PERSPECTIVES

COMPUTER SCIENCE

Special Topic: Machine Learning

Deep Learning for Natural Language Processing: Advantages and Challenges

Hang Li

Noah's Ark Lab, Huawei Techonologies, Hong Kong, China

Email: hangli65@hotmail.com

1. Introduction

Deep learning refers to machine learning technologies for learning and utilizing `deep' artificial neural networks, such as deep neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks (RNN). Recently deep learning has been successfully applied to natural language processing and significant progresses have been made. This paper summarizes the recent advancement of deep learning for natural language processing and discusses its advantages and challenges.

We think that there are five major tasks in natural language processing, including classification, matching, translation, structured prediction, and sequential decision process. For the first four tasks, it is found that the deep learning approach has outperformed or significantly outperformed the traditional approaches.

End-to-end training and representation learning are the key features of deep learning which make it a powerful tool for natural language processing. Deep learning is not almighty, however. It might not be sufficient for inference and decision making, which are essential for complex problems like multi-turn dialogue. Furthermore, how to combine symbolic processing and neural processing, how to deal with the long tail phenomenon, etc. are also challenges deep learning for natural language processing.

2. Progress in Natural Language Processing

In our view, there are five major tasks in natural language processing, namely classification, matching, translation, structured prediction, and sequential decision process. Most of the problems in natural language processing can be formalized as the five tasks, as summarized in Table 1. In the tasks, words, phrases, sentences, paragraphs, and even documents are usually viewed as sequence of tokens (strings), and treated similarly, although they have different complexities. In fact sentences are the most widely used processing units.

It is observed recently that deep learning can enhance the performances in the first four tasks and becomes the state-of-the-art technologies for the tasks (e.g., [1-8]).

Table 1. Five Tasks in Natural Language Processing

Task	Description	Model	Applications
Classification	assign a label to a string	$s \rightarrow c$ s : string, c : label	text classification, sentiment analysis
Matching	matching two strings	$s, t \rightarrow R^{+}$ s : string, t : string R^{+} : non-negative real values	search, question answering, single turn dialogue (retrieval based)
Translation	transform one string to another	$s \rightarrow t$ s : string, t : string	machine translation, automatic speech recognition, single turn dialogue (generation based)
structured prediction	map a string to a structure	$s \rightarrow [s]$ s : string, $[s]$: structure	named entity recognition, word segmentation, part-of-speech tagging, dependency parsing, semantic parsing
sequential decision process	take actions in states in dynamically changing environment	$\pi : s \rightarrow a$ π : policy, s : state, a : action	multi-turn dialogue

Table 2. Performances of Natural Language Processing Problems

Task	Example Problem	Deep Learning	Traditional Approach	Reference
classification	sentiment classification	CNN, acc = **86.8%**	SVM, acc = 79.4%	[1]
matching	single turn dialogue	CNN, p@1=**49.6%**	MLP, p@1=36.1%	[2]
translation	machine translation	NMT, BLEU=**39.0**	SMT, BLEU=37.0	[6]
structured prediction	dependency parsing	acc = **91.8%**	acc = 90.7%	[8]

Table 2 shows the performances of example problems in which deep learning has surpassed traditional approaches. Among all the NLP problems, the progress in machine translation is particularly remarkable. Neural machine translation, i.e., machine translation using deep learning, has significantly outperformed traditional statistical machine translation. The state-of-the art neural translation systems employ sequence-to-sequence learning models comprising RNNs [4-6].

Deep learning has also, for the first time, made certain applications possible. For example, deep learning is successfully applied to image retrieval (also known as text to image), in which query and image are first transformed into vector representations with CNNs, the representations are matched with DNN, and the relevance of image to query is calculated [3]. Deep learning is also employed in generation-based natural language dialogue, in which given an utterance the system automatically generates a response, and the model is trained in sequence-to-sequence learning [7].

The fifth task, sequential decision process such as Markov decision process, is the key issue in multi-turn dialogue, as explained below. It has not been thoroughly verified, however, how deep learning can contribute to the task.

3. Advantages and Challenges

Deep learning certainly has advantages and challenges when applied to natural language processing, as summarized in Table 3.

Table 3. Advantages and Challenges of Deep Learning for Natural Language Processing

Advantages	Challenges
• Good at pattern recognition problems • Data-driven, and performance is high in many problems • End-to-end training: little or no domain knowledge is needed in system construction • Learn of representations: cross-modal processing is possible • Gradient-based learning: learning algorithm is simple • Mainly supervised learning methods	• Not good at inference and decision making • Cannot directly handle symbols • Data-hungry and thus is not suitable when data size is small • Difficult to handle long tail phenomena • Model is usually a black box and is difficult to understand • Computational cost of learning is high • Unsupervised learning methods need to be developed • Still lacks of theoretical foundation

3-1. Advantages

We think that among the advantages, end-to-end training and representation learning really differentiate deep learning from traditional machine learning approaches, and make it a powerful machinery for natural language processing.

It is often possible to perform end-to-end training in deep learning for an application. This is because the model (deep neural network) offers rich representability and information in the data can be effectively `encoded' in the model. For example, in neural machine translation, the model is completely automatically constructed from a parallel corpus, and usually no human intervention is needed. This is clearly an advantage, compared to the traditional approach of statistical machine translation, in which feature engineering is crucial.

With deep learning, the representations of data in different forms, for example, text and image, can all be learned as real-valued vectors. This makes it possible to perform information processing across multiple modality. For example, in image retrieval, it becomes feasible to match the query (text) against images and find the most relevant images, because all of them are represented as vectors.

3-2. Challenges

There are challenges of deep learning that are more common, for example, lack of theoretical foundation, lack of interpretability of model, and requirement of large amount of data and powerful computing resources. There are also challenges that are more unique to natural language processing, namely, difficulty in dealing with long tail, incapability of directly handling symbols, and ineffectiveness at inference and decision making.

Data in natural language always follows a power law distribution. As a result, for example, the size of vocabulary increases as the size of data increases. That means that no matter how much data there is for training, there always exist cases for which the training data cannot cover. How to deal with the long tail problem poses a significant challenge to deep learning. By resorting to deep learning alone, this problem would be hard to solve.

Language data is by nature symbol data, which is different from vector data (real-valued vectors) that deep learning normally utilizes. Currently, symbol data in language is converted to vector data and then is input into neural networks, and the output from neural networks is further converted to symbol data. In fact, a large amount of knowledge for natural language processing is in the form of symbols, including linguistic knowledge (e.g., grammars), lexical knowledge (e.g., WordNet), and world knowledge (e.g., Wikipedia). Currently deep learning methods have not yet made effective use of the knowledge. Symbol representations are easy to interpret and manipulate, and on the other hand vector representations are robust to ambiguity and noise. How to combine symbol data and vector data and how to leverage the strengths of both data types remains an open question for natural language processing.

There are complex tasks in natural language processing, which may not be easily realized with deep learning alone. For example, multi-turn dialogue amounts to a very complicated process. It involves language understanding, language generation, dialogue management, knowledge base access, and inference. Dialogue management can be formalized as sequential decision process and reinforcement learning can play a critical role. Obviously, combination of deep learning and reinforcement learning could be potentially useful for the task, which is beyond deep learning itself.

In summary, there are still a number of open challenges with regard to deep learning for natural language processing. Deep learning, when combined with other technologies (reinforcement learning, inference, knowledge), may further push the frontier of the field.

Acknowledgement

This work is supported in part by the National Basic Research Program of China (2014CB340301).

References

1. P Blunsom, E Grefenstette and N Kalchbrenner. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore USA, 655-65, 2014.

2. B Hu, Z Lu and H Li *et al*. Advances in Neural Information Processing Systems 27, Montreal Canada, 2042-50, 2014.

3. L Ma, Z Lu and L Shang *et al*., Proceedings of International Conference on Computer Vision, Santiago, Chile, 2623-31, 2015.

4. K Cho, B Van Merriënboer and C Gulcehre *et al*. Proceedings of Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 1724-34, 2014.

5. D Bahdanau, K Cho and Y Bengio. Proceedings of the 3rd International Conference on Learning Representations, San Diago USA, 2015.
6. Y Wu, M Schuster and Z Chen *et al*. CoRR, vol. abs/1609.08144, 2016.
7. L Shang, Z Lu and H Li. ACL-IJCNLP'15, Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 1577-86, 2015.
8. D Chen and C D. Manning. Proceedings of Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 740-50, 2014.