Dual Learning

Tao Qin

Dual Learning

Springer

Tao Qin
Microsoft Research Asia (China)
Beijing, China

ISBN 978-981-15-8883-9 ISBN 978-981-15-8884-6 (eBook)
https://doi.org/10.1007/978-981-15-8884-6

© Springer Nature Singapore Pte Ltd. 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

I would like to dedicate this book to my wife and my lovely son!

Preface

Deep neural networks have become the dominant paradigm for artificial intelligence (AI) in the past decade, and deep learning has significantly advanced various areas of AI, spanning from computer vision, natural language and speech, to game playing. A key factor contributing to the success of deep learning is availability of large-scale labeled training data. Correspondingly, a leading challenge (and also a hot research direction) for deep learning is how to learn from limited/insufficient labeled data. Dual learning is a new learning framework proposed to address this challenge by leveraging structural duality between AI tasks.

This book gives a comprehensive review of recent research on dual learning. We introduce its basic principles, including dual reconstruction, joint-probability equation, and marginal probability equation, and cover various learning settings and algorithms, including dual semi-supervised learning, dual unsupervised learning, dual supervised learning, and dual inference. For each setting, we introduce diverse applications, such as machine translation, image-to-image translation, speech synthesis and recognition, question answering and generation, image classification and generation, code summarization and generation, sentiment analysis, etc.

This book is written for researchers and graduate/undergraduate students in machine learning, computer vision, natural language, and speech areas. Background in those areas would be helpful, but it is not essential to read the book. A very brief introduction to basic machine learning and deep learning concepts is provided in Chaps. 2 and 3.

Beijing, China
July 2020

Tao Qin

Acknowledgments

This book would not have been possible without the contributions of many people.

I would like to thank my colleagues and interns at Microsoft, who have been working together with me on the topic of dual learning, including Tie-Yan Liu, Hsiao-Wuen Hon, Wei-Ying Ma, Yingce Xia, Xu Tan, Di He, Li Zhao, Fei Tian, Jiang Bian, Wei Chen, Sheng Zhao, Duyu Tang, Nan Duan, Jianxin Lin, Yijun Wang, Yi Ren, Yiren Wang, Jin Xu, and Tianyu He. I would also like to thank my external collaborators including Nenghai Yu, Zhou Zhao, Zhibo Chen, Liwei Wang, Jian Li, Enhong Chen, and Chengxiang Zhai.

I also want to thank those who allowed me to reproduce images/figures from their publications. I indicate their contributions in the figure captions throughout the text.

Contents

1 Introduction 1
1.1 Motivation 1
1.2 Structure Duality in AI Tasks 3
1.3 Categorization of Dual Learning 4
1.3.1 Classified by Data Settings 4
1.3.2 Classified by Principles 4
1.4 Book Overview 5
References 6

Part I Preparations

2 Machine Learning Basics 11
2.1 Machine Learning Paradigms 11
2.1.1 Supervised Learning 12
2.1.2 Unsupervised Learning 13
2.1.3 Reinforcement Learning 14
2.1.4 More Learning Paradigms 15
2.2 Key Components of a Learning Algorithm 15
2.2.1 An Example: Multiclass Logistic Regression 17
2.3 Generalization and Regularization 18
2.4 Building a Machine Learning Model 19
2.4.1 Data Collection and Feature Engineering 19
2.4.2 Algorithm Selection, Model Training, and Hyper-parameter Tuning 20
References 21

3 Deep Learning Basics 25
3.1 Neural Networks 25
3.2 Convolutional Neural Networks 28
3.3 Sequence Modeling 31
3.3.1 Recurrent Neural Network and Its Variants 31
3.3.2 Encoder-Decoder Architecture 33

3.3.3 Transformer Networks 36
3.4 Training Deep Models 39
3.4.1 Stochastic Gradient Descent 39
3.4.2 Regularization 40
3.5 Why Deep Networks? 42
References 44

Part II The Dual Reconstruction Principle

4 Dual Learning for Machine Translation and Beyond 49
4.1 Introduction to Machine Translation 49
4.1.1 Neural Machine Translation 50
4.1.2 Back Translation 51
4.2 The Principle of Dual Reconstruction 51
4.3 Dual Semi-supervised Learning 53
4.3.1 Zero-Shot Dual Machine Translation 56
4.4 Dual Unsupervised Learning 57
4.4.1 Basic Ideas 57
4.4.2 System Architectures and Training Algorithms 58
4.5 Multi-Agent Dual Learning 61
4.5.1 The Framework 62
4.5.2 Extensions and Comparisons 63
4.5.3 Multi-Agent Dual Machine Translation 64
4.6 Beyond Machine Translation 67
4.6.1 Semantic Parsing 67
4.6.2 Text Style Transfer 67
4.6.3 Conversations 69
References 70

5 Dual Learning for Image Translation and Beyond 73
5.1 Introduction 73
5.1.1 Generative Adversarial Networks 74
5.2 Basic Idea of Unsupervised Image Translation 75
5.3 Image to Image Translation 76
5.3.1 DualGAN 76
5.3.2 CycleGAN 79
5.3.3 DiscoGAN 80
5.4 Fine-Grained Image to Image Translation 82
5.4.1 The Problem of Fine-Grained Image Translation 84
5.4.2 Conditional DualGAN 84
5.4.3 Discussions 86
5.5 Multi-Domain Image Translation with Multi-Path Consistency ... 87
5.6 Beyond Image Translation 88
5.6.1 Face Related Tasks 88
5.6.2 Visual-Linguistic Tasks 90
5.6.3 Other Image Related Tasks 91
References 91

6 Dual Learning for Speech Processing and Beyond 95
6.1 Neural Speech Synthesis and Recognition 95
6.2 Speech Chain with Dual Learning 96
6.3 Dual Learning for Low-Resource Speech Processing 98
6.3.1 Denoising Auto-Encoding with Bidirectional Sequence Modeling 99
6.3.2 Dual Reconstruction with Bidirectional Sequence Modeling 101
6.3.3 Model Training 102
6.4 Dual Learning for Extremely Low-Resource Speech Processing 103
6.4.1 Pre-training and Fine-Tuning 105
6.4.2 Dual Reconstruction 106
6.4.3 Knowledge Distillation 107
6.4.4 Performance of LRSpeech 108
6.5 Dual Learning for Non-native Speech Recognition 108
6.5.1 The Problem of Non-native Speech Recognition 108
6.5.2 The Method Based on the Dual Reconstruction Principle 109
6.6 Beyond Speech Processing 111
References 112

Part III The Probabilistic Principle

7 Dual Supervised Learning 117
7.1 The Joint-Probability Principle 117
7.2 The Algorithm of Dual Supervised Learning 118
7.3 Applications 120
7.3.1 Neural Machine Translation 120
7.3.2 Images Classification and Generation 121
7.3.3 Sentiment Analysis 122
7.3.4 Question Answering and Generation 123
7.3.5 Code Summarization and Generation 124
7.3.6 Natural Language Understanding and Generation 129
7.4 Theoretical Analysis 130
References 131

8 Dual Inference 135
8.1 General Formulation 135
8.2 Applications 137
8.3 Theoretical Analysis 138
References 140

9 Marginal Probability Based Dual Semi-Supervised Learning 143
9.1 Efficient Approximation of Marginal Probability 143
9.2 Marginal Probability as a Constraint 145

9.3 Likelihood Maximization for Unlabeled Data 147
9.4 Discussions 148
References 150

Part IV Advanced Topics

10 Understanding Dual Reconstruction 155
10.1 Overview 155
10.2 Understanding Dual Reconstruction in Unsupervised Settings 156
10.2.1 A Formulation of Dual Unsupervised Mapping 156
10.2.2 Issues and the Simplicity Hypothesis 158
10.2.3 Minimal Complexity 160
10.3 Understanding Dual Reconstruction in Semi-Supervised Settings 162
10.3.1 Algorithm and Notations 162
10.3.2 Translation Between Two Languages 163
10.3.3 Extension: Multi-domain Dual Learning 167
References 169

11 Connections to Other Learning Paradigms 171
11.1 Dual Semi-Supervised Learning and Co-training 171
11.2 Dual Learning and Multitask Learning 173
11.3 Dual Learning, GANs and Autoencoder 173
11.4 Dual Supervised Learning and Bayesian Ying-Yang Learning 174
11.5 Dual Reconstruction and Related Concepts 176
References 176

Part V Summary and Outlook

12 Summary and Outlook 181
12.1 Summary 181
12.2 Future Directions 182
12.2.1 More Learning Settings and Applications 182
12.2.2 Efficient Training 184
12.2.3 Theoretical Understanding 185
References 185

About the Author

Dr. Tao Qin is a Senior Principal Researcher and Manager at the Microsoft Research Asia, and an Adjunct Professor (PhD advisor) at the University of Science and Technology of China. He coined the term "dual learning" together with his colleagues and has published numerous papers on the topic in Neur IPS/ICLR/ICML/AAAI/IJCAI/CVPR/ACL/EMNLP/NAACL. His research interests include machine learning (with the focus on deep learning and reinforcement learning), artificial intelligence (with applications to language understanding, speech processing and computer vision), game theory and multi-agent systems (with applications to cloud computing, online and mobile advertising, and ecommerce), information retrieval and computational advertising.

He has published over 100 refereed papers at prestigious conferences such as Neur IPS, ICML, ICLR, AAAI, IJCAI, AAMAS, ACL, EMNLP, NAACL, KDD, WWW, SIGIR, WSDM, JAIR, EC, WINE and ACM Transactions. He served/is serving as Area Chair for AAAI, IJCAI, EMNLP, AAMAS, SIGIR and ACML, Workshop Chair for WWW 2020, and Industry Chair for DAI 2019. He is a Senior Member of the IEEE and ACM.