Theory and Applications of Natural Language Processing
Monographs

Verena Rieser
Oliver Lemon

Reinforcement Learning for Adaptive Dialogue Systems

A Data-driven Methodology for Dialogue Management and Natural Language Generation

Springer

Theory and Applications
of Natural Language Processing

Series Editors:
Graeme Hirst (Textbooks)
Eduard Hovy (Edited volumes)
Mark Johnson (Monographs)

Aims and Scope

The field of Natural Language Processing (NLP) has expanded explosively over the past decade: growing bodies of available data, novel fields of applications, emerging areas and new connections to neighboring fields have all led to increasing output and to diversification of research.

"Theory and Applications of Natural Language Processing" is a series of volumes dedicated to selected topics in NLP and Language Technology. It focuses on the most recent advances in all areas of the computational modeling and processing of speech and text across languages and domains. Due to the rapid pace of development, the diversity of approaches and application scenarios are scattered in an ever-growing mass of conference proceedings, making entry into the field difficult for both students and potential users. Volumes in the series facilitate this first step and can be used as a teaching aid, advanced-level information resource or a point of reference.

The series encourages the submission of research monographs, contributed volumes and surveys, lecture notes and textbooks covering research frontiers on all relevant topics, offering a platform for the rapid publication of cutting-edge research as well as for comprehensive monographs that cover the full range of research on specific problem areas.

The topics include applications of NLP techniques to gain insights into the use and functioning of language, as well as the use of language technology in applications that enable communication, knowledge management and discovery such as natural language generation, information retrieval, question-answering, machine translation, localization and related fields.

The books are available in printed and electronic (e-book) form:

* Downloadable on your PC, e-reader or iPad
* Enhanced by Electronic Supplementary Material, such as algorithms, demonstrations, software, images and videos
* Available online within an extensive network of academic and corporate R&D libraries worldwide
* Never out of print thanks to innovative print-on-demand services
* Competitively priced print editions for eBook customers thanks to MyCopy service http://www.springer.com/librarians/e-content/mycopy

For further volumes:
http://www.springer.com/series/8899

Verena Rieser • Oliver Lemon

Reinforcement Learning for Adaptive Dialogue Systems

A Data-driven Methodology for Dialogue Management and Natural Language Generation

Springer

Verena Rieser
School of Mathematical
and Computer Sciences
Heriot-Watt University
Edinburgh EH14 4AS
United Kingdom
v.t.rieser@hw.ac.uk

Oliver Lemon
School of Mathematical
and Computer Sciences
Heriot-Watt University
Edinburgh EH14 4AS
United Kingdom
o.lemon@hw.ac.uk

ISSN 2192-032X e-ISSN 2192-0338
ISBN 978-3-642-24941-9 e-ISBN 978-3-642-24942-6
DOI 10.1007/978-3-642-24942-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011942326

Mathematics Subject Classification (2010): 68-XX, 68Txx, 68T05, 68T50, 68T37, 68T42

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The past decade has seen a revolution in the field of spoken dialogue systems. As in other areas of Computer Science and Artificial Intelligence, data-driven methods are now being used to drive new methodologies for system development and evaluation. These methods are proving to be more robust, flexible, and adaptive than the largely rule-based approaches which preceded them.

We hope that this book is a contribution to that ongoing change. It describes, in detail, a new methodology for developing spoken dialogue systems – in particular the Dialogue Management and Natural Language Generation components – which starts with human data, and culminates in evaluation with real users. The journey therefore starts and ends with human behaviour in interaction, and explores methods for learning from the data, for building simulation environments for training and testing systems, and for evaluating the results.

The detailed material covers: Spoken and Multimodal dialogue systems, Wizard-of-Oz data collection, User Simulation methods, Reinforcement Learning, and Evaluation methodologies.

This book is therefore intended as research guide which navigates through a detailed case study in data-driven methods for development and evaluation of spoken dialogue systems. Common challenges associated with this approach are discussed and example solutions provided, for example, how to learn from limited amounts of data. As such, we hope it will provide insights, lessons, and inspiration for future research and development – not only for spoken dialogue systems in particular, but for data-driven approaches to human-machine interaction in general.

Edinburgh,
September 2011

Verena Rieser
Oliver Lemon

Acknowledgements

The research presented here would not have been possible without interaction with our research community, and we owe special thanks to the participants of the TALK and CLASSiC projects for discussing the ideas presented in this book.

There is a lively international research community in Spoken Dialogue Systems whose interest, criticism, and encouragement have helped to shape this work. We are especially indebted to Steve Young and his colleagues at Cambridge University, who have had a deep influence on our ideas. Members of the Interaction Lab[1] at Heriot-Watt University's School of Mathematics and Computer Science have also helped to discuss and develop these ideas over several years.

In particular we would also like to thank Dr. Xingkun Liu, Dr. Helen Hastie, Dr. Ivana Kruijff-Korbayová, Dr. Tilman Becker, and other colleagues from Saarland and Edinburgh Universities for helping with the data collection and evaluation involved in this work. We especially thank Professor Manfred Pinkal for his guidance.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7, 2007-2013) under grant agreement number 216594 ("Computational Learning in Adaptive Systems for Spoken Conversation", CLASSiC project: www.classic-project.org), the EC FP6 project "TALK: Talk and Look, Tools for Ambient Linguistic Knowledge" (IST 507802, www.talk-project.org), from EPSRC grant numbers EP/E019501/1 and EP/G069840/1, and from the International Research Training Group "Language Technology and Cognitive Systems", Saarland University, funded by the German Science Foundation (DFG).

Finally, the first author would like to thank her parents Franz and Tatjana Rieser for their support and encouragement. The second author thanks his family for the decades of rewarding learning experiences which have made this book possible.

[1] http://www.macs.hw.ac.uk/InteractionLab/

Contents

Acronyms

ASR	Automatic Speech Recognition
DA	Dialogue Act
DB	Database
DM	Dialogue Management
GUI	Graphical User Interface
HCI	Human Computer Interaction
IP	Information Presentation
ISU	Information State Update
MDP	Markov Decision Process
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
PARADISE	PARAdigm for DIalogue System Evaluation
POMDP	Partially Observable Markov Decision Process
RL	Reinforcement Learning
SA	Speech Act
SASSI	Subjective Assessment of Speech System Interfaces
SDS	Spoken Dialogue System
SL	Supervised Learning
SLU	Spoken Language Understanding
TTS	Text-to-Speech
VOIP	Voice-Over-Internet Protocol
WER	Word-Error Rate
WOZ	Wizard-of-Oz

Chapter 1
Introduction

The past decade has seen something of a revolution in the field of spoken dialogue systems. As in other areas of Computer Science and Artificial Intelligence, data-driven methods are being used to drive new methodologies for system development and evaluation. These methods are proving to be more robust, flexible, and adaptive than the rule-based approaches which preceded them.

We hope that this book makes a contribution to that revolution. It describes, in detail, a new methodology for developing spoken dialogue systems – in particular the Dialogue Management component – which starts with human data, and ends with evaluation with real users. Related methods are now being developed further by a number of researchers worldwide. The journey begins and ends with human behaviour in interaction, and en route we explore methods for learning from such data, for building simulation environments for training and testing our systems, and methods for evaluating the results.

This book is therefore intended as a guide which navigates through a detailed case study in data-driven methods for development and evaluation of spoken dialogue systems. It focusses on Dialogue Management and Natural Language Generation, rather than speech recognition and spoken language understanding. As such, we hope that it can provide insights and lessons for future research and development – not only for spoken dialogue systems in particular, but for data-driven approaches to building better human-machine interaction in general.

1.1 The Design Problem for Spoken Dialogue Systems

The design of Spoken Dialogue Systems (SDS) is not only concerned with integrating speech and language processing modules such as Automatic Speech Recognition (ASR), Spoken Language Understanding (SLU), Natural Language Generation (NLG), and Text-to-Speech (TTS) synthesis systems. It also requires the development of skills for "what to say next": *dialogue strategies* which take into account the performance of these components, the nature of the user's tasks (e.g.

V. Rieser and O. Lemon, *Reinforcement Learning for Adaptive Dialogue Systems*,
Theory and Applications of Natural Language Processing, DOI 10.1007/978-3-642-24942-6_1,

information-seeking, tutoring, or robot control), and other features of the operating environment such as the user's behaviour and preferences. The great variability and unpredictability of these factors makes dialogue strategy design an extremely difficult task for human developers. In conventional, rule-based, dialogue development many expensive iterations of manual design and re-design are necessary in order to produce good strategies. In addition, such hand-coded strategies are not re-usable from task to task, are not scalable, require a substantial amount of human labour and expertise, and are not guaranteed to be optimal.

For these reasons machine learning methods (such as Reinforcement Learning) for dialogue strategy design have been a leading research area for several years. These statistical computational learning approaches offer several key potential advantages over the standard rule-based hand-coding approach to dialogue systems development (Lemon and Pietquin, 2007):

- a data-driven automatic development cycle
- provably optimal action policies
- a principled mathematical model for action selection
- possibilities for generalisation to unseen states
- reduced development and deployment costs.

However, in cases where a system is designed from scratch, there is often no suitable in-domain data to enable such a design. Collecting dialogue data without a working prototype is problematic, leaving the developer with a classic "chicken-and-egg" problem. One of the main issues that this book addresses is how to use a data-driven development methodology when little or no in-domain data exists.

1.2 Overview

In this book we propose to learn dialogue strategies by simulation-based Reinforcement Learning (RL) (Sutton and Barto, 1998), where a simulated environment is learned from small amounts of Wizard-of-Oz (WOZ) data. Using WOZ data rather than data from real Human-Computer Interaction (HCI) allows us to learn optimal strategies for domains where no working dialogue system already exists. Automatic strategy learning has been applied to dialogue systems which have already been deployed in the real world using handcrafted strategies. In such work, strategy learning was performed based on already present extensive online-operation experience, e.g. (Henderson et al, 2005, 2008; Singh et al, 2002). In contrast to this preceding work, our approach enables strategy learning in domains where no prior system is available. Optimised learned strategies are then available from the first moment of online-operation, and labour-intensive handcrafting of dialogue strategies is avoided. This independence from large amounts of in-domain dialogue data allows researchers to apply RL to new application areas beyond the scope of existing dialogue systems. We call this method "bootstrapping".

This book first provides the general proof-of-concept that RL-based strategies outperform handcrafted strategies which are manually tuned for a wide spectrum of application scenarios. After theoretically motivating our approach, we turn to the practical problem of how to learn optimal strategies for new application domains where no prior system or in-domain data are available. We propose to learn dialogue strategies by simulation-based RL, where the simulated environment is learned from small amounts of WOZ data. We therefore introduce a 5-step procedure:

1. Collect data in a WOZ experiment.
2. Use this data to construct a simulated learning environment using data-driven methods only.
3. Train a RL-based dialogue policy by interacting with the simulated environment. We compare this policy against a supervised baseline. This comparison allows us to measure the relative improvements over the WOZ strategies contained in the training data.
4. Evaluate the learned policy with real users.
5. Show that "bootstrapping" from WOZ data is a valid estimate of real HCI by comparing different aspects of the 3 corpora gathered so far: the WOZ study, the dialogues generated in simulation, and the final user tests.

It should be noted that these steps are not unique to the method introduced in this book, but most of the defined steps are required for any simulation-driven approach to strategy learning (though the last step is specific to our method). A main contribution of this book is that all these steps are now performed starting with a limited WOZ data set, and specific methods are introduced to build and validate the obtained simulations.

We apply this framework to optimise multimodal[1] Dialogue Management strategies and Natural Language Generation. In the first case we consider Dialogue Management and content selection as two closely interrelated problems for information seeking dialogues: the decision of *when* to present information depends on *how many* pieces of information to present and the available options for *how* to present them, and vice versa. We therefore formulate the problem as a hierarchy of joint learning decisions which are optimised together.

The second study describes a new approach to generating Natural Language in interactive systems. Natural Language Generation (NLG) addresses the problem of "how to say" an utterance, once "what to say" has been determined by the Dialogue Manager. We treat NLG as planning under uncertainty for information-seeking dialogue systems, where the strategy for information presentation and its associated attributes are incrementally selected using hierarchical learning. This hierarchical approach to DM and NLG has recently been explored by other researchers (Dethlefs and Cuayahuitl, 2010; Dethlefs and Cuayáhuitl, 2011), and a utility-based approach to NLG is discussed by van Deemter (2009a).

[1] "Multimodal" dialogue systems are those which do not only use speech input and output, but which also use other information modalities, such as graphics (as in our case), or gesture, gaze, facial expressions, and so on.

Our results in both studies show that RL significantly outperforms supervised learning (SL) when interacting in simulation as well as for interactions with real users. For optimising multimodal Dialogue Management, the RL-based policy gains on average 50-times more reward than the SL policy when tested in simulation, and almost 18-times more reward when interacting with real users. Users also subjectively rate the RL-based policy on average 10% higher. For optimising Natural Language Generation, the trained information presentation strategies significantly improve dialogue task completion, with up to a 9.7% increase (30% relative) compared to the deployed dialogue system which uses conventional, hand-coded presentation prompts.

One focus of this book is to optimise dialogue strategies with respect to real user preferences. A major advantage of RL-based dialogue strategy development is that the dialogue strategy can be automatically trained and evaluated using the same objective function (Walker, 2005). Despite its central importance for RL, quality assurance for objective functions has received little attention so far. In fact, the reward function is one of the most hand-coded aspects of RL (Paek, 2006). Clearly, automatic optimisation and evaluation of dialogue policies, as well as quality control of the objective function, are closely inter-related problems: how can we make sure that we optimise a system according to real users' preferences? This book is the first to explore learning with data-driven, non-linear objective functions. We also propose a new method for meta-evaluation of the objective function.

Note that chapters 4 to 8 are significantly revised, updated, and extended versions of material from (Rieser, 2008).

1.3 Structure of the Book

Chapter 2 (Background)

This chapter provides the reader with relevant background knowledge for the research. After introducing some general information about Spoken Dialogue Systems, we contrast different methods applied in research and industry to develop dialogue strategies. We show how these two approaches fail to meet the current challenges for strategy design and argue for the use of statistical methods. In particular, we propose the use of Reinforcement Learning, as it uses trial-and-error exploration with delayed rewards which, we argue, is a natural model for human dialogue.

Chapter 3 (Reinforcement Learning)

This chapter provides technical background on RL for dialogue strategy development and discusses simulation-based learning in particular. We also introduce the application domain.

Chapter 4 (Proof-of-Concept: Information Seeking Strategies)

He we develop the theoretical proof-of-concept that RL-based strategies outperform hand-coded strategies, which are tuned to the same objective function. We show this for a wide range of application scenarios, e.g. for different user types and noise conditions. This chapter also demonstrates how to apply simulation-based RL to solve a complex and challenging problem for information-seeking dialogue systems: which questions to ask the user, how many database search results to present, and when to present them, given the competing trade-offs between the length of the results list, the length of the interaction, the type of database, and noise in the speech recognition environment. In this chapter the reader will receive a deeper understanding of the principles and advantages of Reinforcement Learning for dialogue strategy learning. We also explain why data-driven simulation approaches are preferred over manually constructed simulated environments (as done here for the theoretical proof-of-concept).

Chapter 5 (A Bootstrapping Approach to Develop Reinforcement Learning-based Strategies)

This chapter introduces a 5-step procedure model to bootstrap optimal RL-based strategies for WOZ data. The resulting strategies are tailored to the application environment, do not require a working prototype system, and are optimised with respect to real user preferences. We also explain how we meet the challenges when learning from WOZ data. In particular, we introduce the problem of how to construct a simulated environment from limited amounts of WOZ data. We also discuss the fact that a WOZ study itself is a simulation of real HCI.

Chapter 6 (Data Collection in a Wizard-of-Oz Experiment)

Here we describe the experimental setup of the WOZ experiment. We explain which changes to the conventional WOZ method are necessary for strategy learning. In particular, we introduce an utterance distortion method in order to resemble noise conditions for real dialogue systems. In addition, we explore the “intuitive” strategies that were applied by our human wizards.

Chapter 7 (Building a Simulated Learning Environment from Wizard-of-Oz Data)

This chapter uses the WOZ data to construct a simulated learning environment. We therefore introduce methods suited to build and validate simulations from small amounts of data. In particular, we construct the action set and state space by exploring the wizards’ actions. The user and noise simulations are obtained using frequency-based approaches. The objective function is a predictive model of user ratings obtained by a regression analysis, following the PARADISE framework of (Walker et al, 1997). We then train a RL-based dialogue policy by interacting with the simulated environment, and we compare this strategy against a baseline constructed by Supervised Learning. This comparison allows us to measure the relative improvements over the wizard strategies obtained from the training data.

Chapter 8 (Comparing Reinforcement and Supervised Learning of Dialogue Policies with Real Users)

In this chapter we evaluate the learned strategy with real users. We therefore develop a music-player dialogue system using a rapid development tool, where the learned strategy is implemented using a table look-up between states and learned actions. We report detailed results from the real user tests.

We also post-evaluate our overall "bootstrapping" approach by comparing different aspects of the 3 corpora gathered so far: the WOZ study, the dialogues generated in simulation, and the final user tests. We first evaluate whether strategies learned in simulation do transfer to tests with real users, and we also compare the experimental conditions of the different studies, where we discuss the noise model in particular. Furthermore, we explore whether the objective function used for learning is a realistic estimate of real user preferences.

Chapter 9 (Natural Language Generation)

This chapter further develops the methodology to encompass elements of policy learning for adaptive Natural Language Generation in spoken dialogue systems. This chapter shows that our method can quite easily be applied to new domains and tasks. We show how to develop a data-driven approach to content selection and structuring decisions in NLG, this time in the domain of a restaurant recommendation SDS. We also report results from evaluations with both simulated and real users. The real user evaluation shows that improved NLG can lead to significant improvements in overall dialogue system performance.

Chapter 10 (Conclusion)

Finally, we conclude by summarising the main contributions of this work. We also report on "lessons learned" to provide guidance for future researchers. For example, we discuss the data quantity and quality required for the proposed bootstrapping approach. In the final outlook of this chapter we discuss ongoing challenges for research in this field.

Part I
Fundamental Concepts

Chapter 2
Background

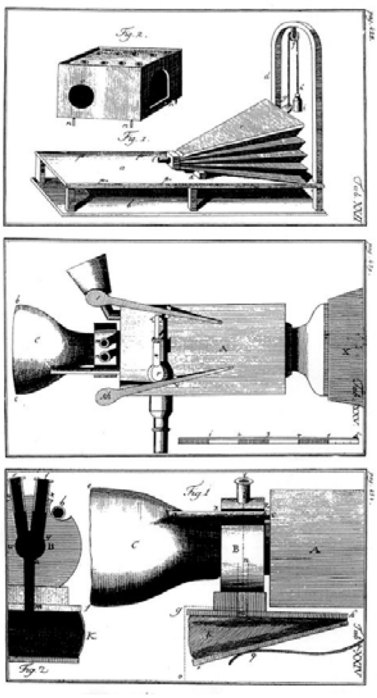

Fig. 2.1 Wolfgang von Kempelen's speaking machine, drawing from *Mechanism of Human Speech* (1791)

Mankind appears to be fascinated by the idea of talking with machines. The first attempts to produce human speech by machine were made in the second half of the 18th century. One of the best known examples is Wolfgang von Kempelen's speak-

V. Rieser and O. Lemon, *Reinforcement Learning for Adaptive Dialogue Systems*,
Theory and Applications of Natural Language Processing, DOI 10.1007/978-3-642-24942-6_2,

ing machine, as described in his book "Mechanism of Human Speech" [1] (1791), see Figure 2. Von Kempelen's machine was the first that produced not only some speech sounds, but also whole words and short sentences. Clearly, in order to converse with a machine in natural language more than just automatic sound production is necessary. Descartes even thought that it would never be possible to engage in dialogue with machines at all. In his book "Discourse on the Method of Rightly Conducting the Reason, and Searching for Truth in the Sciences" [2] (1637) he declares:

> "... but if there were machines bearing the image of our bodies, and capable of imitating our actions as far as it is morally possible, [...] they could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others."

Today, however, it is arguable that such machines now exist, at least for limited application domains, due to major advances in the field of Human-Computer Interaction (HCI) and Spoken Dialogue Systems (SDS) . Still, human-machine dialogue is far from resembling the capabilities of human-human dialogue, as we discuss below.

2.1 Human-Computer Interaction

For computers, holding a conversation is difficult. Engaging in a conversation requires more than just technical language proficiency. When people engage in dialogue, they carry out a purposeful *activity* (Austin, 1962), a *joint action* (Clark, 1996), or a *language game* (Wittgenstein, 1953), which they know how to perform using their communicative skills. Dialogue behaviour is often formally described as a sequence of Speech Acts (SAs) (Searle, 1969). These SAs are organised into structural patterns in the field of Conversation Analysis, e.g. (Levinson, 1983; Sacks et al, 1974). Some of this behaviour follows standardised cultural conventions. For example, the fact that people greet each other is described as the standardised SA "adjacency pair" `greeting-greeting` (Levinson, 1983). Other behaviour is highly context-dependent. For example, in previous work we show that the way people ask for clarification is influenced by various contextual and environmental factors, such as dialogue type, modality, and channel quality (Rieser and Moore, 2005). In addition, people often engage in dialogue to solve a task together. Their behaviour is then driven by the goal of the task as well as by their "mental model" (Johnson-Laird, 1983) of the other dialogue participant.

Humans acquire these communicative skills over time, but for a dialogue system, they need to be developed by a dialogue designer. This usually is an expert who defines a *dialogue strategy* , which "tells" the system what to do in specific situations. The "dialogue strategy" is part of the Dialogue Manager (DM) which controls the

---

[1] Original title: *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine*

[2] Original title: *Discours de la méthode pour bien conduire sa raison, et chercher la verité dans les sciences*

behaviour of the system. Broadly speaking, a dialogue system has three modules, one each for input, output, and control, as shown in Figure 2.2. The input module commonly comprises Automatic Speech Recognition (ASR) and Spoken Language Understanding (SLU). The control module corresponds to the Dialogue Manager, which executes a dialogue strategy. The output module consists of a Natural Language Generation (NLG) system and a Text-To-Speech (TTS) engine. Usually, these modules are placed in a pipeline model (see Figure 2.2). The ASR converts the user's speech input (1) into text (2). SLU parses the text into a string of meaningful concepts, intentions, or Speech Acts (3). The Dialogue Manager maintains an internal state and decides what SA action to take next (4). This is what we call a dialogue strategy. For most applications the DM is also connected to a back-end database. In the output module, NLG renders the communicative acts (4) as text (5), and the TTS engine converts text to audio (6) for the user. Interested readers are referred to introductory texts such as (Bernsen et al, 1998; Huang et al, 2001; Jurafsky and Martin, 2000; McTear, 2004).

Human-Computer Interaction (HCI) is the study of interaction between people (users) and computers (such as dialogue systems). Human-machine dialogue differs from human-human dialogue in various ways. The most prominent features are the lack of deep language understanding and the lack of pragmatic competence (communicative skills) of the system. The lack of language understanding is due to errors introduced by less-than-perfect input processing (ASR and NLU), and the common use of shallow semantic representations. The lack of pragmatic competence is mainly due to the limited capabilities (often hand-coded heuristics) of the control module.

A substantial amount of recent work targets the problem of limited language understanding capabilities with so-called "error handling", e.g. (Bohus, 2007; Frampton, 2008; Skantze, 2007a), or "uncertainty handling" mechanisms, e.g. (Thomson and Young, 2010; Williams, 2006; Williams and Young, 2007a). This book addresses the problem of pragmatic competence: how to improve the communicative skills of a system by providing effective mechanisms to develop better dialogue strategies. In particular, this book explains how to automatically learn these skills from experience using simulated interactions, starting with a small experimental study of human behaviour. We now explain the conventional methods for strategy development.

2.2 Dialogue Strategy Development

There is a wide range of techniques to develop dialogue strategies, and techniques applied in industry are very different from the ones applied in research (Griol et al, 2010; Pieraccini and Huerta, 2005; Williams, 2008). These differences reflect the fact that academia and industry often pursue different objectives. Academic systems often aim to emulate human behaviour in order to generate 'natural' behaviour, whereas commercial systems are required to be robust interfaces in order to solve