THE IMPLEMENTATION OF PARTIAL GRAPH HEURISTIC WITH IMPROVEMENT METHOD TO SOLVE THE UMP EXAMINATION TIMETABLING PROBLEM

LEE JIAN HAO

THESIS SUBMITTED IN FULFILMENT OF THE DEGREE OF COMPUTER SCIENCE (SOFTWARE ENGINEERING)

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING

UNIVERSITI MALAYSIA PAHANG

2013

ABSTRACT

The examination timetabling problem has attracted the interested of many researchers over the years. However, this problem is difficult to solve due to the lack of benchmark dataset and many constraints that need to be satisfied in examination timetabling problem. Currently, Universiti Malaysia Pahang (UMP) use proprietary system to generate the examination timetable but the weakness of this system is unable to define the quality of solution because having no evaluation function. Other than that, the UMP examination timetabling problem consist unique constraints such as distance penalty and split penalty. Having all of these constraints had made the task to solve examination timetabling problem becomes more challenging. To produce examination timetable, all of the exams need to be scheduled into timetable while satisfying the hard constraint and soft constraint. The timetable result should have the minimum penalty value in terms of spread exams, split rooms and distance between rooms. Therefore, the technique partial graph heuristic with hill climbing method should be implemented to solve UMP examination timetabling problem. The graph heuristic method will partially schedule the exam and then improved by hill climbing method. This process will be continued until finish scheduled all of the exams. By using this technique, the solution of timetable result can comply all of the constraint and has a better quality of solution compared to the result of current examination system.

Keywords — graph heuristic, hill climbing, penalty value

ABSTRAK

Masalah jadual waktu peperiksaan telah menarik minat ramai penyelidik selama ini. Walau bagaimanapun, masalah ini sukar untuk diselesaikan kerana kekurangan dataset dan pelbagai jenis kekangan yang perlu dipenuhi dalam masalah jadual waktu peperiksaan. Pada masa ini, Universiti Malaysia Pahang (UMP) menggunakan sistem sendiri untuk menjana jadual peperiksaan tetapi sistem ini mempunyai kelemahannya iaitu tidak dapat menilai kualiti jadual peperiksaan disebabkan kekurangan fungsi penilaian. Selain itu, masalah jadual waktu peperiksaan mempunyai kekangan yang unik seperti jarak antara bilik peperiksaan dan pemecahan bilik peperiksaan. Semua kekangan ini telah menyebabkan tugas penyelesaian masalah jadual waktu peperiksaan menjadi semakin mencabar. Untuk menjana jadual waktu peperiksaan, semua peperiksaan perlu dimasukkan ke waktu dan bilik yang sesuai dengan memenuhi semua kekangan yang ada. Jadual waktu peperiksaan yang dijana haruslah berkualiti dan memenuhi semua kekangan yang ada. Oleh itu, teknik Graph Heuristic bersama Hill Climbing haruslah digunakan untuk menyelesaikan masalah jadual waktu peperiksaan UMP. Teknik Graph Heuristic akan menjadual sebahagian peperiksaan ke waktu dan bilik yang sesuai and seterusnya menggunakan teknik Hill Climbing untuk menjadual semula peperiksaan tersebut ke waktu dan bilik lain yang sesuai. Proses ini akan berulang sehingga semua peperiksaan habis dijadualkan. Dengan penggunaan kedua-dua teknik ini, sebuah jadual waktu peperiksaan yang lebih berkualiti berbanding dengan jadual waktu peperiksaan sistem asal dapat dijana dan jadual waktu baru ini dapat memenuhi semua kekangan yang ada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

UMP - Universiti Malaysia Pahang

GH - Graph Heuristic Method

GA - Genetic Algorithm

HC - Hill Climbing Method

ETP - Examination Timetabling problem

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Chapter 1 will briefly explains about the overview of the project to let us understand deeply about this project. This chapter comprises five sections which are introduction, problem statement, objectives, scopes and thesis organization. Introduction gives an overview of the project. In problem statement, we explain the problem face that motivates us to carry out the project and the objectives explains about the research's goals. Scope describes the area of the research and the restriction for users and project. Finally in thesis organization, we illustrate the flows of each chapter in this thesis.

1.1 Background of Study

The timetabling activity occurs periodically throughout the year and every University need to produce course timetable and examination timetable. Both scheduling problems involve in arranging the courses to satisfy variety of constraints and try as much as possible to satisfy all the objectives or aim in space time (Dimopoulou and Miliotis, 1998). The main difference between them is examination timetabling allows multiple exams to be carried out in the same rooms while it is not possible for course timetabling. Students also are free to enroll to the courses they want but it is compulsory for them to take all exam based on their chosen course (Kahar, 2011). In this research we are going to investigate exam timetabling problem only.

The ETP include two types of constraints, which are the hard and soft constraints. The differences between both are hard constraint refers to those requirements which are strongly enforced and cannot be break while the soft constraint are requirement that should be satisfy as much possible and it is use as a guidelines for quality of the timetable. According to Ayob, Abdullah and Malik (2007), it is not possible to fulfill all of the soft constraints.

The UMP examination timetabling problems are considered as capacitated problem that consider room capacity constraint (Kahar and Kendall, 2010). The capacitated problem which is more resemble the real world situation received less attention from researches due to the lack of benchmark dataset. This capacitated ETP also are more complicated compared to those dataset that does not considered room capacity such as Toronto datasets (Kahar and Kendall, 2010).

The UMP examination problems include other constraints that are distance of room, number of room an exam being split across and spreading (Kahar and Kendall, 2010). The current proprietary software unable to fulfills all the hard constraint which might affect the quality of timetable (Kahar and Kendall, 2010). Additionally, student prefer to have an evenly spread of exam that will allow them to have a better time in preparing themselves between 2 exam (Kahar, 2011).

1.2 Problem Statement

The examination timetabling problem has attracted the interested of many researchers over the years. The reason because, of the complexity of the problem and the demand for better timetable to suit with the nature of the business and satisfaction of customer. The UMP used proprietary system to generate the timetable but this system unable to define the quality of the solution because having no evaluation function. Additionally, the UMP examination problem contains unique constraint different from other constraint reported in the literature (e.g. distance penalty if an exam being assign to

multiple rooms, etc.). Furthermore, the UMP examination problem also insists the exam must be in same building if the exam is being split in multiple rooms. All of these constraints had make it more challenging in producing clash free timetable and producing high quality timetable (i.e. spreading, etc.) for the student.

This motivates us to investigate the UMP examination timetabling problem due it unique timetable constraint and the inefficient of the proprietary system in producing the timetable.

1.3 Objective

The objectives of the research are as follows:

- To study the UMP examination timetabling problem.
- To implement the partial graph heuristic with improvement method in UMP examination timetable problem.
- To compare the results of partial graph heuristic with improvement method with the UMP proprietary software result.

1.4 Scope

The scopes of the work are as follows:

- The investigation concentrates on UMP examination dataset at semester 1-2007/08.
- Investigation focuses on partial graph heuristics with improvement method to solve the UMP examination timetabling problem. The improvement method used is the Hill-Climbing

1.5 Thesis Organization

This thesis includes of six chapters. Initially, Chapter 1 explains about the introduction of the research, problem statement, research objectives and lastly research scopes.

Chapter 2 describes the current UMP examination timetabling problem and also the compared the problem with the previous researches which can be found in the scientific literature.

Chapter 3 explains the methodology of the research that will be used in this research.

Chapter 4 explains the design and implementation of the system which will include some explanation about coding and algorithm parts of system.

Chapter 5 describes the result of system produced and investigates which sorting way can produce better result in UMP examination timetable problem with the use of improvement method.

Chapter 6 discusses about the conclusion of the research.

CHAPTER II

LITERATURE REVIEW

This chapter briefly describes about the common timetabling problem in real world, all of the constraints which need to be considered when solve the exam timetabling problem and lastly the common techniques reported in the literature which had been used recently to solve the timetabling problem.

Lastly, we will conclude the methodologies that can be used to solve examination timetabling problem at the academic institution and we will present the conclusion from it.

2.1 Common Timetabling Problem

Timetabling problem is defined as a problem with R, a finite set of resources; T, a finite of times; C, a finite set of constraints and M, a finite set of meetings as their parameters when solving the timetabling problem (Burke, Kingston and de Werra, 2004). Timetabling also can be determined as combinatorial problem and it is usually be defined as assigning the activities to a finite number of locations and timeslot by fulfilling the constraints and functional objectives as possible (Ahmad, Gourgand and Caux, 2012). Practical timetabling can be categorized into few which are employee timetabling, educational timetabling, communication timetabling, sport timetabling and transport timetabling (Ahmad, Gourgand and Caux, 2012).

The research activity in the timetabling area had increased simultaneously since this few years due to the introduction of new requirements or constraints. Other than that, the end user who insisted for better solutions in timetabling problem also attract researches to carry out research in the timetabling area (Burke and Ross, 1996). There are two types of requirement or constraints in timetabling problem which are hard constraint and soft constraint. In timetabling problem, hard constraints cannot be violated while soft constraints are requirement that need to satisfy it as much as possible (Kahar and Kendal, 2010). The soft constraint is used to measure the solution quality of the timetable and the overall aim is to minimize the total penalty value which related to the satisfaction of the user (Kahar, 2011).

One of the most popular timetabling problems which had been studied is the educational timetabling problem and usually this time-consuming task happens periodically in all institutions. Examples of educational timetabling are course timetabling, school timetabling, exam timetabling, classroom assignment and lastly is the faculty timetabling (Qu, Burke, McCollum, Merlot and Lee. 2008). In next section, we will describe about the timetabling problems at University.

2.2 Timetabling Problems at University

University timetabling problems can be categorized into examination and course timetabling problems (Dimopoulou and Miliotis, 1998). Both of the timetabling problems have the same characteristics and the main problem between them can be assumed similar (Burke, Kingston and de Werra, 2004).

2.2.1 Course Timetabling

The course timetabling problem can be determined as a set of offered courses and schedule these courses to timeslots and classrooms with the conditions that all student, teacher or classroom do not involved more than once per period and the number of students which being assigned in a classroom must be less or equal to the room capacities (Abdullah, 2006).

Similar to other timetabling problem, course timetabling problem also has its own constraints which can be classify as hard and soft constraints. Examples of constraints for the course timetabling problems are as follows (Abdullah, 2006):

- Hard Constraints
    1. A teacher and student should not be assigned to more than one place at the same timeslot.
    2. Each timeslot can only have one course with only one classroom.
    3. Capacity of each classroom must be able to accommodate the total number of students that participating the course at a certain timeslot.
    4. The classroom should have suitable equipment and features to fulfill the course that being assigned in.

- Soft Constraints
    1. Each student should have more than one course per day.
    2. A student should be avoided attending two or more consecutive courses on a day.
    3. Each student should be avoided from being scheduled to participate in a course that is being allocated to the final timeslot of the day.

2.2.2 Examination Timetabling

The growth in the number of student enrolments, combined degree courses and broad variety of courses has make the task of developing examination timetabling becomes more complex (Ayob, Abdullah and Malik, 2007). Besides that, the difficulty level of constructing the examination timetable depends on the level of freedom of choice on students to select their own courses (Kahar, 2011). The examination timetabling problem can be defined as allocating a set of exams into finite rooms and period slots with the aim in fulfilling a set of constraints (Qu, Burke, McCollum, Merlot and Lee. 2008).

The solution of the exam timetable should satisfy all parties and many factors need to be considered such as no clashes for the students and adequate gaps for each exam papers (McCollum, 2007). The examples of hard and soft constraints for the examination timetabling problems are shown as below (Qu, Burke, McCollum, Merlot and Lee. 2008):

- Hard Constraints
    1. There are no collaborative resources (e.g. Students) in exams being assigned simultaneously.
    2. There are sufficient resources to be used for examination timetable (e.g. the number of students that take the exam must be less or equal to the room capacity of that exam.).

- Soft Constraints
    1. The exams should not in any consecutive period slots or days and should spread as much as possible.
    2. The exams in same group must be held at the same period, day or at same place.
    3. Consecutive all of the exams.
    4. Every exam should be scheduled first or the largest exams should be scheduled at early time compared to others small exams.
    5. Satisfied all prior exams condition.

6. Every timeslot should limit the numbers of exams and students.
7. Some specific exams must be place in certain timeslots as request by the school.
8. Located conflicting exams on the same day as near as possible.
9. Might be able to split the exams over nearby or similar places.
10. Combined the exams with the same length into same room as long as got sufficient room capacity for students.
11. Resource requirements should be fulfill as many as possible.

The introduction of new constraints (e.g. Muslim students request do not have any exams at Fridays) also make the generation of examination timetabling becomes harder (McCollum, 2007).

There are un-capacitated and capacitated problem in the examination timetabling problem. Room capacities are not considered for un-capacitated examination timetabling problem while the room capacities need to be counted as hard constraint in the capacitated examination timetabling problem. Therefore, the capacitated examination timetabling problems are much more difficult to be solved as it requires more comprehensive data (Kahar and Kendall, 2010).

2.2.3 Similarity and Differences between Course and Examination Timetabling

Examination timetabling and course timetabling problems in university involved in allocating the students to sit only once (course or exam) in each timeslots. Students are strongly prohibited to sit more than two (course or exam) in each timeslots. However, there are still some differences between both timetabling problems. The main different between the examination and course timetabling are examination timetabling allows multiple exams to be carried out in a rooms while it is not possible for course timetabling. Students also are

free to enroll to the course they want but it is necessary for them to take all the exam based on their registered course (Kahar, 2011).

Basically, examination and course timetabling might be categorized into process environment, modeling and scheduling instances. There are differences between examination and course timetabling in the process environment. Every school has their own ways and requirement to produce their course timetable but for the exam timetable, it is produced centralized by the academic office of each school (McCollum, 2007). From modeling perspective, course timetabling is constructed according to the number of students that will enroll the courses while the exam timetable is according to the number of students that register on certain courses (McCollum, 2007). Lastly in scheduling instances, we need to allocate the individual lectures, tutorial and labs into the limited timeslots for course timetable while exam timetables are constructed based on the offered courses (McCollum, 2007).

2.3 List of constrains and objective of current examination timetabling problems.

There are many constraints and objectives in the examination timetable problem. The constraints and requirements are based on the affected parties in each academic institution such as students and lecturers. To produce a good quality of examination timetable, we need to fulfill as much as possible the constraints and requirements by each parties. Example of constraints by lecturer prefers to have their course schedule earlier within the exam periods while student prefer not to have a consecutive exams. Additionally, they dislike their exams to be packed together within the small gap of timeslot. By this, they will not have sufficient time to do the revision for their exams (Kahar, 2011).

Examples of the dataset are Toronto dataset (Carter, Laporte and Lee, 1996), Nottingham (Burke, Newall and Weare, 1996), Melbourne (Merlot, Boland, Hughes and Stuckey, 2003), Second International Timetabling Competition (ITC2007) (McCollum,

2007), UKM (Ayob, Abdullah and Malik, 2007) and UiTM (Kendall and Hussin, 2004). In the next section, we will discuss about the example of dataset in research community over this few years.

2.3.1 Un-capacitated dataset

As being mentioned before, un-capacitated dataset is dataset which does not considered about the room capacity and therefore it is much easier to solve compared to other dataset. The example of un-capacitated dataset is show in next sub-section.

A. Toronto dataset

The Toronto dataset is the oldest dataset among the datasets which had been listed out at previous section. It established from thirteen real-world examination timetabling problems. The timetabling problems are three from the Canadian highs schools, five from Canadian institution, each one from the London School of Economics, King Fahd University, Dhahran and Purdue University, Indiana (Carter, Laporte and Lee, 1996). The table below shows the information of the Toronto datasets.

Table 2.1: Toronto Dataset (Qu, Burke, McCollum, Merlot and Lee. 2008).

Problem Instance	Exams	Students	Enrollments	Conflict Density	TimeSlots
car91 I	682	16925	56877	0.13	35
car91 II	682	16925	56242/56877	0.13	35
car92 I	543	18419	55522	0.14	32
car92 II	543	18149	55189/55522	0.14	32
ear83 I	190	1125	8109	0.27	24

ear83 II	189	1108	8014	0.27	24
hec92 I	81	2823	10632	0.42	18
hec92 II	80	2823	10625	0.42	18
kfu93	461	5349	25113	0.03	42
lse91	381	2726	10918	0.06	18
pur93 I	2419	30029	120681	0.03	42
pur 93 II	2419	30029	120686/120681	0.03	42
rye92	486	11483	45051	0.07	23
sta83 I	139	611	5751	0.14	13
sta 83 II	138	549	5689	0.14	13
tre92	261	4360	14901	0.18	23
uta92 I	622	21266	58979	0.13	35
uta 92 II	638	21329	59144	0.13	35
ute92	184	2749	11793	0.08	10
yor83 I	181	941	6034	0.29	21
yor83 II	180	919	6012	0.29	21

They introduced the dataset with the aim to decrease the number of timeslots needed and to allocate the conflicting exam within the timeslot (using proximity values of 16, 8, 4, 2, 1) (Carter, Laporte and Lee, 1996). They implemented graph colouring heuristic with clique initialization and back-tracking methods. As the un-capacitated timetabling problem does not really solve the real world timetabling problem, Burke et al. (1999) modified the dataset by including total seating capacity.

2.3.2 Capacitated dataset

For capacitated dataset, it needs to consider room capacities in addition to other hard constraint in the dataset. Examples of capacitated datasets are discussed as follow.

A. Dataset of University of Nottingham

This dataset was introduced by Burke, Newall and Weare in 1996. This dataset differences from other dataset as it contains three timeslots per day from Monday till Friday. There are total 23 timeslots in this dataset with three timeslots per day from Monday till Friday. This dataset aims to reduce the number of conflicts on the same day.

Table below shows the University of Nottingham dataset and this dataset is available at http://www.asap.cs.nott.ac.uk/resources/data.shtml.

Table 2.2: University of Nottingham dataset (Burke, Newall and Weare, 1996)

Exams	Students	Enrolments	Conflict Density	Timeslots	Capacity
800	**7896**	**34265**	**0.03 (3%)**	**23**	**1550**

Merlot et al. (2003) applied the graph heuristic method to the Nottingham dataset in order to reduce the conflicts overnight and same day.

B. Dataset of University of Melbourne

This dataset was presented by Merlot et al. (2003). They presented two different datasets which has difference capacity in each timeslot and contains two timeslots on every weekday. This dataset has the same aim with the University of Nottingham that is to reduce the conflict overnight and same day.

Table below shows the University of Melbourne datasets and this dataset is available at http://www.or.ms.unimelb.edu.au/timetabling.

Table 2.3: University of Melbourne datasets

Problem Instances	Exams	Students	Enrolments	Timeslots
I	**521**	**20656**	**62248**	**23**
II	**526**	**19816**	**60637**	**31**

These dataset consider the overall capacity in a big room. However, a dataset which considered overall capacity can be considered a simplified version of the real world problem. This is because in real world problem, individual room capacity needs to be considered one by one not only the overall capacities (Kahar, 2011).

C. Dataset of Second International Timetabling Competition (ITC2007)

This dataset consists of many constraints which resemble the real world problem such as no students can take more than one exam simultaneously and the number of students who sits for the exams should not exceed the capacities of rooms which accommodate the exam. Besides that, other constraints such as exam which being allocate to a certain timeslots cannot violate that particular timeslot length and lastly the exams arrangement should follow the specified arrangement such as exam A must follow by exam B. Variety of the methodologies had been used by McCollum in this dataset such as hill climbing, great deluge algorithm and iterated forward search (McCollum, McMullan, Parkes, Burke and Abdullah, 2009). Table below shows the dataset of ITC2007 (examination track).