A Method and a Program for the School Timetabling Problem

CASSILDA MARIA RIBEIRO[1]
LANA MARA RODRIGUES DOS SANTOS[2]
JOSÉ FRANCISCO FERREIRA RIBEIRO[1]
[1]Department of Computer Science
University of São Paulo – USP
Caixa postal 668, cep:13560-970 São Carlos, SP
BRAZIL

[2]Department of Mathematics
Federal University of Viçosa- UFV
36571-000 Viçosa, MG
BRAZIL

Abstract: In this paper we present a method for cluster analysis using graph colouring and an application to the school timetabling problem. This problem can be conceived as a partitioning of a set of subjects into a fixed number of schedules. The dissimilarities among subjects and all possible subject distributions are calculated. The solution is chosen according to an efficiency index that benefits the greatest possible number of students, but prioritises those with higher grades. The results achieved are of good quality, despite the studied problem being non-polynomial.

Key words: colouring, cluster analysis, school timetabling problem *CSCC'99 Proceedings*, Pages:2321-2324

1. Introduction

Consider the problem of partitioning a set X with *n* elements provided with a dissimilarity, in a fixed number, p, of classes such that the diameter between them is minimum. The dissimilarity between the elements of X is a value that measures the “distance” between any two objects of the set and, in this sense, is fundamental for the agglomeration of elements in the same set or for their separation into different sets. The problem of partitioning the set X, corresponds to the construction of a minimum threshold graph G_s, partial subgraph of G(X,A), which is p-colourable [6]. A partial subgraph G_s of G(X, A), where X is the set of vertices or elements and A the set of edges, is called threshold graph *s* when all of its edges possess dissimilarities greater than or equal to a fixed value or threshold fixed *s*

For p = 2, the polynomial algorithms of Guénoche [4], Hanson and Jaumard [7], Hubert [8], Leclerc [9] and Rao [12] enumerate all the possible partitions of X in two classes of minimum diameter. For p > 2, Hansen and Delattre [6] demonstrated that the problem is NP-complete [1]. They proposed an algorithm based on an optimal colouring method of a sequence of threshold graphs adapted from the algorithm of Brown [2] and constructed only partition of minimum diameter for each value of p. It should be noted, however, that the problem of optimal colouring is NP-complete [3].

The algorithm of Guénoche [5] enumerates all the partitions of a set in a fixed number of classes of minimum diameter. This algorithm was based on the colouring of a threshold graph in p colours, each colour defining a class. Several heuristics were used to approximate the diameter and the partitions of the minimum diameters were enumerated only in the final step.

In this article, an original way is proposed of constructing a matrix of dissimilarities between the elements of the set (that is, the set of subjects offered for which a timetable has to be established) and of obtaining an index to evaluate the efficiency of several solutions, since the algorithm used gives different solutions for the problem, corresponding to the distinct p-colourings of the associated graph. The method of colouring adopted is based on the algorithm of Guénoche [5]. We also present the computational implementation of the partitioning method developed with the results obtained for the accomplishment of registration.

2. Basic Concepts

Let X = $\{x_1, x_2, \cdots, x_n\}$ be a set of *n* objects and G(X, A) a weighted, undirected graph whose edges have weights provided by a matrix, D=$\{d_{ij}\}$, called a matrix of dissimilarities on X such that: d_{ij} = $d(x_i, x_j) \geq 0$. The dissimilarity in question measures the existing differences between the objects in relation to a series of attributes considered to be relevant.

Let $\{C_1, C_2, \cdots, C_p\}$ be a partition of X in *p* classes. To compare the constructed partitions on X, the following are defined:

- d(C_k) = *max*$\{d_{ij}$, for $x_i, x_j \in C_k\}$, diameter of a class (largest intra-class dissimilarity).
- d(P) = *max*$\{d(C_k)$, $k = 1, ..., p\}$, diameter of a partition (largest diameter of its classes).

Given an index of dissimilarity $s \geq 0$, the threshold graph G*s* (X, A*s*) is the partial subgraph of G(X; A) such that $\exists \ (x_i, x_j) \in A_s \Leftrightarrow d(x_i, x_j) > s$.

3. Lesson Timetable Problem

We tried to determine a lesson timetable for a defined school period. Various factors must be taken into consideration, such as: which subjects will be offered in the semester, how many students wish to take the subjects (for this, pre-registration processing is necessary), how many classes of each subject will be offered, the weekly timetable load, the periodicity of the subjects, whether the subject is obligatory or optional, the physical space necessary, whether the teacher will give more than one subject, etc. To simplify the problem and so that the result of the, example is easy to confirm, only three factors are considered:

1. The number of days per week on which there will be lessons, the number of lessons per day and the timetable load of the subjects.
2. The priority of one student in relation to another, to succeed in registering in a determined subject, is a function of the academic achievement index on a scale ranging from 1 to 10.
3. The school offers various courses. The total number of subjects offered in the period is *n*. In the pre-registration period the student can choose a number $r_i \leq n_i$ of subjects from those offered for their specific course (*i*).

4. Algorithm

The algorithm developed consists of the following fundamental steps:

Step 1: Determination of the matrix of dissimilarities.

Step 2: Application of the *p*-colouring algorithm based on the algorithm of Guénoche [5].

Step 3: Choice of the solution for the problem according to an efficiency index.

4.1 Determination of the Matrix of Dissimilarities

The dissimilarities between the elements of a set can be calculated in different ways, aiming to explore the specific characteristics of each problem. Witte [13] has done an extensive study on this subject.

In this application, the dissimilarity (or distance) between the subjects *i* and *j* is calculated in the original manner [11], using the following formula:

$$d(i, j) = \begin{cases} \sum_{S(i) \cap S(j)} priority \ \ of \ the \, student\,[k] & if \ \ i \neq j \\ \\ 0 & if \ \ i = j \end{cases}$$

where *S(i)* is the number of students pre-registered for the subject *i*

A variant for this formula consists of calculating the dissimilarity between the subjects two by two multiplying the sum above by $|S(i) \cap S(j)|$ / $|S(i) \cup S(j)|$. Other variants for the presented formula can be proposed, with the objective of eliminating or reducing possible distortions existing in the dissimilarity calculation in each problem.

4.2 Algorithm of p-colouring

To process and enumerate all the partitions of X in p classes of minimum diameter, an algorithm is applied which was developed from that of Guénoche [5] and consists of 3 stages:

Stage 1 - *Obtaining the p-colourable threshold graph* G_s : a heuristic method is used to determine the threshold *s*, approximation greater than *d*(P) such that G_s is p-colourable. The threshold *s* is determined by means of a sequence of dichotomous subdivisions of the dissimilarity variation interval. For each trial, a sequential colouring algorithm is used.

Stage 2 - *Enumeration of all the p-colourings of* G_s: A *p*-colourable graph G_s is prepared, with an order on the vertices (obtained in stage 1). The first *p* vertices are coloured with different colours. The other vertices are coloured in all possible ways. If a vertex is adjacent to coloured vertices, these colours will not be able to be used to colour it. In this way, all the partitions in *p* classes of diameter less than *s* are obtained. This algorithm stage is of a complexity proportional to the number of vertices and to the number of partitions in p colours of graph G_s

Stage 3 - *Elimination of the partitions which are not of minimum diameter:* The edges in descending order of the dissimilarity values starting from *s* are considered. An edge may be inserted in the graph while a possible partition exists, that is, where each edge connects two vertices of different classes. Each inserted edge eliminates some partitions previously obtained, if they had not already been eliminated. The first edge which cannot be eliminated has a value equal to *d* (P), which is the largest value of intra-class dissimilarity. The remaining partitions are of minimum diameter (for a fixed number of classes).

4.3 Efficiency Index to Aid Decision Taking

Among the different partitions found by the *p*-colouring algorithm, one has to be chosen for implementation. To assist the process of decision-taking, we propose the calculation of an efficiency index corresponding to each of the given solutions. The proposed formula is given by:

$$\text{Efficiency} = \left(\sum_{i=1}^{nrequests} x_i . na_i \,/\, nstudents \right).100$$

where:

x_i - factor of consideration associated with attending to *i* requests of the student

na_i - number of students attended to in *i* requests, $i = 1, .., nrequests$

nrequests - number of subjects in which the student can pre-register

nstudents - total number of registered students

In the example, described below, the consideration factor chosen, associated with the number of students attended to in *i* requests is :

$x_i = i/nrequests$

5. Example

5.1 Data of the Problem

- The school has 3 different courses, a total of 12 subjects are being offered in the semester and 4 sessions will be distributed.
- The processing of 48 students' pre-registration is done. In the pre-registration, the student must choose 4 subjects to study in the period.
- The priority of each student is pre-established by a specific routine, based on, for example, academic achievement.
- In the case of a student (*i*) who has not been able to take all the subjects chosen in pre-registration, the algorithm offers the options of the subjects which he may register in later.

20,000 examples were generated randomly. The graph produced in Figure 4 depicts the distribution, as a function of efficiency, of these hypothetical timetables using the algorithm.

Figure 4: Distribution of the 20,000 timetables generated as a function of Efficiency

The solution of the test problem illustrates in a detailed manner the steps of the procedure. The relationship (subject x course) is given in Table 1.

Course	**Subject**
Course(1)	1,...,8
Course(2)	1,..,3,5,9,...,12
Course(3)	1,5,...,10

Table 1: Courses and Subjects

The result of the processing of pre-registration by 48 students, as well as the course and the priority of each student, are provided in Table 2.

Pre-registration			
Students	Priority	Course	Subjects