

Embeddings in Natural Language Processing

Theory and Advances in
Vector Representations of Meaning

Synthesis Lectures on Human Language Technologies

Editor

Graeme Hirst, *University of Toronto*

Synthesis Lectures on Human Language Technologies is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

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

Mohammad Taher Pilehvar and Jose Camacho-Collados
2020

Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots

Michael McTear
2020

Natural Language Processing for Social Media, Third Edition

Anna Atefeh Farzindar and Diana Inkpen
2020

Statistical Significance Testing for Natural Language Processing

Rotem Dror, Lotem Peled, Segev Shlomov, and Roi Reichart
2020

Deep Learning Approaches to Text Production

Shashi Narayan and Claire Gardent
2020

Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics

Emily M. Bender and Alex Lascarides
2019

Cross-Lingual Word Embeddings

Anders Søgaard, Ivan Vulić, Sebastian Ruder, Manaal Faruqui
2019

Bayesian Analysis in Natural Language Processing, Second Edition

Shay Cohen
2019

Argumentation Mining

Manfred Stede and Jodi Schneider
2018

Quality Estimation for Machine Translation

Lucia Specia, Carolina Scarton, and Gustavo Henrique Paetzold
2018

Natural Language Processing for Social Media, Second Edition

Atefeh Farzindar and Diana Inkpen
2017

Automatic Text Simplification

Horacio Saggion
2017

Neural Network Methods for Natural Language Processing

Yoav Goldberg
2017

Syntax-based Statistical Machine Translation

Philip Williams, Rico Sennrich, Matt Post, and Philipp Koehn
2016

Domain-Sensitive Temporal Tagging

Jannik Strötgen and Michael Gertz
2016

Linked Lexical Knowledge Bases: Foundations and Applications

Iryna Gurevych, Judith Eckle-Kohler, and Michael Matuschek
2016

Bayesian Analysis in Natural Language Processing

Shay Cohen
2016

Metaphor: A Computational Perspective

Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov
2016

Grammatical Inference for Computational Linguistics

Jeffrey Heinz, Colin de la Higuera, and Menno van Zaanen
2015

Automatic Detection of Verbal Deception

Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
2015

Natural Language Processing for Social Media

Atefeh Farzindar and Diana Inkpen
2015

Semantic Similarity from Natural Language and Ontology Analysis

Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain
2015

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition

Hang Li
2014

Ontology-Based Interpretation of Natural Language

Philipp Cimiano, Christina Unger, and John McCrae
2014

Automated Grammatical Error Detection for Language Learners, Second Edition

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2014

Web Corpus Construction

Roland Schäfer and Felix Bildhauer
2013

Recognizing Textual Entailment: Models and Applications

Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto
2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax

Emily M. Bender
2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing

Anders Søgaard
2013

Semantic Relations Between Nominals

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
2013

Computational Modeling of Narrative

Inderjeet Mani
2012

Natural Language Processing for Historical Texts

Michael Piotrowski
2012

Sentiment Analysis and Opinion Mining

Bing Liu
2012

Discourse Processing

Manfred Stede
2011

Bitext Alignment

Jörg Tiedemann
2011

Linguistic Structure Prediction

Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing

Hang Li
2011

Computational Modeling of Human Language Acquisition

Afra Alishahi
2010

Introduction to Arabic Natural Language Processing

Nizar Y. Habash
2010

Cross-Language Information Retrieval

Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce

Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling

Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems

Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing

Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics

Graham Wilcock
2009

Dependency Parsing

Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval

ChengXiang Zhai
2008

© Springer Nature Switzerland AG 2022
Reprint of original edition © Morgan & Claypool 2021

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning
Mohammad Taher Pilehvar and Jose Camacho-Collados

ISBN: 978-3-031-01049-1 paperback
ISBN: 978-3-031-02177-0 ebook
ISBN: 978-3-031-00188-8 hardcover

DOI 10.1007/978-3-031-02177-0

A Publication in the Springer series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #47
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

Embeddings in Natural Language Processing

Theory and Advances in Vector Representations of Meaning

Mohammad Taher Pilehvar
Tehran Institute for Advanced Studies

Jose Camacho-Collados
Cardiff University

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #47

ABSTRACT

Embeddings have undoubtedly been one of the most influential research areas in Natural Language Processing (NLP). Encoding information into a low-dimensional vector representation, which is easily integrable in modern machine learning models, has played a central role in the development of NLP. Embedding techniques initially focused on words, but the attention soon started to shift to other forms: from graph structures, such as knowledge bases, to other types of textual content, such as sentences and documents.

This book provides a high-level synthesis of the main embedding techniques in NLP, in the broad sense. The book starts by explaining conventional word vector space models and word embeddings (e.g., Word2Vec and GloVe) and then moves to other types of embeddings, such as word sense, sentence and document, and graph embeddings. The book also provides an overview of recent developments in contextualized representations (e.g., ELMo and BERT) and explains their potential in NLP.

Throughout the book, the reader can find both essential information for understanding a certain topic from scratch and a broad overview of the most successful techniques developed in the literature.

KEYWORDS

natural language processing, semantics, vector space model, word embeddings, contextualized embeddings, sense embeddings, graph embeddings, sentence embeddings

Contents

	Preface	xv
1	Introduction	1
1.1	Semantic Representation	4
1.2	Vector Space Models	4
1.3	The Evolution Path of Representations	7
2	Background	9
2.1	Natural Language Processing Fundamentals	9
2.1.1	Linguistic Fundamentals	9
2.1.2	Language Models	10
2.2	Deep Learning for NLP	11
2.2.1	Sequence Encoding	12
2.2.2	Recurrent Neural Networks (RNNs)	12
2.2.3	Transformers	21
2.3	Knowledge Resources	22
2.3.1	WordNet	22
2.3.2	Wikipedia, Freebase, Wikidata, DBpedia, and Wiktionary	23
2.3.3	BabelNet and ConceptNet	24
2.3.4	PPDB: The Paraphrase Database	24
3	Word Embeddings	25
3.1	Count-Based Models	26
3.1.1	Pointwise Mutual Information	27
3.1.2	Dimensionality Reduction	28
3.1.3	Random Indexing	28
3.2	Predictive Models	29
3.3	Character Embeddings	31
3.4	Knowledge-Enhanced Word Embeddings	32
3.5	Cross-Lingual Word Embeddings	33
3.5.1	Sentence-Level Supervision	34

3.5.2	Document-Level Supervision	34
3.5.3	Word-Level Supervision	34
3.5.4	Unsupervised	36
3.6	Evaluation	37
3.6.1	Intrinsic Evaluation	37
3.6.2	Extrinsic Evaluation	39
4	Graph Embeddings	41
4.1	Node Embedding	41
4.1.1	Matrix Factorization Methods	43
4.1.2	Random Walk Methods	43
4.1.3	Incorporating Node Attributes	45
4.1.4	Graph Neural Network Methods	47
4.2	Knowledge-Based Relation Embeddings	49
4.3	Unsupervised Relation Embeddings	51
4.4	Applications and Evaluation	53
4.4.1	Node Embedding	53
4.4.2	Relation Embedding	54
5	Sense Embeddings	55
5.1	Unsupervised Sense Embeddings	56
5.1.1	Sense Representations Exploiting Monolingual Corpora	56
5.1.2	Sense Representations Exploiting Multilingual Corpora	61
5.2	Knowledge-Based Sense Embeddings	61
5.3	Evaluation and Application	66
6	Contextualized Embeddings	69
6.1	The Need for Contextualization	69
6.2	Background: Transformer Model	71
6.2.1	Self-Attention	72
6.2.2	Encoder	74
6.2.3	Decoder	74
6.2.4	Positional Encoding	74
6.3	Contextualized Word Embeddings	77
6.3.1	Earlier Methods	77
6.3.2	Language Models for Word Representation	78
6.3.3	RNN-Based Models	80

6.4	Transformer-Based Models: BERT	81
6.4.1	Masked Language Modeling	82
6.4.2	Next Sentence Prediction	83
6.4.3	Training	83
6.5	Extensions	84
6.5.1	Translation Language Modeling	85
6.5.2	Context Fragmentation	85
6.5.3	Permutation Language Modeling	86
6.5.4	Reducing Model Size	87
6.6	Feature Extraction and Finetuning	87
6.7	Analysis and Evaluation	89
6.7.1	Self Attention Patterns	89
6.7.2	Syntactic Properties	90
6.7.3	Depth-Wise Information Progression	91
6.7.4	Multilinguality	92
6.7.5	Lexical Contextualization	93
6.7.6	Evaluation	95
7	Sentence and Document Embeddings	97
7.1	Unsupervised Sentence Embeddings	97
7.1.1	Bag of Words	97
7.1.2	Sentence-Level Training	98
7.2	Supervised Sentence Embeddings	100
7.3	Document Embeddings	101
7.4	Application and Evaluation	101
8	Ethics and Bias	103
8.1	Bias in Word Embeddings	104
8.2	Debiasing Word Embeddings	105
9	Conclusions	109
	Bibliography	111
	Authors' Biographies	157

Preface

Semantic representation is one of the oldest areas of research in Natural Language Processing (NLP). This area lies at the core of language understanding and looks at ways to express meaning in a machine-interpretable form. Among different types of representation, vectors have always been a popular choice, owing to their simple geometrical intuition and versatility. In fact, vector representations (more recently, *embeddings*) have played a crucial role in improving the generalization ability of various NLP systems.

In the early 2010s, the field of semantic representation was dominated by word embeddings. One reason behind this success was that embeddings proved perfect complements for neural NLP models, which were getting increasing attention around those years. It did not take long for the research attention to be expanded to other areas of representation learning (e.g., graphs, sentences, senses), pushing the boundaries in those areas. More recently, a new branch has emerged, namely contextualized embeddings, in which representations and downstream NLP systems have become even more tied together. Contextualized models have proven effective not only in a wide range of NLP tasks but also in knowledge transfer across different languages.

This book presents an overview of the field of vector representation. As a significant extension and update of a survey article we had on the topic [[Camacho-Collados and Pilehvar, 2018](#)], this book reflects the subsequent development of the field, particularly for contextualized representations. Moreover, while the survey mainly focused on word-level embeddings, the book also discusses other types of embeddings that are commonly used in NLP, such as graph and sentence embeddings. Last but not least, this book was written with a different style and purpose, as our aim was to make it more accessible to a wider audience.

INTENDED READERSHIP

This book should be of interest to all AI researchers who work with natural language, especially those who are interested in semantics. Our goal is to introduce the topic of semantic representation to those who are new to the area and to provide those who are already familiar with the area with a broader perspective and a quick overview of recent developments and the state of the art in various branches. The book synthesizes the diverse literature on semantic representations and provides a high-level introduction to major semantic embedding models.

We note that in our overview of various techniques, we provide details only to the depth that is necessary to sketch the general shape of the field and provide a hint on how the research problem was approached. In these cases, we also provide relevant references to allow the reader

to further investigate specific sub-areas. We hope this book can bring fresh researchers and practitioners up to speed on the recent developments in the field, while pointing out open problems and areas for further exploration.

COVERAGE AND OUTLINE

The book is split into nine chapters. The first two chapters provide basic concepts that can be helpful in understanding the remainder of the book. Chapters 3 to 7 are the core of this book, which discusses different types of embeddings. While there is some connection between these chapters, we tried to make them as self-contained and independent as possible. Chapter 8 discusses some of the implications of the common representation models, while Chapter 9 provides concluding remarks and future research prospects. In the following, we describe each individual chapter in more detail.

1. In Chapter 1, we start with a brief introduction on why **semantic representation** is an important topic in NLP and discuss its evolution path. We also explain the basics of **Vector Space Models**.
2. In Chapter 2, we provide some **background knowledge** on the fundamentals of NLP and machine learning applied to language problems. Then, we briefly describe some of the main knowledge resources that are commonly used in lexical semantics.
3. Chapter 3 discusses **word representations**, starting from a brief overview of conventional count-based models and continuing with the more recent predictive embeddings. We also describe some of the techniques for specializing embeddings, which serve to produce knowledge-enhanced and cross-lingual word embeddings, and common evaluation methods for word representations.
4. Chapter 4 covers various techniques for embedding structural knowledge resources, in particular semantic **graphs**. We will overview major recent methods for embedding graph nodes and edges and conclude with their applications and evaluation.
5. In Chapter 5, we focus on the representation of individual meanings of words, i.e., **word senses**. Two classes of sense representations (unsupervised and knowledge-based) are discussed, followed by evaluation techniques for each type of representation.
6. Chapter 6 is about the recent paradigm of **contextualized embeddings**. We first explain the need for such embeddings and then describe the most popular models and their connection to language models. The chapter also covers some of the efforts to explain and analyze the effectiveness of contextualized models.
7. Chapter 7 goes beyond the level of words, and describes how **sentences and documents** can be encoded into vectorial representations. We cover some of the widely used supervised

and unsupervised techniques and discuss the applications and evaluation methods for these representations. Given the book's main focus on word-level representation, this chapter provides partial coverage but also pointers for further reading.

8. Chapter 8 explains some of the **ethical issues and inherent biases** in vector representations (and, in particular, word embeddings), which recently have been a topic of concern. The chapter also covers proposals for debiasing word embeddings.
9. Finally, in Chapter 9 we present the **concluding remarks and open research challenges**.

We note that given its massive success, there has been a surge of interest in representation learning, resulting in rapid progress over the past few years, essentially making the state-of-the-art representation technique a moving target. Therefore, it is impossible to keep track of all the recent developments in the field; instead, we try to focus on some of the more established techniques and general concepts. In fact, during the writing of the book, the field of contextualized representation had changed so much that, upon the completion of the first draft, we had to go back to the corresponding chapter and significantly extend it with newer content. Nevertheless, as we mentioned, we tried to focus on the more general concepts and ideas that are relevant for understanding rather than individual systems that may change over time.

Finally, we would like to thank our editors, reviewers, colleagues, and all others who provided us with valuable feedback for improving the book. Also, we would like to thank our families for their patient support during the writing of this book.

Mohammad Taher Pilehvar and Jose Camacho-Collados
November 2020