

Algorithms for processing of large data sets using distributed architectures and load balancing

Ing. Ondřej Šubrt

Description

The doctoral thesis deals with the field of large data sets processing and optimisation using different approaches (from linear programming to reinforcement learning).

The work is divided into nine parts including introduction and conclusion. After a very brief introduction, (two pages) chapters 2, 3, 4 and 5 follows, where author introduces a background of his work. Here, it is necessary to state, that student cooperates with CERN, one of the most prestigious workplaces in the world. Therefore, it was (i) necessary to explain the experimental setup in four chapters and (ii) interesting to read about the experiment background. I must say that it was very difficult to read. Maybe more illustrations would be helpful.

The body of author's work is described in chapters 6 and 7. The motivation of the dissertation was to improve the workload across multiple computing resources, maximize the throughput and minimize the response time. This issue is very crucial for projects such as COMPASS Experiment because it uses about 300 000 different detector channels, where the trigger rate can rise up to 50kHz with 36kB average event size.

Author formulates the optimisation problem at the beginning of Chapter 6, subsequently, several methods for solving this problem are predetermined:

- Dynamic programming
- Greedy algorithm
- Integer programming
- Genetic algorithm
- Reinforcement learning

Chapter 7 introduces the numerical results of the optimisation. All above mentioned methods were used in several "case studies". I have to state that all mentioned methods give satisfactory results and both methods and results are clearly described. Moreover, attachments with complete optimisation results are also part of the work.

Formal part

The formal and graphical part of the work is very good having a precise logical structure. It might be useful to include also the list of figures and the list of tables. The language level is very good with several mistakes only.

Actuality of work

The chosen topic of the dissertation is highly topical from a research and practical point of view. The list of used literature includes a number of publications from recent years, which also testifies to the topicality of the topic of the work.

Comments and questions

The study is carefully written describing a very interesting research topic. I have following comments and questions:

- Figure 1.2 is confusing. Can the student describe this picture in more details?
- It is not clear from the work whether it was necessary to make any changes in the programs for communication, or whether this part was taken over and the student optimized over it.
- Could the student describe the strengths and weaknesses of each approach he used? Not only from the point of view of self-optimization, but also implementation, programming or software support

Conclusion

The dissertation thesis is carefully written describing selected research topics and results of own studies. It is possible to summarize that the work is both from the research and formal point of view excellent and it forms a very good basis of further work.

Owing to facts presented above it is possible to state that the dissertation thesis of Ing. Ondřej Šubrt presents important topics and it is possible to **recommend defense of the work**.

22. 4. 2021

Doc. Ing. Jan Mareš, Ph.D.

Department of Computing and Control Engineering
University of Chemistry and Technology
Technická 5, 166 28 Praha 6 - Dejvice