

Report 10-Towards a discipline of experimental algorithmics

Xinyu Chen

April 19, 2017

1 Summary

This paper emphasized the implementation as a principal standards of value and call this approach *Experimental Algorithmics*. The authors gave many examples to argue there have been gap between theory and practice for a long time. These examples included both NP-hard problems and tractable problems. They argued the big O and big Theta notation and worst-case behavior didn't reflect real running time. Some complicated algorithms are only studied by paper-and-pencil and not really be implemented. It is not good for people to get insight from those unimplemented algorithms.

The paper listed directions that worth such efforts. These areas are testing and improving NP-hard heuristics, comparing algorithms and data structures for tractable problems, supporting and refining conjectures, developing tools and so on. They discussed about running time and memory references are good measurements. And beyond the obvious measures, we should think about other measures. They suggested to present and analyze the results carefully and not to discard unexplained results.

The paper used a case study on implementing and testing several minimum spanning tree algorithm to show how to do experimental algorithmics researches. They suggested to use different language, different compiler, test large size of generated data, as well as real world data, on different machines, to compare these algorithms. They normalized result to eliminate cache effect. This is also beneficial for present and analyze the results.

2 Key Takeaway

The authors mentioned a list of pitfalls in implement and testing heuristics and algorithms. The cache effect is very important and we should pay attention to that in our experiments. They normalized their results by comparing with some linear procedures. This is a very good approach that we can learn.

3 Discussions

The paper is interesting. The following questions are important.

- *Benchmark Dataset.* The authors mentioned some research areas have benchmark datasets. But computational biology and chemistry don't have. Do these two areas have some progress now?
- *Hard problems.* The authors listed some NP-hard problems and other tractable problems. What about machine learning problems. How can we use the principals they proposed into this area?
- *Present and analyze results.* In resent use of deep learning and neural networks, it's hard to interpret and visualize the result and process. Is there any method that we can use to help us visualize the learning process better?