WENZHOU KEAN UNIVERSITY

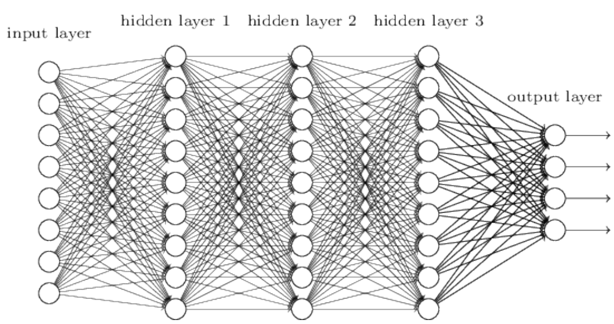
ENG 3091 REPORT

Emerging Trend in Applied Deep Learning Research

BY

Hangzhi (Birkhoff) Guo

1025618



supervised by

Dr.Wendy AUSTIN

December 16, 2019

ABSTRACT

The recent emerging trending of artificial intelligent has drawn a significant attention in academia and industry. Such trend should be given credit to the recent rise of deep learning, a specific type of machine learning method mimicking the neural network behavior in the human brain that achieves significant performance powered by modern super-hardware. Various empirical studies indicate the superiority of deep learning among other machine learning algorithms. In this report, I will discuss the rising trend of applied field of deep learning in the research community. Specifically, I will talk about three trending fields, transfer learning, deep reinforcement learning, and recommender system. The recent innovations as well as challenges and potential applications will be discussed. Although I discuss state-of-the-art research in this report, I target to make readers who have limited knowledge in artificial intelligence to understand the novel applications in this tremendously potential yet inconclusive area.

***K*eywords** Deep Learning · Artificial Intelligence · Machine Learning · Neural Network · Reinforcement Learning · Natural Language Processing · Big Data · AI for Social Good

Contents

Figures

1 Introduction

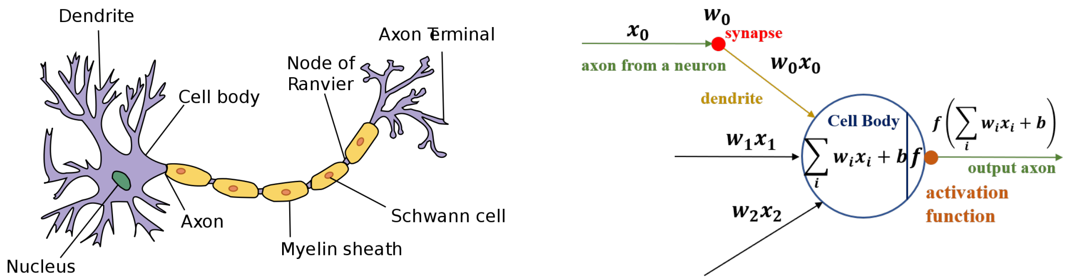
Scientists have dreamed for centuries to allow a machine to have intelligence like a human being. Not long after the initial invention of the electronic computer, in 1950, British scientist Alan Turing (Turing, 2009) published a paper entitled "Computing Machinery and Intelligence", which initially put forward an imaginary question: *can a machine think like a human?* To answer this question, generations of scientists have devised different approaches to study this problem, such as developing expert systems aiming to model the human expertise in certain fields.

Applications, such as medical diagnosis, can be processed by representing human knowledge and thus can be applied with the expert systems. However, there are excessive problems that are unable to be determined by simple facts and rules. For example, problems in distinguishing cats or dogs have baffled scientists for decades until the early 21st century. Human might determine the species of animals by their specific appearance, such as the shape of mouth, color of the hair, etc. If a computer follows the process of identifying objects by human, it would generate a lot of new problems, e.g. how to identify the shape of the object's mouth or eye? What if two animals have similar hair color?

Then, a novel but not completely unorthodox idea was arose: instead of telling the machine how to do it, we can allow the machine to learn the rule by itself. We later called these kinds of method as *machine learning*. Although this idea might sound imaginary, it originated from statistical learning theory, which has been studied for over a century, long before the computer was invented. Statistical learning heavily relies on statistical modeling and mathematics algorithm, which can be considered as consisting of multiple mathematical functions, or in a statistician's view, multiple (mixed) probability distributions. The statistical learning model tries to fit the given data as close as it can, as in middle school when our math teachers asked us to fit a linear function from given data points. When the data gets sufficiently large and the model fits the data cohesively, we generally believe that such statistical model can represent the pattern of particular dataset. Such statistical learning algorithms (e.g. Support Vector Machine, K-Nearest Neighboring, Gradient Boosting Algorithm) work really well in certain real-world problems that are applied in the giant industry to date. However, such methods are formulated on mathematics, making the argument for the scarcity of "intelligence". Engineers should handcraft the features in the dataset, causing redundant laboring work. Moreover, for certain problems such as speech recognition, computer vision, it is hard to model such problems with statistical algorithms since these algorithms suffer from the explosive dimension, meaning the algorithms will perform badly when the input dimension is large.

Meanwhile, some scientists tried to devise models which simulate the thought processes of the human brain. Such models simulating the neural network in human's brain, designed in 1957, was the embryonic form of modern deep learning model. After decades of efforts, LeCun et al. (1989) provided the first practical demonstration of a neural network with back propagation onto read "handwritten" digits, suggesting that it is applicable on deploying deep neural network on real-life problems. The cornerstone of the new generation of deep learning started in 2006, when Hinton, Osindero, and Teh (2006) put forward the deep belief network, which allows machine to overcome previous limitations and train a deeper network. Empowered by a recent explosion of data in the internet era, deep learning exhibits its superior capability. In 2012, a deep learning model, AlexNet (Krizhevsky, Sutskever, and Hinton, 2012), achieved a error of 15.3% with more than 10.8 percentage point lower than the followers in the ImageNet Large Scale Visual Recognition Challenge, an image classification competition requiring competitors to classify millions of images into 1000 classifiers.

To date, deep learning has applied to almost every inch of our daily life. Probably the most prominent application in deep learning is *computer vision (CV)*, which empowers computers to perceive the visual perception as human being do. Multiple autonomous car companies (such as Uber, Tesla, Waymo) are exploiting deep learning to detect pedestrians, cyclists, vehicles on their autonomous car. Computer vision



(a) The structure of a biological neuron; (b) Mathematical model for neural net.

Figure 1: Biological motivation and neural net unit.

technology also assists healthcare professionals saving patients' lives by eliminating inaccurate diagnoses. For example, Rajpurkar et al. (2017) developed Chexnet to detect radiologist-level pneumonia, which outperforms medical professions at Stanford Medical School. Apart from computer vision, the deep reinforcement learning is another thriving approach which achieves tremendous success. The famous AlphaGo (Silver et al., 2016, 2017) has defeated two of the best Go player in human history. Empowered by deep reinforcement learning, the Go agent can make movements that human players will not ever consider. Moreover, natural language processing (NLP) is another well-investigated field for deep learning application. Enhanced by the recurrent neural network, Google Translator (Wu et al., 2016; Johnson et al., 2017) now achieves expert-level in translating human languages. The state-of-art NLP model, BERT (Devlin et al., 2018), is reported to outperform human in various reading comprehension tasks.

In this report, I will describe the recent trends of applications in deep learning. Specifically, I will discuss three well-studied but promising areas in deep learning, *transfer learning*, *reinforcement learning*, *recommender system*. First, I will discuss the recent novel innovations as well as the challenges in these areas. I will also project the future trends in these areas and describe how such technology can empower society and human being. Although this report discusses state-of-the-art (SOTA) applications and technologies in deep learning, I attempt to write this for readers who have only minor background in artificial intelligence to have a basic understanding in this still developing areas in deep learning.

2 Background

As one special type of machine learning, deep learning has achieved significant performance and applied in various application field, such as computer vision, natural language processing, speech recognition. In this section, I will discuss the basic set of deep learning (2.1) and the three important reasons for the push of deep learning in recent years (2.2).

2.1 What is deep learning?

The modern term *Deep Learning* originated from the traditional architecture of neural networks inspired by the biological neural system. The basic computational unit of the brain is a *neuron*, which is connected with a significant amount of *synapses*. Each neuron receives input signals from its *dendrites* and produces output signals along its single *axon*. The axon eventually connects to dendrites of other neurons via synapses. In the computational model of a neuron, or neural net unit, the signals that travel along the axons (e.g. x_0) interact with the dendrites of the other neuron based on the synaptic (e.g. $w_0 x_0$) strength at that synapse (e.g. w_0). With multiple signals connected by synapses to dendrites, the cell unit receives multiple signals, and combines all sorts of signals (e.g. $\sum_i w_i x_i + b$). After receiving those input signals, we model the firing rate of the neuron with an activation function f that represents the frequency of the spikes along the axon (Brunel, Hakim, and Richardson, 2014). Like the biological neuron, the neural unit can connect and be connected by other units. Figure 1 illustrates the basic structure of a biological neuron and the neural unit in today deep learning network.

It is notable that despite the initial motivation of neural networks that come from neural biology, modern deep learning architecture becomes a refined mathematical and engineering model which can approximate any measurable function (Hornik, 1991). In other word, the development of modern neural networks is driven primarily by the findings of mathematics and engineering, rather than biological neuroscience. We significantly simplify the real neuron to our neural unit given the complexity of the biological neuron and scarce knowledge of its panoramic structure.

With various neural units, the deep learning network connected all the neural units by means of different layers. It is the depth of layers that has a notable difference between today's deep learning network versus the initial version. Empirically, greater depth seems to result in better generalization for a wide variety of tasks (Hinton, Osindero, and Teh, 2006; LeCun, Bengio, and Hinton, 2015; He et al., 2016; Cheng et al., 2016). One theory called *universial approximation theorem* states that regardless of what function we are trying to learn, we know that there exists a large network to represent this function. However, such theory does not show how large the network should be, and the parameter size might be unreasonably large. Moreover, one statistical theory called *No Free Lunch* shows that there is no superior machine learning algorithm in all cases. Therefore, many different neural network (illustrated by Figure 2) architectures have been developed for specific tasks. Convolutional neural network (LeCun et al., 1989), one of the most popular network structure in deep learning world, is specifically designed for computer vision tasks. For tasks coping with the sequential data, Hochreiter and Schmidhuber (1997) proposed Long short-term memory (LSTM), which has been proven to be successful in tasks covering sequential dependencies.

2.2 Why does deep learning succeed until now?

Although deep learning is a hot area, unlike most scientific areas that encounter breakthroughs that are typically triggered by innovative novel findings, most fundamental theories and ideas in deep learning are not put forward in recent years. In fact, the original form of deep learning, the neural network, was first presented in 1950s, and perfected in 1980s. The current research community believe this trend should give credit to three important factors: computational resources (hardware, 2.2.1), explosive data (big data, 2.2.2), and increasing understanding of the model (algorithm, 2.2.3), ordered in descending on significance.

2.2.1 Hardware

Although most deep learning models were proposed in the last century, due to the limited computational capability on that period, the performance and accuracy of machine learning could not be proved and determined. As a result, the development of artificial intelligence encountered the "iceberg of AI" in the late 20th century. At present, the personal computer is at least 100 times faster than 20 years ago, which enables developers to train a small network on their personal laptop. More importantly, due to the immense matrix calculation while training a deep learning network, researchers found out that the *Graphical Processing Unit (GPU)* accelerates the training process significantly. Dean et al. (2012) reports that in order to train a huge net, Google built a multiple clusters of CPUs system, costing total 5 billion dollars. However, using GPUs, they can achieve the same processing power but reduce the costs to just $33k. To date, multiple institutions are developing accelerator specifically designed for AI, which speeds up the computation further. In 2017, Google announced their AI accelerator application-specific integrated circuit (ASIC) which has already been deployed in their data center as well as applied in several projects, including the AlphaGo.

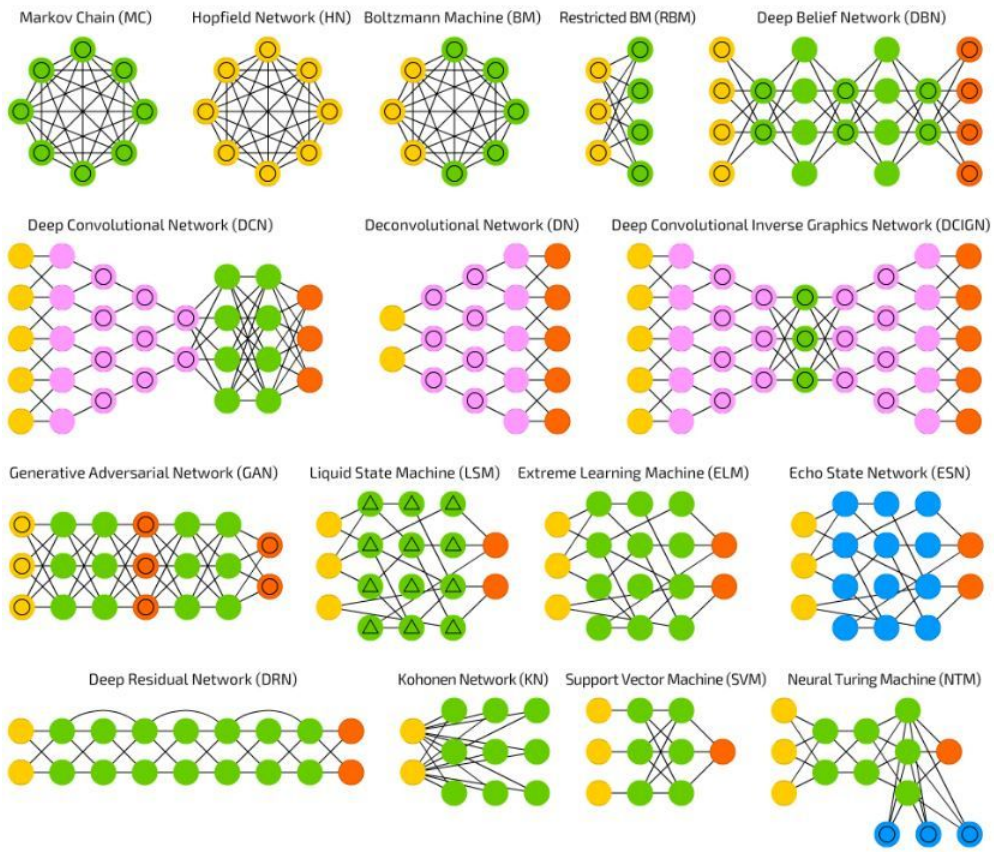


Figure 2: Architectures of different neural network.

2.2.2 Large Volumes of Data

Another crucial aspect is the availability of massive amount of data. Human beings generate tons of data everyday in this digital era; every movement on the internet will be recorded. As a result, technology-based companies such as Google, Facebook, Amazon and Alibaba have collected and maintained data that is measured in exabyte proportions or larger. These large quantities of data contribute to train a powerful neural network. Unlike traditional machine learning method, deep learning continues to enhance its performance while the dataset gets augmented. ImageNet (Deng et al., 2009) in Computer Vision are one fore type to demonstrate the significance of large volume of data. Before the ImageNet era, the computer can hardly distinguish dogs and cats, while the SOTA deep learning model are capable of outperforming human in classifying images into different categories. With tremendous sets of images crawled from the internet, deep networks can capture important patterns and thus achieve great results.

2.2.3 Algorithm

Despite the fact that the prevailing technologies in the deep learning world are invented ahead of this era, researchers still contribute to some key ideas which change the understanding of deep neural networks. First and foremost, researchers have determined that making the *depth* of model is crucial in the network architecture. As illustrated in Figure 3 While AlexNet (Krizhevsky, Sutskever, and Hinton, 2012) constructs has only 8 layers, ResNet (He et al., 2016) are able to train a 101 layers network, resulting considerable performance improvement. Another important achievement is the *Attention Layer* (Vaswani et al., 2017), a novel neural network initially designed for mimicing the human visual attention. This architecture later

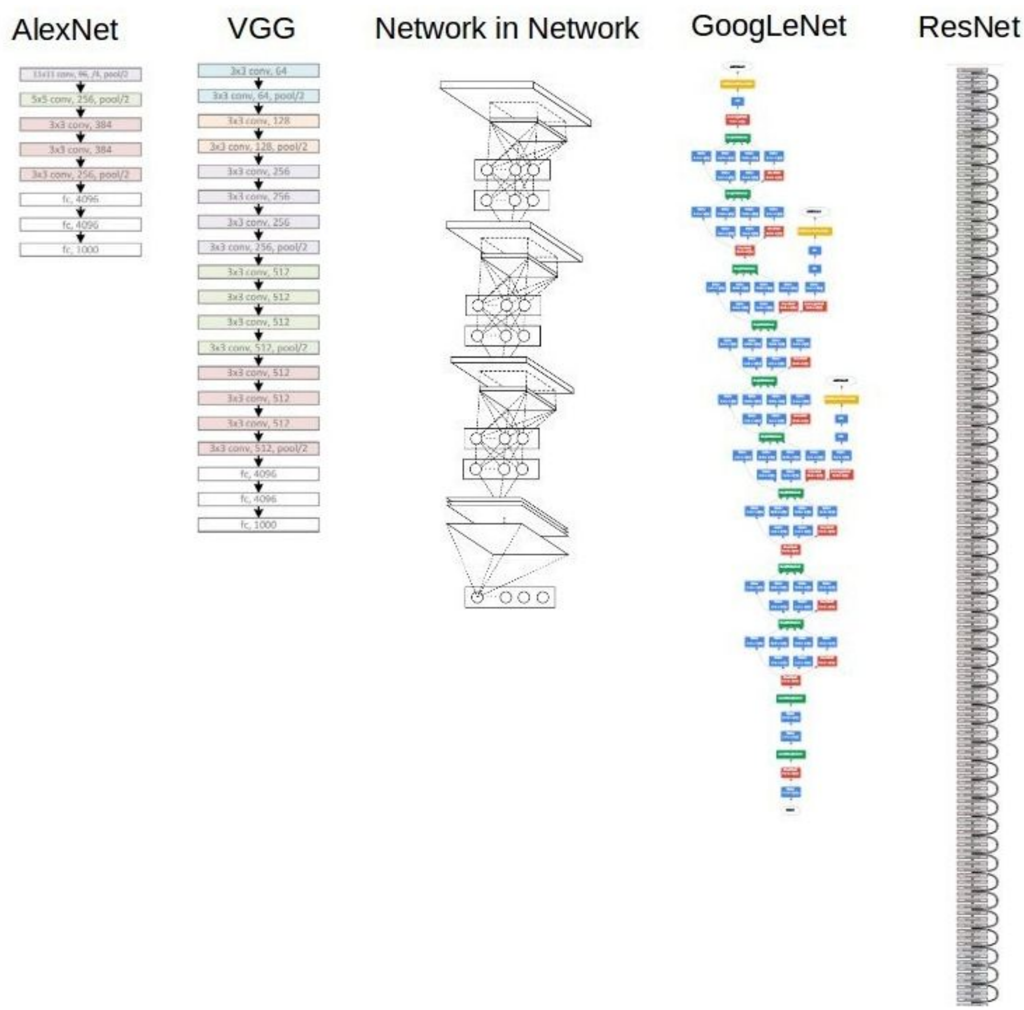


Figure 3: Depth evolution of convolutional neural network.

on achieved a wide success in natural language processing (NLP). Devlin et al. (2018) reports that BERT, a special attention architecture, achieves human-level on various NLP tasks, such as reading comprehension, machine translation, sentiment analysis.

3 Discussion

As stated previously, deep learning algorithms extract meaningful abstract representations of the raw data through the use of an hierarchical multi-level learning approach. As an effective representative learning approach, deep learning algorithms are crucial in extracting complex features, and areas such as computer vision, speech processing and natural language processing, all rely heavily on its powerful representative capability. In this section, I will discuss about three popular topics that still carry profound potential in today's research. Topics including transfer learning (3.1), reinforcement learning (3.2), user modeling (3.3) appeal to large research community, and these topics have considerable potential in future research.

3.1 Learning from Transferring

Unlike supervised learning, transfer learning concentrates on applying knowledge gained from one task to other similar or related problems. This learning strategy is analogues to the learning process of humans, in that we recognize and apply relevant knowledge from previous learning experiences when encountering new tasks. In deep learning, one neural network is firstly trained on a problem similar to the problem that is being solved, and one or more layers can be used in a new model trained on the problem of interest. Erhan et al. (2010) suggest that such transfer learning method, often referred to as pre-training, results in significant reduction of training time and boosting accuracy rate in generalization phrase.

In computer vision, transfer learning is a well-established paradigm that pre-training models using ImageNet (Deng et al., 2009) and then to fine-tune the models on target tasks that often have less training data. Such learning method enables state-of-the-art results in multiple tasks, including object detection, image segmentation, and action recognition. In a nutshell, if the computer can distinguish between cat and dog, computer should also be capable of differentiating tigers and lions. Many researchers contribute to transfer learning in the visual perception for computer. Girshick et al. (2014) proposed a simple and scalable convolutional neural network to detect and segment images; they also applied the supervised pre-training method followed by domain-specific fine-tuning and achieved boosting performance. Long, Shelhamer, and Darrell (2015) proposed a Fully Convolutional Networks (FCN) for semantic segmentation, which refers to link each pixel of image to a classification label. He et al. (2016) proposed the residual neural network, an influential work which enables researchers to train a network in depth of 100 layers, and received the best performance in various tasks, such as image recognition.

In the recent two years (since 2018), natural language processing has experienced a big revolution by pre-training the language model, which results in human-level performance in tasks such as reading comprehension, sentiment analysis, machine translation. To represent words and paragraph of a corpus of texts, Mikolov et al. (2013) introduced a word2vector to annotate extraction of semantic representation of words by denoting each word into an independent vector. Such word vectors can be applied in other NLP tasks. However, the biggest challenge of such word representation is the inability to share representation at sub-word levels. In other words, if one word has different meanings (e.g. bank has both meaning of the land alongside to a river or lake or a place to store money), the representation of word2vector cannot distinguish such differences. Researchers try to explore the option of deep learning using an enormous text corpus in the internet. Peters et al. (2018) introduced a bi-directional LSTM network to create a rich language representative model. Devlin et al. (2018) applied the bi-directional Transformer model training in an extremely large dataset, which leads to an impressive performance boosts.

The future research will not be limited in the area merely in computer vision and natural language processing. For instance, Spruyt (2018) proposed a convolutional neural network framework to represent a geolocation. Such method can be used directly for tasks such as venue mapping or transport classification and contribute to improve the classifier accuracies and generalization capabilities by means of transfer learning.

3.2 Learning from Reinforcing

Reinforcement learning is another area that achieves a wide variety of successes in recent artificial intelligence research. Some striking achievements, especially among sophisticated strategy game (e.g. AlphaGo, Dota), demonstrate the potential of this promising area. It also gains tremendous success in training simulated robots to follow human instructions.

The setting of reinforcement learning differs in the supervised learning methods. As Figure 4 illustrates, the key components of reinforcement learning are *agent* and *environment*. The *environment* is the world that the agent interacts with. At every step of interaction, the agent obtains an observation of the state of the world, and then decides on making an action. The agent also perceives a reward signal from the environment, which indicates whether such action is successful. The goal of the agent is to learn an optimal policy by maximizing

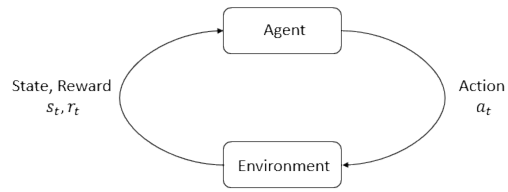


Figure 4: Agent-environment interaction loop.

its cumulative reward. Reinforcement learning methods are ways that the agent can learn behaviors to achieve its goal.

While reinforcement learning could solve difficult problems, algorithms incorporating deep leaning models can achieve super-human performances in sophisticated strategy game, including Go, Dota 2, Atair. Open-Mind (Mnih et al., 2015) demonstrates to surpass the professional human players in Atair by training the deep Q-learning agents. They (Vinyals et al., 2019) also trained a deep-learning-based agent and manage to beat the human world champion in StarCraft, a complicated strategy video game. Perhaps the most famous work for this team is AlphaGo, who have defeated two best Go player in the world. The first AlphaGo version (Silver et al., 2017) utilized the neural network and the state-of-the-art Monte Carlo tree search programs to simulate thousands of random games of self-play. The second version AlphaGo Zero (Silver et al., 2016), compared with the first version, was trained from scratch and competed with each other. AlphaGo Zero was reported to outperform AlphaGo extensively, and defeated World Go Champions Ke Jie.

Deep reinforcement learning is not limited to the areas of strategy game. As one powerful tool to explore the uncertain area of human knowledge, deep reinforcement learning has been demonstrated to be leveraged in breaking the scientific barrier. Popova, Isayev, and Tropsha (2018) integrated two deep neural networks – generative and predictive to generate novel target compounds. Another related work (Segler, Preuss, and Waller, 2018) applied deep reinforcement learning and symbolic artificial intelligence discover retrosynthetic routes. They reported such method performed significantly faster than traditional computers. In quantum physics, Niu et al. (2019) proposed a deep trusted-region reinforcement learning framework to simultaneously optimize the speed and fidelity of quantum computational problems that human scientist had difficulty resolving.

The future research in deep reinforcement learning will extend the area to broader areas, especially in problems related to society and humanity. For example, Yadav et al. (2016) proposed a dynamic social network framework to maximize the influence of homeless individuals for the awareness of HIV-AIDS spreading. Such problem has potential to be solved effectively in deep reinforcement learning manners, which requires further research and investigation.

3.3 Learning from Human Interests

Understanding user's interests are crucial to large enterprises who seek to advertise commodities to customers. By understanding users' preferences, and personalizing user experiences accordingly, companies such as Google, Amazon, Netflix, and Alibaba gain significant revenue increase as well as user's satisfaction and loyalty of certain application. Such system, often referred as *recommender system*, is aimed to narrow the scale of enormous items and facilitate with users in the decision-making process.

The recommender system and user modeling has been studied in the academy and industry for decades. Herlocker et al. (1999) proposed a algorithmic framework for performing collaborative filtering, and Sarwar et al. (2001) presented an item-based collaborative filtering algorithm, which is the most pervasive-used method in industry (Schafer, Konstan, and Riedl, 2001). Amazon (Linden, Smith, and York, 2003) adopted item-to-item collaborative filtering in production, which displayed items similar to user's purchase and interests. However, such method suffered in three issues: 1) *Code start*: with insufficient data, the algorithm cannot make accurate predictions for new users and items; 2) *Sparsity*: the number of items are extremely large, while users only rate very few of items, leading to a sparse rating matrix; and 3) *Scalability*: given the large amount of users and items, excessive computational power is required in real-world production.

Before the rising of deep learning, enterprises applied the traditional machine learning model onto their recommendation services. For example, Facebook (He et al., 2014) introduced a model which combines gradient boosting decision trees with logistic regression, and contributed to significant impact to overall system performance. However, such method relied heavily on feature engineering, which requires engineers and scientists to manually craft cross features. Plus, the traditional machine learning model could not harness the benefits of big data (billion-level). Shan et al. (2016) proposed a deep crossing model, which contains embedding and several multi-layer perception layers. Cheng et al. (2016) proposed a Wide & Deep Network, which explores the prosperity of both memorization and generation. Such model not only outperforms traditional machine learning model, but holds three unique features: 1) such model convert a sparse input feature into a dense matrix, which overcomes the sparsity issue; 2) the deep network could harness explosive data points; as one marvelous property of deep learning, more data generally results in significant performance booming; 3) the deep learning model does not require feature engineering but can do such thing on their own, and it saves significant labor expenses.

Future research will continue to investigate user behavior by combining social knowledge and big data analysis. Zhou et al. (2018) proposed a Deep Interest Evolution Network (DIEN) to simulate the evolving change of user interests under the premises of user interests not remaining consistent.

4 Challenges & Conclusion

Despite the wide variety of applications of deep learning, there exist limitations that hinder humans from achieving artificial general intelligence. The three most notable flaws include: 1) deep learning heavily relies on large amount of labeled data. Deep learning has obtained excessive success in areas having large amounts of data, but such scenario does not happen in every field. Some areas such as medical diagnosis or learning analytic are unable to gather massive volume of data, and it hinders the applications of deep learning in these fields; 2) deep learning lacks interpretability. Although empirical studies indicate its power, scientists can hardly understand how the network makes predictions and decisions. The complexity of the deep neural network exceeds far beyond human knowledge, and the lack of explainability for neural network results in various problems, including fairness and ethical in decision making, lack of trust and accountability, etc. 3) neural networks are vulnerable to be attacked. Su, Vargas, and Sakurai (2019) shows that even the most advanced neural networks can be fooled by only modifying one pixel in an image, giving rise to a totally different result. This research indicates the significance of studying the robustness for neural networks to prevent adversarial attacks in real world applications.

This article presents recent breakthrough of deep neural networks. Deep neural network are a powerful tool that has become immersed in our life. Although the fundamental ideas were put forward decades ago, with recent large volume of data and powerful computing resources, we can apply the model in real-life applications with significant performance boosting. Transfer learning, reinforcement learning and recommender systems are three promising areas that have achieve profound success, and these areas have potential to consistently empower human beings. In recent years, increasing researchers focus on applications for social good. Researchers follow the paradigm that AI should never replace the human being, but empower us. Researches one related topics including the ethics, fairness, accountability are in high demand that attracts large research community.

References

Brunel, N.; Hakim, V.; and Richardson, M. J. 2014. Single neuron dynamics and computation. *Current opinion in neurobiology* 25:149–155.

Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, 7–10. ACM.

Dean, J.; Corrado, G.; Monga, R.; Chen, K.; Devin, M.; Mao, M.; Ranzato, M.; Senior, A.; Tucker, P.; Yang, K.; et al. 2012. Large scale distributed deep networks. In *Advances in neural information processing systems*, 1223–1231.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.-A.; Vincent, P.; and Bengio, S. 2010. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research* 11(Feb):625–660.

Girshick, R.; Donahue, J.; Darrell, T.; and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

He, X.; Pan, J.; Jin, O.; Xu, T.; Liu, B.; Xu, T.; Shi, Y.; Atallah, A.; Herbrich, R.; Bowers, S.; et al. 2014. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, 1–9. ACM.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Herlocker, J. L.; Konstan, J. A.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999*, 230–237. Association for Computing Machinery, Inc.

Hinton, G. E.; Osindero, S.; and Teh, Y.-W. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18(7):1527–1554.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Hornik, K. 1991. Approximation capabilities of multilayer feedforward networks. *Neural networks* 4(2):251–257.

Johnson, M.; Schuster, M.; Le, Q. V.; Krikun, M.; Wu, Y.; Chen, Z.; Thorat, N.; Viégas, F.; Wattenberg, M.; Corrado, G.; et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics* 5:339–351.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature* 521(7553):436.

LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; and Jackel, L. D. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1(4):541–551.

Linden, G.; Smith, B.; and York, J. 2003. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* (1):76–80.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518(7540):529.

Niu, M. Y.; Boixo, S.; Smelyanskiy, V. N.; and Neven, H. 2019. Universal quantum control through deep reinforcement learning. *npj Quantum Information* 5(1):33.

Peters, M. E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; and Zettlemoyer, L. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

Popova, M.; Isayev, O.; and Tropsha, A. 2018. Deep reinforcement learning for de novo drug design. *Science advances* 4(7):eaap7885.

Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

Sarwar, B. M.; Karypis, G.; Konstan, J. A.; Riedl, J.; et al. 2001. Item-based collaborative filtering recommendation algorithms. *Www* 1:285–295.

Schafer, J. B.; Konstan, J. A.; and Riedl, J. 2001. E-commerce recommendation applications. *Data mining and knowledge discovery* 5(1-2):115–153.

Segler, M. H.; Preuss, M.; and Waller, M. P. 2018. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature* 555(7698):604.

Shan, Y.; Hoens, T. R.; Jiao, J.; Wang, H.; Yu, D.; and Mao, J. 2016. Deep crossing: Web-scale modeling without manually crafted combinatorial features. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 255–262. ACM.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. 2016. Mastering the game of go with deep neural networks and tree search. *nature* 529(7587):484.

Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. 2017. Mastering the game of go without human knowledge. *Nature* 550(7676):354.

Spruyt, V. 2018. Loc2vec: Learning location embeddings with triplet-loss networks.

Su, J.; Vargas, D. V.; and Sakurai, K. 2019. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*.

Turing, A. M. 2009. Computing machinery and intelligence. In *Parsing the Turing Test*. Springer. 23–65.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.

Vinyals, O.; Babuschkin, I.; Chung, J.; Mathieu, M.; Jaderberg, M.; Czarnecki, W. M.; Dudzik, A.; Huang, A.; Georgiev, P.; Powell, R.; et al. 2019. Alphastar: Mastering the real-time strategy game starcraft ii. *DeepMind Blog*.

Wu, Y.; Schuster, M.; Chen, Z.; Le, Q. V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Yadav, A.; Chan, H.; Xin Jiang, A.; Xu, H.; Rice, E.; and Tambe, M. 2016. Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 740–748. International Foundation for Autonomous Agents and Multiagent Systems.

Zhou, G.; Zhu, X.; Song, C.; Fan, Y.; Zhu, H.; Ma, X.; Yan, Y.; Jin, J.; Li, H.; and Gai, K. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1059–1068. ACM.