Thesis: Coded Matrix Multiplication – A Literature Survey

Abstract: Matrix multiplication is a fundamental building block in many machine learning models. As the input matrices may be too large to be multiplied on a single server, it is common to split input matrices into multiple sub-matrices and execute the multiplications on different servers. However, in a distributed infrastructure, it is common to observe stragglers whose performance is significantly lower than other servers at some time. Compared to replicating each task on multiple servers, coded matrix multiplication, i.e., a combination of coding theoretic techniques and distributed matrix multiplication, can tolerate the same number of stragglers with much fewer servers. The recent years have witnessed the fast development of research in coded matrix multiplication. Besides alleviating the stragglers effect, there are many new improvements and applications in the coded matrix multiplication area.

In this survey, we first describe the fundamental of coded matrix multiplication. Then we introduce several crucial coded matrix multiplication schemes to mitigate stragglers and optimization of these works on recovery threshold, communication, and computation workload. Furthermore, we review coded matrix multiplication schemes that can leverage resources from homogeneous and heterogeneous workers. In addition, we present coding schemes that can be used to enhance the security and privacy of coded matrix multiplication. Finally, we highlight works that apply the idea of coded matrix multiplication to gradient descent and machine learning models.

Committee:

- Professor Jun Li, Mentor. Queens College
- Professor Kaliappa Ravindran, The City College
- Professor Ping Ji, John Jay College