

Neural Network Methods for Natural Language Processing

Yoav Goldberg

(Bar Ilan University)

Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 37), 2017, xxii+287 pp; paperback, ISBN 9781627052986, \$74.95; ebook, ISBN 9781627052955, \$59.96; doi:10.2200/S00762ED1V01Y201703HLT037

Reviewed by

Yang Liu and Meng Zhang

Tsinghua University

Deep learning has attracted dramatic attention in recent years, both in academia and industry. The popular term *deep learning* generally refers to neural network methods. Indeed, many core ideas and methods were born years ago in the era of “shallow” neural networks. However, recent development of computation resources and accumulation of data, and of course new algorithmic techniques, has enabled this branch of machine learning to dominate many areas of artificial intelligence, first for perception tasks like speech recognition and computer vision, and gradually for natural language processing (NLP) since around 2013.

Natural language is an intricate object for computers to handle. Philosophical debates aside, the field of NLP has witnessed a paradigm shift from rule-based methods to statistical approaches, which have been dominant since the 1990s. Following this background, deep learning goes further down the statistical route, and gradually becomes the *de facto* technique of the mainstream statistical landscape.

This book covers the two exciting topics of neural networks and natural language processing. More specifically, it focuses on how neural network methods are applied on natural language data. With this guideline, the structure of the book appears smoother from a neural network entry: It first lays the background of neural network methods, and then discusses the traits of natural language data, including challenges to address and sources of information that we can exploit, so that specialized neural network models introduced later are designed in ways that accommodate natural language data. On the other hand, some fundamentals in natural language processing are not covered in the book, for example, linguistic theories and backgrounds of the natural language processing tasks, and proper preparation of corpus data. Based on this structure, the book is intended for practitioners from both deep learning and natural language processing to have a common ground and a shared understanding of what has been achieved at the intersection of these two fields. NLP practitioners can become well armed with the neural network tools to work on their natural language data, whereas neural network practitioners may feel that the content of the book is a bit light, although sufficient and effective enough for an entry into working with natural language data.

After the first, introductory chapter, the book is divided into four parts that roughly follow the structure of the book mentioned above.

1. Part I: Supervised Classification and Feed-Forward Neural Networks

This part introduces the basic machinery of neural networks, and contains four chapters. Chapter 2 provides the background of supervised machine learning, including concepts like parameterized functions, train, test, and validation sets, training as optimization, and, in particular, the use of gradient-based methods for optimization. Readers familiar with machine learning may safely skip this chapter. The models presented in this chapter are linear and log-linear models. Their limitations are discussed in Chapter 3, which motivates the need for nonlinear models, and sets the backdrop for the introduction of feed-forward neural networks presented in Chapter 4. Finally, Chapter 5 discusses the training of neural networks. Unlike most presentations from other sources, this chapter comes with a more algorithmic than mathematical flavor, by presenting computation graph abstraction as well as related software. It also provides a handy subsection that discusses practical choices for training neural networks.

As mentioned earlier, neural network practitioners may feel that the neural network content of the book is a bit light, and this part can be almost entirely skipped by these readers. However, for people coming from more traditional branches of statistical learning, Chapter 5 is still well worth reading.

2. Part II: Working with Natural Language Data

This part discusses the traits of natural language data, the object to which we would like to apply neural networks. There are seven chapters in this part. Chapter 6 presents a categorization of natural language classification problems and discusses the information sources that we can exploit in natural language data. Chapter 7 provides concrete examples of natural language features for solving various NLP tasks. These two chapters are probably quite dense for people coming from machine learning, and they serve to prepare them with the familiarity needed to work with natural language data. Chapter 8 is where neural networks come in; the chapter discusses how to represent textual features as inputs for neural network models. Chapter 9 describes the language modeling task and discusses the feed-forward neural language model. The neural language model also produces the byproduct of word representations, which form the subject of Chapters 10 and 11. In particular, Chapter 10 presents approaches to learning word representations, and Chapter 11 discusses the usage of word representations outside the context of neural networks, like word similarity and word analogies. Chapter 12 is an independent chapter that describes a specific feed-forward neural network architecture for the task of natural language inference.

This part of the book, especially Chapter 8, which connects neural networks with natural language data, is the core of the content that distinguishes this book from other materials that cover either neural networks or natural language processing.

3. Part III: Specialized Architectures

This part is composed of five chapters that introduce the specialized architectures of convolutional neural networks (CNNs) (Chapter 13) and recurrent neural networks (RNNs) (Chapters 14–17). Chapter 13 mainly introduces 1D CNNs, which are specialized at learning ngram patterns. Chapter 14 describes the modeling of sequences and stacks with recurrent neural networks in an abstract way. This is to be made concrete by the succeeding two chapters. In Chapter 15, concrete instantiations of RNNs like the Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU) are described,

and in Chapter 16, concrete applications of modeling with the RNN abstraction to NLP tasks are presented, including sentiment classification, grammaticality detection, part-of-speech tagging, document classification, and dependency parsing. Chapter 17 also includes concrete applications of RNNs, but these tasks involve generating natural language, which are usually modeled with a conditioned RNN language model. The most typical example of these tasks is probably machine translation.

As the distribution of the chapters suggests, recurrent neural networks clearly receive more emphases. Indeed, RNNs alleviate the reliance on the Markov assumption and have the potential to model very long sequences. Their capabilities have led to breakthroughs in various sequence processing tasks, making them the celebrated models in research frontiers with proven performance.

4. Part IV: Additional Topics

This part contains four chapters that are relatively independent. Chapter 18 presents recursive neural networks for modeling trees. The capability of modeling trees is important for natural language because of its hierarchical structure. Chapter 19 is devoted to structured prediction, because certain NLP tasks like named entity recognition can be cast in this framework. Chapter 20 discusses multi-task learning and semi-supervised learning. These approaches have not yet grown into a full-fledged stage, but are still important topics for research and offer helpful techniques for many tasks.

The final chapter, Chapter 21, briefly reviews the content presented in the book, and discusses challenges that are yet to be addressed.

5. Summary

The application of neural networks to natural language processing has revolutionized this long-standing research field, pushing forward the state of the art of many tasks. Nonetheless, the goal of equipping computers with human language capability is still far from solved, and the field continues to develop at a fast pace. This book provides valuable materials for newcomers into this exciting arena of cross-disciplinary research, by preparing relevant information of both neural networks and natural language processing. The book mainly presents mature neural network approaches to natural language processing, because it is hardly possible for a book to keep up to date with such fast development—although at 287 pages, the book is already quite long compared with other books in the synthesis lectures series, which are usually monographs of 50 to 150 pages.

Yang Liu is an Associate Professor in the Department of Computer Science and Technology at Tsinghua University. His research interests are in machine learning approaches to natural language processing, particularly machine translation. Liu's e-mail address is liuyang2011@tsinghua.edu.cn. *Meng Zhang* is a fifth-year Ph.D. student in the Department of Computer Science and Technology at Tsinghua University. His primary research interest is deep learning-based natural language processing. Zhang's e-mail address is zmlarry@foxmail.com.