Deep Learning: Effective Tool for Big Data Analytics

Nagwa M. Elaraby
Information System Dept.,Faculty of Computers and Information,
Mansoura University, Egypt.
nagwaelaraby@gmail.com

Mohammed Elmogy
Information Technology Dept., Faculty of Computers and Information,
Mansoura University, Egypt.
elmogy@gmail.com

Shereif Barakat
Information System Dept., Faculty of Computers and Information,
Mansoura University, Egypt
sherifiib@yahoo.com

Abstract— Currently, Deep Learning is a very active research area in pattern recognition and machine learning society. It has achieved unprecedented success in applications of essential fields such as Computer Vision, Speech and Audio Processing, and Natural Language Processing. Every day there are enormous amounts of data generated by multiple sources. So the term of data is converted to Big Data which face challenges in information acquisition and decision-making processes. Dealing with these data can be supported by Deep Learning capabilities, especially its ability to deal with both the labeled and unlabeled data which are often collected abundantly in Big Data. So Deep Learning is recently coming to play a significant role in reaching solutions for Big Data analytics.

This survey discusses how Deep Learning architectures differ from convolutional structured architectures by illustrating the structure and learning aspects of most common used Deep Neural Networks. We will give an overview of Big Data and identify specific data analysis problems that can be addressed by Deep Learning. We will present some studies in Deep Learning that are used as a solution for data analysis. Finally, some Deep Learning challenges due to specific data analysis needs of Big Data will be showed.

Keywords- Deep Learning; Big Data; Boltzmann Machine (BM); Auto-Encoders (AE); Deep Neural Networks (DNN); Deep Belief Networks (DBN); Deep Stacking Networks (DSN); Big Data Analytics.

I. INTRODUCTION

Shallow-structured architectures have been used until now in most signal processing and machine learning techniques [1]. These architectures are simple and consist of a single layer, which transforms the raw input signals into a problem-specific feature space. Unfortunately, these architectures are effective only in solving well-constrained or simple problems. However, they have limitations in dealing with unstructured or complex problems. Therefore, deep architectures are needed to alleviate these limitations and solve more complicated problems.

Deep architectures are composed of feature detector units organized in multi-layers [2]. Simple features are detected by the lower layers and then feed into higher layers, which can detect more complicated features. Deep Learning works through a greedy layer-wise unsupervised pre-training by training each layer individually, using the output of the previous layer as the input to the next layer [3]. After all iterations, the set of layers with learned weights could be used in a supervised fine-tuning stage as initialization for generating a deep supervised predictor as showed in Fig.1 .

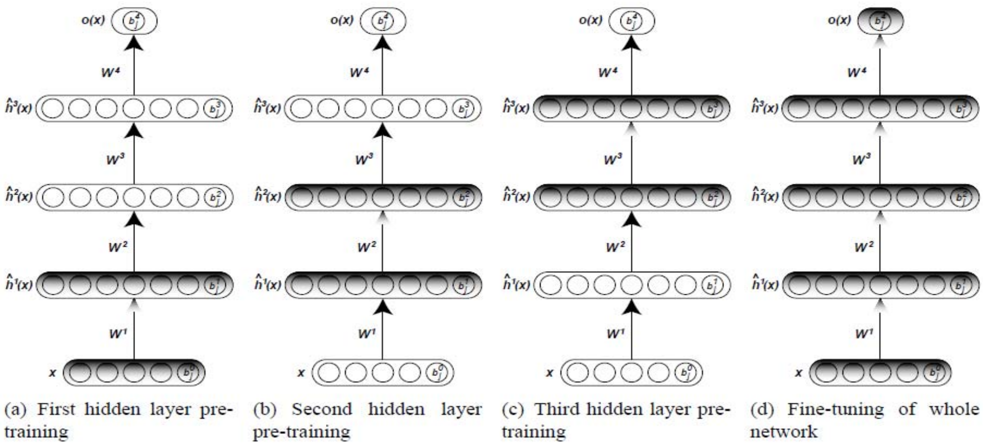


Fig. 1: The greedy unsupervised layer-wise training followed by a supervised fine-tuning stage affecting all layers [4].

A key property of Deep Learning is its ability to provide hierarchal representations for either labeled structured input data or unlabeled unstructured input data [1]. Hierarchal representation concentrates on learning high order representations from low-level data [5]. Recognizing words from audio, recognizing objects from images, and recognizing poses and movement from the video are considered as examples for this scenario. This property is the reason for the movement to Deep Learning in the computational statistics and machine learning community.

Every day, the real world collects massive amounts of unlabeled real-time data coming from multiple resources; these data are referred by the term ‘Big Data’ [6]. Big data analytics poses challenges to the typical storage, processing, and computing capacity of traditional data analysis techniques. So, Big Data needs novel architectures and technologies to become possible to reveal its hidden correlations and complex patterns.

In machine learning community, dealing with Big Data can be supported by Deep Learning according to its ability to extract complex abstractions. It provides high-level data representation from large-scale data especially unlabeled data, which are collected abundantly in Big Data [7] [8]. Big Data analytics problems can be summarized as extracting hidden patterns from massive volumes of data, fast information retrieval, data indexing/tagging, and simplifying discriminative tasks. These problems can be better solved with the aid of Deep Learning. Deep Learning together with Big Data is reflected as the ‘‘big deals and the bases for an American innovation and economic revolution’’ [9].

The remainder of this paper is organized as follows. Section II illustrates the structure and learning aspects of most common used Deep Neural Networks. Section III presents an overview of Big Data and identifies specific data analysis problems that can be addressed by Deep Learning. Section IV presents some studies in Deep Learning that are used as a solution for data analysis. Finally, Section V discusses some Deep Learning challenges due to specific data analysis needs of Big Data.

II. OVERVIEW OF DEEP LEARNING ARCHITECTURES

The objective of Deep Learning is to learn a complex and abstract data representation hierarchically, through passing the data over multiple transformation layers [8]. Therefore, Deep Learning algorithms are considered as Deep architectures of successive layers in which a nonlinear transformation is applied to each layer's input to provide a representation of its output.

There are two unsupervised single layer learning algorithms, which are used as fundamental building blocks to build deeper models: Auto-Encoders (AEs) and Restricted Boltzmann Machines (RBMs) [8]:

- AEs are unsupervised single layer learning algorithms, which are constructed of three layers: input, hidden, and output. AEs try to capture the structure of input data in a manner that makes it possible to rebuild the input in the output layer [10]. Back-propagation algorithm has been used in the training phase, and the target output is the input itself.

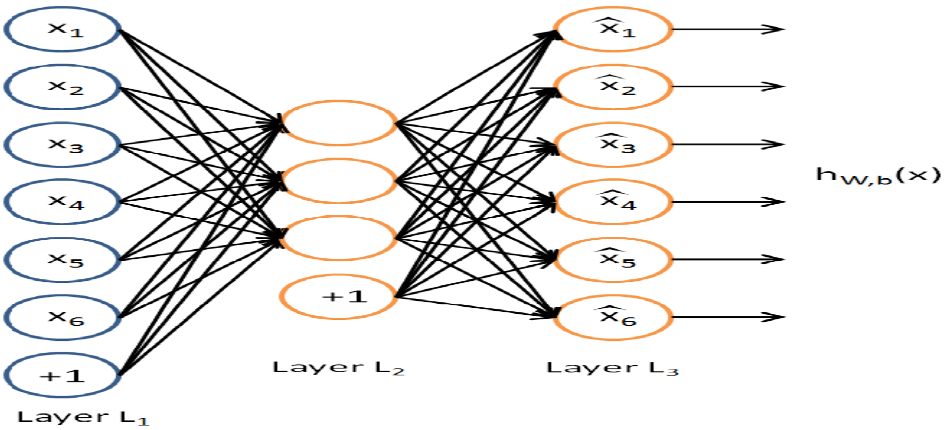


Fig2: The Auto-Encoder neural network in which the output is similar to the input [10].

- RBMs considered a special case of the conventional Boltzmann machines (BMs) [11]. BM is a non-multilayered network of symmetrically, bi-directionally connected units that make stochastic decisions about whether to be on or off. RBM is a BM with no connections between the units within the same layer. The Contrastive Divergence algorithm has mostly been applied in the training phase of BMs [12].

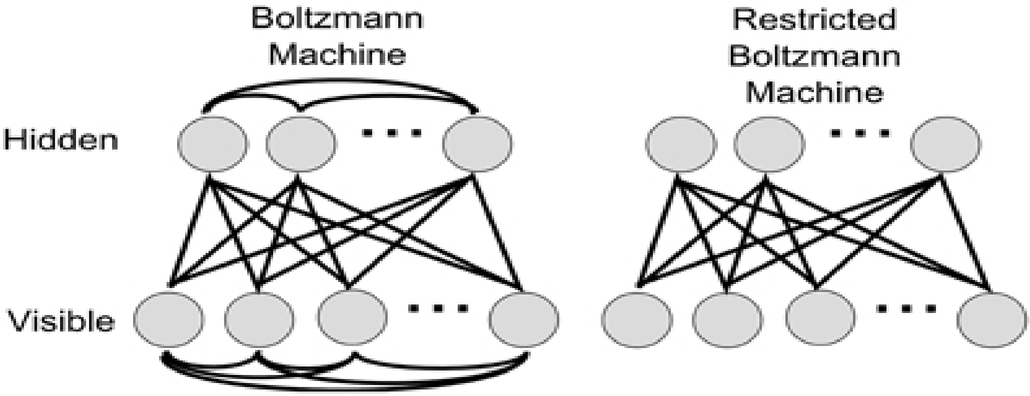


Fig3: The Restricted Boltzmann Machine as a special case of Boltzmann machine [13].

On the other hand, Deep unsupervised networks can be constructed from these unsupervised single layer learning algorithms. Here we explain the structure of mostly developed Deep Neural Networks (DNNs): Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), and Deep Stacking Networks (DSNs).

A. Deep Belief Network (DBN)

DBN consists of multi-layers of random, hidden variables that usually have binary values and used as feature detectors. It is composed of a stack of RBMs with a full set of connections between each two adjacent layers, but in the same layer, there are no connected units [7]. Connections between the two higher layers are symmetric and undirected connections and shape an associative memory. The lower layers receive from the layer above, directed and top-down connections [14].

DBN is a probabilistic generative model that learns a joint probability distribution of training data without using data labels. Therefore, it can control the complexity of highly nonlinear parameter estimation problems by providing efficient unobserved initialization points [1]. The learning phase in DBN is followed by a greedy layer-wise unsupervised pre-training which helps in avoiding local optima and over fitting problems [7].

The learning phase of DBN can be represented in the following steps [15]:

1) The first layer is trained as an RBM, which models the raw input as its visible layer.
2) The input data representation is obtained from the first layer and used as training data for the second layer. This representation can be taken as being the mean activations or samples of conditional probabilities.
3) Now the first layer becomes visible layer to the second layer. However, the second layer is trained as an RBM by taking the transformed data (samples or mean activations) as training examples.

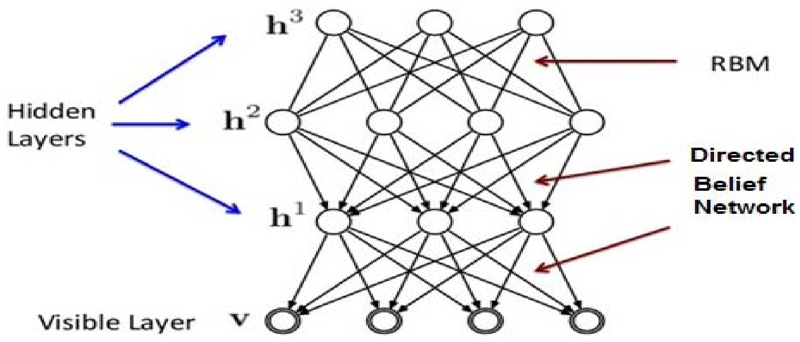


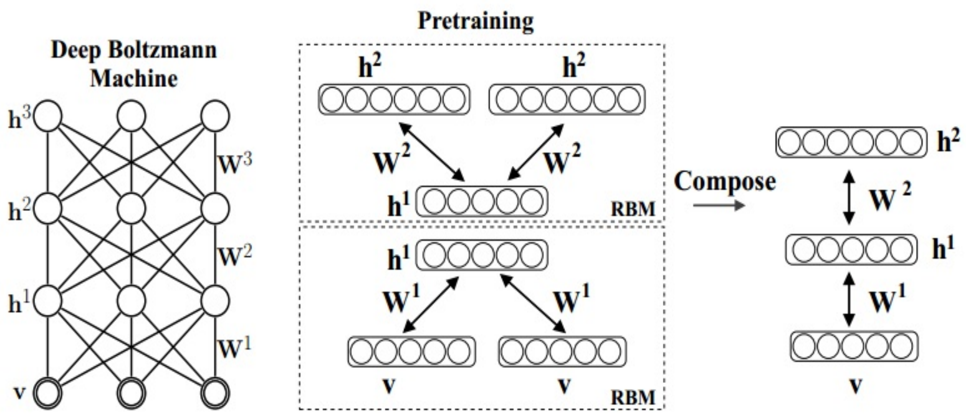
Fig4: The Deep Belief Network as a stack of Restricted Boltzmann Machines [15].

4) Step 2 and 3 will be iterated for all the layers above, each time either samples or mean values are propagated upward.
5) Now all weights of the layers are initialized. After pre-training, fine-tuning for the whole network can be done by adding a final layer to represent the preferred outputs and derivatives of back-propagating errors.

DBN combined strategies of unsupervised pre-training and supervised fine-tuning. This hybrid training strategy increases the generative performance and the discriminative power of the network [7]. So, it becomes increasingly complex and highly desirable for solving classification problems.

B. Deep Boltzmann Machines (DBM)

DBM is a type of binary pairwise Markov Random Field with multiple layers of hidden random variables [16]. It is a stack of RBM in which only adjacent layers are connected, but there are no visible to visible or hidden to hidden connections



.

Fig 5: The Deep Boltzmann Machine and its Pre-training procedure [16].

As DBN, the learning phase in DBM is followed by a greedy layer-wise unsupervised pre-training in which DBM is considered as a stack of RBMs, the output of current trained RBM is used as the input to training the next RBM. After pre-training, the learned RBMs are fine-tuned using back-propagation of error derivatives and then used to regenerate the original input data.

DBM follows a bottom-up pass with an approximate inference procedure which gives it the ability of providing high-level representations for uncertain and ambiguous input data [16].So, DBMs become increasingly highly needed for solving object and speech recognition problems.

C. Deep Stacking Networks (DSN)

DSNs also called deep convex networks and had a different architecture from the general approach of deep neural networks [17]. DSN is built to be a deep architecture with serially connected, interfering, and layered modules (subnets). It consists of several small subnets with the only single hidden layer. Each subnet is considered as a neural network with a single hidden layer and two sets of trainable weights.

Each subnet in the DSN comprises three layers. The first linear layer consists of some linear input units equals to the input feature dimensionality. The hidden layer encompasses a set of non-linear sigmoidal units whose number is a disciplined hyper-parameter. Finally, the second linear layer includes a variety of linear output units that are representative of the target classification classes [17].

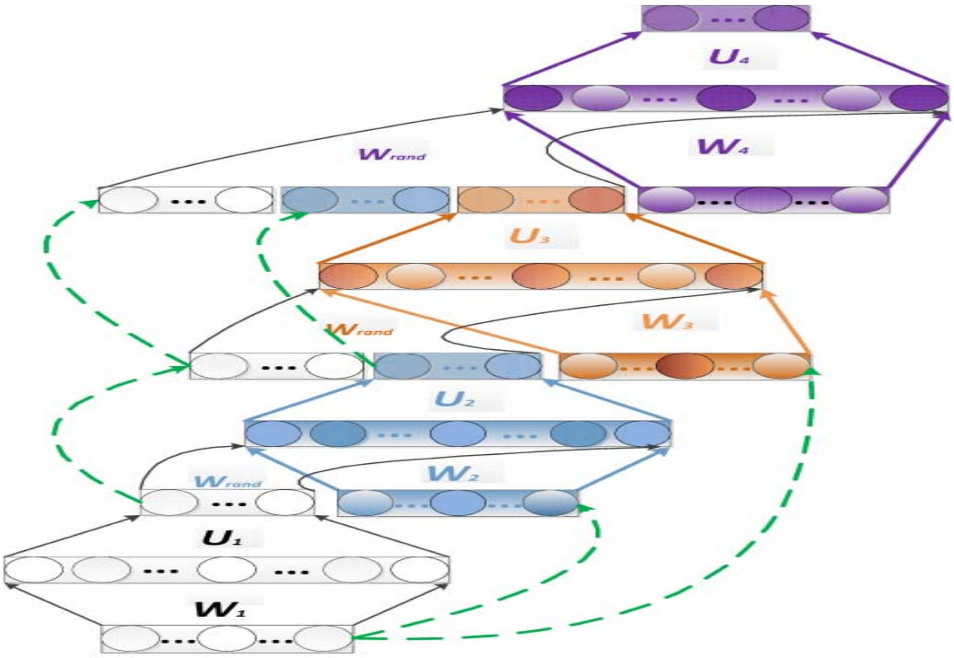


Fig 6: The layout of a DSN with four subnets [17].

The linearity of the output layer in each subnet enables extremely well-organized, parallelizable, and closed-form approximation for the whole network. Learning phase in DSN is estimated in a batch mode manner in which each subnet is trained and fine-tuned independently in parallel. The key idea of batch mode algorithms is to write the upper layers weights of the single hidden-layer neural network as a deterministic function of the lower layers weights [18].

The implementation of learning algorithms for the DSN does not require GPU units unlike most other deep models [19].DSN attacks scalability problem over a vast amount of training data and allows parallel training over many machines. So, DCN can perform information processing tasks successfully especially for large scale data.

III. DEEP LEARNING FOR BIG DATA ANALYTICS

Every day, real world collects massive amounts of real-time labeled and unlabeled data coming from multiple resources [6]. For example, the existing technology based companies, such as Yahoo, Google, and Microsoft, have collected Exabyte of data or larger. In addition, social media communities, such as YouTube, Facebook, and Twitter, have millions of users that permanently create massive amounts of data [8]. These data overpass the typical computing, processing, and storage capacity of traditional database systems and data analysis techniques. Therefore, the term of data is modified to Big Data, which gets big prospects and transformative potential for various sectors.

Big Data poses several challenges that stand as a hinder for Big Data analytics. These challenges can be summarized in three V's: volume, variety, and velocity [20] [21].

- Data volume: The data size that is being processed cannot be limited, but the speed of processing is constant. Analyzing and manipulating big amounts of data entail new resources that can materialize and display the requested results.
- Data variety: Data comes in increasingly diverse and complex formats from a variety sources and probably with different distributions. The challenge is how to combine data that is dissimilar in source or structure and do it at a reasonable cost.
- Data velocity: Fast coming off real-time data is a big challenge. Transfer rates can be limited, but requests are unlimited. The traditional systems are not capable enough of performing the analytics on the data, which is constantly in motion.

Now, performing parallel and distributed data processing, high-dimensionality and data reduction, integrating heterogeneous data, tracking and analyzing data provenance and real-time analysis, and decision-making become fundamental problem areas in Big Data analytics [8]. So, there is a need for a cohesive set of new analytical tools and techniques to provide solutions for these areas.

In machine learning community, Deep Learning can be used as a new tool that helps in Big Data analytics [7]. Deep Learning algorithms are exposed to do well compared to relatively shallow learning architectures at extracting global and non-local patterns and relationships in the data. The extracted representations by Deep Learning can be reflected as a real source of knowledge for decision-making, information retrieval, semantic indexing, and for other purposes in Big Data analytics [8].

Deep Learning algorithms and architectures are more aptly suitable to address issues related to volume and variety of Big Data. Deep Learning can deal with massive amounts of labeled and unlabeled input data [8]. In addition, it can handle its incompleteness and noise properties and provide high-level representations for it. Deep Learning can learn factors of data variation and make effective integrating of different data formats. It also can provide online learning, learning one instance at a time for handling fast coming of data streams. All these capabilities are an aid to make Deep Learning an effective Big Data analytic tool.

IV. DEEP LEARNING APPLICATIONS for BIG DATA ANALYTICS

Deep Learning algorithms extract significant abstract representations of the raw data by using a hierarchical multi-level learning approach. The extracted representations can be reflected as a real source of knowledge for tasks of Big Data analytics, such as data tagging and recognition, information retrieval, and natural language processing.

A. Object Recognition

Computer Vision is the art of making useful decisions for the real physical objects and scenes based on images [22]. Object recognition, 3D-modeling, medical imaging, and smart cars are all examples of what current computer vision systems can do. A fundamental challenge of large scale object recognition is how to attain proficiency in both feature extraction and classifier training without conceding performance [23]. It is found that feature detection by using deep networks is more powerful in performing object recognition tasks [14].

Nair and Hinton [24] presented a third-order Boltzmann Machine (BM) as a new type of top-level Deep Belief Network (DBN) model for 3D objects recognition tasks. A hybrid training algorithm is used which incorporates both generative and discriminative gradients. Generative training makes more accurate object recognition and extracts more abstract image representation and discriminative training provides better classification accuracy. This model is applied to NORB database (normalized-uniform version), which holds stereo-pair images of objects in dissimilar lighting conditions and viewpoints. The error rate reached to 6.5%, which is less than other state-of-the-art error rates. So, they proved that DBNs extraordinarily outperforms shallow models, such as Supper Vector Machines (SVM). However, third-order BM needed to be more factorized with the purpose of making the top-level features can be shared across classes.

Zhou et al. [25] addressed the problem of making image classification for large variance datasets with the existence of only limited labeled data. A Discriminative DBN (DDBN) is presented as a novel semi-supervised learning algorithm to solve this problem, which is built by using a set of RBMs. In the learning phase, the greedy layer-wise unsupervised learning algorithm is applied to the network using the limited labeled data with plenty unlabeled data. In fine tuning phase, gradient descent based supervised learning algorithm is applied to the whole network by using an exponential loss function for maximizing the existence of the labeled data. The performance of DDBN is demonstrated on MNIST and Caltech 101 standard artificial datasets. Results showed that DDBN achieves less error rates compared with typical classifiers.

Krizhevsky et al. [26] trained one of the largest Deep Convolutional Neural Networks (DCNN) to classify ImageNet LSVRC-2010 contest which comprises 1.2 million high-resolution images belonging to 1000 different image classes. This large DCNN consists of 650,000 neurons with 60 million parameters and eight layers. Five of layers are convolutional which may be followed by max-pooling layers and the remaining three are fully connected with a final 1000-way softmax. To speed up the training process a rectified linear units with a very efficient GPU implementation are used .After pre-training, ‘dropout’, regularization method is applied to prevent over-fitting in the fully-connected layers. On the test set, results showed that the error rates of the large DCNN model significantly lower than the previous state-of-the-art .But the network’s performance is directly proportional with number of convolutional layer, thus led to complex computations.

Frome et al. [27] proposed A Deep Visual-Semantic Embedding model (DeViSE) model for overcoming the weaknesses of modern visual recognition systems that can be summarized in the difficulty in dealing with large scale images with only limited training data. (DeViSE) is trained by asynchronous stochastic gradient descent algorithm and worked with not only the labeled images but also with a relatively independent and large dataset of semantic information from un-annotated text data. So, the semantic relationships between labels can be learned easily and images can be mapped obviously into a rich semantic embedding space with fewer limitations. This model is applied to the 1000-class ImageNet dataset and results showed that the semantic information aided in making better predictions about tens of thousands of image labels that not observed during training.

B. Speech & Audio Processing and Tagging

One of the state-of-the-art techniques for Speech Recognition is the Hidden Markov Models (HMMs) in which the observation probabilities are exhibited using Gaussian mixture models (GMMs). However, GMM-HMMs models suffered from independent unrealistic assumptions. They have limited representational capacity of their hidden states [28]. Several research groups proved that DNNs can outperform previous state-of-the-art models for speech recognition on a variety of large datasets [14].

Mohamed et al. [28] proposed an acoustic model for large scale phone recognition by using two different types of DBNs. These models are the back propagation DBN (BP-DBN) architecture and the associative memory DBN

(AM-DBN) architecture. Learning phase in both types is followed by a greedy layer-wise contrastive divergence algorithm. After pre-training the network, a final layer is added to the end of the network by the first architecture (BP-DBN) for adapting the desired outputs and using back-propagation to perform discriminative fine-tuning. The second architecture (AM-DBN) is used to model the joint density of the labels and inputs by applying an RBM associative memory for the final layer. During the fine-tuning stage, only the discriminative component of the weight updates is propagated back through the earlier layers in the network. DBN-HMM is applied with a variety of choices of the number of hidden layers and the number of units per layer to the TIMIT core test set. The results reported that DBNs regularly outperformed other techniques and decreased phone error rates. This work was improved in a later work by using the conditional random field (CRF) instead of the HMMs [1].

Deng and Yu [17] presented a new DSN architecture for addressing the challenge of scalability in large vocabularies speech recognition. This DSN consists of a set of overlapping, serially connected, and layered modules. Learning phase in DCN is convex within each module. In addition, it is batch-mode based instead of stochastic to enable parallel training that can be distributed over many machines. Then, fine tuning is restricted within each module, rather than totally across all layers. DCN-HMM is applied to both MNIST and TIMIT task. The experimental results proved that DCN confirmed superior performance over DBN. This model showed that DCN was not only scalable in training and computations but also accurate in classification tasks. However, the DSN architecture is needed to become flexible.

Deng et al. [28] proposed an acoustic model for speech recognition using a DBN. This network consists of eight hidden layers, and each layer has a few thousand hidden units. Learning phase in this network aimed to model the statistical structure of the input data patterns. It is followed by greedy layer-wise unsupervised learning algorithm, which learns one hidden layer of binary stochastic features at a time. After pre-training, discriminative back-propagation algorithm is used to calculate slopes and stochastic gradient descent with momentum to perform the final fine-tuning in the whole network. DNN-HMM is applied to TIMIT database, and results reflect the capabilities of DNN approach in reducing error patterns. This model varies in several ways from previous attempts to use neural networks for acoustic modeling but is deeper and takes a long processing time.

Kang et al. [29] made the first attempt in using DBN for speech synthesize. DBN is utilized in modeling and generating the speech parameters with spectrum and the fundamental frequency (F0) simultaneously. A fixed number of frames are sampled equally within the defined syllable boundary with high resolution to represent the time dynamics of each syllable rather than using a variable sequence of states. The continuous spectrum and the multi-space F0 pattern are modeled concurrently instead of only the continuous spectrum. They showed that DBN models spectrum outperformed and achieved high performance comparable to context-independent HMM synthesis. However, F0 contour with higher resolution of frame sampling requires better modeling.

C. Information Retrieval

Information retrieval is the process of providing users with the multimedia objects that will satisfy their information need. Performing fast retrieval of information becomes a problem area in Big Data as there are massive amounts of data, such as text, audio, image, and video. These multimedia files are collected and readily available across several domains. Deep networks are mainly utilized in information retrieval for extracting semantically meaningful features for subsequent object ranking stages [14].

Hinton and Salakhutdinov [30] described a generative Deep Learning model for the purpose of making fast information retrieval by learning the binary codes of documents. In this model, a DBN is constructed in which the lowest layer denotes the word-count vector of the document and the top layer denotes the binary code for that document. The documents' binary codes that are semantically similar are confirmed to place closer in the Hamming space. They found that binary codes need slight storage space, and relatively allow quicker searches. They achieved more perfect and faster retrieval than latent semantic analysis technique. However this model deals with short binary codes not very large and very sparse codes.

Wan et al. [31] addressed the ability of Deep Learning in overpassing the semantic gap in content-based image retrieval (CBIR) system. It revolves in high-level semantic concepts perceived by human and low-level image pixels captured by machines. They investigated a DNN model for learning representations of feature and similarity measures. This model is applied to Caltech256 dataset. The results proved that it can capture high semantic information in the raw pixels and Deep Learning feature representations constantly outperform traditional hand-crafted features on all datasets. However, this is just a beginning for Deep Learning with application to CBIR, and there are still many open challenges.

D. Natural Language Processing

The state-of-the-art techniques for language modeling were the N-gram language models. However, data sparseness is one of these models' problems, even with extremely small or very large training corpora, probabilities be assigned to many valid word sequences [31]. Without time-intensive feature engineering or external hand-designed resources, Deep Learning methods can perform natural language processing tasks [14]. Recently, Deep Learning based systems have recorded unprecedented successes on these tasks, such as machine translation, named entity recognition, part-of-speech tagging, sentiment analysis, and paraphrase detection.

Arısoy et al. [32] explored the success of DNN in performing language modeling tasks. A feed-forward DNN is constructed and used for capturing higher-level discriminative information about input features. Their model is trained on the baseline language model training text. It is applied to a Wall Street Journal (WSJ) task. Results showed that DNN Languages models outperformed a single hidden layer NN language models and offer improvements over N-gram language models. However, training strategies for DNN LM have to be more investigated.

Collobert and Weston [33] developed a convolutional DBN as a language model to separately solve some standard problems, such as named entity tagging, part-of-speech tagging, chunking, similar word identification, and semantic role identification. They trained the model to active these tasks simultaneously by using weight-sharing as an instance of multitask learning. The learning phase is not totally unsupervised or totally supervised. A novel form of semi-supervised learning for the shared tasks is used. This model achieved high performance on performing these tasks compared to other state-of-the-art performance. However, they used non-gold standard annotations in their model setup. However, they could not evaluate named entity tagging error rates.

V. DEEP LEARNING CHALLENGES in BIG DATA ANALYTICS

There are several areas of Big Data where Deep Learning needs extra exploration to cover them. Deep Learning achieves limited progress in Learning with streaming data, distributed computing, and dealing with high-dimensional data.

A. Real-time Non-stationary Data

Real-time data refers that there is a high-speed generation of data, which needed to be processed promptly. Non-stationary data means that the distribution of data is varying over time. Providing efficient analysis for these data is profitable in monitoring tasks, such as fraud detection [8]. Real-time nonstationary data are frequently collected and present a challenging area in Big Data analytics. Finding an ideal feature set size becomes difficult for large-scale online datasets whose distribution may change over time.

Online incremental learning is considered a solution for learning from such data [7]. Incremental feature learning offers an effective manner to learn from large datasets through starting with a slight set of initial features [34]. The idea of such learning is to add new feature mappings to the existing feature set and then merge them redundantly. In recent years, only limited progress has been made on deep online learning, it is important to acclimate Deep Learning algorithms to be able to handle large scale of online, real-time data streams.

B. Data parallelism

Big data often retains large scale of inputs, high dimensionality attributes, and great varieties of output. These properties lead to high complexity of running time and proposed models. Deep learning algorithms with a central processor and storage face a challenge in dealing with these properties. Instead, distributed frameworks with parallelized machines are desired.

Popular mini-batch stochastic gradient techniques of Deep Learning are well-known to be difficult to be parallelized over computers. Novel architectures of sound parallel learning algorithms needed to be further developed to make Deep Learning techniques scalable to large scale input data. Parallel implementation utilizes clusters of CPUs or GPUs in increasing speed of training without decreasing the accuracy learning algorithms [35].

C. Multimodal Data

Big Data is usually collected from multi-modalities. Multimodal data reside several input modalities come from different sources. Each modality has a different kind of representation and correlational structure. For example, an image is usually represented by real values of pixel intensities, but the text is usually represented as vectors of discrete sparse word count [36]. It is difficult task to realize the highly non-linear relationships, which exist between low-level features through dissimilar modalities.

Deep Learning is more suitable for heterogeneous data integration due to its capability of learning variation factors of data and providing abstract representations for it. However, its capability is limited to integrate only bi-modalities in which data come from two modalities [7]. Deep Learning also cannot deal with confliction and fusion, which may normally be existed due to the variation of data. There is a need for more progress in Deep Learning algorithms to be optimal models for integrating multi-modal data.

VI. CONCLUSION

Deep Learning algorithms extract significant abstract representations of the raw data by using a hierarchical multi-level learning approach. The extracted representations can be reflected as a real source of knowledge for tasks of Big Data analytics, such as data tagging and recognition, information retrieval and natural language processing. However, there are several areas of Big Data where Deep Learning needs extra exploration to cover them, such as learning with streaming data, distributed computing, and dealing with high-dimensional data. We need to address these technical challenges with new ways of thinking and transformative solutions to realize the full prospective of Big Data.

References

[1] L. Deng, "An Overview of Deep-Structured Learning for Information Processing". In Proceedings of Asian-Pacific Signal and Information Processing Annual Summit and Conference (APSIPAASC), 2011.

[2] H. Lee, R. Gross, R. Ranganath, and A. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations". In ICML, 2009.

[3] R. Socher, Y. Bengio, and C. Manning, "Deep Learning for NLP". Tutorial at ACL,http://www.socher.org/index.php/DeepLearnngTutorial/DeepLearningTutoril, 2012.

[4] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks", The Journal of Machine Learning Research, vol. 10, pp. 1–40, 2009.

[5] B. Tannahill, C. Maute, Y. Yetis, M. Ezell, A. Jaimes, R. Rosas, A. Motaghi, H. Kaplan, and M. Jamshidi, "Modeling of system of systems via data analytics — Case for Big Data in SoS". In System of Systems Engineering International Conference, pp.177, 183, "doi: 10.1109/SYSoSE.2013.6575263", 2013.

[6] S. Sagiroglu, and D. Sinanc, "Big Data: A review". In Collaboration Technologies And Systems (CTS),2013 IEEE International Conference on San Diego,CA: IEEE, 2013.

[7] X. Chen, and X. Lin, "Big Data Deep Learning: Challenges and Perspectives". In Access, IEEE, vol.2, pp.514, 525, doi: 10.1109/ACCESS.2014.2325029, 2014.

[8] M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N, Seliya, R. Wald, and E. Muharemagic, "Deep Learning applications and challenges in Big Data analytics". Journal of Big Data, vol.2, doi: 10.1186/s40537-014-0007-7, 2015.

[9] E. Crego, G. Munoz, and F. Islam, (2016). 'Big Data and Deep Learning: Big deals or big delusions?'[Online] Available at: http://www.huffingtonpost.com/george-munoz frank-islam-and-ed-crego/big-data-anddeep-learnin_b_3325352.html. (Accessed on 2 /1/ 2016).

[10] A. Ng, " Sparse auto encoder", CS294A Lecture notes: Stanford University, 2011.

[11] L. Deng, "A tutorial survey of architectures, algorithms, and applications for Deep Learning", In APSIPA Transactions on Signal and Information Processing, 2013.

[12] G.Hinton, "Training products of experts by minimizing contrastive divergence", Neural Computation, vol.14, pp.1771, 1800, 2002.

[13] P. O'Connor, D. Neil, SC. Liu, T. Delbruck, and M. Pfeiffer, "Real-time classification and sensor fusion with a spiking deep belief network". Frontiers in Neuroscience, vol. 7, doi: 10.3389/fnins.2013.00178, 2013.

[14] L. Deng, and D. Yu, "Deep Learning: Methods and Applications". Foundations and Trends in Signal Processing, vol. 7, no. 3–4, 2014.

[15] G.E. Hinton, S. Osindero, and Y. Teh ," A fast learning algorithm for deep belief nets". Neural Computation, vol. 18, pp. 1527-1554," doi:10.1162/neco.2006.18.7.1527", 2006.

[16] R. Salakhutdinov, and G. Hinton, "A better way to pre train deep Boltzmann machines". Proc, NIPS, 2012.

[17] L. Deng, and D. Yu," Deep Convex Network: A scalable architecture for Speech Pattern Classification". In Proc.Interspeech, 2011.

[18] D. Yu, and L. Deng, "Accelerated parallelizable neural networks learning algorithms for speech recognition". In Proc. Interspeech, 2011.

[19] L. Deng, B. Hutchinson, and D. Yu, " Parallel training of deep stacking networks". In Proc. Interspeech, 2012.

[20] A. Katal , M. Wazid ,and R.H. Goudar ," Big data: Issues, challenges, tools and Good practices". In Contemporary Computing (IC3), Sixth International Conference, pp.404-409, 2013.

[21] A.Adrian,"Big Data Challenges", Database Systems Journal, no. 3, pp. 31–40, 2013.

[22] R. Rao,(2015). " Computer Vision", [online] available at https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect1.pdf.(accessed on 20/1/2016).

[23] Y.Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification: Fast feature extraction and SVM training". In Computer Vision and Pattern Recognition (CVPR),IEEE Conference, pp.1689-1696,doi: 10.1109/CVPR.2011.5995477, 2011.

[24] V. Nair , and G. Hinton," 3-d object recognition with deep belief nets". In Neural Information Processing Systems, vol.22, pp. 1339-1347, 2009.

[25] S. Zhou, Q. Chen, and X. Wang," Discriminative Deep Belief Networks for image classification". In Image Processing (ICIP), 17th IEEE International Conference, pp.1561-1564, doi:10.1109/ICIP.2010.5649922, 2010.

[26] A. Krizhevsky, I. Sutskever , and G. Hinton, "Image Net classification with deep convolutional neural Networks". In Neural Information Processing Systems, pp. 1106–1114, 2012.

[27] A. Frome, G. Corrado, J. Shlens, S. Bengio, J. Dean, M. Ranzato, and T. Mikolov, "Devise: A deep visual-semantic embedding model". Advances in Neural Information Processing Systems, 2013.

[28] A. Mohamed, G. Dahl, and G. Hinton," Deep belief networks for phone recognition". In Proc, NIPS Workshop Deep Learning for Speech Recognition and Related Applications, 2009.

[29] S. Kang, X. Qian, and H. Meng, "Multi-distribution deep belief network for speech synthesis". In Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 8012–8016, 2013.

[30] G. Hinton, and R. Salakhutdinov , "Discovering binary codes for documents by learning deep generative models". Topics in Cognitive Science, pp. 1-18, 2011.

[31] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study". In Proceedings of the 22nd ACM international conference on Multimedia, pp.157-166, doi:10.1145/2647868.2654948, 2014.

[32] E. Arisoy, T. Sainath, B. Kingsbury, and B. Ramabhadran, " Deep neural network language models ". In Proceedings of the Joint Human Language Technology Conference and the North American Chapter of the Association of Computational Linguistics (HLT-NAACL) Workshop, pp. 20-28, 2012.

[33] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning". In Proceedings of the 25th international conference on Machine learning, pp. 160-167, doi: 10.1145/1390156.1390177, 2008.

[34] G. Zhou, K. Sohn, and H. Lee," Online incremental feature learning with denoising autoencoders". In Proceedings of the 15th International Conference on Artificial Intelligence and Statistics, pp 1453–1461, 2012.

[35] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, Q. Le, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, and A. Ng," Large scale distributed deep networks". Advances in Neural Information Processing Systems, Vol. 25, pp. 1232–1240, 2012.

[36] N. Srivastava, and R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines". In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems, vol.25, pages 2231–2239, 2012 .