A Brief History of Machine Learning

Norbert Cristian BERECZKI

April 2018

Abstract

One of the most actual and interesting topics of the new millennium is how to make machines learn and help our society. In this paper we will present a brief history of machine learning and how it came to be the hot topic that it is today. We will also put much emphasis on Deep Learning and its potential.

1 Motivation

In today's world, Machine Learning is an indispensable part of our life. Many of the technologies that are build in this time are based on Machine Learning. Moreover, it creates new jobs and automates many simple-to-do tasks. So, it looks more and more that people will stop doing mundane tasks and start concentrating on complex ones. Consequently, it is natural to investigate how it came to the capabilities that ML-tools provide and the direction the domain is heading towards.

This paper consists mostly of 2 chapters. The first one presents a more general view of how ML tools developed and the most important milestones that occurred. The second chapter puts emphasis upon Deep Learning(DL), and why, even though neural networks were invented a long time ago, it became so popular only in the past two to three decades.

2 A broader view of ML

This section comprises of more subsections that describe parts of the Machine Learning.

2.1 Statistical Learning

We begin with this topic to emphasize its importance to the field in discussion. It is fair to say that without statistics there would be no Machine Learning nor there would be any advanced AI algorithm that we have today. However, we will only concentrate upon the part which includes *learning*.

Interesting enough, the first tabulating machine made by Herman Hollerith was created for the government to offer support for some statistical computations related to the population (1890). Statistical Learning is preoccupied with the employing of statistics into modeling data. Most of the state-of-the-art techniques use at some degree statistics, but in this part we will deal only with the ones that make much use of them.

The revolutionary work made by Lagrange and Gauss (in the 19th century) lead to the discovery of the *linear least square method* which was a sort of **regression** method. By the time of approximately 1950, more and more linear classifications and regression methods were developed. However only after 1970 most people started using these methods in practical applications such as predictive analysis, anomaly detection and classification. This is mostly due to the rise of the more powerful computers which could support complex computations and had much more memory and to the rise of non-linear classification and regression techniques. Not to forget the importance of the creation of the R language which boosted the field of statistics and provided very good (and optimized) tools for the development of statistical models and inference models. [4]

2.2 Unsupervised Learning

Methods provided by this topic are widely used in subfields such as Data Science, Data Analysis and Business Intelligence. Unsupervised learning implies finding structure and ordering data where there is unlabeled data. So you will not be able to know for certain how well your model works.

The most trivial example for this is **clustering**. The apparition of clustering techniques (including here the most known clustering algorithm: **k-means**) is a little bit debatable many people referring to it as clustering or using k-means but not publishing it. Although, the first one to publish work related to the k-means algorithm is Stuart Lloyd in 1957. In the early 2000s the rise of data mining led to the invention of yet another clustering algorithm, namely the hierarchical clustering algorithm. Although it is not as used as other, more advanced algorithms, the hierarchical clustering algo had much popularity until approximately 2010s.[1]

Other areas where unsupervised learning shows its prowess are the fields of Anomaly Detection, Structural Prediction and feature reduction (mainly PCA). The PCA method was created in 1901 by Karl Pearson (as [2] states) and since it has been used in many fields. However, the years 2000-2013 have known a large increase in papers upon PCA, mainly the year of 2013 where most of the papers on PCA were published, as stated in [1].

2.3 Supervised Learning

3 Looking into DL

4 Conclusion and Fun Facts

[3]

References

[1] URL: https://f1000research.com/articles/6-2012.

[2] URL: https://en.wikipedia.org/wiki/Principal_component_analysis.

[3] URL: http://www.mellanox.com/interconnected-planet/related-docs/machine-learning-infographic.pdf.

[4] Gareth James et al. An Introduction to Statistical Learning. Springer, 2013. ISBN: 978-1-4614-7137-0.