Performing classification on edge nodes using Edge-centric Distributed Deep Learning and data reduction

by Manan Khasgiwale and Vasu Sharma

Traditional Machine Learning models are usually run on clouds which leads to a couple of issues like network congestion and significantly higher latencies for real-time applications such as human activity recognition (HAR) or facial recognition. The latest research in this field is inclined towards performing Distributed Deep Learning on the edge nodes but the nodes have considerably smaller computational power compared to their cloud counterparts to perform heavy Machine Learning tasks.

We are planning to solve the above problem by extending the solution provided in the paper [1]. This paper has explored a combination of Cloud ML and edge-centric data reduction techniques to perform HAR tasks. The techniques used for data reduction in the cited research reduces the amount of data sent over the network to remote cloud servers substantially. Also, there has been significant research in optimizing Deep Learning on edge nodes using optimal container distribution and Deep Neural Network (DNN) slicing.

We are now proposing to move the Deep Learning tasks from remote cloud servers to the distributed edge nodes and along with that use the data reduction techniques (sliding window) cited in the paper [1]. To achieve distributed Deep Learning infrastructure at the edge we are going to explore the methods cited in the papers [2] and [3]. Paper [2] tries to minimize resource utilization of Deep learning tasks by using efficient container allocation and [3] tries to do Deep Neural Network slicing at the edge. We plan to achieve lower latencies for real-time applications like HAR as well as lower the memory footprint for running the Deep Learning tasks on edge nodes.

References:-

1) https://ieeexplore-ieee-org.colorado.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=9139356
2) https://www.usenix.org/system/files/hotedge20_paper_park.pdf
3) https://dl.acm.org/doi/10.1145/3318216.3363309
4) http://archive.ics.uci.edu/ml/datasets/mhealth+dataset