

Contents

| | |
|--|-----|
| Foreword by Antal van den Bosch | vii |
| 1 Introduction | 1 |
| 1.1 Structure Discovery for Language Processing | 1 |
| 1.1.1 Structure Discovery Paradigm | 3 |
| 1.1.2 Approaches to Automatic Language Processing | 4 |
| 1.1.3 Knowledge-intensive and Knowledge-free | 5 |
| 1.1.4 Degrees of Supervision | 6 |
| 1.1.5 Contrasting Structure Discovery with Previous Approaches . | 7 |
| 1.2 Relation to General Linguistics | 7 |
| 1.2.1 Linguistic Structuralism and Distributionalism | 8 |
| 1.2.2 Adequacy of the Structure Discovery Paradigm | 9 |
| 1.3 Similarity and Homogeneity in Language Data | 11 |
| 1.3.1 Levels in Natural Language Processing | 11 |
| 1.3.2 Similarity of Language Units | 12 |
| 1.3.3 Homogeneity of Sets of Language Units | 12 |
| 1.4 Vision: The Structure Discovery Machine | 13 |
| 1.5 Connecting Structure Discovery to NLP tasks | 15 |
| 1.6 Contents of this Book | 16 |
| 1.6.1 Theoretical Aspects of Structure Discovery | 16 |
| 1.6.2 Applications of Structure Discovery | 17 |
| 1.6.3 The Future of Structure Discovery | 17 |
| 2 Graph Models | 19 |
| 2.1 Graph Theory | 19 |
| 2.1.1 Notions of Graph Theory | 19 |
| 2.1.2 Measures on Graphs | 23 |
| 2.2 Random Graphs and Small World Graphs | 27 |
| 2.2.1 Random Graphs: Erdős-Rényi Model | 27 |
| 2.2.2 Small World Graphs: Watts-Strogatz Model | 28 |
| 2.2.3 Preferential Attachment: Barabási-Albert Model | 29 |

| | | |
|----------|---|-----------|
| 2.2.4 | Ageing: Power-laws with Exponential Tails | 30 |
| 2.2.5 | Semantic Networks: Steyvers-Tenenbaum Model | 32 |
| 2.2.6 | Changing the Power-Law's Slope: (α, β) Model | 32 |
| 2.2.7 | Two Regimes: Dorogovtsev-Mendes Model | 35 |
| 2.2.8 | Further Remarks on Small World Graph Models | 35 |
| 2.2.9 | Further Reading | 37 |
| 3 | Small Worlds of Natural Language | 39 |
| 3.1 | Power-Laws in Rank-Frequency Distribution | 39 |
| 3.1.1 | Word Frequency | 40 |
| 3.1.2 | Letter N -grams | 41 |
| 3.1.3 | Word N -grams | 41 |
| 3.1.4 | Sentence Frequency | 44 |
| 3.1.5 | Other Power-Laws in Language Data | 44 |
| 3.1.6 | Modelling Language with Power-Law Awareness | 45 |
| 3.2 | Scale-Free Small Worlds in Language Data | 46 |
| 3.2.1 | Word Co-occurrence Graph | 46 |
| 3.2.2 | Co-occurrence Graphs of Higher Order | 51 |
| 3.2.3 | Sentence Similarity | 55 |
| 3.2.4 | Summary on Scale-Free Small Worlds in Language Data | 57 |
| 3.3 | An Emergent Random Generation Model for Language | 57 |
| 3.3.1 | Review of Emergent Random Text Models | 58 |
| 3.3.2 | Desiderata for Random Text Models | 59 |
| 3.3.3 | Testing Properties of Word Streams | 60 |
| 3.3.4 | Word Generator | 60 |
| 3.3.5 | Sentence Generator | 63 |
| 3.3.6 | Measuring Agreement with Natural Language | 65 |
| 3.3.7 | Summary for the Generation Model | 70 |
| 4 | Graph Clustering | 73 |
| 4.1 | Review on Graph Clustering | 73 |
| 4.1.1 | Introduction to Clustering | 73 |
| 4.1.2 | Spectral vs. Non-spectral Graph Partitioning | 77 |
| 4.1.3 | Graph Clustering Algorithms | 77 |
| 4.2 | Chinese Whispers Graph Clustering | 83 |
| 4.2.1 | Chinese Whispers Algorithm | 84 |
| 4.2.2 | Empirical Analysis | 88 |
| 4.2.3 | Weighting of Vertices | 91 |
| 4.2.4 | Approximating Deterministic Outcome | 92 |
| 4.2.5 | Disambiguation of Vertices | 95 |
| 4.2.6 | Hierarchical Divisive Chinese Whispers | 96 |
| 4.2.7 | Hierarchical Agglomerative Chinese Whispers | 98 |
| 4.2.8 | Summary on Chinese Whispers | 99 |

| | | |
|----------|---|-----|
| 5 | Unsupervised Language Separation | 101 |
| 5.1 | Related Work | 101 |
| 5.2 | Method | 102 |
| 5.3 | Evaluation | 103 |
| 5.4 | Experiments with Equisized Parts for 10 Languages | 104 |
| 5.5 | Experiments with Bilingual Corpora | 107 |
| 5.6 | Case study: Language Separation for Twitter | 109 |
| 5.7 | Summary on Language Separation | 111 |
| 6 | Unsupervised Part-of-Speech Tagging | 113 |
| 6.1 | Introduction to Unsupervised POS Tagging | 113 |
| 6.2 | Related Work | 114 |
| 6.3 | System Architecture | 117 |
| 6.4 | Tagset 1: High and Medium Frequency Words | 118 |
| 6.5 | Tagset 2: Medium and Low Frequency Words | 122 |
| 6.6 | Combination of Tagsets 1 and 2 | 124 |
| 6.7 | Setting up the Tagger | 125 |
| 6.7.1 | Lexicon Construction | 125 |
| 6.7.2 | Constructing the Tagger | 126 |
| 6.7.3 | Morphological Extension | 127 |
| 6.8 | Direct Evaluation of Tagging | 127 |
| 6.8.1 | Influence of System Components | 128 |
| 6.8.2 | Influence of Parameters | 131 |
| 6.8.3 | Influence of Corpus Size | 132 |
| 6.8.4 | Domain Shifting | 133 |
| 6.8.5 | Comparison with Clark [66] | 134 |
| 6.9 | Application-based Evaluation | 137 |
| 6.9.1 | Unsupervised POS for Supervised POS | 137 |
| 6.9.2 | Unsupervised POS for Word Sense Disambiguation | 139 |
| 6.9.3 | Unsupervised POS for NER and Chunking | 141 |
| 6.10 | Conclusion on Unsupervised POS Tagging | 144 |
| 7 | Word Sense Induction and Disambiguation | 145 |
| 7.1 | Related Work on Word Sense Induction | 145 |
| 7.2 | Task-oriented Definition of WSD | 146 |
| 7.3 | Word Sense Induction using Graph Clustering | 147 |
| 7.3.1 | Graph Clustering Parameterisation | 148 |
| 7.3.2 | Feature Assignment in Context | 149 |
| 7.4 | Evaluation of WSI Features in a Supervised WSD System | 149 |
| 7.4.1 | Machine Learning Setup for Supervised WSD System | 149 |
| 7.4.2 | SemEval-07 Lexical Sample Task | 151 |
| 7.4.3 | Lexical Substitution System | 152 |
| 7.4.4 | Substitution Acceptability Evaluation | 153 |
| 7.5 | Conclusion on Word Sense Induction and Disambiguation | 155 |

| | | |
|----------|---|-----|
| 8 | Conclusion | 157 |
| 8.1 | Current State of Structure Discovery..... | 157 |
| 8.2 | The Future of Structure Discovery | 159 |
| | References | 161 |



<http://www.springer.com/978-3-642-25922-7>

Structure Discovery in Natural Language

Biemann, C.

2012, XX, 180 p., Hardcover

ISBN: 978-3-642-25922-7