You Can Check Others’ Work More Quickly Than Doing It Yourself

Graham Cormode *1, Christopher Hickey *2

*University of Warwick
Coventry, UK
1 g.cormode@warwick.ac.uk
2 c.hickey@warwick.ac.uk

***Abstract*—Much of computer science involves problems where it is considered to be easier to check that an answer is correct than to find a correct answer (the complexity class NP). In this talk, we outline results that apply this notion to checking outsourced computations for data analytics.**

I. Background

The increasing popularity of data analtics means dramatic increases in scale: much larger models with many parameters to set optimally, and much larger training data sets to determine these parameters. This presents a challenge to data owners who do not have a convenient data centre at their disposal: the size of data and computational cost in order to extract accurate models begins to look prohibitive. So we look to outsourced computation where computation can be ‘rented’ on demand to run analytics workloads.

One doubt remains. If we send our data off to the cloud, what guarantee do we receive that the processing has been done to our satisfaction? The provider has an economic incentive to cut corners: to perform the computation on only a sample of provided data, or to terminate an iterative parameter search before convergence has occurred, for example. Such short cuts yield plausible but suboptimal answers. So how could we be assured that the correct model has been found, without repeating the computation ourself or paying independent providers to repeat the work, substantially driving up the costs?

II. Results

Ideas of “interactive proofs” originally developed as a thought-experiment in computational complexity have led to methods that can check outsourced computations very effectively [1], [2], [3]. Applied here, they require the cloud to give a “proof” that can be checked easily by the data owner, shown schematically in Figure 1 [4], [5]. This lightning talk will give some examples of data analytics problems for which there are ultra-efficient proof protocols. The overhead for the cloud provider is minimal – often, the required information is a relatively low cost function of the input data or natural by-products of the target computation. These do not restrict the cloud to use any particular implementation or algorithm; just that they demonstrate that the output meets certain necessary properties. The key part of these protocols is that the information required is very easy for the original data owner to check, based on appropriately defined fingerprints computed flexibly from the input. If the data owner’s checks pass, then they are assured that the computation has been performed by the cloud satisfactorily, with a very high degree of certainty. One can then think of these protocols as providing effective “checksums for computation”.

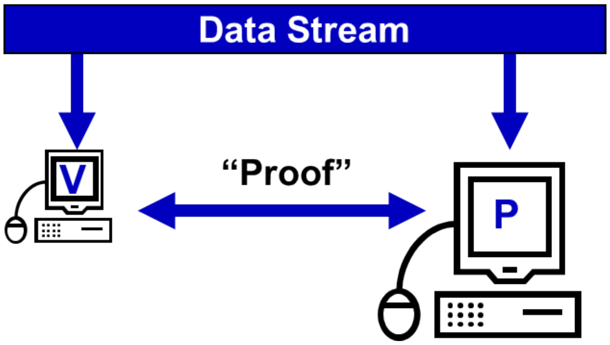

Fig. 1. Annotated streaming model: Prover (P) and Verifier (V) observe the same input stream, and interact to determine if V will accept P’s claims.

Acknowledgments

This talk is based on a sequence of works, the most recent of which is [6].

This work is supported by European Research Council grant ERC-2014-CoG 647557.

References

[1] L. Babai, “Trading group theory for randomness,” in *Proceedings of the seventeenth annual ACM symposium on Theory of computing*. ACM, 1985, pp. 421–429.

[2] S. Goldwasser and M. Sipser, “Private coins versus public coins in interactive proof systems,” in *Proceedings of the eighteenth annual ACM symposium on Theory of computing*. ACM, 1986, pp. 59–68.

[3] A. Shamir, “IP=PSPACE,” *Journal of the ACM (JACM)*, vol. 39, no. 4, pp. 869–877, 1992.

[4] S. Goldwasser, Y. T. Kalai, and G. N. Rothblum, “Delegating computation: interactive proofs for muggles,” in *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM, 2008, pp. 113–122.

[5] A. Chakrabarti, G. Cormode, and A. Mcgregor, “Annotations in data streams,” *Automata, Languages and Programming*, pp. 222–234, 2009.

[6] G. Cormode and C. Hickey, “Cheap Checking for Cloud Computing: Statistical Analysis via Annotated Data Streams,” in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018.