

Contents

Acknowledgements	5
Contents	7
1 Introduction	11
1 Theoretical background	12
1.1 Problem description	12
1.2 Data representation	13
1.3 Positive and negative learning examples	14
1.4 Innate knowledge	14
2 Experiment setup	15
2.1 Goals	16
2.2 The training and test data	17
2.3 Data complexity	18
2.4 The linguistic initialization model	20
2.5 Elementary statistics	22
3 Related work	23
3.1 The work by Ellison	23
3.2 The work by Daelemans et al.	24
3.3 Other work	25
2 Statistical Learning	27
1 Markov models	27
1.1 General description of Markov models	27
1.2 The forward procedure	29
1.3 The Viterbi algorithm	31
2 Hidden Markov Models	34
2.1 General description of Hidden Markov Models	34
2.2 The extended forward procedure	35
2.3 The extended Viterbi algorithm	37
2.4 Learning in a Hidden Markov Model	39
2.5 Using Hidden Markov Models in practice	44

3	Initial Experiments	45
3.1	A test experiment	45
3.2	Orthographic data with random initialization	46
3.3	Orthographic data with linguistic initialization	48
3.4	Discussion	50
4	Experiments with bigram HMMs	52
4.1	General bigram HMM experiment set-up	53
4.2	Orthographic data with random initialization	53
4.3	Orthographic data with linguistic initialization	55
4.4	Phonetic data with random initialization	58
4.5	Phonetic data with linguistic initialization	60
5	Concluding remarks	63
3	Connectionist Learning	65
1	Feed-forward networks	65
1.1	General description of feed-forward networks	66
1.2	Learning in a feed-forward network	68
1.3	Representing non-numeric data in a neural network	71
2	The Simple Recurrent Network (SRN)	72
2.1	General description of SRNs	73
2.2	Learning in SRNs	75
2.3	Using SRNs for language experiments	76
3	Experiments with SRNs	77
3.1	General experiment set-up	78
3.2	Finding network parameters with restricted data	80
3.3	Orthographic data with random initialization	83
3.4	Orthographic data with linguistic initialization	85
4	Discovering the problem	87
4.1	The influence of the number of valid successors of a string	87
4.2	Can we scale up the Cleeremans et al. experiment?	88
4.3	A possible solution: IT-SRNs	90
4.4	Experiments with IT-SRNs	91
5	Concluding remarks	93
4	Rule-based Learning	95
1	Introduction to Rule-based Learning	95
1.1	Positive versus negative examples	96
1.2	The expected output of the learning method	97
1.3	Available symbolic learning methods	97
2	Inductive Logic Programming	99
2.1	Introduction to Inductive Logic Programming	99
2.2	The background knowledge and the hypotheses	102
2.3	Deriving hypotheses	105
2.4	The hypothesis models and grammar theory	107

3	Experiments with Inductive Logic Programming	110
3.1	General experiment setup	110
3.2	Handling orthographic and phonetic data	112
3.3	Adding extra linguistic constraints	113
3.4	Discussion	116
4	Alternative rule-based models	118
4.1	Extending the model	118
4.2	Deriving extended hypotheses	120
4.3	Experiments with the extended model	122
4.4	Compressing the models	124
5	Concluding Remarks	128
5	Concluding remarks	131
1	Experiment results	131
2	Recent related work	134
3	Future work	136
	Bibliography	139
	Samenvatting	145