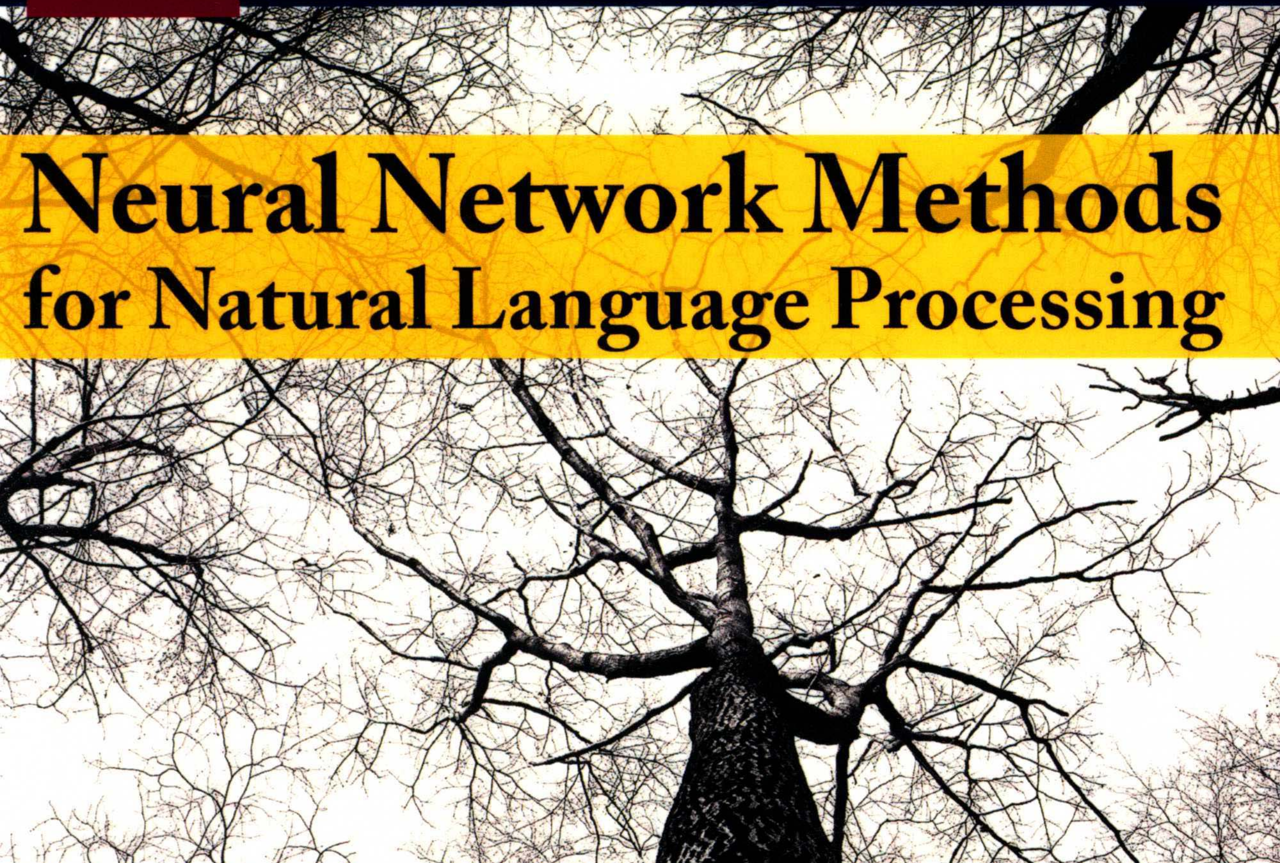
MORGAN & CLAYPOOL PUBLISHERS

Neural Network Methods for Natural Language Processing

Yoav Goldberg



Synthesis Lectures on Human Language Technologies

Graeme Hirst, *Series Editor*

Neural Network Methods for Natural Language Processing

Neural Network Methods for Natural Language Processing

Yoav Goldberg

www.morganclaypool.com

ISBN: 9781627052986 paperback
ISBN: 9781627052955 ebook

DOI 10.2200/S00762ED1V01Y201703HLT037

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #37
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

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.

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

Neural Network Methods for Natural Language Processing

Yoav Goldberg
Bar Ilan University

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #37

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

Neural networks are a family of powerful machine learning models. This book focuses on the application of neural network models to natural language data. The first half of the book (Parts I and II) covers the basics of supervised machine learning and feed-forward neural networks, the basics of working with machine learning over language data, and the use of vector-based rather than symbolic representations for words. It also covers the computation-graph abstraction, which allows to easily define and train arbitrary neural networks, and is the basis behind the design of contemporary neural network software libraries.

The second part of the book (Parts III and IV) introduces more specialized neural network architectures, including 1D convolutional neural networks, recurrent neural networks, conditioned-generation models, and attention-based models. These architectures and techniques are the driving force behind state-of-the-art algorithms for machine translation, syntactic parsing, and many other applications. Finally, we also discuss tree-shaped networks, structured prediction, and the prospects of multi-task learning.

KEYWORDS

natural language processing, machine learning, supervised learning, deep learning, neural networks, word embeddings, recurrent neural networks, sequence to sequence models

Preface

Natural language processing (NLP) is a collective term referring to automatic computational processing of human languages. This includes both algorithms that take human-produced text as input, and algorithms that produce natural looking text as outputs. The need for such algorithms is ever increasing: human produce ever increasing amounts of text each year, and expect computer interfaces to communicate with them in their own language. Natural language processing is also very challenging, as human language is inherently ambiguous, ever changing, and not well defined.

Natural language is symbolic in nature, and the first attempts at processing language were symbolic: based on logic, rules, and ontologies. However, natural language is also highly ambiguous and highly variable, calling for a more statistical algorithmic approach. Indeed, the current-day dominant approaches to language processing are all based on *statistical machine learning*. For over a decade, core NLP techniques were dominated by linear modeling approaches to supervised learning, centered around algorithms such as Perceptrons, linear Support Vector Machines, and Logistic Regression, trained over very high dimensional yet very sparse feature vectors.

Around 2014, the field has started to see some success in switching from such linear models over sparse inputs to *nonlinear neural network models* over dense inputs. Some of the neural-network techniques are simple generalizations of the linear models and can be used as almost drop-in replacements for the linear classifiers. Others are more advanced, require a change of mindset, and provide new modeling opportunities. In particular, a family of approaches based on *recurrent neural networks* (RNNs) alleviates the reliance on the Markov Assumption that was prevalent in sequence models, allowing to condition on arbitrarily long sequences and produce effective feature extractors. These advances led to breakthroughs in language modeling, automatic machine translation, and various other applications.

While powerful, the neural network methods exhibit a rather strong barrier of entry, for various reasons. In this book, I attempt to provide NLP practitioners as well as newcomers with the basic background, jargon, tools, and methodologies that will allow them to understand the principles behind neural network models for language, and apply them in their own work. I also hope to provide machine learning and neural network practitioners with the background, jargon, tools, and mindset that will allow them to effectively work with language data.

Finally, I hope this book can also serve a relatively gentle (if somewhat incomplete) introduction to both NLP and machine learning for people who are newcomers to both fields.

INTENDED READERSHIP

This book is aimed at readers with a technical background in computer science or a related field, who want to get up to speed with neural network techniques for natural language processing. While the primary audience of the book is graduate students in language processing and machine learning, I made an effort to make it useful also to established researchers in either NLP or machine learning (by including some advanced material), and to people without prior exposure to either machine learning or NLP (by covering the basics from the grounds up). This last group of people will, obviously, need to work harder.

While the book is self contained, I do assume knowledge of mathematics, in particular undergraduate level of probability, algebra, and calculus, as well as basic knowledge of algorithms and data structures. Prior exposure to machine learning is very helpful, but not required.

This book evolved out of a survey paper [Goldberg, 2016], which was greatly expanded and somewhat re-organized to provide a more comprehensive exposition, and more in-depth coverage of some topics that were left out of the survey for various reasons. This book also contains many more concrete examples of applications of neural networks to language data that do not exist in the survey. While this book is intended to be useful also for people without NLP or machine learning backgrounds, the survey paper assumes knowledge in the field. Indeed, readers who are familiar with natural language processing as practiced between roughly 2006 and 2014, with heavy reliance on machine learning and linear models, may find the journal version quicker to read and better organized for their needs. However, such readers may also appreciate reading the chapters on word embeddings (10 and 11), the chapter on conditioned generation with RNNs (17), and the chapters on structured prediction and multi-task learning (MTL) (19 and 20).

FOCUS OF THIS BOOK

This book is intended to be self-contained, while presenting the different approaches under a unified notation and framework. However, the main purpose of the book is in introducing the neural-networks (deep-learning) machinery and its application to language data, and not in providing an in-depth coverage of the basics of machine learning theory and natural language technology. I refer the reader to external sources when these are needed.

Likewise, the book is not intended as a comprehensive resource for those who will go on and develop the next advances in neural network machinery (although it may serve as a good entry point). Rather, it is aimed at those readers who are interested in taking the existing, useful technology and applying it in useful and creative ways to their favorite language-processing problems.

Further reading For in-depth, general discussion of neural networks, the theory behind them, advanced optimization methods, and other advanced topics, the reader is referred to other existing resources. In particular, the book by Bengio et al. [2016] is highly recommended.

For a friendly yet rigorous introduction to practical machine learning, the freely available book of Daumé III [2015] is highly recommended. For more theoretical treatment of machine learning, see the freely available textbook of Shalev-Shwartz and Ben-David [2014] and the textbook of Mohri et al. [2012].

For a strong introduction to NLP, see the book of Jurafsky and Martin [2008]. The information retrieval book by Manning et al. [2008] also contains relevant information for working with language data.

Finally, for getting up-to-speed with linguistic background, the book of Bender [2013] in this series provides a concise but comprehensive coverage, directed at computationally minded readers. The first chapters of the introductory grammar book by Sag et al. [2003] are also worth reading.

As of this writing, the progress of research in neural networks and Deep Learning is very fast paced. The state-of-the-art is a moving target, and I cannot hope to stay up-to-date with the latest-and-greatest. The focus is thus with covering the more established and robust techniques, that were proven to work well in several occasions, as well as selected techniques that are not yet fully functional but that I find to be established and/or promising enough for inclusion.

Yoav Goldberg
March 2017

Acknowledgments

This book grew out of a survey paper I've written on the topic [Goldberg, 2016], which in turn grew out of my frustration with the lack organized and clear material on the intersection of deep learning and natural language processing, as I was trying to learn it and teach it to my students and collaborators. I am thus indebted to the numerous people who commented on the survey paper (in its various forms, from initial drafts to post-publication comments), as well as to the people who commented on various stages of the book's draft. Some commented in person, some over email, and some in random conversations on Twitter. The book was also influenced by people who did not comment on it per-se (indeed, some never read it) but discussed topics related to it. Some are deep learning experts, some are NLP experts, some are both, and others were trying to learn both topics. Some (few) contributed through very detailed comments, others by discussing small details, others in between. But each of them influenced the final form of the book. They are, in alphabetical order: Yoav Artzi, Yonatan Aumann, Jason Baldridge, Miguel Ballesteros, Mohit Bansal, Marco Baroni, Tal Baumel, Sam Bowman, Jordan Boyd-Graber, Chris Brockett, Ming-Wei Chang, David Chiang, Kyunghyun Cho, Grzegorz Chrupala, Alexander Clark, Raphael Cohen, Ryan Cotterell, Hal Daumé III, Nicholas Dronen, Chris Dyer, Jacob Eisenstein, Jason Eisner, Michael Elhadad, Yad Faeq, Manaal Faruqui, Amir Globerson, Fréderic Godin, Edward Grefenstette, Matthew Honnibal, Dirk Hovy, Moshe Koppel, Angeliki Lazaridou, Tal Linzen, Thang Luong, Chris Manning, Stephen Merity, Paul Michel, Margaret Mitchell, Piero Molino, Graham Neubig, Joakim Nivre, Brendan O'Connor, Nikos Pappas, Fernando Pereira, Barbara Plank, Ana-Maria Popescu, Delip Rao, Tim Rocktäschel, Dan Roth, Alexander Rush, Naomi Saphra, Djamé Seddah, Erel Segal-Halevi, Avi Shmidman, Shaltiel Shmidman, Noah Smith, Anders Søgaard, Abe Stanway, Emma Strubell, Sandeep Subramanian, Liling Tan, Reut Tsarfaty, Peter Turney, Tim Vieira, Oriol Vinyals, Andreas Vlachos, Wenpeng Yin, and Torsten Zesch.

The list excludes, of course, the very many researchers I've communicated with through their academic writings on the topic.

The book also benefited a lot from—and was shaped by—my interaction with the Natural Language Processing Group at Bar-Ilan University (and its soft extensions): Yossi Adi, Roee Aharoni, Oded Avraham, Ido Dagan, Jessica Ficler, Jacob Goldberger, Hila Gonen, Joseph Keshet, Eliyahu Kiperwasser, Ron Konigsberg, Omer Levy, Oren Melamud, Gabriel Stanovsky, Ori Shapira, Micah Shlain, Vered Shwartz, Hillel Taub-Tabib, and Rachel Wities. Most of them belong in both lists, but I tried to keep things short.

The anonymous reviewers of the book and the survey paper—while unnamed (and sometimes annoying)—provided a solid set of comments, suggestions, and corrections, which I can safely say dramatically improved many aspects of the final product. Thanks, whoever you are!

And thanks also to Graeme Hirst, Michael Morgan, Samantha Draper, and C.L. Tondo for orchestrating the effort.

As usual, all mistakes are of course my own. Do let me know if you find any, though, and be listed in the next edition if one is ever made.

Finally, I would like to thank my wife, Noa, who was patient and supportive when I disappeared into writing sprees, my parents Esther and Avner and brother Nadav who were in many cases more excited about the idea of me writing a book than I was, and the staff at The Streets Cafe (King George branch) and Shne'or Cafe who kept me well fed and served me drinks throughout the writing process, with only very minimal distractions.

Yoav Goldberg
March 2017

Contents