Title: "Exploiting Neuromorphic Networks to predict the future based on the past"

Abstract:

The ever-increasing demand to extract temporal correlations across sequential data and perform context-based learning in this era of big data has led to the development of long short-term memory (LSTM) networks. Furthermore, there is an urgent need to perform these time-series data-dependent applications, including speech/video processing and recognition, and language modeling and translation, on compact Internet-of-Things (IoT) edge devices with limited energy. To this end, for the first time, we propose an extremely area- and energy-efficient LSTM network implementation exploiting the passive resistive random access memory (RRAM) crossbar array. We developed a hardware-aware LSTM network simulation framework and performed an extensive analysis of the proposed LSTM implementation considering the nonideal hardware artifacts such as spatial (device-todevice) and temporal variations, nonlinearity, and noise, utilizing an experimentally calibrated comprehensive phenomenological model for passive RRAM crossbar array. Our results indicate that the proposed passive RRAM crossbar-based LSTM network implementation not only outperforms the prior digital and active 1Transistor-1RRAM (1T-1R) crossbar-based LSTM implementations by more than three orders of magnitude in terms of area and two orders of magnitude in terms of training energy for identical network accuracy but also exhibits robustness against spatial and temporal variations and noise, and a faster convergence rate.