Iraqi Journal of Electrical and Electronic Engineering
Volume 14, No.2-2018
المجلة العراقية للهندسة الكهربائية والالكترونية
المجلد 14، العدد 2- 2018

Current Big Data Issues and Their Solutions via Deep Learning: An Overview

Asif Ali Banka*, Roohie Naaz Mir
Department of Computer Science and Engineering, National Institue of Technology, Srinagar, Jammu and Kashmir 190006 India

Abstract: *The advancements in modern day computing and architectures focus on harnessing parallelism and achieve high performance computing resulting in generation of massive amounts of data. The information produced needs to be represented and analyzed to address various challenges in technology and business domains. Radical expansion and integration of digital devices, networking, data storage and computation systems are generating more data than ever. Data sets are massive and complex, hence traditional learning methods fail to rescue the researchers and have in turn resulted in adoption of machine learning techniques to provide possible solutions to mine the information hidden in unseen data. Interestingly, deep learning finds its place in big data applications. One of major advantages of deep learning is that it is not human engineered. In this paper, we look at various machine learning algorithms that have already been applied to big data related problems and have shown promising results. We also look at deep learning as a rescue and solution to big data issues that are not efficiently addressed using traditional methods. Deep learning is finding its place in most applications where we come across critical and dominating 5Vs of big data and is expected to perform better.*

***Index Terms:* Big data, machine learning, deep learning**

I. Introduction

The advancements in modern day computing and architecture focuses on harnessing parallelism and achieve high performance computing resulting in generation of massive amounts of data. The interconnected devices have practically resulted into data intensive computing where data is now essential part of human life. Sharing data, professional or non-professional due to high availability of anywhere-computing and usage of internet oriented application like social media, internet of things, GPS, Cloud computing etc. have resulted in generation of massive amounts of data. Every minute Facebook records more than 3.2 million likes, stores more than 3.4 million posts and generates around 4 GB of data. In March 2013, Facebook launched a graph search feature that enables its users to search the social graph for users with similar locations or hobbies. Also in 1 min, Google answers about 300K searches, 126 h uploaded to YouTube and more than 140K video views, about 700 users created in Twitter and more than 350K tweets generated, and more than 11K searches on LinkedIn performed [1]. The massive data is being produced, consumed and transferred and is expected to increase with more people gaining access to technology.

Computing technology alone is not the

reason for big data explosion. Inter-disciplinary research in areas like astronomy, aviation, particle physics, genome sequences have also invested in projects that result in generation of massive amounts of data. Like Moore's Law, where number of transistor double roughly every two years we have a similar trend in data volumes that are expected to double every 2 years over the next decade. According to IBM, we are currently creating 2.5 quintillion bytes of data every day. IDC predicts that the worldwide volume of data will reach 40 Zetta-bytes by 2020 where 85% of all of this data will be of new data types and formats including server logs and other machine generated data, data from sensors, social media data, and many more other data sources [1]. There are more and more surveys in support of the facts that data generated in near future will be concern of every research domain.

The core challenge in research today is to harvest this data and seek innovations with data analysis and decision-making processes. The ability to analyze and understand large-scale information is current trend for database management organizations with desire to make it as easy as it is for smaller volumes of structured data. The big data term was initially coined and described by 3Vs that were later changed to 4Vs and then 5Vs. These complexities are major attributes that need to be addressed while big data processing. The Volume describing the massive amount of data that can be billions of rows and millions of columns, the Variety representing the variety on formats, data sources, and structures; the Velocity referring to the very high speed on data generation, ingestion, and near real-time analysis; the Veracity refers to the trustworthiness of the data and the Value refers to conclusions drawn from the big data [2, 3, 4]. There is a latency trade-off between state of art computers and mechanical hard drives that hampers the modern day data intensive computing [4]. A new trend should instead focus on supporting cheaper cluster of computers to manage and process all the data instead of focusing on having the biggest and fastest single computer. A need was felt and researchers found it interesting to invest in warehouse scale computing and commodity off-the-shelf high performance computing to support big data processing and analysis.

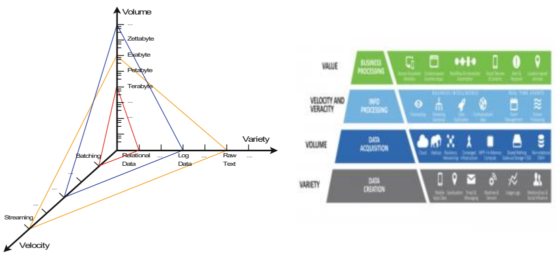

Fig1: Big Data Complexities

Despite of so many confounding factors, big data provides a major break- through for companies such as Facebook, Yahoo, Google who have recently started to harness into its benefits. A general theme in Big Data systems is that massive volumes of the raw data is increasingly diverse and complex, consisting of largely un-categorized/unsupervised data along with, perhaps, a small quantity of categorized/supervised data on top of it Variety among different data representations in a given repository poses unique challenges with Big Data, which requires preprocessing of unstructured data in order to extract structured/ordered representations of the data for human and/or downstream consumption. Real time analyses for data-intensive computations require handling the Velocity (the increasing rate at which data is collected and obtained) which is just as important as the Volume and Variety characteristics of Big Data.

Data loss in time sensitive systems are caused as streaming data is generally not immediately processed and analyzed and a separate conversion into usable information adds on to challenges of handling real time data. Twitter, Yahoo, and IBM have developed applications that address the analysis of streaming data [5]. Veracity brings in challenges in trustworthiness or usefulness of results for decision-making in business domain obtained from data analysis. Besides these major challenges, other factors like data quality and validation, data cleansing, feature engineering, high-dimensionality and data reduction, data representations and distributed data sources, data sampling, scalability of algorithms, data visualization, parallel and distributed data processing, real- time analysis and decision making, crowdsourcing developing new models for massive data computation have to be addressed.

Data sets are massive and complex hence traditional learning methods fail to rescue the researchers, which has resulted in adoption of machine learning techniques to provide possible solutions in order to mine the information hidden in unseen data. Though analyzing massive data is expected to improve quality of technology that will reflect in all domains of human life [5], traditional learning models are not designed to handle large volumes of data. Mostly data would be completely loaded into memory [6], which is a challenge in big data aspect.

II. MACHINE LEARNING AS A TOOL FOR BIG DATA

Machine leaning gives computers the ability to learn without being explicitly programmed. Backed up by ideas from many different kinds of fields such as artificial intelligence, optimization theory, information theory, statistics, many other disciplines of science, engineering, and mathematics [7, 8, 9, 10] relate its implementation in almost every scientific domain, which has revolutionized science, technology and society [11]. Broadly we classify learning as supervised learning, unsupervised learning, and reinforcement learning [12].

Supervised learning is learning from

labeled data, which has inputs and desired outputs. While as unsupervised learning does not require labeled training data. It learns from test data only, without desired output targets. Reinforcement learning is learning from feedback received through interactions with an external environment [13, 14, 15]. Companies like Google for different application domains have rigorously exploited all these categories. Supervised learning and unsupervised learning mainly relate to data analysis and decision-making problems, which are preferably addressed using reinforcement learning.

As the data keeps getting bigger and bigger, existing machine learning techniques encounter great difficulties to handle future data as completely loading data into memory for centralized processing is not possible. Due to failure of traditional algorithms to handle massive data efficiently, more advanced learning methods have been adopted. These broadly focus on the idea of learning and are identified as:

Representation Learning: or feature learning is transformation of raw data input to a representation that can be effectively exploited in machine learning tasks. Representation learning [16, 17], is a promising solution to learn the meaningful and useful representations of the data that make it easier to extract useful information when building classifiers or other predictors. It has achieved impressive performance on high dimensionality reduction tasks [18]. There are mainly three subtopics on representation learning: feature selection, feature extraction, and distance metric learning [18]. It has demonstrated remarkable successes in real-world applications, such as speech recognition, natural language processing, and intelligent vehicle systems [18, 19, 20].

Deep learning: Interestingly, a class of techniques called deep learning finds its place in big data applications. One of advantages of deep learning is that it is not human engineered. Though deep learning is yet to mature, it has attained huge attention from researchers. Supervised and/or unsupervised strategies are used in deep architectures to automatically learn hierarchical representations [21]. Speech recognition, computer vision, language processing, and information retrieval [22, 23, 24, 60] have attracted much attention from the academic community in recent years.

Distributed and parallel learning: unprecedented volumes of data result in a bottleneck, preventing ability of learning algorithms to use all the data within a reasonable time. Allocating the learning process among several work- stations is a natural way of scaling up learning algorithms [25], which sounds promising. With the advantage of distributed computing for managing big volumes of data, distributed learning avoids the necessity of gathering data into a single workstation. Earlier research has resulted in development of several popular distributed machine learning algorithms such as, decision rules [26], stacked generalization [27], meta-learning [28], and distributed boosting. With the advancements in efficacy of multicore processors and cloud computing platforms, parallel and distributed computing systems have become readily available for parallelization [25].

Transfer learning: In order to handle knowledge generated from multiple sources with great heterogeneity and lot of variety, transfer learning has been proposed. Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned [29, 30]. This helps solving new problems faster and intelligently from knowledge learned from previously tasks. It can be categorized into: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning [30]. Many real-world data processing applications, such as cross- domain text classification, constructing informative priors, and large-scale document classification [31, 32, 33] have been addressed using transfer learning.

Active learning: There are real-world applications where we have lot of data available but labels are scarce or expensive to obtain. Learning from unlabeled data is difficult and time-consuming which is addressed by active learning, which selects most critical instances for labelling [34]. This helps in selecting few labels and minimizing cost [35]. Image classification and biological DNA identification have extensively applied these methods [36, 37].

Kernel-based learning: Kernel-based learning has been used over years and has emerged as very powerful technique to increase the computational capability based on a breakthrough in the design of efficient nonlinear learning algorithms [38]. Implicitly mapping samples from the original space into a potentially infinite-dimensional feature space is their elegant property in which inner products can be calculated directly via a kernel function [39]. Kernel functions provide the nonlinear means to address many challenging applications e.g., online classification [40].

III. CURRENT ISSUES IN BIG DATA AND MACHINE LEARNING APPROACHES

Different domains of machine learning can address the 5V challenges of big data:

Volume: It is impossible to train machine-learning algorithm with a central processor and storage on massive datasets. Distributed frameworks with parallel computing are, therefore preferred. Various algorithms working on top of large scale standaradized datasets, various learning have been proposed. MapReduce [41, 42], a powerful programming framework, enables the automatic paralleling and distribution of computation applications on large clusters of commodity machines. To deal with the massive volume challenge of big data, cloud computing assisted learning method has also been adopted. It can enhance computing and storage capacity through cloud infrastructure.

Variety: The variety of data that is generally produced from various sources are of different types. Data with different representation forms - structured, semi-structured, and even entirely unstructured data add to the degree of complexity. Good data representations and integration from each individual data source works as rescue [43]. Besides representation learning, deep learning methods have also been successfully identified for effective data integration Transfer learning is also a good choice to address the challenge of different data types.

Velocity: Handling the data in real time is one of major concerns for data science people. For big data, speed or velocity really matters, which is another emerging challenge for learning. Online learning is one strong solution for learning from high speed of data sources [7, 44, 45, 46]. Extreme learning machine (ELM)[47] is one novel-learning algorithm proposed to speed up learning. ELM provides extremely faster learning speed, better generalization performance, and least human intervention [48]. Representative streaming processing systems and many other architectures have been proposed recently to provide real-time analytics over big data [49].

Veracity: Massive data generation result in questioning the precision and trust of the data source. We have begun to realize importance of addressing and managing the uncertainty and incompleteness on data quality. A simple way to handle data uncertainty is to apply summary statistics such as means and variances to abstract sample distributions [50]. Advanced deep learning methods to handle imperfect, noisy data and tolerate some messiness, have also been proposed.

Value: Extraction of valuable information to analyze big datasets is the core reason for data science. The inferences are not straightforward. Knowledge discovery in databases (KDD) and data mining technologies [51, 52] play a pivotal role to extract information hidden in the massive data. The diversity of data meaning and the context in which data is analyzed provides room for cognitive computing [53].

IV. Deep Learning for Big Data - Revisited

Major breakthroughs have been achieved in the field of artificial intelligence as the concept of deep learning matures over time. Since deep learning is not human engineered it is more suitable to address most of application domains. Deep Learning algorithms learns over the time through hierarchical learning process. Simpler abstractions at preceding levels form the basis for more complex abstractions. Deep learning finds its application in big data analysis because of its ability to handle large amounts of unsupervised and raw data that is mostly not labeled and categorized. Basic problems of finding patterns in large volumes of massive data, tagging relevant data, semantic indexing, retrieval of information in real time are nicely addressed by deep learning algorithms which is known to work well with high-dimensional data, scalable models, and distributed computing. Mehdi Gheisari et al in [59] have highlighted some of application areas of deep learning in big data, which are particularly directed towards volume and variety of data that are inherently suitable for deep learning scenarios.

For unsupervised data, deep learning algorithms with their approach to learn data representations in a greedy layer-wise fashion [54, 55] have been exploited with success. Stacking up non-linear feature extractors in fields of speech recognition [56, 57], computer vision[55, 56, 58], and natural language processing[17] deep learning solutions have yielded outstanding results.

Conventional machine learning and feature learning have not shown significant results with large-scale data whereas deep learning has shown significant refinement in approaches to handle semantic indexing, data tagging, fast information retrieval. But deep Learning has not been able to address problems related to streaming data, high dimensionality of data, distributed and parallel computing. Deep learning has come up as rescue to most big data problems. Handling unsupervised data without human interference has been addressed well by deep learning approaches. The variation in data from various sources is addressed by abstract representations of deep learning. The abstract representations provided by deep learning algorithms can separate the different sources of variations in data. Non-linear transformations are applied at each layer and stacking up these result in more complicated nonlinear transformations, which represent data. These transformations tend to disentangle factors of variations in data [17]. The output of the final layer provides essential information for classification and other applications that designates value of the information. Inherent support to handle volume and variety provide one such solution venue for data analytics experts and practitioners. Big data technology needs to adapt to new incoming stream of an online large-scale data that are not addressed by deep networks because of continuous memory requirement. High-dimensional data adds to the complexity and volume of the raw data. Time sensitive learning using deep learning needs to be addressed.

Large-scale Deep Learning models are quite suited to handle massive volumes of input associated with Big Data, and as demonstrated in the above works they are also better at learning complex data patterns from large volumes of data. Fast inferences, which are one major goal of big data analytics, can be acquired using the deep learning methods besides using less storage capacity. Deep learning models are readily accepted in domains where handling massive amounts of data, feature learning for heterogeneous data, real time and low quality data are considered difficult tasks [61].

V. Conclusions and Future Scope

Determining the context of data from various sources may not be exactly the same, which may significantly impact the quality of the machine learning results. Valuable patterns can be determined based on ontology modeling which help us to be proactive about data that is being generated. Since the ontology and semantic web technologies lack maturity we need to understand that implementing deep learning to process big data will be meaningful. Labelled patterns play a pivotal role for the learning algorithms; but on the other hand, labeling patterns is often expensive in terms of the computation time or cost, particularly for the large-scale streaming data, which is intractable. These two issues can be addressed by deep learning, hence, addressing the issue of pattern training. Finding accurate results by analyzing the personal information is one of major concerns of privacy in big data. Mining has been used for profiling which is conducive to change marketing tactics. Privacy is an issue that makes people feel at risk if effective measures are not taken to secure it. Privacy policies for profiling to encourage security issues for social data can be addressed using deep learning especially when addressing high dimensional data from images and videos. The aim of study is to realize and focus on building the bridge from theory to practice and implementation of deep learning for big data analysis.

VI. References

[1] S. Sakr, F. M. Orakzai, I. Abdelaziz, Z. Khayyat, Large-Scale Graph Processing Using Apache Giraph, Springer, 2017.
[2] E. Dumbill, What is big data (2012).
[3] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, E. Muharemagic, Deep learning applications and challenges in big data analytics, Journal of Big Data 2 (1) (2015) 1.
[4] E. Alpaydin, Introduction to machine learning, MIT press, 2014.
[5] N. R. Council, et al., Frontiers in massive data analysis, National Academies Press, 2013.
[6] K. Slavakis, G. B. Giannakis, G. Mateos, Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge, IEEE Signal Processing Magazine 31 (5) (2014) 18–31.
[7] X.-W. Chen, X. Lin, Big data deep learning: challenges and perspectives, IEEE Access 2 (2014) 514–525.
[8] T. M. Mitchell, Machine learning and data mining, Communications of the ACM 42 (11) (1999) 30–36.

[9] S. Russell, P. Norvig, A. Intelligence, A modern approach, Artificial Intelligence. Prentice-Hall, Egnlewood Cli↲s 25 (1995) 27.

[10] V. Cherkassky, F. M. Mulier, Learning from data: concepts, theory, and methods, John Wiley & Sons, 2007.

[11] T. M. Mitchell, The discipline of machine learning, Vol. 9, Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.

[12] C. Rudin, K. L. Wagsta↲, Machine learning for science and society, Machine Learning 95 (1) (2014) 1–9.

[13] B. Adam, I. F. Smith, Reinforcement learning for structural control, Journal of Computing in Civil Engineering 22 (2) (2008) 133–139.

[14] N. Jones, The learning machines, Nature 505 (7482) (2014) 146.

[15] J. Langford, Tutorial on practical prediction theory for classification, Journal of machine learning research 6 (Mar) (2005) 273–306.

[16] R. Bekkerman, R. El-Yaniv, N. Tishby, Y. Winter, Distributional word clusters vs. words for text categorization, Journal of Machine Learning Research 3 (Mar) (2003) 1183–1208.

[17] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence 35 (8) (2013) 1798–1828.

[18] F. Huang, A. Yates, Biased representation learning for domain adaptation, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, 2012, pp. 1313–1323.

[19] A. Bordes, X. Glorot, J. Weston, Y. Bengio, Joint learning of words and meaning representations for open-text semantic parsing., in: AISTATS, Vol. 22, 2012, pp. 127–135.

[20] N. Boulanger-Lewandowski, Y. Bengio, P. Vincent, Modeling temporal de- pendencies in high-dimensional sequences: Application to polyphonic music generation and transcription, arXiv preprint arXiv:1206.6392.

[21] K. Dwivedi, K. Biswaranjan, A. Sethi, Drowsy driver detection using representation learning, in: Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 995–999.

[22] D. Yu, L. Deng, Deep learning and its applications to signal and information processing [exploratory dsp], IEEE Signal Processing Magazine 28 (1) (2011) 145–154.

[23] G. E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Trans- actions on Audio, Speech, and Language Processing 20 (1) (2012) 30–42.

[24] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., Deep neural networks for acoustic modeling in speech recognition: The shared views

of four research groups, IEEE Signal Processing Magazine 29 (6) (2012) 82–97.

[25] D. C. Cire ̧san, U. Meier, L. M. Gambardella, J. Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition, Neural computation 22 (12) (2010) 3207–3220.

[26] D. Peteiro-Barral, B. Guijarro-Berdiñas, A survey of methods for dis- tributed machine learning, Progress in Artificial Intelligence 2 (1) (2013) 1–11.

[27] H. Chen, T. Li, C. Luo, S.-J. Horng, G. Wang, A rough set-based method for updating decision rules on attribute values coarsening and refining, IEEE Transactions on Knowledge and Data Engineering 26 (12) (2014) 2886–2899.

[28] J. Chen, C. Wang, R. Wang, Using stacked generalization to combine svms in magnitude and shape feature spaces for classification of hyperspectral data, IEEE Transactions on Geoscience and Remote Sensing 47 (7) (2009) 2193–2205.

[29] E. Leyva, A. Gonz ́alez, R. P ́erez, A set of complexity measures designed for applying meta-learning to instance selection, IEEE Transactions on Knowledge and Data Engineering 27 (2) (2015) 354–367.

[30] E. W. Xiang, B. Cao, D. H. Hu, Q. Yang, Bridging domains using world wide knowledge for transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (6) (2010) 770–783.

[31] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (10) (2010) 1345–1359.

[32] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, A survey of machine learning for big data processing, EURASIP Journal on Advances in Signal Processing 2016 (1) (2016) 1–16.

[33] R. Raina, A. Y. Ng, D. Koller, Constructing informative priors using trans- fer learning, in: Proceedings of the 23rd international conference on Ma- chine learning, ACM, 2006, pp. 713–720.

[34] J. Zhang, Deep transfer learning via restricted boltzmann machine for document classification, in: Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on, Vol. 1, IEEE, 2011, pp. 323–326.

[35] Y. Fu, B. Li, X. Zhu, C. Zhang, Active learning without knowing individual instance labels: A pairwise label homogeneity query approach, IEEE Transactions on Knowledge and Data Engineering 26 (4) (2014) 808–822.

[36] B. Settles, Active learning literature survey, University of Wisconsin, Madison 52 (55-66) (2010) 11.

[37] M. M. Haque, L. B. Holder, M. K. Skinner, D. J. Cook, Generalized query- based active learning to identify differentially methylated regions in dna, IEEE/ACM Transactions on Computational Biology and Bioinformatics 10 (3) (2013) 632–644. 14

[38] D. Tuia, M. Volpi, L. Copa, M. Kanevski, J. Munoz-Mari, A survey of active learning algorithms for

supervised remote sensing image classification, IEEE Journal of Selected Topics in Signal Processing 5 (3) (2011) 606–617.

[39] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, Y. Chen, Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions, IEEE Signal Processing Magazine 30 (4) (2013) 126–136.

[40] C. Li, M. Georgiopoulos, G. C. Anagnostopoulos, A unifying framework for typical multitask multiple kernel learning problems, IEEE Transactions on Neural Networks and Learning Systems 25 (7) (2014) 1287–1297.

[41] K. Slavakis, S. Theodoridis, I. Yamada, Online kernel-based classification using adaptive projection algorithms, IEEE Transactions on Signal Pro- cessing 56 (7) (2008) 2781–2796.

[42] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Communications of the ACM 51 (1) (2008) 107–113.

[43] J. Dean, S. Ghemawat, Mapreduce: a flexible data processing tool, Communications of the ACM 53 (1) (2010) 72–77.

[44] S. Shalev-Shwartz, et al., Online learning and online convex optimization, Foundations and Trends R in Machine Learning 4 (2) (2012) 107–194.

[45] J. Wang, P. Zhao, S. C. Hoi, R. Jin, Online feature selection and its applications, IEEE Transactions on Knowledge and Data Engineering 26 (3) (2014) 698–710.

[46] J. Kivinen, A. J. Smola, R. C. Williamson, et al., Online learning with kernels, in: NIPS, 2001, pp. 785–792.

[47] M. Bilenko, S. Basil, M. Sahami, Adaptive product normalization: Using online learning for record linkage in comparison shopping, in: Data Mining, Fifth IEEE International Conference on, IEEE, 2005, pp. 8–pp. 15

[48] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.

[49] W. Yang, X. Liu, L. Zhang, L. T. Yang, Big data real-time processing based on storm, in: Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on, IEEE, 2013, pp. 1784–1787.

[50] R. Ranjan, Streaming big data processing in datacenter clouds, IEEE Cloud Computing 1 (1) (2014) 78–83.

[51] I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, A. Y. Ng, Measuring invariances in deep networks, in: Advances in neural information processing systems, 2009, pp. 646–654.

[52] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases, AI magazine 17 (3) (1996) 37.

[53] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data mining with big data, IEEE transactions on knowledge and data engineering 26 (1) (2014) 97–107.

[54] J. Kelly III, S. Hamm, Smart Machines: IBMÕs Watson and the Era of Cognitive Computing, Columbia University Press, 2013.
[55] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (7) (2006) 1527–1554.
[56] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al., Greedy layer-wise training of deep networks, Advances in neural information processing systems 19 (2007) 153.
[57] G. Dahl, A.-r. Mohamed, G. E. Hinton, et al., Phone recognition with the mean-covariance restricted boltzmann machine, in: Advances in neural information processing systems, 2010, pp. 469–477.
[58] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.
[59] Gheisari, Mehdi, Guojun Wang, and Md Zakirul Alam Bhuiyan. "A survey on deep learning in big data." In Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on, vol. 2, pp. 173-180. IEEE, 2017.
[60] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
[61] Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. Information Fusion, 42, 146-157.