HornProposal

Abstract

Horn [h:n] (korean meaning of Horn is a "Spirit") is a neuron-centric programming APIs and execution framework for large-scale deep learning, built on top of Apache Hama.

Proposal

It is a goal of the Horn to provide a neuron-centric programming APIs which allows user to easily define the characteristic of artificial neural network model and its structure, and its execution framework that leverages the heterogeneous resources on Hama and Hadoop YARN cluster.

Background

The initial ANN code was developed at Apache Hama project by a committer, Yexi Jiang (Facebook) in 2013. The motivation behind this work is to build a framework that provides more intuitive programming APIs like Google's MapReduce or Pregel and supports applications needing large model with huge memory consumptions in distributed way.

Rationale

While many of deep learning open source softwares such as Caffe, DeepDist, DL4j, and NeuralGiraph are still data or model parallel only, we aim to support both data and model parallelism and also fault-tolerant system design. The basic idea of data and model parallelism is use of the remote parameter server to parallelize model creation and distribute training across machines, and the BSP framework of Apache Hama for performing asynchronous mini-batches. Within single BSP job, each task group works asynchronously using region barrier synchronization instead of global barrier synchronization, and trains large-scale neural network model using assigned data sets in BSP paradigm. Thus, we achieve data and model parallelism. This architecture is inspired by Google's DistBelief (Jeff Dean et al, 2012).

Initial Goals

Some current goals include:

- builds new community
- provides more intuitive programming APIs
- needs both data and model parallelism support
- must run natively on both Hama and Hadoop2
- needs also GPUs and InfiniBand support (FPGAs if possible)

Current Status

Meritocracy

The core developers understand what it means to have a process based on meritocracy. We will provide continuous efforts to build an environment that supports this, encouraging community members to contribute.

Community

A small community has formed within the Apache Hama project community, universities, and companies such as deep learning startup, instant messenger service company, and mobile manufacturing company. And many people are interested in the large-scale deep learning platform itself. By bringing Horn into Apache, we believe that the community will grow even bigger.

Core Developers

Edward J. Yoon, Thomas Jungblut, Jungin Lee, and Minho Kim

Known Risks

Orphaned Products

Apache Hama is already a core open source component at Samsung Electronics, and Horn also will be used by Samsung Electronics and Cldi Inc., and so there is no direct risk for this project to be orphaned.

Inexperience with Open Source

Some are very new and the others have experience using and/or working on Apache open source projects.

Homogeneous Developers

The initial committers are from different organizations such as, Microsoft, Samsung Electronics, Seoul National University, Technical University of Munich, KAIST, LINE plus, and Cldi Inc.

Reliance on Salaried Developers

Few will be worked as a full-time open source developer. Other developers will also start working on the project in their spare time.

Relationships with Other Apache Products

- Horn is based on Apache Hama
- Apache Zookeeper is used for distributed locking service
- Natively run on Apache Hadoop and Mesos
- Horn can be somewhat overlapped with Singa podling (If possible, we'd also like to use Singa or Caffe to do the heavy lifting part).

An Excessive Fascination with the Apache Brand

Horn itself will hopefully have benefits from Apache, in terms of attracting a community and establishing a solid group of developers, but also the relation with Apache Hadoop, Zookeeper, and Hama. These are the main reasons for us to send this proposal.

Documentation

Initial plan about Horn can be found at http://blog.udanax.org/2015/06/googles-distbelief-clone-project-on.html

Initial Source

The initial source code has been release as part of Apache Hama project developed under Apache Software Foundation. The source code is currently hosted at https://svn.apache.org/repos/asf/hama/trunk/ml/src/main/java/org/apache/hama/ml/ann/

Cryptography

Not applicable.

Required Resources

Mailing Lists

- horn-private
- horn-dev

Subversion Directory

- Git is the preferred source control system: git://git.apache.org/horn

Issue Tracking

- a JIRA issue tracker, HORN

Initial Committers

- Thomas Jungblut (tjungblut AT apache DOT org)
- Edward J. Yoon (edwardyoon AT apache DOT org)
- Dongjin Lee (dongjin.lee.kr AT gmail DOT com)
- Minho Kim (minwise.kim AT samsung DOT com)
- Jungin Lee (jilee AT clid DOT io)
- Kyunghyun Paeng (khpaeng AT kaist DOT ac DOT kr)
- Chia-Hung Lin (chl501 AT apache DOT org)
- Behroz Sikander (behroz.sikander AT tum DOT de)
- Kisuk Lee (ks881115 AT gmail DOT com)

Affiliations

- Thomas Jungblut (Microsoft)
- Edward J. Yoon (Samsung Electronics)
- Donjin Lee (LINE Plus)
- Minho Kim (Samsung Electronics)

- Jungin Lee (Cldi Inc.)
- Kyunghyun Paeng (KAIST)
- Chia-Hung Lin (Self)
- Behroz Sikander (Technical University of Munich)
- Kisuk Lee (Seoul National University)

Sponsors

Champion

- Edward J. Yoon <ASF member, Samsung Electronics>

Nominated Mentors

- Luciano Resende <ASF member, IBM>
- Robin Anil <ASF member, Tock>
- Edward J. Yoon <ASF member, Samsung Electronics>
- Rich Bowen <ASF member, Red Hat>

Sponsoring Entity

The Apache Incubator