

The Second International Timetabling Competition (ITC-2007): Curriculum-based Course Timetabling (Track 3)

— preliminary presentation —

Luca Di Gaspero and Andrea Schaerf

DIEGM, University of Udine
via delle Scienze 208, I-33100, Udine, Italy
{l.digaspero|schaerf}@uniud.it

Barry McCollum

School of Electronics, Electrical Engineering
and Computer Science, Queen's University
SARC Building, Belfast, United Kingdom
b.mccollum@qub.ac.uk

Abstract

Following the success of the First International Timetabling Competition in 2002, the timetabling research community is organising a new competition on this problem (opening July 1st). This new competition will be on three different timetabling problems, and one of the tracks concerns the course timetabling formulation that applies to Italian universities (called Curriculum-based Course Timetabling). The dataset is composed by real-world instances provided by the University of Udine. In this work, we overview the general rules of the competition and we describe in details the problem formulation and the instances proposed for this track.

Introduction

Timetabling within Universities has long been recognised as a difficult combinatorial problem of practical relevance. Whether it be timetabling exams or courses, much (human or computing) effort is spent in producing solutions that are both workable and of a high quality (see, e.g., Schaerf, 1999).

Recently, timetabling has also been a subject for a competition. In 2002, the International Metaheuristic Network organised the First International Timetabling Competition (ITC-2002), which attracted 24 feasible submissions from all over the world. Information relating to the problem definition, instances, rules, and solution evaluation of ITC-2002 is available at the webpage: <http://www.idsia.ch/Files/ttcomp2002/>.

Based on a specific problem model proposed for the competition, the formulation contained many of the characteristics found in certain Universities. Artificially generated instances were used in the competition. Thanks to the competition, this formulation has successively become a standard within the research area, and many researchers have used it within their work (Kostuch 2005; Chiarandini *et al.* 2006; Di Gaspero & Schaerf 2006; Abdullah, Burke, & McCollum 2007). ITC-2002 therefore has had a positive effect of generating common ground for cross-fertilisation of ideas within research groups in the timetabling community.

The Second International Timetabling Competition (ITC-2007), opening July 1st, follows the main ethos of the first

edition, but also aims at advancing upon it in a number of respects.

In this paper, we describe one of the competition tracks, namely the Curriculum-based Course Timetabling, which is under the responsibility of the authors. We present the general rules of ITC-2007, the problem formulation of the track, and the description of the instances. The paper closes with some discussion.

The information presented here is preliminary, as the organisation is on-going and some rules may still change up to the actual opening of the competition. Updated information will appear in the ITC-2007 web site: <http://www.cs.qub.ac.uk/itc2007/>.

Competition rules

The competition has a set of rules that the participants have to satisfy. The rules of ITC-2007 are in large part taken from those of ITC-2002, but obviously some modifications have been made based on the previous experience.

1. The competition has an opening day and a deadline (approximately 6 months later). All coding and experimenting must be finished by the deadline.
2. Participants have to implement an algorithm to tackle the problem on a single processor machine. It can be expressed using any programming language.
3. The goal is to produce feasible timetables, in which a number of hard constraints are satisfied, and to minimise the number of broken soft constraints. If feasibility cannot be reached, information outlined on the solution produced should be provided, and the number of violated hard constraints is used for evaluation.
4. The dataset will be split into three sets of instances:

Early instances: A first set of instances will appear on the web at the opening of the competition.

Late instances: A second set of instances will be published two weeks before the deadline.

Hidden instances: A third set of instances will be revealed only after the competition has closed and the participants have submitted their solvers. These will be used internally to rank the solvers submitted.

5. All solvers will be granted a fixed CPU time, which will be in the range of 10 minutes on a recent computer. Participants have to benchmark their machine with a program provided to them in order to establish how long they can run their program on their machines.
6. The algorithms should run on a single processor machine, take as input a problem file in the format described, and produce as output a timetable with a minimum number of hard and soft constraint violations in the allowed CPU time. The output timetable must adhere to the data format determined by the organisers.
7. The algorithm should not take account of additional knowledge about the instance (e.g., results from previous runs). The same version (and fixed parameters) of the algorithm must be used for all instances. That is, the algorithm should not “know” which instance it is solving: although it might analyse the instance and set parameters accordingly, it should not “recognise” the particular instance.
8. The solver can be either deterministic or stochastic. In both cases, participants must be prepared to show that those results are reproducible in the given computer time. In particular, the participants that use stochastic techniques should make their program in such a way that the *exact* run that produced each solution submitted can be repeated (by providing the random seed). They can try several runs to produce each submitted solution (each with the allowed computer time), but they must be able to repeat the specific run for any solution submitted.
9. Participants should submit for each instance (early and late ones) the best score found by their algorithm in the specified computer time, by uploading it onto the competition web site.
10. Participants should also submit a concise and clear description of their algorithm, so that in principle others can reproduce it.
11. Classification will be based on the scores provided. The actual list will be based on the *ranks* of the solvers on each single instance. Ranks will be based hierarchically on hard constraint violations and scores on the soft ones. The average of the ranks on all instances will produce the final place-list.
12. Based on the place-list a set of top solvers, the *finalists*, will be asked to provide the executable that will be run and tested by the organisers. The finalists’ solver will be rerun by the organisers on all instances (including the hidden ones).
13. In some circumstances, finalists may be required to show source code to the organisers. This is simply to check that they have stuck to the rules.
14. Finalists’ place-list will be again based on the ranks on each single instance for a set of trials on the hidden instances.

Problem formulation and Instances

The Curriculum-based timetabling problem consists of the weekly scheduling of the lectures for several university courses within a given number of rooms and time periods, where conflicts between courses are set according to the curricula published by the University and not on the basis of enrolment data.

This formulation applies to University of Udine (Italy) and to many Italian and indeed International Universities, although it is slightly simplified with respect to the real problem to maintain a certain level of generality.

The problem consists of the following entities:

Days, Timeslots, and Periods. We are given a number of *teaching days* in the week (typically 5 or 6). Each day is split in a fixed number of *timeslots*, which is equal for all days. A *period* is a pair composed of a day and a timeslot. The total number of scheduling periods is the product of the days times the day timeslots.

Courses and Teachers. Each course consists of a fixed *number of lectures* to be scheduled in distinct periods, it is attended by given *number of students*, and is taught by a *teacher*. For each course there is a minimum number of days that the lectures of the course should be spread in, moreover there are some periods in which the course cannot be scheduled.

Rooms. Each *room* has a *capacity*, expressed in terms of number of available seats. All rooms are equally suitable for all courses (if large enough).

Curricula. A *curriculum* is a group of courses such that any pair of courses in the group have students in common. Based on curricula, we have the *conflicts* between courses and other soft constraints.

The solution of the problem is an assignment of a period (day and timeslot) and a room to all lectures of each course.

The hard constraints are the following:

Lectures: All lectures of a course must be scheduled, and they must be assigned to distinct periods. A violation occurs if a lecture is not scheduled.

RoomOccupancy: Two lectures cannot take place in the same room in the same period. Two lectures in the same room at the same period represent one violation . Any extra lecture in the same period and room counts as one more violation.

Conflicts: Lectures of courses in the same curriculum *or taught by the same teacher* must be all scheduled in different periods. Two conflicting lectures in the same period represent one violation. Three conflicting lectures count as 3 violations: one for each pair.

Availabilities: If the teacher of the course is not available to teach that course at a given period, then no lectures of the course can be scheduled at that period. Each lecture in a period unavailable for that course is one violation.

The soft constraints are the following:

RoomCapacity: For each lecture, the number of students that attend the course must be less or equal than the number of seats of all the rooms that host its lectures. *Each student above the capacity counts as 1 point of penalty.*

MinimumWorkingDays: The lectures of each course must be spread into the given minimum number of days. *Each day below the minimum counts as 5 points of penalty.*

CurriculumCompactness: Lectures belonging to a curriculum should be adjacent to each other (i.e., in consecutive periods). For a given curriculum we account for a violation every time there is one lecture not adjacent to any other lecture within the same day. *Each isolated lecture in a curriculum counts as 2 points of penalty.*

RoomStability: All lectures of a course should be given in the same room. *Each distinct room used for the lectures of a course, but the first, counts as 1 point of penalty.*

Instances and File Formats

There will be 21 instances available: 7 for each set (early, late, and hidden). All instances are real data and come from the University of Udine. For all instances there exists at least one feasible solution (no hard constraint violations), but at present it is not known which is the optimal value for the soft constraints.

In order to model cases in which the number of timeslots is not the same for all days (e.g. Saturday afternoon free), there might be periods unavailable for all courses. Furthermore, for all instances there cannot be two curricula composed by exactly the same courses.

Each instance comes in a single file, containing a file header and four sections: courses, rooms, curricula, and constraints. The header provides all scalar values and each section provides the arrays for that specific aspect of the problem. The exact format is described in the web site.

The output also must be provided in a single file such that each line represents the assignment of the room and the timeslot to one lecture (lines can be in any order).

Solution validation

We provide the C++ source code of a solution *validator*, so that the participants can compile it themselves at their machine and also inspect the code. In case it is necessary, executables for various platforms could be provided on request.

The validator takes two command-line arguments: the input file and the output file and it produces on the standard output the evaluation of the solution along with the detailed description of all violations (hard and soft). The very same validator will be used by the organizers for the validation of the scores of the finalists during their final phase.

If the output file is not formatted correctly, the validator produces an error message on the standard error and aborts. Conversely, the input is assumed always correct. A validator for the input file, in case the participants want to create new instances for their internal use, is also available and can be provided upon request.

Discussion

We now discuss the main characteristics of the competition. We start highlighting the novelties w.r.t. the previous ITC-2002; then, we discuss the key issues underlying selected the rules; and, we conclude discussing the motivations for the definition of the specific formulation of the problem for our track.

Differences with ITC-2002

The first innovation consists in being subdivided in three tracks so as to better cover the main formulations of the field of educational timetabling problems. Specifically, the tracks will be on: *Examination timetabling*, *Post Enrolment Course Timetabling* (the evolution of the ITC-2002 formulation), and *Curriculum-based Course Timetabling*.

The second innovation aims at bridging the gap between research and practice: the competition introduces a significant degree of complexity in all tracks so that the new formulations employed are closer (in more aspects, although not all) to those of 'real world' problems (McCollum 2006) and data is coming from the real world.

The third main innovation regards feasibility of solutions. In ITC-2002 only feasible solutions were accepted and it was purposely rather simple to produce a feasible one for all instances. This time, participants that reach only infeasible solutions for some instances can submit their solution, although all instances are guaranteed to have at least a feasible one.

In order to compare different solvers in cases of unfeasible solutions for some instances, we use an evaluation based on ranking of solutions on each instance, rather than on the actual scores (which might be incomparable). Due to this scoring based on rankings, an infeasible solution on one instance does not necessarily prejudice the overall performance. In addition, instances for the competition can be selected from a larger set of interesting cases, without the limitation of 'easy feasibility'.

Finally, in ITC-2002 the ranking was fully based on the solution provided by the participants. In case of stochastic solvers, this CPU time was to grant the maximum time *for each single trial*. Therefore, the participants could take advantage of what we call the *Mongolian Horde* approach ((Schaerf & Di Gaspero 2006)): "Run as many trials as you can and report only the best of all of them". In ITC-2007, the re-running of finalist solvers on organisers machine (with new seeds) and the use of hidden instances should be able to provide against this approach. Moreover we believe that, in the case of stochastic algorithms, this fosters the design of *robust* solvers.

Although they will not be used for the place-list, the organizers plan to use principled statistical tools to analyse in more details the performances of the solvers, especially for the stochastic ones.

It is worth mentioning that in order to provide against the excessive use of the Mongolian Horde approach, the competition organisers of ITC-2002 tested the best few algorithms also on unseen instances, and indeed the results were found to be broadly in-line with the known instances.

Comments about the running time rules

One of the main issues about the competitions in general is the running time limits. The reason why it has been decided to have a fixed running time is mainly to remove one degree of variability from the scoring system. We hope that future competitions will take into account in some principled way the trade-off between solution quality and running times.

For the selection of the fixed amount of running time, the question is which is a realistically feasible running time for the actual timetabling. Given that the timetabling is performed usually a few times a year, one might think that a much longer running time would also be reasonable.

In practical cases however, as many researchers have pointed out, the solution of a real case is an interactive process, during which it is necessary to solve a large number of instances. In fact, constraints and objectives are usually manually adjusted between runs of a working session for one single case (for various reasons: what-if scenarios, last minute changes, etc.). As a rule of thumb, a running time longer than a few minutes makes the process very tiresome and difficult for the human operator.

Comments about the problem formulation

The actual formulation used at the University of Udine, with respect to the one used for ITC-2007, has the following extra features:

- A cost component dealing with the lunch break for students: at least one free slot among those around the lunch time.
- The curriculum compactness feature is more complex, and specific patterns are more penalized than others.
- There is a maximum daily student load for each curriculum.
- Some specific lectures must be (must not be) in consecutive periods.
- Rooms might not be available in certain periods, and they must be not suitable for specific lectures.
- If a room is too big for a class, this is also penalized (this is not only for the unpleasant feeling that an empty room provoke, but also to same big rooms for unforeseen activities).
- Weight assigned to soft violations are not complex, and they depend also on the number of students in the curriculum.
- Teacher preferences on periods and rooms are only included as soft constraints.

The only reason for which we have decided to remove all the above features is to maintain a certain degree of generality, so as to do not inflict to the participant the burden to understand all the details of the formulation. The selection of the features to include in the formulation has been based on the aim to balance different types of constraint. Needless to say, if in the future this formulation will prove to be too simple, some features could be reintroduced for future research.

References

- Abdullah, S.; Burke, E. K.; and McCollum, B. 2007. Using a randomised iterative improvement algorithm with composite neighbourhood structures for the university course timetabling problem. In Doerner, K.; Gendreau, M.; Greistorfer, P.; Gutjahr, W.; Hartl, R.; and Reimann, M., eds., *Metaheuristics - Progress in Complex Systems Optimization*, Springer Operations Research / Computer Science Interfaces Book. To appear.
- Chiarandini, M.; Birattari, M.; Socha, K.; and Rossi-Doria, O. 2006. An effective hybrid approach for the university course timetabling problem. *Journal of Scheduling* 9(5):403–432.
- Di Gaspero, L., and Schaerf, A. 2006. Neighborhood portfolio approach for local search applied to timetabling problems. *Journal of Mathematical Modeling and Algorithms* 5(1):65–89. DOI: 10.1007/s10852-005-9032-z.
- Kostuch, P. 2005. The university course timetabling problem with a three-phase approach. In Burke, E., and Trick, M., eds., *Proc. of the 5th Int. Conf. on the Practice and Theory of Automated Timetabling (PATAT-2004)*, selected papers, volume 3616 of *Lecture Notes in Computer Science*, 109–125. Berlin-Heidelberg: Springer-Verlag.
- McCollum, B. 2006. University timetabling: Bridging the gap between research and practice (invited paper). In Burke, E., and Rudová, H., eds., *Proc. of the 6th Int. Conf. on the Practice and Theory of Automated Timetabling (PATAT-2006)*, 15–35.
- Schaerf, A., and Di Gaspero, L. 2006. Measurability and reproducibility in timetabling research: State-of-the-art and discussion (invited paper). In Burke, E., and Rudová, H., eds., *Proc. of the 6th Int. Conf. on the Practice and Theory of Automated Timetabling (PATAT-2006)*, 53–62.
- Schaerf, A. 1999. A survey of automated timetabling. *Artificial Intelligence Review* 13(2):87–127.