World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:3, No:6, 2009

Automatic Choice of Topics for Seminars by Clustering Students According to Their Profile

J.R. Quevedo, E. Montañés, J. Ranilla and A. Bahamonde

***Abstract*—**The new framework the Higher Education is immersed in involves a complete change in the way lecturers must teach and students must learn. Whereas the lecturer was the main character in traditional education, the essential goal now is to increase the students' participation in the process. Thus, one of the main tasks of lecturers in this new context is to design activities of different nature in order to encourage such participation. Seminars are one of the activities included in this environment. They are active sessions that enable going in depth into specific topics as support of other activities. They are characterized by some features such as favoring interaction between students and lecturers or improving their communication skills. Hence, planning and organizing strategic seminars is indeed a great challenge for lecturers with the aim of acquiring knowledge and abilities. This paper proposes a method using Artificial Intelligence techniques to obtain student profiles from their marks and preferences. The goal of building such profiles is twofold. First, it facilitates the task of splitting the students into different groups, each group with similar preferences and learning difficulties. Second, it makes it easy to select adequate topics to be a candidate for the seminars. The results obtained can be either a guarantee of what the lecturers could observe during the development of the course or a clue to reconsider new methodological strategies in certain topics.

***Keywords*—**artificial intelligence, clustering, organizing seminars, student profile

I. Introduction

THE European Space of Higher Education clearly involves a profound change in both learning and teaching processes. Some challenges have to be beaten to make the convergence really effective. Firstly, the subjects must be conceived as a set of abilities the students must achieve. Secondly, a mixture of methodologies and activities must be efficiently engaged to make students satisfactorily reach such abilities. Finally, several evaluation strategies must be carefully designed to obtain reliable information about the students' skills in terms of the abilities they must reach.

This paper focuses on the second goal. Great efforts are made to improve the quality of teaching methods at national levels, at local levels and at all levels in between. Through such evolutionary process of incremental adjustments, some elements could be implemented by the academic staff; some require decisions at course level, some at department level, some at faculty level and some at other institutional level [1]. Although all the decisions have an influence on what the most adequate teaching methods must be carried out, this paper places emphasis on what academic staff are able to do with the resources available and assuming other decisions taken at other levels. Our mission is to seek viable alternatives to direct our energies in a proactive way. So far, the students have spent most of the time being taught in a large group. They study several contents and a final evaluation takes place. Only individual tuitions have allowed lecturers to interact with students, but lecturers hardly ever have information about the whole learning process of each student. Before analyzing such alternatives, which lead to an improvement of learning, a clear answer to the question about what the learning outcomes lecturers wish for their students must be provided. Therefore, some of the main goals should be to develop the ability to use, to test and to generate ideas, information and evidence, the ability to plan and manage their own learning and the ability of interacting with other students to discuss topics, to acquire knowledge or to get projects off the ground. Seminars, tutorials or workshops are some methodologies that have demonstrated success in developing the abilities mentioned above [2]. Particularly, they enhance and encourage certain abilities that companies used to demand from university graduates, like dynamism, communication, critical skills or customer interaction, in short, abilities related to team work. These statements prove what Hague claims [3]: *The key role of the university lies in interaction; the most important element is interaction with tutors, tutorials and social activities.* Fortunately, nowadays, we have at one's disposal technologies provided by the Web 2.0 that facilitate interactivity. Google Groups is an example of a platform widely adopted to share material, to discuss topics and definitively to work in groups. Despite of that the traditional

J.R. Quevedo works in the Artificial Intelligence Center, Campus de Viesques, Gijón, Asturias, Spain (corresponding author to provide phone: +34 985-18-24-80; e-mail:quevedo@aic.uniovi.es).

E. Montañés works in the Computer Science Department, Campus de Viesques, Gijón, Asturias, Spain (e-mail:montaneselena@uniovi.es).

J. Ranilla works in the Computer Science Department, Campus de Viesques, Gijón, Asturias, Spain (e-mail:ranilla@uniovi.es).

A. Bahamonde works in the Artificial Intelligence Center, Campus de Viesques, Gijón, Asturias, Spain (e-mail:antonio@aic.uniovi.es).

The support of the University of Oviedo to the project entitled *La minería de datos como mecanismo de ayuda para la toma de decisiones en la actividad docente dentro del marco del Espacio Europeo de Educación Superior* is gratefully acknowledged.

face-to-face meeting must not be taken away.

II. SEMINARS AS LEARNING METHODS

A workshop or seminar could be defined as a scenario where both students and lecturers deal with a specific topic in depth through the interaction among all the participants [2]. Several ways of organizing them depend on the objectives, physical conditions, the structure and the context in which they are developed. However, the essential feature of these teaching methods is the interactivity, the exchange of experience, the criticism, the experimentation, the application, the dialogue, the discussion and the reflection among the participants.

This active and shared participation entails a careful organization and design, establishing the necessary conditions to promote its correct development. Besides, it is necessary to ensure that the number of participants is not too large to facilitate the interaction.

Other aspects to mention about this methodology are the focus on the activities the students carry out and an organization based on small groups.

Seminars could be seen as support for other methodologies, like lectures, or as the core of the university system. In a lecture, concepts are related, hypotheses are formulated, demonstrations are carried out, etc… but the main character is the lecturer, who manages the speech with his/her own abilities. Unlike lectures, in the seminars the center of attention is the activity carried out by the group. The main achievement is the common benefit from the contribution of all members of the group. The role of the lecturer is to promote the active participation of the students, to provide them the necessary material to facilitate the learning progress of the students and to establish beforehand a mechanism that provides information about the degree of achievement of their abilities.

There are a lot of strategies to tackle during a seminar. They range from case studies or simulations to the elaboration of specific projects. The key is to encourage communication and interaction and to activate the group. This may be performed by assigning roles, partial tasks or other alternatives such as brainstorming [4]. However, the organizational aspects this paper makes emphasis on are related to the distribution of the students rather than the methodologies, although one may depend on the other. Particularly, the goal is to organize them according to the topics to deal with, which will be selected depending on the interests and/or weaknesses of the students. Notice that the length and the periodicity are not taken into account, but the number of them is indeed considered.

III. THE METHOD

Previous section mentions the importance and advantages of organizing seminars. Such organization is not a straight-forward one; otherwise it must be carefully designed to guarantee their success. The method proposed in this paper automatically performs a distribution of the students according to their profiles and it helps to decide the topics the seminar must deal with. The purpose is to make the seminars as helpful as possible for the students to acquire the abilities required in a subject.

A. Computing the improvement a student could reach in a topic through the affinity of a student to the topic

A priori one could think that the topic a student must work in depth on is that in which he/she obtains lower marks. However, this would be true if all the topics in a subject had the same relevance. Let us consider that the scale to assign marks in all topics of a subject is the same and that it ranges from 0 to 10. Then, if the relevance of a topic A is 80% and that of the topic B is 20% and if a student whose marks in A and B are respectively 5 and 3, then he/she would obtain a final mark of 4.6. Hence, he/she would be better to participate in a seminar about topic A to improve the mark in this topic rather than about B to improve the mark in this other topic. In fact, obtaining the mark 10 in topic A leads to get a global mark of 8.6, whereas obtaining the mark 10 in topic B leads to reach a global mark of 6.0. Therefore, the affinity of a student to attend a seminar about a certain topic could be quantified depending on the improvement of the global mark if he/she is able to obtain the best mark in such topic. This measure relates the marks of the students in each topic and the relevance the teacher grants to each topic. Formally, let be n the number of topics and r_i the relevance of topic t_i in percentage such that

$$\sum_{i=1}^{n} r_i = 100\% \qquad (1)$$

Let be m the number students and $m_{ji} \in [0,10]$ the mark a student s_j has obtained in topic t_i. Hence, the affinity of student s_j to certain topic t_i is defined as follows

$$a_{ji} = r_i \frac{10 - m_{ji}}{10} \qquad (2)$$

This expression gives an idea of the improvement a student could reach if he/she comes to a seminar about certain topic.

B. Using preferences to group students in seminars

Once defined the affinity of a student to a certain topic, they will be grouped in seminars.

One could think of grouping the students depending just on the affinity defined in (2). But this could be incoherent. Let us illustrate it using an example. Suppose that the students s_1, s_2, s_3 and s_4 obtain from topics t_1 and t_2 the values a_{ij} shown in Table I.

TABLE I
EXAMPLE OF CLOSENESS

Student/Topic	t_1	t_2
s_1	1	0
s_2	0	1
s_3	5	4
s_4	4	5

Then, if one tries to group the students according to the closeness of Table I, the clusters that group the students with similar preferences will be $C_1=\{s_1, s_2\}$ and $C_2=\{s_3, s_4\}$. But, actually the topics the students would wish to assist to improve the marks are topic t_1 for $C_1=\{s_1, s_3\}$ and topic t_2 for $C_2=\{s_2, s_4\}$. Hence, this means that the adequate groups to form would be $C_1=\{s_1, s_3\}$ and $C_2=\{s_2, s_4\}$ instead of $C_1=\{s_1, s_2\}$ and $C_2=\{s_3, s_4\}$.

Therefore, the key is to cluster the students according to their preferences about the topics rather than according to the affinity. Now, the question is how the preferences are provided in order to cluster the students in the second way. A possible solution can be to deduce the preferences from the closeness, as s_i prefers topic t_j to t_k if and only if $a_{ij} \geq a_{ik}$. From now on, let us denote by $t_j > t_k$ the preference of topic t_j with regard to t_k. Such preference will be positive (+1) if $a_{ij} \geq a_{ik}$ and negative (-1) otherwise. Then, Table I turns into Table II.

TABLE II
EXAMPLE OF PREFERENCES

Student/Topic	$t_1 > t_2$
s_1	+1
s_2	-1
s_3	+1
s_4	-1

In this manner, trying to cluster the students according to the values of Table II, the desirable splitting is obtained.

Extending this approach to the general case, let be $t_1,\ldots,t_n$ the topics of a subject. Then, a table of as many columns as the number of combinations of n elements taken in two by two is built, that is $(n^2-n)/2$ columns of the type $t_j>t_k$ such that $j>k$. The number of rows is delimited by the number of students (m). Each cell in row i and column p will be +1 if the p^{th} preference of the form $t_j>t_k$ is positive and -1 otherwise, where the index i ranges from 1 to m and the index p ranges from 1 to $(n^2-n)/2$. The building of this table leads to obtain a representation of each student through a set of +1 and -1 values. Table III illustrates the general case, where each cell is denoted by $c(s_k, t_i>t_j)$.

TABLE III
REPRESENTATION OF STUDENTS

S/T	$t_1> t_2$	...	$t_1> t_n$	$t_2> t_3$	...	$t_2> t_n$	...
s_1	+1	…	-1	-1	…	+1	…
s_2	-1	…	-1	+1	…	-1	…
...	…	…	…	…	…	…	…
s_m	-1	…	-1	+1	…	+1	…

C. Clustering students from their preferences

From the preferences which have been already built, it is necessary to choose a clustering technique. There are a lot of them in the literature from different nature. But, it is hard to find one able to obtain a balance among the number of students per group. This is a desirable feature for the seminars in order to keep the teaching and learning quality and avoiding saturated groups. This paper proposes a modification of the agglomerative hierarchical clustering algorithm [5].

The original algorithm starts grouping each element (student in this case) in a different cluster, so at the beginning the number of clusters equals the number of students. Then, it looks for the two nearest clusters according to the Euclidean metric and joins them in one. The process continues by performing a hierarchy until all the students are in a unique cluster. In fact, this process builds a tree from leaves to the root and a different configuration of the clusters could be obtained from each node in the tree.

The modification proposed in this paper consists of adding a threshold that allows cutting the hierarchy in an adequate point. This paper adopts such threshold to limit the maximum number of students per cluster. An adequate choice can be $h=ceil\ (m/c)$, where m is the number of students, c is the number of clusters which has to be defined beforehand and which will be the seminars teachers must organize, and *ceil* is the function that for a real number r it returns the nearest integer q such that $q \geq r$. In this manner, the system obtains c groups with h or $h-1$ students.

Some possible situations could arise during this process. One of them occurs when a cluster already has the maximum number of students allowed (h students). In this case the process continues with the rest of the clusters. Another situation happens when the algorithm proposes to join two clusters and the sum of their students exceeds the value of h. In this case, the proposal involves splitting the resultant cluster to transfer some students from one of the clusters to another one in order to make the latter have h students. The election of one cluster to shift students to the other is based on the quality of the cluster which reaches the h students. Such quality is obtained by evaluating the average distance of all the students to the central point of the cluster. The central point of the cluster is the average of the elements of the cluster. Obviously, these averages are obtained according to the representation of the students proposed in section II.B. The cluster with lower distance is chosen as candidate to complete the h students, and hence it is removed from the process. The other cluster continues in the process, since it still has fewer students than h.

Applying that algorithm from the preferences of the students, the clusters obtained will contain students with similar preferences, which is the goal of the paper.

D. Choosing the topics for each seminar

Once all students are split into clusters according to their preferences, it is necessary to choose the topics closer to each group. The proposal now is to take into account the preferences of a topic rather that those of the students. Let us detail how to evaluate such preferences. For each topic t_i, let consider from Table III the columns of the form $t_i>t_j$ and of the form $t_k>t_i$. Notice that the former preferences mean that t_i is preferred to t_j and the latter preferences mean that t_k is preferred to t_i. Hence, the preferences of the former keep equal, but it is necessary to change the sign of the preferences of the latter to make all preferences being of the form $t_i > t_l$. In turn, for each cluster C_q let consider the rows in Table III referring to the students of cluster C_q. Now, just adding the values of all those preferences and students to obtain the fitness of topic t_i to cluster C_q.

TABLE IV
EXAMPLE OF REPRESENTATION OF STUDENTS

S/T	$t_1 > t_2$	$t_1 > t_3$	$t_1 > t_4$	$t_2 > t_3$	$t_2 > t_4$	$t_3 > t_4$
s_4	+1	+1	+1	-1	+1	+1
s_9	+1	-1	+1	+1	+1	-1
s_{20}	+1	+1	+1	-1	-1	-1
s_{36}	-1	-1	+1	+1	+1	-1

Let now show an example to clarify this calculus. Imagine a subject with topics t_1, t_2, t_3 and t_4 and a cluster C_q which contains students s_4, s_9, s_{20} y s_{36} with the preferences shown in Table IV. From them, the preferences p_i of each topic t_i when i ranges from 1 to 4 are obtained as follows

$$p_1 = \sum_{s \in C_q} \sum_{j=2,\dots,4} c(s, t_1 > t_j) = 9 \cdot (+1) + 3(-1) = 6$$

$$p_2 = \sum_{s \in C_q} \sum_{j=3,4} c(s, t_2 > t_j) - \sum_{s \in C_q} c(s, t_1 > t_2) = \dots$$

$$\dots = 5 \cdot (+1) + 3 \cdot (-1) - 3 \cdot (+1) - 1 \cdot (-1) = 0$$

$$p_3 = \sum_{s \in C_q} c(s, t_3 > t_4) - \sum_{s \in C_q} \sum_{j=1,2} c(s, t_j > t_3) = \dots$$

$$\dots = 1 \cdot (+1) + 3 \cdot (-1) - 4 \cdot (+1) - 4 \cdot (-1) = -2$$

$$p_4 = -\sum_{s \in C_q} \sum_{j=1,\dots,3} c(s, t_j > t_4) = -\big(8 \cdot (+1) + 4(-1)\big) = -4$$

Hence, the topic for cluster C_q is chosen to be t_1, since it reaches a higher score of preference.

IV. EXPERIMENTS

The data set has been taken from four courses related to databases of the Computer Science degree. They are subjects of different years whose main contents are grouped into several topics. The assessment involved the completion of several activities of case-solving under certain restrictions for the different topics. The four subjects include respectively 10, 9, 8 and 9 topics and 72, 55, 44 and 44 students.

Table V shows the relevance of each topic for every subject.

TABLE V
RELEVANCE OF EACH TOPIC (%)

Subject	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Sub1	25	25	10	10	10	10	2.5	2.5	2.5	2.5
Sub2	10	10	10	15	10	15	10	10	10	-
Sub3	15	15	10	15	15	10	10	10	-	-
Sub4	5	10	20	10	5	10	20	10	10	-

According to the availability of teachers, they reached the consensus of organizing five seminars per subject.

For the four subjects, the topics, the number of students, the average mark of the students before attending any seminar and the average of the maximum marks students are able to obtain if they take advantage of the seminar they attend are shown in Tables VI to IX.

TABLE VI
SEMINARS OF SUBJECT 1

Seminar	Topic	Students	Initial mark	Expected mark
Sem1	2	14	6.4	7.5
Sem2	1	15	3.3	5.6
Sem3	5	15	6.9	8.0
Sem4	1	14	1.3	3.8
Sem5	2	14	4.8	7.1

One can deduce that students of subject 1 have preferences and difficulties in topics 1, 2, and 5, especially in the two first ones. Besides, students are able to improve their marks between 1.1 and 2.5 points, in average.

TABLE VII
SEMINARS OF SUBJECT 2

Seminar	Topic	Students	Initial mark	Expected mark
Sem1	6	11	3.2	3.9
Sem2	6	11	4.0	4.8
Sem3	1	11	4.7	4.8
Sem4	4	11	2.4	3.1
Sem5	4	11	2.0	2.7

The marks in average of the students of subject 2 may only increase in about 0.7 or 0.8 points for the more conflictive topics 4 and 6, and the eleven students of seminar 3 would hardly improve the marks.

TABLE VIII
SEMINARS OF SUBJECT 3

Seminar	Topic	Students	Initial mark	Expected mark
Sem1	5	9	1.9	2.6
Sem2	5	9	2.2	2.8
Sem3	1	9	0.9	1.7
Sem4	8	9	3.5	3.9
Sem5	8	8	3.3	3.7

The improvement students could achieve in subject 3 (again in average) is not as much as in subject 1, since they could just increase their marks between 0.4 and 0.8 points. In this case, topics 5 and 8 seem to be more problematic.

There is no doubt that topic 7 is the most conflictive in subject 4, since 4 from 5 seminars deal with topic according to the system. In fact, students could improve their marks in between 1.3 and 2 points if they attend one of the seminars of topic 7, whereas the students that participate in seminar 1 about topic 3 might improve the marks in 1 point.

TABLE IX
SEMINARS OF SUBJECT 4

Seminar	Topic	Students	Initial mark	Expected mark
Sem1	3	8	5.3	6.3
Sem2	7	9	5.7	7.0
Sem3	7	9	4.9	6.6
Sem4	7	9	1.8	3.8
Sem5	7	9	4.1	6.0

Table X presents for each subject the average marks of the students originally (first column), given that an individual tutorial action is carried out for each student in the topic each one prefers (second column) and given that the students attend one of the seminars according to their preferences (third column).

TABLE X
SEMINARS OF SUBJECT 4

Subject	Original	Tutorship	Seminar
Sub1	4.61	6.50	6.39
Sub2	3.24	3.89	3.86
Sub3	2.32	3.00	2.92
Sub4	4.34	6.05	5.95

Clearly, attending seminars of subject 1 and 4 could lead to improve the marks of the students in more degrees than in subjects 2 and 3. In any case, although a personalized tutorial slightly improves the marks with regard to seminars of several students, such improvement is just 0.11, 0.03, 0.8 and 0.1 points, which is quite low comparing to the human effort of performing an individual tutorial action for every student.

V. CONCLUSION

The goal is to organize seminars according to the topics to cope with which will be selected according to the interests and/or weakness of the students. Teachers work under the hypothesis that clustering students into small groups to deal with topics of common interests would lead to improve the quality of certain abilities they do not do best. The contribution is an automatic system that helps teachers to split the students into seminars and to decide the best topics to tackle in them.

An extension of this work could be to include more information about the students to make the system plan the seminars more accurately.

REFERENCES

[1] T. Bourner and S. Flowers “Teaching and learning methods in higher education: A glimpse of the future”, Reflections on Higher Education, vol. 9, pp. 77-102, 1997.

[2] G. Brown and M. Atkins, “Effective Teaching in Higher Education”. London: Routledge, 1988.

[3] D. Hague, “Beyond Universities: A new republic of the intellect, London, Institute of Economic Affairs, 1991.

[4] A.U. Bolin and G.A. Neuman “Personality, process, and performance in interactive brainstorming groups” J. Bus. Psychol., Vol. 20, No. 4. (2006), pp. 565-585.

[5] S. C. Johnson "Hierarchical Clustering Schemes" Psychometrika, vol. 2, pp. 241-254, 1967.