



INSTITUTE OF COMPUTER SCIENCE

Academy of Sciences of the Czech Republic

Martin Holeňa

Pod Vodárenskou věží 2, 182 07 Praha 8, phone: +420 266052921, fax: +420 286585789, e-mail: martin@cs.cas.cz
web: www.cs.cas.cz/~martin

Review of the PhD thesis

Integer Optimization Heuristics

by Matej Mojzeš

The reviewed thesis investigates heuristic methods for integer optimization, paying particular attention to measures of their time complexity and to specific suitability of particular heuristics to specific kinds of problems. Three alternative heuristics have been developed and a number of experiments have been reported, both on benchmark problems and in real-world case studies.

The thesis is, apart from the Introduction (Chapter 1), statement of its objectives in Chapter 3 and Conclusion (Chapter 8), structured into 5 chapters (Chapters 2 and 4–7). Chapter 2 recalls important traditional integer optimization tasks as well as approaches to dealing with them, and it surveys relevant performance measures.

In Chapter 4, a general view of the search process in integer optimization is first thoroughly formalized. In the context of that formalization, time complexity measures are then studied, primarily measures based on quantiles and on mean. In my opinion, this chapter is the main methodological contribution of the thesis.

Also Chapter 5 starts with introducing a formal framework, this time a general framework for integer optimization. Within that framework, three alternatives to existing integer optimization methods are then proposed. I have a question concerning the second of them, competitive integer mutation search: What is the justification for artificially increasing the number of 1-coordinate mutations over the considered distribution $\text{Bi}(d, p_{\text{mut}})$ through performing a 1-coordinate mutation also in situations when the sampled number of mutated coordinates was 0?

In Chapter 6, a number of particular integer optimization tasks is discussed, most attention being paid to submodel selection in regression.

Chapter 7, which is the most extensive chapter of the thesis (comprising 36 from the overall 100 thesis pages) presents results of experiments in 10 different case studies. I appreciate the large number of various comparisons performed in those experiments. On the other hand, I am disappointed that no significance results are reported for those comparison, although enough data for significance testing were available at least for the experiments comparing

different values of T_0 in the study of time complexity measures, experiment comparing the 2 instances A a B in the study of optimal restarting period, experiments comparing different values of the mutation temperature in the study tuning that temperature, experiments comparing different heuristics as well as experiments comparing different values of the MFT temperature in the RMFIF efficiency study, experiments comparing different heuristics in the Poisson regression study, and for polite generators also some of the experiments comparing different heuristics in the politeness study. Can results of significance testing be presented at least during the thesis defence?

As far as the text of the thesis is concerned, I appreciate its structure and its language. On the other hand, it seems that when deciding which text to include into the thesis and which not, and when choosing the employed notation, the author was sometimes rather negligent about avoiding reader's confusion. In particular:

- The reader will likely be confused when encountering on page 20 two and half paragraphs read already three pages earlier.
- Equation (6.1) is confusing without recalling what GLM are, and explaining why they are considered as random variables on \mathbb{R}^{n+1} .
- Equation (6.9) is confusing without recalling that 1.96 is approximately the 97.5% quantil of $N(0,1)$.
- It is not a good idea to use the same symbol in two subsequent chapters for two different concepts (a neighbourhood in Chapter 5 and a solution set in Chapter 6), even if it is explicitly stated.
- The symbol K occurring in the equation (5.48) has been nowhere explained.
- In Section 6.1, it is confusing that $y^* \in \mathbb{R}^n$. From the context, I would expect $y^* \in \mathbb{R}^m$.

Apart from the two questions concerning Chapters 5 and 7, mentioned above, I have another, third question to be answered during the thesis defence: In situations as in your thesis, that there is a set of algorithms and/or algorithm settings, and a set of problems and/or data such that for each of them, a specific algorithm or setting is the most suitable, the most general approach to constructing a mapping of the problems/data set to the algorithms/setting set is metalearning. Which metalearning methods have you used in your research and how?

In spite of the few above mentioned problems that I have with the thesis, I definitely recognize that it clearly shows the ability of Matej Mojzeš to perform creative, independent and systematic research, fulfilling high standards required for the PhD degree at the CTU Faculty of Nuclear and Physical Engineering.

Prague, December 30, 2017

Martin Holeňa