SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at http://www.springer.com/series/10028

Lili Mou · Zhi Jin

Tree-Based Convolutional Neural Networks

Principles and Applications

Springer

Lili Mou
AdeptMind Research
Toronto, ON, Canada

Zhi Jin
Institute of Software
Peking University
Beijing, China

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-13-1869-6 ISBN 978-981-13-1870-2 (eBook)
https://doi.org/10.1007/978-981-13-1870-2

Library of Congress Control Number: 2018953306

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

To all

Preface

In recent years, neural networks have become one of the most popular models in various applications of artificial intelligence, including image recognition, speech processing, and natural language processing. The convolutional neural network and the recurrent neural network are among the most popular neural architectures. The former uses a sliding window to capture translation invariant features; it typically works with signals in a certain dimensional space (e.g., 1D speech or 2D image). The latter is suitable to process time-series data as it iteratively aggregates information.

However, such models cannot explicitly incorporate more complicated structures, e.g., the parse tree of a sentence. Socher et al. propose a recursive neural network that propagates information recursively bottom up along a tree structure. Although the recursive network can encode the tree structure to some extent, it has a long propagation path and may suffer from the problem of gradient exploding or vanishing during training.

In this book, we propose a new neural architecture, the *tree-based convolutional neural network* (TBCNN). It combines the merits of convolutional neural networks and recursive neural networks. Our key idea is to design a subtree feature detector, applied to different regions of a tree. Then, a dynamic pooling layer aggregates information to a fixed-size vector for further processing. In this way, TBCNN has short propagation path as a convolutional neural network, but is structure sensitive as a recursive neural network.

We would like to take you a tour to three applications of TBCNN: the abstract syntax trees of program source code, the constituency trees of natural language sentences, and the dependency trees of natural language sentences. In each application, we will first introduce the background of the domain, and then design a TBCNN variant especially suited to the trees in that domain. In the design, we will particularly address three technical difficulties, namely, the representations of nodes, the weight assignment of convolution, and the way of pooling. These provide useful philosophy of designing TBCNN variants as well as other neural models. We will also present detailed experiments, showing that TBCNN has achieved high performance in all these domains.

Our book addresses both rigorous mathematics and intuitive understanding. The book is suitable for researchers and graduate students in the fields of machine learning and natural language processing. It is also suitable for general public who are interested in machine learning, or more generally, artificial intelligence. The foundations of generic neural networks are covered in Chap. 2, and for each application domain, we have made efforts to make it self-contained by introducing sufficient background knowledge. We hope this book would be an interesting read for a wide range of audience.

Toronto, Canada — Lili Mou
Beijing, China — Zhi Jin
July 2018

Acknowledgements

This research is supported by the National Basic Research Program of China (the 973 Program) under Grant No. 2015CB352201, and the National Natural Science Foundation of China under Grant Nos. 61620106007 and 61751210. The research in the book started in early 2014, and was conducted mainly at the Key Laboratory of High Confidence Software Technologies (Peking University), Ministry of Education; and the Institute of Software, Peking University. During the writing of the book, the first author had a post-doctoral fellow position at the University of Waterloo, and then took a scientist position at AdeptMind Research. We would like to thank the University of Waterloo and AdeptMind Inc. (Toronto) for their support of the writing of the book as well as research in general.

The authors would like to thank Prof. Lu Zhang, Dr. Ge Li, Dr. Yan Xu, Dr. Yangyang Lu, Hao Peng, Yuxuan Liu, Hao Jia, Ran Jia, Rui Men, and Zhao Meng at Peking University for their help in the research. We would also like to thank Dr. Lixing Li, Dr. Yunchuan Chen, and Dr. Weizhuo Li from the University of Chinese Academy of Sciences; together with them and Dr. Yan Xu and Dr. Yangyang Lu, we self-organized weekly seminars, during which we have learned both foundations and frontiers of machine learning research.

In our research, we have also discussed with many other colleagues, including Dr. Hang Li from Toutiao AI Lab, Dr. Zhengdong Lu from DeeplyCurious.ai, and Yiping Song, Bolin Wei, Jingsi Wen, Wenhao Huang, and Zhenxin Fu from Peking University. We would like to thank them for their insightful thoughts.

Last but not least, we would like to thank our families and friends for their long-time help and support.

Thank you all!

Contents

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
AST	Abstract Syntax Tree
BoW	Bag-of-Words
CBoW	Continuous Bag-of-Words
CCG	Combinatory Categorical Grammar
CNN	Convolutional Neural Network
CRF	Conditional Random Field
DAG	Directed Acyclic Graph
DNN	Deep Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
MLP	Multilayer Perceptron
MSE	Mean Square Error
OJ	Open Judge
PCA	Principle Component Analysis
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
RGB	Red, Green, and Blue
RBF	Radius-Basis Function
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TBCNN	Tree-Based Convolutional Neural Network
c-TBCNN	Constituency Tree-Based Convolutional Neural Network
d-TBCNN	Dependency Tree-Based Convolutional Neural Network