NORTHEASTERN UNIVERSITY

EECE5698

PARALLEL PROCESSING FOR DATA ANALYTICS

Final Project Proposal

Author:
Tong Jian
Zifeng Wang

Student Number:
001224147
001211627

March 12, 2019

1 Problem

Given temporal radio signals in the form of IQ samples coming from different devices, design a classifier to predict exactly which devices they are from.

2 Dataset

532905 examples from 10000 devices. Each example is a $l \times 2$ vector. l, the length of each example may vary. We encode each device to a unique number represents the label, $e.g.$, for the 10K devices we have, we encode them to $0 \sim 10K - 1$.

3 Solution

We design a Convolutional Neural Network(CNN) as the classifier using Keras framework, an example architecture could be Fig 1.

The whole solution is designed as following. We first partition the examples to slices of equal length, so that we have a fixed input size. Then we feed batches of slices to our designed CNN with a softmax output and categorical cross entropy loss. Then we run SGD/ADAM on that loss to optimize weights of our CNN.

This could be done by multiple GPUs, which means both parallelism and distributed systems are involved.

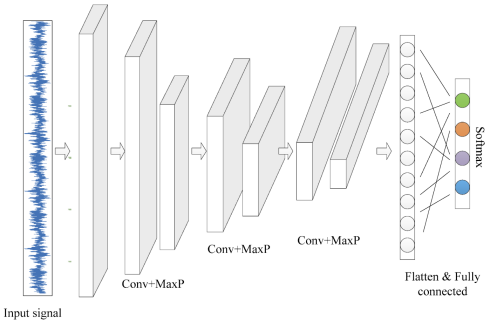

Figure 1: A possible architecture of proposed CNN model

4 Performance Analysis

We will divide the whole dataset into training, validation and test sets, use training and validation sets during training phase and test set to do final evaluation.

We will compare the running time of the same experiment (both training and testing) on 1 GPU and X GPUs, to analysis performance in terms of parallelism.

5 Work Split

We will share work on the code base and split tasks in terms of performance analysis, specifically,

- Tong will run experiments and do analysis on 1 GPU

- Zifeng will run experiments and do analysis on multiple GPUs