Book Reviews

Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning

Mohammad Taher Pilehvar and Jose Camacho-Collados
(Tehran Institute for Advanced Studies & Cardiff University)

Reviewed by
Marcos Garcia
CiTIUS, University of Santiago de Compostela

Word vector representations have a long tradition in several research fields, such as cognitive science or computational linguistics. They have been used to represent the meaning of various units of natural languages, including, among others, words, phrases, and sentences. Before the *deep learning tsunami*, count-based vector space models had been successfully used in computational linguistics to represent the semantics of natural languages. However, the rise of neural networks in NLP popularized the use of word embeddings, which are now applied as pre-trained vectors in most machine learning architectures.

This book, written by Mohammad Taher Pilehvar and Jose Camacho-Collados, provides a comprehensive and easy-to-read review of the theory and advances in vector models for NLP, focusing specially on semantic representations and their applications. It is a great introduction to different types of embeddings and the background and motivations behind them. In this sense, the authors adequately present the most relevant concepts and approaches that have been used to build vector representations. They also keep track of the most recent advances of this vibrant and fast-evolving area of research, discussing cross-lingual representations and current language models based on the Transformer. Therefore, this is a useful book for researchers interested in computational methods for semantic representations and artificial intelligence. Although some basic knowledge of machine learning may be necessary to follow a few topics, the book includes clear illustrations and explanations, which make it accessible to a wide range of readers.

Apart from the preface and the conclusions, the book is organized into eight chapters. In the first two, the authors introduce some of the core ideas of NLP and artificial neural networks, respectively, discussing several concepts that will be useful throughout the book. Then, Chapters 3 to 6 present different types of vector representations at the lexical level (word embeddings, graph embeddings, sense embeddings, and contextualized embeddings), followed by a brief chapter (7) about sentence and document embeddings. For each specific topic, the book includes methods and data sets to assess the quality of the embeddings. Finally, Chapter 8 raises ethical issues involved

in data-driven models for artificial intelligence. Each chapter can be summarized as follows.

Chapter 1 makes a brief introduction to some challenges of NLP, both from understanding and from generation perspectives, including different types of linguistic ambiguity. The main part of the chapter introduces vector space models for semantic representation, presenting the distributional hypothesis and the evolution of vector space models.

The second chapter starts by giving a quick introduction of some linguistic fundamentals for NLP (syntax, morphology, and semantics) and of statistical language models. Then, it gives an overview of deep learning, presenting the fundamental differences between architectures, and concepts which will be referred along with the book. Finally, the authors present some of the most relevant knowledge resources to build semantically richer vector representations.

Chapter 3 is an extensive review of word embeddings. It first presents different count-based approaches and dimensionality reduction techniques and then discusses predictive models such as Word2vec and GloVe. Additionally, it also describes character-based and knowledge-based embeddings as well as supervised and unsupervised approaches of cross-lingual vector representations.

Chapter 4 illustrates the principal methods to build node and relation embeddings from graphs. First, it presents the key strategies to build node embeddings, from matrix factorization or random walks to methods based on graph neural networks. Then, two approaches regarding relation embeddings are presented: those built from knowledge graphs, and unsupervised methods which exploit regularities in the vector space.

The next chapter (5) starts by presenting the *Meaning Conflation Deficiency* of static word embeddings, which motivates research on sense representations. This chapter discusses two main approaches to build sense embeddings: unsupervised methods to induce senses from corpora, and knowledge-based approaches which take advantage of lexical resources.

Chapter 6 addresses contextualized embeddings and describes the main properties of the Transformer architecture and the self-attention mechanism. It includes an overview of these types of embeddings, from early methods that represent a word by its context, to current language models for contextualized word representation. In this respect, the authors present contextualized models based on recurrent neural networks (e.g., ELMo), and on the Transformer (GPT, BERT, and some derivatives). The potential impact of several parameters, such as subword tokenization or the training objectives, is also explained, and the authors discuss various approaches to use these models in downstream tasks, such as feature extraction and finetuning. Finally, they also summarize some interesting insights regarding the exploration of the linguistic properties encoded by neural language models.

Chapter 7 comprises a brief sketch of vector representations of longer units, such as sentences and documents. It presents the bag of words approach and its limitations as well as the concept of compositionality and its significance for the unsupervised learning of sentence embeddings. Some supervised strategies (e.g., training on natural language inference or machine translation datasets) are also discussed.

Ethical aspects and biases of word representations are the focus of Chapter 8. Here, the authors present some risks of data-driven models for artificial intelligence and use examples of gender stereotypes to show biases present in word embeddings, followed by several methods aimed at reducing those biases. Overall, the authors emphasize the growing interest in the NLP community to critically analyze the social impact of these models.

The book concludes by highlighting some of the major achievements of current vector representations and calling for more rigorous evaluations to measure their progress, especially in languages other than English, and with an eye on interpretability.

In summary, this book brings a high-level synthesis of different types of embeddings for NLP, focused on the general concepts and the most established techniques, and includes useful pointers to delve deeper into specific topics. As the book also discusses the most recent contextualized models (up to November 2020), it results in an attractive combination of the foundations of vector space models with current approaches based on artificial neural networks. As suggested by the authors, because of the explosion and rapid development of deep learning methods for NLP, maybe "it is necessary to step back and rethink in order to achieve true language understanding."

Marcos Garcia is a postdoctoral researcher at CiTIUS, the Research Center in Intelligent Technologies of the University of Santiago de Compostela. He has worked on NLP topics such as PoS-tagging, dependency parsing, and lexical semantics, and has developed resources and tools for different languages in both industry and academia. His e-mail address is `marcos.garcia.gonzalez@usc.gal`.