CENTRO NACIONAL DE MICROELECTRONICA IMSE

Scalability and Robustness of Artificial Neural Networks. Case Study on SpiNNaker

Abstract

Over many decades researchers from diverse scientific areas have used simulated neural networks in their experimentation. For computational neuroscientists large-scale simulations of neural tissue offer an attractive alternative methodology for investigating the functionality of different brain regions as it allows greater observability, experimental control and reproducibility. Computer scientists, on the other hand, in the field of machine learning aim to solve object recognition, speech recognition tasks by employing large scale simulations of neural networks such as deep learning architectures.

Nowadays, increasingly large deep learning architectures, such as Deep Belief Networks (DBNs) are the focus of current machine learning research and achieve state-of-the-art results in different domains. However, both training and execution of large-scale Deep Networks require vast computing resources, leading to high power requirements and communication overheads. Supercomputers offer significant parallelism and great opportunity for model flexibility, however they suffer from large electrical power demands, which are rarely reported, and from communication bottlenecks that limit the real-time performance of the simulation.

Neuromorphic engineering originally aimed at exploiting sub-threshold transistor dynamics to emulate neurons in silicon, efficiently and in real-time. The neuromorphic approach can be very power-efficient, as neuron dynamics are implemented directly in silicon but many neuromorphic systems are highly optimised to a particular neural model and offer minimal configurable interconnectivity, often limited by wiring density. In this context SpiNNaker constitutes a novel parallel fully programmable architecture, with communication and memory accesses optimised for spike-based computation, permitting simulation of large spiking neural networks in real-time.

The on-going work on design and construction of spike-based hardware platforms offers an alternative for running deep neural networks with significantly lower power consumption and lower latencies, but has to overcome hardware limitations in terms of noise and limited weight precision, as well as noise inherent in the sensor signal. In this context SpiNNaker is used as an exploratory neurally-inspired platform, to assess the feasibility, the power requirements and the limitations of deep learning approaches. Using SpiNNaker's programmability different trade offs can be explored, and these can guide the development of future Neuromorphic platform designs.

First we begin by characterising the power requirements and communication latencies of the SpiNNaker platform while running large-scale spiking neural network simulations. The results of this investigation lead to the derivation of a power estimation model for the SpiNNaker system and the characterisation of the intra- and inter-SpiNNaker chip spike latencies imposed by the fabric and the software overheads. The second part of this research focuses on a full characterisation of spiking DBNs by developing a set of case studies to determine the impact of the hardware bit precision, the input noise, weight noise, and combinations on the

classification performance of a deep network for handwritten digit recognition. The results demonstrate that spiking DBNs can be realised on limited precision hardware platforms without drastic performance loss, and thus offer an excellent compromise between accuracy and low-power, low-latency execution. SpiNNaker is used as an exploration platform to verify the correctness of the results, classification latencies and to estimate the scalability, in terms of power requirements, of running deep learning models on the SpiNNaker platform. These studies provide important guidelines for informing current and future efforts to develop custom large-scale digital and mixed-signal spiking network platforms.

EVANGELOS STROMATIAS

SpiNNaker group at University of Manchester.

IMSE-CNM, September 11, 2015