Energy-Efficient Machine Learning Algorithms *

Elsa Dupraz[1], **Lav Varshney**[2]

[1,2,4] *IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France*
[3] *University of Illinois at Urbana Champain, USA*
E-mail: [1] elsa.dupraz@imt-atlantique.fr,
[2] varshney@illinois.edu

Keywords

Machine Learning; Faulty Hardware; Energy optimization; Binary Recursive Estimation

Summary

Machine Learning algorithms are known to be highly energy consuming due to the large amount of data and computation operations required by these algorithms. It is of high importance to lower this energy consuption in order to reduce the environmental impact and also to improve the learning performance under limited computational ressources.
When designing electronic systems, a standard technique to reduce the energy consumption consists of agressively downscaling the voltage supply. However, due to physical limitations, further reducing the power supply of next generations of electronic devices will make computational units unreliable, which may introduce faults in the computation operations realized on these chips [3]. On the other hand, tolerating faults in the computation operations gives us the opportunity to address a tradeoff between algorithm performance and energy consumption. This is the issue we consider in this talk.
The first part of this talk reviews existing works on fault-tolerant computation and learning. Linear computation requires to (almost) retrieve the exact value of the function output, and fault-tolerant linear computation was studied in [5, 6, 10, 15]. In addition, many machine learning problems have been considered recently under faulty hardware. For instance, noisy hypothesis testing and noisy parameter estimation were considered in [1], logistic regression was studied in [14], and neural networks were described in [12]. In the field of error-correction, noisy Low Density Parity Check (LDPC) decoders have also been widely investigated in [4, 7, 8, 13]. Unlike linear computation, the above problems are naturally robust to errors introduced by the hardware. But the above works mainly focus on fault-tolerance, and do not make the connection with energy consumption.
In a second part of this talk, we consider the problem of recursive binary estimation under faulty hardware. Recursive binary estimation [2, 9] consists of estimating a sequence of statistically dependent hidden states from their noisy observations. It is considered in many applications such as target tracking, speech, or image processing, see [11] for a review. To the best of our knowledge, the problem of noisy recursive binary estimation was not studied yet in the literature.
We first focus on studying the robustness to faults of recursive binary estimation. We propose a theoretical analysis that bounds the expected gap between the noisy recursion and the noiseless one. We prove that this gap converges to a fixed point, which shows the robustness of the recursive binary estimation. Finite-length simulations show the accuracy of the proposed analysis. Then, we derive a model that relates the amount of faults in the computation to the energy required to performed the computation. We consider two energy allocation strategies. In a first case, we assume that energy allocation can vary from time to time, while in the second case, we assume that energy allocation can vary from bit to bit. In the two cases, we exhibit optimal energy allocation strategy in order to maximize the performance of the recursion under energy constraints.

References

[1] Hao Chen, Lav R Varshney, and Pramod K Varshney. Noise-enhanced information systems. *Proceedings of the IEEE*, 102(10):1607–1621, 2014.

[2] Andrew Critch. Binary hidden Markov models and varieties. *Journal of Algebraic Statistics*, 4(1):1–30, 2013.

[3] Ronald G Dreslinski, Michael Wieckowski, David Blaauw, Dennis Sylvester, and Trevor Mudge. Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits. *Proceedings of the IEEE*, 98(2):253–266, 2010.

[4] E. Dupraz, D. Declercq, B. Vasić, and V. Savin. Analysis

* This work was supported by grant ANR-17-CE40-0020 of the French National Research Agency (ANR) and the French MEAE and MESRI through the Pavle Savic program.

and design of finite alphabet iterative decoders robust to faulty hardware. *IEEE Transactions on Communications*, 63(8):2797–2809, 2015.

[5] Elsa Dupraz, Valentin Savin, Satish Kumar Grandhi, Emanuel Popovici, and David Declercq. Practical LDPC encoders robust to hardware errors. In *IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2016.

[6] P. Gács and A. Gál. Lower bounds for the complexity of reliable boolean circuits with noisy gates. *IEEE Transactions on Information Theory*, 40(2):579–583, March 1994.

[7] Chu-Hsiang Huang, Yao Li, and L. Dolecek. Gallager B LDPC decoder with transient and permanent errors. *IEEE Transactions on Communications*, 62(1):15–28, 2014.

[8] F. Leduc-Primeau and W.J. Gross. Faulty Gallager-B decoding with optimal message repetition. In *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing*, pages 549–556, Oct. 2012.

[9] Erik Ordentlich and Tsachy Weissman. On the optimality of symbol-by-symbol filtering and denoising. *IEEE transactions on information theory*, 52(1):19–40, 2006.

[10] N. Pippenger. On networks of noisy gates. In *26th Annual Symposium on Foundations of Computer Science*, pages 30–38, Oct. 1985.

[11] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[12] Cesar Torres-Huitzil and Bernard Girau. Fault and error tolerance in neural networks: A review. *IEEE Access*, 5:17322–17341, 2017.

[13] L.R. Varshney. Performance of LDPC codes under faulty iterative decoding. *IEEE Transactions on Information Theory*, 57(7):4427–4444, 2011.

[14] Yaoqing Yang, Pulkit Grover, and Soummya Kar. Fault-tolerant distributed logistic regression using unreliable components. In *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*, pages 940–947. IEEE, 2016.

[15] Yaoqing Yang, Pulkit Grover, and Soummya Kar. Computing linear transformations with unreliable components. *IEEE Transactions on Information Theory*, 63(6):3729–3756, 2017.