A Study on the Short-Term Prohibition Mechanisms in Tabu Search for Examination Timetabling

Luca Di Gaspero[1], Marco Chiarandini[2], and Andrea Schaerf[1]

1. Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica, Università di Udine, via delle Scienze 208, I-33100, Udine, Italy,
2. Dept. of Mathematics & Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M – Denmark.
email: `l.digaspero@uniud.it` `marco@imada.sdu.dk` `schaerf@uniud.it`

1 Introduction

Tabu Search (TS) is a well known local search method [11] which has been widely used for solving timetabling problems. Different versions of TS have been proposed in the literature, and many features of TS have been considered and tested experimentally. They range from long-term tabu, to dynamic cost functions, to strategic oscillation, to elite candidate lists, to complex aspiration criteria, just to mention some (see [10] for an overview).

The feature that is included in virtually all TS variants is the so called *(short-term) tabu list*. The tabu list is indeed recognized as the basic ingredient for the effectiveness of a TS-based solution, and its behaviour is a crucial issue of TS.

Unfortunately, despite the fact that the importance of a correct empirical analysis has been recognised in the general context of heuristic methods [1,12] and even in the specific case of TS [15], the definition of the parameters associated with the tabu list remains in most research work still a handcrafted activity.

Often, the experimental work behind the parameter setting remains hidden or is condensed in a few lines of text reporting only the final best configuration. Even the recently introduced *racing* methodology for the tuning of algorithms [3] only allows to determine the best possible configuration. This procedures are certainly justified from a practical point of view, but a description of the behavior of the algorithm with respect to its different factors and parameters is surely of great interest in the research field.

In this work, we aim at determining which factors of basic TS are important and responsible for the good behaviour of the algorithm. Instead of the one-factor-at-a-time approach used in [15], our approach uses experimental design techniques [14] combining the racing methodology for the definition of quantitative factors and the analysis of variance for the study of qualitative factors. We focus our analysis on the EXAMINATION TIMETABLING problem, for which there is a consistent literature and many benchmark instances. In particular, we consider the formulation proposed by Burke *et al* [5], which considers first

E. K. Burke, H. Rudová (Eds.): PATAT 2006, pp. 406–411. ISBN 80-210-3726-1.

and second order conflicts (exams in adjacent periods of the same day), but no capacity of rooms.

We plan to extend the analysis to other formulations, and to other timetabling problems so as to have a more general picture of the outcome.

2 Tabu Search Basic Features

At each iteration, TS explores the full neighborhood and selects as the new current state the neighbor that gives the best value of the cost function, independently of whether its cost is less or greater than the current one. This selection allows the search to *escape* from local minima, but creates the risk of cycling among a set of states. In order to prevent cycling, TS uses a prohibition mechanism based on the tabu list. This list stores the most recently accepted moves, so that the *inverses* of the moves in the tabu list are forbidden (i.e., the moves that are leading again towards the just visited local minimum). The two main features related to the tabu list are the following:

Prohibition power: The prohibition power determines which moves are prohibited by the fact that a move is in the tabu list. A move is normally composed of several attributes; depending on the power, the prohibition can be applied only to the move with the same values for all attributes or to the set of moves that have one or more attribute equal to it.

List dynamics: The list dynamics determines for how many iterations a move remains in the tabu list. This can be either a fixed value, or a value selected randomly in an interval, or value selected adaptively on the basis of the current state of the search.

For the Examination Timetabling problem, we consider the search space and the neighborhood relation as defined in [8]. That is, we create one variable per exam with domain equal to the set of periods, and change the value of one single exam at a time. In this setting, a local search move m has three attributes: an exam e, its old period o and its new period n. We identify m with the triple $\langle e, o, n \rangle$.

For the prohibition power, assuming that the move $\langle e, o, n \rangle$ is in the tabu list, we consider the following three alternatives (where the underscore means "any value"):

Strong: All moves of the form $\langle e, \_, \_ \rangle$ are prohibited.
Medium: All moves of the form $\langle e, \_, o \rangle$ are prohibited
Weak: Only the single move $\langle e, n, o \rangle$ is prohibited

Intuitively, in the first case, it is not possible to move the exam anywhere in the tabu iterations. In the second case, it is not possible to move the exam back to the old period. In the third case it is not possible only to make the reverse of the tabu move.

For the list dynamics we also consider three possibilities:

Fixed: The tabu list is a queue of fixed size t. At each iteration, the accepted move gets in and the oldest one gets out. All moves remain in the list for exactly t iterations.

Dynamic: The size of the list can vary within the range $[t_m, t_M]$. Each accepted move remains in the list for a number t of iterations equal to $Random(t_m, t_M)$, where $Random(a, b)$ is the uniform integer-valued random distribution in $[a, b]$

Adaptive: The value t depends on the current state. We use the formula (proposed in [9] for graph coloring) $t = \lfloor Random(0, t_b) + \alpha * c \rfloor$ where t_b and α (real value) are parameters, and c is the number of conflicts in the current state.

3 Experimental Methodology and Results

There are two types of factors present in the analysis: quantitative and qualitative. The two qualitative factors are the prohibition power and the list dynamics, and consist each of three levels. The quantitative factors are the numerical parameters of the list dynamics strategy and may assume an unlimited number of values. There are two issues that complicate the factorial design: (i) the qualitative parameters do not cross with all other factors (for example, there is no α parameter with fixed list dynamics); (ii) the importance of each of the two qualitative factors strongly depends on the values assigned to the underlying quantitative parameters.

In order to understand the relative influence of the qualitative tabu list features a possible way is to split the analysis in two phases. This approach allows to maximise the fairness in the analysis, although perhaps it is not the most economical in terms of number of experiments.

In the first phase, for each combination of qualitative factors, we tune the numerical parameters: $\langle t \rangle$, $\langle t_m, t_M \rangle$, or $\langle t_b, \alpha \rangle$. We tested the following values for the parameters: $t \in \{5, 10, 15, 20, 25\}$, $t_m \in \{5, 10, 15, 20, 25\}$, $t_M \in \{t | t = t_m + 5 \vee t = t_m + 10\}$, $t_b \in \{5, 10, 20, 30\}$, and $\alpha \in \{0.3, 0.5, 0.8\}$.

We perform this task by means of the RACE algorithm developed by Birattari [4]. All experiments are run on 7 instances employed in [5] and each configuration was granted 120 seconds of CPU time on an AMD Athlon 1.5GHz computer running Linux. A t-test ($p < 0.05$) is used to discriminate between the configurations. The best parameter settings found across the tested instances are reported in Table 1. Further details on the racing process will be provided in the forthcoming full paper.

In the second phase, after the values of the parameters have been determined, we perform another experiment whose aim is to understand whether there are main effects or interactions between the two factors, list dynamics and prohibition power, and how these affect the performances of the algorithm. To this aim, we run a full factorial design with 25 replicates per instance and perform an analysis of variance [14]. Each experimental unit exploits a different combination of list dynamics, prohibition power and test instance. In the analysis the test instances are then treated as blocks and hence their influence on the performance

	Weak	*Medium*	*Strong*
Fixed	$t = 15$	$t = 10$	$t = 5$
Dynamic	$t_m = 20, t_M = 25$	$t_m = 5, t_M = 10$	$t_m = 5, t_M = 10$
Adaptive	$\alpha = 0.3, t_b = 30$	$\alpha = 0.3, t_b = 30$	$\alpha = 0.5, t_b = 10$

Table 1. Best parameter settings for the various combinations of features

of the algorithms, though recognised, is not taken into account (this entails that the results of the analysis are robust with respect to the set of instances).

Specifically, for the analysis of the $3 \times 3 \times 25$ ({*Strong, Medium, Weak*} $\times$ {*Fixed, Dynamic, Adaptive*} $\times$ replicates) combinations we used two statistical tests. The well known parametric ANOVA, through the F ratio, and the non-parametric Friedman two ways analysis of variance by ranks. Though based on less assumptions, the Friedman test is not able to recognise interaction between the two factors [7]. On the other hand, the F ratio can detect also interactions and is apparently robust even in cases of deviation from the assumptions. In the parametric case we transformed each numerical result, expressed in terms of cost violation of the solution, by standardization of the value within the results per instance. In the non-parametric case the results are instead ranked within the instances, as usual in the Friedman test procedure.

Surprisingly both tests indicated the absence of a significant influence of both main and interaction effects (technically, the F ratio gave a p-value of 0.98 and the Friedman test gave p-value $= 0.54$).

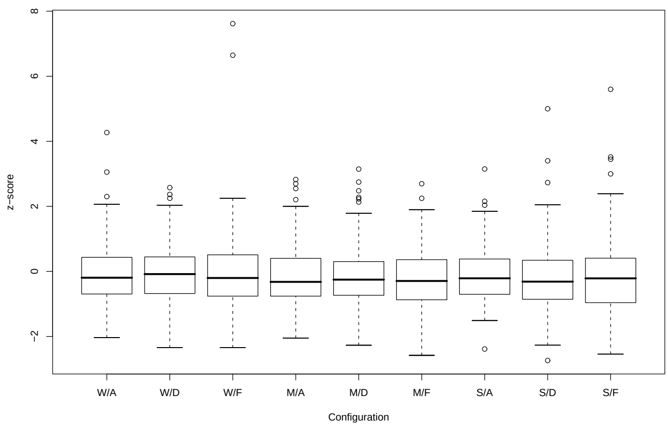

Figure 1. Results of the 9 configurations for the qualitative features (W = Weak, M = Medium, S = Strong, F = Fixed, D = Dynamic, A = Adaptive)

The results are shown in Figure 1 by means of box-and-whiskers plots. A closer insight in the numerical results revealed that indeed there is no important difference in the results. The conclusion is that, if the quantitative parameters are tuned by means of a statistically sound procedure, all configurations of the qualitative parameters perform equally well.

In the future, we plan to investigate the difference in the robustness of the qualitative features, by analysing the sensitivity of the tuning of the quantitative parameters for different configurations. A response surface approach as suggested in [2] would be more appropriate for the selection of quantitative parameters (although it can be computationally more expensive).

We conclude by comparing, in Table 2, our overall best results (in bold) with the currently published ones. From the table we first see that we improved significantly w.r.t. our previous best results ([8]). In addition, although our results are still far from the best ones of Merlot *et al* [13], they are in most cases the second best ones. This improvement is achieved mainly thanks to our statistically sound parameter tuning.

Instance	p	W/A	W/D	W/F	M/A	M/D	M/F	S/A	S/D	S/F	DS [8]	BNW [5]	CDI [6]	MBHS [13]
car-f-92	40	271	292	265	263	278	275	297	**224**	242	424	331	268	158
car-s-91	51	46	38	38	53	54	**32**	37	40	33	88	81	74	31
kfu-s-93	20	1148	1027	1103	1047	**914**	984	1172	1104	932	512	974	912	247
nott	23	112	89	109	103	106	**69**	109	112	73	123	269	—	7
nott	26	17	17	17	**7**	15	15	16	13	20	11	53	—	—
tre-s-92	35	**0**	0	0	0	0	0	0	0	0	4	3	2	0
uta-s-92	38	534	544	526	584	572	**507**	565	537	567	554	772	680	334

Table 2. Best results found for each configuration and comparison with the best known results.

References

1. R. Barr, B. Golden, J. Kelly, M. Resende, and W. Stewart. Designing and reporting on computational experiments with heuristic methods. *J. of Heur.*, 1:9–32, 1995.
2. T. Bartz-Beielstein. Experimental analysis of evolution strategies – overview and comprehensive introduction. Technical Report Reihe CI 157/03, SFB 531, Universität Dortmund, Germany, 2003.
3. M. Birattari. *The Problem of Tuning Metaheuristics as Seen from a Machine Learning Perspective.* PhD thesis, Université Libre de Bruxelles, Belgium, 2004.
4. M. Birattari. *The RACE pacakge*, April 2005.
5. E. Burke, J. Newall, and R. Weare. A memetic algorithm for university exam timetabling. In *Proc. of ICPTAT-95*, pages 241–250, 1995.
6. M. Caramia, P. Dell'Olmo, and G. F. Italiano. New algorithms for examination timetabling. In S. Nher and D. Wagner, editors, *Proc. of WAE 2000*, volume 1982 of *LNCS*, pages 230–241. Springer-Verlag, 2000.
7. W.J. Conover. *Practical Nonparametric Statistics.* John Wiley & Sons, New York, USA, third edition, 1999.

8. L. Di Gaspero and A. Schaerf. Tabu search techniques for examination timetabling. In E. Burke and W. Erben, editors, *Proc. of PATAT-2000*, volume 2079 of *LNCS*, pages 104–117. Springer-Verlag, 2001.
9. R. Dorne and J. Hao. A new genetic local search algorithm for graph coloring. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Proc. of PPSN V, 5th Int. Conference*, volume 1498 of *LNCS*, pages 745–754. Springer-Verlag, 1998.
10. Fred Glover and Manuel Laguna. *Tabu search.* Kluwer Ac. Publ., 1997.
11. H. H. Hoos and T. Stützle. *Stochastic Local Search Foundations and Applications.* Morgan Kaufmann Publishers, San Francisco, 2005.
12. D. S. Johnson. A theoretician's guide to the experimental analysis of algorithms. In M. H. Goldwasser, D. S. Johnson, and C. C. McGeoch, editors, *Data Structures, Near Neighbor Searches, and Methodology: 5th and 6th DIMACS Impl. Challenges*, pages 215–250. AMS, 2002.
13. L. Merlot, N. Boland, B. Hughes, and P. Stuckey. A hybrid algorithm for the examination timetabling problem. In E. Burke and P. De Causmaecker, editors, *Proc. of PATAT-2002*, volume 2740 of *LNCS*, pages 207–231. Springer-Verlag, 2003.
14. D. C. Montgomery. *Design and Analysis of Experiments.* John Wiley & Sons, sixth edition, 2005.
15. J. Xu, S. Chiu, and F. Glover. Fine-tuning a tabu search algorithm with statistical tests. *Intl. Tran. in OR*, 5(3):233–244, 1998.