Distributed Ensemble

Andrea Burns   Gavin Brown   Iden Kalemaj

December 9, 2019

1 Links

- Github link: https://github.com/gavinrbrown1/distributed-ensemble-project

- Slice name: Project5

2 Introduction

The general goal of image recognition is to recognize/classify an object(s) given an image. This can be applied to a variety of images, such as those in ImageNet, which consist of natural scenes containing objects such as cars, humans, animals, and other inanimate objects. Image recognition is an important machine learning task, as it is an initial step toward visual scene understanding – a skill needed to interact with and reason about an environment (use cases include robot assistants, autonomous vehicles, etc).

Ensembling is a popular way to improve the performance of a machine learning task; an ensemble is a set of different classifiers that are 'combined' to make a collective, better decision on the task. The core idea is that combining many weak learners results in a strong learner. Two common ways to ensemble different classifiers are bagging and boosting. In bagging, classifiers are learned independently, whereas boosting tries to learn new models which succeed in cases that previous classifiers failed.

In this experiment we distribute 4 independent classifiers across different network nodes (servers) to be used as an ensemble, similar to bagging, in which we take a majority vote of their decisions and use that as the final classification result. This distributed system allows us to investigate latency versus performance benefits when introducing additional classifiers to the ensemble. Lastly, we try hand-crafted caching tools that can act as short-cut decisions, to see the impact on end-to-end delay.

A client wishing to receive classification for an image connects to a manager and sends the image to the manager. The manager then sends copies of the image to the classifiers

3 Experimental methodology

4 Results

4.1 Usage instructions

4.2 Analysis results

5 Conclusion

6 Division of labor