Contributions to Efficient Machine Learning


Doctoral Dissertation
Doctoral Program in Electronic and Telecommunications Engineering (33rd cycle)

Contributions to Efficient Machine Learning

Giulio Franzese
* * * * * *

Supervisor
Prof Monica Visintin

Doctoral Examination Committee:
Prof. Giuseppe Caire, Referee, Technische Universitat Berlin
Prof. Maria Alejandra Zuluaga Valencia , Referee, EURECOM

Politecnico di Torino

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Giulio Franzese
Turin,

Summary

In recent years, the words machine learning and artificial intelligence have become common words in news and pop culture.

Although the field traces back its origins into the 50's of past century, the research area has seen highs and lows, the latter colloquially referred as *AI winter*. Experts and researchers in the field have partially attributed the recent renaissance to the increase of computational power that has unlocked the possibility of performing large scale experiments and consider more complex models. Recently, however, researchers has started to warn that the trend of simply throwing more computational power to be able to achieve better results is no longer sustainable both from a computational [2] and an environmental point of view [1]. In this work we focus in particular on a usually neglected aspect of the computational cost balance: the hyperparameters and hardware topology optimization loop. We do argue, in fact, that comparing efficiency of methods as, for example, training time required to reach a target accuracy is not a completely correct procedure. Actually, in fact, the training/design complexity should be computed as the cost of training a given model with a given topology times the number of different trials that are necessary to choose a good performing training scheme. This multiplicative factor, seldom reported in scientific publications, can easily be in the order of the hundreds, if not more. Experience and the access to automatic searching tools can lessen the burden of the number of different trials one has to do, but the current status is far from being solved.

We argue that an interesting direction, from a practical point of view, is the one in which we potentially sacrifice optimality of performance in favour of more efficient schemes. In the first part of the thesis, we focus on the large domain of stochastic gradient Markov Chain Montecarlo methods (SG-MCMC), a class of algorithms designed to collect samples from complex probability distributions, in which many hyperparameters need to be selected. We consider an alternative algorithm that, on a sound theoretical basis, lessens the burden of the need to find accurate schedules for learning rates. In the second part of the thesis, we explore instead the realm of classification trees and consider a novel algorithm that builds "reversed trees" based on the information theoretic concept of Information Bottleneck. The proposed algorithm is extremely simple, robust to architectural choices and built

using modular, parallelizable and easy to scale basic blocks.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.