are presented next.

Step 1. Input data is read. Input data consist in the results to the test and their interpretation.

Step 2. The students are grouped in the format of ns_1, ns_2, …, ns_K based on the results on the initial test.

Step 3. The students from each group are verified if they know the concept w from the set T. If any student wishes to recapitulate the concept w, it is introduced in the group M.

Step 4. The cluster P is determined as a difference between the set of students and the cluster M.

Step 5. The two clusters are formed.

Practically, it can be summarized in the next pseudocode:

for each $w \in T$ *do*
    $M \leftarrow \Phi$
    for $i = 1, K$ *do*
        $M \leftarrow M \cup S_i$
        for each $j \in S_i$ *do*
            //if student i from S_i wishes to recapitulate the notion w, although
            //the results showed that the student knows the notion w

if j ∈ S_i wishes to recapitulate the notion w *then*

M ← M ∪ {j}

endif

endfor

endfor

P ← N_{St} \ M

write w, M, P

endfor

Practically, the algorithm can be schemed in the form shown in Figure 2.

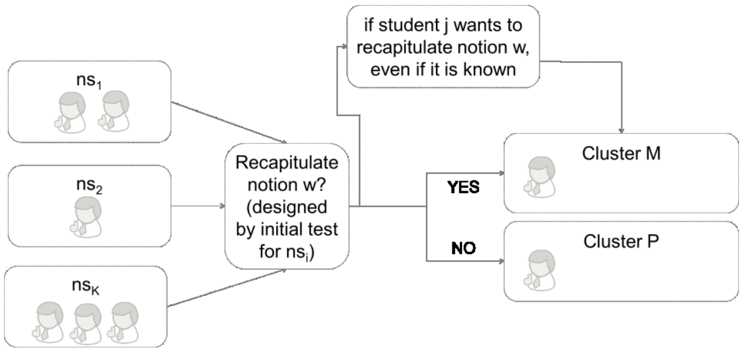

Fig. 2: The scheme of the algorithm

3 Group learning

Learning can be a dull activity if methods used are not interesting. That is, a method to improve learning efficiency is group learning. This method can have two forms:

- classroom group learning (CGL), made in the classroom; the students from the same cluster can relate to each other in the classroom;
- online-based group learning (OGL), which would continue indifferent to the time and which will be presented in the next lines.

OGL uses structures and methods such as databases, keywords or searching operations in order to update the actual information and offer valid information to persons which access the system.

The structure used in OGL is the piece of information (PI), which is formed of two components: a keyword T and the actual information I. Practically, PI is a pair of a keyword *kw* and a piece of information I (PI = [*kw*, I]). We will also use the variables:

- N: the number of pieces of information PI_i, i = 1, N;
- SPI: the PI searched by the user.

The model of functioning is a simple one and it will be presented in the next rows. We must mention that an initial database of PI is established in time. An illustrative scheme of the system can be seen in Figure 1.

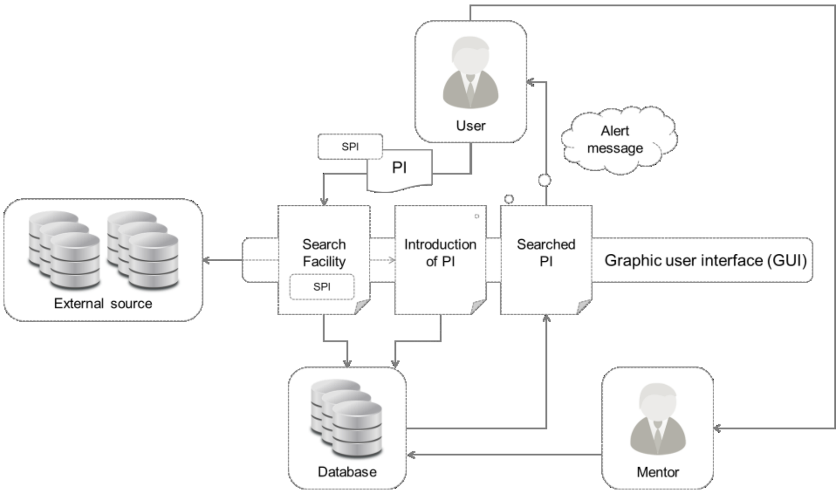

Fig. 1: The OGL method proposed by the authors

Firstly, a keyword desired to be searched in the system is set. Information about this word (whether it can be found in the database, how many PIs contain the keywords etc.) is output. Every PI that contains the keyword is output.

If at the previous phase the information was not enough, complete or the user does not find them useful, the new piece of information is search externally (on the Internet, on books, papers etc.), then these pieces of information are introduced in the system in the format provided by the system: [keyword, piece of information].

If the user does not find any information related to the searched keyword, then pieces of information are searched externally, then systematized, completed with examples and introduced in the system in order to be used by other members of the group.

Another method of introduction of pieces of information is the introduction of an admin or a mentor who has the responsibility to avoid the intrusion of incorrect or redundant information and to fill will competent information the system whether it is needed. An important facility of the system is represented by a component which notifies the users that a new piece of information was added in the system, whether is the case.

The modality of functioning of the system is presented in Figure 2.

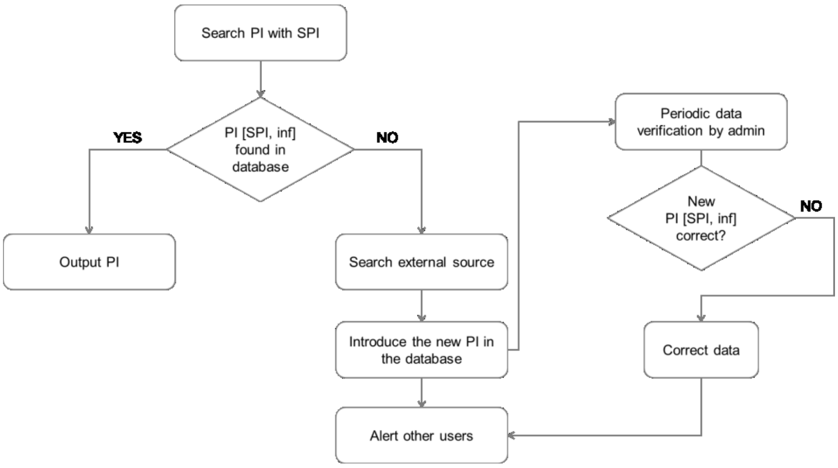

Fig. 2: The modality of the functioning of OGL

This method is very useful, because it extends the learning through online tools and, practically, it is an informal form of peer mentorship. The system can be continuously updated by students and every interested student can find information related to desired subjects, giving good results for both parties.

4 Peer mentorship: learning by teaching

As said at the end of section 3, OGL is kind of an informal peer mentorship. It is informal because peers interact only through online means and the formal rules of mentorship are not fully kept. Thus, a more formal type of mentorship should be used, in which students from cluster P are mentors and those from cluster M are mentees. This is made under the strict supervision of the teacher.

Peer-mentorship has been declared to have beneficial effects on learning. According to a study presented in paper [9], the peer mentorship led to the increase of submitted projects for junior level researchers, but also to the enhancement of the focus on more specific projects for the chosen specialty in case of each researcher.

Why is peer mentorship beneficial? Researches show that five instances of a peer make this method successful: connecting link, peer leader, learning coach, student advocate and trusted friend [10]. With all the risks that come from this method (vulnerability, non-acceptance of mentors etc.), this is known to be beneficial for the main reason that feedback from students received by the mentor is more honest and clearer than the feedback received directly by the teacher, according to the same source.

The peer mentoring in case of our system is simple: the students from the cluster P are trained to mentor those from cluster M. In this way, participants from both clusters would profit: those from cluster M teach in a different way the needed notions, while those from cluster M form a more solid knowledge base for the future notions, thus, they learn by teaching to others.

5 Unifying the three methods: system description

The three methods presented in the previous section can be united in a system that would consist in a powerful tool used in the classroom. Some methods of the system are already used in programming classes within the National College "Radu Greceanu" from Slatina (e.g., formation of clusters, peer mentors). Basically, an extended form of the system would behave in this way:

- the clusterization of the students is made at the beginning of a period;
- after the clusters are formed, the student can learn in groups or clusters (M and P), using both CGL and OGL;
- parallel with the group learning, the peer mentorship can be an important tool in the system.

A scheme of the system that would reunite the three methods presented above is presented in Figure 3.

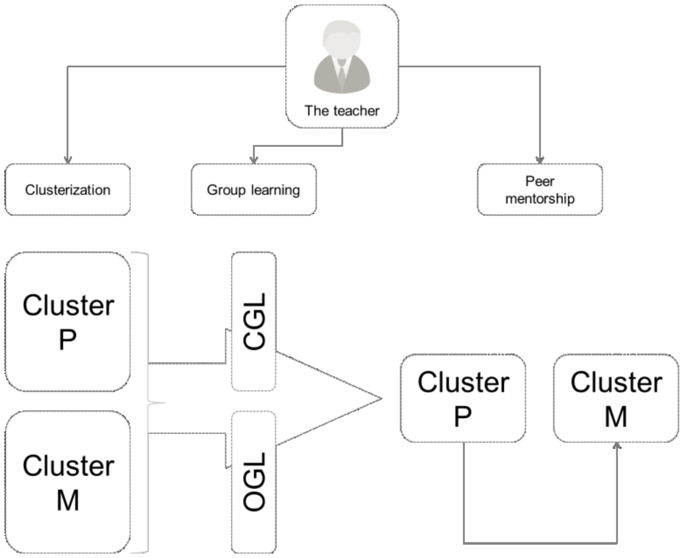

Fig. 3: Unifying the methods in the system

The methods can be used in complementarity or in different combinations and can led to good results, especially regarding the qualitative aspect of learning. This means that, within a classroom, the difference between the clusters is decreasing, creating a balance between the two clusters from the skills gain point of view. As student in high-school, one of the authors had been beneficiated by the benefits of a type of this system.

6 Results

The model is used in classrooms with mathematics-informatics profile in high schools, at an empirical level. We shall take an example of a high school classroom of

9^{th} grade at the Web Programming class. The class is made from 25 students. The number of institutions will be 7 (K=7), thus the number of groups (ns) will be 7. LKs for each student based on three notions W_1 (*the usage of Internet*), W_2 (*the tag elements in HTML*) and W_3 (*creation of simple websites*) are shown in Table 1, considering that all the students are aware of notion W_1. The DC of every institution is calculated by dividing the average of the means of all students from the respective institution to 100. The means are obtained at the exam of National Evaluation, which is given in the 8^{th} grade [11].

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
LK	0.72	0.72	0.72	0.47	0.72	0.72	0.72	0.47	0.72	0.72	0.72	0.72	0.43
Student	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*ALK*
LK	0.43	0.71	0.71	0.41	0.43	0.43	0.44	0.44	0.44	0.44	0.43	0.43	0.55

Table 1: Levels of knowledge for each student

The average LK (ALK) is calculated as an average of LKs of all 25 students. In our case, the average has the value of 0.55. This value divides the students in the two clusters in this way: if $LK_i < ALK$, then the student *i* is in the cluster M and if $LK_i > ALK$, the student *i* is in the cluster P. The groups depending on institutions are found in Table 2. The concept W_2 will be considered *the tag elements in HTML*. Some of them learned in the general school this notion and reached the next one W_3 (*creation of simple websites*).

Institution (SC_i)	DC	Number of students (ns_i)	T_i for W_2	Cluster P	Cluster M
"Eugen Ionescu" School	0.90	12	YES / 1	10	2
School no. 7	0.74	2	NO / 0	0	2
"Radu Greceanu" National College	0.84	2	YES / 1	2	0
Ianca School	0.66	1	NO / 0	0	1
School no. 3	0.72	2	NO / 0	0	2
"Constantin Brancoveanu" School	0.76	4	NO / 0	0	4
Sport-Scheduled High school	0.74	2	NO / 0	0	2
Total	0.77	25	-	12	13

Table 2: Clustering a classroom depending the notion W_2

The cluster P will be formed of 12 students, after the initial test is given and the results are interpreted. The two students coming from "Eugen Ionescu" school wanted to reiterate the notion T_1, thus will be considered in cluster M. The students from the cluster P will learn to develop simple websites, while the students from cluster M will learn the tags from HTML, in a future the two clusters being assessed accordingly to the knowledge gained, both clusters being brought to a good level of knowledge in a certain amount of time.

7 Conclusions

The LMSC shows an adaptive model of teaching, according to the needs of the student. Thus, the separation made in the two clusters has an educational purpose, the learning having a continuous flow for every individual student. We must also mention that for every notion T_i, clusters M and P change, thus these clusters are dynamic. This method is also based on the native talent of the teacher and of his managerial and organizational competences, his talent compensating the weaknesses of this method (determined by the projection of the model or the methods used for testing). A future work would be the integration of this model into a more complex model of assessment and teaching which is currently built.

8 References

[1] Popescu Elvira, *Adaptation Provisioning with respect to Learning Styles in a Web-Based Educational System: An Experimental Study*, Journal of Computer Assisted Learning, Vol. 26(4), Wiley, pp. 243-257, 2010

[2] Daniel Nijloveanu, Bold Nicolae, Bold Alina-Claudia, *A Hierarchical Model of Test Generation within a Battery of Tests*, International Conference on Virtual Learning, pp. 147-153, Timişoara, 2015.

[3] Doru Popescu Anastasiu, Nicolae Bold, Daniel Nijloveanu, *A method based on genetic algorithms for generating assessment tests used for learning*, 17th International Conference on Intelligent Text Processing and Computational Linguistics, April 3–9, 2016.

[4] Doru Anastasiu Popescu, Nicolae Bold, and Ovidiu Domşa, *Generating assessment tests with restrictions using genetic algorithms*, 12th IEEE International Conference on Control & Automation, Kathmandu, Nepal, June 1-3, 2016.

[5] Ovidiu Domşa, Nicolae Bold, *Generator of variants of tests using the same questions*, The 12th International Scientific Conference eLearning and Software for Education, Bucharest, April 21-22, 2016.

[6] Ştefănescu Valeriu, Ştefănescu Cornelia, Roşu Stoican Oana, *The influence of using ICT on the quality of learning*, International Conference on Virtual Learning – ICVL, pp.169-172, 2015

[7] Effie MacLellan, *Assessment for Learning: The differing perceptions of tutors and students*, Assessment & Evaluation in Higher Education Volume 26, Issue 4, pp. 307-318, 2001

[8] Sabine Niebuhr and Niels Pinkwart, *Usability requirements for exploratory learning environments: The case of educational argumentation systems*, 11th International Conference on Intelligent Tutoring Systems, June 15, 2012, Chania, Crete, Greece, pp. 64-71.

[9] Santucci, A. K., Lingler, J. H., Schimdt, K. L., Nolan, B. A. D., Thatcher, D., Polk, D. E., *Peer-Mentored Research Development Meeting: A Model for Successful Peer Mentoring Among Junior Level Researchers*, Academic Psychiatry, 32:6, November-December 2008, pp. 493-497.

[10] Colvin, J. W., Ashman, M., *Roles, Risks, and Benefits of Peer Mentoring Relationships in Higher Education, Mentoring & Tutoring: Partnership in Learning*, Vol. 18, No. 2, May 2010, pp. 121-134.

[11] Website of Ministry of National Education and Scientific Research, http://www.edu.ro/index.php/articles/weblinks/20408

[12] Doru Constantin, *The Independent Component Analysis of Clusterized Data on Principal Directions - Basis*, Proceedings of the 3rd Annual South East European Doctoral Student Conference, Thessaloniki, Greece, 26-27 June, 2008.

[13] Doru Constantin, Evan Hamzh, *Clustering Data by using PCA and ICA model*, "Finance. Economy. Strategy" Proceedings no. 5, 2011, Voronezh State Technological Academy, 2011, Rusia.