World Transactions on Engineering and Technology Education
Vol.1, No.2, 2002

Using the fuzzy clustering algorithm for the allocation of students

Sani Susanto, Ignatius Suharto & Paulus Sukapto

Parahyangan Catholic University
Bandung, Indonesia

ABSTRACT: It is a common practice to allocate students from certain subjects into different classes based on their ID number. A new approach for this allocation activity is proposed in this article. This proposed approach, which applies the fuzzy clustering algorithm, is based on each student's achievement in the prerequisite subjects. Utilising this method, students with similar achievements are pooled into the same class. On the other hand, students with slightly different levels of achievement will be in a different class. The approach proposed in this article for student allocation is expected to improve the effectiveness of the daily learning process.

INTRODUCTION

There is a unique yet interesting phenomenon to be observed that usually happens at the beginning of each semester of the academic calendar: the allocation of students into classes. Consider that about 180 students are about to take ITI 372 Operational Research II, a compulsory subject in the Department of Industrial Engineering at Parahyangan Catholic University, Bandung, Indonesia, which students have to take in their 6^{th} semester. Those students must be allocated to, say, three classes. This raises the question: how are students to be allocated in those three classes? There is, in fact, a common way to do this at Parahyangan Catholic University. It is based on the students' ID number in ascending order, which are allocated into the first, second and third groups of 60 students, respectively, to classes A, B and C. And that is it!

There are four potential research questions related to the above-mentioned phenomenon that need to be answered:

1. Does this common method of allocating students to classes need to be changed?
2. What are the reason(s) if such a change is needed?
3. How can it be changed, if such a change is needed?
4. What are the factors to be considered if the faculty decide to make such a change?

QUALITY LEARNING-TEACHING ACTIVITY

The issues of quality are discussed in almost every agenda. Improving product quality is a compulsory task of all institutions, including educational institutions. However, quality is still considered an *enigmatic* concept [1]. The word *enigmatic* is defined as follows: *Something that is* enigmatic *is mysterious, puzzling, not obvious and difficult to understand* [2].

The concept of quality learning-teaching is also *enigmatic*. It seems that educators agree that improving quality is a worthy undertaking. However, there is no such agreement on the definition of quality and how to measure it [3].

The learning-teaching activity involves two parties, namely lecturers and students. Students are unique individuals: they differ from each other. As such, each student's level of mastery of the prerequisite subjects for ITI 372 Operational Research II will be unique as well. In this light, the authors propose a definition of the *quality teaching-learning process* as that which enables us to facilitate students with regard to the uniqueness of their level of mastery of the prerequisite subjects. If an agreement to this definition is reached, then the first and the second potential research questions mentioned above are implicitly answered. The common way of allocating students to classes has to be changed in order to enable educators to facilitate students with regard to the uniqueness of their level of mastery of the prerequisite subjects. The definition at present is too student-oriented; however, they are the University's main and direct customers.

Ideally, the quality learning-teaching activity demands the availability of one lecturer for one student per subject. Of course, such a demand is financially unfeasible, especially for a private university like Parahyangan Catholic University. Thus, a compromise solution, which is still based on quality learning-teaching, must be offered.

EFFORT TO ESTABLISH QUALITY LEARNING-TEACHING ACTIVITY

A quality learning-teaching activity can be established if serious attention is given to the allocation of students to classes. Mathematically, the process of allocating students to classes is called *clustering* and the class obtained is called a *cluster*.

Clustering of the 180 students for the subject ITI 372 Operational Research II, by definition, is based on students' level of mastery of the prerequisite subjects. In the case of ITI 372 Operational Research II, its prerequisite subjects are ITI 371 Operational Research I, AMA 214 Multivariable Calculus, AMA 213 Matrices and Vector Spaces, AMA 102 Calculus II (Integral Calculus) and AMA 101 Calculus I (Differential Calculus).

How are the students' levels of mastery of prerequisite subjects measured? It is currently based on the grade that students have obtained. At Parahyangan Catholic University, students' grades are divided into five categories, ie. A (for excellent), B (for good), C (for satisfactory), D (for less than satisfactory) and E (for fail). The grades A, B, C, D and E correspond to the scores 4, 3, 2, 1 and 0, respectively. Scores are the sole input for clustering. Successful clustering results in clusters of students with similar mastery levels of the prerequisite subjects.

In terms of mechanical way of speaking, successful *clustering* gathers students with a similar *initial velocity*. Since the *distance* students have to cover is the same (ie they have to undergo 13-16 weeks of learning-teaching activity), a lecturer must set an *acceleration* that is appropriate to the mastery level of his/her students. Thus, clustering students based on their mastery level of prerequisite subjects is the answer to the third potential research questions, the mechanism of which will be discussed in further below.

THE MECHANISM OF FUZZY CLUSTERING

As an illustration, a result from a small study of 20 students undertaking the subject ITI 372 Operational Research II in the Department of Industrial Engineering at Parahyangan Catholic University is presented. As stated before, the prerequisites for this subject are ITI 371 Operational Research I, AMA 214 Multivariable Calculus, AMA 213 Matrices and Vector Spaces, AMA 102 Calculus II (Integral Calculus) and AMA 101 Calculus I (Differential Calculus). Those 20 students' scores of these prerequisites, as presented in Table 1, are their *attributes*. The clustering of these 20 students to, say, three classes will be based only on these attributes.

It is known from Table 1 that the attribute for the 8^{th} student is represented as vector X_8, where

$$X_8 = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 2 \\ 2 \end{pmatrix} \qquad (1)$$

This means that the respective student received D (less than satisfactory) for AMA 102; C (satisfactory) for AMA 214, ITI 371 and AMA 213 and B for AMA 101

The *fuzzy clustering* technique, first introduced by James C. Bezdek in 1973, is used here to allocate these 20 attribute vectors X_1, X_2, …, X_{20} into three classes or clusters. The technique is called fuzzy clustering, since it gives a degree of membership to each cluster for each attribute vector. This means that the fuzzy clustering technique gives a *suitability level* to each student belonging to each of those three classes. A student having the highest *degree of membership* to a cluster is assigned to be a member of this cluster.

Once the fuzzy clustering technique received an input in terms of the 20 attribute vectors, it gives two types of vectors as its outputs. The first vector, called the degree of membership vector, is:

$$\mathbf{U_i} = \begin{pmatrix} u_{1i} \\ \vdots \\ u_{ki} \\ \vdots \\ u_{20,i} \end{pmatrix}, \; i = 1,2,3 \qquad (2)$$

of which its value in the k^{th} row represents the degree of membership (or the suitability level) of the k^{th} student to undertake the subject ITI 372 Operational Research II in the cluster (or class) i.

The second vector, called the cluster centre vector, is:

$$\mathbf{v_i} = \begin{pmatrix} v_{1i} \\ \vdots \\ v_{ji} \\ \vdots \\ v_{5i} \end{pmatrix}, \; i = 1,2,3 \qquad (3)$$

Here, v_{ji} represents the (weighted) average of students' grades achieved by students who belong to the cluster (or class) i for the j^{th} prerequisite subject for the subject ITI 372 Operational Research II. The value of each component in vector $\mathbf{v_i}$ plays an important role for the institution, since it generates information regarding:

- The students' levels of mastery of prerequisite subjects for ITI 372 Operational Research II in each class since, in this case, students are allocated into three classes wherein the mastery level can be described as high, normal or low.
- Guidance in assigning lecturers into classes.

Table 1: Attributes of 20 students based on their grade achieved for the prerequisite subjects of ITI 372 Operational Research II (Note: the score in student's attribute comes from such a conversion where A=4,B=3,C=2,D=1 and E=0).

Prerequisites	x_i (the attributes of the i^{th} student)																			
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
AMA 214	1	4	2	3	3	4	3	2	2	2	4	3	4	3	4	3	1	3	1	3
AMA 102	1	2	2	3	2	4	4	1	2	1	4	2	4	2	4	3	2	3	2	4
AMA 101	2	4	2	3	3	4	3	3	2	2	4	3	3	3	4	2	3	2	2	3
ITI 371	2	4	2	2	1	4	3	2	2	2	4	3	3	1	4	2	3	3	1	3
AMA 213	2	2	2	2	3	4	4	2	1	2	3	3	4	3	2	2	4	3	1	4

The component value of vectors $\mathbf{U_i}$ and $\mathbf{v_i}$ are obtained by solving the fuzzy clustering problem, which is basically a constrained optimisation problem in the form as follows:

$$\min J_2(\mathbf{U},\mathbf{v})=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^2\left\|\mathbf{x_k}-\mathbf{v}_i\right\|^2 \quad (4)$$

subject to:

$$\sum_{i=1}^{c} u_{ki}=1, \forall k=1,2,...,n \quad (5)$$

$$\sum_{k=1}^{n} u_{ki}>0, \forall i=1,2,...,c \quad (6)$$

A description of each item of notation, equation and inequation follows:

- The variable n represents the number of students registered as a participant of the subject ITI 372 Operational Research II, who will be allocated into c classes or clusters.
- The variable c represents the number of classes or clusters, the value of this variable can be determined by the institution policy.
- The matrix U = (uki)nxc consists of n rows and c columns, of which the element (uki) represents the degree of membership (or the suitability level) of the kth student to undertake the subject ITI 372 Operational Research II in the cluster (or class) i.
- The matrix v = (vji)mxc consists of m rows and c columns, of which the element represents the (weighted) average of students' grade achieved by students belong to the cluster (or class) i for the jth prerequisite subject for the subject ITI 372 Operational Research II.
- In extreme conditions, the value of the functional J2(U,v) in equation (4) is 0 (zero), which indicates the obtained clusters are ideal, since they consist of students with the same level of mastery for the prerequisite subjects for ITI 372 Operational Research II. Principally, the lower the value of J2(U,v) is, then the better the clustering process.
- The equation (5) requires that the total degree of membership of each student (in order to belong to the available c classes or clusters) is 1 or 100%.
- The inequation (6) requires that there must be at least one student in every class or cluster.

Bezdek has developed an algorithm, called the *Fuzzy C-means Algorithm* (FCM), to solve the fuzzy clustering problem ([4] in [5]). The application of the FCM algorithm is illustrated by a case described as data in Table 1. Table 2 gives the value of the elements of vector $\mathbf{U_i}$ (i = 1,2,3). As an illustration, the values in the 8th row of Table 2 can be interpreted as:

$$u_{81}=0.742 \quad u_{82}=0.210 \quad u_{83}=0.048$$

From those three values, the 8th student is the most suitable to be in class or *cluster* 1 (or class A), since he/she has the highest degree of membership to this class or *cluster* compared to the other two. By the same interpretation, the following class allocation was obtained for students participating in the subject Operational Research II as follows:

- The first class or class A consists of student numbers 1, 3, 8, 9, 10 and 19.
- The second class or class B consists of student numbers 4,5,12,14,16,17 and18.
- The second class or class C consists of student number 2,6,7,11,13,15 and 20.

Table 2: The values of u_{ki} of the 20 students participating in the subject ITI 372 Operational Research II to be allocated into three classes.

u_{ki}		i		
		1	2	3
k	1	0.840	0.122	0.038
	2	0.200	0.391	0.409
	3	0.852	0.127	0.021
	4	0.139	0.760	0.101
	5	0.267	0.621	0.112
	6	0.035	0.082	0.883
	7	0.066	0.191	0.743
	8	0.742	0.210	0.048
	9	0.775	0.178	0.047
	10	0.880	0.097	0.023
	11	0.038	0.092	0.870
	12	0.149	0.693	0.158
	13	0.044	0.127	0.829
	14	0.267	0.621	0.112
	15	0.103	0.223	0.674
	16	0.241	0.646	0.113
	17	0.361	0.408	0.231
	18	0.158	0.591	0.251
	19	0.729	0.201	0.070
	20	0.066	0.191	0.743

What have not thus far been discussed are the differences of students' levels of mastery in the prerequisite subjects for ITI 372 Operational Research II among those three classes or clusters. The answer to this question is given implicitly by interpreting the values of vector $\mathbf{v_i}$ (i =1,2,3), as presented in Table 3.

Table 3: The values of v_{ji} of the five prerequisite subjects of ITI 372 Operational Research II in the three classes.

v_{ji}		I		
		1	2	3
j	1	1.775	2.904	3.655
	2	1.568	2.492	3.842
	3	2.226	2.795	3.500
	4	1.913	2.220	3.498
	5	1.849	2.639	3.472

An interpretation of the values in the 5th row of Table 3 serve as an illustration:

$$v_{51}=1.849, \; v_{52}=2.639 \text{ and } v_{53}=3.472.$$

From those three values, it can be concluded that the highest mastery level for the 5th prerequisite subject for ITI 372 Operational Research, ie AMA 213 Matrices and Vector Spaces, is achieved by students of class C, followed by classes

B and C. The same order applies for the other four prerequisites. This information is important for the Department in assigning suitable lecturers to classes.

The next section will indicate some points to be considered if student allocation into classes is to be based on fuzzy clustering.

CONCLUSIONS AND POINTS FOR CONSIDERATION

There has been a significant paradigm shift in education: we are moving from a lecturer centred approach towards a learner centred learning process. Once an education institution decides to recruit students, it has to realise that students are unique.

One of the unique quantities to be considered is the students' mastery level of some prerequisite subjects. It is suggested that the institution allocate their students into classes based on this mastery level issue so that the institution is able to help students cope with the subjects they have to undertake in appropriate acceleration facilitated by suitable lecturers.

Points to be considered in deciding to allocate students based on their mastery level of prerequisite subjects are as follows:

- The socialisation process among students so that they better understand that the aim of this approach is to facilitate students' levels of coping with subjects in accordance to their acceleration rate in learning.
- Socialisation among lecturers for them to be open for a new learning-teaching situation in which the number of students in a class is no longer determined by the ratio between the number of students and the number of lecturers, but by the similarity in students' mastery level of some prerequisite subjects.
- Socialisation among the administrative staffs, especially those dealing with timetabling or scheduling.

REFERENCES

1. Sallis, E., *Total Quality Management in Education*. London: Kogan Page (1993).
2. *Collins COBUILD English Language Dictionary*. London: HarperCollins (1994).
3. Ibrahim, A., Current issues in engineering education quality. *Global J. of Engng. Educ.*, 3, **3**, 301-305 (1999).
4. Bezdek, J.C., Fuzzy Mathematics in Pattern Recognition. PhD Thesis, Cornell University (1973).
5. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press (1981).