

# Contents

	<b>Preface</b> .....	<b>xvii</b>
	<b>Acknowledgments</b> .....	<b>xix</b>
<b>1</b>	<b>Introduction</b> .....	<b>1</b>
	1.1 Classic Machine Learning Paradigm .....	1
	1.2 Motivating Examples .....	3
	1.3 A Brief History of Lifelong Learning .....	6
	1.4 Definition of Lifelong Learning .....	9
	1.5 Types of Knowledge and Key Challenges .....	14
	1.6 Evaluation Methodology and Role of Big Data .....	16
	1.7 Outline of the Book .....	18
<b>2</b>	<b>Related Learning Paradigms</b> .....	<b>21</b>
	2.1 Transfer Learning .....	21
	2.1.1 Structural Correspondence Learning .....	22
	2.1.2 Naïve Bayes Transfer Classifier .....	23
	2.1.3 Deep Learning in Transfer Learning .....	24
	2.1.4 Difference from Lifelong Learning .....	25
	2.2 Multi-Task Learning .....	26
	2.2.1 Task Relatedness in Multi-Task Learning .....	26
	2.2.2 GO-MTL: Multi-Task Learning using Latent Basis .....	27
	2.2.3 Deep Learning in Multi-Task Learning .....	29
	2.2.4 Difference from Lifelong Learning .....	30
	2.3 Online Learning .....	31
	2.3.1 Difference from Lifelong Learning .....	31
	2.4 Reinforcement Learning .....	32
	2.4.1 Difference from Lifelong Learning .....	33
	2.5 Meta Learning .....	33
	2.5.1 Difference from Lifelong Learning .....	34
	2.6 Summary .....	34

<b>3</b>	<b>Lifelong Supervised Learning</b> . . . . .	<b>35</b>
3.1	Definition and Overview . . . . .	36
3.2	Lifelong Memory-Based Learning . . . . .	37
3.2.1	Two Memory-Based Learning Methods . . . . .	37
3.2.2	Learning a New Representation for Lifelong Learning . . . . .	37
3.3	Lifelong Neural Networks . . . . .	38
3.3.1	MTL Net . . . . .	38
3.3.2	Lifelong EBNN . . . . .	39
3.4	ELLA: An Efficient Lifelong Learning Algorithm . . . . .	40
3.4.1	Problem Setting . . . . .	41
3.4.2	Objective Function . . . . .	41
3.4.3	Dealing with the First Inefficiency . . . . .	42
3.4.4	Dealing with the Second Inefficiency . . . . .	44
3.4.5	Active Task Selection . . . . .	45
3.5	Lifelong Naive Bayesian Classification . . . . .	46
3.5.1	Naïve Bayesian Text Classification . . . . .	46
3.5.2	Basic Ideas of LSC . . . . .	48
3.5.3	LSC Technique . . . . .	49
3.5.4	Discussions . . . . .	50
3.6	Domain Word Embedding via Meta-Learning . . . . .	51
3.7	Summary and Evaluation Datasets . . . . .	53
<b>4</b>	<b>Continual Learning and Catastrophic Forgetting</b> . . . . .	<b>55</b>
4.1	Catastrophic Forgetting . . . . .	55
4.2	Continual Learning in Neural Networks . . . . .	57
4.3	Learning without Forgetting . . . . .	59
4.4	Progressive Neural Networks . . . . .	61
4.5	Elastic Weight Consolidation . . . . .	62
4.6	iCaRL: Incremental Classifier and Representation Learning . . . . .	64
4.6.1	Incremental Training . . . . .	64
4.6.2	Updating Representation . . . . .	65
4.6.3	Constructing Exemplar Sets for New Classes . . . . .	66
4.6.4	Performing Classification in iCaRL . . . . .	67
4.7	Expert Gate . . . . .	67
4.7.1	Autoencoder Gate . . . . .	68
4.7.2	Measuring Task Relatedness for Training . . . . .	69

4.7.3	Selecting the Most Relevant Expert for Testing	69
4.7.4	Encoder-Based Lifelong Learning	70
4.8	Continual Learning with Generative Replay	70
4.8.1	Generative Adversarial Networks	70
4.8.2	Generative Replay	71
4.9	Evaluating Catastrophic Forgetting	72
4.10	Summary and Evaluation Datasets	73
<b>5</b>	<b>Open-World Learning</b>	<b>77</b>
5.1	Problem Definition and Applications	78
5.2	Center-Based Similarity Space Learning	79
5.2.1	Incrementally Updating a CBS Learning Model	79
5.2.2	Testing a CBS Learning Model	81
5.2.3	CBS Learning for Unseen Class Detection	82
5.3	DOC: Deep Open Classification	85
5.3.1	Feed-Forward Layers and the 1-vs.-Rest Layer	85
5.3.2	Reducing Open-Space Risk	86
5.3.3	DOC for Image Classification	88
5.3.4	Unseen Class Discovery	88
5.4	Summary and Evaluation Datasets	89
<b>6</b>	<b>Lifelong Topic Modeling</b>	<b>91</b>
6.1	Main Ideas of Lifelong Topic Modeling	91
6.2	LTM: A Lifelong Topic Model	94
6.2.1	LTM Model	95
6.2.2	Topic Knowledge Mining	96
6.2.3	Incorporating Past Knowledge	97
6.2.4	Conditional Distribution of Gibbs Sampler	99
6.3	AMC: A Lifelong Topic Model for Small Data	100
6.3.1	Overall Algorithm of AMC	100
6.3.2	Mining Must-link Knowledge	101
6.3.3	Mining Cannot-link Knowledge	103
6.3.4	Extended Pólya Urn Model	104
6.3.5	Sampling Distributions in Gibbs Sampler	106
6.4	Summary and Evaluation Datasets	108

<b>7</b>	<b>Lifelong Information Extraction</b>	<b>111</b>
7.1	NELL: A Never-Ending Language Learner	111
7.1.1	NELL Architecture	114
7.1.2	Extractors and Learning in NELL	114
7.1.3	Coupling Constraints in NELL	117
7.2	Lifelong Opinion Target Extraction	117
7.2.1	Lifelong Learning through Recommendation	118
7.2.2	AER Algorithm	119
7.2.3	Knowledge Learning	120
7.2.4	Recommendation using Past Knowledge	121
7.3	Learning on the Job	123
7.3.1	Conditional Random Fields	123
7.3.2	General Dependency Feature	124
7.3.3	The L-CRF Algorithm	126
7.4	Lifelong-RL: Lifelong Relaxation Labeling	127
7.4.1	Relaxation Labeling	127
7.4.2	Lifelong Relaxation Labeling	128
7.5	Summary and Evaluation Datasets	129
<b>8</b>	<b>Continuous Knowledge Learning in Chatbots</b>	<b>131</b>
8.1	LiLi: Lifelong Interactive Learning and Inference	132
8.2	Basic Ideas of LiLi	134
8.3	Components of LiLi	136
8.4	A Running Example	137
8.5	Summary and Evaluation Datasets	138
<b>9</b>	<b>Lifelong Reinforcement Learning</b>	<b>139</b>
9.1	Lifelong Reinforcement Learning through Multiple Environments	141
9.1.1	Acquiring and Incorporating Bias	141
9.2	Hierarchical Bayesian Lifelong Reinforcement Learning	142
9.2.1	Motivation	142
9.2.2	Hierarchical Bayesian Approach	143
9.2.3	MTRL Algorithm	143
9.2.4	Updating Hierarchical Model Parameters	144
9.2.5	Sampling an MDP	146
9.3	PG-ELLA: Lifelong Policy Gradient Reinforcement Learning	146

9.3.1	Policy Gradient Reinforcement Learning .....	147
9.3.2	Policy Gradient Lifelong Learning Setting .....	148
9.3.3	Objective Function and Optimization .....	149
9.3.4	Safe Policy Search for Lifelong Learning.....	150
9.3.5	Cross-domain Lifelong Reinforcement Learning .....	151
9.4	Summary and Evaluation Datasets .....	152
<b>10</b>	<b>Conclusion and Future Directions .....</b>	<b>153</b>
	<b>Bibliography .....</b>	<b>159</b>
	<b>Authors' Biographies .....</b>	<b>187</b>