

Chapter 1

Deep Learning for Big Data Analytics

Priti Srinivas Sajja
Sardar Patel University, India

Rajendra Akerkar
Western Norway Research Institute, Norway

ABSTRACT

Traditional approaches like artificial neural networks, in spite of their intelligent support such as learning from large amount of data, are not useful for big data analytics for many reasons. The chapter discusses the difficulties while analyzing big data and introduces deep learning as a solution. This chapter discusses various deep learning techniques and models for big data analytics. The chapter presents necessary fundamentals of an artificial neural network, deep learning, and big data analytics. Different deep models such as autoencoders, deep belief nets, convolutional neural networks, recurrent neural networks, reinforcement learning neural networks, multi model approach, parallelization, and cognitive computing are discussed here, with the latest research and applications. The chapter concludes with discussion on future research and application areas.

INTRODUCTION

Deep learning refers to a kind of machine learning techniques in which several stages of non-linear information processing in hierarchical architectures are utilized for pattern classification and for feature learning. Recently, it also involves a hierarchy of features or concepts where higher-level concepts are defined from lower-level ones and where the same lower-level concepts help to define higher-level ones. With the enormous amount of data available today, big data brings new opportunities for various sectors; in contrast, it also presents exceptional challenges to utilize data. Here deep learning plays a key role in providing big data analytics solutions. The chapter discusses in brief fundamentals of big data analytics, neural network, deep learning. Further, models of deep learning are analyzed with their issues and limitations along with possible applications. Summary of the literature review is also provided in this chapter. Further, future possible enhancements are also listed in the domain. This chapter is organized as follows.

DOI: 10.4018/978-1-5225-5852-1.ch001

Section 1 introduces various fundamental topics such as big data analytics, artificial neural network, and deep learning. Section 2 highlights big data analytics by discussing large scale optimization, high dimensional data handling, and handling dynamic data. Section 3 discusses different deep models such as autoencoders, deep belief nets, deep convolutional neural networks, recurrent neural networks, reinforcement learning neural networks, multi model approach, parallelization, and cognitive computing with latest research and applications. Section 4 discusses some successful applications of deep learning for big data analytics. Section 5 discusses the issues and problems with the deep learning. Section 6 concludes the paper with summary and provides discussion on the work done so far and future research and application areas in the domain.

Big Data Analytics

Data word came from ‘Datum’ which means ‘thing given’. It has been used since early 1500 i.e. since beginning of computing. With the evolution of computing technology, the word has become more popular. Data are raw observations from domain of interest. It can be collection of numbers, words, measurements, or textual description of things. Obviously, data is everywhere and serves as an important base for business related decision making. It is also said that data is currency of knowledge as it provides basis of reasoning and analysis. Every business generates lots of data, which further act as a good resource to analyze, understand and improve the business. It is really an irony that the data which can help in improving quality of business makes the life miserable just because of our limitations to understand and use it properly. Such data creates a big problem due to its size, unstructuredness and redundancy. Some researchers identify the parameters like volume, velocity and variety as main reasons of the problem to handle data. According to Eric Horvitz and Tom Mitchell (2010) and James Manyika et al., (2011) such data when analyzed and used properly, offers a chance to solve problems, accelerates economic growth, and improve quality of life.

Big data is also a kind of data but big enough to handle. For example, consider a medical store having cough syrup bottles on shelves. The labels on the bottles show medicine name, type, components, batch number and expiry date. These data are very structured, small in amount (even for thousands of bottles in many batches, and such many medicines in the store) and homogeneous in nature. On the other hand, why the particular medicine expires so early requires really a big amount of data (and big amount of time and effort also). Such big data can act as a good resource to increase productivity and hence improves businesses in terms of quality, brand image and customer surplus. Effective use of big data can be one of the key factors of competition and growth for individual firm. It is predicted that by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions (Manyika, et al., 2011).

Big data handling encompasses activities in five dimensions. The first step is the data procurement, where data are identified and procured as per the requirement and available sources. Prior to the data collection, it is required to have proper vision and plan to make sure that how these data will be handled and what are the expectations. The volume, velocity and variety of the data make these procedures more challenging. Since the data is big, procurement would not be an easy job. It may leave some gap and results in missing data. To compensate the gap, data processing is needed, which is the second step. One may use soft computing techniques (such as fuzzy logic) to predict the missing data and fill in the gap. Next step is to find out a proper representation scheme so that the complete set of data can be presented

and preserved in a suitable format. This phase is known as data curation. The word curation comes from the concept that valuable artifacts are preserved in a museum and the museum curator preserves and facilitates artifact for further use such as analysis. Here, data curators do this job. Most of the time, the data are represented into electronic form to facilitate efficient analysis. Otherwise, due to volume of data, it is very difficult to analyze them manually. After proper representation of data, the data are ready to be analyzed. Further, the data and the analysis made can be presented in visual forms. Figure 1 represents the four dimensional activities to manage big data.

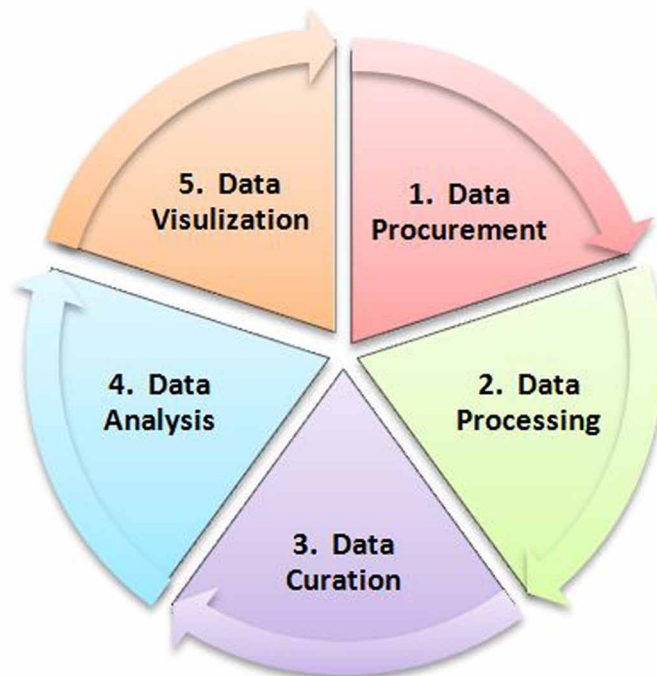
For different activities in these five dimensions, many tools are used. Statistical tools, mathematical tools, soft computing and machine learning tools, etc. can be used to manage different activities related to the big data.

Artificial Neural Network

Artificial Neural Network (ANN) is a methodology that mimics the working of nervous system in a narrow domain. Basic component of a neural network is neuron. ANN generally have many neurons connected with each other in a predefined order and are capable of learning (from the data or without data) using various paradigms. Hopfield model with parallel relaxation learning, perceptron with fixed increment learning, multi-layer perceptron with back propagation learning and Kohonen with unsupervised learning are a few popular models with the respective learning algorithm. There are many situations where data is available but it is really difficult to derive generalize logic from the data. In this case, ANNs help most.

The typical multiplayer neural networks, particularly multi-layer perceptron, consist of one (or at most two) hidden layer(s) with input and output layer. If there is no hidden layer, it can solve only linearly

Figure 1. Five dimensional activities in big data management



separable problems as illustrated in first part of the Figure 2. If there is one hidden layer, the ANN can approximate a function that having a continuous mapping from one finite space to another. If there are two hidden layers it can approximate any smooth mapping to any accuracy. The hidden layers do not interact with the external interface, but they can influence the working of ANN. Rather, introduction of the hidden layer helps the network to exhibit non-linear behavior. Figure 2 illustrates structure of multiplayer perceptron with one hidden layer.

Such typical (or shallow) neural network works well and remarkably useful for standard applications. Adding more hidden layers can help learning in perfect non-linear manner, but simultaneously it over fits the network and increase the issues related to efficiency. That is, the earlier layers can learn better, but learning later layer may get stuck.

Deep Learning

Deep Learning is a technique of machine learning that consists of many hierarchical layers to process the information in non-linear manner, where some lower level concepts help to define the higher level concepts. Deep learning is defined as “Deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs...” by Nitish Srivastava et al., (2014). The shallow artificial neural network, as stated above are not capable of handling big amount of complex data, which are obvious in many mundane applications such as natural speech, images, information retrieval and other human like information processing applications. Deep learning is suggested for such applications. With deep learning, it is possible to recognize, classify and categorize patterns in data for a machine with comparatively less efforts. A word ‘less engineering labor’ is used by Mark Bergen (2015) to mention less effort by machines. He further stated at the same source that the deep learning enables the process to unfold huge reams of previously unmanageable data. Google is pioneer to experiment deep learning, which is initiated by the Stanford Computer Scientist named Andrew Ng (now at Baidu as Chief Scientist). Experiment Google’s “deep dream” images floating around when you are Googling! Figure 3 is generated by Google (On Circulation, n.d.). The concept is just like what you see in clouds; and believe me, everybody’s perception is different.

Figure 2. Single and multilayer perceptron

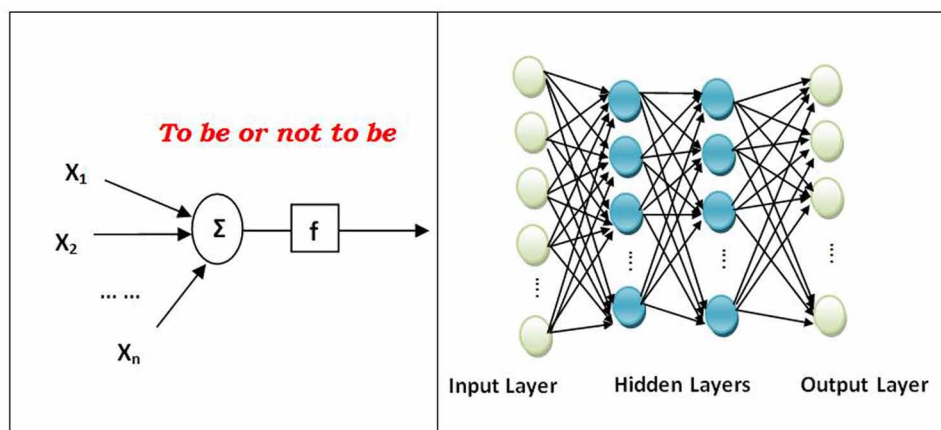


Figure 3. Deep dream image generated by Google

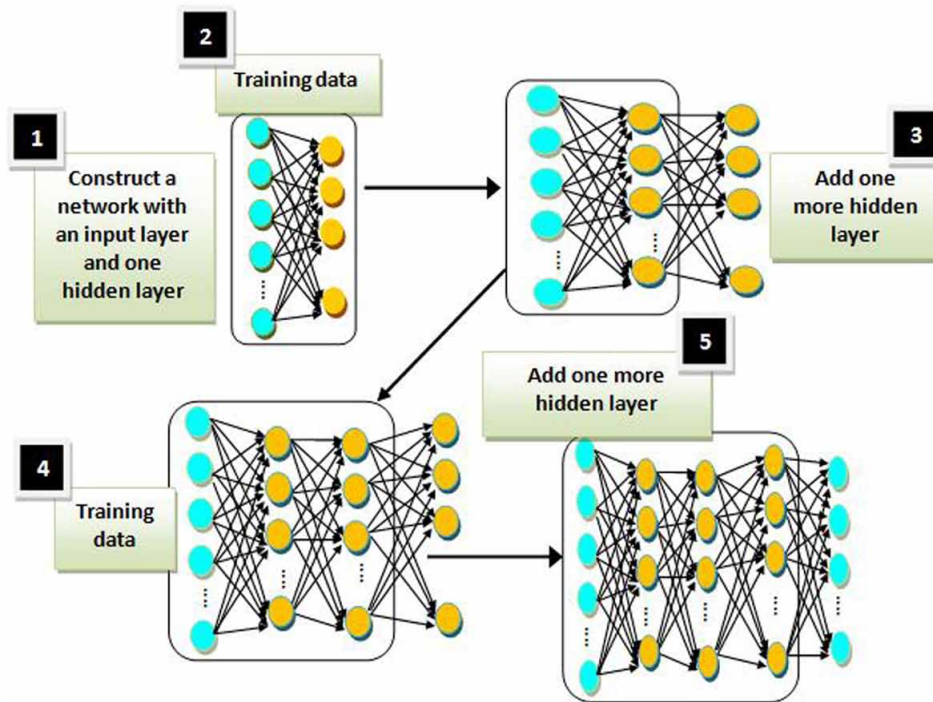


Deep learning offers human like multi layered processing in comparison with the shallow architecture. The basic idea of deep learning is to employ hierarchical processing using many layers of architecture. The layers of the architecture are arranged hierarchically. Each layer's input is provided to its adjacent layer after some pre-training. Most of the time, such pre-training of a selected layer is done in unsupervised manner. Deep learning follows distributed approach to manage big data. The approach assumes that the data are generated considering various factors, different time and various levels. Deep learning facilitates arrangement and processing of the data into different layers according to its time (occurrence), its level, or nature. Deep learning is often associated with artificial neural network.

There are three categories of deep learning architectures namely (i) generative, (ii) discriminative, and (iii) hybrid deep learning architectures (Deng, 2014). Architectures belong to the generative category focus on pre-training of a layer in unsupervised manner. This approach eliminates the difficulty of training the lower level architectures, which relay on the previous layers. Each layer can be pre-trained and later included into the model for further general tuning and learning. Doing this resolves the problem of training neural network architecture with multiple layers and enables deep learning. Neural network architecture may have discriminative processing ability by stacking output of each layer with the original data or by various information combinations and thus forming deep learning architecture. According to Li Deng (2014) the descriptive model often considers the neural network outputs as conditional distribution over all possible label sequences for the given input sequence, which will be optimized further through an objective function. The hybrid architecture combines the properties of the generative and discriminative architecture. The typical structure and mechanism of deep learning ANN is shown in Figure 4. The typical deep learning can be done as follows.

- Construct a network consisting of an input layer and a hidden layer with necessary nodes.
- Train the network.
- Add another hidden layer on the top of the previously learned network to generate a new network.

Figure 4. ANN for deep learning



- Re-train the network.
- Repeat adding more layers and after every addition, retrain the network.

ANALYZING BIG DATA

The big data analytics is required to manage huge amounts of data efficiently. The major aspects that can be considered while dealing with big data are large scale optimization, high dimensional data handling and dynamical data handling.

Large Scale Optimization

Optimization deals with finding the most effective solution to the problems using well defined procedures and models. Everybody deals with the problem of optimization either in direct (systematic) way or indirect (informal) way. The problems which need optimization include travelling salesperson problem, selecting a course from available courses under given stream (say science stream) and level (12th grade), e-commerce activities (selection of best mobile through various online shopping sites), etc.

Optimization helps in finding the cost-effective alternatives to perform the task. Commonly, it is considered as maximization or minimization of a function of resources. Some examples are maximization of profit, minimization of cost and errors. For domains with finite dimension, various models for such problem are available; however, when it comes to the big data, task of optimization becomes

challenging. In case of the big data, not only the size of transactions is voluminous, but the number of variables and number of constraints are also high. On the contrary, sometime data and constraints are moderate but their structure is complex. Such complexity in the structure increases difficulty of applying current methods on them. For example, feature learning from the large repository of medical images in optimum manner will be difficult with support of the traditional method. Further, it would require manual tuning of some parameters. Besides traditional approaches, machine learning and parallel optimization methods are also becoming popular. Some traditional methods for the optimization methods are Newton-Raphson's method, Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, conjugate gradient method and stochastic gradient descent method and stochastic gradient descent method (SGD). Many researchers have worked in this domain (Dean & Ghemawat, 2008), (Chu, et al., 2007), (Teo, Le, Smola, & Vishwanathan, 2007), (Mann, McDonald, Mohri, Silberman, & Walker, 2009), (Zinkevich, Weimer, Smola, & Li, 2010). In the work of Jeffrey Dean, et al., (2012), a problem of training a deep network with billions of parameters using tens of thousands of CPU cores is discussed. The model is successfully tested on large repository of image and then on speech recognition application. Experiments on Speech recognition was also carried out using the deep learning approach for large scale optimization by George Dahl, et al., (2012) and Geoffrey Hinton, et al., (2012). Yann Dauphin, et al., (2014) have also used deep learning for efficient optimization.

High Dimensional Data Handling

Complexity increases when the problem has more dimensions. Limited dimension of the problem makes the problem easy to solve, however, the solution is not powerful and does not provide any high-level knowledge. Increased number of dimensions results in tremendous growth of data which are difficult to handle, visualize and solve. It is said that due to exponential growth of number of possible values with each dimension, complete enumeration of all subspaces becomes intractable with increased dimensionality, which is known as the curse of dimensionality. Deep learning will be helpful in managing such high dimensional data and help in clustering, processing and visualizing such data. Bio informatics, sensor web, vision and speech recognition are fields where one can find such high dimensional data. Deep learning models are used for managing such high dimensional data (Krizhevsky, Sutskever, & Hinton, 2012), (Hinton, et al., 2012). Deep learning technique is also useful in extracting meaningful representations besides dimensionality reduction (Bengio, Courville, & Vincent, 2013). Yann Le Cun, et al., (2015) have also documented and proved the utility of deep learning technology to handle high dimensional data in various fields such as industry, business and research.

Handling Dynamical Data

Besides the volume and structure, time is another major factor that increases the complexity in the data and hence makes the job of managing data more difficult. Dynamic data are varying in terms of size, volume, and underlying structure. Large scale and dynamic data are generated and manipulated in many areas such as fluid dynamics, material science, modular dynamics and bio-inspired systems. An example domain is human speech generation. Human speech generation follows hierarchical structure. Deep learning would be useful in modeling structured speech. Researchers like Li Deng, et al., (2013) and Sabato Marco Siniscalchi, et al., (2013) have claimed that such models are really deeper and effective than the existing solutions. Deep learning is also useful in managing hidden dynamic models as

well as network also (Carneiro & Nascimento, 2013). While creating a neural network with an aim to learn deep, a node of the network can also be created dynamically. This approach can handle the aforementioned aspect of the dynamic input data. Such approach is experimented long back by Timur Ash (1989). Deep neural network approach is also used by Itamar Arel, et al., (2009) for dynamic pattern representation in combining various concepts in unsupervised manner. Researchers have used the deep learning approach to handle log time gaps between events using the non-linear and adaptive extensions (Mozer, 1989), (Hihi & Bengio, 1996).

DIFFERENT DEEP MODELS

This section introduces various models of deep learning.

Autoencoders

An autoencoder is an artificial neural network capable of learning various coding patterns. The simple form of the autoencoder is just like the multilayer perceptron containing an input layer, one or more hidden layers, and an output layer. The major difference between the typical multilayer perceptron and feed forward neural network and autoencoder is in the number of nodes at the output layer. In case of the autoencoder, the output layer contains same number of nodes as in the input layer. Instead of predicting target values as per the output vector, the autoencoder has to predict its own inputs. The broad outline of the learning mechanism is as follows.

For each input x ,

- Do a feed-forward pass to compute activation functions provided at all the hidden layers and output layers.
- Find the deviation between the calculated values with the inputs using appropriate error function.
- Back propagate the error in order to update weights.

Repeat the task till satisfactory output.

If the number of nodes in the hidden layers is fewer than the input/output nodes, then the activations of the last hidden layer are considered as a compressed representation of the inputs. When the hidden layer nodes are more than the input layer, an autoencoder can potentially learn the identity function and become useless in majority of the cases. The classic use of the autoencoder is described by Li Deng, et al., (2010).

Deep Belief Net

Deep belief network is a solution to the problem of handling non-convex objective functions and local minima while using the typical multilayer perceptron. The concept of the deep belief network was first suggested by Geoffrey Hinton, et al., (2006). This is an alternative type of deep learning consisting of multiple layers of latent variables with connection between the layers. The deep belief network can be viewed as Restricted Boltzmann Machines (RBM) where each sub-network's hidden layer acts as the visible input layer for the adjacent layer of the network. It makes the lowest visible layer training set

for the adjacent layer of the network. This way, each layer of the network is trained independently and greedily. The hidden variables are used as the observed variables to train each layer of the deep structure. The training algorithm for such deep belief network is provided as follows.

- Consider a vector of inputs.
- Train a restricted Boltzmann machine using the input vector and obtain the weight matrix.
- Train the lower two layers of the network using this weight matrix.
- Generate new input vector by using the network (RBM) through sampling or mean activation of the hidden units.
- Repeat the procedure till the top two layers of the network are reached.

The fine tuning of the deep belief network is very similar to the multilayer perceptron. Such deep belief networks are useful in acoustic modeling (Mohamed, Dahl, & Hinton, 2012), speech recognition (Mohamed, Yu, & Deng, 2010), phone recognition (Mohamed, Dahl, & Hinton, 2009) and other hierarchical process requiring deep learning.

Convolutional Neural Networks

Convolutional neural network is another variant of the feed forward multilayer perceptron. It is a type of feed forward neural network, where the individual neurons are arranged in such a way that they respond to overlapping regions in the visual field. Such network follows the visual mechanism of the living organisms. The cells in the visual cortex are sensitive to small sub-regions of the visual field, called a receptive field. The sub-regions are arranged to cover the entire visual field and the cells act as local filters over the input space. According to David Hubel and Torsten Wiesel (1968) such cells are well-suited to exploit the strong spatially local correlation present in natural images. Earlier, similar approach had been used for recognition of hand written characters called neocognitron. According to Yann Le Cun, et al., (2015), the neocognitron is considered as a predecessor of the convolutional networks. Back-propagation algorithm is used to train the parameters of each convolution kernel. Further, each kernel is replicated over the entire image with the same parameters. There are convolutional operators which extract different features of the input. Besides the convolutional layer, the network contains rectified linear unit layer, pooling layers to compute the max or average value of a particular feature over a region of the image, and a loss layer consisting of application specific loss functions. Image recognition and video analysis and natural language processing are major applications of such neural network. Some of the latest work in the area is discussed below.

Ranking convolutional neural network for the age estimation is experimented by Shixing Chen, et al., (2017). ResNet-101 deep network model for 3DMM regression is proposed by Anh Tuan Tran, et al., (2017) which uses deep learning and available for download with the network technical details and values of the chosen parameters.

Deep convolutional neural networks are also used in work of Hueu-Fang Yang, et al., (2018). The paper highlights use of the deep learning model for construction of binary hash codes from labeled data in order to search large scale images. Deep neural networks have been used for unlabeled or poorly labeled data also. Oscar Keller, et al., (2016) discusses convolutional network, its training and its performance for 1 billion hand images, where data are continuous and poorly labelled.

Deep networks have also been used for geographic information analysis such as analysis of images of earth (Audebert, Saux, & Lefevre, 2016). Recently, a very effective deep learning algorithm is introduced for channel pruning without degradation and efficiently accelerates learning of very deep network. Work of Yihui He, et al., (2017) explains it in detail.

Recurrent Neural Networks

The convolutional model works on fixed number of inputs, generates a fix sized vector as output with predefined number of steps. The recurrent networks allow us to operate over sequences of vectors in input and in output. In case of recurrent neural network, the connection between units forms a directed cycle. Unlike the traditional neural network, the recurrent neural network input and output are not independent, but related. Further, the recurrent neural network shares the common parameters at every layer. One can train the recurrent network in a way which is similar to the traditional neural network using back-propagation method. Here, calculation of gradient depends not on the current step, but on previous steps also. A variant called bidirectional recurrent neural network is also used for many applications. The bidirectional neural network not only considers the previous but also the expected future output. In bidirectional and simple recurrent neural networks, deep learning can be achieved by introducing multiple hidden layers. Such deep networks provide higher learning capacity with lots of learning data. Speech, image processing and natural language processing are some of the candidate areas where recurrent neural networks can be used. Applications of the recurrent neural network are described in various papers (Mikolov, Karafiat, Burget, Cernocky, & Khudanpur, 2010), (Mikolov, Kombrink, Burget, Cernocky, & Khudanpur, 2011), (Sutskever, Martens, & Hinton, 2011), (Graves & Jaitly, 2014).

Reinforcement Learning to Neural Networks

Reinforcement learning is a kind of hybridization of dynamic programming and supervised learning (Bertsekas & Tsitsiklis, 1996). Typical components of the approach are environment, agent, actions, policy and cost functions. The agent acts as a controller of the system; policy determines the actions to be taken; and the reward function specifies the overall objective of the reinforcement learning problem. In Reinforcement Learning, an agent plays a key role learning next course of action through trial and error-based experiments. The overall idea of the reinforcement learning is as follows.

“An agent iteratively interacts with an environment by carrying out an action u_t based on its observed state information x_t , which can be smaller (partially observable) or equal (fully observable) to the environmental state s_t . In return it retrieves a feedback in form of a reward R_{t+1} , which it uses to improve its policy and thereby increase its future sum of rewards ($t = 1, \dots, \infty$)” (Sutton & Barto, 1998).

Application of this approach can be found in robotics, games, swarm intelligence, multi agent systems and control theory.

Multi-Modalities for Deep Learning

So far, we have seen the approaches, models and applications for single modalities of the neural network, specifically deep neural networks. However, the deep neural networks can also be applied to learn features over multi modalities. Multi-model learning involves relating information from multiple sources. For example, manipulating audio and visual sources at a time falls into this category. This can be done

in two ways. First way is to consider data from all modalities and make them available at every phase. Such work is demonstrated in work of Garasimos Potamianos, et al., (2004) regarding audiovisual speech recognition. Another way follows cross modality learning. Here data from multiple modalities is available for feature learning phase only. Typically, supervised training and testing phases deal with the data from a single modality; however, Jiquan Ngiam, et al., (2011) describe a work on cross modality and present a series of tasks for multi model learning and shows training of deep networks that learn features to address these tasks. Many applications (Huang & Kingsbury, 2013), (Kim, Lee, & Provost, 2013), (Ngiam, et al., 2011) employ the approach in the field of audio visual data. Other applications of the approach include work of many researchers (Lai, Bo, Ren, & Fox, 2013), (Lenz, Lee, & Saxena, 2013) in the field of robotics with visual and depth data; and medical applications with visual and temporal data (Shin, Orton, Collins, Doran, & Leach, 2013).

Parallelization

To reduce the time of training, parallelization of the learning processes is essential. By this way, training of deep neural network becomes fast and efficient. Parallelization is done via distributing the model and applying same data sets for learning. For example, the $1000 * 1000$ weight matrix can be divided into two processing units with effective size of $1000 * 500$. Such data parallelism is fast for small network and slow for big network. Parallelism can be applied on model as well as data. In case of model parallelism, different parts of models are trained in parallel. In case of the data parallelism, data are distributed on models. In case of later choice, the model parameters must be synchronized.

Cognitive Computing

Cognitive computing is defined as a simulation of human thinking process with the help of machine learning algorithms and other areas of artificial intelligence. As stated earlier, deep learning is suitable of simulation of many human dominated processes such as perception, natural language processing, etc. helping in general thinking, and decision making. Such processes are hierarchical in nature; hence it is suitable for deep learning technique. Cognitive computing facilitated via deep learning techniques is also used for analyzing the so called dark data, which are yet to be analyzed. Dark data are generally data which are in bulk, unstructured in nature and difficult to analyze and visualize. Another aspect of the dark data is concerned with its location. They are stored generally in locations which are difficult and challenging to explore. To identify, process, analyze and visualize such dark and big data, deep learning can be utilized. IBM and other companies have started employing deep learning techniques for understanding, speech, vision, language translation, dialog and other tasks. An application for healthcare is also developed by IBM Research and Memorial Sloan Kettering Cancer Center for detecting skin cancer using smart phones and cloud computing platform. The application works in conjunction with deep learning to flag cancer images and analyze them on demand.

SUCCESSFUL APPLICATIONS OF DEEP LEARNING

The generative deep model using unsupervised manner was introduced by work of Geoffrey Hinton, et al., (2006), which was named as Deep Belief Network (DBN). The network learns with the help of a

greedy learning algorithm in order to optimize the weights. These weights are provided later to a multi-layered network as an initial weight set. Work done by Abdel Mohamed, et al., (2010), Geoffrey Hinton et al., (2012) and Abdel Mohamed, et al., (2012) also fall in the category of pre-trained back-propagation learning. A real-world application of prelearning is discussed by Veselin Stoyanov et al., (2011) using deep generative graphical models. Work of Ray Kurzweil (2012) also illustrates a generative deep learning model. A deep learning architecture is designed by Graham Taylor et al., (2006) for human motion modeling. For natural language processing and natural scene parsing, a deep learning model is proposed by Richard Socher, et al., (2011). Mohammad Havaei et al., (2015) used deep learning concept for large data set containing medical imaging.

Spoken language identification and phone recognition is done through deep discriminative learning by Dong Yu et al., (2010) and Dong Yu and Li Deng (2010) respectively. Later Dong Yu, et al., (2010) suggested an improved model for natural language processing. Nelson Morgan (2012) has also proposed a deep learning approach for automatic speech recognition. He has used discriminatively trained, feed forward neural networks for automatic speech recognition involving large and complex data.

The hybrid deep learning refers to the utilization of generative and discriminative deep learning concept in architecture. The generative concept and discriminative concept may work in cooperative or concurrent manner with the ultimate aim of getting dual advantages from both the concepts. Such hybridization is experimented in the work of Tara Sainath, et al., (2013) managing deep convolutional neural network for managing vocabulary for continuous speech recognition, which is practically very large and complex to handle with typical approaches. Similar work is presented by George Dahl, et al., (2012) while the latest work in this category is the work of Cha Zhang and Zhengyou Zhang (2014) with the basic objective to improve the multi-view face detection.

There are some other noticeable works on big data manipulation using the deep learning approach. A survey is presented in the work of Yann Le Cun, et al., (2015) as well as Maryam Najafabadi, et al., (2015). As discussed, the deep learning networks are slow and difficult to train. Further, they have tendency to over-fit themselves. To solve these problems, residual framing deep learning network is proposed by Kaiming He, et al., (2016). As per the claims of the authors, the network layers are reformulated using residual functions. The above-mentioned paper uses 152 layers and proposes really very deep learning.

Table 1 enlists some applications in various domains with reference.

Those who want to experiment ready models of the big data for various domains, pre-trained models are available at model zoo (Hasanpour, 2018). These models can be used to experiment deep learning in selected domains. The models available in this repository are ready to use, pre-learned, and can be downloaded by running following script.

```
scripts/download_model_binary.py<dirname>
```

where <dirname> is a name of directory associated with concern model in the zoo repository.

ISSUES AND LIMITATIONS OF DEEP LEARNING

Deep learning has started late in comparison with other machine learning techniques. A few of the many reasons can be lack of data and lack of able infrastructure, particularly slow processors in comparison with the modern era. It is said by Yann Lecun, et al., (2015) that the data set were approximately 1,000

Table 1. Latest work done in the area of big data manipulation using deep learning

| Reference | Domain | Year |
|--|--|------------|
| Cireřan, D., Meler, U., Cambardella, L. & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. <i>Neural Computation</i> , 22(12), 3207-3220. | Handwritten digit recognition | 2010 |
| Seide, F., Li, G. & Yu, D. (2011). Conversational speech transcription using context-dependent deep neural networks. In <i>Proceedings of Interspeech</i> , Florence, Italy (pp. 437-440). | Conversational speech transcription | 2011 |
| Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P. (2011). Natural language processing almost from scratch. <i>Journal of Machine Learning & Research</i> , 12, 2493–2537. | Natural language processing | 2011 |
| Gantz, J. & Reinsel, D. (2011). <i>Extracting value from chaos</i> . Hopkinton, MA: EMC. | Chaos Management | 2011 |
| Lin, J. & Kolcz, A. (2012). Large-scale machine learning at twitter. In <i>Proceeding of ACM SIGMOD</i> , Scottsdale, Arizona (pp. 793-804). | Twitter data handling | 2012 |
| Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition. <i>IEEE Signal Processing Magazine</i> , 29(6), 82-97. | Speech Recognition | 2012 |
| Panda, B., Herbach, J., Basu, S. & Bayardo, R. (2012). <i>MapReduce and its application to massively parallel learning of decision tree ensembles. Scaling Up Machine Learning: Parallel and Distributed Approaches</i> . Cambridge, UK: Cambridge University. | Parallel learning of decision tree | 2012 |
| Srivastava, N. & Salakhutdinov, R. (2012). Multimodal learning with deep Boltzmann machines. In <i>Proceedings of Neural Information Processing Systems</i> , Lake Tahoe, Nevada. | Multi model learning | 2012 |
| Crego, Muñoz, & Islam, (2013). Big data and deep learning: Big deals or big delusions? | Comments on big data | 2013 |
| Mikolov, T., Le, QV, Sutskever, I. (2013). Exploiting similarities among languages for machine translation. <i>Computation and Language</i> . arXiv:1309.4168 | Machine translation | 2013 |
| Zhang, K. & Chen, X. (2014). Large-scale deep belief nets with MapReduce. <i>IEEE Access</i> , 2, 395-403. | Deep belief network using MapReduce | 2014 |
| Jones, N. (2014). Computer science: The learning machines. <i>Nature</i> , 505(7482), 146-148. | Learning machines | 2014 |
| Fei-Fei, L., Karpathy, A., Leung, T., Shetty, S., Sukthankar, R., & Toderici, G. (2014). Large-scale video classification with convolutional neural networks. In <i>IEEE Conference on Computer Vision and Pattern Recognition</i> , Columbus, Ohio. | Large scale video classification | 2014 |
| Lapedriza, À., Oliva, A., Torralba, A., Xiao, J., & Zhou, B. (2014). Learning deep features for scene recognition using places database. In <i>Conference on Neural Information Processing Systems</i> , Montreal, Canada. | Scene learning | 2014 |
| He, K., Zhang, X., Ren, S. & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on image net classification. In <i>International Conference on Computer Vision</i> , Santiago, Chile. | Image classifiers from very large set of images | 2015 |
| He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. In <i>European Conference on Computer Vision</i> , Zurich. and He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. <i>IEEE Conference on Computer Vision and Pattern Recognition</i> , (pp. 770-778). Las Vegas, USA. | Image recognizer, and uses more than 150 layers | 2014, 2016 |
| Larsson, G., Maire, M., Shakhnarovich, G. (2016). Learning representations for automatic colorization. In <i>European Conference on Computer Vision</i> , Amsterdam. | Recolouring black and white images | 2016 |
| Koller, O., Ney, H., & Bowden, R. (2016). Deep Hand: How to Train a CNN on 1 Million Hand Images. In <i>IEEE International Conference on Computer Vision and Pattern Recognition</i> , Las Vegas, NV. | Convolutional network, its training and its performance for 1 billion hand images, where data are continuous and poorly labelled | 2016 |

continued on following page

Table 1. Continued

| Reference | Domain | Year |
|--|---|------|
| Audebert, N., Saux, B., & Lefevre, S. (2016). Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks. In <i>Asian Conference on Computer Vision</i> , Taipei, Taiwan. | Geographic information analysis such as analysis of images of earth | 2016 |
| He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In <i>IEEE Conference on Computer Vision and Pattern Recognition</i> , Las Vegas, NV (pp. 770-778). | Channel pruning without degradation and efficiently accelerates learning of very deep network | 2016 |
| Chen, S., Caojin, Z., Ming, D., Jialiang, L., & Mike, R. (2017). Using ranking-CNN for age estimation. Honolulu, Hawaii. | Ranking convolutional neural network for the age estimation | 2017 |
| Anh Tuan, T., Tal, H., Iacopo, M., & Gerard, M. (2017). Regressing robust and discriminative 3D morphable models with a very deep neural network. In <i>IEEE Conference on Computer Vision and Pattern Recognition</i> , Honolulu, HI. | ResNet-101 deep network model for 3DMM regression | 2017 |
| Huei-Fang, Y., Kevin, L., & Chu-Song, C. (2018). Supervised learning of semantics-Preserving hash via deep convolutional neural networks. In <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> (pp. 437-451). | Construction of binary hash codes from labeled data in order to search large scale images | 2017 |
| Aksan, E., Pece, F. & Hilliges, O. (2018). Deep Writing: Making digital ink editable via deep generative modeling. In <i>SIGCHI Conference on Human Factors in Computing Systems</i> , Montréal, Canada. | Deep Generative Modeling | 2018 |
| Wang, H., Zhang, F., Xie, X. & Guo, M. (2018). DKN: Deep knowledge-aware network for news recommendation. In <i>The Web Conference</i> , Lyon, France. | Online news recommender systems | 2018 |
| Lee, J., Park, J., Kim, K.L. & Nam, J. (2018). Sample CNN: End-to-end deep convolutional neural networks using very small filters for music classification. <i>Applied Sciences</i> , 8(150). | Music classification | 2018 |

times smaller and computing infrastructure were approximately 1,000,000 times slower in comparison with today's volume of data sets and computing infrastructure. Further, no proper consideration was given to the domain of deep learning. It was believed that adding many layers to an artificial neural network, that is increasing dimensionality of the neural network, over-fits the network and also degrades its performance. Due to really big data overflow in various domains led experts and researchers to think about the deep leaning.

Deep learning can be put in category of empirical success, just "simple parametric models trained with gradient descent" as mentioned by Francois Chollet (2017) and many real-life applications still can be solved with the deep learning techniques. However, because of simplicity and its belongingness in the category of the simple parametric and empirical model, many researchers consider it is not suitable for the real life complex problem solving. There are some other limitations of deep learning in general, which are discussed below.

- Deep neural networks are hard to train as each layer is learning at vastly different rate. Many a times the later layers learn at fast rate and starting layers still busy in learning. Reverse situation may also occur surprisingly, in which the initial layers learn in fast manner and later layers are still struggling for the same.

- Applications that require reasoning and explanation are not the candidate applications for deep learning technologies. The main reason of this is the knowledge of any neural network-based system is stored in its connection in indirect form, which is not easy to be converted in verbal form. Here one can think for neuro-fuzzy system to get dual advantages of fuzzy logic-based systems such as uncertainty and explanation; besides advantages of neural network based systems. Some example applications are planning, scheduling, automatic programming, advisory systems, etc. are the example of the same. Further, such hybridization can incorporate advance fuzzy technology such as type 2 fuzzy inference system.
- Often deep learning models misclassify the inputs to some absurd classes. This can be made possible through some tricky training sets called “adversarial examples”. Such examples are designed to confuse the model and trickedly lead the model to understand them in other way and classify into different than intended category.
- It should be noted that the deep leaning models cannot understand the inputs, hence human like understanding is bit far to achieve. By providing them validated big training data sets, one can only make the deep leaning models to learn geometric transformation from input to output without any human like understanding of the images.
- Deep models are generally experimented on narrow domains, for the concepts they are taught for; rather than the concepts they are not aware of. Humans can handle with partial, incompetent and ambiguous data also.
- The deep learning requires lots of data. Human beings are having capabilities to learn required knowledge from the very small amount of data and can come up with a physical model for problem solving. Deep learning models cannot do this. They require large amount of data. Further, in case of deep learning, even if it is trained with a sufficiently large amount of data, it may fail if it sees slightly different data.

CONCLUSION

From the discussion and aforementioned references, it is clear that deep learning might be helpful to effectively manage the great volume of complex data in a business. Many different models and algorithms have made it practically possible to implement deep learning in various domains. Some of the possible applications of deep learning to manage big data are mentioned below which can be developed with the help of tools such as MatLab using Octave toolbox for deep learning or dedicated software such as cuda-convnet¹ offering a fast C++ implementation of convolutional neural networks.

- Natural language processing and natural query.
- Film industry where dubbing of film, re-colouring film prints and adding sounds to the silent films.
- News aggregators.
- Machine Translation in partly or fully automatic manner; also sentence corrections.
- Classification, abstraction and identification of multimedia objects.
- Automatic prescription of handwriting production and understanding.
- Managing knowledge wallet.
- Chatting agents.

- Intelligent games.
- Google's automatic statistician project.
- Intelligent web applications including searching and intelligent crawling.
- Image, speech and multimedia mining.
- Utilization of social network platform for various activities.
- Development for resources and sustainable development, population information, governance (weather forecasting, infrastructural development, natural resources).
- Sensor web, agricultural information, decision support systems in domains like forestry and fisheries.

The above-mentioned domains have possibility to generate and deal with lot many data, which are big in terms of volume, velocity and variety. Typical approach may not solve the problem with the desired effectiveness. Further, from the literature survey, it is observed that only a few of them are traversed. There is a good research opportunity in the above-mentioned applications and domains. One can also consider exploring the area such as privacy, security and intellectual copy rights of the big data and its analyzed results. Beside the above-mentioned applications, the generalized deep network that are self-evolutionary in nature can also be considered as a major future enhancement. In future, models of deep learning may be more generic in nature. The generated solutions need to be not only generic, but reusable too. As stated, deep learning requires lots of data and human control when data feeding and data interpreting is concerned. So, in the future, hybrid deep learning-based systems will become more popular. The key technologies that can be used for hybridization is fuzzy logic, genetic algorithms, rough sets and other modern artificial intelligent techniques. The results also lead to general and broader Artificial Intelligence techniques. Automatic machine learning and growing reusable generic systems are the future of deep learning.

REFERENCES

- Anh Tuan, T., Tal, H., Iacopo, M., & Gerard, M. (2017). Regressing robust and discriminative 3D morphable models with a very deep neural network. In *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI.
- Arel, I., Rose, D., & Coop, R. (2009). DeSTIN: A scalable deep learning architecture with application to high-dimensional robust pattern recognition. In *AAAI Fall Symposium* (pp. 11-15).
- Ash, T. (1989). Dynamic node creation in backpropagation neural networks. *Connection Science*, 1(4), 365–375. doi:10.1080/09540098908915647
- Audebert, N., Saux, B., & Lefevre, S. (2016). Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Asian Conference on Computer Vision*, Taipei, Taiwan.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. doi:10.1109/TPAMI.2013.50 PMID:23787338

- Bergen, M. (2015, July 15). Deep learning AI is taking over tech what is it? *Recode.net*. Retrieved from <http://recode.net/2015/07/15/deep-learning-ai-is-taking-over-tech-what-is-it/>
- Bertsekas, D., & Tsitsiklis, J. (1996). *Neuro-dynamic programming*. Belmont, MA: Athena Scientific.
- Carneiro, G., & Nascimento, J. C. (2013). Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11), 2592–2607. doi:10.1109/TPAMI.2013.96 PMID:24051722
- Chen, S., Caojin, Z., Ming, D., Jialiang, L., & Mike, R. (2017). *Using ranking-CNN for age estimation*. doi:10.1109/CVPR.2017.86
- Chollet, F. (2017). *Deep Learning with Python*. New York, USA: Manning Publications.
- Chu, C. T., Kim, S. K., Lin, Y. A., Yu, Y., Bradski, G., Ng, A. Y., & Olukotun, K. (2007). *Map reduce for machine learning on multicore* (pp. 281–288). Advances in Neural Information Processing Systems Vancouver.
- Crego, E., Muñoz, G., & Islam, F. (2013, May 26). Big Data and Deep Learning: Big Deals or Big Delusions? *Huffington Post*. Retrieved from https://www.huffingtonpost.com/george-munoz-frank-islam-and-ed-crego/big-data-and-deep-learnin_b_3325352.html
- Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 30–42. doi:10.1109/TASL.2011.2134090
- Dauphin, Y., Pascanu, R., Gulcehre, C., Cho, K., Ganguly, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *27th International Conference on Neural Information Processing Systems*, Montreal, Canada (pp. 2933-2941).
- Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M., Le, Q. V., ... Ng, A. Y. (2012). Large scale distributed deep networks. In *Conference on Neural Information Processing Systems*, Lake Tahoe, NV.
- Dean, J., & Ghemawat, S. (2008). Mapreduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. doi:10.1145/1327452.1327492
- Deng, L. (2014). *A tutorial survey of architectures, algorithms, and applications for deep learning*. In *APSIPA Transactions on Signal and Information Processing*.
- Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. *International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada. 10.1109/ICASSP.2013.6639344
- Deng, L., Seltzer, M., Yu, D., Acero, A., Mohamed, A., & Hinton, G. (2010). Binary coding of speech spectrograms using a deep auto-encoder. In *INTERSPEECH 2010*, Makuhari, Chiba.
- Graves, A., & Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. In *31st International Conference on Machine Learning*. Beijing, China.
- Hasanpour, S. H. (2018). Model Zoo. Retrieved from <https://github.com/BVLC/caffe/wiki/Model-Zoo>

- Havaei, M., Davy, A., Farley, D., Biard, A., Courville, A., Bengio, Y., . . . Larochelle, H. (2015). *Brain tumor segmentation with deep neural networks*. Retrieved from <http://arxiv.org/abs/1505.03540>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV (pp. 770-778).
- Hihi, S. E., & Bengio, Y. (1996). Hierarchical recurrent neural networks for long-term dependencies. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), (pp. 493-499). *Advances in Neural Information Processing Systems*. MIT Press.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29(6), 82-97. doi:10.1109/MSP.2012.2205597
- Hinton, G., Osidero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554. doi:10.1162/neco.2006.18.7.1527 PMID:16764513
- Horvitz, E., & Mitchell, T. (2010). *From data to knowledge to action: A global enabler for the 21st century* (White Paper). A Computing Community Consortium.
- Huang, J., & Kingsbury, B. (2013). Audio-visual deep learning for noise robust speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada. 10.1109/ICASSP.2013.6639140
- Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 215-243. doi:10.1113/jphysiol.1968.sp008455 PMID:4966457
- Huei-Fang, Y., Kevin, L., & Chu-Song, C. (2018). Supervised Learning of Semantics-Preserving Hash via Deep Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 437-451. PMID:28207384
- Kim, Y., Lee, H., & Provost, E. (2013). Deep learning for robust feature generation in audiovisual emotion recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada. 10.1109/ICASSP.2013.6638346
- Koller, O., Ney, H., & Bowden, R. (2016). Deep hand: how to train a CNN on 1 million hand images. *IEEE International Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1106-1114.
- Kurzweil, R. (2012). *How to create a mind: The secret of human thought revealed*. Audio Book.
- Lai, K., Bo, L., Ren, X., & Fox, D. (2013). RGB-D object recognition: Features, algorithms, and a large scale benchmark. In A. Fossati, J. Gall, H. Grabner, X. Ren, & K. Konolige (Eds.), *Consumer depth cameras for computer vision* (pp. 167-192). London: Springer-Verlag. doi:10.1007/978-1-4471-4640-7_9
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539 PMID:26017442

- Lenz, I., Lee, H., & Saxena, A. (2013). Deep learning for detecting robotic grasps. *International Conference on Learning Representations*. Scottsdale, Arizona.
- Mann, G., McDonald, R., Mohri, M., Silberman, N., & Walker, D. D. (2009). *Efficient large-scale distributed training of conditional maximum entropy models*. In Advances in Neural Information Processing Systems.
- Manyika, J., Chui, M., Brad, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011, May). Big data: The next frontier for innovation, competition, and productivity. *McKinsey*. Retrieved March 4, 2015, from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- Mikolov, T., Karafiat, M., Burget, L., Cernocky, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In *11th Annual Conference of the International Speech Communication Association*, Chiba, Japan (pp. 1045-1048).
- Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., & Khudanpur, S. (2011). Extensions of recurrent neural network language model. In *36th International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic (pp. 5528-5531). 10.1109/ICASSP.2011.5947611
- Mohamed, A., Dahl, G., & Hinton, G. (2009). Deep belief networks for phone recognition. In *NIPS Workshop*. Vancouver, Canada.
- Mohamed, A., Dahl, G., & Hinton, G. (2012). Acoustic Modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 14–22. doi:10.1109/TASL.2011.2109382
- Mohamed, A., Yu, D., & Deng, L. (2010). *Investigation of full-sequence training of deep belief networks for speech recognition*. Makuhari, Japan: Interspeech.
- Morgan, N. (2012). Deep and wide: Multiple layers in automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 7–13. doi:10.1109/TASL.2011.2116010
- Mozier, M. C. (1989). A focused back-propagation algorithm for temporal sequence recognition. *Complex Systems*, 3, 349–381.
- Najafabadi, M., Villanustre, F., Khoshgoftaar, T., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. (2011). Multimodal deep learning. In *28th International Conference on Machine Learning*, Bellevue.
- On Circulation. (n.d.). Deep Dreaming with Google. Retrieved from <https://oncirculation.com/2015/07/14/deep-dreaming-with-google/>
- Potamianos, G., Neti, C., Luetttin, J., & Matthews, I. (2004). Audio-visual automatic speech recognition: An overview. In G. Potamianos, C. Neti, J. Luetttin, & I. Matthews (Eds.), *Issues in visual and audio-visual speech processing* (pp. 356–396). Cambridge: MIT Press.

- Sainath, T., Mohamed, A., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada.
- Shin, H., Orton, M., Collins, D., Doran, S., & Leach, M. (2013). Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1930–1943. doi:10.1109/TPAMI.2012.277 PMID:23787345
- Siniscalchi, S. M., Yu, D., Deng, L., & Lee, C. (2013). Exploiting deep neural networks for detection-based speech recognition. *Neurocomputing*, 106, 148–157. doi:10.1016/j.neucom.2012.11.008
- Socher, R., Lin, C., Ng, A., & Manning, C. (2011). Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *International Conference on Machine Learning*. Bellevue, Washington, USA.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhtdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Stoyanov, V., Ropson, A., & Eisner, J. (2011). Empirical risk minimization of graphical model parameters given approximate inference, decoding, and model structure. In 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL (pp. 725-733). .
- Sutskever, I., Martens, J., & Hinton, G. (2011). Generating text with recurrent neural networks. In *28th International Conference on Machine Learning*. Bellevue, USA.
- Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction (Adaptive computation and machine learning)*. Cambridge, MA: MIT Press.
- Taylor, G., Hinton, G., & Roweis, S. (2006). Modeling human motion using binary latent variables. In *Annual Conference on Neural Information Processing Systems*, Vancouver, Canada (pp. 1345-1352).
- Teo, C. H., Le, Q., Smola, A., & Vishwanathan, S. V. (2007). A scalable modular convex solver for regularized risk minimization. In *Conference on Knowledge Discovery and Data Mining*, San Francisco, CA.
- Yann, L., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, 521(7553), 436–444. doi:10.1038/nature14539 PMID:26017442
- Yihui, H., Xiangyu, Z., & Sun, J. (2017). Channel Pruning for Accelerating Very Deep Neural. In *IEEE International Conference on Computer Vision*, Venice, Italy (pp. 1389-1397).
- Yu, D., & Deng, L. (2010). Deep-structured hidden conditional random fields for phonetic recognition. In Eleventh Annual Conference of the International Speech Communication Association. Makuhari, Japan: Interspeech.
- Yu, D., Wang, S., & Deng, L. (2010). Sequential labeling using deep structured conditional random fields. *IEEE Journal of Selected Topics in Signal Processing*, 4(6), 965–973. doi:10.1109/JSTSP.2010.2075990

Yu, D., Wang, S., Karam, Z., & Deng, L. (2010). Language recognition using deep-structured conditional random fields. In *IEEE International Conference on Acoustics Speech and Signal Processing*, Dallas, TX (pp. 5030-5033). 10.1109/ICASSP.2010.5495072

Zhang, C., & Zhang, Z. (2014). Improving multiview face detection with multi-task deep convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, Steamboat Springs, CO (pp. 1036-1041). 10.1109/WACV.2014.6835990

Zinkevich, M., Weimer, M., Smola, A., & Li, L. (2010). Parallelized stochastic gradient descent. In *Conference on Neural Information Processing Systems*, Vancouver, Canada.

ENDNOTES

¹ <https://code.google.com/p/cuda-convnet/>