

## Clustering Approach to Examination Scheduling

Tang Van To, Soe San Win

Faculty of Science and Technology

Assumption University

Bangkok, 10240, Thailand

e-mails: tvto@scitech.au.edu, g4719557@scitech.au.edu

**Abstract**— The difficulty in examination scheduling is to draw an examination timetable by taking a number of different constraints into account. This paper attempts to optimize two major constraints – to minimize the examination time conflicts and the number of consecutive examinations for a student in a day. Clustering method is applied by analyzing exam enrollments and then groups students into clusters so that students within a cluster are having more similar subject combination to each others than those in another clusters. The purpose of clustering is to effectively arrange the order of exams to be allocated into a suitable exam period. Since exams are scheduled by clusters, it solves the exam conflicts for the students within the same cluster whose are closely related to each other.

**Keywords**—*Examination Scheduling; Data Mining; Clustering; Heuristic*

### I. INTRODUCTION

Scheduling the examinations is a problematic and time consuming task for the institutions which allow students to choose any combination of desire subjects in each semester. Examination scheduling problem has been an active research for last few decades. As referred in [1], Welsh and Powell built the bridge between graph coloring and timetabling; Burke, Newall and Weare developed graph based heuristic color degree and largest degree to solve examination scheduling; fuzzy logic was employed by Asmuni et al to schedule the exams based on graph coloring. Salazer and Oakford [2] presented school scheduling by using graph formulation which influenced the early years of solving the timetabling problem. Again, Zhaohui and Lim [3] used greedy heuristic to solve exam scheduling of the National University of Singapore. Wong, Cote and Gely [4] developed the final exam timetable generator by applying Genetic Algorithm to the data of Université du Québec, Canada. Various approaches of Genetic Algorithm are used to solve examination scheduling for instance in [5] using population as the past experience.

This paper uses knowledge discovery and data mining technique, particularly data clustering, which is relatively new in examination scheduling methodologies. The first part of the proposed methodology is to group students into clusters according to the similarity of subjects they have learned. Agglomerative hierarchical clustering is used to create clusters from flat data into the patterns that help the second portion of the algorithm which actually generates the timetable. When the number of timeslots and available

number of seats per timeslot is predefined, the algorithm will provide only less-conflict solution rather than no-conflict one in the worst case.

### II. EXAMINATION SCHEDULING PROBLEM

The examination scheduling problem generally states that the assignment of a given number of exams into a limited number of exam periods in such a way that no student is required to sit more than one examination at the same time. The conflict between one subject and other occurred because every student is choosing subjects arbitrarily.

There are critical constraints (hard constraints) and desire constraints (soft constraints) in the scheduling. Only the most crucial conditions are used as hard constraints in this paper as follow.

1. Any two exams in which there are common students cannot be allocated at the same timeslot as the most important constraint.
2. In order to give students more study time, exams are spread out by minimizing the number of consecutive exams for a student in a day.
3. Total number of students in each period must not exceed the number of available seats. Algorithms always ensure not to violate this constraint because it is not possible at all to arrange an exam over the seat capacity.

Unlike critical hard constraints which are essential, soft constraints are only desirable. Following soft constraints are taken into account in this algorithm.

1. Total number of examination timeslots for all subjects should be minimal.
2. Examinations with more students should be scheduled in earlier exam period in order to give teachers more time for grading [3].

Although its satisfaction can measure how good the scheduling is, it is not always possible to have a solution that does not violate all soft constraints.

### III. CLUSTERING

Hierarchical clustering concept has been applied to solve the examination time conflict problem. Clustering is one of the fundamental tasks of data mining that generally finds the similarity between data without any historical knowledge, and groups the most similar data into clusters. A cluster, in turn, can be defined as a collection of data that are similar to one another within the same cluster. There are many methods