CONTENTS

1 INTRODUCTION 1
- 1.1 Motivation 1
- 1.2 The successful trilogy 11
- 1.3 Some remarks on the approach to discover novelties 14
- 1.4 Road map 19
  - 1.4.1 Problem statement and formalization 19
  - 1.4.2 Proposed algorithmic solution 22
  - 1.4.3 Experimentation and results 23
  - 1.4.4 Conclusions and future work 24

I Problem Statement and Formalization 25

2 SOME APPLICATION DOMAINS 27
- 2.1 Text segmentation 28
- 2.2 Part of speech tagging 29
- 2.3 Text chunking 29
- 2.4 (Monotone) Automatic translation 30
- 2.5 Spoken language understanding 32
- 2.6 Ambiguous keyboards, predictive text 33
- 2.7 Handwritten text recognition 37
- 2.8 Speech 41
- 2.9 Summary and some conclusions 49
  - 2.9.1 Underlying representation 50
  - 2.9.2 Preprocessing 52
  - 2.9.3 Ambiguous front-ends and the pipeline problem 52

3 SOME PROBLEM TYPES RELATED TO SEQUENCES 57
- 3.1 Some problem types 58
  - 3.1.1 Sequence classification 59
  - 3.1.2 Segment Classification 60
  - 3.1.3 Joint classification 61
  - 3.1.4 Segmentation and over-segmentation 62
  - 3.1.5 Recognition 64
  - 3.1.6 Parsing 64
  - 3.1.7 Language modeling 65
  - 3.1.8 Prediction/Forecasting 65
  - 3.1.9 Translation 65
  - 3.1.10 Alignment 66
  - 3.1.11 KWS, indexing, querying by example 69
  - 3.1.12 Unsupervised term discovery 72
  - 3.1.13 Filtering/Denoising and Smoothing 73
  - 3.1.14 Summarization 73
  - 3.1.15 Joint Segmentation and Classification 74
- 3.2 Some extensions 80
  - 3.2.1 Several input sequences 80
  - 3.2.2 Interactive systems 82

3.3 Limitations, how to face them, new proposal 87
3.4 Questionnaire when faced with a new problem 91
3.5 Analysis and evaluation measures 93
3.5.1 Quality assessment 96
3.5.2 Performance 99
3.5.3 Interaction 100
3.5.4 Other measures 102
3.6 Summary and some conclusions 102

4 MODELS 105
4.1 Some preliminary ML concepts 106
4.1.1 (Probabilistic) graphical models 115
4.2 Two stage generative model 130
4.2.1 Hierarchy for the second stage 133
4.2.2 Limitations, extensions and generalizations 137
4.3 Classical problems of two stage generative models 141
4.3.1 Probability of observation 142
4.3.2 Decoding 144
4.3.3 Model estimation 145
4.4 Some alternative models 146
4.4.1 Relationship with Dynamic Graphical Models 146
4.4.2 Fixed dimension feature segments 150
4.4.3 Estimation of frame-wise segment posteriors 155
4.4.4 Graph transformer networks 160
4.4.5 Some non-probabilistic frameworks 163
4.5 Summary and some conclusions 165

5 RECOGNITION, DECODING, PARSING, TRANSLATION 167
5.1 Introduction 167
5.2 Recognition, parsing, decoding 170
5.3 Weighted languages and semirings 172
5.4 Some formalisms 177
5.4.1 Formal/generative grammars 177
5.4.2 Finite state automata and transducers 184
5.4.3 Recurrent transition networks 187
5.5 Deriving the composition of a regular and a CF model 194
5.5.1 State-pair transducer composition 195
5.5.2 Extension to null-transitions 200
5.5.3 Extension to model reference transitions 209
5.5.4 Transformation to homogeneous epsilon form 221
5.6 Review of parsing approaches, decoders and algorithms 226
5.7 From composition to recognition/decoding 233
5.7.1 Acyclic inputs 234
5.7.2 Semiring specific optimizations 236
5.8 Summary and some conclusions 236

6 SEQUENCE MODELS 239
6.1 Probabilistic decompositions 242
6.1.1 Chain rule, clustering histories 243
6.1.2 Whole sentence LMs 244
6.1.3 Combining spans of the sequence 245
6.2 Assesment measures for LMs 246

6.3 Lexical units 249
6.3.1 Open vs closed lexicons 250
6.3.2 Most common types of lexical unit 250
6.3.3 Categories 253
6.3.4 Lexicon representation 254
6.4 LM combination 257
6.5 Dynamic capabilities 259
6.6 Common LM types 261
6.6.1 N-gram models 261
6.6.2 Weighted finite automata 265
6.6.3 Recurrent NN LMs 266
6.6.4 Other LM types 267
6.7 On semantics 268
6.8 Dealing with OOV words 269
6.9 LM Wrappers 271
6.9.1 Based on transducer composition 271
6.9.2 Changing lexical units 273
6.9.3 Partial/imperfect transcription 274
6.10 Some LM decoding features 275
6.10.1 Pruning, bounds, sub-lexicons, look-ahead and scales 276
6.10.2 LM histories, long-span LMs 278
6.11 Heterogeneous lattice-based LM 280
6.12 Summary and some conclusions 282

7 SEGMENT MODELS 285
7.1 Introduction 285
7.2 Overcoming the limitations of HMMs 288
7.2.1 Piecewise stationary 289
7.2.2 Frame independence assumption 290
7.2.3 Markov modeling of intra-segmental regions 292
7.2.4 Weak duration modeling of HMMs 293
7.3 Implicit vs explicit duration 293
7.3.1 Implicit duration models (and topologies) 294
7.3.2 Explicit duration models 299
7.4 Frame sequence emission 300
7.4.1 Intermediate levels of description 301
7.4.2 Trajectories, state-space and templates 303
7.5 Frame emission 304
7.5.1 From discrete features 304
7.5.2 From continuous features 305
7.5.3 Estimation from frame posteriors 307
7.5.4 A general scheme of frame emission 309
7.6 Context dependency 312
7.7 Allowing discriminative models 314
7.8 Some steps towards a general scheme of Segment Models 315
7.9 Summary and some conclusions 317

II Proposed Algorithmic Solutions 321

8 ARCHITECTURE 323
8.1 Dataflow architecture 324
8.2 DAG serialization protocol 326
8.3 Overview of dataflow components 331
8.3.1 Generic dataflow components 331
8.3.2 Domain specific components 332
8.4 Recognizer examples 334
8.4.1 One pass 334
8.4.2 Decoupled and semi-decoupled architectures 334
8.4.3 Semi-decoupled with feedback 337
8.4.4 Multipass recognizer 337
8.4.5 One-stage dag recognizer 339
8.4.6 Multimodal recognizer 339
8.4.7 Interactive assisted recognition 339
8.5 Summary and some conclusions 341

9 LM INTERFACES AND REPRESENTATIONS 343
9.1 Language Model Interface 344
9.1.1 n-gram interface 345
9.1.2 Automaton interface 345
9.1.3 RTN interface 353
9.1.4 Heterogeneous lattice LM interface 355
9.2 Count based n-gram implementations 355
9.3 Automata representation 360
9.3.1 Proposed "fan-out based" representation 361
9.3.2 Memory mapping 363
9.3.3 Proposed RTN representation 364
9.4 n-gram history manager 367
9.5 Neural Network LM 369
9.5.1 Fast evaluation with skipping NN LMs 370
9.5.2 Normalization constant estimation 373
9.5.3 NN LMs specialized for small vocabulary 376
9.5.4 "LM look-ahead" NN LMs 378
9.6 Assisted transcription representation 380
9.7 Summary and some conclusions 382

10 SEGMENT AND LEXICON ESTIMATORS 385
10.1 Introduction 386
10.2 Segment Estimator Interface 388
10.2.1 Incremental computation 389
10.2.2 Persistent and ephemeral operations 390
10.2.3 LM vs time conditioned decoders and Token Passing 394
10.2.4 Relationship with across-word context dependency 397
10.2.5 Relationship with pruning 400
10.3 Classical Viterbi implementations 400
10.3.1 Array swapping 402
10.3.2 Hash swapping 403
10.3.3 Reverse topological traversal 404
10.4 Dealing with null transitions and edition operations 405

10.5 Lexicon organization 408
10.5.1 Linear/flat lexicon 409
10.5.2 Prefix tree lexicon 412
10.5.3 General lexicon networks (a.k.a. DAWG) 413
10.6 Techniques to reduce the cost 415
10.6.1 Active states pruning 415
10.6.2 Fast-match look-ahead 422
10.6.3 Variable frame rate and frame dropping 422
10.6.4 External over-segmentation 424
10.6.5 Information from previous passes 425
10.6.6 Bundle-search and word-pair approximation 426
10.6.7 Language Model Look-Ahead 428
10.6.8 Submodel dominance recombination 433
10.6.9 Other techniques to reduce the cost 434
10.7 Detecting and emitting OOV words 438
10.8 Specialized linear lexicon decoders 440
10.9 Specialized tree lexicon decoders 450
10.9.1 Model representation 450
10.9.2 Representation of active states 453
10.9.3 Different traversal possibilities 455
10.9.4 Forward (root to leaves) traversal 457
10.9.5 Decoding sub-word DAGs with error correction 462
10.9.6 Backwards (leaves towards the root) traversal 464
10.9.7 Combining Forward and Backwards 465
10.9.8 Sub-word-centric decoding 466
10.10 Lexicon Network specialized decoders 469
10.10.1 Adapting forward traversal to DAGWs 469
10.10.2 Adapting the sub-word-centric approach to DAGWs 471
10.11 Summary and some conclusions 472

11 DAG GENERATION 475
11.1 From a sequence (`seq_graph_gen`) 477
11.1.1 Basic model 478
11.1.2 Pruning 482
11.1.3 Using feedback information 483
11.1.4 External over-segmentation 486
11.1.5 Control automaton and across-word dependency 489
11.1.6 Clustering of frame boundaries 493
11.2 From another graph (`graph_graph_gen`) 495
11.2.1 Over-segmenter information 498
11.2.2 Feedback channel 501
11.2.3 Edition operations 502
11.3 Combining DAGs 504
11.4 Summary and some conclusions 504

12 DECODING 507
12.1 Introduction 507
12.1.1 Parallel, multi-threaded distributed recognizers 508
12.2 Two stage decoder 509
12.2.1 Internal components 510
12.2.2 Behavior 511
12.3 One Pass decoder 514

12.3.1 One pass without dictionary look-ups 516
12.3.2 Dealing with silence/space models 518
12.4 Dealing with external over-segmentation 520
12.5 Summary and some conclusions 521

III Experimentation and results 523

13 LM EXPERIMENTATION 525
13.1 Some brief empirical support to our automata representation 525
13.2 Fall back skipping NN LMs 530
13.2.1 Emulating lower order n-gram NNLMs 530
13.2.2 Models in a machine translation system 533
13.3 NN LM sofmax estimation with auxiliary LMs 535
13.4 Summary and some conclusions 536

14 HTR PREPROCESSING 539
14.1 Image cleaning 541
14.2 Tracking reference lines 544
14.2.1 Extraction and classification of "interesting points" 549
14.2.2 Active learning of classes of "interesting points" 550
14.3 Slope removal 551
14.4 Slant removal 553
14.4.1 Non-uniform slant removal 556
14.5 Size normalization 558
14.5.1 Vertical/Weight normalization 559
14.5.2 Width normalization 560
14.6 Feature extraction 563
14.6.1 Avoiding the feature extraction process 566
14.7 External word over-segmentation 566
14.8 Online preprocessing 569
14.8.1 Online preprocessing stages 569
14.8.2 Adapting the proposed offline preprocessing 571
14.8.3 Feature Extraction 571
14.8.4 External word over-segmentation 573
14.9 Summary and some conclusions 573

15 HTR EXPERIMENTATION 577
15.1 HTR with HMM/ANNs and count-based n-grams 579
15.1.1 Experimental setting (corpora, dictionary and LM) 579
15.1.2 Baseline experiments with HMM/GMMs 582
15.1.3 Experiments with HMM/ANNs 584
15.1.4 Analysis of results 588
15.2 HTR with HMM/ANN models and connectionist LMs 592
15.3 HTR by using CNNs for preprocessing 597
15.4 Lexicon-free recognition 599
15.5 Combining HMM/ANNs with holistic ANN 604
15.6 Some (isolated word) bimodal experiments 609
15.7 Summary, conclusions and future work 611

IV Conclusions and future work 615

16 CONCLUSIONS AND FUTURE WORK 617
16.1 Brief overview highlighting our contributions 619
16.1.1 Relating the first part 620
16.1.2 Relating the second part 622
16.1.3 Relating the third part 626
16.2 Some conclusions 627
16.2.1 Some expected critiques 629
16.2.2 Some final thoughts 630
16.3 Future Work 632
16.3.1 Related to decoding 632
16.3.2 Related to language modeling 634
16.3.3 Related to handwriting 635
16.4 Publications 636
16.4.1 Contributions derived from this thesis 636
16.4.2 Collaborations with other authors 638
16.4.3 Other publications 640

V Appendix 643

A CORPORA 645
A.1 Text 645
A.1.1 Lancaster Olso Bergen (LOB) 645
A.1.2 One Billion Word Benchmark 646
A.2 Handwriting 646
A.2.1 IAMdb 647
A.2.2 IAM-OnDB 647
A.2.3 biMod-IAM-PRHLT 649
A.2.4 RIMES 649

BIBLIOGRAPHY 653