Book Review

Deep Learning Approaches to Text Production

Shashi Narayan and Claire Gardent
(University of Edinburgh; CNRS/LORIA, Nancy)

Reviewed by
Yue Zhang
Westlake University, Westlake Institute for Advanced Study

Text production (Reiter and Dale 2000; Gatt and Krahmer 2018) is also referred to as *natural language generation* (NLG). It is a subtask of natural language processing focusing on the generation of natural language text. Although as important as natural language understanding for communication, NLG had received relatively less research attention. Recently, the rise of deep learning techniques has led to a surge of research interest in text production, both in general and for specific applications such as text summarization and dialogue systems. Deep learning allows NLG models to be constructed based on neural representations, thereby enabling end-to-end NLG systems to replace traditional pipeline approaches, which frees us from tedious engineering efforts and improves the output quality. In particular, a neural encoder-decoder structure (Cho et al. 2014; Sutskever, Vinyals, and Le 2014) has been widely used as a basic framework, which computes input representations using a neural *encoder*, according to which a text sequence is generated token by token using a neural *decoder*. Very recently, pre-training techniques (Broscheit et al. 2010; Radford 2018; Devlin et al. 2019) have further allowed neural models to collect knowledge from large raw text data, further improving the quality of both encoding and decoding.

This book introduces the fundamentals of neural text production, discussing both the mostly investigated tasks and the foundational neural methods. NLG tasks with different types of inputs are introduced, and benchmark datasets are discussed in detail. The encoder-decoder architecture is introduced together with basic neural network components such as convolutional neural network (CNN) (Kim 2014) and recurrent neural network (RNN) (Cho et al. 2014). Elaborations are given on the encoder, the decoder, and task-specific optimization techniques. A contrast is made between the neural solution and traditional solutions to the task. Toward the end of the book, more recent techniques such as self-attention networks (Vaswani et al. 2017) and pre-training are briefly discussed. Throughout the book, figures are given to facilitate understanding and references are provided to enable further reading.

Chapter 1 introduces the task of text production, discussing three typical input settings, namely, generation from meaning representations (MR; i.e., *realization*), generation from data (i.e., *data-to-text*), and generation from text (i.e., *text-to-text*). At the end of the chapter, a book outline is given, and the scope, coverage, and notation convention

are briefly discussed. I enjoyed the examples and figures demonstrating the typical NLG tasks such as abstract meaning representation (AMR) to text generation (May and Priyadarshi 2017), the E2E dialogue task (Li et al. 2018), and the data-to-text examples. It would have been useful if more examples had been given for some other typical tasks such as summarization and sentence compression, despite the fact that they are intuitively understandable without examples and are discussed later in the book. I find Section 1.3 particularly useful for understanding the scope of the book.

Chapter 2 briefly summarizes pre-neural approaches to text production. It begins with data-to-text generation, where important components for a traditional pipeline, such as content selection, document planning, lexicalization, and surface realization, are discussed. Then it moves on to discuss the MR-to-text generation, for which two major approaches are discussed. The first approach is grammar-centric, where rules are used as a basis and much care is taken for pruning a large search space. The second approach is statistical, where features are used to score candidate outputs. Finally, the chapter discusses text-to-text generation, introducing major techniques for sentence simplification, sentence compression, sentence paraphrasing, and document summarization. This chapter presents a rich literature review on text-to-text methods, which can be helpful. It would have been useful if more references had been given to data-to-text methods, such as modular approaches and integrated approaches for implementing the pipeline.

Chapter 3 discusses the foundational neural model—a basic encoder-decoder framework for text generation. It consists of three main sections. The first section introduces the basic elements of deep learning, discussing feed-forward neural networks, CNNs, RNNs, and their variants LSTM (Hochreiter and Schmidhuber 1997) and GRU (Cho et al. 2014). It also briefly discusses word embeddings (i.e., word2vec [Mikolov et al. 2013] and GloVe [Pennington, Socher, and Manning 2014]) and contextualized embeddings (i.e., ELMo [Peters et al. 2018], BERT [Devlin et al. 2019], and GPT [Radford 2018]). The second section introduces the encoder-decoder framework using a bidirectional RNN encoder and a simple RNN decoder. Training and decoding issues are also discussed, including training techniques for neural networks in general. The final section makes a comparison between pre-neural approaches and neural approaches, highlighting robustness and freedom from feature engineering as two major advantages of the latter, while also discussing their potential limitations. This chapter is rich in figures and references, which helps the reader understand the big picture. On the other hand, it can be difficult for beginners to fully absorb, and they should refer to the reference materials such as the Goodfellow book on deep learning (Goodfellow, Bengio, and Courville 2016) cited at the beginning of Section 3.1 for further reading.

Chapters 4 to 6 form the central part of this book. They discuss major techniques for improving the decoding module, the encoding module, and for integrating task-specific objectives, respectively. Chapter 4 begins with a survey of seminal work using encoder-decoder modeling for text-to-text (i.e., machine translation and summarization), MR-to-text, and data-to-text tasks, and then lays out four main issues, namely, accuracy, repetitions, coverage, and rare/unknown words. It devotes three sections to introducing major solutions to these issues, which include attention (Bahdanau, Cho, and Bengio 2015), copy (Vinyals, Fortunato, and Jaitly 2015), and coverage (Tu et al. 2016) mechanisms. For each method, similar or alternative approaches are also discussed. The chapter gives a concise introduction to these techniques, which are essential to know in the neural NLG literature. Though using RNN as a base model, these techniques are also useful for self-attention networks.

Chapter 5 discusses how to deal with long text and graph-structured data. It begins with a review of methods using the standard encoder-decoder structure for encoding

documents and linearized graphs (e.g., AMR, RDF triples, dialogue moves, and Infoboxes in Wikipedia), showing the main limitation: lack of structural information and weakness in capturing long-range dependencies. It then spends a section discussing typical models for long-text structures, which include hierarchical network structures using RNNs and CNNs for modeling both word-sentence structures and sentence-document structures, and collaborative modeling of paragraphs for representing documents. The final section of the chapter discusses the modeling of graph structures using graph LSTMs (Song et al. 2018) and GCNs (Bastings et al. 2017). Techniques discussed in this section receive much more attention in current NLG research.

Chapter 6 discusses techniques for integrating task-specific communication goals such as summarizing a text and generating a user-specific response in dialogue. To this end, two types of methods are introduced. The first focuses on augmenting the encoder-decoder architecture with task-specific features, and the second focuses on augmenting the training objective with task-specific metrics. The chapter consists of three main sections. The first section discusses content selection in the encoder module for summarization. Several representative models are detailed while a range of other models are surveyed briefly. The second section discusses reinforcement learning, describing a general algorithm of policy gradient and its applications in many tasks with different reward functions. The third section discusses user modeling in neural conversational models. I find the reinforcement learning section particularly informative. For example, the case study demonstrating the disadvantage of cross-entropy loss for extractive summarization is insightful.

Chapter 7 describes the most prominent datasets used in neural text production research. It is organized in three main sections, which focus on data-to-text generation, MR-to-text generation, and text-to-text generation, respectively. The origin, size, data source, format, and other characteristics are given for each dataset, and examples are shown in figures. This chapter covers a range of datasets, including most benchmarks that I am aware of and also some I am unfamiliar with. It can be highly useful for researchers and students as a reference, adding much to the value of the book.

Chapter 8 summarizes the book, reviewing the main techniques and discussing the remaining issues and challenges, before mentioning recent trends. In particular, the authors identify semantic adequacy and explainability as two major issues with neural NLG, highlighting the limitation of existing evaluation methods. Additionally, they raise three main challenges, namely, long inputs and outputs, cross-domain and cross-lingual transfer learning, and knowledge integration. Finally, Transformer (Vaswani et al. 2017) and pre-training are briefly discussed as recent trends.

Overall, this book presents a succinct review of the most prominent techniques in foundational neural NLG. It can serve as a great introductory book to the field for the NLP research community and NLP engineers with basic relevant background. It features rich reference materials and figures. Although I enjoyed reading its content, I feel that it would have been more valuable if Transformer and pre-training had been elaborated in more detail, with relevant literature surveys being included, since they are the dominant methods in the current literature. Given the fast-moving pace of the research field, maybe subsequent editions will meet such expectations.

References

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA.

Bastings, Jasmijn, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. 2017. Graph convolutional encoders for

syntax-aware neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1957–1967, Copenhagen. DOI: `https://doi.org/10.18653/v1/D17 -1209`

Broscheit, Samuel, Massimo Poesio, Simone Paolo Ponzetto, Kepa Joseba Rodriguez, Lorenza Romano, Olga Uryupina, Yannick Versley, and Roberto Zanoli. 2010. BART: A multilingual anaphora resolution system. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 104–107, Uppsala.

Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder– decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha. DOI: `https://doi.org/10.3115 /v1/D14-1179`

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, MN.

Gatt, Albert and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170. DOI: `https://doi.org/10.1613/jair .5477`

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`, DOI: `https://doi.org/10.1162/neco.1997.9 .8.1735`, PMID: 9377276

Hochreiter, Sepp and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780. DOI: `https://doi.org/10.1162/neco.1997 .9.8.1735`, PMID: 9377276

Kim, Yoon. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha. DOI: `https://doi.org/10.3115 /v1/D14-1181`

Li, Xiujun, Sarah Panda, Jingjing Liu, and Jianfeng Gao. 2018. Microsoft dialogue challenge: Building end-to-end task-completion dialogue systems. *arXiv preprint arXiv:1807.11125*.

May, Jonathan and Jay Priyadarshi. 2017. SemEval-2017 task 9: Abstract Meaning Representation parsing and generation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 536–545, Vancouver. DOI: `https://doi.org/10.18653/v1 /S17-2090`

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, Lake Tahoe, NV.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543, Doha. DOI: `https://doi.org/10.3115/v1/D14 -1162`

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2227–2237, New Orleans, LA. DOI: `https://doi.org/10 .18653/v1/N18-1202`

Radford, Alec. 2018. Improving language understanding by generative pre-training.

Reiter, Ehud and Robert Dale. 2000. *Building Natural Language Generation Systems*. Studies in Natural Language Processing. Cambridge University Press.

Song, Linfeng, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. A graph-to-sequence model for AMR-to-text generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1616–1626, Melbourne. DOI: `https://doi.org/10.18653/v1/P18 -1150`, PMID: 30540182

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112, Montreal.

Tu, Zhaopeng, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. 2016. Modeling coverage for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 76–85, Berlin. DOI: `https://doi.org/10.18653/v1/P16-1008`, PMCID: PMC5358015

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., pages 5998–6008.

Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., pages 2692–2700.

Yue Zhang is an associate professor of Computer Science and Technology at Westlake University. His main research goal is to investigate robust open domain human language understanding and synthesis technologies, together with their downstream applications. Yue Zhang's e-mail is: `yue.zhang@wias.org.cn`.