Tutorial NSPLib benchmark dataset

In this manuscript, the characteristics of the benchmark dataset of nurse scheduling problem instances and different sets of case-specific constraints proposed in the paper "NSPLib – A Nurse Scheduling Problem Library: A tool to evaluate (meta-)heuristic procedures" (by Mario Vanhoucke and Broos Maenhout) are explained. This paper can be downloaded from our website.

NSPLib: Benchmark dataset of problem instances

Input File Characteristics

Diverse set

We have 4 different data sets generated under a controlled design, as follows:

Set 1: N25 set

	Size		Preferences		Coverage
N	25	**NPD**	0.25, 0.50 or 0.75	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.25, 0.50 or 0.75	**SCD**	0.25, 0.50 or 0.75
D	7	**DPD**	0.25, 0.50 or 0.75	**DCD**	0.25, 0.50 or 0.75

Using 10 instances for each setting, we have generated 3^6 x 10 = 7290 NSP instances.

Set 2: N50 set

	Size		Preferences		Coverage
N	50	**NPD**	0.25, 0.50 or 0.75	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.25, 0.50 or 0.75	**SCD**	0.25, 0.50 or 0.75
D	7	**DPD**	0.25, 0.50 or 0.75	**DCD**	0.25, 0.50 or 0.75

Using 10 instances for each setting, we have generated 3^6 x 10 = 7290 NSP instances.

Set 3: N75 set

	Size		Preferences		Coverage
N	75	**NPD**	0.25, 0.50 or 0.75	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.25, 0.50 or 0.75	**SCD**	0.25, 0.50 or 0.75
D	7	**DPD**	0.25, 0.50 or 0.75	**DCD**	0.25, 0.50 or 0.75

Using 10 instances for each setting, we have generated 3^6 x 10 = 7290 NSP instances.

Set 4: N100 set

Size		Preferences		Coverage	
N	100	**NPD**	0.25, 0.50 or 0.75	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.25, 0.50 or 0.75	**SCD**	0.25, 0.50 or 0.75
D	7	**DPD**	0.25, 0.50 or 0.75	**DCD**	0.25, 0.50 or 0.75

Using 10 instances for each setting, we have generated 3^6 x 10 = 7290 NSP instances.

Realistic set

We have 2 different data sets generated under a controlled design, as follows:

Set 5: N30 set

Size		Preferences		Coverage	
N	30	**NPD**	0.3 or 0.7	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.3 or 0.7	**SCD**	0.3 or 0.7
D	28	**DPD**	0.3 or 0.7	**DCD**	0..3 or 0.7

Using 10 instances for each setting, we have generated 2^5 x 3 x 10 = 960 NSP instances.

Set 6: N60 set

Size		Preferences		Coverage	
N	60	**NPD**	0.3 or 0.7	**TCC**	0.20, 0.35 or 0.50
S	4 (including the free shift)	**SPD**	0.3 or 0.7	**SCD**	0.3 or 0.7
D	28	**DPD**	0.3 or 0.7	**DCD**	0..3 or 0.7

Using 10 instances for each setting, we have generated 2^5 x 3 x 10 = 960 NSP instances.

In the files, no differentiation is made towards nurses' skill categories or towards nurses' contracts or working regulations (e.g. part-time nurses, full-time nurses) to maintain the generality of the proposed dataset. These features can be incorporated in the dataset as follows:

Different contracts or working regulations

For the incorporation of different contract, the user needs to specify:

- The types of contracts: E.g. three types of contracts: full-time (100%), 80% part-time and 50% part-time.
- The distribution of contract among nurses: E.g. contracts are distributed in a uniform way, i.e. 33% of the nurses have a fulltime contract, 33% of the nurses a 80% contract and the

remaining pool a 50% contract.the types of contract (full-time: 100%; part-time: 80%, 60% 50%, etc…)

Different skill categories

For the incorporation of different skill categories, the user needs to specify:

- The number of skills and the corresponding substitution possibilities: E.g. three skill types where nurses of a higher skill (grade) can substitute lower skilled nurses, but not vice versa
- The day/shift coverage requirements for each skill: E.g. the original day/shift coverage requirements can be re-distributed among skill types in proportion with the number of nurses that are able to fulfil the required coverage for that shift/day.
- The distribution of skills among nurses: E.g. equally distributed among the nurses.

Input File Format

The nurse scheduling problem instances for each dataset are numbered and all files have the same extension *'.nsp'*.

The input data structure is the same for all files having the following layout:

Problem size

N	D	S
4	4	4

Coverage matrix

	Shift 1	Shift 2	Shift 3	Shift 4
Day 1	**2**	**1**	**1**	**0**
Day 2	**1**	**1**	**1**	**0**
Day 3	**2**	**1**	**1**	**0**
Day 3	**2**	**1**	**1**	**0**

Preference matrix

	Day 1				Day 2				Day 3				Day 4			
	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
Nurse 1	**2**	**4**	**4**	**4**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**2**	**4**	**2**	**3**	**1**
Nurse 2	**1**	**2**	**2**	**2**	**1**	**2**	**3**	**3**	**4**	**2**	**4**	**4**	**4**	**2**	**4**	**4**
Nurse 3	**3**	**4**	**4**	**4**	**1**	**4**	**4**	**4**	**4**	**2**	**4**	**4**	**4**	**2**	**4**	**4**
Nurse 4	**4**	**2**	**4**	**4**	**4**	**2**	**4**	**4**	**1**	**2**	**3**	**4**	**3**	**3**	**3**	**3**

Sets of case-specific constraint files

Input File Characteristics

In order to solve nurse scheduling problem instances, many constraints can be incorporated in the problem description (Burke et al., 2004). We have extended the NSPLib with a number of case-constraint files with case-specific constraints identified by Cheang et al. (2003) as appearing most frequently in literature. The daily coverage requirements are an inherent characteristic to all personnel rostering problems, and are incorporated in the coverage complexity indicators. The minimal free time between working shifts and the fact that nurses can take only one assignment per day occur almost always in literature and are supposed to characterize all nurse scheduling problems. Some constraints typify nurse scheduling practices but are not always incorporated. Not only the presence of these constraints are an indicator of the restrictiveness of nurse scheduling instances, but also the parameter setting for each constraint contributes to the restrictiveness of the problem. Both, the 'case-specific' constraints and the parameters which characterize them, are presented in the following table along with the case number they can be identified with:

	Number of assignements	Number of assignements per shift	Consecutive working shifts	Consecutive same working shifts
Diverse Set: N25, N50, N75 and N100 instances				
Case 1	[5,5]	x	x	x
Case 2	[4,6]	x	x	x
Case 3	[5,5]	[1,3], [1,3], [1,2]	x	x
Case 4	[4,5]	[0,5], [0,5], [0,4]	x	x
Case 5	[5,5]	x	[2,5]	x
Case 6	[4,6]	x	[1,5]	x
Case 7	[5,5]	[0,5], [0,5], [0,3]	[2,5]	[2,3], [2,3], [2,3]
Case 8	[2,6]	[0,6], [0,6], [0,3]	[2,4]	[1,4], [1,4], [2,4]
Realistic Set: N30 and N60 instances				
Case 9	[20,20]	x	[1,7]	[1,7], [1,7], [1,7]
Case 10	[16,24]	x	[1,7]	[1,7], [1,7], [1,7]
Case 11	[20,20]	[4,12], [4,12], [4,8]	[1,7]	[1,7], [1,7], [1,7]
Case 12	[16,20]	[0,20], [0,20], [0,16]	[1,7]	[1,7], [1,7], [1,7]
Case 13	[20,20]	x	[2,5]	[1,7], [1,7], [1,7]
Case 14	[16,24]	x	[1,5]	[1,7], [1,7], [1,7]
Case 15	[20,20]	[0,20], [0,20], [0,12]	[2,5]	[2,3], [2,3], [2,3]
Case 16	[16,24]	[0,24], [0,24], [0,12]	[2,4]	[1,4], [1,4], [2,4]

Table. The case-specific constraints per CC with [min, max] settings

In the proposed case constraint files we assume that the fourth shift is the free, non-working shift. Since no restrictions are imposed on this fourth shift, the respective minimum and maximum requirements for this shift are not mentioned in the second and fourth column of the table.

The cases from 1 to 8 concern the diverse set of files where the planning horizon is limited to seven days. The cases from 9 to 16 refer to the realistic set where nurses need to be schedule over a period of four weeks.

Input File Format

The nurse scheduling problem instances for each dataset are numbered from 1 to 16 and all files have the same extension '*.cc*'.

The input data structure is the same for all files having the following layout:

	7	4	**→ Problem Size (D and S)**
	5	5	**→ Number of Assignments (min and max)**
	1	7	**→ Consecutive Working Shifts (min and max)**

	Consecutive same working shifts		**Number of assignements per shift**	
	Min	**Max**	**Min**	**Max**
Shift 1	1	7	0	7
Shift 2	1	7	0	7
Shift 3	1	7	0	7
Shift 4	1	7	0	7

References

Burke E.K., De Causmacker P., Vanden Berghe G., and Van Landeghem H., 2004, The state of the art of Nurse Rostering, Journal of Scheduling, 7, 441-499

Cheang, B., Li, H., Lim, A., and Rodrigues, B., 2003, "Nurse rostering problems – a bibliographic survey", European Journal of Operational Research, 151, 447-460

Vanhoucke, M., and Maenhout, B., 2005, NSPLib – A Nurse Scheduling Problem Library: A tool to evaluate (meta-)heuristic procedures, submitted to the ORAHS Proceedings 2005