Constructive versus Improvement Heuristics: An Investigation of Examination Timetabling

Edmund Burke, Graham Kendall
University of Nottingham, United Kingdom
Barry McCollum, Paul McMullan [1]
Queen's University of Belfast, United Kingdom, p.p.mcmullan@qub.ac.uk

1. Introduction

Examination timetabling has been a widely studied research problem for over 30 years, with many techniques being used to produce good quality solutions. The majority of the techniques rely on producing an initial solution using a constructive heuristic, and then improving it via a local search [1,2,3], relying on the correct setting of internal parameters to allow generally acceptable results for different data sets. Unfortunately, in attempting to provide good quality results for all data sets, in many cases we may not achieve the best possible solutions. From the author's practical experience, it has been observed that human intervention is often necessary to adapt the scheduling techniques used for the particular characteristics of each data set. The goal of this paper is to remove this human intervention where possible, and allow full automation for all problem instances. We investigate a number of benchmark problem data sets, of varying degrees of complexity, to see if there is a relationship between the difficulty of the problem and the effects of combining constructive and improvement heuristics in achieving the best possible solution under specified time constraints. In this way we can attempt to generalise the link between the problem instances and the approach taken in gaining a solution. We use the Carter benchmark datasets [4], which represent real world instances and whose characteristics in terms of measure of difficulty are well known. We use the commercial examination timetabling system, Optime [5] to generate the results. Our results show that there is a relationship between the difficulty of the problem and the time spent in each part of the solution methodology.

2. Data Sets

We are using the Carter benchmark data sets [4] as these are widely used in other work. The difficulty measure is based on the *conflict density* value, calculated as the average number of all other exams that each exam conflicts with, divided by the total number of exams. For instance, a density of 0.25 denotes that, on average, each exam conflicts with 25% of the other exams. It is proposed that this is an obvious initial starting point in establishing a correlation between difficulty and appropriately tailored scheduling heuristics. If a link can be established this, could lead to a more fine-grained analysis of the characteristics of data sets in relation to various parameter settings for existing scheduling techniques, and can build on current research into the categorisation and similarity measure analysis of examination data sets [6].

3. Solution Methodology

In general, successful scheduling techniques have employed a two-phase approach to establishing a timetable; construction to produce a feasible solution and improvement, employing intelligent search techniques to find high quality solutions given specified objectives. It has been observed that if we continue construction beyond the point at which we have produced a feasible solution that we often achieve a better quality solution on which the improvement phase can then operate

[1] Corresponding Author

[7]. In conjunction with an improvement phase, an even higher quality solution could be possible. However, in practice the same relative time spent on construction and improvement respectively may not yield the best results (given limiting factors such as time constraints) for each data set. For instance, one data set may favour more construction over improvement, while another will favour a greater emphasis on improvement. At this point human intervention or trial and error are required, and we cannot say the process is truly automatic.

The main purpose of this study is to ascertain if the relative time spent in each of the two phases relates to the difficulty of the problem instance, with the aim of generalising the process without the trade-off of solution quality. The data sets used have a wide range of conflict density values, which will help to identify clear distinctions between difficulty and results based on the tested combinations of construction / improvement heuristic under analysis.

The Optime examination system uses this two phase approach when generating timetable solutions for examination data sets. Initial solution construction is carried out by an Adaptive (Squeaky-Wheel) ordering heuristic [7] technique. This utilises a weighted order list of the exams to be scheduled, the initial ordering based on the *degree* (number of conflicts) of each exam. This allows an initial estimation of a list with the most difficult exams being placed first. Each exam weighting is then increased, depending on the difficulty or penalty of its placement in the schedule, which allows the ordering to adapt as difficult exams are encountered. Although construction techniques in general are simply used to establish a feasible solution, the adaptive heuristic can continue to improve a feasible solution as the exam ordering changes due to the weighting.

The improvement phase takes the feasible timetable from the constructive heuristic and implements the Great Deluge Algorithm [8] as a local search method. The Great Deluge (also known as Degraded Ceiling) was introduced by Dueck [9] as an alternative to Simulated Annealing [10,11]. This uses a boundary condition to accept worse solutions, in order to escape local optima. The boundary is initially set slightly higher than the initial solution cost, and gradually reduced throughout the improvement process. New solutions are only accepted if they improve on the current cost evaluation or are within the boundary value. This approach has previously been successfully applied to construction and improvement techniques in timetabling problems [8].

The evaluation function to drive both construction and improvement in the search for improved solutions is that used within the commercial examination timetabling system, Optime [5]. It is used in preference to that traditionally used for the Carter data sets [4], as the experiments were run through the Optime software. This allows many more constraints than those included within the Carter evaluation function to be included, in order to extend the analysis to further data sets. The Carter evaluation function is primarily concerned with minimization of the proximity penalties caused by exams which are scheduled a specified number of periods apart. Each occurrence of a student taking exams within proximity limits will add to the total penalty. The Optime evaluation function [5] takes this proximity measure into account, as well as accommodating many other soft constraints considered in real-world scheduling problems.

4. Results

Initial results are encouraging in being able to establish a link between construction and improvement. The experiments have involved allocating a given number of total solution generations for the entire process, and dividing this allocation by stepped percentages for construction and improvement respectively. For example, one test run for a single data set would involve an initial setting of 0% of the total evaluations spent on construction, 100% spent on improvement, then 10% on construction and 90% on improvement and so on. This is repeated for a single data set to get an average or typical set of results, and performed for each of the data sets under consideration.

Figure 1 presents the initial results achieved from the average of three test runs for each data set. The construction value is presented as a value between 0 and 10, representing 0% to 100% of time spent on construction within the entire test run. The construction value achieved is that setting at which the best quality solution was found for each of the 11 different combinations of construction and improvement time. As can be seen, there is a direct (inversely proportional) correlation between construction time to the difficulty (in terms of conflict density) of the data set.

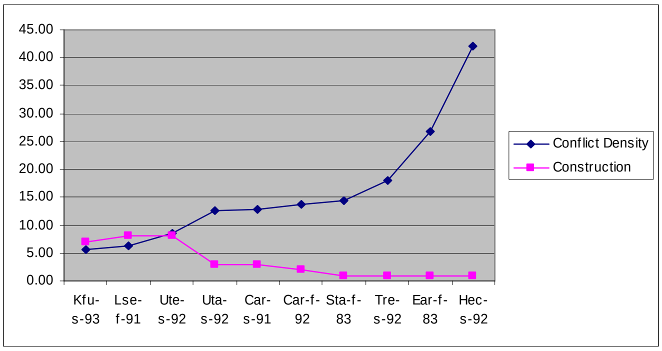

Figure 1 – Data Sets used in this Study

5. Discussion and Conclusions

We have shown there is a link between the difficulty (in terms of conflict density) of examination data sets and the relative time spent on construction and improvement when creating timetable solutions. Each different scheduling technique has many parameter settings, with a certain dependence on the human user to provide the right balance. The work outlined provides scope for further analysis in removing this dependency. Further work will be to investigate whether this observation holds for actual real-world examination data sets, drawn from our commercial scheduling package.

References

[1] Burke E. K., Trick, M. (eds), (2005), The Practice and Theory of Automated Timetabling V: Revised Selected Papers from the 5th International conference, Pittsburgh 2004, Springer Lecture Notes in Computer Science, Vol. 3616, Springer 2005.
[2] White G.M., Xie B. S., Zonjic S., (2004), "Using tabu search with longer-term memory and relaxation to create examination timetables", European Journal of Operational Research, 153(16): 80-91.
[3] Qu R., Burke E. J., McCollum B., Merlot L., Lee S., (2006), "A Survey of Search Methodologies and Automated Approaches for Examination Timetabling", Computer Science Technical Report No. NOTTCS-TR-2006-4.
[4] Carter, M. W., Laporte, G., Lee, S., (1996), "Examination timetabling: Algorithmic strategies and applications", Journal of the Operational Research Society, 47(3): 373-383.
[5] Burke E. K., Kendall G., McCollum, B., McMullan, P. J. P., Newall, J. P., (2006), "A Preference Based Measurement of Optimisation", ASAP Technical Report, eMAP/2006/02a.
[6] Yang Y., Petrovic, S., (2004) "A Novel similarity measure for heuristic scheduling in examination timetabling", Selected Papers from 5th International conference on the Practice and Theory of Automated Timetabling, Springer Lecture Notes in Computer Science, 3616, 377-396.
[7] Burke, E. K., Newall, J. P., (2004), "Solving Examination Timetabling Problems through Adaptation of Heuristic Orderings, Annals of Operational Research, 129: 107-134.
[8] Burke, E. K., Bykov, Y., Newall, J. P., Petrovic, S., (2003) "A Time-Predefined Approach to Course Timetabling", Yugoslav Journal of Operational Research (YUJOR), 13(2): 139-151.
[9] Dueck, G., (1990), "Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing", Journal of Computational Physics, 90: 161-175.
[10] Kirkpatrick, S., Gellat, J. C. D., Vecci, M. P., (1983), "Optimization by Simulated Annealing", Science, 220: 671-680.
[11] Aarts E., Korst J., Michiels W. (2005), "Simulated Annealing", In Burke E. K., Kendall G. (eds), (2005) "Search Methodologies: Introductory Tutorials in Optimization and Decision Support Methodologies", Springer 2005, Chapter 7, 187-210.