Book Review

Natural language processing (NLP) has become one of the most dynamic and fast developing interdisciplinary research fields over the past decade, which highlights the robustness of computational intelligence. Studies central to NLP modelling have gone through three main stages, specifically, rationalism, empiricism and deep learning. As a core part of machine learning, particularly deep learning, finding good representations with neural networks for natural language is an indispensable component in the paradigmatic shift which moves from computational science towards data-intensive scientific discovery (see, e.g., Szalay and Gray 2006). In this sense, White et al.'s recent work *Neural Representations of Natural Language* in the *Studies in Computational Intelligence* series appears to be an interesting publication responding to the needs of researchers who wish to master the techniques of representing different structures in natural language. The volume is a practical introduction to natural language engineering, which elaborates on the development of neural networks in NLP and focuses specifically on word representations, word sense representations and sentence representations. A variety of cutting-edge topics related to language modelling are addressed, which would appeal to a large audience such as NLP researchers, applied linguistics researchers, computer programmers and even general readers showing an interest in artificial intelligence. Notwithstanding the broad spectrum of readership, the volume, as the authors note, is essentially practice-oriented and might require readers to associate the contents of the book with extensive materials (particularly online resources). In addition, potential readers are advised to obtain basic knowledge of linguistics, statistics and computer programming in order to achieve a better understanding of the book.

The volume under review is composed of six chapters, which can be divided into two major parts. The first part (Chapters 1 and 2) presents an introduction to neural networks (including recurrent networks), while the second part (Chapters 3 to 5) deals chiefly with representations of different levels. Chapter 1 opens with the introduction to entry knowledge of neural networks required for readers, covering a wide range of machine learning techniques. A neural network is essentially a 'machine learning algorithm' (p. 2) and is constructed to 'represent the transformation of the input to the output as links between neurons' (p. 3) in a multi-layered architecture, which can be used to approximate functions (Sonada and Murata 2017) by employing hidden layers. A neural network normally includes parameters (e.g., weights and biases) and hyper-parameters (e.g., activation functions and training methods). The authors delineate the mechanism of activation functions such as identity, sigmoid, softmax, tanh and Rectified Linear Unit and explain in detail the different methods of training networks. Well-crafted examples of neural network architectures are also provided with a particular emphasis on classifier and bottlenecking autoencoder. As the authors note, this chapter aims to set out the context for the remaining contents of the book, and therefore, it might not be so valuable for experienced NLP researchers.

Chapter 2 narrows down its topic to the introduction of recurrent neural networks (RNNs), a family of deep learning algorithms. In general, an RNN is 'a chain of feed-forward neural networks, each one being identical in terms of their weight and bias parameters' (p. 24) and is a type

CrossMark

of recursive neural network, in which all the recurrent units (RUs) are connected in chains (see, e.g., Goodfellow, Bengio and Courville 2016). White et al. introduce four RNN structures (i.e., matched sequence, encoder, decoder, and encoder–decoder) and delineate different types of RUs in terms of basic RU (e.g., Elman RU and Jordan RU), the gated RU (GRU) and the long short-term memory (LSTM) RU. A number of deep algorithms are also illustrated, such as stacked RNNs and bidirectional RNNs. The 'depth' in such algorithms can be twofold, that is, the depth in sequence progression and the depth in input and output at every time step, with the former being determined by the length of input sequence and the latter by the quantity of 'chains'. Such RNNs are now widely applied to addressing core issues in NLP, such as speech recognition, speech synthesis and language modelling. In particular, as one of the mainstream algorithms in machine translation, RNNs can help language engineers to construct neural machine translation models that are distinguished from statistical machine translation. For instance, Sutskever et al. (2014) examined the validity of end-to-end learning LSTM in French-to-English translations; Kalchbrenner and Blunsom (2013) suggest the combination of convolutional n-gram models and RNNs in machine translation.

Focusing on word embeddings (i.e., vector representations of words, p. 38), Chapter 3 addresses the issue of creating numerical vectors to capture word features. White et al. first discuss representations for language modelling by reviewing the development of the neural probabilistic language model, delineating the rationales of simplified neural trigram language models (with input or/and output embeddings), Bayes-like reformulation and n-gram models. More sophisticated RNN models such as LSTM and GRU-based networks are also briefly introduced in this section. Then, they analyse acausal language modelling prior to the advent of word2vec, with a particular focus on the continuous bag of words (CBOW) model, the skip-grams model and analogy tasks. With regard to the defects of direct CBOW/skip-grams model training, White et al. emphasise the implementation of global vectors and conclude that 'neural predictive co-location models are functionally very similar to matrix factorisation of co-location counts with suitable weightings, and suitable similarity metrics' (p. 55). In addition, they introduce two other methods, that is, hierarchical softmax and negative sampling, explaining that they 'allow for large speed-up in any task which involves outputting very large classification probabilities' (p. 56). At the end of this chapter, White et al. discuss the alignment of vector spaces across different languages and outline, with examples, recent literature on training embeddings from different angles (e.g., modalities).

Chapter 4 deals with the techniques for representing a word's different meanings in various contexts, since polysemy is a ubiquitous phenomenon in most languages. Following a brief introduction to some key notions pertinent to word sense embedding (e.g., part of speech, lemma, lexeme, synset and gloss), White et al. review several analyses regarding word sense disambiguation and provide a variety of word processing techniques which specifically include lemmatisation, unlemmatisation and semantic syllepsis. With an extensive account of the most frequent sense method, they emphasise the importance of creating vector representations for word senses and evaluate two methods, namely, the directly supervised method and the word embedding-based disambiguation method. Next, White et al. discuss word sense induction systems that are designed to 'discover the word senses at the same time as they find their representations' without 'reference to a standard set of senses' (p. 82). They make a distinction between context clustering-based approaches (e.g., offline clustering and online clustering methods) and co-location prediction-based approaches (e.g., the expectation maximisation method). According to the authors, word sense representations are very useful for addressing core technical issues in machine translation, and they can also be used for processing languages where polysemy and homonymy are common.

Chapter 5 extends the topic of capturing and manipulating representations by looking at larger structures on phrasal and sentential levels in natural language, which, according to White et al., is an important yet tough task in artificial intelligence. In this chapter, four major categories of

such representations are generalised, specifically, unordered and weakly ordered representations, sequential models, structured models and matrix vector models. When discussing unordered and weakly ordered representations, White et al. note the importance of word order and introduce two models, namely, sums of word embeddings (the product of bag of words with an embedding matrix, p. 95) and paragraph vector models (e.g., the paragraph vector distributed memory model and the paragraph vector distributed bag of words model, see Le and Mikolov 2014). In sequential models such as the variational autoencoder model and encoder–decoder model and the skip-thought model, an 'RNN learns a representation of all its input and output in its state' and we can use 'RNN encoders and decoders' to 'generate representations of sequences by extracting a coding layer' (p. 97). In comparison with sequential models, structured models take account of parsing and incorporate tree/graph structures such as recursive neural networks and recursive autoen-coders (p. 101). A number of examples are offered in this context, including constituency parse trees, dependency trees and so forth. This chapter ends with the introduction to more sophisti-cated matrix vector models (i.e., structured matrix vector models and sequential matrix vector models). The final chapter wraps up the book with a brief summary of its contents and outlines the potential research directions in relevant future studies.

In respect to what this volume can offer NLP engineers, corpus linguists, as well as quanti-tative and computational linguistics researchers, a number of major strengths of the volume are worth noting. First of all, the book is technically sound and offers evaluations for the strengths and limitations of recent contributions to natural language engineering. A wide range of popular deep neural networks for NLP is covered with an admirably succinct account of both theoretical attestation and practical application, thus showing readers the state of the art in this discipline and providing future research directions central to deep learning. In addition, novel observations in support of or against available neural representation techniques are illustrated, and critical com-ments are well presented with regard to language modelling and model training. In this sense, such a thought-provoking guidebook is particularly helpful for novice language engineers, computer programmers and even general language researchers who attempt to embark on NLP studies. Second, the volume is compiled logically with precise academic English (although we note a few typos and grammatical inconsistencies that have been missed in the editing process) and organised in a manner that makes it accessible to most language engineers and researchers. The introduction to elementary neural network knowledge and the practical language modelling on lexical, seman-tic and phrasal levels are appropriately portioned, introducing readers to new ideas that appear promising or might stimulate them to develop promising alternatives. In particular, 'grey boxes' are used, as the authors note, to delineate unfamiliar notations, highlight technical issues central to neural networks and offer referencing resources, all of which will help readers achieve a better understanding about the book's contents.

Overall, this book makes a strong academic contribution to the development of both NLP and computational linguistics, and this, coupled with the wide potential readership previously mentioned, makes it worthy of our enthusiastic recommendation.

Haoda Feng
School of Foreign Languages, Bohai University, 19 Keji Road, Songshan District, Jinzhou, Liaoning Province, China
Email: leblanc.feng@yahoo.co.nz

References

Goodfellow I., **Bengio Y.** and **Courville A.** (2016). *Deep learning*. Cambridge, MA: MIT Press.

Kalchbrenner N. and **Blunsom P** (2013). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1700–1709.

Le Q. and **Mikolov T.** (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1188–1196.

Sonada S. and **Murata N.** (2017). Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis* **43**(2), 233–268.

Sutskever I., **Vinyals O.** and **Le Q.V**. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pp. 3104–3112.

Szalay A. and **Gray J.** (2006). Science in an exponential world. *Nature* **440**, 413–414.