PERSPECTIVES

INFORMATION SCIENCE

Special Topic: Machine Learning

Model-driven deep-learning

Zongben Xu* and Jian Sun*

Deep learning has been widely recognized as the representative advances of machine learning or artificial intelligence in general nowadays [1,2]. This can be attributed to the recent breakthroughs made by deep learning on a series of challenging applications. A deep-learning approach improves the accuracy rate of face recognition to be higher than 99%, beating the human level [3]. For speech recognition and machine translation, deep learning is approaching the performance level of a simultaneous interpreter [4]. For the game of 'go', it successfully beats the human world champion [5]. For diagnosis of some specific diseases, it has matched the level of medium or senior professional physicians [6]. Until now, it has been hard to find areas in which the deep-learning technique has not been tried in their respective tasks.

One can observe that these breakthroughs always take place in large IT companies or specialized R&D institutes, such as Google, Microsoft, Facebook, etc. This is because deep-learning applications require some prerequisites, such as a huge volume of labeled data, sufficient computational resources and the engineering experiences in determining the network topology, including the number of layers, number of neurons per layer and non-linear transforms of neurons. Due to these prerequisites, it requires sufficient knowledge and engineering experience in neural network design, and takes a long time in accumulating and labeling data. Professional IT companies and specialized R&D institutions can obviously match these requirements.

Figure 1. Model-driven deep-learning approach.

With the arrival of the big data era, data requirements are gradually no longer an obstacle (at least for many areas), but the determination of network topology is still a bottleneck. This is mainly due to the lack of theoretical understandings of the relationship between the network topology and performance. In the current state, the selection of network topology is still an engineering practice instead of scientific research, leading to the fact that most of the existing deep-learning approaches lack theoretical foundations. The difficulties in network design and its interpretation, and a lack of understanding in its generalization ability are the common limitations of the deep-learning approach. These limitations may prevent its widespread use in the trends of 'standardization, commercialization' of machine learning and artificial intelligence technology.

A natural question is whether we can design network topology with theoretical foundations, and make the network structure explainable and predictable. We believe that it is possible to provide a positive answer to this question through combing the model-driven approach and data-driven deep-learning approach. Here we take the deep-learning approach as a data-driven approach because it uses a standard network architecture as a black box, heavily relying on huge data to train the black box. In contrast, the model-driven approach here refers to the method using a model (e.g. a loss function) constructed based on the objective, physical mechanism and domain knowledge for a specific task. A prominent feature of the model-driven approach is that, when the model is sufficiently accurate, the solution can be generally expected to be optimal, and the minimization algorithm is commonly deterministic. A fatal flaw of the model-driven approach lies in the difficulty in accurately modeling for a specific task in real applications, and sometimes the pursuit of accurate modeling is a luxury expectation. In recent years, we have studied and implemented a series of model-driven deep-learning methods [7–10] combining the modeling-based and deep-learning-based approaches, which showed their feasibilities and effectiveness in real applications.

Given a specific task, the basic procedures of our model-driven deep-learning method are shown in Fig. 1 and explained as follows:

(1) *A model family* is first constructed based on the task backgrounds (e.g. objective, physical mechanism and

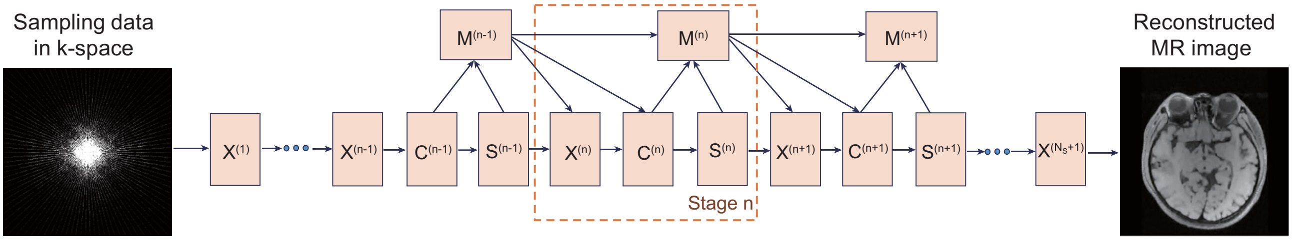

Figure 2. Topology of ADMM-Net [7]: given under-sampled *k*-space data, it outputs the reconstructed MRI image after *T* stages of processing.

prior knowledge). The model family is a family of functions with a large set of unknown parameters, amounting to the hypothesis space in machine learning. Differently from the accurate model in the model-driven approach, this model family only provides a very rough and broad definition of the solution space. It has the advantage of a model-driven approach but greatly reduces the pressure of accurate modeling.

(2) *An algorithm family* is then designed for solving the model family and the convergence theory of the algorithm family is established. The algorithm family refers to the algorithm with unknown parameters for minimizing the model family in the function space. The convergence theory should include the convergence rate estimation and the constraints on the parameters that assure the convergence of the algorithm family.

(3) The algorithm family is unfolded to *a deep network* with which parameter learning is performed as a deep-learning approach. The depth of the network is determined by the convergence rate estimation of the algorithm family. The parameter space of the deep network is determined by the parameter constraints. All the parameters of the algorithm family are learnable. In this way, the topology of the deep network is determined by the algorithm family, and the deep network can be trained through back-propagation.

Taking [7] as an example, we apply the above model-driven deep-learning approach to compressive sensing magnetic resonance imaging (CS-MRI), i.e. recovering the high-quality MR image using sub-sampled *k*-space data lower than the Nyquist rate. The model family is defined as:

$$\hat{x} = \arg\min_{x} \left\{ \frac{1}{2} \| Ax - y \|_2^2 + \sum\nolimits_{l=1}^{L} \lambda_l g(D_l x) \right\}, \quad (1)$$

where $A = PF$ is the measurement matrix, P is the sampling matrix, F is the Fourier transform matrix, D_l is linear transform for convolution, $g(\cdot)$ is the regularization function, λ_l is the regularization parameter and L is the number of linear transforms. All the parameters of (D_l, g, λ_l, L) are unknown and reflect the uncertainty in modeling (notice that these parameters are known and fixed in traditional CS-MRI models). According to the ADMM (Alternating Direction Method of Multipliers) method, the algorithm family for solving the model family can be designated as:

$$\begin{cases} x^{(n)} = F^T \left(P^T P + \sum_l \rho_l F D_l^T D_l F^T \right)^{-1} \\ \times \left[P^T y + \sum_l \rho_l F D_l^T \left(z_l^{(n-1)} + \beta_l^{(n-1)} \right) \right], \\ z_l^{(n)} = S\left(D_l x^{(n)} + \beta_l^{(n-1)}; \frac{\lambda_l}{\rho_l} \right) \\ \beta_l^{(n)} = \beta_l^{(n-1)} + \eta_l \left(D_l x^{(n)} - z_l^{(n)} \right) \end{cases} \quad (2)$$

where $S(\cdot)$ is a non-linear transform relating to $g(\cdot)$. According to the ADMM convergence theory, this algorithm is linearly convergent. By unfolding the algorithm family to a deep network, we design an ADMM-Net composed of T successive stages, as shown in Fig. 2. Each stage consists of a reconstruction layer (R), a convolution layer (C), a non-linear transform layer (Z) and a multiplier update layer (M). We learn the parameters of $(S, D_l, \lambda_l, \rho_l, \eta_l)$ using a back-propagation algorithm. In [7], we reported the state-of-the-art CS-MRI results using this model-driven deep-learning method.

The above model-driven deep-learning approach obviously retains the advantages (i.e. determinacy and theoretical soundness) of the model-driven approach, and avoids the requirement for accurate modeling. It also retains the powerful learning ability of the deep-learning approach, and overcomes the difficulties in network topology selection. This makes the deep-learning approach designable and predictable, and it balances well versatility and pertinence in real applications.

We point out that the model-driven approach and data-driven approach are not opposed to each other. If the model is accurate, it provides the essential description of the problem solutions, from which infinite ideal samples can be generated, and vice versa: when the sufficient samples are provided, the model of the problem is fully (but in discretized form) represented. This is the essential reason for the effectiveness of the model-driven deep-learning approach.

Please refer to [2,8] for the previous investigations of the model-driven deep-learning approach. The recent advances can be found in [7,9–11]. Most of these successful applications lie in the inverse problems in imaging sciences, for which there exists domain knowledge that can be well modeled in the model family. We believe that this model-driven deep-learning approach can be widely applied to the applications where we can design the model family by incorporating domain knowledge and then the deep architecture can be correspondingly designed following the above procedures.

Zongben Xu* and Jian Sun*
Xi'an International Academy for Mathematics & Mathematical Technology, Xi'an Jiaotong University, China
*Corresponding authors.
E-mails: zbxu@mail.xjtu.edu.cn; jiansun@mail.xjtu.edu.cn

REFERENCES

1. LeCun Y, Bengio Y and Hinton G. *Nature* 2015; **521**: 436–44.
2. Gregor K and LeCun Y. *ICML* 2010.
3. Schroff F, Kalenichenko D and Philbin J. *CVPR* 2015.
4. Yonghui W, Schuster M and Zhifeng Chen *et al.* arXiv:1609.08144, 2016.
5. Silver D, Aja Huang and Chris J. Maddison *et al.* *Nature* 2016; **529**: 484–9.
6. Gulshan V, Peng L and Coram M *et al.* *Jama* 2016; **316**: 2402–10.
7. Yang Y, Sun J and Li H *et al.* *NIPS* 2016.
8. Sun J and Tappen M. *CVPR* 2011.
9. Sun J and Tappen M. *IEEE T Image Process* 2013; **22**: 402–8.
10. Sun J, Sun J and Xu Z. *IEEE T Image Process* 2015; **24**: 4148–59.
11. Sprechmann P, Bronstein AM and Sapiro G. *IEEE TPAMI* 2015; **37**: 1821–33.

National Science Review
5: 22–24, 2018
doi: 10.1093/nsr/nwx099
Advance access publication 25 August 2017

COMPUTER SCIENCE

Special Topic: Machine Learning

Deep learning for natural language processing: advantages and challenges

Hang Li

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 progress has 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 the 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 that 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 of deep learning for natural language processing.

PROGRESS IN NATURAL LANGUAGE PROCESSING

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

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

Table 2 shows the performances of example problems in which deep learning has surpassed traditional approaches. Among all the NLP problems, 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 has been 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 the image to the 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