Synthesis Lectures on
Human Language Technologies

SYNTHESIS
COLLECTION OF TECHNOLOGY

Sahisnu Mazumder · Bing Liu

Lifelong and Continual Learning Dialogue Systems

Springer

Synthesis Lectures on Human Language Technologies

The series publishes topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Sahisnu Mazumder · Bing Liu

Lifelong and Continual Learning Dialogue Systems

Springer

Sahisnu Mazumder
Intel Labs
Santa Clara, CA, USA

Bing Liu
University of Illinois Chicago
Chicago, IL, USA

ISSN 1947-4040 ISSN 1947-4059 (electronic)
Synthesis Lectures on Human Language Technologies
ISBN 978-3-031-48188-8 ISBN 978-3-031-48189-5 (eBook)
https://doi.org/10.1007/978-3-031-48189-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Sahisnu dedicates this book to his parents Ramesh Ch. Mazumder and Dulali Mazumder, wife Mohana, sister Snigdha, brother-in-law Sudip, and (late) aunt Malaya Goswami.

Bing dedicates this book to his wife Yue He, his children, Shelley and Kate, and his parents Yuhan Liu and Xianzhen Dong.

Preface

The purpose of writing this book is to introduce the emerging topic of lifelong/continual learning dialogue systems. Dialogue systems, also commonly known as chatbots, are computer programs that can converse with humans to perform intended tasks. They typically employ text or speech modes. As deep learning has improved the accuracy of speech recognition, text generation, and processing dramatically since 2012 or so, dialogue systems are becoming increasingly ubiquitous and are used in all types of applications such as smartphones, cars, home appliances, company websites, and mobile robots. They provide a very rich set of services, e.g., performing some specific tasks and chit-chatting with human users. However, the user experiences have not been very satisfactory so far (discussed in Chap. 1). It is clearly unfair to expect a deployed dialogue system to be able to understand everything that users may say, but these systems should be able to learn during conversations by interacting with users to become more and more knowledgeable and powerful. That is the goal of building lifelong learning dialogue systems. That is also the key motivation for us to write this book to introduce and promote the research of lifelong learning dialogue systems.

The project of writing this book started with a tutorial titled *Continual Learning Dialogue Systems—Learning on the Job after Model Deployment* that we gave at the *2021 International Joint Conference on Artificial Intelligence* (IJCAI-2021), August 21–26, 2021, Montreal, Canada. As we believe that lifelong learning dialogue systems is a very important topic for the future of dialogue systems and AI, we decided to develop the tutorial into a book. Our original interest in the topic stemmed from our research in lifelong/continual machine learning, dialogue systems, and natural language processing. Over the years, we have used many dialogue systems in smartphones and customer support websites. Our experiences have mostly been less than satisfactory. It is very natural to ask the question of why the dialogue system cannot communicate with users and learn to improve itself when it cannot understand what the user says as we humans do. Most of the deployed dialogue systems work in multi-user environments, e.g., Amazon Alexa, Apple Siri, and Google Assistant. If these systems can learn even a tiny amount of knowledge from each user when they get stuck, they will become very knowledgeable and smart over time. Another reason for our interest in lifelong learning dialogue systems is that

our research group has been working on lifelong/continual learning for many years. This combination of factors encouraged us to work on the topic and to write this book.

As lifelong learning dialogue systems sit at the intersection of dialogue systems in natural language processing and lifelong/continual learning in machine learning, this book will touch both fields. We aim to present a comprehensive survey and review of the important research results and latest ideas in these areas. We also want to propose a theoretical framework to be used to guide future research and development in the field. This framework is called SOLA (*Self-initiated Open-world continual Learning and Adaptation*), which was originally proposed for building autonomous and continual learning AI agents. Since lifelong learning dialogue systems are such agents, the framework is naturally suited for the topic. Presently, there are several research topics in dialogue systems that are closely related to lifelong learning dialogue systems. This book will bring all these topics under one roof and discuss their similarities as well as differences. Through this book, we would also like to motivate and encourage researchers to work on and practitioners to build lifelong learning dialogue systems that can be deployed for practical use to improve user experiences and to make dialogue systems smarter and smarter over time. Without the capability of continually learning, accumulating knowledge, making inference about it, and using the knowledge to help future learning and problem solving, achieving true intelligence for AI agents is unlikely.

Two principles have guided the writing of this book. First, it should contain strong motivations for conducting research in lifelong learning dialogue systems in order to encourage graduate students and researchers to work on the problem. Second, the writing should be accessible to practitioners and upper-level undergraduate students who have basic knowledge of natural language processing and machine learning. Yet there should be sufficient in-depth materials for graduate students who plan to pursue Ph.D. degrees in dialogue systems, lifelong learning, or their integration of lifelong learning dialogue systems. We also strongly believe that lifelong learning dialogue systems can be built and deployed for practical applications.

This book is suitable for students, researchers, and practitioners who are interested in dialogue systems, natural language processing, and machine learning. Lecturers can readily use the book in class for courses in any of these related fields.

Santa Clara, USA — Sahisnu Mazumder
Chicago, USA — Bing Liu
June 2023

Acknowledgments

We would like to thank the current and former graduate students in our group and our collaborators: Joydeep Biswas, Jiahua Chen, Zhiyuan Chen, Daniel A. DeLaurentis, Sepideh Esmaeilpour, Neta Ezer, Geli Fei, Hasan Ghadialy, Yiduo Guo, Scott Grigsby, Estevam R. Hruschka Jr., Wenpeng Hu, Minlie Huang, Reid Hyland, Adam Kaufman, Zixuan Ke, Gyuhak Kim, Tatsuya Konishi, Mori Kurokawa, Huayi Li, Jian Li, Tianrui Li, Yanni Li, Haowei Lin, Zhou Lin, Lifeng Liu, Qian Liu, Guangyi Lv, Chuhe Mei, Piero Molino, Chihiro Ono, Arjun Mukherjee, Nianzu Ma, Shaoshuai Mou, Alexander Politowicz, Qi Qin, Steven Rizzi, Eric Robertson, Gokhan Tur, Vipin Vijayan, Lei Shu, Hao Wang, Mengyu Wang, Yijia Shao, Shuai Wang, Hu Xu, Yueshen Xu, Rui Yan, Yan Yang, Tim Yin, Tim Yuan, Philip S. Yu, Lei Zhang, Dongyan Zhao, Hao Zhou, and Xiaoyan Zhu for their contributions of numerous research ideas and helpful discussions over the years.

We are especially grateful to the anonymous reviewer who read the draft manuscripts very carefully and gave us so many excellent comments and suggestions, which have helped us improve the book tremendously. This book was also benefited significantly from numerous discussions in DARPA SAIL-ON Program meetings.

On the publication side, we thank the editor of Synthesis Lectures on Natural Language Processing, Graeme Hirst, for his valuable guidance in the project. The President and CEO of Morgan & Claypool Publishers, Michael Morgan, and his staff Christine Kiilerich, Springer manuscript editor Susanne Filler, and production editor Bharath Kumar Dhamodharan have given us all kinds of help promptly whenever requested, for which we are very grateful.

Our greatest gratitude goes to our own families. Sahisnu Mazumder would like to thank his parents Ramesh Ch. Mazumder and Dulali Mazumder, wife Mohana, elder sister Snigdha, brother-in-law Sudip, and (late) aunt Malaya Goswami for their invaluable support and encouragements. Bing Liu would like to thank his wife Yue, children Shelley and Kate, and parents Yuhan Liu and Xianzhen Dong. They have helped in so many ways.

The writing of this book was partially supported by a SAIL-ON Program Contract (HR001120C0023) of the Defense Advanced Research Projects Agency (DARPA), Northrop Grumman research gifts, three National Science Foundation (NSF) grants (1650900, IIS-1910424 and IIS-1838770), an NCI grant R01CA192240, and a Research

Contract with KDDI. The content of the book is solely the responsibility of the authors and does not necessarily represent the official views of DARPA, Northrop Grumman, NSF, NCI, KDDI, UIC, or Intel.

The Department of Computer Science at the University of Illinois at Chicago provided computing resources and a very supportive environment for this project. We would particularly like to thank Patricia Brianne Barrera, Sheri Lyn Joscelyn, Denise Marie Kelly, Emily Lam, Sherice Nelson, and Ivy Yuan for their valuable support in the research projects related to this book.

Sahisnu Mazumder would also like to thank Intel Corporation, his colleagues at Multimodal Dialogue and Interaction (MDI) group and Intelligent Systems Research (ISR) division at Intel Labs, and, specifically, Saurav Sahay and Lama Nachman for their invaluable support, advice, and encouragement.

June 2023

Sahisnu Mazumder
Bing Liu

Contents

About the Authors

Sahisnu Mazumder is an AI Research Scientist at Intel Labs, Santa Clara, CA, USA, where he works on Human-AI Collaboration and Dialogue & Interaction Systems research. His research interests include Lifelong and Continual Learning, Dialogue and Interactive Systems, Open-world AI/Learning, Knowledge Base Reasoning, and Sentiment Analysis. He obtained his Ph.D. in Computer Science at the University of Illinois at Chicago, USA, and his Masters in Computer Science from the Indian Institute of Technology (IIT)—Roorkee, India. During his Ph.D., he worked as a Research Intern at Huawei Research, USA, on projects related to user activity and interest mining and at Microsoft Research—Redmond on Natural Language Interaction (NLI) system design. He has published several research papers in leading AI, NLP, and Dialogue conferences like AAAI, IJCAI, ACL, EMNLP, NAACL, SIGDIAL, and CIKM and delivered tutorials in SIGIR—2022, IJCAI—2021, and Big Data Analytics (BDA—2014). He has also served as a Session Chair in COLING 2022 and PC Member/Reviewer of premier conferences like AAAI, IJCAI, ACL, EMNLP, NAACL, EACL, and COLING and journals like ACM TALLIP, Computational Linguistics, and IEEE TNNLS.

Bing Liu is a Distinguished Professor of Computer Science at the University of Illinois at Chicago, USA. He received his Ph.D. in Artificial Intelligence from the University of Edinburgh, UK. His current research interests include Lifelong and Continual Learning, Lifelong Learning Dialogue Systems, Open-World Learning, Sentiment Analysis and Opinion Mining, Machine Learning, and Natural Language Processing. His previous interests include Fake Review Detection, Web Mining, and Data Mining. He has published extensively in top conferences and journals in these areas and authored four books: one about Lifelong/Continual Machine Learning, two about Sentiment Analysis, and one about Web Mining. Three of his papers have received Test-of-Time awards and another one received Test-of-Time honorable mention. Some of his works have been widely reported in popular and technology press internationally. He has served as the Chair of ACM SIGKDD from 2013 to 2017, as program chair of many leading data mining conferences including KDD, ICDM, CIKM, WSDM, SDM, and PAKDD, and as

associate editor of many leading journals such as TKDE, TKDD, TWEB, and DMKD. He is the winner of 2018 ACM SIGKDD Innovation Award, and is a Fellow of AAAI, ACM, and IEEE.

Introduction

1

Building *dialogue systems* or *conversational agents* capable of conversing with humans in natural language (NL) and understanding human NL instructions is a long-standing goal of AI (Winograd 1972). These systems, also known as *chatbots*, have become the front runner of AI advancement due to wide-spread applications such as assisting customers in buying products, booking tickets, reducing stress, and executing actions like controlling house appliances and reporting weather information. However, the user experiences have not been fully satisfactory so far (Huang et al. 2020; Li et al. 2016a; Schaub et al. 2021). There are many weaknesses with the current research and fielded dialogue systems. One of the major weaknesses is that they do not learn continuously during conversation (i.e., post-deployment) with the user after they are deployed in practice. Building lifelong learning dialogue systems that possess the capability of continuous learning during conversation is the main topic of discussion of this book. This chapter aims to motivate and provide the foundational idea of building such dialogue systems. Note that, we use the term *chatbots* to refer to all kinds of conversational agents, such as dialogue systems, personal assistants, conversational question-answering systems etc., onward.

In the following sections, we first provide some background of modern dialogue systems (Sect. 1.1) and discuss their general weaknesses (Sect. 1.2), which provide the motivations for studying and building lifelong learning dialogue systems (introduced in Sects. 1.3 and 1.4). Finally, we conclude the chapter with a discussion on the organization of this book in Sect. 1.5.

S. Mazumder and B. Liu, *Lifelong and Continual Learning Dialogue Systems*,
Synthesis Lectures on Human Language Technologies,
https://doi.org/10.1007/978-3-031-48189-5_1

1.1 Dialogue and Interactive Systems: Background

Dialogue systems can broadly be categorized into two main types (Gao et al. 2019; Jurafsky and Martin 2020):

(1) **Chit-chat systems** (Chen et al. 2017; Li et al. 2016a, c; Mei et al. 2017; Pandey et al. 2018; Serban et al. 2016, 2018; Shang et al. 2015; Shen et al. 2017, 2018; Vinyals and Le 2015; Wu et al. 2017, 2018; Xing et al. 2017; Yao et al. 2015) are chabots designed to engage users by conducting the chit-chat type of conversation on a wide range of topics *without having a specific goal to complete.* Examples include *Social Chatbots* like ELIZZA (Weizenbaum 1966), PARRY (Colby 1975), ALICE (Wallace 2009), Microsoft XiaoIce (Zhou et al. 2020), AliMe Chat (Qiu et al. 2017), and many systems in Alexa Prize Socialbot Grand Challenge (Khatri et al. 2018). Such chatbots are built with the goal of supporting seamless conversation with users, and helping them with useful recommendations and mental supports.

(2) **Task-oriented chatbots** (Bordes et al. 2017; Budzianowski et al. 2018; Lowe et al. 2017; Luo et al. 2019; Shah et al. 2018; Wen et al. 2017a, b; Williams and Young 2007; Zhao et al. 2017) are chatbots designed to assist users to complete tasks based on users' requests, e.g., providing the requested information and taking actions. Most of the popular personal assistants such as Alexa, Siri, Google Home, and Cortana, are task-oriented chatbots. Besides, these types of chatbots are also built as QA Bots to support Question-answering (QA) over knowledge bases, conversational recommendation systems for online product or service recommendations to end-users and as Natural Language Interaction (NLI) systems to enable natural language (NL) driven task completion.

Although the broad goal of task-oriented chatbots is to perform actions or tasks on users' behalf, they can also be of two types based on their design and the nature of interaction with users whom they are meant to support: **(1)** Systems that achieve task completion through *multi-turn dialogues* with users where users express their intents and refine them (based on the feedback from the chatbot) over a sequence of dialogue turns with the system. They are formally known as **Task-oriented Dialogue Systems** (**ToDS**). **(2)** The other kind of systems that intend to accomplish tasks through a *single-turn dialogue* where the user provides a NL instruction (command) and the system's goal is to just interpret it by translating it into some actions to be executed by the underlying application. Such systems are formally known as **Natural Language Interfaces** (**NLIs**). NLIs are also sometimes referred to as Natural Language Interaction systems in general. Although NLIs are mostly built to support a single-turn interaction with user per task completion goal, they can engage in muti-turn dialogues with user as well (similar to traditional ToDS systems) to better understand the NL instruction and resolve ambiguities (if any) to serve the user better.

Recently, some works have been done to bridge the gap of the two types of chatbots by fusing them through mode switching (Yoshino et al. 2022; Young et al. 2022), which is typically done using a classification model. The system in (Yoshino et al. 2022) also has a module that generates utterances to bridge the two types of chats to ensure a seamless transition.

1.1.1 Task-Oriented Dialogue Systems

A full-fledged Task-oriented Dialogue Systems (ToDS) is mostly designed as a modular system, having six modules, viz., Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Dialogue State Tracking (DST), Dialogue Policy (DP) Learning, Natural Language Generation (NLG) and Text-to-speech (TTS) Synthesis. Figure 1.1 shows the architecture of a typical ToDS with all the modules integrated with each other. Here, the ASR module is responsible for translating the spoken utterance from the user into text, which is fed to the NLU module for language understanding and the TTS module generates the speech from text which is the NL response generated by the NLG module. The DST and DP modules are often unified and referred to as Dialogue Manager (DM) that is responsible for the progression of the dialogue by managing the dialogue turns.

Often, ASR and TTS modules are studied as a separate field of research and existing works mostly focus on the remaining four components and their interactions in a typical ToDS framework. In particular, they assume ASR and TTS are available to use and deal with only text-based user input and text-based generated ouput by the rest four modules. Thus, in the rest of the book, we mainly focus on NLU, DST, DP and NLG as the four main modules of a typical ToDS system.

(1) **Natural Language Understanding (NLU)**: The goal of the NLU module is to identify the user intents and extract associated information (slots) from the user utterance. In general, NLU involves solving three subtasks:

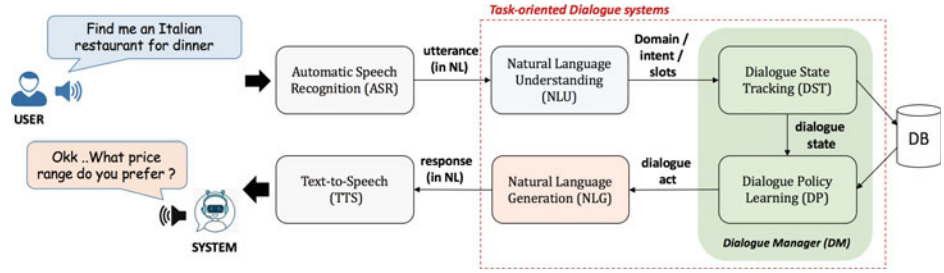

Fig. 1.1 A typical task-oriented dialogue system with its modules