

MXNetProposal

⚠️ FINAL ⚠️

This proposal is now complete and has been submitted for a VOTE.

MXNet: Apache Incubator Proposal

Abstract

MXNet is a Flexible and Efficient Library for Deep Learning

Proposal

MXNet is an open-source deep learning framework that allows you to define, train, and deploy deep neural networks on a wide array of devices, from cloud infrastructure to mobile devices. It is highly scalable, allowing for fast model training, and supports a flexible programming model and multiple languages. MXNet allows you to mix symbolic and imperative programming flavors to maximize both efficiency and productivity. MXNet is built on a dynamic dependency scheduler that automatically parallelizes both symbolic and imperative operations on the fly. A graph optimization layer on top of that makes symbolic execution fast and memory efficient. The MXNet library is portable and lightweight, and it scales to multiple GPUs and multiple machines.

Background

Deep learning is a subset of Machine learning and refers to a class of algorithms that use a hierarchical approach with non-linearities to discover and learn representations within data. Deep Learning has recently become very popular due to its applicability and advancement of domains such as Computer Vision, Speech Recognition, Natural Language Understanding and Recommender Systems. With pervasive and cost effective cloud computing, large labeled datasets and continued algorithmic innovation, Deep Learning has become the one of the most popular classes of algorithms for machine learning practitioners in recent years.

Rational

The adoption of deep learning is quickly expanding from initial deep domain experts rooted in academia to data scientists and developers working to deploy intelligent services and products. Deep learning however has many challenges. These include model training time (which can take days to weeks), programmability (not everyone writes Python or C++ and like symbolic programming) and balancing production readiness (support for things like failover) with development flexibility (ability to program different ways, support for new operators and model types) and speed of execution (fast and scalable model training). Other frameworks excel on some but not all of these aspects.

Initial Goals

MXNet is a fairly established project on [GitHub](#) with its first code contribution in April 2015 and roughly 200 contributors. It is used by several large companies and some of the top research institutions on the planet. Initial goals would be the following:

1. Move the existing codebase(s) to Apache
2. Integrate with the Apache development process/sign CLAs
3. Ensure all dependencies are compliant with Apache License version 2.0
4. Incremental development and releases per Apache guidelines
5. Establish engineering discipline and a predictable release cadence of high quality releases
6. Expand the community beyond the current base of expert level users
7. Improve usability and the overall developer/user experience
8. Add additional functionality to address newer problem types and algorithms

Current Status

Meritocracy

The MXNet project already operates on meritocratic principles. Today, MXNet has developers worldwide and has accepted multiple major patches from a diverse set of contributors within both industry and academia. We would like to follow ASF meritocratic principles to encourage more developers to contribute in this project. We know that only active and committed developers from a diverse set of backgrounds can make MXNet a successful project. We are also improving the documentation and code to help new developers get started quickly.

Community

Acceptance into the Apache foundation would bolster the growing user and developer community around MXNet. That community includes around 200 contributors from academia and industry. The core developers of our project are listed in our contributors below and are also represented by logos on the mxnet.io site including Amazon, Baidu, Carnegie Mellon University, Turi, Intel, NYU, Nvidia, MIT, Microsoft, [TuSimple](#), University of Alberta, University of Washington and Wolfram.

Core Developers

(with [GitHub](#) logins as an FYI)

- Tianqi Chen (@tqchen)
- Mu Li (@mli)
- Junyuan Xie (@piiswrong)
- Bing Xu (@antinucleon)
- Chiyuan Zhang (@pluskid)
- Minjie Wang (@jermainewang)
- Naiyan Wang (@winstywang)
- Yizhi Liu (@javelinjs)
- Tong He (@hetong007)
- Qiang Kou (@thirdwing)
- Xingjian Shi (@sxjscience)
- Yutian Li (@hotpxl)
- Yuan Tang (@terrytangyuan)

Alignment

ASF is already the home of many distributed platforms, e.g., Hadoop, Spark and Mahout, each of which targets a different application domain. MXNet, being a distributed platform for large-scale deep learning, focuses on another important domain for which there still lacks a scalable, programmable, flexible and super fast open-source platform. The recent success of deep learning models especially for vision and speech recognition tasks has generated interests in both applying existing deep learning models and in developing new ones. Thus, an open-source platform for deep learning backed by some of the top industry and academic players will be able to attract a large community of users and developers. MXNet is a complex system needing many iterations of design, implementation and testing. Apache's collaboration framework which encourages active contribution from developers will inevitably help improve the quality of the system, as shown in the success of Hadoop, Spark, etc. Equally important is the community of users which helps identify real-life applications of deep learning, and helps to evaluate the system's performance and ease-of-use. We hope to leverage ASF for coordinating and promoting both communities, and in return benefit the communities with another useful tool.

Known Risks

Orphaned products

Given the current level of investment in MXNet and the stakeholders using it - the risk of the project being abandoned is minimal. Amazon, for example, is in active development to use MXNet in many of its services and many large corporations use it in their production applications.

Inexperience with Open Source

MXNet has existed as a healthy open source project for more than a year. During that time, the project has attracted 200+ contributors.

Homogenous Developers

The initial list of committers and contributors includes developers from several institutions and industry participants (see above).

Reliance on Salaried Developers

Like most open source projects, MXNet receives a substantial support from salaried developers. A large fraction of MXNet development is supported by graduate students at various universities in the course of research degrees - this is more a "volunteer" relationship, since in most cases students contribute vastly more than is necessary to immediately support research. In addition, those working from within corporations are devoting significant time and effort in the project - and these come from several organizations.

A Excessive Fascination with the Apache Brand

We choose Apache not for publicity. We have two purposes. First, we hope that Apache's known best-practices for managing a mature open source project can help guide us. For example, we are feeling the growing pains of a successful open source project as we attempt a major refactor of the internals while customers are using the system in production. We seek guidance in communicating breaking API changes and version revisions. Also, as our involvement from major corporations increases, we want to assure our users that MXNet will stay open and not favor any particular platform or environment. These are some examples of the know-how and discipline we're hoping Apache can bring to our project.

Second, we want to leverage Apache's reputation to recruit more developers to create a diverse community.

Relationship with Other Apache Products

Apache Mahout and Apache Spark's MLlib are general machine learning systems. Deep learning algorithms can thus be implemented on these two platforms as well. However, in practice, the overlap will be minimal. Deep learning is so computationally intensive that it often requires specialized GPU hardware to accomplish tasks of meaningful size. Making efficient use of GPU hardware is complex because the hardware is so fast that the supporting systems around it must be carefully optimized to keep the GPU cores busy. Extending this capability to distributed multi-GPU and multi-host environments requires great care. This is a critical differentiator between MXNet and existing Apache machine learning systems.

Mahout and Spark ML-LIB follow models where their nodes run synchronously. This is the fundamental difference to MXNet who follows the parameter server framework. MXNet can run synchronously or asynchronously. In addition, MXNet has optimizations for training a wide range of deep learning models using a variety of approaches (e.g., model parallelism and data parallelism) which makes MXNet much more efficient (near-linear speedup on state of the art models). MXNet also supports both imperative and symbolic approaches providing ease of programming for deep learning algorithms.

Other Apache projects that are potentially complimentary:

Apache Arrow - read data in Apache Arrow's internal format from MXNet, that would allow users to run ETL/preprocessing in Spark, save the results in Arrow's format and then run DL algorithms on it.

Apache Singa - MXNet and Singa are both deep learning projects, and can benefit from a larger deep learning community at Apache.

Documentation

Documentation has recently migrated to <http://mxnet.io>. We continue to refine and improve the documentation.

Initial Source

We currently use Github to maintain our source code:

- <https://github.com/dmlc/mxnet> (Core code)

Will need to discuss code migration with Infra.

Source and Intellectual Property Submission Plan

MXNet Code is available under Apache License, Version 2.0. We will work with the committers to get CLAs signed and review previous contributions.

External Dependencies

- required by the core code base: GCC or CLOM, Clang, any BLAS library (ATLAS, OpenBLAS, MKL), dmlc-core, mshadow, ps-lite (which requires lib-zeromq), TBB
- required for GPU usage: cudnn, cuda
- required for python usage: Python 2/3
- required for R module: R, Rcpp (GPLv2+ licensing)
- optional for image preparation and preprocessing: opencv
- optional dependencies for additional features: torch7, numba, cython (in NNVM branch)

Rcpp and lib-zeromq are expected to be licensing discussions.

Cryptography

Not Applicable

Required Resources

Mailing Lists

There are currently no mailing lists.

The usual mailing lists are expected to be setup when entering incubation:

- dev@mxnet for general development discussion and user interaction
- private@mxnet for internal PPMC (and later PMC) discussions
- commits@mxnet for all source repository commits

It's imagined that the discussion on Issue Tracking below has potential to lead to an issues@mxnet list.

Issue Tracking

Currently uses [GitHub](#) to track issues. Would like to continue to do so. Will need to discuss migration possibilities with Infra.

Committers and Affiliations

- Tianqi Chen (UW)
- Mu Li (AWS)
- Junyuan Xie (AWS)
- Bing Xu (Apple)
- Chiyuan Zhang (MIT)
- Minjie Wang (NYU)
- Naiyan Wang (Tusimple)
- Yizhi Liu (Qihoo 360)
- Tong He (Simon Fraser University)
- Qiang Kou (Indiana U)
- Xingjian Shi (HKUST)
- Joe Spisak (AWS)
- Naveen Swamy (AWS)
- Indhu Bharathi (AWS)
- Chris Olivier (AWS)
- Yutian Li (Stanford)
- Yu Zhang (MIT)
- Ziheng Jiang (AWS/Fudan University)
- Hongliang Liu (Nominum)
- Shiwen Hu (tbd)
- Zihao Zheng (Alibaba Group)
- Liang Xiang (Toutiao Lab)
- Tsuyoshi Ozawa (NTT)
- Terry Chen (Novumind)
- Yifeng Geng (Horizon Robotics)
- Jian Zhang (Horizon Robotics)
- Liang [DePeng](#) (Sun Yat-sen University)
- Yuan Tang (Uptake)
- Nan Zhu (Microsoft/Apache)
- Felix Cheung (Microsoft/Apache)
- Sandeep Krishnamurthy (AWS)

Sponsors

Champion

Henri Yandell (bayard at apache.org)

Nominated Mentors

- Sebastian Schelter (ssc at apache.org)
- Suneel Marthi (smarthi at apache.org)
- Markus Weimer (weimer at apache.org)

Sponsoring Entity

We are requesting the Incubator to sponsor this project.