Blekinge Institute of Technology
Doctoral Dissertation Series No. 2020:02
ISSN 1653-2090
ISBN 978-91-7295-396-3

ENERGY EFFICIENCY IN MACHINE LEARNING

APPROACHES TO SUSTAINABLE DATA STREAM MINING

Eva García Martín

Akademisk avhandling
för avläggande av teknologie doktorsexamen vid Blekinge Tekniska Högskola kommer att offentligt försvaras i sal J1650, Campus Gräsvik, den 31 januari 2020, kl. 13:15.

Handledare:
Professor Håkan Grahn,
Institutionen för datavetenskap

Professor Veselka Boeva,
Institutionen för datavetenskap

Dr. Emiliano Casalicchio (docent),
Institutionen för datavetenskap

Opponent:
Professor Jesse Read, École Polytechnique

Betygsnämnd:
Dr. Miquel Pericàs (docent),
Chalmers Tekniska Högskola

Professor Christoph Kessler,
Linköpings Universitet

Professor Panagiotis Papapetrou,
Stockholms Universitet

Blekinge Tekniska Högskola
Institutionen för Datavetenskap

Abstract

Energy efficiency in machine learning explores how to build machine learning algorithms and models with low computational and power requirements. Although energy consumption is starting to gain interest in the field of machine learning, the majority of solutions still focus on obtaining the highest predictive accuracy, without a clear focus on sustainability.

This thesis explores green machine learning, which builds on green computing and computer architecture to design sustainable and energy-efficient machine learning algorithms. In particular, we investigate *how to design machine learning algorithms that automatically learn from streaming data in an energy-efficient manner*.

We first illustrate how energy can be measured in the context of machine learning, in the form of a literature review and a procedure to create theoretical energy models. We then use this knowledge to analyze the energy footprint of Hoeffding trees, presenting an energy model that maps the number of computations and memory accesses to the main functionalities of the algorithm. We also analyze the hardware events correlated to the execution of the algorithm, their functions and their hyper parameters.

The final contribution of the thesis is showcased by two novel extensions of Hoeffding tree algorithms, the Hoeffding tree with *nmin adaptation* and the Green Accelerated Hoeffding Tree. These solutions are able to reduce the energy consumption of the original algorithms by twenty and thirty percent, with minimal impact on accuracy. This is achieved by setting an individual splitting criteria for each branch of the decision tree, spending more energy on the fast growing branches and saving energy on the rest.

This thesis shows the importance of evaluating energy consumption when designing machine learning algorithms, proving that we can design more energy-efficient algorithms and still achieve competitive accuracy results.

Keywords: machine learning, energy efficiency, data stream mining, green machine learning, edge computing