Neural Networks at the Edge

ABC
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

I. Introduction

Problem: Federated setting, different parties exist and different resources available with constraints on these resources. How to split a neural network to address these resource constraints?

Challenge in a federated setting and edge is that the edge has limited resources. For example, party A may provide access to party B for only limited amount of resources to ensure that its clients' needs are met. Neural networks have become quite prevelant, however are resource intensive. One way to improve the resource consumption is to reduce the footprint of the neural network (e.g., TFLite). This does not always solve the problem as depending on the resource availability, the network may not get executed at all.

We examine the problem of scoring neural networks in a federated setting with resource constraints. Idea: Splitting of the neural network layers across multiple nodes. Evaluate the initial part in party A and then use the output with confidence metric to determine partial information on the task at hand. This is based on the resource consumption for the portion of the task. Then, depending on the need and availability of resources, move to the next portion of the computation on a different node (possibly in one's own party, when it becomes available). More details of this will be in scenario section (TODO: Richard).

The splitting and conditional evaluation is explained in Splitting section (TODO: Kaushik)

Can this be extended to using Spiking (TODO: Kaushik)

Conclusions

II. Scenario and Problem Statement

Explain the scenario in further detail with a picture (perhaps).

Advantages and usefulness in a federated/coalition setting.

III. Conditional Deep Neural Networks

Explain the conditional, pointing to the existing work. https://arxiv.org/pdf/1509.08971.pdf

Identify any issues/drawbacks with this approach. How much information/confidence can be derived? Will interpretability be an issue and how much extra information be gathered from the task being partially completed?

IV. Spiking extensions

Can spiking help reduce further energy consumption? Will the splitting/conditional evluation extend to spiking and what challenges are faced?

V. Conlusions