# HONEYBEE ALGORITHMS BASED ON FORAGING BEHAVIOUR FOR EXAMINATION TIMETABLING PROBLEMS

MALEK ALI AHMAD FEHAID

THESIS SUBMITTED IN FULFILMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2012

# ALGORITMA LEBAH BERDASARKAN KELAKUAN PERBURUAN UNTUK MASALAH PENJADUALAN PEPERIKSAAN

MALEK ALI AHMAD FEHAID

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH IJAZAH
DOKTOR FALSAFAH

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2012

## DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

May 2012 MALEK ALI FEHAID

P49063

## ACKNOWLEDGEMENT

Firstly I would like to state my sincere appreciation to my supervisors Assoc. Prof. Dr. Salwani Abdullah for her patience guidance, encouragement, understanding, and excellent advice throughout this study.

I would like to dedicate this work to my father and to the memory of my mother.

I am also deeply and forever indebted to the person in my life that touched my heart and gave me strength to move forward to something better. The person who inspire me to breathe, who encourage me to understand who I am, and who believe in me when no one else did. I am deeply devoted to my wife for the help, patience and confidence she had in me, throughout the times.

Finally, I am thankful to all my colleagues and friends for their support, with whom I shared pleasant times.

## ABSTRACT

Examination timetabling problems (ETTPs) become an intensive research in optimisation field. In ETTP, examinations are scheduled to timeslots and rooms based on the constraints and universities resources. Many approaches in the literature have addressed this difficult optimisation problem. Literature review shows that most of the population-based meta-heuristic approaches concentrated in finding only one high quality solution. This motivated the investigation of honeybee algorithms that are based on foreign behaviour (Honeybee Based Foraging Behaviour, HBFB) for ETTPs that try to bring all the solution in the population to be as good as possible. To date, HBFB algorithms have not been applied to the ETTPs. HBFB algorithms mimic the real foreign behaviour of honeybee in searching for food sources. HBFB algorithms are population based algorithms which consist of three main processes i.e. exploration, selection and exploitation. Some of the important challenges include finding a right selection strategy that bring the population to converge together without sacrificing the quality of the solutions, deal with adaptive mechanism for a better exploitation during the search process, and creating a balance between the exploration and exploitation to avoid premature convergence and get trapped in a local optimum. The research firstly aims to drive the population to better solutions by considering the principles of selection strategies. Secondly, to enhance the exploration of the search space by adaptively change the neighbourhood operator during the search process. Finally, to balance the exploration and exploitation processes, overcome the premature convergence and avoid from easily trapped into a local optimum by employing hybridisation methods. Towards these aims, three different models of the foreign behaviour in honeybee algorithms i.e. Artificial Bee Colony (ABC), Bees Algorithm (BA) and Bee Colony Optimisation (BCO) have been proposed and tested on two categories of datasets i.e. uncapacitated examination timetabling and International Timetabling Competition datasets (ITC2007). Three selection strategies, namely, tournament, rank and disruptive have been tested. The results demonstrate that the disruptive selection performed better than tournament and rank selections when embedded with ABC, BA and BCO (coded as DABC, DBA and DBCO, respectively). In addition, a self-adaptive mechanism for the neighbourhood search has been employed within DABC, DBA and DBCO algorithms and able to improve the quality of the solution (coded as Self-Adaptive DABC, Self-Adaptive DBA, Self-Adaptive DBCO). These are then incorporated with two local search algorithms (i.e. Simulated Annealing, SA and Late Acceptance Hill Climbing, LAHC) which show that the LAHC can further enhance the quality of the solutions in comparison with SA (coded as Self-Adaptive $DABC_{LAHC}$, Self-Adaptive $DBA_{LAHC}$, Self-Adaptive $DBCO_{LAHC}$). Overall comparisons indicate that Self-Adaptive $DBCO_{LAHC}$ works well across all datasets and able to obtain two best results in comparison with best known results in the literature particularly on the uncapacitated examination timetabling problem.

# ABSTRAK

Masalah penjadualan peperiksaan (MPP) merupakan satu penyelidikan yang intensif dalam bidang pengoptimuman. Dalam MPP, peperiksaan dijadualkan kepada slot masa dan bilik peperiksaan berdasarkan kepada pelbagai kekangan. Terdapat banyak pendekatan yang telah cuba menangani masalah pengoptimuman yang sukar ini. Tinjauan literatur menunjukkan bahawa kebanyakan kaedah meta-heuristik berasaskan populasi tertumpu kepada penghasilan hanya satu penyelesaian yang berkualiti tinggi. Ini memotivasikan penyelidikan algoritma lebah berdasarkan kelakuan perburuan (Lebah Berdasarkan Kelakuan Perburuan, LBKP) untuk MPP yang cuba menghasilkan semua penyelesaian dalam populasi sebaik mungkin. Buat masa ini, algoritma LBKP belum diaplikasikan untuk MPP. Algoritma LBKP mimik kelakuan sebenar lebah dalam mencari sumber makanan. Algoritma LBKP merupakan algoritma berasaskan populasi yang terdiri daripada tiga proses utama iaitu eksplorasi, pemilihan dan eksploitasi. Beberapa cabaran penting seperti mencari strategi pemilihan yang betul agar dapat membawa populasi untuk menumpu bersama tanpa mengorbankan kualiti penyelesaian, berurusan dengan mekanisma adaptif untuk eksploitasi yang lebih baik semasa proses carian, mewujudkan keseimbangan antara eksplorasi dan eksploitasi bagi mengelak daripada penumpuan pra-matang dan terperangkap dengan masalah optimum tempatan. Penyelidikan ini bertujuan pertamanya untuk memacu populasi kepada penyelesaian yang lebih baik dengan mengambil kira prinsip strategi pemilihan. Kedua, untuk meningkatkan eksplorasi ruang carian dengan menukar operator kejiranan secara adaptif semasa proses pencarian. Akhir sekali, kaedah penghibridan digunakan bagi mengimbangi proses eksplorasi dan eksploitasi, mengatasi penumpuan pra-matang dan mengelak daripada terperangkap dengan mudah dalam optimum tempatan. Bagi mencapai matlamat ini, tiga model algoritma LBKP iaitu Koloni Lebah Buatan (KLB), Algoritma Lebah (AL) dan Pengoptimuman Koloni Lebah (PKL) telah dicadangkan dan diuji ke atas dua kategori set data iaitu penjadualan peperiksaan tidak berkapasiti dan set data pertandingan penjadualan antarabangsa (ITC2007). Tiga strategi pemilihan, iaitu kejohanan, pangkat dan gangguan telah diuji. Keputusan ujian menunjukkan bahawa strategi pemilihan gangguan menunjukkan prestasi yang lebih baik berbanding strategi pemilihan kejohanan dan pangkat apabila digabungkan dengan KLB, AL dan PKL (dikodkan sebagai GKLB, GAL dan GPKL). Di samping itu, satu mekanisma adaptif-kendiri untuk carian kejiranan telah dilaksanakan pada GKLB, GAL dan GPKL, dan berupaya meningkatkan kualiti penyelesaian (dikodkan sebagai Adaptif-kendiri GKLB, Adaptif-kendiri GAL dan Adaptif-kendiri GPKL). Ini kemudiannya digabungkan dengan dua algoritma carian tempatan (iaitu Penyepuhlindapan Simulasi, PS, dan Lewat Penerimaan Pendakian Bukit, LPPB) yang menunjukkan bahawa LPPB boleh meningkatkan lagi kualiti penyelesaian berbanding dengan PS (dikodkan sebagai Adaptif-kendiri $GKLB_{LPPB}$, Adaptif-kendiri $GAL_{LPPB}$ dan Adaptif-kendiri $GPKL_{LPPB}$). Perbandingan keseluruhan menunjukkan bahawa Adaptif-kendiri $GPKL_{LPPB}$ berfungsi dengan baik merentasi semua set data dan berupaya menghasilkan dua penyelesaian terbaik berbanding dengan penyelesaian yang sedia ada dalam kajian kesusasteraan, khususnya dalam masalah penjadualan peperiksaan tidak berkapasiti.

# CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

## 1.1 BACKGROUND AND MOTIVATION

Examination timetabling problems (ETTPs) belong to a huge family of scheduling problems which are concerned with distributing a collection of exams to a limited number of timeslots (periods of time) and locations, so that students can sit for the exams (Qu et al. 2009). ETTP is considered as a major administrative activity for a wide variety of educational institutions. ETTPs have been addressed as a difficult optimisation problem, in which the amount of computation required to solve the problem increases exponentially with the problem size (Cooper & Kingston 1995).

ETTP is considered as an active research area that has acquired the attention of the researchers over artificial intelligence and operational research fields. It is a NP-Complete problem (Cooper & Kingston 1995), in which exact approaches are not applicable in finding a (near) optimal solution due to the computational time needed is exponentially increased with respect to the size of the problem. Thus, meta-heuristic approaches become an alternative solution approach for ETTPs due to their ability to search over a very large search space.

Several meta-heuristic approaches have been developed for solving ETTPs which can be classified into two main types, i.e., single-based approaches (e.g., tabu search, simulated annealing, great deluge and variable neighbourhood search) and population-based approaches (e.g., genetic algorithms, ant colony optimisation and memetic algorithms) (Qu et al. 2009). Burke et al. (1995), Carter & Laporte (1996), and Schaerf (1999) have conducted surveys and overviews on a number of algorithmic approaches adapted to solve timetabling problems until the end of 1990s, followed by other survey papers proposed by Qu et al. (2009). Single-based approaches have gained interest by many researchers due to the ability of these approaches to exploit the search space in a short time, but these approaches have some limitations such as a weak exploration and it is easy to get stuck in a local optima (Qu et al. 2009).

In addition, researchers have introduced population-based approaches to solve the examination timetabling problems. The main idea behind the population-based is that the algorithms iteratively improve a number of solutions (Talbi 2009). However, these approaches have some limitations such as they are more concerned with exploration rather than exploitation, premature convergence and low convergence speed.

To overcome the limitations of the single-based and population-based approaches, the hybridisation between population-based approaches with a single-based approach has been addressed for timetabling problems. The aim of the hybridisation is to utilize the benefit of population-based approaches that has the ability of identifying promising areas in the search space and single-based approaches that are good in exploiting the promising area Burke et al. (1995); Blum & Roli (2003); Duong & Lam (2004); Thanh (2007). It is believed that the hybridisation approach is able to give a better performance in obtaining a preferred solution for a given problem (Blum & Roli 2003).

Population-based approaches can be categorised as either Evolutionary Algorithms or Swarm Intelligence based algorithms (Yang 2008; Dreo et al. 2006). These two categories depend on the nature of the phenomenon simulated by the algorithm. Most common Evolutionary Algorithms that are introduced for timetabling problems can be found in, Burke et al. (1994), Colorni et al. (1991), Erben (1996) and Ross et al. (1995). Swarm Intelligence relies on the cooperative behaviour of self-organised systems to develop meta-heuristics that mimic such a system's problem solving (Farooq 2008).

Local communication between individuals and with their environment contributes to the collective intelligence of the social colonies (Kamil et al. 1987). These swarm intelligence characteristics motivated a number of researchers to employ such behaviour in algorithms for timetabling problems, including Ant Colony Algorithm (Dowsland & Thompson 2005; Socha et al. 2003), Fish Swarm Optimisation algorithm (Turabieh & Abdullah 2011), and Honey-bee Mating Optimisation (Sabar et al. 2009, 2012).

With regard to swarm intelligence, researchers are concerned with developing algorithms that model the behaviour of honeybees. Honeybee algorithms are classified into three different groups (Baykasoglu et al. 2007) i.e. marriage behaviour, queen bee behaviour and foraging behaviour. Marriage behaviour starts with a waggle dance by the queen then the mating with the drones is performed. The example of the marriage behaviour of a honeybee algorithm for ETTP can be found in Sabar et al. (2009, 2012). The technique in the queen bee behaviour represents the improvement of the Genetic Algorithm, where the main changes have been made during crossover (Sung 2003).

The foraging behaviour has been recently applied to different complex optimisation problems (Karaboga 2005; Lucic & Teodorovic 2001; Pham et al. 2005). This behaviour tries to model the natural behaviour of real honey bee in finding nectar and sharing the information of food sources to the bees in the hive. Honeybees use some methods like waggle dance to locate the food sources and to search new ones. This makes them a good candidate for developing new intelligent search algorithms.

This motivates the employment of such behaviour for ETTP. Such as honeybee algorithms that are based on foraging behaviour (i.e. Artificial Bee Colony (ABC), Bees Algorithm (BA) and Bee Colony Optimisation (BCO)). Honeybee foraging algorithms are a relatively new member of swarm intelligence which is classified under population-based search algorithms. The similarities and differences between them are presented in detail in Chapter II. The hybridisation of the honeybee foraging algorithms have been proposed in this work that is treated as alternative solution approaches that try to overcome as much as possible the limitation of the single approaches in isolation, and improve the quality of the available solutions in the literature.

## 1.2 PROBLEM STATEMENT

The examination timetabling problem is one of the most common problems that academic institutions face in every semester. The problem is in allocating a set of examinations to a limited number of timeslots. Carter & Laporte (1996) defined the examination timetabling problem as:

> "*The assigning of examinations to a limited number of available time periods in such a way that there are no conflicts or clashes*".

The huge number of students in a university makes the examination timetabling problem more complex. Even in moderately sized universities, the manual timetable usually requires many person-days of effort. In some universities, faculties scheduled the examinations for specialised courses only and the schedule for the common courses is prepared by the university. This sometimes causes the generated timetable to be unsuitable for students because examinations are scheduled too close to each other (but at the same time administrators want to shorten the duration of the examination period). Even if the examination timetable is prepared centrally, it may still contain elements which some students find unacceptable. These situations show the complexity in generating an examination timetable that can fulfill preferences of different groups of people.

In order to generate a feasible timetable, it must satisfy a predetermined set of hard constraints (which cannot be violated in any circumstances) and satisfying as much as possible a given set of soft constraints (Carter & Laporte 1996; Silva et al. 2004). But, satisfying all the soft constraints is very difficult, or may even be impossible (Qu et al. 2009). Rudova & Murray (2003) stated that an examination timetable should be preserved as a timetable over hard constraints and optimising the soft constraints as much as possible.

Literature review shows that the researchers have recently focused on local search approaches rather than population-based approaches for timetabling problems (Chiarandini et al. 2006). This may due to the quality of the solutions produced by local search approaches being better than the ones produced by population-based approaches. This may be caused by the fact that the population-based approaches are more concerned with exploration rather than exploitation. While in the exploitation process, the selected solutions may get trapped in local optima.

The selection strategy is one of the main search components in population-based algorithms that decides which solutions are chosen for exploitation (Talbi 2009). The principle of the selection methods in population-based approaches is biased to the better solutions where these solutions have more chances to be selected for exploitation. Such a selection procedure will drive the population to better solutions. However, worse solutions should not be eliminated and they also should have a chance to be selected (Talbi 2009). Most of the population-based meta-heuristic approaches concentrated on finding only one high quality solution at the end of the search.

Talbi (2009) highlighted another issue in the search algorithms, which mention that: "*For an intermediate landscape, the search algorithms should balance the intensification and the diversification, but this task is uneasy. According to this assumption, a promising approach consists in modifying the structure of the landscape to make it easier, that is, composed of a few deep valleys. The landscape can be modified by changing the neighbourhood operator or the objective function.*"

In the light of the above problems (i.e., population-based approaches are more concerned with exploration rather than exploitation, and selection strategy in the population-based approaches is biased to the better solutions), the research questions are as follow:

1. What is the right selection strategy to bring the population to converge together and lead to improvement without sacrificing the quality of the solutions?

2. How to enhance the neighbourhood search in the population-based approaches during the search process?

3. How to attain a balance between exploration and exploitation of the search that leads to better quality of the solutions.

4. Which model of foraging behaviour can work well for examination timetabling problems?

The main interest is to obtain good enough quality solutions by trying to answer the stated research questions through the employment of the honeybee algorithms based on foraging behaviour. Honeybee foraging algorithms still need more investigation in terms of selection strategy involved in order to bring all the solutions in the population to be as good as possible simultaneously.

## 1.3 RESEARCH OBJECTIVES

This study concerns primarily the investigation of different models of honeybee foraging behaviour algorithms (i.e. Artificial Bee Colony (ABC), Bees Algorithm (BA) and Bee Colony Optimisation (BCO)) that could improve the available search methodologies for examination timetabling problems. The main goal is to bring all the solutions in the honeybee foraging algorithms population to be as good as possible , and to provide a suitable balance between exploration and exploitation (between honeybee foraging algorithms and local search algorithms) in order to improve the quality of the examination timetable. To achieve the main goal, several objectives are outlined:

i. To investigate the impact of using three selection strategies, namely, tournament, rank and disruptive selection strategies in honeybee foraging algorithms over examination timetabling problems to drive the population to better solutions.

ii. To enhance the neighbourhood search in the honeybee foraging algorithms by introducing a self-adaptive mechanism to adaptively select a neighbourhood structure based on the search progress.

iii. To enhance the honeybee foraging algorithms through the hybridisation with local search algorithms, including the Late Acceptance Hill Climbing and Simulated Annealing algorithms, in order to have a balance between the exploration and exploitation.

iv. To compare different modeling of foraging behaviour, and identify which foraging behaviour model suits most to ETTPs.

## 1.4 RESEARCH SCOPE

The scope of this study is on honeybee algorithms based on foraging behaviour for examination timetabling problems. The algorithms are tested on two standard benchmark datasets, including uncapacitated examination timetabling problems (12 datasets) proposed by Carter et al. (1996), and examination timetabling problems in the International Competition datasets ITC2007-Track 1 (8 competition datasets) (McCollum et al. 2010).

## 1.5 RESULT SUMMARY

The comparisons of results are conducted in two stages. The first stage is the comparison between different modifications of the proposed approaches, including the experimental comparison of different selection strategies, self-adaptive mechanism and followed by the experimental comparison on different types of hybridisation for honeybee foraging algorithms. In the second stage, the experimental comparison with state-of-the-art approaches is conducted.

First experimental results show that the disruptive selection strategy performed better than tournament and rank selections when embedded with ABC, BA and BCO (coded as DABC, DBA and DBCO, respectively). In addition, a self-adaptive mechanism for the neighbourhood search was employed within DABC, DBA and DBCO algorithms and was able to improve the quality of the solution (coded as Self-Adaptive DABC, Self-Adaptive DBA and Self-Adaptive DBCO). These were then incorporated with two local search algorithms (i.e. Simulated Annealing, SA and Late Acceptance Hill Climbing, LAHC) which showed that the LAHC can further enhance the quality of the solutions in comparison with SA (coded as Self-Adaptive $DABC_{LAHC}$, Self-Adaptive $DBA_{LAHC}$, Self-Adaptive $DBCO_{LAHC}$). Overall comparisons in the second stage indicate that Self-Adaptive $DBCO_{LAHC}$ works well across all tested datasets and is able to obtain some best results in comparison with best known results in the literature.

In conclusion, the research carried out here managed to answer the stated research questions and achieve the listed objectives as summarised in Figure 1.1.


Figure 1.1 Research structure

## 1.6 THESIS ORGANISATION

This thesis contains total of eight chapters including the current chapter. Chapter I is the introduction that covers the research background and motivation, research problem and overview of the research idea.

Chapter II presents the literature review in examination timetabling problems. It introduces the timetabling problem in general, and then concentrates upon reviews and analyses the current published researches on this problem. The available tested datasets are also presented together with the best-known results in the literature.

Chapter III demonstrates the research methodology used in this thesis. It consists of three main phases starting from the identification of the problem domain, identification of the possible solution approaches, and finally the performance evaluation phase.

Chapter IV investigates the implementation of three basic honeybee foraging algorithms (artificial bee colony, ABC, bee algorithm, BA, and bee colony optimisation algorithm, BCO) over examination timetabling problems. The aim of the preliminary experiment is to study the behaviour of the standard selection strategy that is embedded with the basic honeybee foraging algorithms in selecting the solutions in the population, where the roulette wheel selection strategy is used in the ABC and BCO algorithms, and the fitness function is used in the BA algorithm.

Chapter V tries to overcome the limitation occurring in Chapter IV, where the standard selection strategies within the basic honeybee algorithms fail to bring all the solutions to converge together at the end of the search process. In this chapter, three different selection strategies (i.e., rank, tournament and disruptive) with the basic honeybee algorithms are investigated in order to see their performance in comparison with the standard selection strategies. Furthermore, a self-adaptive mechanism is examined to further improve the neighbourhood search in finding neighbouring solutions.

Chapter VI hybridises the best performance modified honeybee algorithm (that has been embedded with a selection strategy and a self-adaptive mechanism) with local search algorithms to quickly explore the search and further enhance the quality of the solution. Two local search algorithms considered in this chapter are simulated annealing (SA) and late acceptance hill climbing (LAHC) algorithms, since they have a capability in accepting a worse solution compared to honeybee foraging algorithms that only accept an improved solution. This capability is believed can avoid the algorithm from getting stuck into local optima. Thus, better solutions can be obtained.

Chapter VII presents an analysis and evaluation based on the results obtained from the best performed algorithm in this work in comparison with other available approaches in the literature with respect to the examination timetabling problem.

Finally the overall conclusions of the work presented in this thesis and research directions for future work in this area are presented in Chapter VIII.

# CHAPTER II

# BACKGROUND AND LITERATURE REVIEW

## 2.1 INTRODUCTION

This chapter emphasis particularly on different available approaches in the literature related to examination timetabling problems. It also describes the definition and the classification of examination timetabling problems, together with the standard benchmark datasets used in this work. Note that, the discussion on the available approaches in this chapter is arranged based on the different classification of the examination timetabling problem. Due to the huge number of published work in this area, this chapter only focuses on the most significant works in the literature. In addition, an overview of honeybee algorithms is also provided, since these algorithms are investigated in this work.

The definition of examination timetabling problems with their specification and formulation are given in the next section. Section 2.3 discusses the approaches applied on different categories of examination timetabling problems i.e., uncapacitated and capacitated timetabling problems. The finding in terms of the strength and limitation on the employed approaches to date is summarised and also presented in this section. Section 2.4 presents an overview of honeybee algorithms. Finally, Section 2.5 presents a brief summary of this chapter.

## 2.2 EXAMINATION TIMETABLING PROBLEMS

The examination timetabling problem is an important problem for academic institutions. It gives a great challenge to the institution in developing a good solver since it affects different groups of peoples including administrators, academic staff and students (Romero 1982). It is a difficult process due to the fact that constraints and demands required by educational institutions are increasing and changing over time, and is different between institutions.

Over the years, researchers have provided their own definition on the examination timetabling problem. Balakrishnan (1991) gave the definition as "*the examination timetabling problem typically involves the assignment of exams to specific periods and classrooms in order to obtain a schedule that uses a minimum number of periods and satisfies a number of different objectives*".

Wren (1996) defined timetabling as "*Timetabling is the allocation, subject to constraints, of given resources to objects being placed in space-time, in such a way as to satisfy as nearly as possible a set of desirable objectives*".

Carter and Laporte (1996) defined the examination timetabling as "*the assigning of examinations to a limited number of available time periods in such a way that there are no conflicts or clashes*".

According to Schaerf (1999), examination timetabling can be defined as the scheduling for the examinations of a set of university courses, avoiding overlapping examinations of courses having common students, and spreading the examinations for the students as much as possible.

Qu et al. (2009) stated that "*examination timetabling problem involve assigning a set of exams* $E=e_1, e_2, \ldots, e_e$ *into a limited number of available timeslots* $T = t_1, t_2, \ldots, t_t$ *in such a way that there are no conflicts or clashes*".

According to the above definitions of the examination timetabling problem by different researchers, it can be determined that examination timetabling problem deal with an allocation of given resources (students and rooms) to objects (examinations) placed in timeslots to satisfy all the constraints associated with some resources.

A number of research papers have proposed several models and formulations for examination timetabling problem. For example, de Werra (1985) illustrates how a timetabling problem can be modelled using a graph. The main aim of graph colouring is to minimise the number of colours that can be used to colour the graph's vertices as much as possible, by taking into account that none of the linked adjacent vertices have the same colour (normally there are a limited number of colours available). The number of colours is known as the 'chromatic number' of a graph. This can be linked to the examination timetabling problem (in its simplest form), where the vertices represent the examinations, colours represent the timeslots and the edges represent the conflict between examinations (students taking both corresponding examinations at one timeslot) (Burke et al. 2004). In the light of the above, the examination timetabling problem can be considered as a graph colouring problem, in which the goal is to find the minimum number of timeslots which are able to contain all the examinations without any clashes.

The graph colouring problem and its relationship to timetabling in constructing a no-clash timetable is widely discussed in the literature (see examples in Welsh & Powell 1967; Brelaz 1979; De Werra 1985; Carter 1986; Carter et al. 1994; Carter & Laporte 1996; Burke & Ross 1996; Burke et al. 1994; Burke et al. 2004).

Generally, examination timetabling problem can be classified into two categories i.e., uncapacitated and capacitated examination timetabling problems. The description of both categories with their associated benchmark datasets are presented in the following sections. In this work we considers both categories of examination timetabling problems

### 2.2.1 Uncapacitated Examination Timetabling Problems

Uncapacitated examination timetabling problems refer to the assignment of examinations to a limited number of timeslots in such a way that there is *no clash* between examinations. No conflict refers to the situation where no student is needed to take more than one examination simultaneously.

## A. Standard Benchmark Dataset

The standard benchmark for the uncapacitated examination timetabling problems was introduced by Carter et al. (1996) which are also known as Toronto Benchmark dataset that can be freely downloaded from ftp://ftp.mie.utoronto.ca/pub/carter/testprob/. They introduced 13 real world case studies collected from three Canadian highs schools, five Canadian universities, one American university, one British university and one university in Saudi Arabia. However, there was some confusion in the literature due to some datasets appearing in different versions but under the same names as reported by Qu et al. (2009).

These versions are classified as version I, II and IIc where the main differences between them are in terms of number of students, number of enrolments, number of examinations and conflict density as shown in Table 2.1 (Qu et al. 2009).

Table 2.1 Differences between versions I, II and IIc for the uncapacitated datasets

| Datasets | Number of timeslots | Number of examinations | Number of students | Conflict density | Enrolment |
|---|---|---|---|---|---|
| ear83 I | 24 | 190 | 1125 | 0.27 | 8109 |
| ear83 II | 24 | 189 | 1108 | 0.27 | 8014 |
| ear83 IIc | 24 | 189 | 1108 | 0.27 | 8057 |
| hec92 I | 18 | 81 | 2823 | 0.42 | 10632 |
| hec92 II | 18 | 80 | 2823 | 0.42 | 10625 |
| pur93 I | 42 | 2419 | 30032 | 0.03 | 120681 |
| pur93 II | 42 | 2419 | 30032 | 0.03 | 120686 |
| sta83 I | 13 | 139 | 611 | 0.14 | 5751 |
| sta83 II | 13 | 138 | 549 | 0.14 | 5689 |
| sta83 IIc | 35 | 138 | 549 | 0.19 | 5417 |
| uta92 I | 35 | 622 | 21266 | 0.13 | 58979 |
| uta92 II | 35 | 638 | 21329 | 0.13 | 59144 |
| yor83 I | 21 | 181 | 941 | 0.29 | 6034 |
| yor83 II | 21 | 180 | 919 | 0.29 | 6012 |
| yor83 IIc | 21 | 180 | 919 | 0.30 | 6002 |

Qu et al. (2009) have carefully examined the data and in order to avoid future confusion, they listed the definitive versions of these datasets which are available at http://www.asap.cs.nott.ac.uk/resources/data.shtml.

Table 2.2 Characteristics of the uncapacitated datasets

| Datasets | Institution | Number of timeslots | Number of examinations | Number of students | Conflict density |
|---|---|---|---|---|---|
| car91 | Carleton University, ( | 35 | 682 | 16925 | 0.13 |
| car92 | Ottawa) | 32 | 543 | 18419 | 0.14 |
| ear83 I | Earl Haig Collegiate Institute, Toronto | 24 | 190 | 1125 | 0.27 |
| hec92 I | Ecole des Hautes Etudes Commercials, Montreal | 18 | 81 | 2823 | 0.42 |
| kfu93 | King Fahd University, Dharan | 20 | 461 | 5349 | 0.06 |
| lse91 | London School of Economics | 18 | 381 | 2726 | 0.06 |
| rye92 | Ryeson University, Toronto | 23 | 486 | 11483 | 0.07 |
| sta83 I | St. Andrew's Junior High School, Toronto | 13 | 139 | 611 | 0.14 |
| tre92 | Trent University, Peterborough, Ontario | 23 | 261 | 4360 | 0.18 |
| uta92 I | Faculty of Arts and Sciences, University of Toronto | 35 | 622 | 21266 | 0.13 |
| ute92 | Faculty of Engineering, University of Toronto | 10 | 184 | 2749 | 0.08 |
| yor83 I | York Mills Collegiate Institute, Toronto | 21 | 181 | 941 | 0.29 |

In this work, datasets from version I have been considered as shown in Table 2.2. Note that the pur93 I dataset is not considered here due to the expensive computational time to generate the solution.

### B. Data Specification and Formulation

The problem description employed in this work is adopted from the description presented in Burke et al. (2004), where the inputs for the problem are stated as follows:

- $N$ is the number of examinations.
- $E_i$ is an examination, $i \in \{1\ldots N\}$.
- $T$ is the given number of available timeslots.

- $M$ is the number of students.
- $C = (c_{ij})_{NxN}$ is the conflict matrix with each element denoted by $c_{ij}$, $i,j \in \{1,\ldots,N\}$ is the number of students taking examinations $i$ and $j$.
- $t_k$ ($1 \leq t_k \leq T$) specifies the assigned timeslot for examination $k$ ($k \in \{1,\ldots,N\}$).

In this problem, an objective function is to space out students' examinations throughout the examination period. The problem can be formulated as the minimisation of the sum of proximity costs as formulated below (Burke & Newall 2004):

$$Min\ \frac{\sum_{i=1}^{N-1} F_1(i)}{M} \tag{2.1}$$

where

$$\mathrm{F_1(i)} = \sum_{\mathrm{j=i+1}}^{\mathrm{N}} \mathrm{C_{ij}} \,.\, \mathrm{proximity}\left(t_i\,, t_j\right) \tag{2.2}$$

and

$$proximity\left(t_i\,, t_j\right) = \begin{cases} 2^5 / \, 2^{|t_i - t_j|} & if\ 1 \leq |t_i - t_j| \leq 5 \\ 0 & otherwise \end{cases} \tag{2.3}$$

subject to:

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} C_{ij} \,.\, \gamma\left(t_i\,, t_j\right) = 0$$

where

$$\gamma\left(t_i\,, t_j\right) = \begin{cases} 1 & if\ t_i = t_j \\ 0 & otherwise \end{cases} \tag{2.4}$$

Equation 2.2 represents the examination cost, which is the proximity value, multiplied by the number of conflicting students. Equation 2.3 presents a proximity value between two examinations (Carter et al. 1996). Equation 2.4 represents a *no clash* requirement so no student is asked to sit for two examinations at the same time.

### C. Solution Representation

The solution is presented as a 2 dimensional array where the first column represents the examination index, and the second column represents the timeslot as shown in Figure 2.1. For example, examination 12 is scheduled at timeslot 8, examination 45 is scheduled at timeslot 2, and so on.

| | |
|---|---|
| 12 | 8 |
| 45 | 2 |
| .<br>.<br>. | .<br>.<br>. |
| 67 | 12 |

Figure 2.1 Solution representation for uncapacitated examination timetabling problems

## 2.2.2 Capacitated Examination Timetabling Problems

This problem deals with the assignment of examinations to a limited number of timeslots and rooms, in such a way that no student is required to sit for more than one examinations at the same time, and the seating capacity in the room (where the examination is scheduled in it) must be at least equal to or greater than the number of students taking the examination. The requirement of a seating capacity differentiates between uncapacitated and capacitated examination timetabling problems. Thus, this problem considers a room capacity requirement along with a *no clash* requirement while constructing a timetable. In this work, the capacitated examination timetabling problems are referred as competition examination timetabling problems.

### A. Standard Benchmark Dataset

There are different datasets that are categorised under capacitated examination timetabling problems such as the University of Nottingham benchmark data (Burke et al. 1996) and University of Melbourne benchmark data (Merlot et al. 2003). However, newly created datasets that were introduced by McCollum et al. (2010) are experimented within this work in order to examine the performance of the proposed approaches. From now on we refer to the standard benchmark datasets for the capacitated examination timetabling problems as competition datasets.

Competition datasets represent an examination timetabling model that incorporates a significant number of real-world constraints. This formulation was introduced as part of the second International Timetabling Competition (competition datasets)-Track 3. Competition datasets contain real-world constraints, which are considered as complex and more practical datasets than the uncapacitated datasets. The benchmark instances for this problem are taken from http://www.cs.qub.ac.uk/itc2007/index.htm. Table 2.3 shows the characteristics of these datasets.

Table 2.3 Characteristics of the competition datasets

| Datasets | D1 | D2 | D3 | D4 | D5 | D6 | CD |
|---|---|---|---|---|---|---|---|
| Exam_1 | 7891 | 607 | 54 | 7 | 12 | 0 | 5.05 |
| Exam_2 | 12743 | 870 | 40 | 49 | 12 | 2 | 1.17 |
| Exam_3 | 16439 | 934 | 36 | 48 | 170 | 15 | 2.62 |
| Exam_4 | 5045 | 273 | 21 | 1 | 40 | 0 | 15.0 |
| Exam_5 | 9253 | 1018 | 42 | 3 | 27 | 0 | 0.87 |
| Exam_6 | 7909 | 242 | 16 | 8 | 23 | 0 | 6.16 |
| Exam_7 | 14676 | 1096 | 80 | 15 | 28 | 0 | 1.93 |
| Exam_8 | 7718 | 598 | 80 | 8 | 20 | 1 | 4.55 |

Where: D1= Number of students reported in McCollum et al. (2007). D2= Number of exams. D3=Number of timeslots. D4= Number of rooms. D5= Period hard constraints. D6=Room hard constraints. CD= Conflict Density.

A complete description of the datasets and the objective function are described below which are also available in McCollum et al. (2010).

## B. Data Specification and Formulation

The feasibility of the timetable in the competition datasets relates to assigning all examinations to a period and room and not violates the hard constraints. The hard constraints are listed below (McCollum et al. 2010):

- No student sits for more than one exam at the same time.
- The total number of students assigned to each room cannot exceed the room capacity.
- The length of exams assigned to each timeslot should not violate the timeslot length.
- The exam sequences must be respected; for example, *Exam_A* must be scheduled after *Exam_B*.
- Room-related hard constraints must be satisfied; for example, *Exam_A* must be scheduled in *Room 2*.

The objective function minimises the violation of soft constraints, as in Equation 2.5 (McCollum et al. 2010):

$$min \sum_{s \in S} (W^{2R}\, C_S^{2R} + W^{2D}\, C_S^{2D} + W^{PS}\, C_S^{PS}) + (W^{NMD}\, C_S^{2NMD} + \mathrm{W}^{\mathrm{FL}}\, \mathrm{C}^{\mathrm{FL}} + \mathrm{W}^{\mathrm{P}}\, \mathrm{C}^{\mathrm{P}} + \mathrm{W}^{\mathrm{R}}\, \mathrm{C}^{\mathrm{R}}) \tag{2.5}$$

Each dataset has its own weight(W), as shown in Table 2.4.

Table 2.4 Associate weight of the competition datasets

| Data sets | $W^{2D}$ | $W^{2R}$ | $W^{PS}$ | $W^{NMD}$ | $\mathrm{W}^{\mathrm{FL}}$ | $\mathrm{W}^{\mathrm{P}}$ | $\mathrm{W}^{\mathrm{R}}$ |
|---|---|---|---|---|---|---|---|
| Exam_1 | 5 | 7 | 5 | 10 | 100 | 30 | 5 |
| Exam_2 | 5 | 15 | 1 | 25 | 250 | 30 | 5 |
| Exam_3 | 10 | 15 | 4 | 20 | 200 | 20 | 10 |
| Exam_4 | 5 | 9 | 2 | 10 | 50 | 10 | 5 |
| Exam_5 | 15 | 40 | 5 | 0 | 250 | 30 | 10 |
| Exam_6 | 5 | 20 | 20 | 25 | 25 | 30 | 15 |
| Exam_7 | 5 | 25 | 10 | 15 | 250 | 30 | 10 |
| Exam_8 | 0 | 150 | 15 | 25 | 250 | 30 | 5 |

The soft constraints for the competition datasets are listed and coded as below (McCollum et al. 2010):

- ($C_S^{2R}$): Student has to sit two examinations in a row
- ($C_S^{2D}$): Student has two examinations in a day.

- ($C_S^{PS}$): Students should have a fair distribution of examinations over their timetable.
- ($C_S^{2NMD}$): Mixed durations of examinations scheduled in the same room.
- ($C^{FL}$): Larger examinations appearing later in the timetable.
- ($C^{P}$): Period-related soft constraint.
- ($C^{R}$): Room-related soft constraint.

### C. Solution Representation

The solution is also represented as a 2 dimensional array where the first column represents the timeslot and the second column represents the room as shown in Figure 2.2. For example, the first examination is scheduled at timeslot 10 in room 3, the second examination is scheduled at timeslot 5 in room 2, and the last examination is scheduled at timeslot 7 in room 6.

| | |
|---|---|
| 10 | 3 |
| 5 | 2 |
| .<br>.<br>. | .<br>.<br>. |
| 7 | 6 |

Figure 2.2 Solution representation for capacitated examination timetabling problems (particularly for competition datasets)

## 2.3 APPROACHES APPLIED ON EXAMINATION TIMETABLING PROBLEMS

Several approaches have been previously introduced to tackle examination timetabling problems. These approaches have been highlighted and discussed in the comprehensive surveys presented by Carter (1986); Carter & Laporte (1996); Burke et al. (1995); Schaerf (1999); Qu et al. (2009), are studies one or more standard benchmark datasets were included.