INTEL NEWS FACT SHEET

Bringing Artificial Intelligence to Life with Intel

August 23, 2016

The Path to Artificial Intelligence

Artificial Intelligence (AI) may seem like a vision for a distant future, but in truth, AI is all around us as machines are increasingly learning to sense, learn, reason, act and adapt in the real world. This is transforming industries and changing our lives in amazing new ways, by amplifying human capabilities, automating tedious or dangerous tasks, and solving some of our most challenging societal problems.

Machine Learning

While less than 10 percent of servers worldwide were deployed in support of machine learning last year[1], machine learning is the fastest growing field of AI and a key computational method for expanding the field of AI. At its core, machine learning is the use of computer algorithms to make predictions based on data, allowing machines to act or think without being explicitly directed to perform specific functions. The machine is trained to recognize insightful patterns and connections between complex data, and then score or classify new, incoming data to perform tasks. Currently, it can take weeks to train machine learning models, impeding the ability for models to learn from new data and information in real time. However, with the explosion of data in our smarter and more connected world that the models can learn from, along with increased compute power, machine learning models are becoming significantly more accurate and useful. Today, Intel processors power 97 percent of servers deployed to support machine learning workloads[2].

Deep Learning

Deep learning, a branch of machine learning, is a nascent and fast-growing field. Deep learning uses neural networks to comprehend more complex and unstructured data and is delivering breakthroughs in areas like image recognition, speech recognition, natural language processing and other complex tasks. Deep learning emulates neurons and synapses in the brain, learning through iteration and the formation of complex pathways in a neural network. Many of us already benefit from these algorithms, which are used for the facial recognition/tagging feature on social media, voice recognition on our smartphones, semi-autonomous vehicle control, and many more applications.

Intel Technologies for Machine Learning

To learn and act quickly, machine learning requires tremendous computational capability to run complex mathematical algorithms and process huge amounts of data. Reducing the time to train machine models, while also improving how fast they can score data, requires a paradigm shift to distributed computing, using a robust, multi-node cluster infrastructure. Intel offers a consistent programming model and common architecture that can be used across high-performance computing, data analytics and machine learning workloads.

- Intel® Xeon Phi™ processor family enables data scientists to train complex machine algorithms faster and to run a wider variety of workloads than GPUs. The next-generation Intel® Xeon™ Phi processor, codename Knights Mill, will include enhancements for high performance machine

learning training and is expected to be available in 2017. It will integrate mixed precision performance to reduce deep learning training time and offer high memory bandwidth to increase performance for complex neural datasets.
- The Intel® Xeon® processor E5 family is the most widely deployed processor for machine learning inference, with the added flexibility to run a wide variety of data center workloads. Combining it with Altera Arria 10 FPGAs delivers excellent performance per watt and the ability to reconfigure the device to manage various workloads.
- The Intel® Scalable System Framework offers comprehensive reference architectures and designs that enhance technology interoperability and reduce deployment complexity, offering a path to broad adoption of distributed deep learning algorithms and significant reduction in time to model.
- Nervana Systems*, a recognized leader in deep learning, was acquired by Intel to further advance Intel's AI portfolio and enhance the deep learning performance and total cost of ownership of Intel Xeon and Intel Xeon Phi processors.

An Open Approach to Machine Learning

Intel actively works with the open source community and also offers a variety of libraries and APIs to accelerate AI progress and broaden access to powerful tools.

Machine Learning Framework Optimizations
- For machine learning, Intel worked with the open source community to optimize industry standard frameworks, including Caffe* and Theano*, so that customers can tap into the full performance of Intel technologies using their existing infrastructure. Customers using the optimized version of Caffe are now able to realize up to 30 times increase in performance compared to the mainstream version running on Intel architecture[3].
- Intel will optimize other major machine learning frameworks for Intel architecture by the end of 2016.

Machine Learning-Optimized Libraries
- Intel optimized the widely used Intel® Math Kernel Library (Intel® MKL) for common machine learning primitives, allowing deeper access to optimized code through a standard set of APIs at no cost.
- Intel plans to release the Intel® Math Kernel Library – Deep Learning Neural Network (Intel® MKL-DNN), offering an open source implementation of MKL's deep learning neural network layers. This will facilitate integration and adoption of popular open source deep learning frameworks.
- The Intel® Data Analytics Acceleration Libraries (Intel® DAAL) are now available in open source and support deep learning.
- Intel plans to release a deep learning software development kit by the end of 2016.

Intel Developer Zone
- Intel's developer zone provides comprehensive training on machine learning tools to more than 100,000 developers and will continue to expand training opportunities in 2016.
- Intel provides select academic institutions with access to the Intel Xeon Phi processor family to foster research on next-generation deep learning algorithms.

[1] Source: Intel estimate

[2] Source: Intel estimate

[3] Up to 30x software optimization improvement claim by Intel is based on customer CNN training workload running 2S Intel® Xeon® processor E5-2680 v3 running Berkeley Vision and Learning Center* (BVLC) Caffe + OpenBlas* library and then run tuned on the Intel® Optimized Caffe (internal development version) + Intel® Math Kernel Library (Intel® MKL).

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.