

Signals and Communication Technology

Baihan Lin

Reinforcement Learning Methods in Speech and Language Technology

 Springer

Signals and Communication Technology

Series Editors

Emre Celebi , Department of Computer Science, University of Central Arkansas, Conway, USA

Jingdong Chen, Northwestern Polytechnical University, Xi'an, China

E. S. Gopi, Department of Electronics and Communication Engineering, National Institute of Technology, Tiruchirappalli, India

Amy Neustein, Linguistic Technology Systems, Fort Lee, USA

Antonio Liotta, University of Bolzano, Bolzano, Italy

Mario Di Mauro, University of Salerno, Salerno, Italy

This series is devoted to fundamentals and applications of modern methods of signal processing and cutting-edge communication technologies. The main topics are information and signal theory, acoustical signal processing, image processing and multimedia systems, mobile and wireless communications, and computer and communication networks. Volumes in the series address researchers in academia and industrial R&D departments. The series is application-oriented. The level of presentation of each individual volume, however, depends on the subject and can range from practical to scientific.

Indexing: All books in “Signals and Communication Technology” are indexed by Scopus and zbMATH

For general information about this book series, comments or suggestions, please contact Mary James at mary.james@springer.com or Ramesh Nath Premnath at ramesh.premnath@springer.com.

Baihan Lin

Reinforcement Learning Methods in Speech and Language Technology

 Springer

Baihan Lin
Berkman Klein Center for Internet &
Society
Harvard University
Cambridge, MA, USA

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-031-53719-6 ISBN 978-3-031-53720-2 (eBook)
<https://doi.org/10.1007/978-3-031-53720-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed 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 Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

*In loving memory of my cat Neptune,
faithful companion and silent supporter of
countless research hours.
Your furry presence is deeply missed.*

Preface

Dear Reader,

Welcome to “Reinforcement Learning Methods in Speech and Language Technology.” This book represents a labor of love, a distillation of years of research, teaching, and passionate exploration of a field that continues to fascinate and inspire me. It’s a journey that has taken me from curious student to researcher, and now as a faculty member, where I have the privilege of sharing this knowledge with you.

My fascination with reinforcement learning (RL) in speech and language technologies (SLT) began with a simple yet profound question: How can we create AI systems that learn and adapt to the complex, nuanced world of human language? This question has been my North Star, guiding me through research labs, conference halls, and countless late-night coding sessions. It led me to organize tutorials at both *INTERSPEECH* and *ICASSP*, where I had the opportunity to engage with brilliant minds from around the world, all of us united in our quest to push the boundaries of what’s possible in AI.

Writing this book has been both a monumental challenge and an immense joy. Capturing a field that evolves as rapidly as RL+SLT is no small feat. Indeed, during the writing process, we’ve witnessed seismic shifts in the landscape of AI, with the emergence of ChatGPT and other large language models that have redefined our understanding of what’s possible. These developments, far from derailing our work, have only served to underscore the importance and potential of reinforcement learning in this domain. It’s precisely this dynamism that makes our work so exciting and important.

In these pages, my goal with this book is to make these complex ideas accessible. Yes, there are mathematical formulations and technical details, but I’ve strived to present them in a way that illuminates rather than intimidates. I want everyone, from students to seasoned researchers, to understand that RL can be a universal mechanism for autonomous technologies in the future.

This book covers a wide spectrum of topics, from the fundamental principles of multi-armed bandits to the cutting-edge applications of RL in large language models. We’ll explore how RL can create adaptive, interactive, and scalable models for speech recognition, natural language understanding, dialogue systems, and

more. Through case studies and practical examples, I hope to bridge the gap between theory and application, showing how these powerful techniques can be deployed in real-world scenarios.

One of the unique aspects of this book is its emphasis on the reward-driven adaptability, state representations, temporal structures, and generalizability of RL. While many current applications focus on improving the training of deep neural networks, I believe the true power of RL in speech and language processing lies in these inherent properties. Throughout the book, we'll explore how these characteristics make RL particularly well-suited to the context-dependent, temporally structured nature of human language.

As my first book, and a major work within my first faculty position, this text is both a milestone and a work in progress. It reflects the nascent and rapidly advancing nature of our field. I see it not as a definitive tome but as an invitation to dialogue and further exploration. The field of RL+SLT is evolving at breakneck speed, and I have no doubt that new breakthroughs will emerge even as this book goes to print.

That's why I warmly welcome your comments, suggestions, and insights. This book is the beginning of a conversation, not the end. Your feedback will not only help improve future editions but will also contribute to the collective knowledge of our field. Whether you're a student embarking on your AI journey, a researcher pushing the boundaries of what's possible, or a practitioner applying these techniques in industry, your perspective is valuable and appreciated.

As we stand on the brink of a new era in AI, with language models that can engage in human-like conversation and speech systems that can understand and respond to nuanced commands, the potential of RL+SLT has never been greater. I believe that reinforcement learning will play a crucial role in creating the next generation of AI systems: systems that can truly understand us, adapt to our needs, and seamlessly integrate into our daily lives.

Thank you for joining me on this exciting journey. I hope this book challenges you, inspires you, and opens up new possibilities in your own work and research. May it serve as a springboard for your own explorations and innovations in the fascinating world of reinforcement learning for speech and language technologies.

Here's to the future of AI. Let's shape it together.

Warm regards,

New York City, NY, USA
December, 2023

Baihan Lin

Acknowledgments

The journey of writing this book has been both challenging and rewarding, and it would not have been possible without the support and contributions of many individuals.

This book is the culmination of a journey that began with my tutorial course “Reinforcement Learning and Bandits for Speech and Language Processing” that I had the privilege of presenting at *INTERSPEECH 2022* and *ICASSP 2024*. I am deeply grateful to the conference organizers for this opportunity and to the attendees whose engagement and feedback significantly shaped this work.

I owe a debt of gratitude to the pioneers in this field whose work has inspired and informed both the tutorial and this book. In particular, I would like to acknowledge Sergey Levine, David Silver, Xiangyu Zhao, Andrea Barraza-Urbina, Dorota Glowacka, Chelsea Finn, Aviral Kumar, Lilian Weng, and Emma Brunskill. Your contributions to the field of reinforcement learning have been invaluable.

I am profoundly thankful to my editors at Springer Nature, Mary James and Bakiyalakshmi RM, for their unwavering support and constructive feedback. Your expertise has significantly enhanced the quality and clarity of this work.

My heartfelt thanks go to my many colleagues and fellow researchers in the field of reinforcement learning and speech and language technologies. I want to especially express my gratitude to my long-term collaborators and mentors, Guillermo Cecchi, Djallel Bouneffouf, and Irina Rish, whose wisdom have been instrumental in shaping my research and this book. Your insights and encouragement have been invaluable. Our discussions, collaborations, and shared passion for pushing the boundaries of AI have been a constant source of inspiration.

A special thanks to my students, whose curiosity and fresh perspectives continually challenge me to explain complex concepts more clearly. You have been my best teachers.

I owe a debt of gratitude to my family and friends for their unwavering support. A special thanks to my wife, Esther, whose love, patience, and encouragement have been my anchor and kept me going during the long nights and challenging moments, especially as we prepare to welcome our first child. Your strength and understanding during this journey have been invaluable. And to our cats, Jupiter and Cosmo, thank

you for your constant companionship during those long writing sessions. And to Neptune, who we lost to lymphoma last year: your memory continues to inspire and comfort me. This book is dedicated in part to your loving spirit.

Finally, I want to thank you, the reader, for your interest in this exciting field. It is my hope that this book will inspire you as much as the journey of writing it has inspired me.

To all who have contributed directly or indirectly to this work, my sincere thanks. Any errors or oversights in this book are entirely my own.

New York City, NY, USA
June, 2024

Baihan Lin

Contents

Part I A New Learning Paradigm in Speech and Language Technology

- 1 RL+SLT: Emerging RL-Powered Speech and Language Technologies** 3
 - 1.1 Introduction..... 3
 - 1.2 Defining RL+SLT 4
 - 1.3 RL+SLT in Action 5
 - 1.4 Challenges and Opportunities in RL+SLT 7
 - 1.5 Concluding Thoughts 8
 - References..... 10
- 2 Why Is RL+SLT Important, Now and How?** 11
 - 2.1 The Labyrinth of Speech and Language Technologies 11
 - 2.2 The Many Faces of Uncertainty 14
 - 2.3 RL Addresses the Exploration vs. Exploitation Trade-Off 15
 - 2.4 RL Addresses Challenges in Real-World Deployment 17
 - 2.5 The Rise of Large Language Models and the Need for RL 17
 - 2.6 Transforming Education and Healthcare with RL+SLT 18
 - 2.7 Ethical Considerations and Alignments in RL+SLT..... 18
 - 2.8 Concluding Thoughts 19
 - References..... 20

Part II Bandits and Reinforcement Learning: A Gentle Introduction

- 3 Introduction to the Bandit Problems** 23
 - 3.1 Introduction..... 23
 - 3.2 Exploration vs. Exploitation Trade-Off 24
 - 3.3 Multi-armed Bandits 25
 - 3.4 Contextual Bandits 30

3.5	Concluding Thoughts	33
	References	33
4	Reinforcement Learning: Preliminaries and Terminologies	37
4.1	Introduction	37
4.2	The RL Landscape: Settings and Scenarios	38
4.3	Markov Decision Process	40
4.4	Reinforcement Learning Formulations	41
4.5	Concluding Thoughts	43
	Reference	43
5	The RL Toolkit: A Spectrum of Algorithms	45
5.1	Introduction	45
5.2	Categorization of Reinforcement Learning Approaches	46
5.3	Value-Based Reinforcement Learning	47
5.4	Policy-Based Reinforcement Learning	49
5.5	Hybrid Approach: Actor-Critic Algorithm	50
5.6	Other Types of Reinforcement Learning	51
5.7	Concluding Thoughts	52
	References	52
6	Inverse Reinforcement Learning Problem	53
6.1	Introduction	53
6.2	Problem Setting	54
6.3	Strategies of Inverse Reinforcement Learning	56
6.4	Connections to Generative Adversarial Networks	59
6.5	Concluding Thoughts	59
	References	61
7	Behavioral Cloning and Imitation Learning	63
7.1	Introduction	63
7.2	Problem Setting	64
7.3	Distribution Shift Problem	64
7.4	Strategies to Mitigate the Challenge	65
7.5	Concluding Thoughts	67
	References	67
 Part III Reinforcement Learning in SLT Applications		
8	Reinforcement Learning Formulations for Speech and Language Applications	71
8.1	Introduction	71
8.2	Markov Decision Processes and RL in Speech and Language	72
8.3	Reinforcement Learning for Sequential Decision-Making	73
8.4	Metrics and Evaluation for RL in Speech and Language	73
	8.4.1 Task-Specific Metrics	74
	8.4.2 Reward-Based Metrics	74

8.4.3	User-Centric Evaluations	74
8.5	Chapter Preview	75
8.6	Concluding Thoughts	75
9	Reinforcement Learning in Automatic Speech Recognition (ASR): The Voice-First Revolution	79
9.1	Introduction	79
9.2	Challenges and Opportunities in ASR with RL	82
9.3	Cutting Through the Noise: RL for Robust ASR	83
9.4	Breaking Language Barriers: RL for Low-Resource ASR	84
9.5	Fine-Tuning for Success: RL-Driven Hyperparameter Optimization	86
9.6	Learning from the Unspoken: RL with Unlabeled Data	86
9.7	End-to-End ASR and Beyond with RL	88
9.8	Practicality and Limitations	88
9.9	Concluding Thoughts	89
	References	89
10	Reinforcement Learning in Speaker Recognition and Diarization: Decoding the Voices in the Crowd	91
10.1	Introduction	91
10.2	Challenges and Opportunities in Speaker Recognition and Diarization with RL	94
10.3	Real-Time Voice Tracking: RL for Online Diarization	96
10.4	RL for Unified Online and Offline Speaker ID	98
10.5	Multi-modal Speaker Identification with RL	100
10.6	Emerging Frontiers of RL-Powered Speaker Recognition and Diarization	100
10.7	Practicality and Limitations	102
10.8	Concluding Thoughts	103
	References	103
11	Reinforcement Learning in Natural Language Understanding (NLU): Teaching Machines to Comprehend	105
11.1	Introduction	105
11.2	Challenges and Opportunities in NLU with RL	107
11.3	From Pixels to Understanding: RL Unlocks Meaning in Text-Based Games	109
11.4	Your AI, Your Language: RL-Powered Personalized Understanding	110
11.5	Teaching Machines to Learn Better: NLU as a Catalyst for RL ...	111
11.6	Emerging Frontiers of RL x NLU	112
11.7	Practicality and Limitations	113
11.8	Concluding Thoughts	113
	References	114

12 Reinforcement Learning in Spoken Language Understanding (SLU): Giving Machines an Ear for Understanding	115
12.1 Introduction	115
12.2 Challenges and Opportunities in SLU with RL	118
12.3 Bridging Language Gaps: RL’s Role in Preserving Meaning Across Tongues	119
12.4 Siri’s Learning Journey: RL-Powered Decisions in Voice Assistants	121
12.5 Learning from Every Conversation: Human-Guided RL for SLU Enhancement	122
12.6 Interactive SLU for Communication Disorders	123
12.7 Practicality and Limitations	125
12.8 Concluding Thoughts	126
References	126
13 Reinforcement Learning in Text-to-Speech (TTS) Synthesis: Giving Machines a Voice	127
13.1 Introduction	127
13.2 Challenges and Opportunities in Sequence Generation and TTS with RL	129
13.3 From Text to Talk: RL’s Role in Sequence-to-Sequence Speech Synthesis	131
13.4 Speaking on the Fly: RL-Powered Real-Time Speech Generation	132
13.5 Emerging Frontiers of RL-Powered TTS	133
13.6 Practicality and Limitations	134
13.7 Concluding Thoughts	134
References	135
14 Reinforcement Learning in Natural Language Generation (NLG): Machines as Wordsmiths	137
14.1 Introduction	137
14.2 Challenges and Opportunities in NLG with RL	139
14.3 Crafting Conversations: RL and Inverse RL in Dialogue Creation	140
14.4 The AI Improv Duo: Balancing Rewards in Multi-agent Dialogue Systems	142
14.5 Emerging Frontiers of RL-Powered NLG	143
14.6 Practicality and Limitations	144
14.7 Concluding Thoughts	144
References	145
15 Reinforcement Learning in Large Language Models (LLMs): The Rise of AI Language Giants	147
15.1 Introduction	147
15.2 Challenges and Opportunities in LLM with RL	149
15.3 RL with Human Feedback: Human Touch in Shaping AI Giants	149

15.4	AI Teaching AI: The Self-Improving Ecosystem of Language Models	151
15.5	The Best of Both Worlds: Hybrid Feedback for Safer, Smarter LLMs	152
15.6	Emerging Frontiers of RL-Powered LLMs	154
15.7	Practicality and Limitations	155
15.8	Concluding Thoughts	155
	References	156
16	Reinforcement Learning in Conversational Recommendation Systems (CRS): AI's Personal Touch	157
16.1	Introduction	157
16.2	Challenges and Opportunities in CRS with RL	159
16.3	Exploring vs. Exploiting: The Bandit's Dilemma in Conversational AI	160
16.4	AI in the Therapist's Chair: RL-Powered Clinical Recommendations	161
16.5	Emerging Frontiers of RL-Powered CRS	164
16.6	Practicality and Limitations	165
16.7	Concluding Thoughts	166
	References	166
 Part IV Advanced Topics and Future Avenues		
17	Emerging Strategies in Reinforcement Learning Methods	171
17.1	Introduction	171
17.2	Recent Advances in Deep Reinforcement Learning and Bandits ..	172
17.3	Off-Policy Evaluation and Learning in Reinforcement Learning ..	173
17.4	Transfer Learning Methods in Reinforcement Learning	176
17.5	Concluding Thoughts	177
	References	178
18	Navigating the Frontiers: Key Challenges and Opportunities in RL-Powered Speech and Language Technology	181
18.1	Introduction	181
18.2	The AI Conversation Party: Exploring Multi-agent RL in Language Dynamics	182
18.3	Policy Juggling: Balancing Multiple Objectives and Systems in RL-Driven Language Processing	185
18.4	Decoding the Black Box: Making RL+SLT Models Transparent and Understandable	187
18.5	The Human Element: Ethical Frontiers in AI-Human Collaborative Language Systems	188
18.6	Concluding Thoughts	190
	References	190

19	Reflections, Resources, and Future Horizons in RL+SLT	193
19.1	Introduction	193
19.2	Summary of RL+SLT Concepts and Applications	194
19.3	Summary of Emerging Frontiers and Open Challenges in RL+SLT	195
19.4	Resources for Further Exploration	196
19.4.1	Relevant Texts and Recent Surveys	196
19.4.2	Practical Resources and Code Repositories	196
19.5	Future Horizons of RL+SLT	197
19.6	Concluding Thoughts	199
	References	200
Index		201

Part I
**A New Learning Paradigm in Speech and
Language Technology**

Chapter 1

RL+SLT: Emerging RL-Powered Speech and Language Technologies



Abstract Imagine a world where your virtual assistant understands your preferences, adapts to your needs, and provides you with a truly personalized experience. Or picture a language learning app that tailors its lessons to your unique learning style, ensuring you stay engaged and motivated throughout your journey. These are just a couple of the exciting possibilities that emerge when we combine reinforcement learning (RL) with speech and language technologies (SLT). The intersection of reinforcement learning (RL) and speech and language technologies (SLT) has given rise to a new paradigm in creating autonomous, adaptive, and intelligent systems that can learn and improve through interaction with their environments. In this chapter, we'll introduce you to the fascinating world of RL+SLT, a new paradigm that's set to revolutionize the way we design and deploy speech and language systems in real-world, real-time scenarios, as well as the way we interact with technology. By leveraging the power of RL as a universal learning mechanism, these systems can continuously adapt and optimize their behavior based on feedback and rewards, leading to improved performance and user experience. We'll explore the key concepts behind RL+SLT, share some captivating real-world examples, and discuss the challenges and opportunities that lie ahead. Welcome, and get ready to embark on an exciting adventure into the future of speech and language technologies.

1.1 Introduction

Picture this. You're having a conversation with your smartphone's virtual assistant, and it feels like you're talking to a close friend who understands your every need and preference. It remembers your favorite restaurants, suggests movies based on your watching history, and even cracks jokes that make you laugh out loud. This level of personalization and adaptability might seem like something out of a sci-fi movie, but with the rise of reinforcement learning in speech and language technologies (RL+SLT), it's quickly becoming a reality.

In recent years, artificial intelligence (AI) and machine learning (ML) have taken the world by storm, transforming industries and reshaping the way we live and work, and enabling the development of increasingly sophisticated and capable systems. One of the most exciting applications of AI and ML is in the field of speech and language technologies (SLT), where machines are learning to understand, interpret, and generate human language with unprecedented accuracy and fluency. From virtual assistants and chatbots to automatic speech recognition and machine translation, SLT have become an integral part of our daily lives, facilitating more natural and efficient human-machine interaction.

However, traditional approaches to SLT often rely on supervised learning techniques, which require large amounts of labeled data and can struggle to adapt to the dynamic and unpredictable nature of real-world environments. This is where reinforcement learning (RL) comes into play, offering a powerful framework for creating autonomous and adaptive systems that can learn and improve through interaction and feedback, just like humans do.

The fusion of RL and SLT, which we call *RL+SLT*, is a game-changer in the world of AI. It opens up a whole new realm of possibilities, from personalized virtual assistants that can anticipate your needs to adaptive language learning systems that can help you master a new language in record time. In this chapter, we'll take you on a journey through the exciting world of RL+SLT, introducing you to its key principles, showcasing some tangible real-world examples, and exploring the challenges and opportunities in this nascent field, highlighting the importance of interdisciplinary collaboration and the potential for RL+SLT to shape the future of human-machine interaction. As we delve into the world of RL+SLT, we invite you to consider the potential impact of this technology on your own work and research and to join us in exploring the exciting possibilities that lie ahead.

1.2 Defining RL+SLT

Before diving into the applications and potential of RL+SLT, it is essential to define what we mean by this term. In a nutshell, RL+SLT refers to the integration of reinforcement learning techniques with speech and language technologies to create smart, adaptive systems that can learn and improve through interaction.

Reinforcement learning is a branch of machine learning that's all about learning by doing. It focuses on how intelligent agents should take actions in an environment to maximize cumulative rewards [1]. In an RL setting, an agent interacts with its environment by observing the current state, taking an action, and receiving a reward signal based on the outcomes of those actions. The goal of the agent is to learn a strategy (called a policy), which maps states to actions that maximize the expected cumulative reward over time. Based on the reward feedback attributed by the underlying reward function of this environment, the agent learns how to perform the task of interest through trial and error. It's kind of like learning to play a game:

you try different moves, see what works and what doesn't, and gradually improve your skills until you can beat the competition.

Now, when we apply RL to speech and language technologies, things get more interesting. Imagine an AI agent that can learn to understand and respond to human speech, adapt to different accents and speaking styles, and even carry out complex conversations. By using RL techniques, the agent can continuously learn and improve its language skills based on the feedback it receives from users.

The field of SLT encompasses two cornerstones of modern-day analytic approaches, speech processing and natural language processing (NLP), and is innately sequence learning problems to extract information from these linguistic or speech signals and provide insights into interactive systems to communicate in human understandable languages. The sequential and interactive nature of these problems can make them well-suited into the sequential decision-making framework of RL.

For example, let's say you're chatting with a virtual assistant powered by RL+SLT. The assistant might try different responses to your queries, gauge your reaction (e.g., through your tone of voice or the words you use), and adjust its approach accordingly. If you respond positively to a particular type of answer, the assistant will learn to favor that approach in future conversations. Over time, the assistant will become increasingly attuned to your preferences and communication style, providing a truly personalized experience. In the language of RL, by receiving these *reward* signals based on user satisfaction, task completion, or other relevant metrics, the system can continuously update its *policy* to maximize these *rewards* over time, which in this case is a more engaging and effective interaction experience for you.

The beauty of RL+SLT is that it allows machines to *learn from real-world interactions*, rather than relying solely on pre-programmed rules or limited datasets. This makes RL+SLT systems incredibly adaptable and capable of handling the complexities and nuances of human language. By exploring the space of possible actions and receiving feedback in the form of rewards, RL agents can discover optimal strategies that may not be apparent from traditional supervised learning approaches. This is because in many of the SLT applications, the desired behavior is complex, context-dependent, and difficult to specify a priori. And as we'll see in the next section, the potential applications of this technology are nothing short of incredible.

1.3 RL+SLT in Action

To illustrate the transformative potential of RL+SLT, let's explore two compelling examples of its application in real-world scenarios: personalized virtual assistants and adaptive language learning systems.