

Contents

Part I Entropy Guided Transformation Learning: Algorithms

1	Introduction	3
1.1	Motivation	3
1.2	Applications	4
1.3	Overview of the Book	6
	References	6
2	Entropy Guided Transformation Learning	9
2.1	Transformation Based Learning	9
2.2	TBL Bottleneck	11
2.3	Entropy Guided Template Generation	12
2.3.1	Information Gain	13
2.3.2	Decision Trees	14
2.3.3	Template Extraction	15
2.3.4	True Class Trick	15
2.3.5	High Dimensional Features	16
2.4	Template Evolution	17
2.5	Template Sampling	18
2.6	Redundant Transformation Rules	18
2.7	Related Work	19
	References	20
3	ETL Committee	23
3.1	Ensemble Algorithms	23
3.2	Training Phase	24
3.2.1	Bootstrap Sampling	24
3.2.2	Feature Sampling	25
3.2.3	ETL Training	26

3.3	Classification Phase	26
3.4	Related Work	27
	References	28

Part II Entropy Guided Transformation Learning: Applications

4	General ETL Modeling for NLP Tasks	31
4.1	Modeling	31
4.2	Basic Parameter Setting	32
4.3	Committee Parameter Setting	33
4.4	Performance Measures	33
4.5	Software and Hardware	34
	References	34
5	Part-of-Speech Tagging	35
5.1	Task and Corpora	35
5.2	POS Tagging Modeling	36
5.2.1	Morphological Stage	36
5.2.2	Contextual Stage	37
5.3	Machine Learning Modeling	37
5.4	Mac-Morpho Corpus	38
5.5	Tycho Brahe Corpus	38
5.6	TIGER Corpus	39
5.7	Brown Corpus	39
5.8	Summary	40
	References	41
6	Phrase Chunking	43
6.1	Task and Corpora	43
6.2	Phrase Chunking Modeling	44
6.2.1	Derived Features	45
6.3	Machine Learning Modeling	45
6.4	SNR-CLIC Corpus	46
6.5	Ramshaw and Marcus Corpus	46
6.6	CoNLL-2000 Corpus	47
6.7	SPSAL-2007 Corpus	47
6.8	Summary	48
	References	49
7	Named Entity Recognition	51
7.1	Task and Corpora	51

7.2	Named Entity Recognition Modeling	52
7.2.1	Derived Features	53
7.3	Machine Learning Modeling	53
7.4	HAREM Corpus	54
7.5	SPA CoNLL-2002 Corpus	56
7.6	DUT CoNLL-2002 Corpus	56
7.7	Summary	58
	References	58
8	Semantic Role Labeling	59
8.1	Task and Corpora	59
8.2	Semantic Role Labeling Modeling	61
8.2.1	Derived Features	62
8.2.2	Preprocessing	62
8.2.3	Postprocessing	63
8.3	Machine Learning Modeling	64
8.4	CoNLL-2004 Corpus	65
8.5	CoNLL-2005 Corpus	66
8.6	Summary	68
	References	68
9	Conclusions	71
9.1	Final Remarks on ETL	71
9.2	Final Remarks on ETL Committee	72
9.3	Future Work	73
	References	73
	Appendix A: ETL Committee Behavior	75