A Research Survey Report On Deep Learning Concepts

P Venkateshwarlu, Dr. B.Manjula

Abstract: The area of "machine learning" is into its golden era because of its features and aspects due to which "deep learning" has became the ruler in the domain as it uses numerous layers where each layer represents data abstractions required to construct computational models by enabling "deep learning" algorithms that are "generative adversarial networks, conventional neural networks, and model transfers" that entirely customizes our insight of how the data is processed and information is attained as the opening of consideration behind it is extreme because the domain is never previously symbolize multi scope perspective due to be deficient in core understanding that leads to controlling methods such as black box machines that restrain development of basic levels as "deep learning" which is repeatedly perceived as tentative blocks in machine learning. In this paper we present a thorough assessment of past and present state of art that does not retreat in visual or audio and text processing aspects.

Index Terms: Parallel algorithms, Distributed algorithms;, Deep Learning, machine learning theory, Neural networks, Deep Learning Network, Conventional Neural Networks

1. INTRODUCTION

In the present era "machine learning is becoming popular in the research that is incorporated in a maximum number of applications that includes multimedia concepts that extensively use machine-learning algorithms called as deep learning known as representation learning [1] based applications. There is a huge growth in availability of data that depicts remarkable advancement in hardware technologies that leads to newer studies imparted to distributed and deep learning aspects. The inception of deep learning is through the conventional neural networks that tends to outperforms its predecessors by utilizing graph technologies imparted on neurons to create layered learning models that promises results over applications related to Natural Language Processing (NLP), visual data processing, speech and audio processing, and many other well-known applications" [2, 3]. When we specify about the comettence of "machine learning algorithms" which depends on the depiction of the input data as the bad data represents lower performance when compared with good data representation. Hence this leads the feature engineering or the research trends in "machine learning" for an comprehensive time by imparting raw data over many research domains with significant human effort.

"Deep learning" algorithms are used to implement feature extraction in a automated way with a limited field knowledge along with the human struggle [4] as these algorithms comprises of layered architecture of representing data with many features such as data extraction be capable of performing over preceding layers that exists in the networks even when the features are attained from the bottom layers are inspired as "Artificial Intelligence (AI)" over a human brain impact.In a general scenario a human brains will be repeatedly extract data from distinct scenes in distinct ways by getting them from end to end furthers that classify data objects which are main sources of information to ""deep learning"" which further represents the IQ level of human brain. Due to this in almost all areas "deep learning" is considered to be the preferred research area by most of the present day researchers and in this paper we provide a survey report on the "machine learning" aspects over "deep learning".

"Developing a machine that can replicate human brains is a mere dream since ages for many centuries as the deep learning has been initiated in early 300 B.C. when Aristotle proposed associationism which started the history of humans' ambition in trying to understand the brain, since such an idea requires the scientists to understand the mechanism of human recognition systems. The modern history of deep learning started in 1943 when the McCulloch-Pitts (MCP) model was introduced and became known as the prototype of artificial neural models [5]. They created a computer model based on the neural networks functionally mimicking neo cortex in human brains [6]. The combination of the algorithms and mathematics called threshold logic was used in their model to mimic the human thought process but not to learn. Since then deep learning has evolved steadily with a few significant milestones in its development".

1 DEEP LEARNING NETWORKS

Many popular "deep learning" networks exists as represented in below table, Table 1,that comprises of key points and networks information, due to immense research in this area many new networks and their architectures appear on a weekly basis.

2.1 Recursive Neural Network (RvNN) [15]

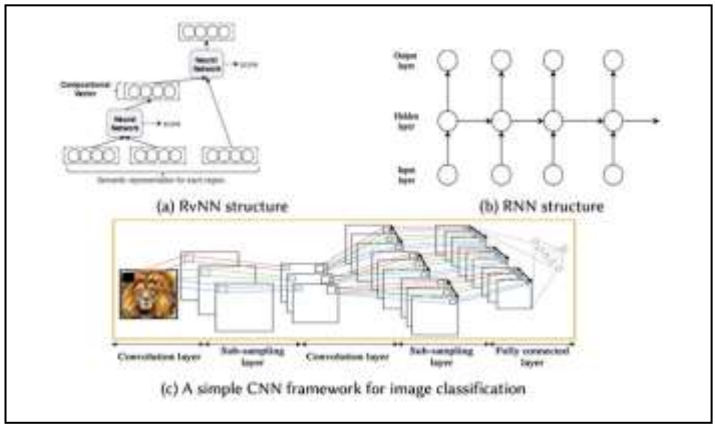
- P. Venkateshwarlu, Research Scholar, Dept. of CS, Kakatiya University, Warangal, Telangana, India. Email: venkateshwarlupurumala@gmail.com
- Dr. B. Manjula, Assistant Professor, Dept. of CS, Kakatiya University, Warangal, Telangana, India. Email: manjulabairam@kakatiya.ac.in

TABLE 1

"DEEP LEARNING NETWORK REPRESENTATION"

DEEP LEARNING NETWORK	DESCRIPTION	REFERENCE
CONVENTIONAL NEURAL NETWORK (CNN)	"CONVOLUTIONAL NETWORKS FOR IMAGES, SPEECH, AND TIME SERIES"	YANN LECUN, ET AL. [7], ALEX KRIZHEVSKY, ET AL.[8]
DEEP BELIEF NETWORKS (DBN)	"DEEP NEURAL NETWORKS FOR ACOUSTIC MODELING IN MONAURAL SPEECH RECOGNITION WHERE THE SHARED VIEWS OF RESEARCH GROUPS"	PO-SEN HUANG, ET AL [9, 10]
DEEP BOLTZMANN MACHINE (DBM)	"AN EFFICIENT LEARNING PROCEDURE FOR DEEP BOLTZMANN MACHINES AND STATISTICS"	RUSLAN SALAKHUTDINOV, ET AL. [11, 12]
RECURRENT NEURAL NETWORKS (RNN)	"CONSTRUCTING LONG SHORT-TERM MEMORY BASED DEEP RECURRENT NEURAL NETWORKS FOR LARGE VOCABULARY SPEECH RECOGNITION USING RNN ENCODER DECODER FOR STATISTICAL MACHINE TRANSLATION"	KYUNGHYUN CHO, ET AL. [13], XIANGANG LI, ET AL. [14]
RECURSIVE NEURAL NETWORKS (RVNN)	"LEARNING TASK DEPENDENT DISTRIBUTED REPRESENTATIONS BY BACK PROPAGATION THROUGH STRUCTURE FOR PARSING NATURAL SCENES AND NATURAL LANGUAGE WITH RECURSIVE NEURAL NETWORKS"	CHRISTOPH GOLLER, ET AL [15], RICHARD SOCHER, ET AL [16]
GENERATIVE ADVERSARIAL NETWORKS (GAN)	"GENERATIVE ADVERSARIAL NETS WITH UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS"	IAN GOODFELLOW, ET AL [17], ALEC RADFORD, ET AL [18]
VARIATIONAL AUTO ENCODER (VAE)	"AUTO-ENCODING VARIATIONAL BAYES WITH UNSUPERVISED LEARNING PROBABILISTIC GRAPHICAL MODEL"	DIEDERIK P. KINGMA, ET AL [19]

The RvNN [15] will perform hierarchical predictions in a constructive manner for performing classification to attain the outputs using compositional vectors as it is mostly inspired by "Recursive Auto associative memory (RAAM)" [15] the architecture that creates process objects in a structured subjective shape similar to trees or graphs where the process will utilize iterative nature of data structures with distinct volume and generates constant width that is distributes the process using "Back propagation Through Structure (BTS) learning scheme" used to train the network [15] using standard back propagation algorithm by supporting tree-like structure by imposing auto association training set to re generate the intended pattern over the input layer with the output layer desired. Socher et al [16] has developed a RvNN architecture that is capable of handling inputs with distinct modalities by implementing results over two examples using RvNN for classification of natural images to generate natural language sentences by separating images into various segments with distinct interest where a sentence is partitioned into well formed words by calculating score of a possible pairs to merge and generated a syntactic tree as the pair with the maximum score is merged into a compositional vectors that represents a region and provides a unique class label for every the new region till we represent for the entire region as shown in figure 1



2.2 Recurrent Neural Network (RNN)

One of the most popular algorithms in "deep learning" is NLP and speech processing called as RNN [13, 14] that utilizes the ordered information in a network due to which many applications used embedded structure in the probable data series to generate desired knowledge with the use of time bound memory units which tend to include x the input layer with s the hidden layer and y the output layer for obtaining an input sequence. In [14] three deep RNN approaches that include deep "Input-to-Hidden," "Hidden-to-Output," and "Hidden-to-Hidden" will lead to propose advantages of deeper RNN that reduces difficult learning process in deep networks by exploding gradients [16] which may decay due to multiplication operations are performed over tiny or maximum derivatives in the training process for reducing the sensitivity over time when compared with the preliminary inputs called as "Long Short-Term Memory (LSTM)" [14] is used to resolve issues by generating memory blocks using recurrent connections by using memory cells to store network temporal states into very deep networks.

2.3 Convolutional Neural Network (CNN)

CNN is one of the most widely used algorithm in "deep learning [7] that uses applications like NLP [21] for speech processing [26] and computer vision [20]" as they are inspired by neurons for simulating the visual cortex in a lion's brain with complex sequences in cells [22] where the major advantage is "parameter sharing for sparse interactions and equivalent representations of cells" with multi dimensional data obtained through a fully connected network. Sampling of layers is performed by pooling various layers that are connected where the input x is denoted into three dimensions: "$m \times m \times r$" where m is denoted for height and width and r represents depth. Kernel is represented by K comprises of various filters: $n \times n \times q$, where q denotes size and n represents a compressed image with weight W and bias b for generating feature map (hk) over a convolational layer is:

$$h^k = f\,(W^k * x + b^k\,). \tag{1}$$

2.4 Deep Generative Networks

“Deep generative networks such as DBN, Deep Boltzmann Machine (DBM), Generative Adversarial Network (GAN), and Variational Auto encoder (VAE) are hybrid probabilistic generative model using RBM over undirected connections are formed by the top two layers where the lower layers are the directed connections that receives inputs from lowest layer which is the visible layer that represents input states as a data vector in an unsupervised approach and in others”. Greedy algorithm is added to improve the generative model to form a “DBN that allows a sub-network to consecutively obtain distinct depictions of data as the initial weights W0 are mapped with the transposed weighted matrix WT0 to generate the maximum level “data” for the proceeding layer to log prospect of every input data vector with less than approximate distribution while appending a new layer into the DBN is improvised for every novel RBM block in the right process”. DBM [24, 25] has the capability to study complex inner depicted as a stronger “deep learning” model for performing object and speech recognition tasks as the approximate reasoning procedure allows handling of duplicate DBM inputs, due to this property a DBM is different from DBN as the complete process is based on directed belief networks rather than undirected called as RBMs. GAN [17, 18] comprises of productive model G and discriminative model D as G captures distribution pg over data at real time t by modeling data m rather than pg over every iteration for performing back propagation to generate maximum sensible data to deceive and randomize the discriminator for identifying deceived data produced by G with a value function V. VAE [19] employees the log likelihood of the data that influences the approach for obtaining minor bound estimator in a directed graphical model with implementation of continuous latent variables with generative parameters θ for implementing “Auto Encoding Variational Bayes (AEVB) algorithm” for optimizing parameters ϕ and θ over the probabilities encoder qϕ (z|x) in a possible neural network with approximation generative model pθ (x,z), where z denotes latent variable with normalization N(0)

2 DEEP LEARNING TECHNIQUES AND FRAMEWORKS

“Deep learning algorithms” allows to improve learning process to simplify calculation process with a longer training time due to which model remains as a key issue with researchers as the classification accuracy of training data and parameters represent techniques and frameworks:

3.1 Unsupervised Transfer Learning

In recent era generative models such as GANs and VAEs are the predominant techniques over unsupervised “deep learning” as GANs are trained to reuse the fixed feature extractor using supervised tasks as the networks are based on CNNs for representing their incomparability as unsupervised learning while performing visual data analysis using sparse auto encoder over very large scale image dataset [25]. The data that is generated as unlabeled over network to extract data that can be further used for face detection by which we can detect high level objects to generate stochastic network based on transition operators called as transfer learning.

3.2 Online Learning

In the present day era stream of data with a huge time complexity is a major concern over various network topologies in “deep learning” with time static or time invariant [26] which is the mainstream research area with the advancement of online “deep learning” as the conventionally DNNs are constructed over “Stochastic Gradient Descent (SGD)” that update individually all the parameters with unique labels for processing streams in a sequential manner or as a batch processing over “Identically Distributed (IID)” with which computing resources and execution time are decreased drastically with high velocity over varying distributions with a certain degree of association with linear learning pace on every input sample.

3.3 Optimized deep learning

Optimization process over DNN is performed by finding the parameters in a network to reduce loss function as most of the fundamental algorithms are imparted over “deep learning” as the process adjust various parameters in a iterative manner

TABLE 2
“DEEP LEARNING FRAMEWORK REPRESENTATION”

Deep Learning Framework	Description	CNN & RNN Support	DBN Support	Reference
TensorFlow	“Large-scale machine learning on heterogeneous distributed systems where the core language used is C++ and interfaces supported are Python & MATLAB”	Yes	Yes	Martín Abadi, et al. [28]
Theano	“A Python framework for fast computation of mathematical expressions with BSD license”	Yes	Yes	Rami Al-Rfou, et al. [29]
MXNet	“A flexible and efficient machine learning library for heterogeneous distributed systems where the core languages used are C++, python, R, Scala, Perl”	Yes	Yes	Tianqi Chen, et al. [30]
Torch	“A Modular machine learning Software Library being implemented in C++ and python”	Yes	Yes	Ronan Collobert, et al. [31]
Neon	“Neon deep learning framework being implemented in Python”	Yes	Yes	Intel Nervana Systems [32]
Caffe	“Convolutional architecture for fast feature embedding being implemented in Python and MATLAB”	Yes	No	Yang qing Jia, et al. [33]

over training sample as the computational complexity of SGD is tiny than the original gradient as the whole dataset is considered when ever parameters are updated. In the learning progression the updating velocity is prohibited by the hyper parameter learning rate as the lower learning rates will eventually lead to a state which is optimal though the data is fluctuated due to decay or loss [27] due to which the idea of momentum is introduced to determine the proper learning rate where the weight decay implements as penalty coefficient in cost function to reduce over fitting and improvisation of performance. The learning rate is further amplified whenever the parameters are updated by recording the generated gradient squares which are always positive.

$$\mathrm{E}[g^2]_t = \beta_2 \mathrm{E}[g^2]_{t-1} + (1-\beta_2)g^2_t \quad (2)$$

where $E[g^2]t$ is the accumulated squared gradient at stage t and g^2t is the squared gradient at stage t which is improved further by adding decay fraction β1 to record the accumulation using Adams[27] l-2 norm is reinstated to make the algorithm stable.

3.4 Distributed System based Deep learning

Competence of training model is accelerated in distributed “deep learning” techniques over training process using data parallelism and model parallelism replicated over computational nodes where model is trained within assigned subset of data following certain period of time to synchronize the nodesWhere as in model parallelism data is practiced with a model in which each node is accountable for executing inference of parameters in the model. “Let Wt,i represents a parameter in neural network node i at a specific time t with slave nodes N used for training with master node”:

$$w_{t+1} = \frac{1}{N}\sum_{i=1}^{N} W_{t+1} \quad (3)$$

Scalability of model parallelism is inferior as the framework takes the embedding represents each operations over different devices when compared with human experts.

4 FRAMEWORKS OF DEEP LEARNING

Table 2 represents “the list of popular deep learning frameworks for implementing architecture designs where the table represents CNN & RNN and DBN frameworks supported are listed”: By observing the Table 2 is “usually implemented using C++ for implementing deep learning frameworks that accelerate the training speed as it uses GPU which is significantly improvise speed up process of matrix evaluation using the interface presented by CuDNN [34] as python has emerging to be a preferable language for implementing deep learning architecture as python is more efficient programming language and simple to implement process due to which the distributed calculation become more easy in some of the latest frameworks like TensorFlow and MXNet tend to improvise the processing speed and efficiency while performing deep learning”. TensorFlow contains the support to be provided to adapt “deep learning Application-Specific Integrated Circuit (ASIC) called Tensor Processing Unit (TPU) to help increase the efficiency and decrease the power consumption”. TensorFlow is instigated as customized “deep learning” process that provides sequence of internal functions to implement any deep neural network oriented static processing graph attainment [28] “Keras started to support Tensorflow via a high level interface that is used to develop architecture without considering internal design as the framework is implemented using parallel and distributed operations with fatal tolerance due to which most of the developers adopted TensorFlow popular deep learning” framework. Theano [29] and Neon [32] are the frameworks that are developed in Python to perform code optimization in the developed system with detailed utilization of kernel level due to which the training speeds typically outperform when compared with existing frameworks as Python extensively support parallelism and multi GPU environment but the major disadvantage in this framework is the multi node calculation is not designed in these framework. MXNet “supports several interfaces, including C++, Python, R, Scala, Perl, MATLAB, Javascript, Go, and Julia [30] as it supports both computation graph declarations and imperative computation while performing architecture design as MXNet extensively supports data and model parallelism with distinct parameter over various server schemes to support distributed calculations with most comprehensive functionality. But the major disadvantage in this frame work is performance is not optimized as that of other existing frameworks”. Torch [31] has its “deep learning” features that are merged with “Facebook’s deep learning CUDA library (fbcunn)” [35] as Torch can operate over model and data level computation over parallel systems due to which it is built on a dynamic graph denote instead of a static graph as to be a dynamic graph that allows us to update the computational graph at runtime by defining functions to generate advanced graphs. Due to all of these advantages Torch is considered to be the most utilized framework. Caffe framework is implemented using the Berkeley Vision and Learning Center due to which it is considered to be most extensively used framework [33] as the most extensively used layers for CNN and RNN and the disadvantage of this framework is it doesn’t utilize DBN framework. The main advantage of Caffe is the structure of computation graphs that are based on convolutional layers as pre-trained models in neural networks. Another limitation of Caffe framework is it is single-machine framework as it cannot support execution in multimode but the exception is while executing multi-GPU calculations.

3 DEEP LEARNING APPLICATIONS

“Deep learning” applications are implemented using NLP Natural Language processing where data is processed using visual tools and speech is used for performing audio processing and many other application make use of social network to analyze social impact and health analysis where each application uses its own tools and methodology.

5.1 Natural Language Processing (NLP)

NLP is a collection of techniques and algorithms that are used to train the computer machines for performing various tasks by using human language as input where the process includes various phases such as “document classification, translation, paraphrase identification, text similarity, summarization, and question answering” as shown in Table 3.

NLP process is considerably complex and with ambiguous structure and highly context specific where a change in single word will lead to change in the whole context. Where the NLP follows steps involved 1) division of input text into words using

TABLE 3

"DEEP LEARNING APPLICATIONS REPRESENTATION"

DEEP LEARNING APPLICATIONS	DESCRIPTION	NLP SUPPORT	REFERENCE
SENTIMENTAL ANALYSIS	"CONVOLUTIONAL NEURAL NETWORKS FOR SENTENCE CLASSIFICATION USED TO PERFORM SENTIMENTAL ANALYSIS AND GENERAL CLASSIFICATION ARE IMPOSED"	YES	YOON KIM, ET AL. [36]
TRANSLATION	"NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATION."	YES	DZMITRY BAHDANAU, ET AL. [37]
PARA PHRASE IDENTIFICATION	"DYNAMIC POOLING AND UNFOLDING RECURSIVE AUTO ENCODERS FOR PARAPHRASE DETECTION"	YES	RICHARD SOCHER, ET AL. [38]
SUMMARIZATION	"EXTRACTIVE SUMMARIZATION USING CONTINUOUS VECTOR SPACE MODELS"	YES	MIKAEL KÅGEBÄCK, ET AL. [39]
QUESTION & ANSWER	"QUESTION ANSWERING OVER FREEBASE WITH MULTI-COLUMN CONVOLUTIONAL NEURAL NETWORKS"	YES	LI DONG, ET AL. [40]

tokenization process (2) reproduction of words into vectors or n-grams and the major issue in this process is to calculate word length.

Sentiment Analysis is a branch of NLP to perform text classification based on the inputs given by writer as the sentiment analysis are represented with the natural phrases such as positive or negative by eliminating classifications related to subjectivity methods as the "Recursive Neural Tensor Network (RNTN)" [41] represents word vectors and parses by constructing a tree of phrases with captured interactions between various elements in a recursive manner to attain sentence level classifications called grammar. Machine Translation is performed in "deep learning" by improvising conventional automatic conversion methods that are suggested by Cho et al. [13] used RNN based encoding and decoding architectures over a "Neural Machine Translation (NMT)" with "RNN Encoder Decoder frameworks" used to map input sequences into fixed length vectors. Bahdanau et al. [37] implemented dynamic-length vector that translates text using translation procedures using binary search operation as a predictive translation process which is computationally expensive and inefficient while handling rare words. "Google's Neural Machine Translation (GNMT)" [42] proposed character level models as it is a deep LSTM network with eight encoder decoder layers connected with attention based mechanism.

Paraphrase identification analyzes two sentences and projecting based on the similarity in their fundamental hidden semantics as one of the key feature that is advantageous over numerous NLP jobs like "plagiarism detection, answers to questions, context detection, summarization, and domain identification". Socher et al. [38] implemented the use of unfolding "Recursive Auto encoders (RAEs)" for measuring the comparison of two sentences. By using syntactic trees that are used to develop the feature space by measuring both phrase and word level matches though it is similar to RvNN as a whole and RAE plays a major key role in implementing unsupervised classification by computing and reconstruction error instead of generating the supervised score while performing merge operation over two vectors yielding a compositional vector. This paper also introduces the dynamic pooling layer which is used to balance and categorize two sentences of distinct sizes as either a paragraph or any other such.

5.2 Visual Data Processing

CNN techniques comprises of issue handling techniques such as segmentation of images by performing classification leads to attract most of the data mining and "machine learning" researchers groups where the major research is performed on computer vision communities AlexNet [20] comprises of image classification results over a very large dataset with the GPU implementation using augmentation and dropout techniques to decrease over fitting problems. VGGNet [43] proposed a 19 layer CNN methods with the spatial size as input to reduce the depth of network is achieved by increasing the achieves with 7.4% top five error rate using simplicity and depth. Microsoft deep residual network (ResNet) [44] proposed the process by including ILSVRC and COCO segmentation and detection methods with residual connections attained 4% top five error rate by using vanishing gradients used to resolve deprivation issue for generating saturated accuracy in deep networks as ResNeXT [45] proposed the original version called (ResNet) which significantly utilizes half of layers of ImageNet dataset for performing image categorization over a definite period of time by utilizing supervised image classification techniques that exists. Object Detection and Semantic Segmentation in complex systems with many lower level features for performing object detection over a Region-based CNN (R-CNN) [46] that performs object detection using image classification over a selected region by taking a large dataset of small objects with labeled data to train a large data sets over CNN networks and ultra-deep networks YOLO (You Only Look Once) [47] is a online image detection technique that implements bounding box detection using 45 frames per second for comparing the existing real time systems as they fully utilizes convolutional networks which shares techniques such as object detecton and achieves used to speed up the process. Single-Shot MultiBox Detector (SSD) [48] uses YOLO as its performance is accurate over region-based techniques for generating set of fixed sized bounding boxes with corresponding object scores at pixel level. Video Processing is considered to be a challenging task because the process includes spatial and temporal data over the CNN model [49] with multi-resolution architectures with local motion information along with context stream implemented over low-resolution image modeling techniques. Recurrent Convolution Networks (RCNs) [50] proposed video processing techniques using CNNs over video frames for imparting visual feeds their frames with transitional layers of CNNs with gated iterative unit based datasets like "YouTube2Text datasets". Visual Datasets fully depends over the improvisation of novel learning algorithms that make use of powerful hardware systems for processing very large scale datasets to train "deep learning" algorithms by considering influential datasets.

4 CHALLENGES OF DEEP LEARNING

Most of the domains are yet to be researched due to its challenging nature that lacks data which is present in general public that creates significant opportunities for performing future research like the lingering black box perception over DNNs to perform decisions without analyzing the domain knowledge [51] specially when data is generated without physical manifestation by mapping layers of a neural networks with yeast cell of DNA attained through microscopic nucleotides as the process takes instructions from the DNA to generate proteins due to which DNA is updated. Google Brain [52] is a unique technique that implements the synthetic brain of DNN called “inceptionism” where each neuron’s estimate values that are grouped with technique called the “deep dream” used to map network’s generated response. Manning et al. [53] represents similar methods with semantic dataset by comprises of distinct network paths that are activated by various data parts that are largely attributed by various statisticians and “machine learning” professionals using “deep learning” to relate neural networks with physical or biological phenomenon to develop metaphysical relationships with DNN brain for simplifying interfaces with low processing overheads. The major issue in “machine learning” is that training samples are not sufficiently available with labels [54] as the data in the present era is ranging from zetta bytes to peta bytes of data being generated on hourly basis with a huge exponential growth due to which the aspect of labeled data is a issue need to be resolved by implementing supervised learning using sentimental analysis by dividing huge data sets into smaller ones. Due to huge increase in size and complexity of data unsupervised learning is a predominant solution with the issues such as data scarcity and cleaning of data is another issue as we have to clean the data based on observations rather than any approximated values which leads to impart “deep learning” methods. Maryam M [55] implemented their methodology with 80 million low resolution images and executed queries by reducing noisy labels and increasing total number of applications with streaming live formats such as time series with social networks.

5 CONCLUSION

“Deep learning” in the present era is the most renown topic in “machine learning” defined as a various layers that implement nonlinear processing with the existence of multiple levels of data that is discovered with distinct patterns as the data is represented in the form of raw data. “Machine learning and data mining techniques” tend to generate knowledge at a higher level of data that represents in the form of streams of raw data over maximum real world applications. In this paper we have reviewed and presented optimization techniques with popular frameworks in this area which is a major challenge to perform, we have take 55 papers or research articles to illustrate the existing solutions and show insight on challenges by considering maximum issues in the present era.

REFERENCES

[1] Li Deng. 2014. A tutorial survey of architectures, algorithms, and applications for deep learning.APSIPA Transactions on Signal and Information Processing3 (2014), 1–29.

[2] Yilin Yan, Min Chen, Saad Sadiq, and Mei-Ling Shyu. 2017. Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters. International Journal of Multimedia Data Engineering and Management8, 1 (2017), 1–20.

[3] Yilin Yan, Min Chen, Mei-Ling Shyu, and Shu-Ching Chen. 2015. Deep learning for imbalanced multimedia data classification. In The IEEE International Symposium on Multimedia. IEEE, 483–488.

[4] Maryam M. Najafabadi, Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and EdinMuharemagic. 2015. Deep learning applications and challenges in big data analytics.Journal of Big Data2, 1 (2015), 1–21.

[5] Warren S. McCulloch and Walter Pitts. 1943. A logical calculus of the ideas immanent in nervous activity.Bulletin of Mathematical Biophysics5, 4 (1943), 115–133.

[6] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview.Neural Networks61 (2015), 85–117.

[7] Yann LeCun and Yoshua Bengio. 1995. Convolutional networks for images, speech, and time series. Handbook of Brain Theory and Neural Networks3361, 10 (1995), 255–257.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. InAdvances in Neural Information Processing Systems 25, F.Pereira, C.J.C.Burges, L.Bottou, and K. Q. Weinberger (Eds.). Curran Associates, 1097–1105.

[9] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry P. Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. InThe 22nd ACM International Conference on Information and Knowledge Management. ACM, 2333–2338.

[10] Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis. 2014. Deep learning for monaural speech separation. InIEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 1562–1566.

[11] Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Deep Boltzmann machines. InArtificial Intelligence and Statistics. PMLR, 448–455.

[12] Ruslan Salakhutdinov and Geoffrey Hinton. 2012. An efficient learning procedure for deep Boltzmann machines. Neural Computation24, 8 (2012), 1967–2006.

[13] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. InThe Conference on Empirical Methods in Natural Language Processing. 1724–1734.

[14] Xiangang Li and Xihong Wu. 2015. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. InIEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 4520–4524.

[15] Christoph Goller and Andreas Kuchler. 1996. Learning task-dependent distributed representations by backpropagation through structure. InIEEE International Conference on Neural Networks, Vol. 1. IEEE, 347–352.

[16] Richard Socher, Cliff C. Lin, Chris Manning, and Andrew Y. Ng. 2011. Parsing natural scenes and natural language with recursive neural networks. InInternational Conference on Machine Learning. Omnipress, 129–136.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. InAdvances in Neural Information Processing Systems. Curran Associates, 2672–2680.

[18] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks.CoRRabs/1511.06434 (2015).

Retrieved from http://arxiv.org/abs/1511.06434.
[19] Diederik P. Kingma and Max Welling. 2013. Auto-encoding variational bayes.CoRRabs/1312.6114 (2013). Retrieved from http://arxiv.org/abs/1312.6114.
[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. InAdvances in Neural Information Processing Systems 25,F.Pereira,C.J.C.Burges,L.Bottou,and K. Q. Weinberger (Eds.). Curran Associates, 1097–1105.
[21] Zhiwei Zhao and Youzheng Wu. 2016. Attention-based convolutional neural networks for sentence classification. InThe 17th Annual Conference of the International Speech Communication Association. ISCA, 705–709.
[22] David H. Hubel and Torsten N. Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.Journal of Physiology160, 1 (1962), 106–154.
[23] Ruslan Salakhutdinov and Geoffrey Hinton. 2012. An efficient learning procedure for deep Boltzmann machines. Neural Computation24, 8 (2012), 1967–2006.
[24] Dominik Scherer, Andreas Müller, and Sven Behnke. 2010. Evaluation of pooling operations in convolutional architectures for object recognition.International Conference on Artificial Neural Networks6354 (2010), 92–101.
[25] Quoc V. Le. 2013. Building high-level features using large scale unsupervised learning. InIEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 8595–8598.
[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning.Nature521, 7553 (2015), 436–444.
[27] Samira Pouyanfar and Shu-Ching Chen. 2017. T-LRA: Trend-based learning rate annealing for deep neural networks. In The 3rd IEEE International Conference on Multimedia Big Data. IEEE, 50–57.
[28] Martín Abadi, Ashish Agarwal, Paul Barham, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems.CoRRabs/1603.04467 (2016). Retrieved fromhttp://arxiv.org/abs/1603.04467.
[29] Rami Al-Rfou, Guillaume Alain, Ying Zhang. 2016. Theano: A Python framework for fast computation of mathematical expressions.CoRRabs/1605.02688(2016). Retrieved from http://arxiv.org/abs/1605.02688.
[30] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang. 2015. MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems. CoRRabs/1512.01274 (2015). Retrieved from http://arxiv.org/abs/1512.01274.
[31] Ronan Collobert, Samy Bengio, and Johnny Mariéthoz. 2002.Torch: A Modular Machine Learning Software Library. Idiap-RR Idiap-RR-46-2002. Idiap.
[32] Intel Nervana Systems. 2017. Neon deep learning framework. Retrieved from https://www.nervanasys.com/ technology/neon. Accessed April 4, 2017.
[33] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. InACM International Conference on Multimedia. ACM, 675–678.
[34] Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. 2014. cuDNN: Efficient primitives for deep learning.CoRRabs/1410.0759 (2014). Retrieved fromhttp:// arxiv.org/abs/1410.0759.
[35] Nicolas Vasilache, Jeff Johnson, Michaël Mathieu, Soumith Chintala, Serkan Piantino, and Yann LeCun. 2014. Fast convolutional nets with fbfft: A GPU performance evaluation.CoRRabs/1412.7580 (2014). Retrieved fromhttp://arxiv.org/abs/1412.7580.
[36] Yoon Kim. 2014. Convolutional neural networks for sentence classification.CoRRabs/1408.5882 (2014). Retrieved from http://arxiv.org/abs/1408.5882.
[37] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate.CoRRabs/1409.0473 (2014). Retrieved from http://arxiv.org/abs/1409.0473.
[38] Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011. Dynamic pooling and unfolding recursive auto encoders for paraphrase detection. In Advances in Neural Information Processing Systems, Vol. 24. Neural Information Processing Systems Foundation, 801–809.
[39] Mikael Kågebäck, Olof Mogren, Nina Tahmasebi, and Devdatt Dubhashi. 2014. Extractive summarization using continuous vector space models. In2nd Workshop on Continuous Vector Space Models and their Compositionality. Citeseer, Association for Computational Linguistics, 31–39.
[40] Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2015. Question answering over freebase with multi-column convolutional neural networks. In53rd Annual Meeting of the Association for Computational Linguistics, Vol. 1. Association for Computational Linguistics, 260–269.
[41] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. InConference on Empirical Methods in Natural Language Processing. Citeseer, Association for Computational Linguistics, 1631–1642.
[42] Yonghui Wu, Mike Schuster, Zhifeng Chen, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. CoRRabs/1609.08144 (2016). arxiv:1609.08144. Retrieved from http://arxiv.org/abs/1609.08144.
[43] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. CoRRabs/1409.1556 (2014). Retrieved fromhttp://arxiv.org/abs/1409.1556.
[44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 770–778.
[45] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2016. Aggregated residual transformations for deep neural networks.CoRRabs/1611.05431 (2016). Retrieved fromhttp://arxiv.org/abs/1611.05431.
[46] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. InIEEE Conference on Computer Vision and Pattern Recognition. IEEE, 580–587.
[47] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 779–788.
[48] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. 2016. SSD: Single shot multibox detector. InEuropean Conference on Computer Vision. Springer, 21–37.

[49] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Largescale video classification with convolutional neural networks. InIEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 1725–1732.
[50] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. InIEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2625–2634.
[51] Hayit Greenspan, Bram van Ginneken, and Ronald M. Summers. 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35, 5 (2016), 1153–1159.
[52] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. 2015. Inceptionism: Going deeper into neural networks. Google Research Blog. Retrieved fromhttps://research.googleblog.com/2015/06/inceptionism-going-deeperinto-neural.html. Accessed March 26, 2018
[53] Christopher Manning. 2016. Understanding human language: Can NLP and deep learning help? InThe 39th International ACM SIGIR Conference on Research and Development in Information Retrieval.ACM,1–1.
[54] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning.Nature521, 7553 (2015), 436–444.
[55] Antonio Torralba, Rob Fergus, and William T. Freeman. 2008. 80 million tiny images: A large data set for nonparametric object and scene recognition.IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 11 (2008), 1958–1970.