Book Review

Spoken Dialogue Systems

Kristiina Jokinen and Michael McTear
(University of Helsinki, University of Ulster)

Princeton, NJ: Morgan & Claypool (Synthesis Lectures on Language Technologies, edited by Graeme Hirst, volume 5), 2009, xiv+151pp; paperback, ISBN 978-1-59829-599-3, $40.00; ebook, ISBN 978-1-59829-600-6, doi 10.2200/S00204ED1V01Y200910HLT005, $30.00 or by subscription

Reviewed by
Mary Ellen Foster
Heriot-Watt University

This book gives a short but comprehensive overview of the field of spoken dialogue systems, outlining the issues involved in building and evaluating this type of system and making liberal use of techniques and examples from a wide range of implemented systems. It provides an excellent review of the research, with particularly relevant discussions of error handling and system evaluation, and is suitable both as an introduction to this research area and as a survey of current state-of-the-art techniques.

The book is structured into seven chapters. Chapter 1 provides an introduction to the research area and briefly introduces the topics covered in the book. The end of the chapter consists of a list of links to tools and components that can be used for dialogue system development, which—although currently useful—seems likely to go out of date quickly.

Chapter 2 addresses the task of dialogue management, beginning by describing simple graph- and frame-based methods for dialogue control, and continuing with a discussion of VoiceXML. The chapter ends with an extended discussion of recent work in statistical approaches to dialogue control and modeling. It is unfortunate that the discussion of other methods such as the information state approach and plan-based models is postponed to Chapter 4, but otherwise this chapter provides a good summary of both classic and recent approaches.

Chapter 3 discusses error handling, which is divided into three processes: error detection, error prediction (i.e., the online prediction of errors based on dialogue features), and error recovery. After surveying a range of previous approaches to these three subtasks, the authors go on to discuss several more recent, data-driven approaches. Error handling is both a vital component of any spoken dialogue system designed for real-world use and an active area of current research, so this compact summary of techniques and issues is welcome.

Chapter 4 contains case studies illustrating a range of dialogue control strategies and models. It begins with a description of the information state approach, and then moves on to discuss plan-based approaches as exemplified in the TRAINS and TRIPS projects. This is followed by a discussion of two systems that employ software agents for dialogue management: the Queen's Communicator and the AthosMail system. Finally, two systems which make use of statistical models are presented: the Microsoft Bayesian receptionist, which models conversation as decision making under uncertainty, and the DIHANA system, which employs corpus-based dialogue management. The case studies in this chapter provide detailed examples of a range of techniques, along with

some discussion of the advantages and disadvantages of each approach, adding to the relevance of the book for developers of future dialogue systems.

Chapter 5 discusses four aspects of spoken dialogue systems that go beyond the straightforward information exchange scenarios considered in the preceding chapters. The first section covers aspects of collaborative planning along with Jokinen's (2009) Constructive Dialogue Modeling approach. The section on adaptation and user modeling provides a brief but useful survey of approaches to this task. The discussion of multimodality is longer and more concrete, but concentrates almost entirely on multimodal input processing; it would have been helpful to include a similar summary of the issues involved in multimodal output generation. The final section on "natural interaction" briefly discusses two topics: non-verbal communication for embodied conversational agents and multimodal corpus annotation.

Chapter 6—the longest in the book—gives a thorough treatment of the issues involved in evaluating spoken dialogue systems, beginning with an historical overview. It continues with a discussion of terminology and techniques and describes a wide range of subjective and objective evaluation measures that have been applied to the evaluation task. Next, two frameworks are presented that are designed to provide a general methodology for evaluation: PARADISE (Walker et al. 1997) and Quality of Experience (Möller 2005). This is followed by a discussion of concepts from HCI-style usability evaluations and how they can be applied to spoken dialogue systems, and then a section dealing with semiautomatic evaluation and the role of standardization. The final section discusses challenges that arise when evaluating advanced dialogue systems incorporating multimodality and adaptivity, and when evaluating systems designed for real-world applications.

Finally, Chapter 7 briefly discusses conversational dialogue systems (i.e., companion/chatbot systems), whose emphasis is on social communication rather than the information exchange tasks considered for most of the book. It also addresses the relationship between commercial and academic approaches to spoken dialogue.

It is interesting to note that both of the authors have also written books of their own that address aspects of spoken dialogue systems: McTear (2004) gives a comprehensive overview of the research area, including a series of hands-on tutorials on building systems using tools such as the CSLU toolkit and VoiceXML, and Jokinen (2009) provides a detailed description of a particular style of dialogue management, Constructive Dialogue Modeling.

This book fills a different niche: It has neither the exercises and tutorials of the McTear book, nor the in-depth description of a single formalism provided by Jokinen. Instead, it concentrates on outlining the issues involved in building a spoken dialogue system and on describing a wide range of existing techniques and systems. Although the book is short, it provides an excellent starting point for researchers new to the field, and every section has a good selection of references which would easily allow the interested reader to follow up any particular topic in more depth. The chapters on error handling and evaluation provide particularly useful summaries of the issues and techniques in these active research areas.

Although the preface says that the book is aimed at both engineering and humanities students, I suspect that readers without any mathematical background would have difficulty with some of the more formal sections. However, the necessary background knowledge is not great, and in general the book is clearly written and understandable; it is suitable as both an introduction to this research area and a survey of current state-of-the-art techniques.

References

Jokinen, K. 2009. *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. Wiley Series in Agent Technology. John Wiley & Sons, Ltd., Hoboken, NJ.

McTear, M. F. 2004. *Spoken Dialogue Technology: Toward the Conversational User Interface*. Springer-Verlag, Berlin.

Möller, S. 2005. *Quality of Telephone-Based Spoken Dialogue Systems*. Springer, New York.

Walker, M. A., D. J. Litman, C. A. Kamm, and A. Abella. 1997. PARADISE: a framework for evaluating spoken dialogue agents. In *Proceedings of EACL 1997*, pages 271–280, Madrid.

This book review was edited by Pierre Isabelle.

Mary Ellen Foster is a research fellow in the Interaction Lab of the School of Mathematical and Computer Sciences at Heriot-Watt University. Her research interests include embodied communication, natural language generation, and multimodal dialogue systems. Her address is School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom; e-mail: M.E.Foster@hw.ac.uk.