THE GREAT DELUGE ALGORITHM ... OD TO SOLVE ITC 2007 EXAMINATION TIMETABLING PROBLEM

POH YI KENG

THESIS SUBMITTED IN FULLFILLMENT OF THE DEGREE OF COMPUTER SCIENCE (COMPUTER SYSTEMS & NETWORKING)

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING
UNIVERSITI MALAYSIA PAHANG
2014

Abstract

There are several problems in the examination timetabling (exam-timeslot-room assignment) that involves assigning exams to a specific or number of timeslots and rooms, with the aim of fulfill the soft and hard constraints as much as possible. In the scientific literature, there are many techniques that have been used to solve the simplified examination benchmark datasets.

The examination track of the second International Timetabling Competition (ITC 2007) dataset is used for this research and this dataset had several novel constraints and also to those commonly used in the literature. The International Timetabling Competition (ITC 2007) have total of twelve dataset and each of them have different features. For example, number of exam, number of room and also number of timeslot.

Great deluge algorithm is being used to solve the problem of timetabling.

ABSTRAK

Masalah sistem peperiksaan meliputi mengaturkan peperiksaan ke dalam bilik dan masa yang tertentu, dengan memenuhi kekangan sebaik mungkin. Kebanyakan teknik yang dilaporkan dalam kesusasteraan telahpun digunakan untuk menyelesaikan masalah peperiksaan ini. Bagi penyelidikan ini, data adalah diambil daripada trek peperiksaan bagi International Timetabling Competition (ITC2007) yang kedua. Dataset ini mempunyai beberapa kekangan yang tidak biasa dijumpai bagi dataset yang lain. Kesemuanya terdapat dua belas dataset yang mempunyai ciri-ciri yang berbeza, contohnya kapasiti bilik, bilangan masa dan kuantiti perperiksaan.

Gabungan bagi kaedah Graph Heuristic dan Great Deluge Algorithm akan diguankan untuk menyelasaikan masalah perperiksaan ini. Bagi kaedah Graph Heuristic, ia akan digunakan untuk menghasilkan penyelesaian yang paling awal. Selepas itu, penyelesaian itu akan diperbaiki oleh kaedah Great Deluge Algorithm. Akhirnya, akan menghasilkan satu penyelesaian yang lebih baik dan memenuhi kekangan yang diperlukan.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In chapter 1, there are total of five subtopics where they are background of study, problem statement, research objective, scopes and thesis organization. Background of study will discuss about background of the project. Problem statement will explain about the situation of my motivation to do this project. Project goals will be discussed in objective part and the restriction for user and project, the area of research will be talked in scope part. Finally, the last part, thesis organization will talk about the flow of each chapter in this project.

1.1 Background of Study

Many researchers have widely study the examination timetabling problem of the Toronto benchmark dataset where it is introduced by Carter in year 1996. The distribution of the university exams and timeslot were concerned in this examination timetabling problem. Besides that, the examination timetabling problems are divided into two types where they are capacitated and un-capacitated. For solving the un-capacitated exams problem quickly and effectively, the algorithm and algorithmic performance were focusing by many researchers said by Burke and Petro Vic in year 2002 and 2009. For example, room capacities for un-capacitated problems will not be included but room capacities for capacitated problems were one of the hard constraints that should be fulfilled so that it will

matched with the real world problem said by Pillay and Benzhaf in year 2009; Abdullah in year 2006.

But, at the same time, the capacitated problem was ignored by the researchers. This situation may cause the lack of capacitated benchmark dataset. Moreover, if compare to the un-capacitated problems, it will be more difficult to solve the capacitated problem which had more constraints. There are more constraint like the amount of room and the size of room when solving the capacitated problems and the complexity of the problem will be increased. Based on Burke in 1996a, most of the university agrees that it is a hard task to schedule examination timetable.

In addition, there are several constrains that should be fulfilled to solve the timetabling problems. The constraints can be divided into two parts where they are soft and hard constraints. For requirement in soft constraint, it is not necessary but the quality of the timetable is needed to improve by soft constraint as far as possible. For example, a student should not take more than two exams in a day. In the other hand, hard constraints are the most importance constraint that must be fulfilled. If there are one of the hard constraints is not being fulfilled, then the time table is considered fail. For example, at the same time, there is no participant can sit for two examinations. Besides soft and hard constraints, there are several constraints must also be taken into consideration according to the solutions of iterative algorithms that are being traversed. For example, reduce students' conflict to have different exams at the same time and prevent two exams are pick at time so that students have some free time to rest are more important constraints based on James Carlo T. Mendoza in year 2006.

In this paper, we will use the International Timetabling Competition 2007 (ITC 2007)'s capacitated problem. There are some new constraints are added in this dataset. So, we will use great deluge algorithm method to solve these problems.

1.2 Problem Statement

Many researchers have been attracted by the examination timetabling in this few years. An un-capacitated dataset is used by many reported in the literature but it does not imitate the real world examination timetabling problem. Toronto dataset is an example of un-capacitated dataset. However, capacitated dataset like Nottingham and Melbourne dataset were introduced by some researchers but as the constraints, only a largest number of seats are included in a day. As the individual room capacity normally needed to be considered, this is not really look like the real world problem. So, a gap is formed between research and practice.

Thus, we need to focus on solving ITC 2007 examination dataset as this dataset is a capacitated examination dataset where it is look like real world problems which different constraints from other dataset are contained like Toronto, Melbourne and Nottingham that seen in the literature .

1.3 Research Objective

In this research, we had set a total of three objectives:-

1. The examination track of the Second International Timetabling Competition (ITC 2007) needed to be study.
2. To implement great deluge algorithm method for solving the timetabling problem.
3. To verify all the hard constraints and soft constraints are fulfilled by develop a schedule.

1.4 Research Scope

The ITC 2007's dataset would be investigated in this research. The ITC 2007 is a capacitated dataset so for the hard constraints like the room size and the number of rooms will be considered by it. Thus, we will implement the great deluge method in order to develop a schedule for the dataset.

1.5 Thesis Organization

This thesis consists of seven chapters. Introduction of the system will be discussed in Chapter 1. Chapter two will be the literature review and we will talk about methodology in Chapter 3. For Chapter four and five, there will be more on the design and implementation. The result and the discussion of the research will be shown in Chapter six. Last but not least, conclusion of the research will be shown on Chapter seven.

Chapter 2

LITERATURE REVIEW

2.1 Timetabling Outlined

A schedule is defined as a timetable that people do it to list down all events with the time that they will take place. Personal timetabling, transportation timetabling and educational timetabling are the categories of timetabling (Qu el al.2009). They might have different constraint or requirement for all of these timetabling problems that need to be satisfied (Burke, Kingston and deWerra 2004).

Constraints in the timetabling problem are divided into two types where they are soft and hard constraint. Soft constraints should be satisfied as much as possible but it is not necessary. The quality of the timetable produced will increase by reducing the violations of soft constraints. For example, exam will be equally distributed angle of view as a student for the soft constraints. Next are hard constraints, hard constraints are fulfilled constraints and it should not be break. If there is no break of hard constraints, then a timetable is considered usable. For an example, more than one exam is not allowed to be taken at the same time.

2.2University timetabling problems

Based on M. Dimopoulou and P. Miliotis in year 1998, there are two types of problems that will occur in university timetabling problems where they are course timetabling and examination timetabling problems. Due to Burke E K, Kingston J H and de Werra D in year 2004, the characteristics and the main problem between course timetabling and examination timetabling problems can be considered as the same. Besides that, two scheduling problems involved shall cause to change constraints and try to meet all the objectives and goals in the space of time.

2.2.1 Problems of course timetabling

Based on Abdullah S in year 2006, in the course scheduling problem can be defined as a group provided by the University to be distributed to all students, teachers or classroom does not participate in more than one time slot period and the number of students and classroom conditions curriculum, well-being in a classroom assignment must be less than or equal to every room capacity.

Of course timetable has its limitations; it can be separated into hard constraints and soft constraints. The following are some examples of constraints

Table 2.1 Course timetabling problem's constraints

Hard Constraints • A teacher and student should not be assigned in more than one place at the same timeslot. • Each timeslot can only have one course with only one schoolroom. • Capacity of each classroom must be able to accommodate the total number of students that attend the course at a certain timeslot by having equal capacity or more than that. • The classroom should have the suitable equipment and features to fulfill the course that being assigned in.
Soft Constraints • Each student should have more than one course per day. • A student should be avoiding attending two or more consecutive courses on a day. • Each student should be avoiding to be scheduled to attend a course which is being allocated to the final timeslot of the day.

2.2.2 Examination timetabling

Part of the academic institution management activities is means by examination timetabling problem. According to Masri Ayob, Salwani Abdullah and Ariff Md Ab Malik in year 2007, for developing examination timetabling will becomes complex when the amount student enrolments, broad variety of courses and combined degree courses are growing. Moreover, based on M.N.M. Kahar in year 2011, the level of freedom of choice on students is depended by the difficulty level to construct the examination timetable to select their own courses. When generate the examination timetable, some academic institutions to achieve the open registration system (OR) may have additional difficulties compared with other academic institutions that do not implement the OR.

According to Masri Ayob, Salwani Abdullah and Ariff Md Ab Malik in year 2007, normally, define a set of exam is the exam timetable arrange in a limited amount of slots allocated within and implement with all the hard constraints. The examination timetabling also is assigned a set of exam rooms and a limited duration of slots conditions are met with constraints based on R. Qu, E.K. Burke, B. McCollum, L.T.G. Merlot and S.Y. Lee. in year 2008.

Some of constraints in examination timetabling problems are shown at Table 2.2.

Table2.2 Example of constraints for the examination timetabling problems (R. Qu, E.K. Burke, B. McCollum, L.T.G. Merlot and S.Y. Lee. 2008)

Hard Constraints

- There are no collaborative resources (e.g. Students) in exams being assigned simultaneously.
- 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

- The exams should not in any consecutive period slots or days and should spread as much as possible.
- The exams in same group must be held at the same period, day or at same place.
- Consecutive all of the exams.
- Every exam should be scheduled first or the largest exams should be scheduled at early time compared to others small exams.
- Satisfied all prior exams condition.
- Every timeslot should limit the numbers of exams and students.
- Some specific exams must be place in certain timeslots as request by the school.
- Located conflicting exams on the same day as near as possible.
- Might be able to split the exams over nearby or similar places.
- Combined the exams with the same length into same room as long as got sufficient room capacity for students.
- Resource requirements should be fulfill as many as possible.

It is harder to arrange exam timetable if students are free to arrange the course as they like. Besides that, according to McCollum B in year 2007, the generation of examination timetabling becomes harder when Muslim students request do not have any exams at Fridays. All party should be satisfied by the solution of the exam timetable and many factors need to be considered, as there is no conflict, and enough space for students each exam papers.

2.3 Objective and constrains inquired into examination timetabling problems

There are a few objectives and constraints when arrange the examination timetable. The affected parties like students and administrator were based by the constraints and requirements. The constraints and requirements needed to be fulfilled by each party as much as possible to form a high quality of the examination timetable. For example, there is no chance for all students to take their respective exam at the same time where this responsibility is done by the examination timetable administrator. For students, they do not like their own examinations are packaged together in a small gap between the timeslots. By this way, time is not enough for them to do their revision for the exams. One of the parties' requirement and constraints will be opposed if we satisfy the other party. So, we just only can fulfill both of the parties but not just concentrates at one of the parties.

There are several most common used datasets in the community of examination timetabling research and they are Toronto dataset introduced by Carter, Laporte and Lee in year 1996, Nottingham introduced by Burke, Newall and weare also in year 1996 and Melbourne introduced by Merlot in year 2003. The researchers mostly focus at Toronto dataset compare with other two datasets. In year 2008, McCollum et al (2008) had introduced the Second International Timetabling Competition (ITC2007) dataset which more realistic problems are more than benchmark problems. In addition to these two, there are other examination datasets like UKM introduced by Ayoub in year 2007 and UiTM introduced by Kendall and Hussin in year 2004, Hu Xin in year 2005.

2.4 Capacitated and un-capacitated examination timetabling problem

There are many literatures that had survey the un-capacitated problem which effectively and quickly produce solutions focused on algorithms and algorithm performance. Based on Carter and Laporte in year 1996 and McCollum in year 2007, all aspects of most researchers have not dealt with the problems and they only work a simplified version of the examination problems. Those survey papers only address a few common hard constraints. For example, the room capacity should be greater or equal to the exam capacity; No student can take more than two subjects at the same time and etc. As the soft constraints, will be used to check possible to spread the examination, or not in a plurality of consecutive time slots or several days.

Assume that there is no capacity limitation problems, the capacitated problem has consists restrictions on the capacity of the room, which makes it more similar to the real world. However, the researchers have pay less attention through the capacitated problem because of the lack of benchmark dataset. Besides that, it is difficult to solve the capacitated problem. Based on Burke survey paper, most of the universities agree that addressing exams is a hard problem. For the capacity's problem, it needs more complete data, because they must consists data capacity of the room and the data of easier problem like exam and student list. According to McCollum in year 2007, this may be hard to collect additional information. Due to the lack of available halls and the problem of split exams between multiple rooms, other constraints will be affected. For example, splitting an exam into different sites or taking into account between rooms.

Based on Burke, Newall and Weare, in year 1996, some modification had made on the benchmark dataset like Toronto dataset. This is because due to make the dataset more look like the real world problem which including the overall capacity, as if all the exams occur in a large room. But, this represents a simplified Timetable problem, because we must consider the capacity of each room based on Mello in year 2003.

2.4.1 Toronto dataset

Thirteen real-world examination timetabling problems are having in Toronto dataset where one from King Fahd University, Dhahran, five from the Canadian institution, three from the Canadian highs schools, one from the London School of Economics and one more from Purdue University, Indiana according to Carter M W, Laporte G and Lee S Y in year 1996. The information of Toronto dataset is showed in Table 2.3.

Table 2.3: Toronto Dataset

Problem Instance	Exams	Students	Enrollments	Conflic 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
pur93 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

The Toronto dataset was presented by Carte, Laporte and Lee, (1996) and with the objective to reduce the number of timeslots and to arrange the exam which is conflict within the timeslot by using proximity values of 16, 8, 4, 2 and 1. They are using the graph colouring heuristic with clique initialization and back-tracking methods.

Because of the un-capacitated timetabling problem does not mimicking the real world timetabling problem, so the dataset are being modified from un-capacitated problem to capacitated problem because of the capacitated problem more mimicking the real world timetabling problem.

2.4.2 Dataset of University of Melbourne

In year 2003, the Melbourne dataset was presented by Merlot. Melbourne dataset can be divided into two dataset where the first is had two timeslots on weekday and another one is the different between each timeslot's capacities. The objective of this dataset is same with University of Nottingham that is to minimize the conflict overnight and same day.

The University of Melbourne datasets are shown in the table below.

Table 2.4 University of Melbourne datasets

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

2.4.3 Dataset of University of Nottingham

In year 1996, the Nottingham dataset were presented by Burke, Newall and Weare. Because from Monday till Friday, there are three timeslots per day, so this dataset is different from other dataset. The objective of this dataset is to minimize the number of conflicts on the same day and in this dataset, there are total of 23 timeslots. The University of Nottingham dataset is shown in the below table.

Table 2.5 University of Nottingham datasets

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

Besides Burke and Newall, the graph heuristic method was used by Merlot in year 2003 to the Nottingham dataset so that it can minimize the conflicts overnight and same day.

2.4.4 Dataset of University Kebangsaan Malaysia (UKM)

In year 2007, the UKM dataset was presented by Ayob and allocate all exams are required in this dataset. The constraints of this dataset almost the same with other dataset like students are not allowed to take more than one exam at the same time. But, there are extra constraints in this dataset which a maximum of two exams only can take by a student continuously in one day and the students who take exams consecutively must arrange in the same room. The year 2006 semester 1 dataset of University Kebangsaan Malaysia (UKM) and the room's information for dataset UKM06-1 is shown at table below.

Table 2.6 University Kebangsaan Malaysia datasets (UKM06-01)

Exams	Students	Enrolments	Timeslots	Capacity
818	14047	75857	42	1550

Table 2.7 Room capacity of dataset UKM06-01

Room	Room Capacity
DPBestari	850
DGemilang	610
Dewan (DECTAR)	610
LobiUtama (DECTAR)	270
PSeni (DECTAR)	152
LobiA (DECTAR)	70
LobiB (DECTAR)	70

2.4.5 Dataset of University Teknologi MARA (UiTM)

In year 2004, the UiTM dataset was presented by Kendall and Hussin which is from Uitm Malaysia. The constraints of this dataset are same with other dataset which is all exams are arranged in examination timetable and so on.

The information of the UiTM examination dataset is shown at Table 2.8.

Table 2.8 University Teknologi MARA (UiTM) dataset

Exams	Students	Enrolments	Timeslots
2063	84675	357761	40

2.4.6 Second International Timetabling Competition (ITC2007) dataset

The second international timetabling competition (ITC 2007) is separated into two parts where are course and examination timetabling but the examination dataset we will only focus on. The aim of ITC 2007 is for researchers to access their algorithms on real world timetabling problems by creating a platform. There are few constraints that are contained in the ITC 2007 examination where first is no student sits more than one exam at the same time and second is the exams should not exceed the room capacity. A timeslot have been assigned in an exam that should not violet the timeslot length and a specified arrangement is needed to be followed in the exams. For example, arrange exam B after exam A and exam B must be room 10 and etc. There are few objectives in this dataset like reduce second-order conflicts on the same day, the duration of exams will be minimized within a timeslot, reduce the usage of a particular timeslots or room and arrange the examination which is large more early. The information of the examination competition track can be found in McCollum et al., (2008). McCollum et al., (2009) which is one of the researchers that investigate this dataset by using iterated forward search, hill climbling and great deluge algorithm. A multistage approach has been used by Gogos, AleFragis and Housos, (2008) which include GRASP, simulated annealing and mathematical programming. Table 2.9 shows the information of the ITC2007 datasets (examination track) whereas Table 2.10 and 2.11 show the hard and also the soft constraints of the dataset. Lastly Table 2.12 shows the summary of some of the dataset.

Table 2.9 International Timetabling competition dataset

Instance	Conflict Density (%)	Exams	Students	Periods	Rooms	Period HC	Room HC
Exam-1	5.05	607	7891	54	7	12	0
Exam-2	1.17	870	12743	40	49	12	2
Exam-3	2.62	934	16439	36	48	170	15
Exam-4	15	273	5045	21	1	40	0
Exam-5	0.87	1018	9253	42	3	27	0
Exam-6	6.16	242	7909	16	8	23	0
Exam-7	1.93	1096	14676	80	15	28	0
Exam-8	4.55	598	7718	80	8	20	1
Exam-9	7.48	169	655	25	3	10	0
Exam-10	4.97	214	1577	32	48	58	0
Exam-11	2.62	934	16439	26	40	170	15
Exam-12	18.45	78	1653	12	50	9	7

Table 2.10 Hard constraints of ITC 2007.

Hard Constraints	
H1	Student cannot sit more than one exam at the same time.
H2	The exams capacity should not exceed the room capacity.
H3	The exam length should not violate the timeslot length.
H4	A sequence or ordering of an exams must be respected, e.g. schedule Exam A after Exam B;
H5	Schedule exam into specified room (room related hard constraints) e.g. Exam A should schedule to Room 11.