The University of Nottingham School of Computer Science

Benchmark Data Sets in Exam Timetabling

R. Qu, E. K. Burke, B. McCollum, L.T.G. Merlot, and S.Y. Lee. A Survey of Search Methodologies and Automated System Development for Examination Timetabling. Journal of Scheduling, 12(1): 55-89, 2009. online publication Oct 2008. doi: 10.1007/s10951-008-0077-5. .pdf

University of Toronto Benchmark Data

Carter, Laporte and Lee in 1996 introduced a set of 13 real-world exam timetabling problems from 3 Canadian highs schools, 5 Canadian, 1 American, 1 UK and 1 mid-east universities.

During the years, however, two versions of the 5 problem instances in the data (presented in the table below) were derived. Also within the 5 problem instances 3 contain duplicated exams for some students.

- The original data: ftp://ftp.mie.utoronto.ca/pub/carter/testprob (version I).
- The version II of the 5 instances: here (version II).
- The corrected version II of the 3 instances: here (version IIc).

Problem Description

A Conflict Matrix C was defined where each element $C_{ij} = 1$ if exam i conflict with exam j (have common students), or $C_{ij} = 0$ otherwise. The *Conflict Density* represents the ratio between the number of elements of value "1" to the total number of elements in the conflict matrix.

Two variants of objectives were defined:

- to minimise the number of timeslots needed for the problem (graph coloring); and
- to minimise the sum of approximate costs per student. i.e. $åw_i/S$, i = 0, 1, 2, 3, 4.

Evaluation Function executable program

For students sitting two exams *s* apart, the approximate costs are w_s i.e. $w_0=16$, $w_1=8$, $w_2=4$, $w_3=2$ and $w_4=1$. The aim is to space out the conflicting exams within a limited number of timeslots. There is no restriction on the capacity of the rooms in the original data.

An executable program which calculates the penalty (using the above eveluation function) and the source code are available for download.

Variants in the Literature

In "x/y/z" in the table, x: characteristics of the original; y: characteristics of the derived data; z: characteristics of the corrected data. *: total number of enrolments conflict in data files. The single value in the table indicates the same value in different versions.

We are collecting the solutions from the researchers who published the best results in the literature for this dataset. We will make the solutions available as soon as we can.

Problems	Exams I/II/IIc	Students I/II/IIc	Enrolments I/II/IIc	Density	Timeslots	BestSolution/penalty
CAR91	682	16925	56877/56242	0.13	35	car91.sol / 4.42
CAR92	543	18419	55522/55189	0.14	32	car92.sol / 3.74
EAR83	190/189/189	1125/1108/1108	8109/8092/8057	0.27/*/0.27	24	ear83.sol /
HEC92	81/80	2823	10632/10625	0.42/0.42	18	hec92.sol /
KFU93	461	5349	25113	0.06	20	kfu93.sol / 12.96
LSE91	381	2726	10918	0.06	18	lse91.sol /
PUR93	2419	30029	120681*	0.03	42	pur93.sol / n/a
RYE92	482	11483	45051	0.07	23	rye92.sol /
STA83	139/138/138	611/549/549	5751/5689/5417	0.14/*/0.19	13/13/35	sta83.sol /
TRE92	261	4360	14901	0.18	23	tre92.sol / 7.75
UTA92	622/638	21266/21329	58979/59144	0.13/0.12	35	uta92.sol /
UTE92	184	2750	11793	0.08	10	ute92.sol /
YOR83	181/180/180	941/919/919	6034/6012/6002	0.29/*/0.3	21	yor83.sol / 34.84

University of Nottingham Benchmark Data

Burke, Newall and Weare in 1996 also introduced the exam timetabling data at University of Nottingham. In the problem, the total room capacity 1550 cannot be exceeded by the total number of students in the exams assigned for each timeslot. The objective is to minimise the students sitting into two consecutive exams on the same day. Later in 1998 the objective function was extended to also consider two consecutive exams overnight.

Problem	Exams	Students	Enrolments	Density	Timeslots	Capacity	Objective
a	800	7896	34265	0.03	23	1550	to minimise consecutive exams on the same day.
b	800	7896	34265	0.03	23	1550	to minimise consecutive exams on the same day and overnight.

University of Melbourne Benchmark Data

Merlot, Boland, Hughes and Stuckey introduced exam timetabling data sets from the University of Melbourne at the PATAT conference in 2002. Two data sets were introduced (with 521 exams, 28 sessions, 20656 students and 62248 enrolments and 562 exams, 31 sessions, 19816 students and 60637 enrolments). For these data sets, there were two exam sessions on each day for each of the five workdays, and the capacity for each session varied. The objective for these data sets was to minimize the number of occasions of students having two exams consecutively either on the same day or overnight. The availability of sessions for some of the exams was restricted. In one problem instance this prevented all feasible solutions so an alternate data set was created which allowed feasible solutions. These data sets can be downloaded from http://www.or.ms.unimelb.edu.au/timetabling. This website also presents results for these data sets on other problems, such as those presented for the University of Toronto problems.

Problem	Exams	Students	Enrolments	Timeslots

| I | 521 | 20656 | 62248 | 28 |
| II | 526 | 19816 | 60637 | 31 |

A Random Exam Timetabling Problem Generator

An exam timetabling instance generator was developed and a benchmark exam timetabling problem data set was generated that consists of 18 different problem instances. The 18 problems consist of 9 small (< 100 exams) and 9 large problems (>= 500 exams). The problems are generated with conflict density values starting from 6% to 47% using 5% intervals. The number of students and their enrolments are variable according to the problem size and conflict density. The problems have similar constraints and use the same objective function as the Toronto benchmark data set stated in the previous section.

The problem generator and the 18 new random problems are available for public to encourage comparisons and feedback.

Last updated date 1/June/2008