

Table of Contents

| | |
|--|------|
| Introduction | xiii |
| Eric GAUSSIER and François YVON | |
| | |
| PART 1: INFORMATION RETRIEVAL | 1 |
| | |
| Chapter 1. Probabilistic Models for Information Retrieval | 3 |
| Stéphane CLINCHANT and Eric GAUSSIER | |
| | |
| 1.1. Introduction | 3 |
| 1.1.1. Heuristic retrieval constraints | 6 |
| 1.2. 2-Poisson models | 8 |
| 1.3. Probability ranking principle (PRP) | 10 |
| 1.3.1. Reformulation | 12 |
| 1.3.2. BM25 | 13 |
| 1.4. Language models | 15 |
| 1.4.1. Smoothing methods | 16 |
| 1.4.2. The Kullback-Leibler model | 19 |
| 1.4.3. Noisy channel model | 20 |
| 1.4.4. Some remarks | 20 |
| 1.5. Informational approaches | 21 |
| 1.5.1. DFR models | 22 |
| 1.5.2. Information-based models | 25 |
| 1.6. Experimental comparison | 27 |
| 1.7. Tools for information retrieval | 27 |
| 1.8. Conclusion | 28 |
| 1.9. Bibliography | 29 |

| | |
|---|--------|
| Chapter 2. Learnable Ranking Models for Automatic Text Summarization and Information Retrieval | 33 |
| Massih-Réza AMINI, David BUFFONI, Patrick GALLINARI, Tuong Vinh TRUONG, and Nicolas USUNIER | |
| 2.1. Introduction | 33 |
| 2.1.1. Ranking of instances | 34 |
| 2.1.2. Ranking of alternatives | 42 |
| 2.1.3. Relation to existing frameworks | 44 |
| 2.2. Application to automatic text summarization | 45 |
| 2.2.1. Presentation of the application | 45 |
| 2.2.2. Automatic summary and learning | 48 |
| 2.3. Application to information retrieval | 49 |
| 2.3.1. Application presentation | 49 |
| 2.3.2. Search engines and learning | 50 |
| 2.3.3. Experimental results | 53 |
| 2.4. Conclusion | 54 |
| 2.5. Bibliography | 54 |
| PART 2: CLASSIFICATION AND CLUSTERING | 59 |
| Chapter 3. Logistic Regression and Text Classification | 61 |
| Sujeevan ASEERVATHAM, Eric GAUSSIER, Anestis ANTONIADIS, Michel BURLET, and Yves DENNEULIN | |
| 3.1. Introduction | 61 |
| 3.2. Generalized linear model | 62 |
| 3.3. Parameter estimation | 65 |
| 3.4. Logistic regression | 68 |
| 3.4.1. Multinomial logistic regression | 69 |
| 3.5. Model selection | 70 |
| 3.5.1. Ridge regularization | 71 |
| 3.5.2. LASSO regularization | 71 |
| 3.5.3. Selected Ridge regularization | 72 |
| 3.6. Logistic regression applied to text classification | 74 |
| 3.6.1. Problem statement | 74 |
| 3.6.2. Data pre-processing | 75 |
| 3.6.3. Experimental results | 76 |
| 3.7. Conclusion | 81 |
| 3.8. Bibliography | 82 |
| Chapter 4. Kernel Methods for Textual Information Access | 85 |
| Jean-Michel RENDERS | |
| 4.1. Kernel methods: context and intuitions | 85 |
| 4.2. General principles of kernel methods | 88 |

| | |
|--|-----|
| 4.3. General problems with kernel choices (kernel engineering) | 95 |
| 4.4. Kernel versions of standard algorithms: examples of solvers | 97 |
| 4.4.1. Kernel logistic regression | 98 |
| 4.4.2. Support vector machines | 99 |
| 4.4.3. Principal component analysis | 101 |
| 4.4.4. Other methods | 102 |
| 4.5. Kernels for text entities | 103 |
| 4.5.1. “Bag-of-words” kernels | 104 |
| 4.5.2. Semantic kernels | 105 |
| 4.5.3. Diffusion kernels | 107 |
| 4.5.4. Sequence kernels | 109 |
| 4.5.5. Tree kernels | 112 |
| 4.5.6. Graph kernels | 116 |
| 4.5.7. Kernels derived from generative models | 119 |
| 4.6. Summary | 123 |
| 4.7. Bibliography | 124 |
| | |
| Chapter 5. Topic-Based Generative Models for Text | |
| Information Access | 129 |
| Jean-Cédric CHAPPELIER | |
| 5.1. Introduction | 129 |
| 5.1.1. Generative versus discriminative models | 129 |
| 5.1.2. Text models | 131 |
| 5.1.3. Estimation, prediction and smoothing | 133 |
| 5.1.4. Terminology and notations | 134 |
| 5.2. Topic-based models | 135 |
| 5.2.1. Fundamental principles | 135 |
| 5.2.2. Illustration | 136 |
| 5.2.3. General framework | 138 |
| 5.2.4. Geometric interpretation | 139 |
| 5.2.5. Application to text categorization | 141 |
| 5.3. Topic models | 142 |
| 5.3.1. Probabilistic Latent Semantic Indexing | 143 |
| 5.3.2. Latent Dirichlet Allocation | 146 |
| 5.3.3. Conclusion | 160 |
| 5.4. Term models | 161 |
| 5.4.1. Limitations of the multinomial | 161 |
| 5.4.2. Dirichlet compound multinomial | 162 |
| 5.4.3. DCM-LDA | 163 |
| 5.5. Similarity measures between documents | 164 |
| 5.5.1. Language models | 165 |
| 5.5.2. Similarity between topic distributions | 165 |
| 5.5.3. Fisher kernels | 166 |

| | |
|--|------------|
| 5.6. Conclusion | 168 |
| 5.7. Appendix: topic model software | 169 |
| 5.8. Bibliography | 170 |
| Chapter 6. Conditional Random Fields for Information Extraction | 179 |
| Isabelle TELLIER and Marc TOMMASI | |
| 6.1. Introduction | 179 |
| 6.2. Information extraction | 180 |
| 6.2.1. The task | 180 |
| 6.2.2. Variants | 182 |
| 6.2.3. Evaluations | 182 |
| 6.2.4. Approaches not based on machine learning | 183 |
| 6.3. Machine learning for information extraction | 184 |
| 6.3.1. Usage and limitations | 184 |
| 6.3.2. Some applicable machine learning methods | 185 |
| 6.3.3. Annotating to extract | 186 |
| 6.4. Introduction to conditional random fields | 187 |
| 6.4.1. Formalization of a labelling problem | 187 |
| 6.4.2. Maximum entropy model approach | 188 |
| 6.4.3. Hidden Markov model approach | 190 |
| 6.4.4. Graphical models | 191 |
| 6.5. Conditional random fields | 193 |
| 6.5.1. Definition | 193 |
| 6.5.2. Factorization and graphical models | 195 |
| 6.5.3. Junction tree | 196 |
| 6.5.4. Inference in CRFs | 198 |
| 6.5.5. Inference algorithms | 200 |
| 6.5.6. Training CRFs | 201 |
| 6.6. Conditional random fields and their applications | 203 |
| 6.6.1. Linear conditional random fields | 204 |
| 6.6.2. Links between linear CRFs and hidden Markov models | 205 |
| 6.6.3. Interests and applications of CRFs | 208 |
| 6.6.4. Beyond linear CRFs | 210 |
| 6.6.5. Existing libraries | 211 |
| 6.7. Conclusion | 214 |
| 6.8. Bibliography | 215 |
| PART 3: MULTILINGUALISM | 221 |
| Chapter 7. Statistical Methods for Machine Translation | 223 |
| Alexandre ALLAUZEN and François YVON | |
| 7.1. Introduction | 223 |
| 7.1.1. Machine translation in the age of the Internet | 223 |

| | |
|--|------------|
| 7.1.2. Organization of the chapter | 226 |
| 7.1.3. Terminological remarks | 227 |
| 7.2. Probabilistic machine translation: an overview | 227 |
| 7.2.1. Statistical machine translation: the standard model | 228 |
| 7.2.2. Word-based models and their limitations | 230 |
| 7.2.3. Phrase-based models | 234 |
| 7.3. Phrase-based models | 235 |
| 7.3.1. Building word alignments | 237 |
| 7.3.2. Word alignment models: a summary | 245 |
| 7.3.3. Extracting bisegments | 246 |
| 7.4. Modeling reorderings | 250 |
| 7.4.1. The space of possible reorderings | 250 |
| 7.4.2. Evaluating permutations | 255 |
| 7.5. Translation: a search problem | 259 |
| 7.5.1. Combining models | 259 |
| 7.5.2. The decoding problem | 261 |
| 7.5.3. Exact search algorithms | 262 |
| 7.5.4. Heuristic search algorithms | 267 |
| 7.5.5. Decoding: a solved problem? | 272 |
| 7.6. Evaluating machine translation | 272 |
| 7.6.1. Subjective evaluations | 273 |
| 7.6.2. The <i>BLEU</i> metric | 275 |
| 7.6.3. Alternatives to <i>BLEU</i> | 277 |
| 7.6.4. Evaluating machine translation: an open problem | 279 |
| 7.7. State-of-the-art and recent developments | 279 |
| 7.7.1. Using source context | 279 |
| 7.7.2. Hierarchical models | 281 |
| 7.7.3. Translating with linguistic resources | 283 |
| 7.8. Useful resources | 287 |
| 7.8.1. Bibliographic data and online resources | 288 |
| 7.8.2. Parallel corpora | 288 |
| 7.8.3. Tools for statistical machine translation | 288 |
| 7.9. Conclusion | 289 |
| 7.10. Acknowledgments | 291 |
| 7.11. Bibliography | 291 |
| PART 4: EMERGING APPLICATIONS | 305 |
| Chapter 8. Information Mining: Methods and Interfaces for Accessing Complex Information | 307 |
| Josiane MOTHE, Kurt ENGLMEIER, and Fionn MURTAGH | |
| 8.1. Introduction | 307 |
| 8.2. The multidimensional visualization of information | 309 |

| | |
|---|------------|
| 8.2.1. Accessing information based on the knowledge of the structured domain | 309 |
| 8.2.2. Visualization of a set of documents via their content | 313 |
| 8.2.3. OLAP principles applied to document sets | 317 |
| 8.3. Domain mapping via social networks | 320 |
| 8.4. Analyzing the variability of searches and data merging | 323 |
| 8.4.1. Analysis of IR engine results | 323 |
| 8.4.2. Use of data unification | 325 |
| 8.5. The seven types of evaluation measures used in IR | 327 |
| 8.6. Conclusion | 331 |
| 8.7. Acknowledgments | 332 |
| 8.8. Bibliography | 332 |
| Chapter 9. Opinion Detection as a Topic Classification Problem | 337 |
| Juan-Manuel TORRES-MORENO, Marc EL-BÈZE, Patrice BELLOT, and Frédéric BÉCHET | |
| 9.1. Introduction | 337 |
| 9.2. The TREC and TAC evaluation campaigns | 339 |
| 9.2.1. Opinion detection by question–answering | 340 |
| 9.2.2. Automatic summarization of opinions | 342 |
| 9.2.3. The text mining challenge of opinion classification (DEFT (Défi Fouille de Textes)) | 343 |
| 9.3. Cosine weights - a second glance | 347 |
| 9.4. Which components for a opinion vectors? | 348 |
| 9.4.1. How to pass from words to terms? | 349 |
| 9.5. Experiments | 352 |
| 9.5.1. Performance, analysis, and visualization of the results on the IMDB corpus | 354 |
| 9.6. Extracting opinions from speech: automatic analysis of phone polls | 357 |
| 9.6.1. France Télécom opinion investigation corpus | 358 |
| 9.6.2. Automatic recognition of spontaneous speech in opinion corpora | 360 |
| 9.6.3. Evaluation | 363 |
| 9.7. Conclusion | 365 |
| 9.8. Bibliography | 365 |
| Appendix A. Probabilistic Models: An Introduction | 369 |
| François YVON | |
| A.1. Introduction | 369 |
| A.2. Supervised categorization | 370 |
| A.2.1. Filtering documents | 370 |
| A.2.2. The Bernoulli model | 372 |
| A.2.3. The multinomial model | 376 |

| | |
|---|------------|
| A.2.4. Evaluating categorization systems | 379 |
| A.2.5. Extensions | 380 |
| A.2.6. A first summary | 383 |
| A.3. Unsupervised learning: the multinomial mixture model | 384 |
| A.3.1. Mixture models | 384 |
| A.3.2. Parameter estimation | 386 |
| A.3.3. Applications | 390 |
| A.4. Markov models: statistical models for sequences | 391 |
| A.4.1. Modeling sequences | 391 |
| A.4.2. Estimating a Markov model | 394 |
| A.4.3. Language models | 395 |
| A.5. Hidden Markov models | 397 |
| A.5.1. The model | 398 |
| A.5.2. Algorithms for hidden Markov models | 399 |
| A.6. Conclusion | 410 |
| A.7. A primer of probability theory | 411 |
| A.7.1. Probability space, event | 411 |
| A.7.2. Conditional independence and probability | 412 |
| A.7.3. Random variables, moments | 413 |
| A.7.4. Some useful distributions | 418 |
| A.8. Bibliography | 420 |
| List of Authors | 423 |
| Index | 425 |

