Deep Osmosis: Holistic distributed deep learning in osmotic computing

Ahsan Morshed, Prem Prakash Jayaraman, Timos Sellis, and Dimitrios Georgakopoulos
Swinburne University of Technology

Massimo Villari
University of Messina

Rajiv Ranjan
Newcastle University

Emerging availability (and varying complexity and types) of Internet of Things devices, along with the large data volumes that such devices (can potentially) generate, can have a significant impact on our lives, fueling the development of critical next-generation services and applications in a variety of application domains (e.g., health care, smart grids, finance, disaster management, agriculture, transportation, and water management). Deep learning technology, which has been used successfully in computer vision and language modeling, is finding application in new domains driven by the availability of diverse and large datasets. One such example is the advances in medical diagnostics and prediction that use deep learning technology to improve human health. However, timely and reliable transfer of large data streams (a requirement of deep learning technologies for achieving high accuracy) to centralized locations, such as cloud datacenter environments, is being seen as a key limitation of expanding the application horizons of such technologies. To this end, various paradigms, including osmotic computing, have been proposed that promote distribution of data analysis tasks across cloud and edge computing environments. However, these existing paradigms fail to provide a detailed account of how technologies such as deep learning can be orchestrated and take advantage of the cloud, edge, and mobile edge environments in a holistic manner. This Blue Skies piece analyzes the research challenges involved with developing a class of holistic distributed deep learning algorithms that are resource and data aware and are able to account for underlying heterogeneous data models, resource (cloud vs. edge vs. mobile edge) models, and data availability while executing—trading accuracy for execution time, etc.

ADVANCES in hybrid high-end computing—processing capabilities offered by a combination of central processing unit (CPU) and graphics processing unit (GPU)—and cloud computing have fueled the deployment and adoption of machine learning technology that powers many aspects of modern society, from social media to recommendation systems on websites.[1,2] Deep learning is one such branch of machine learning that "achieves great power and flexibility by learning to represent the knowledge as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones."[3]

Traditionally, deep learning has successfully been used in many application domains, including computer vision (e.g., facial recognition) and language modeling (e.g., speech recognition), and medical image analysis. However, these complex and well-engineered approaches required substantial effort in selecting handcrafted features.[4] The recent advent of technologies such as the Internet of Things (IoT), high-speed communication, and mobile devices with the capability of running machine learning frameworks (e.g., the Apple iPhone Core Machine Learning [Core ML] toolkit) has dramatically increased the availability of different types of medical data. These include electronic health records (EHRs), imaging (e.g., x-ray and ultrasound), sensor data (including that from wearable devices), text data (e.g., doctors' scripts), social media, blogs, online surveys, and traditional repositories. Increased access to such biomedical data underpinned by advances in deep learning technologies (such as the new Inception v3 model based on GoogLeNet) has renewed interest in deep learning in the world of biomedicine by providing solutions and helping researchers analyze medical data to understand, treat, and predict diseases.[4,5]

Deep-learning-based applications (applications that have been developed using deep learning technologies) have been highly reliant on the availability of hybrid high-end machines with an array of GPUs. Cloud computing has played an important role by providing the necessary high-end processing capabilities on demand to support the growing array of deep-learning-based applications. Moreover, though deep learning technologies have been proven to produce higher levels of accuracy, especially when analyzing medical imaging datasets, the algorithms generally need to be trained with significantly large amounts of data. For example, well-known Apple Siri and Google Now[6] are typical examples of cloud-reliant, deep-learning-based applications. These cloud-only approaches require large amounts of data to be sent to the cloud over wireless networks. This not only places enormous stress on the wireless network but—as we move into more complex applications, such as in the biomedical domain—also raises several security and privacy concerns (e.g., sharing data over public wireless networks). The approach to move data to the cloud to facilitate deep analysis is expensive and proving to be infeasible because of limitations in Internet bandwidth, as well as because of concerns pertaining to data security and privacy.

To overcome the issue imposed by a cloud-only architecture, distributed deep learning approaches have started to emerge in the literature.[7,8] Some of these approaches take advantage of edge computing, wherein the deep learning model is distributed across edge and end devices. These developments are underpinned by advances in edge and mobile edge device technologies such as federated learning by Google,[9] Apple's latest Core ML toolkit that provides capability for Apple's mobile devices to run machine learning algorithms using pretrained models, compressed versions of deep learning frameworks such as tensor flow and caffe (with some of these expected to be available on mobile platforms soon), ARM's improvement to its GPU platform, and Intel's movidius platform.[10] In parallel, there have been several advances in the cloud and edge computing area[11,12] with the introduction of several paradigms to orchestrate cloud- and edge-based applications. Osmotic computing is one such paradigm that was discussed in the recent IEEE cloud computing Blue Skies column.[13] Though these paradigms provide high-level architectural principles of developing and deploying cloud- and edge-based applications, they fail to provide a detailed account of how technologies such as deep learning can be orchestrated and take advantage of the cloud, edge, and mobile edge environments.

Springboarding on advances in cloud, edge, and mobile edge frameworks, coupled with developments in distributed deep learning, we present in this Blue Skies column our vision of a distributed deep learning approach for cloud, edge, and mobile edge environments. In particular, we present a detailed account of issues and challenges in developing and deploying deep-learning-based applications in an osmotic computing environment.[13] Specifically, we focus on biomedical applications because of the intricate nature of such new types of applications. However, the challenges identified in this paper apply to a wider class of distributed deep-learning-based applications (and with some level of abstraction, include other distributed machine learning).

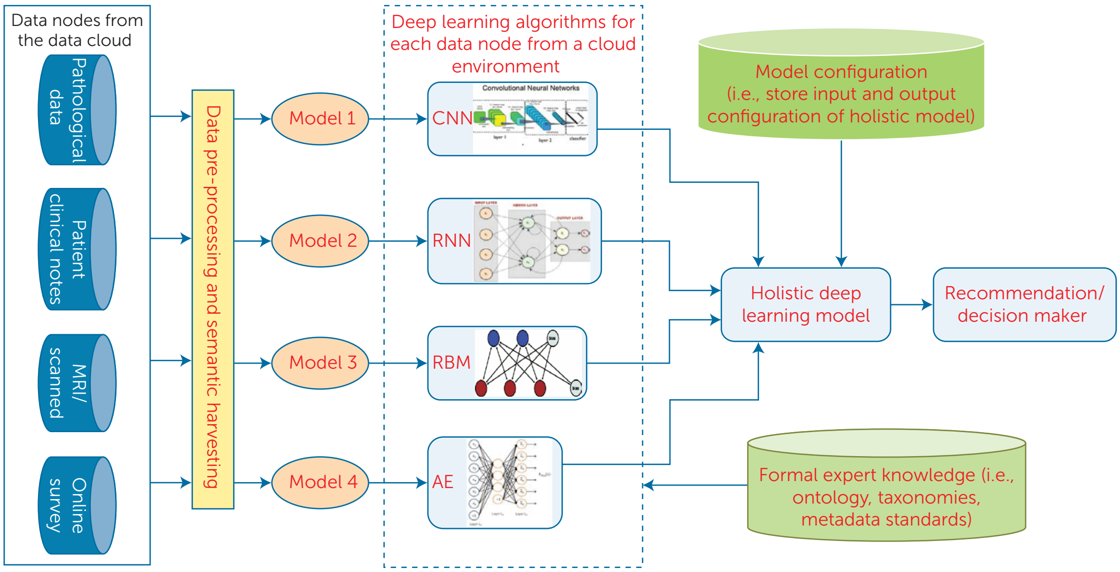

FIGURE 1. Vision of an HDDL approach.

Motivation: The Case for Distributed Deep Learning in an Osmotic Computing Environment

Consider a typical biomedical application, such as telemedicine. Sam Robert, age 35, suddenly suffers from chest pain. Upon visiting his regular general practitioner (GP), he is prescribed an electrocardiogram (ECG), blood test, and x-ray. The GP also logs the patient's information into his EHR. Generally, the EHR is stored locally within the clinic and is only shared with other clinics or hospitals upon return request because of concerns related to privacy. The results from various tests are in several forms: ECG report with corresponding data (time series), blood test report (key value pairs), and x-ray (image). Coupled with the EHR, this enables the GP to diagnose the issue (e.g., muscle sympathetic nerve activity [MSNA]),[14] make further checks (e.g., determine body mass index [BMI] and blood pressure), and recommend a treatment plan. In such a scenario, a deep learning model trained on x-ray images may accurately classify the x-ray but may fail to detect the MSNA problem.[15] The IoT also throws in a new dimension in available datasets as increasingly smart health devices produce data that can be used to understand the personal context of the user (e.g., sleep data from Fitbit). To realize the complete potential of deep learning, such as diagnosing MSNA problems that may lead to heart attack, there is the need for a holistic approach that (1) can distribute different layers of the deep learning model, (2) provide integration capability to fuse output of one deep learning model with another, and (3) be able to do this in an edge, mobile edge, and cloud environment to manage geographical distribution and reduce movement of data.

A Holistic Distributed Deep Learning Approach in Cloud, Edge, and Mobile Edge

Figure 1 presents our vision of a holistic distributed deep learning (HDDL) approach that provides the necessary integration capability for different types of data stemming from a variety of sources across cloud, edge, and mobile edge environments. Our vision is to develop a usable HDDL approach that can be deployed and orchestrated on (1) mobile edge, such as mobile smartphones that can perform local processing of user data (e.g., sleep data from Fitbit); (2) edge, such as Cisco IOX routers, OpenFlow-based software-defined networks providing the necessary data preprocessing and local deep learning capability (e.g., deep learning model for x-ray image processing); and (3) cloud, providing the necessary integration between various deep learning models and underlying heterogeneous datasets.

However, developing such an HDDL approach imposes several complications:

1. A key issue with deep learning is the complexity introduced by increasing amounts of heterogeneous datasets stemming from several geographically distributed data sources. As deep learning technologies try to learn high-level features from data by constructing low-level, midlevel, and high-level features from the underlying datasets, heterogeneity in the underlying datasets introduces significant complexity in training and developing deep learning models.
2. Different terminologies have been used to express the same context by different deep learning models. For example, a patient diagnosis of MSNA can be identified by blood test, x-ray, and clinical notes made by a doctor. A designated domain expert can look at the clinical variables and find the patterns. But, it is difficult to process these clinical data and text data when using a deep learning algorithm to identify patterns and produce predictive models for real-world applications.
3. If we assume each site (e.g., doctor's clinic and radiologist clinic) to be an edge node, the question is, How can a distributed deep-learning-based application be deployed and configured across the edge and cloud nodes? Works that propose distributed deep learning approaches[7,8] consider the distribution of a centrally developed deep learning model to different locations across edge and cloud. For such approaches to be successfully, the centrally developed model will need to have access to an enormous amount of data that is distributed geographically (e.g., doctor's clinic and radiologist clinic). The challenge is to take advantage of the layered approach of deep learning and be able to dynamically generate a holistic model that can integrate inputs and outputs from several independently trained and developed deep learning approaches.
4. It is difficult to determine which deep learning framework is suitable for the application, because several deep learning frameworks (i.e., tensor flow, Theano, CNTK, MXnet, Chainer, Torch, Caffe and Keras)already exist. However, the harder challenge is to capture the notion of the semantics behind the model and data among the different edge nodes.
5. With access to different data sources, deep learning models developed by a particular institution may not be accessible by another institution. For example, the doctor's clinic may not have access to the deep learning model for x-ray used by the radiologist clinic (because of various factors, including geographical distribution, privacy, and security). Hence, developing an integrated, centralized deep learning model may no longer be feasible. Thus, it may make sense to share the outcomes rather than the model used for computation.

Challenges in Realizing the Vision of an HDDL Approach in an Osmotic Computing Environment

In the previous section, we presented our vision of an HDDL approach (exemplified using a biomedical application scenario). In this section, we elaborate on the challenges of realizing such a vision from two key perspectives: (1) an HDDL perspective and (2) the osmotic computing environment (i.e., cloud, edge, and mobile edge) perspective. These two perspectives introduce several diverse challenges that cover a range of areas, from distributed data management, semantic data representation, distributed deep learning, and big data to cloud, edge, and mobile edge resource orchestration, monitoring, and management.

Challenges in Developing HDDL

Many applications have been developed[14,16,17] by using distributed deep learning. However, building the training model from raw datasets is a key challenge. In the medical domain, sharing patient data often has limitations because of technical, legal, or ethical concerns, and the number of patients worldwide with a specific disease is limited.[8] To overcome the situation, an HDDL model[18] can be an alternative path for any organization to share information for better health outcomes (in the case of medical applications).

The following challenges are involved in developing an HDDL:

- *Semantics of the learning model and interoperability.* Understanding the underlying semantics that help to identify the input and output parameters of the deep learning models is critical so that other models can plug in and out easily.[8] Hence, we need an ecosystem that has model integration capabilities based on the semantic taxonomies. Clinical information is sensitive, and organizations often do not want to share their information. However, sharing the outcomes of the deep learning model can address any privacy issues. The key here is interoperability. The interoperable models can save computational complexity and time.[5,18] Such an interoperable model can also be configured individually to reflect organizational rules.
- *Data volume.* Deep learning is a highly computational model that connects the deep layers' neural network with tons of network parameters required to consider and compute the final outcome. For that reason, a huge amount of data and many high-performance machines are needed for the model computation process that can find and save patterns from the input data. Though there is no standard guideline about the training dataset, it is better to consider a decent volume of data. That is why deep learning has been successful in the domains of computer vision, speech recognition, and natural language processing. However, the medical domain is the opposite; we have approximately 7.5 billion people worldwide, and a significant number of people do not have access to basic healthcare facilities. The number of patients is limited, and many institutions do not believe in the open data concept. As a result, it is difficult to fit the small number of data when developing a comprehensive deep learning model. Moreover, understanding diseases and their variability is more complicated than other tasks, such as image and speech recognition. Consequently, from a big data perspective, the amount of medical data that is needed to train an effective and robust deep learning model would be too low compared with other applications of deep learning (e.g., speech recognition).[5,6,16]
- *Data quality.* The data in the medical domain, and other domains, are highly heterogeneous, ambiguous, noisy, and incomplete. For example, monitoring the patient's heart through the ECG machine often gives a huge number of missed signals that can create significant challenges to machine learning algorithms such as those used in deep learning. Because medical data are sensitive, it is important to understand the semantics of the data. However, doctors' notes are not clearly explained and not publicly available; apart from that, health professional are reluctant to add associated metadata. So, a critical challenge is to have good metadata that can guide further analysis. Training a good deep learning model with such massive and verify datasets is challenging and needs to consider several issues, such as data sparsity, redundancy, and missing values.[5,18]
- *Lack of semantic ontology and expert knowledge.* The context of expert knowledge is an invaluable part of any dataset. More specifically, expert knowledge is an important part of health care, because a limited amount of medical data is shared for research and often exhibits poor quality (e.g., incomplete data and noise). To overcome the shortcomings, incorporating expert knowledge in distributed deep learning can significantly affect the accuracy of the final output.[5,14] For example, publicly available medical encyclopedias (e.g., https://medlineplus.gov/encyclopedia.html), DBpedia (http://wiki.dbpedia.org), and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) should be analyzed to extract valuable content that can guide the distributed deep learning architecture. Furthermore, experienced health professional knowledge needs to be captured in a formal way and used to build an ontology based on the existing ontology—i.e., Cyc (http://www.opencyc.org), Gene Ontology (http://www.geneontology.org), and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (http://www.loa.istc.cnr.it/old/DOLCE.html)—that can help to analyze the quality of data. However, no deep learning techniques have made use of the ontology constraint between labels in classification

and recognition tasks or the potential semantic relations in features of the space of object recognition tasks. Furthermore, no deep learning model has addressed the problem of modeling high-level cognitive intelligence ability, such as reasoning, inferencing, or validating.

- *Temporality.* Deep learning models cannot handle temporality in data. For example, the 2016 thunderstorm flu in Melbourne, Australia, affected 35,940 people in this year.[19] Existing deep learning models cannot handle the emergence of new knowledge influenced by time factors. The state of the art of deep learning says it needs to design a new solution or method of deep learning that can handle temporal healthcare data.[1,5,16]
- *Privacy.* Privacy is an important feature of any system, particularly in approaches such as distributed deep learning over healthcare data, which needs to deal with extremely private information. The attack on all authentication and access control mechanisms breaks the model and personal privacy. Preserving privacy of distributed deep learning models is challenging because of its inherent scale and number of possible attacks.[7,10,11]

Challenges in Deploying HDDL in an Osmotic Computing Environment

Figure 2 provides an overview of a typical osmotic computing environment[16] that comprises a cloud, edge, and mobile edge ecosystem. Deep learning uses cascades of multiple layers of nonlinear units for feature extraction and transformation to be able to develop an HDDL approach wherein a layer could be an independently trained deep learning model whose output will be integrated by the holistic deep learning model (at a different layer) or could be a distributed part of a centrally trained deep learning model. Figure 3b provides an illustration of a holistic deep learning approach deployed over an osmotic computing environment. Next, we highlight challenges that are imposed by the development and deployment of distributed deep learning algorithms in an osmotic computing environment.

Deep Learning Model Performance Benchmarking

Figure 3a provides an abstract graph view of the deployed holistic deep learning approach. Each node in the graph is responsible for running and managing a particular layer of the deep learning model. Hence, understanding the performance of this model is important to provision and make decisions about the types and scale of each node (e.g., the mobile edge, edge, and cloud). The mobile edge in this case could be static sensor platforms or gateways or mobile smart devices such as smartphones.

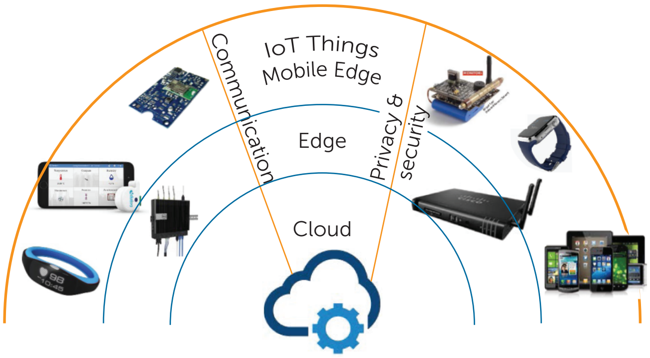

FIGURE 2. Cloud, edge, and mobile edge ecosystem.

Deep Learning Application Composition and Deployment

Consider the distributed deployment of deep learning layers as presented in Figure 3a. We take a graph-based illustration to exemplify the distribution of the multiple layers of the holistic deep learning model. In such a scenario, we introduce the term *microservice,* which represents each deep learning model that runs on the mobile edge, edge, or cloud node. Depending on the application, and the inputs and outputs of the deep learning model, one or more of the microservices (deployed on the nodes) could be used to process the data. These microservices would generally be selected based on their geographical distribution within the network, with the aim of reducing latency and bandwidth consumption. Furthermore, the microservices will drive the resource selection and optimization problem based on the demands imposed by the deep learning application and underlying data security and privacy concerns. The holistic model will provide the necessary interoperability capability for multiple deep learning models residing on the nodes to work toward a common outcome.

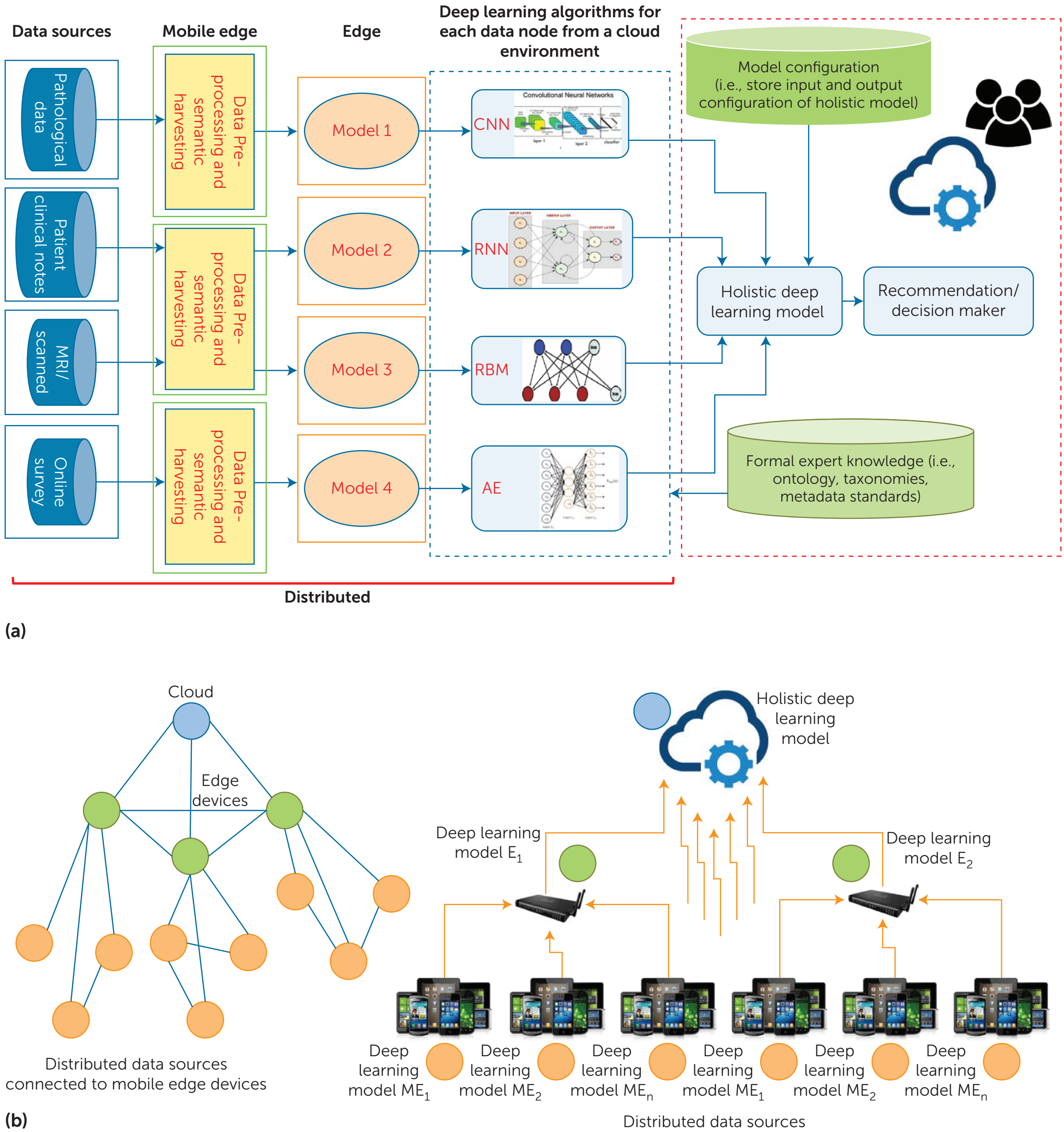

FIGURE 3. Osmotic computing environment. (a) Graph representation of the holistic deep learning approach. (b) Overview of holistic deep learning approach deployment.

Holistic Data Management and Indexing

In the era of resource-constrained edge devices (IoT gateways) and high-velocity, high-volume, and high-variety data, no single database approach is optimal for data management challenges relevant to the deep learning applications that need to be provisioned in a highly distributed osmotic computing environment. Hence, there is a need to investigate hybrid big data management platforms for which different database architectures are optimal for managing disparate data sources that feed into the learning models. To support seamless access to data distributed across different architectures (mapped to different parts of the osmotic computing environment), an integrated, federated application programming interface (API) will need to be developed[18] that can (1) seamlessly provide access to heterogeneous data sources required for training and retraining the deep learning models and (2) dynamically respond to changes in data volume and velocity by controlling the scalability features of the underlying database architecture and resource infrastructure (edge vs. cloud vs. mobile edge). In addition, a query language will also need to be developed that can talk across the different deep learning models.

State of the Art

Machine learning is an old branch of artificial intelligence that can learn relationships from the data.[1,5,8,17,20] Generally, machine learning algorithms contain four steps: data harmonization, representation of learning, model fitting, and evaluation.[8] In contrast, deep learning[1,2,8] is different from traditional machine learning in how multiple layers (i.e., input and hidden layers) are learned from the raw data. Deep learning consists of multiple hidden layers based on the tradition that a neural network allows models to learn the representation of data with multilevel abstraction and automatically extract the multiview features from the data without human involvement.

The literature[5,8,16,17,21] shows various deep learning algorithms that have been used for various data in the medical field. An example would be the deep belief network used for analyzing microRNA, protein structure, fundus images, wearable devices, electroencephalography (EEG), ECG, implantable devices, EHRs, and red, green, and blue plus depth (RGB-D) camera data. The deep autoencoder method is used for analyzing gene expression, magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI), microscopy, and social media data. Another method, called the convolution neural network, is used for analyzing MRI and computed tomography (CT) images, video, wearable devices, geotagged images, and blood and lab test data. The recurrent neural network is used for analysis of EHRs and mobile device data. Restricted Boltzmann machines (RBMs) are used for MRI and CT images and mobile and EHR data. Moreover, one review[5] illustrates that there are no studies using deep learning to combine all of these data sources or part of them in a joint representation for medical analysis and prediction.

The preceding deep learning algorithms have already been used for analyzing data locally in the cloud environment. But this has some associated challenges, such as communication cost, latency issues, data semantic and privacy concerns, that cannot be ignored. As a result, a distributed machine learning approach[3,7,22] has been considered. In addition, a single dataset contains a limited number of features—not enough to extract useful information. More specifically, the number of patients is limited worldwide, and we need more features to build a better predictive model.

Studies show that deep learning models[7,17,21] are a new trend to gain complex and highly dimensional knowledge. Figure 4 shows the growth of the learning layers in a deep learning framework, which vigorously affects various tasks, such as a huge amount of image analysis in the medical domain.[7] The literature[5,7] does not provide studies that attempt to combine the different data sources by using deep learning across an osmotic computing environment (cloud, edge, and mobile edge).

Conclusion

In the last few years, we have seen significant research efforts from the distributed systems community toward developing resource management tools and techniques for cloud, edge, and mobile edge computing environments. Similarly, the artificial (machine) intelligence community has been exploring virtually all aspects of machine learning, including deep learning, as well as enhancing classical algorithms. With the emergence of big data, we are witnessing an increasing number of machine learning algorithms being ported to cloud datacenters, supported by tools such as Watson Machine Learning, MLbase, Google TensorFlow, and Apache Mahout.

However, limited methodologies and tools are available in the literature that can exploit the

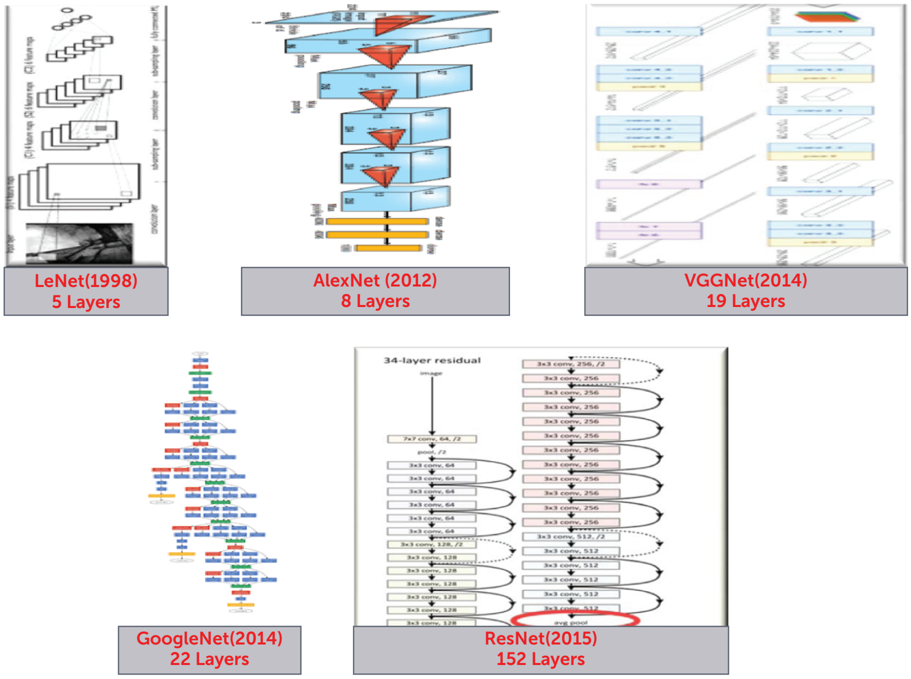

FIGURE 4. Different deep learning network structures.

convergence of the osmotic computing environment with the new generation of resource- and data-aware machine learning algorithms.

In this Blue Skies column, we have identified some challenges in realizing an edge-cloud-driven distributed deep learning approach. We exemplified the need for such an approach using a medical scenario, but numerous other use cases could benefit from such an approach. We proposed a high-level architecture of our approach that we envisage could serve as a blueprint for further research and development.

References

1. J. Dean et al., "Large Scale Distributed Deep Networks," *Proc. 25th Int'l Conf. NIPS*, vol. 1, 2012, pp. 1223–1231; http://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf.
2. Y. LeCun et al., "Deep Learning," *Nature*, vol. 521, 2015, pp. 436–444.
3. I. Goodfellow et al., *Deep Learning*, MIT Press, 2016.
4. G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Med. Image Anal.*, vol. 42, 2017, pp. 60–88.
5. R. Miotto et al., "Deep Learning for Healthcare: Review, Opportunities and Challenges," *Brief. Bioinform.*, vol. 1, no. 11, 2017, bbx044.
6. Y. Kang et al., "Neurosurgeon: Collaborative Intelligence between the Cloud and Mobile Edge," *ACM Sigplan Not.*, vol. 45, no. 1, pp. 615–629, 2017.
7. S. Teerapittayanon et al., "Distributed Deep Neural Networks over the Cloud, the Edge and End Devices," *Proc. IEEE 37th Int'l Conf. ICDCS*, 2017.
8. K. Chang et al., "Institutionally Distributed Deep Learning Networks," 2017; arXiv:1709.05929.
9. B. McMahan and D. Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," *Google Research Blog*; https://research.googleblog.com/2017/04/federated-learning-collaborative.html.
10. A. Kaul, "Artificial Intelligence Processing Moving from Cloud to Edge," 2017; https://www.tractica.

com/artificial-intelligence/artificial-intelligence-processing-moving-from-cloud-to-edge/.
11. M.H.u. Rehman et al., "Opportunistic Computation Offloading in Mobile Edge Cloud Computing Environments," *Proc. 17th IEEE Int'l Conf. MDM*, 2016; http://ieeexplore.ieee.org/document/7517797.
12. P. Mach and Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, 2017, pp. 1628–1656.
13. M. Villari et al., "Osmotic Computing: A New Paradigm for Edge/Cloud Integration," *IEEE Trans. Cloud Comput.*, vol. 3, no. 6, 2016, pp. 76–83.
14. Z. Zhu et al., "Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding," *Proc. IEEE 16th Int'l Conf. Data Mining*, 2016.
15. P.G. Guyenet, "The Sympathetic Control of Blood Pressure," *Nature Rev. Neuroscience*, vol. 7, no. 5, 2006, pp. 335–346.
16. M. Nardelli et al., "Osmotic Flow: Osmotic Computing + IoT Workflow," *IEEE Trans. Cloud Comput.*, vol. 4, no. 2, 2017, pp. 68–75.
17. P.P. Jayaraman et al., "Privacy Preserving Internet of Things," *Future Gener. Comput. Syst.*, vol. 76, no. C, 2017, pp. 540–549.
18. C. Hardy et al., "Distributed Deep Learning on Edge-Devices: Feasibility via Adaptive Compression," *16th IEEE International Symposium on Network Computing and Applications* (NCA), 2017.
19. Immunization Coalition, "Influenza Activity Surveillance 2017;" http://www.immunisationcoalition.org.au/news-media/2017-statistics/.
20. L.R. Miotto et al., "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," *Sci. Rep.*, vol. 6, 2016, 26094.
21. D. Ravì et al., "Deep Learning for Health Informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, 2017, pp. 4–21.
22. H. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," 2016; arXiv:1602.05629.
23. E. Rahm, "The Case for Holistic Data Integration," *East European Conference on Advances in Databases and Information Systems*, 2016, pp. 11–27.

AHSAN MORSHED *is a research fellow at the University of Swinburne, Melbourne, Australia. His research interests include Deep Learning, Machine Learning, big data analytics, Semantic web and Medical informatics. Morshed has a Master's in Computer Science from Royal Institute of Technology, Stockholm, Sweden and a PhD in Computer Science from the University of Trento, Italy. Contact him at amorshed@swin.edu.au*

PREM PRAKASH JAYARAMAN *is a research fellow at the Swinburne University of Technology. His research interests include Internet of Things, cloud computing, big data analytics, and mobile computing. Jayaraman has a PhD in computer science from Monash University. Contact him at prem.jayaraman@gmail.com or http://www.premjayaraman.com.*

TIMOS SELLIS *is Director of Swinburne's Data Science Research Institute. His research interests include big data, data streams, personalization, data integration, and spatio- temporal database systems. Professor Sellis is an IEEE Fellow for his contributions to database query optimization and spatial data management. He is also an ACM Fellow for his contributions to database query optimization, spatial data management and data warehousing. He holds an MSc degree from Harvard University, a PhD from the University of California at Berkeley, and has served as president of the National Council for Research and Technology of Greece. Contact him at tsellis@swin.edu.au*

DIMITRIOS GEORGAKOPOULOS *is a professor and director of the Internet of Things Key Laboratory at Swinburne University of Technology. His research interests include the Internet of Things, data management, and process management. Georgakopoulos has a PhD in computer science from the University of Houston, Texas. He's a member of IEEE and ACM. Contact him at dgeorgakopoulos@swin.edu.au.*

MASSIMO VILLARI *an associate professor of computer science at the University of Messina. His research interests include cloud computing, Internet of Things, big data analytics, and security systems.*

Villari has a PhD in computer engineering from the University of Messina. He's a member of IEEE and IARIA boards. Contact him at mvillari@unime.it.

RAJIV RANJAN *is a reader in the School of Computing Science at Newcastle University, UK; chair professor in the School of Computer, Chinese University of Geosciences, Wuhan, China; and a visiting scientist at Data61, CSIRO, Australia. His research interests include grid computing, peer-to-peer networks, cloud computing, Internet of Things, and big data analytics. Ranjan has a PhD in computer science and software engineering from the University of Melbourne. Contact him at raj.ranjan@ncl.ac.uk or http://rajivranjan.net.*