Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-29 (2016) 180–189

Deep Learning vs. Wise Learning: A Critical and Challenging Overview

Peter P. Groumpos *

* *Departement of Electrical and Computer Engineering, University of Patras, Greece (e-mail: groumpos@ece.upatras.gr).*

Abstract:
Learning is the most important thing that living creatures do. An organism cannot properly animate itself without first learning how to. Knowledge is the basis for all natural and human made systems. Using knowledge to model and control a complex dynamic system (CDS) is considered. Wisdom is carefully reviewed and related to knowledge. Deep Learning (DL) and Wise Learning (WL) are presented and analyzed as two approaches to address the challenging problem of modelling and controlling CDS. Strong and weak issues of both methods are discussed. Future research directions are provided.

Keywords: Learning, Knowledge, Wisdom, Deep Learning, Wise Learning, Fuzzy Cognitive Maps

1. INTRODUCTION

In this Plenary paper the most important scientific challenge of knowledge learning is reviewed thought two different approaches: Deep Learning (DL) and Wise Learning (WL).

In recent years, deep artificial neural networks have won numerous contests in pattern recognition and machine learning. Shallow and Deep Learners are distinguished by the depth of their credit assignment paths, which are chains of possibly learnable, causal links between actions and effects. I review deep supervised learning (also recapitulating the history of back propagation), unsupervised learning, reinforcement learning and evolutionary computation, and indirect search for short programs encoding deep and large networks.

Learning is the most important thing that living creatures do. As far as any living creature is concerned, any action that does not involve learning is pretty much a waste of time. This is especially so for a human one. An organism cannot properly animate itself without first learning how to. Humans, before they can satisfy their own needs, first have to learn these needs, to understand and carefully evaluate them before they decide how to satisfy them. Indeed not an easy task.

There have been many times when people thought that they no longer needed knowledge of the past. The bright new future was to be conquered with new tools, based only on new knowledge, and everything from the past was thought to have exceeded its expiry date, and hanging on to it would only slow down the evolution of the new and previously unimaginable future. Occasional unresolved issues in the strategies implemented could be ignored; someone somewhere would come up with a new way of solving them if necessary. Precautions based on lessons learned in the past were not only unnecessary; they were retrogressive and detrimental to the development of society. Old practices, habits and experiences had to be buried and forgotten. The new can be born without any seed from the past. However this kind of ignorance has led humankind to take catastrophic and unwise decisions. We all know that most systems have memory that cannot be ignored. And this for sure is the case for all natural and human made systems. On the classical state space approach we know that given the linear time invariant system.

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1}$$

$$y(t) = Cx(t) + Du(t) \tag{2}$$

where $x(t)$ is a n x 1 state vector, $u(t)$ is the m x 1 input of the system and $y(t)$ is a r x 1 vector describing the system's output and the A, B, C and D are the known matrices with the appropriate dimensions of the dynamic system the total response equals:

Total response = the response due to initial conditions + the response due to the input.

Or in the mathematical form:

$$y(t) = C(t)\Phi(t_0, x_0)x_0 + C(t)\int_{t_0}^{t} \Phi(t,\tau)B(\tau)u(\tau)d\tau + D(t)u(t) \tag{3}$$

Where x_0 includes all past history of the system. However today we insist to forget this fundamental mathematical theory. As a result of this practice the world today is faced with many difficult and unsolved problem such as climate change, overuse of resources, misuse of energy, waste of food, pollution, illegal immigration, exploitation of the human person,international conflicts and economic and financial instabilities.

Past knowledge and experience exists all around us. We only need to study, analyze and used them carefully, systematically but above all wisely. In this research effort two different approaches are used to address the difficult

10.1016/j.ifacol.2016.11.099

issues of the problem been outlined above. The one is DL and the second is WL.

In Section 2 an overview of DL is provided while Section 3 relates knowledge to wisdom. Section 4 discusses the relationship of the human brain to DL while Section 5 presents some of the discussions and criticisms regarding DL. An extensive historical overview of DL is provided in Section 6. In Section 7 the new concept of WL is introduced, defined and its relation to knowledge and wisdom is analyzed. Section 8 provides a discussion of DL and WL and Section 9 draws conclusions and future research directions.

2. DEEP LEARNING (DL)

2.1 Introductory remarks

DL is the new big trend in machine learning. It had many recent successes in computer vision, automatic speech recognition and natural language processing. DL is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers, with complex structures or otherwise, composed of multiple non-linear transformations. Research in this area attempts to make better representations and create models to and particularly software tools that learn these representations from large-scale unlabeled data. DL software attempts to mimic the activity in layers of neurons in the neocortex, the wrinkly 80 percent of the brain where thinking occurs. The software learns, in a very real sense, to recognize patterns in digital representations of sounds, images, and other data.

The basic idea - that software can simulate the neocortex's large array of neurons in an artificial "neural network" - is decades old, and it has led to as many disappointments as breakthroughs. But because of improvements in mathematical formulas and increasingly powerful computers, computer scientists can now model many more layers of virtual neurons than ever before. DL actually is visiting again Artificial Intelligence (AI) and Artificial Neural Networks (ANN) with the objective to reformulate them given all recent scientific developments. Indeed DL has been characterized as a buzzword, or a rebranding of neural networks and Artificial Intelligence. Gomes [2014]

Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in a nervous system, such as neural coding which attempts to define a relationship between various stimuli and associated neuronal responses in the brain. There are various DL architectures such as deep neural networks, convolutional deep neural networks, deep belief networks, recurrent neural networks among other ones which are actually the same ones as those was developed in AI and ANN. A number of reviews of DL and for certain branches of AI has been reported recently. A recent one, "Deep Learning in neural networks: An overview", by [Schmidhuber, 2015] provides a thorough and extensive overview. This historical survey compactly summarizes relevant work, much of it from the previous millennium. Shallow and Deep Learners are distinguished by the depth of their credit assignment paths, which are chains of possibly learnable, causal links between actions and effects. The review covers deep supervised learning (also recapitulating the history of backpropagation), unsupervised learning, reinforcement learning & evolutionary computation, and indirect search for short programs encoding deep and large networks.

2.2 Architectures and Methods

The two key aspects of DL are: (1) models consisting of multiple layers or stages of nonlinear information processing; and (2) methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers. DL is in the intersections among the research areas of neural networks, artificial intelligence, graphical modeling, optimization, pattern recognition, and signal processing. Three important reasons for its popularity of today are the drastically increased chip processing abilities (e.g., general-purpose graphical processing units or GPGPUs), the significantly increased size of data used for training, and the recent advances in machine learning and signal/information processing research. These advances have enabled the DL methods to effectively exploit complex, compositional nonlinear functions, to learn distributed and hierarchical feature representations, and to make effective use of both labeled and unlabeled data.

There are various DL Methods and architectures:

(1) Artificial neural networks (ANN)
(2) Deep neural networks (DNN)
(3) Convolutional deep neural networks (CDNN)
(4) Deep belief networks (DBN)
(5) Recurrent (or recursive) neural networks (RNN) and
(6) Long short term memory (LSTM).

We will briefly review the above six DL methods for the sake of completeness of this overview, in Appendix A. Then more DL methods will be just named. However from the first look, all of them are revised old methods of AI and ANN.

The applications of DL are many and especially in: (1) Computer vision, (2) Natural language processing, (3) Drug discovery and toxicology, (4) Automatic speech recognition, (5) Customer relationship management, (6) Recommendation systems, (7) Audio recognition, (8) Automatic driving of vehicles and (9) Bioinformatics where they have been shown to produce state-of-the-art results on various tasks.

3. HOW KNOWLEDGE AND WISDOM ARE RELATED

Wisdom and knowledge are linked. Wisdom is enhanced by knowledge and the ability to acquire knowledge effectively. But wisdom is also the ability to use knowledge in a practical and productive manner. Knowledge is often considered to be "externally generated," meaning that it comes primarily from outside sources, such as books, classroom lectures, videos, etc. On the other hand, wisdom is deemed to come primarily from "internal sources," meaning one's own introspective thinking, analysis, and judgment. Wisdom cannot be acquired and applied without knowledge, but knowledge isn't necessarily guided or enhanced by wisdom. The application of knowledge is often