SOLVING TORONTO EXAMINATION TIMETABLING USING HEURISTIC METHOD

LIM RUEY LONG

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. Toronto benchmark data contains 13 real-world examination timetabling problem which have different conflict density for every dataset. Many researchers solved Toronto benchmark data using different method in order to produce a timetable which is feasible and solve all the constraints. To produce a feasible examination timetable, all the exams need to be scheduled into timeslot while satisfying the hard constraint and soft constraint. The timetable result should have the minimum penalty value in term of spread exams. Therefore, the technique partial graph heuristic with hill climbing method should be implemented to solve Toronto examination timetabling problem. The graph heuristic method will partially schedule the exam and then improved by hill climbing method. This process will be repeated until all the exams are scheduled. By using this technique, the solution of timetable result can comply all of the constraints and has a competitive result compared to other researchers' result.

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. Toronto dataset mengandungi 13 dataset yang mempunyai nilai konflik yang berbeze bagi setiap dataset. Ramai penyelidik telah mengguna pelbagai cara untuk menyelesaikan menghasilkan jadual waktu peperiksaan yang berguna dan menyelesaikan semua kekangan. Untuk menjana jadual waktu peperiksaan, semua peperiksaan perlu dimasukkan ke dalam waktu dengan memenuhi semua kekangan. Oleh itu, teknik Graph Heuristic bersama Hill Climbing haruslah digunakan untuk menyelesaikan masalah jadual waktu peperiksaan Toronto. 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 mampu berbanding dengan jadual waktu peperiksaan dihasil oleh penyelidik lain dan jadual waktu baru ini dapat memenuhi semua kekangan yang ada

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

STA-F-83	St. Andrew's Junior High School, Toronto Fall Semester
TRE-S-92	Trent University, Peterborough, Ontario Spring Semester
GH -	Graph Heuristic Method
GA	Genetic Algorithm
HC -	Hill Climbing Method

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter will briefly discuss about the overall of the project. It has five sections in this chapter. Background of the project will discuss in first section while problem statements of will explain in second section. Objectives of project will explain in third section. Next, the scopes and discussion of limitation for user and the projects will be explain in fourth section. Thesis organization will explain in the last section.

1.1 BACKGROUND OF THE PROJECT

Timetabling problem is the problem that contains four factors which is a finite set of resources, times, meetings and constraints (Burke et al, 2004d). Timetabling problems come in several types including nurse scheduling, transportation timetabling, educational timetabling (Burke et al, 2004d) and sport timetabling . All of these have show a important problem and challenging tasks for the researchers. Educational timetabling is widely studied among all the timetabling problems. The main factor of affecting a wide range of various stakeholders is the quality of timetabling. There have a relatively close problems between course and exam timetabling. This paper more focus on the exam timetabling.

Exam timetabling problems is restricted number of timeslots assign by an amount of exams which focus on hard and soft constraints. Hard constraints cannot be violated in any situation and a sufficient timetable will be produce when all hard constraints are solved. For example, a student cannot sit two examinations simultaneously. Besides, soft constraints are desire which hard to get a solution for all the soft constraints be satisfy. For example, a student should not sit for the exam consecutive and should have time to do revision.

Exam timetabling problem can be grouped into capacitated and un-capacitated problems. In un-capacitated problems, amount of room will not concerned. While the room capacities are considered as hard constraints for the capacitated problems. Based on Burke, Newall and Weare, (1996), the most difficulty in examination timetabling is to acquire a conflict-free schedule within a limited number of time periods and room availability. Therefore, capacitated problem is much more difficult than un-capacitated problem due to its close resemblance to the real world problem .

A lot of techniques have been employed to make a good quality solution in last ten years. Such methods involve graph colouring heuristics, meta-heuristics, case-based reasoning, tabu search etc. This motivates us to investigate the Toronto dataset using the hybrid graph heuristic method. The Toronto dataset consist of 13 exam timetabling problems.

1.2 PROBLEM STATEMENTS

The graph heuristics method has been widely used over the years to produce an initial solution. The initial solution is then improved using meta-heuristics method (e.g. hill-climbing, great deluge algorithm, tabu search etc). We have created hybrid graph heuristic methods that combine the graph heuristic and hill-climbing. The method involves partially scheduling the exam based on graph heuristic and the partially scheduled exam used hill-climbing to improve. The rest of the exams will be scheduled by the algorithm. The method able to produces good quality solution when applied to the examination dataset from ITC 2007.

Basically, all timetabling problems contains different side constraints which related with them and it is practical purposes of each university. In this thesis, we consider a real-world exam timetabling problem which consists of amount of constraints that not yet been investigated in previous scientific literature. The hard constraints include spread the exams evenly within the limitation of timeslot.

This motivates us to investigate the Toronto dataset using the hybrid graph heuristic method. The Toronto dataset consist a set of 13 exam timetabling problems. The dataset contain numerous constraint different from ITC2007 dataset which is worth of investigations.

1.3 OBJECTIVES

The objectives of the research are as follows:

- **i.** To implement the graph heuristics with hill-climbing to the Toronto datasets.
- **ii.** To satisfy all the hard constraint of the Toronto datasets.
- **iii.** To minimize the penalty cost of the generate timetable.

1.4 SCOPES

The scopes of this project are:

- i. Focus on the Toronto examination datasets.
- ii. Focus on the graph heuristic and hill climbing method.

1.5 THESIS ORGANIZATION

This thesis composed of seven chapters. Chapter 1 briefly discuss the system. The system demonstrate the problem statements, objectives and scopes. Chapter 2 will describe the examination timetabling problem and presents various examination datasets and constraints from the scientific literature. In this chapter, it will illustrate the method, technique, technology and equipment that carried out in this case studies. Overall of the project design and implementing will be review in Chapter 3. Development of the project design will be discuss in Chapter 4. Next, Chapter 5 presents the implementation, describe how research structure and record all processes involve in research development. The results and the summarized of the project will explain in Chapter 6. Lastly, Chapter 7 will present the conclusion and also the future work. It come together with the appendices which consists of Gantt Chart and reference links.

CHAPTER 2

LITERATURE REVIEW

This chapter provides details of the fundamental aspects of the research area tackled in this thesis. This chapter comprises seven sections. Section 2.1 briefly discuss the definition of timetabling and the general timetabling problem. Section 2.2 discusses the classification of the university timetabling problem. Section 2.3 provides further details of the examination timetabling problem. The variations of the examination timetabling constraints and objectives experimented within the scientific research are discuss in section 2.4. Section 2.5 describes the difference between the un-capacitated and the capacitated examination timetabling problem. Lastly in section 2.7 and section 2.8, we summaries the methodologies that have been applied to examination timetabling problem and we present our conclusions.

2.1 Overview of Timetabling

A timetable is a table that show certain events should occur at specified time. There is a variety types of timetabling such as educational timetabling, sport timetabling, and transportation timetabling. each of these vary in their structure, constraints and requirements (Burke, Kingston and deWarra 2004). A general timetabling definition given by Burke, Kingston and deWarra (2004):

> "A timetabling problem is a problem with four parameters: T, a finite set of times; R, a finite set of resources; M, a finite set of meetings; and C, a finite

set of constraints. The problem is to assign times and resources to the meetings so as to satisfy constraints as far as possible. "

Based on the definitions, timetabling problems involve allocating events into the suitable timeslots and resources whilst satisfying the constraints of the problem. The constraints usually categories into hard constraints and soft constraints. Hard constraints cannot be violated under any circumstances. For example, no student is allowed to take two or more exams at the same times. While soft constraints are critical but need to solve as much as possible. The exams should spread as evenly as possible throughout the exam periods.

2.2 Classification of university timetabling problems

University timetabling problems can be separated into examination and course timetabling. Carter and Laporte (1996) state that both timetabling have the same characteristics in the general timetabling problem. Carter and Laporte (1998) stated the course timetabling as a multi-dimensional assignment problem in which students, lecturers are assigned to courses, course sections or classes. Carter and Laporte (1996) defined the examinations timetabling as the exams are assigned to a limited number of available time periods without any clashes.

Examination and course timetabling problems are concerned with prevent assigning students sitting two (or more) exams or courses at the same time. Though, significant differences are exist which include differences in constraints. Table 2.1 and table 2.2 shows an example of hard and soft constraints for course timetabling (Abdullah,2006) and examination timetabling (Qu et al, 2009) problems respectively.

Moreover the differences in constraints, both timetabling also vary in the way they are constructed. It can divided into process environment, scheduling instances and modeling. In the process environment, course timetable normally produced separately and

independently by each school, not like exam timetable, which is produced centrally by the academic office (McCollum,2007; Burke et al., 1996). In scheduling instances, course and exam are used different instances even though it is from the same source. In the course timetable, we have to schedule the individual lectures, tutorial and labs from offered course. While the examination timetable are produced based on the offered course (McCollum, 2007).

Although there is a differences between the course and examination timetabling problem, the complexity of examination timetabling problem depends on the amount of freedom of choice on students selecting their course timetable (Laporte and Desroches, 1984). The more freedom a student has the difficulty in producing a feasible timetable. This research focuses on the examination timetabling problem and more details will discuss on next sections.

Table 2.1 Example of hard and soft constraints for the course timetabling problems (Abdullah, 2006)

Hard Constraints
1. A lecturer and student cannot be in different places simultaneously. 2. Each timeslot only allow assign one course. 3. The classroom capacity should be fulfill the number of students registered the course. 4. The classroom assigned to the course should satisfy the features required by the course.
Soft Constraints
1. Students should not only attend one class in a day. 2. Students should not be assigned more than two consecutive courses on a day. 3. Students should not attend a course that is the last timeslot of the day.

2.3 Examination Timetabling

Examination timetabling problem can be described as allocating exams to a limited number of timeslots and rooms, satisfying hard constraints and minimising the soft constraints. Schaerf (1999) stated that examination timetabling is scheduled a number of exams into a specified time. According to Qu et al., (2009), examination timetabling problem is assigned a set of exams E = e_1 ,e_2...e_e to a limited number of available time periods T = t_1 ,t_2...t_t without any clashes. Table 2.2 show an example of constraints.

Examination timetabling is important and time-consuming tasks which occur periodically (i.e. annually, quarterly, etc.) in all academic institutions. It considered as time-consuming tasks because it involves data collection, constraint modeling, algorithmic modeling and solution modeling. According to Burke et al. (1996), 75% of timetables are altered between draft and final version. This is due to the data being made available late, poor quality timetables being generated and incorrect data. Hence, a precise and close interaction with all parties (e.g. administrator constraint modeling) should be carried out to avoid any problems. Any misinterpretation and miscommunication during the early stages could lead to changes being required in the generated solution. Freedom of students choose their courses to suit their own preference make the examination timetable more difficult to generate. Amount of students and examination offered also will affect the examination timetable to generate.

Exam timetabling problem can be grouped into capacitated and un-capacitated problems. In un-capacitated problems, amount of room will not concerned. While the room capacities are considered as hard constraints for the capacitated problems. In section 2.5 will further discussion about the capacitated and un-capacitated problems.

Table 2.2 Example of hard and soft constraints for the examination timetabling problems (Qu et al., 2009)

Hard Constraints

1. A student should not attend more than one exam at the same time.
2. Exams resources should be sufficient such as rooms enough for the exams.

Soft Constraints

1. Conflict exams should spread as evenly as possible.
2. Groups of exams required to take place at the same time , on the same day or at one location.
3. Exams to be consecutive.
4. Largest exams should schedule early.
5. Satisfy ordering of the exams.
6. Number of students in any timeslot be limited.
7. Locate nearby the conflict exams .
8. Exams may be split over similar location.
9. Same length exams can be located at the same place.

2.4 Constrains and objective investigated in examination timetabling problems.

Different academic institutions have different constraints to suit their examination timetabling. It is shown in many literature. Besides, a good quality of examination timetable also affected by different parties. For example, a student might hope that their exams can spread as much as possible so that they have time to do revision between the exams. For an administrator side, they prefer no student take two exams at the same period. At here, we refer some of the common constraints appear in the examination timetabling problems.

Most of the research used the datasets from Nottingham (Burke, Newall and Weare, 1996), and Melbourne (Merlot et al. 2003). Examination datasets from Second International Timetabling Competition (McCollum et al. 2008) also used as a reference.

2.4.1 Toroto Datasets

The datasets from Toronto consists of thirteen real-world exam timetabling problems with five from Canadian institutions, three from Canadian high schools, one from London School of Economics, one from King Fahd University, Dhahran and one from Purdue University, Indian. All these datasets can be downloaded from ftp://ftp.mie.utoronto.ca/pub/carter/testprob/. Table 2.3 show the some of the information of Toronto datasets.

Table 2.3 Toronto datasets

Problem Instance	Exams	Students	Enrolments	Conflict Density	Timeslots
car91 II	682	16925	56877	0.13	35
car92 II	543	18419	55522	0.14	32
ear83 II	189	1108	8014	0.27	24
hec92 II	80	2823	10625	0.42	18
kfu93	461	5349	25113	0.06	20
1se91	381	2726	10918	0.06	18
pur93 II	2419	30029	120681	0.03	42
rye92	486	11483	45051	0.07	23
sta83	138	549	5689	0.14	13
tre92	261	4360	14901	0.18	23
uta92	638	21329	59144	0.12	35
yor83	180	919	6012	0.29	21

The Toronto datasets were introduced by Carter, Laporte and Lee on 1996. They investigated two objectives which are reduce the number of timeslots needed and to spread the conflict exam within the timeslots. They used graph heuristic to test all the datasets. On 2001, Gaspero and Schaerf investigated the datasets by using tabu search which consider the first and second order conflict. First order conflict (hard constraint) is when a student has to take two exams at the same time while second order conflict (soft constraint) is when a student take two exams in consecutive periods. On 1996, Burke et al. study the datasets by considering the maximum room capacity per timeslot and second-order conflict of same day constraints.

On 2003, Merlot et al. study the datasets with the aim to reduce the number of timeslots needed, spreading the conflict exams within limited number of timeslots, to reduce second-order conflict of the same day and overnight. They used several methodologies during the investigation which include programming, simulated annealing and hill climbing. Kendall and Hussin (2005) applied tabu search hyper-heuristics that work with high level heuristics.

2.4.2 University of Nottingham

The dataset from University of Nottingham was introduced by Burke et al. (1996b) as benchmark. It contains 23 timeslot and their objective is to reduce the number of students taking two exams at the same period. Table 2.4 show the information of the dataset from Nottingham. In 2004, Burke and Newall investigated the dataset using using heuristic modifier with the aim to minimize the adjacent exams at the same time.

Table 2.4 University of Nottingham dataset (Burke, Newall and Weare, 1996)

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

2.4.3 University of Melbourne

The Melbourne dataset were introduced by Merlot et al., (2003). They introduced two different datasets which has two timeslots on each weekdays and each capacity varied. The objectives of the dataset is to minimize adjacent exams on the same day or overnight. Table 2.5 show the summarized of dataset from Melbourne which can download from http://www.or.ms.unimelb.edu.au/timetabling. In 2005, Cote, Wong and Saboun investigate the dataset using evolutionary algorithms with bi-objective constraint satisfaction to minimize the adjacent exams on the same day or overnight.

Table 2.5 University of Melbourne datasets

Problem Instance	Exams	Timeslots	Students	Enrolments	Objective
I	521	28	20656	62248	Minimize adjacent exams on the same day or overnight
II	562	31	19816	60637	Minimize adjacent exams on the same day or overnight

2.5 Un-capacitated and capacitated examination timetabling problems

The un-capacitated examination timetabling problem was show in many literature which concerned on the algorithm and the performance of produce an effectively solution (Qu et al., 2009). Even though the un-capacitated datasets are popular at that time, McCollum believed most of the researchers are not dealing with all the aspect of the problem. This is because the researchers only worked on the examination problems which are simplified version. In 2009, Qu et al. expose that most of the research only attend to some common hard constraints. For example, students should not take two or more exams at the same period. A student should have enough time to do revision between the exams are the example for the commonly used of soft constraints.

The capacitated problems which consist of room capacity constraint are more likely to the real world problem. But it still less concern by the researchers which due to the lack of benchmark datasets. Since the capacitated problem consist of room capacity, it require more complete data which include student and exam list for the less complex problem. Capacitated problem are hardly to solve due to lack of exam rooms and the splitting exams between more than one room (Burke et al., 1996).

A modification of benchmark datasets have been made by involving an overall capacity as if all exams were taking at the same place (Burke et al., 1996). This is because the capacitated problems more closely resemble the real world problem even current benchmark datasets lack some information on the seating capacity for each room.

2.6 Methodologies applied to the examination timetabling problem

In the last ten years, there has been a significant amount of research on exam timetabling. We can found that a variety of algorithms have been proposed, which include graph heuristic, tabu search, simulated annealing, memetic algorithms and many other approaches, in order to produce a feasible timetable. Carter and Laporte (1996) divided the techniques used into cluster methods, sequential methods, constraints-based methods and meta-heuristics. Petrovic and Burke (2004) added multi-criteria, case-based reasoning and hyper-heuristics approaches.

2.6.1 Graph heuristics (GH)

In examination timetabling problems, the exams are represented by vertices in a graph, and the hard constraints are represented by the edge between the vertices. They assigned different colour to the vertices so that no vertices have the same colour. Then it will correspond to assign timeslots to the exams.

Originally graph heuristics are constructive methods which ordering the exams by how difficult they are to be scheduled. Many ordering strategies and their modified variants with the aim to produce a good solution appear in the timetabling literature (Carter 1986). Graph heuristics are able to produce a good quality solutions in shorter time and easy to apply. Table 2.6 show some of the widely employed ordering strategies.

Table 2.6 Widely studied ordering strategies for examination timetabling problems.

Heuristics	Ordering strategy
Largest degree (Broder 1964)	Schedule the exams that have the most conflict with other exams.
Largest weighted degree (Carter et al. 1996)	Schedule the exams that have the most number of students who are involved in the conflict.
Largest enrolment (Wood 1968)	Schedule the exams that have the highest number of registered students.
Random ordering	Randomly order the exam

Burke et al. (1998c) investigated the effect of random elements into the employment of graph heuristics (Saturation Degree, Colour Degree and Largest Degree) by using (1) tournament selection that randomly choose one from a subset of the first exams in the ordered list; and (2) bias selection that takes the first exam from an ordered list of a subset of all of the exams. These method able to produce a good results on Toronto datasets.

In 2009, Qu and Burke studied the used of graph heuristic within hyper-heuristic methodology. Hyper-heuristic is used to construct the timetables by choosing the graph heuristic. This motivation is due to the graph heuristics cannot appropriate address the complex timetabling problems and sometimes failed to generate feasible solutions. However, Muller (2008) shown that they are effective as producing initial for meta-heuristics.

2.6.2 Hill Climbing (HC)

Hill climbing is one of the local search technique. The candidates solution is random selected from the neighbouring solution. If the candidates solution is accepted then it will replace the current solution. Figure 2.1 show hill climbing procedure.

Hill climbing is easy to implement but also easily trapped in local optima. Hence, many researcher subject to hybridise hill climbing with other search methods. In 1996, Burke et al. hybridised hill climbing with genetic algorithm. Kendall and Hussin (2005b) investigated hill climbing and hyper-heuristic for solving the examination timetabling problem. Muller (2007) applied great deluge algorithm and hill climbing on the ITC2007 problem.