Ankara Yildirim Beyazit University-Faculty of Engineering and Natural Science

DEEP LEARNING

Busra ICOZ[*], **Seyma Nur KARAYAGLI**[**], **Sevil Busra KANLITEPE**[***]

***** Ankara Yildirim Beyazit University, Faculty of Engineerin and Natural Science, Computer Engineering, 1205012030@ybu.edu.tr

****** Ankara Yildirim Beyazit University, Faculty of Engineerin and Natural Science, Computer Engineering, 1205012013@ybu.edu.tr

******* Ankara Yildirim Beyazit University, Faculty of Engineerin and Natural Science, Computer Engineering, 1205012016@ybu.edu.tr

ABSTRACT

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech. In this paper, main topics about deep learning have been covered. The relationship between artificial intelligence, machine learning and deep learning has been mentioned briefly. Detailed information about deep learning has been given, ie. History and future of deep learning. Artificial neural networks has been reviewed. The importance of GPU and deep learning in big data have been shown deeply. Using areas of deep learning have been explained. Benefits and weaknesses of deep learning have been covered. The informations about deep learning algorithms, libraries and tools have been given.

Keywords: Deep learning; Big data, Artificial Intelligence, Machine Learning, Artificial Neural Networks, GPU

Contents

1. INTRODUCTION

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

2.ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DEEP LEARNING

2.1. Artificial Intelligence

AI is intelligence exhibited by machi-nes. In computer science, the field of AI research defines itself as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal. Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving" (known as Machine Learning). As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. For instance, optical character recognition is no longer perceived as an example of "artificial intelligence", having become a routine technology. Capabilities currently classified as AI include successfully understanding human speech, competing at a high level in strategic game systems (such as Chess and Go), self-driving cars, intelligent routing in content delivery networks, and interpreting complex data.

AI research is divided into subfields that focus on specific problems or on specific approaches or on the use of a particular tool or towards satisfying particular applications. The central problems (or goals) of AI research include reasoning, knowledge, planning, learning, natural language processing (communication), perception and the ability to move and manipulate objects. General intelligence is among the field's long-term goals. Approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics. The AI field draws upon computerscience, mathematics, psychology, linguistics, philosophy, neuroscience and ar tificial psychology.

2.2. Machine Learning

Machine learning is the subfield of computer science that gives "computers the ability to learn without being explicitly programmed." Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or unfeasible. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

2.3. Relationships between Artificial Intelligence, Machine Learning and Deep Learning

Google DeepMind’s AlphaGo program defeated South Korean Master Lee Se-dol in the board game Go, the terms AI, machine learning, and deep learning were used in the media to describe how DeepMind won. And all three are part of the reason why AlphaGo trounced Lee Se-Dol. But they are not the same things. The easiest way to think of their relationship is to visualize them as concentric circles with AI the idea that came first the largest, then machine learning which blossomed later, and finally deep learning which is driving today’s AI explosion fitting inside both.

Machine Learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world. So rather than hand-coding software routines with a specific set of instructions to accomplish a particular task, the machine is "trained" using large amounts of data and algorithms that give it the ability to learn how to perform the task.

Machine learning came directly from minds of the early AI crowd, and the algorithmic approaches over the years included decision tree learning, inductive logic programming. Clustering, reinforcement learning, and Bayesian networks among others. None achieved the ultimate goal of General AI, and even Narrow AI was mostly out of reach with early machine learning approaches. As it turned out, one of the very best application areas for machine learning for many years was computer vision, though it still required a great deal of hand-coding to get the job done. People would go in and write hand-coded classifiers like edge detection filters so the program could identify where an object started and stopped; shape detection to determine if it had eight sides; a classifier to recognize the letters "S-T-O-P." From all those hand-coded classifiers they would develop algorithms to make sense of the image and "learn" to determine whether it was a stop sign.

Deep Learning has enabled many practical applications of Machine Learning and by extension the overall field of AI. Deep Learning breaks down tasks in ways that makes all kinds of machine assists seem possible, even likely. Driverless cars, better preventive healthcare, even better movie recommendations, are all here today or on the horizon. AI is the present and the future. With Deep Learning's help.

3. DEEP LEARNING

Deep Learning, as a branch of Machine Learning, employs algorithms to process data and imitate the thinking process, or to develop *abstractions*. Deep Learning (DL) uses layers of algorithms to process data, understand human speech, and visually recognize objects. Information is passed through each layer, with the output of the previous layer providing input for the next layer. The first layer in a network is called the input layer, while the last is called an output layer. All the layers between the two are referred to as hidden layers. Each layer is typically a simple, uniform algorithm containing one kind of activation function.

Feature extraction is another aspect of Deep Learning. Feature extraction uses an algorithm to automatically construct meaningful "features" of the data for purposes of training, learning, and understanding. Normally the Data Scientist, or programmer, is responsible for feature extraction.

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

3.1. History Of Deep Learning

Certain early NNs (McCulloch and Pitts, 1943) did not learn at all. Hebb (1949) published ideas about unsupervised learning. The following decades brought shallow unsupervised NNs and supervised NNs (e.g., Rosenblatt, 1958). Early supervised NNs were essentially variants of linear regressors dating back two centuries (Gauss, Legendre).
Deep Learning networks originated in the 1960s when Ivakhnenko and Lapa (1965) published the first general, working learning algorithm for supervised deep feedforward multilayer perceptrons. Their units had polynomial activation functions combining additions and multiplications in Kolmogorov-Gabor polynomials. Ivakhnenko (1971) already described a deep network with 8 layers trained by the "Group Method of Data Handling," still popular in the new millennium. Given a training set of input vectors with corresponding target output vectors, layers are incrementally grown and trained by regression analysis, then pruned with the help of a separate validation set, where regularisation is used to weed out superfluous units. The numbers of layers and units per layer can be learned in problem-dependent fashion.

Like later deep NNs, Ivakhnenko’s nets learned to create hierarchical, distributed, internal representations of incoming data. Many later non-neural methods of Artificial Intelligence and Machine Learning also learn more and more abstract, hierarchical data representations. For example, syntactic pattern recognition methods (Fu, 1977) such as grammar induction discover hierarchies of formal rules to model observations.

3.2. Future Of Deep Learning

Unsupervised learning had a catalytic effect in reviving interest in deep learning, but has since been overshadowed by the successes of purely supervised learning. Although we have not focused on it in this Review, we expect unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object. Human vision is an active process that sequentially samples the optic array in an intelligent, task-specific way using a small, high-resolution fovea with a large, low-resolution surround. We expect much of the future progress in vision to come from systems that are trained end-to-end and combine ConvNets with RNNs that use reinforcement learning to decide where to look. Systems combining deep learning and rein-forcement learning are in their infancy, but they already outperform passive vision systems at classification tasks and produce impressive results in learning to play many different video games. Natural language understanding is another area in which deep learning is poised to make a large impact over the next few years. We expect systems that use RNNs to understand sentences or whole documents will become much better when they learn strategies for selectively attending to one part at a time.Ultimately, major progress in artificial intelligence will come about through systems that combine representation learning with complex reasoning. Although deep learning and simple reasoning have been used for speech and handwriting recognition for a long time, new paradigms are needed to replace rule-based manipulation of symbolic expressions by operations on large vectors.

4. ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANNs) or connectionist systems are a computational model used in computer science and other research disciplines, which is based on a large collection of simple neural units (artificial neurons), loosely analogous to the observed behavior of a biological brain's axons. Each neural unit is connected with many others, and links can enhance or inhibit the activation state of adjoining neural units. Each individual neural unit computes using summation function. There may be a threshold function or limiting function on each connection and on the unit itself, such that the signal must surpass the limit before propagating to other neurons. These systems are self-learning and trained, rather than explicitly programmed, and excel in areas where the solution or feature detection is difficult to express in a traditional computer program.

Neural networks typically consist of multiple layers or a cube design, and the signal path traverses from the first (input), to the last (output) layer of neural units. Back propagation is the use of forward stimulation to reset weights on the "front" neural units and this is sometimes done in combination with training where the correct result is known. More modern networks are a bit more free flowing in terms of stimulation and inhibition with connections interacting in a much more chaotic and complex fashion. Dynamic neural networks are the most advanced, in that they dynamically can, based on rules, form new connections and even new neural units while disabling others.

The goal of the neural network is to solve problems in the same way that the human brain would, although several neural networks are more abstract. Modern neural network projects typically work with a few thousand to a few million neural units and millions of connections, which is still several orders of magnitude less complex than the human brain and closer to the computing power of a worm.

New brain research often stimulates new patterns in neural networks. One new approach is using connections which span much further and link processing layers rather than always being localized to adjacent neurons. Other research being explored with the different types of signal over time that axons propagate, such as Deep Learning, interpolates greater complexity than a set of boolean variables being simply on or off.

An interesting facet of these systems is that they are unpredictable in their success with self-learning. After training, some become great problem solvers and others don't perform as well. In order to train them, several thousand cycles of interaction typically occur.

Like other machine learning methods – systems that learn from data – neural networks have been used to solve a wide variety of tasks, like computer vision and speech recognition, that are hard to solve using ordinary rule-based programming.

5. GPU IN DEEP LEARNING

Computational Neuroscience is an emerging field that provides unique opportunities to study complex brain structures through realistic neural simulations. However, as biological details are added to models, the execution time for the simulation becomes longer. Graphics Processing Units (GPUs) are now being utilized to accelerate simulations due to their ability to perform computations in parallel. As such, they have shown significant improvement in execution time compared to Central Processing Units (CPUs). Most neural simulators utilize either multiple CPUs or a single GPU for better performance, but still show limitations in execution time when biological details are not sacrificed. Therefore, we present a novel CPU/GPU simulation environment for large-scale biological networks, the NeoCortical Simulator version 6 (NCS6). NCS6 is a free, open-source, parallelizable, and scalable simulator, designed to run on clusters of multiple machines, potentially with high performance computing devices in each of them. It has built-in leaky-integrate-and-fire (LIF) and Izhikevich (IZH) neuron models, but users also have the capability to design their own plug-in interface for different neuron types as desired.

NCS6 is currently able to simulate one million cells and 100 million synapses in quasi real time by distributing data across eight machines with each having two video cards.

6. DEEP LEARNING IN BIG DATA

Big data term is being applied to large data sets, which cannot be processed by traditional data processing techniques. The big data area is growing very rapidly because with fact growth of technology with mobile devices, intelligent sensors, etc. It is much easier now to collect huge amount of data, which need to be processed. Big data usually has multiple dimensions and this make it much more difficult to process because data the processing complexity grows rapidly with dimensionality increase.

Neural networks are common tools for processing large number of data, because with neural networks the design process can be replaced with learning. However it was already demonstrated that capabilities of neural networks are growing linearly with their width and exponentially with their depth.

The prime reason for difficulties of training deep neural architectures is the vanishing gradient problem, where with an increase of number of hidden the error gradient is significantly reduced so the commonly used gradient based methods cannot be used. As the consequence the deep learning community is trying to use a combination of several of unsupervised and supervised methods. Often special data preprocessing and transformation are used for specific problem. With this approach highly-skilled researchers of artificial intelligence has to be engaged to design a problem specific approach.

Recently it has been shown that it is possible to have an universal learning systems where human involvement can be minimal. This was possible by introduction additional concertinas in neural networks across layers in MLP architectures. Such architectures were named as BMLP bridged multi-layer perceptron architectures.

6.1. Supervised Learning

Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.

6.2. Unsupervised Learning

Unsupervised machine learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data (a classification or categorization is not included in the observations). Since the examples given to the learner are unlabeled, there is no evaluation of the accuracy of the structure that is output by the relevant algorithm—which is one way of distinguishing unsupervised learning from supervised learning and reinforcement learning.

Problems associated with Big Data, including high dimensionality, streaming data analysis, scalability of Deep Learning models, improved formulation of data abstractions, distributed computing, semantic indexing, data tagging, information retrieval, criteria for extracting good data representations, and domain adaptation. Future works should focus on addressing one or more of these problems often seen in Big Data, thus contributing to the Deep Learning and Big Data Analytics research corpus.

7. APPLICATIONS

7.1 Applications of deep learning in big data analytics

As stated previously, Deep Learning algorithms extract meaningful abstract representations of the raw data through the use of an hierarchical multi-level learning approach, where in a higher-level more abstract and complex representations are learnt based on the less abstract concepts and representations in the lower level(s) of the learning hierarchy. While Deep Learning can be applied to learn from labeled data if it is available in sufficiently large amounts, it is primarily attractive for learning from large amounts of unlabeled/unsupervised data, making it attractive for extracting meaningful representations and patterns from Big Data.

Once the hierarchical data abstractions are learnt from unsupervised data with Deep Learning, more conventional discriminative models can be trained with the aid of relatively fewer supervised/labeled data points, where the labeled data is typically obtained through human/expert input. Deep Learning algorithms are shown to perform better at extracting non-local and global relationships and patterns in the data, compared to relatively shallow learning architectures . Other useful characteristics of the learnt abstract representations by Deep Learning include: relatively simple linear models can work effectively with the knowledge obtained from the more complex and more abstract data representations, increased automation of data representation extraction from unsupervised data enables its broad application to different data types, such as image, textural, audio, etc., and relational and semantic knowledge can be obtained at the higher levels of abstraction and representation of the raw data. While there are other useful aspects of Deep Learning based representations of data, the specific characteristics mentioned above are particularly important for Big Data Analytics.

Considering each of the four Vs of Big Data characteristics, i.e., Volume, Variety, Velocity, and Veracity, Deep Learning algorithms and architectures are more aptly suited to address issues related to Volume and Variety of Big Data Analytics. Deep Learning inherently exploits the availability of massive amounts of data, i.e. Volume in Big Data, where algorithms with shallow learning hierarchies fail to explore and understand the higher complexities of data patterns. Moreover, since Deep Learning deals with data abstraction and representations, it is quite likely suited for analyzing raw data presented in different formats and/or from different sources, i.e. Variety in Big Data, and may minimize need for input from human experts to extract features from every new data type observed in Big Data. While presenting different challenges for more conventional data analysis approaches, Big Data Analytics presents an important opportunity for developing novel algorithms and models to address specific issues related to Big Data. Deep Learning concepts provide one such solution venue for data analytics experts and practitioners. For example, the extracted representations by Deep Learning can be considered as a practical source of knowledge for decision-making, semantic indexing, information retrieval, and for other purposes in Big Data Analytics, and in addition, simple linear modeling techniques can be considered for Big Data Analytics when complex data is represented in higher forms of abstraction.

In the remainder of this section, we summarize some important works that have been performed in the field of Deep Learning algorithms and architectures, including semantic indexing, discriminative tasks, and data tagging. Our focus is that by presenting these works in Deep Learning, experts can observe the novel applicability of Deep Learning techniques in Big Data Analytics, particularly since some of the application domains in the works presented involve large scale data. Deep Learning algorithms are applicable to different kinds of input data; however, in this section we focus on its application on image, textual, and audio data.

8. BENEFITS AND WEAKNESSES

8.1.Benefits

Robust: No need to design the features ahead of time – features are automatically learned to be optimal for the task at hand. Robustness to natural variations in the data is automatically learned.

Generalizable: The same neural net approach can be used for many different applications and data types.

Scalable: Performance improves with more data, method is massively parallelizable.

8.2.Weaknesses

Deep Learning requires a large dataset, hence long training period. In term of cost, Machine Learning methods like SVMs and other tree ensembles are very easily deployed even by relative machine learning novices and can usually get you reasonably good results, deep learning methods tend to learn everything. It's better to encode prior knowledge about structure of images (or audio or text), the learned features are often difficult to understand. Many vision features are also not really human-understandable (e.g, concatenations/combinations of different features) and requires a good understanding of how to model multiple modalities with traditional tools.

9. ALGORITHMS, LIBRARIES AND TOOLS

9.1. Platform

Ersatz Labs - cloud-based deep learning platform [http://www.ersatz1.com/]

H20 – deep learning framework that comes with R and Python interfaces [http://www.h2o.ai/verticals/algos/deep-learning/]

9.2. Framework

Caffe - deep learning framework made with expression, speed, and modularity in mind. Developed by the Berkeley Vision and Learning Center (BVLC) [http://caffe.berkeleyvision.org/]

Torch - scientific computing framework with wide support for machine learning algorithms that puts GPUs first. Based on Lua programming language [http://torch.ch/]

9.3. Library

Tensorflow - open source software library for numerical computation using data flow graphs from Google [https://www.tensorflow.org/]

Theano - a python library developed by Yoshua Bengio's team

CONCLUSION

In this paper, a brief information about deep learning has been given. Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

Nowadays, deep learning is used in lots of area like face recognition, image capturing, speech analysis, bioinformatics. And an increasing of using deep learning is predicted. Things can be done effectively using deep learning, instead of using artificial intelligence and machine learning. In contrast to more conventional machine learning and feature engineering algorithms, Deep Learning has an advantage of potentially providing a solution to address the data analysis and learning problems found in massive volumes of input data. Also, Deep Learning, a special form of artificial neural network (ANN), is widely used in classification and analysis of picture objects.

REFERENCES

1. Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). "Deep Learning". Nature. 521: 436–444. *doi:10.1038/nature14539*

2. Schmidhuber, Jürgen *(2015).* "Deep Learning". *Scholarpedia. 10 (11): 32832.* doi*:*10.4249/scholarpedia.32832
3. Deep Machine Learning – A New Frontier in Artificial Intelligence Research – a survey paper by Itamar Arel, Derek C. Rose, and Thomas P. Karnowski. IEEE Computational Intelligence Magazine, 2013
4. A Brief History of Deep Learning By Keith D. Foote / February 7, 2017
5. Deep Learning Juergen Schmidhuber (2015), Scholarpedia, 10(11):32832.
6. Foundations and Trends R in Machine Learning Vol. 2, No. 1 (2009) 1–127 c 2009 Y. Bengio DOI: 10.1561/2200000006
7. Russell, Ingrid. "Neural Networks Module". Retrieved 2012
8. Deep belief networks at Scholarpedia
9. Hinton, G. E.; Osindero, S.; Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets" (PDF). *Neural Computation*. 18 (7): 1527–1554.
10. Intelligent Engineering Systems (INES), 2016 IEEE 20th Jubilee International Conference on Roger V. Hoang, Devyani Tanna, Laurence C. Jayet Bray, Sergiu M. Dascalu and Frederick C. Harris Jr.
11. Brain Computation Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA
12. Brain Computation Laboratory, Department of Bioengineering, George Mason University, Fairfax, VA, USA
13. The intelligent agent paradigm:
    Russell & Norvig 2003, pp. 27, 32–58, 968–972
    Poole, Mackworth & Goebel 1998, pp. 7–21
    Luger & Stubblefield 2004, pp. 235–240
    Hutter 2005, pp. 125–126
    Learning Deep Architectures for AI By Yoshua Bengio
14. Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications"(PDF).
15. Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016). Deep Learning. MIT Press.
16. Deng L., Li, J., Huang, J., Yao, K., Yu, D., Seide, F. et al. Recent Advances in Deep Learning for Speech Research at Microsoft. ICASSP, 2013.
17. P. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck (2013) "Learning Deep Structured Semantic Models for Web Search using Clickthrough Data," Proc. CIKM.