20 - 7.

Referee's report on the PhD thesis

Advanced Methods and Models for Employee Timetabling Problems

by Z. Bäumelt

Overall recommendation: Accept.

General assessment. The thesis deals with two standard employee scheduling problems, which are well studied in Operations Research. Both of them are forms of assignment problems, which can be easily modeled by (M)ILPs. However, well-known NP-hardness results imply that instances important for practice cannot be solved exactly due to their sizes. This justifies why it is necessary to study suboptimal approaches working in tractable computation time. The main work of the thesis is in formulation of heuristics and a detailed study of implementation issues taking the advantage of parallel architectures.

As far as I am aware, the thesis presents currently the best computational results. This is one of the reasons why I recommend its acceptance. The positive opinion is also supported by the fact that a part of the thesis has been published in *Comput Oper Res* (IF 1.861, rank 29/102 in Computer Science, Interdisciplinary Applications; 9/43 Engineering, Industrial; 19/81 Operations Research & Management Science).

The results are presented in Chapter 3 *Employee Timetabling with a High Diversity of Shifts* and Chapter 4 *GPU-based Parallel Algorithm for Nurse Rerostering.*

Chapter 3. The problem is explained with a sufficient level of rigor. The survey of existing literature is very good. The basic idea is clever: to reduce the original "huge" problem to a problem with a smaller state space. I have just a pair of remarks and questions.

- Pg 35. L_k is defined as "average shift length". Here it is not clear why taking *averages* is correct (or reasonable). Is it true that averaging indeed does not matter, even if the shift lengths have a high variance?

- Table 3.1 (pg. 38) is misleading — the search is done over $|E| \cdot |D| \cdot |K|$ 0-1 variables; the continuous variables W_e and $RS_{kd}^{(\mathcal{K})}$ are superfluous since their definitions can be plugged into the objective function.

- The description of Stages $1-3$ is, in my opinion, unbalanced. On the one hand, much effort is devoted to the description of Stage 1. We can read a lot e.g. about the particular encoding of data into genes. On the other hand, Stage 3 is commented on p. 45 only by a single sentence "... the result of the second stage is further optimized in the third stage which can be based on common techniques, e.g. Tabu Search...".

  Here I do not understand why it is so essential to describe Stage 1 in such a detail, compared to the single-sentence description of Stage 3. The point is that I cannot see any particular idea or trick in the author's version of the evolutionary algorithm of Section 3.4. For example, I understand that he had to choose *an* encoding of data into genes; one always has to when using EA. But is this encoding in a sense tricky, nontrivial, or does it bring some particular advantages?

- What is indeed interesting is the evaluation methodology of Section 3.6.4. This methodology can be considered as one of original results of the thesis.

- The computational results are convincing. However, I would also acknowledge some discussion on possible weaknesses. In practice one could meet also different instances from those evaluated by the author; so I would be interested in a discussion which classes of instances are suitable for the presented approach and which are *not*.

Chapter 4. The presentation of the rerostering problem is also good. I would prefer writing a formal ILP model (similarly as in Chapter 3); usually a formal model tells a reader more than a pair of pages of verbal description.

I have encountered some recent papers on GPU computation in OR (e.g. for the Traveling Salesman). The idea is still challenging and the author shows how nontrivial work is required when the basic idea is to be utilized for a particular practical problem.

- Chapter 4 can be seen as a comprehensive engineering work, starting from a formulation of a practical OR problem, followed by a design of a sequential heuristic, which is then parallelized. Finally, many low-level implementation details must be worked out to achieve a truly efficient algorithm taking advantages of the particular parallel architecture (and respecting its hardware limitations).
- This work has been submitted to *Eur J Oper Res.* This is a journal focused on methodology of OR. But here, it seems to me that the main contribution is in the GPU implementation details. So wouldn't it be more appropriate to choose an engineering journal, such as an IEEE journal?
- I am not an expert in parallel programming (and not in GPU implementations). However, it is clear that the achieved results are convincing: as shown in Section 4.7, the author has achieved a significant computational speed-up preserving the quality of solutions found.
- My basic question is similar to that of Section 3: which classes of instances of the rerostering problem are suitable for the parallelized algorithm of Section 4, and which are not? Which instances are hard-to-parallelize?

Summary. The thesis presents a good engineering work in two related employee timetabling problems. As far as I am aware, the author has achieved the best currently reported computation times. The work presents and elaborates on a clever combination of ideas, which results in a design of highly efficient algorithms. One of the main ideas is a reduction of a complex problem into a smaller state space; the other idea is a decomposition suitable for parallel processing. Many implementation details for the used GPU architecture must have been worked out.

Moreover, the work of Chapter 3 has been published in a well-ranked OR journal and the work of Chapter 4 has been submitted.

In my opinion, the work has a high scientific quality and I support the promotion of Mr Bäumelt to the PhD degree.

doc. RNDr. Ing. Michal Černý, Ph.D.
Department of Econometrics & DYME Research Center
University of Economics in Prague
June 30, 2015