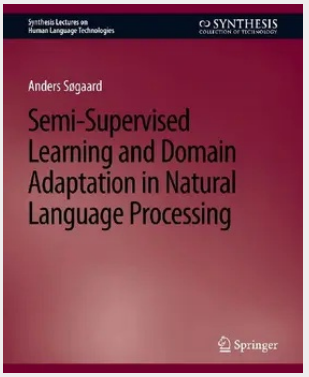
Søgaard

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing



This book introduces basic supervised learning algorithms applicable to natural language processing (NLP) and shows how the performance of these algorithms can often be improved by exploiting the marginal distribution of large amounts of unlabeled data. One reason for that is data sparsity, i.e., the limited amounts of data we have available in NLP. However, in most real-world NLP applications our labeled data is also heavily biased. This book introduces extensions of supervised learning algorithms to cope with data sparsity and different kinds of sampling bias. This book is intended to be both readable by first-year students and interesting to the expert audience. My intention was to introduce what is necessary to appreciate the major challenges we face in contemporary NLP related to data sparsity and sampling bias, without wasting too much time on details about supervised learning algorithms or particular NLP applications. I use text classification, part-of-speech tagging, and dependency parsing as running examples, and limit myself to a small set of cardinal learning algorithms. I have worried less about theoretical guarantees ("this algorithm never does too badly") than about useful rules of thumb ("in this case this algorithm may perform really well"). In NLP, data is so noisy, biased, and non-stationary that few theoretical guarantees can be established and we are typically left with our gut feelings and a catalogue of crazy ideas. I hope this book will provide its readers with both. Throughout the book we include snippets of Python code and empirical evaluations, when relevant.

This book introduces basic supervised learning algorithms applicable to natural language processing (NLP) and shows how the performance of these algorithms can often be improved by exploiting the marginal distribution of large amounts of unlabeled data. One reason for that is data sparsity, i.e., the limited amounts of data we have available in NLP. However, in most real-world NLP applications our labeled data is also heavily biased. This book introduces extensions of supervised learning algorithms to cope with data sparsity and different kinds of sampling bias. This book is intended to be both readable by first-year students and interesting to the expert audience. My intention was to introduce what is necessary to appreciate the major challenges we face in contemporary NLP related to data sparsity and sampling bias, without wasting too much time on details about supervised learning algorithms or particular NLP applications. I use text classification, part-of-speech tagging, and dependency parsing as running examples, and limit myself to a small set of cardinal learning algorithms. I have worried less about theoretical guarantees ("this algorithm never does too badly") than about useful rules of thumb ("in this case this algorithm may perform really well"). In NLP, data is so noisy, biased, and non-stationary that few theoretical guarantees can be established and we are typically left with our gut feelings and a catalogue of crazy ideas. I hope this book will provide its readers with both. Throughout the book we include snippets of Python code and empirical evaluations, when relevant.

26,74 €
24,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783031010217
Medium: Buch
ISBN: 978-3-031-01021-7
Verlag: Springer
Erscheinungstermin: 22.05.2013
Sprache(n): Englisch
Auflage: Erscheinungsjahr 2013
Serie: Synthesis Lectures on Human Language Technologies
Produktform: Kartoniert
Gewicht: 212 g
Seiten: 93
Format (B x H): 191 x 235 mm

Kundenservice Verlag Dr. Otto Schmidt KG
Neumannstraße 10, 40235 Düsseldorf | kundenservice@otto-schmidt.de | 0221-93738-997 (Inland)