Book Review

Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots

Michael McTear
(Ulster University)

Reviewed by
Olga Seminck
Laboratoire Langues, Textes, Traitements Informatiques, Cognition - UMR 8094

This book has appeared in the series Synthesis Lectures on Human Language Technologies: monographs from 50 up to 150 pages about specific topics subjects in computational linguistics. The intended audience of the book are researchers and graduate students in NLP, AI, and related fields. I define myself as a computational linguist; my review is from a perspective of a "random" computational linguistics researcher wanting to learn more about this topic or looking for a good guide to teach a course on dialogue systems. I found the book very easy to read and interesting and therefore I believe that McTear fully achieved his purpose to write "a readable introduction to the various concepts, issues and technologies of Conversational AI." He succeeds remarkably well in staying on the right level of technical details, never losing the purpose of giving an overview, and the reader does not get lost in numerous details about specific algorithms. Additionally, for people who are experts in Conversational AI, the book could still be very useful because its bibliography is exceptionally complete: a very large number of early works and recent studies are cited and commented through the whole book.

The book is well structured into six chapters. After an introduction, there are two chapters about specific types of dialogue systems: rule-based systems (Chapter 2) and statistical systems (Chapter 3). This is followed by a chapter about evaluation methods (Chapter 4), after which the more recent neural end-to-end systems are reviewed (Chapter 5). The book ends with a chapter on various challenges and future directions for the research on Conversational AI (Chapter 6). I found that it was meaningful to distinguish the three types of dialogue systems: rule-based systems, statistical but modular systems, and end-to-end neural systems. It might, at first, seem strange that the topic on system evaluation methods is placed between the chapter about modular statistical dialogue systems and neural end-to-end systems, but as a reader, I believe that the discussion about system evaluation comes around at the right place in the book, because it helps to better understand the difference between modular and sequence to sequence systems. In this review, I will discuss the chapters one by one in the same order as they appear in the book.

The first chapter, the introduction, explains clearly what a dialogue system is and in what cases it can be introduced to perform tasks. It sketches the historical and present context of the domain and illustrates the different types of existing systems with many examples. The chapter clearly introduces the subject of the book, but as a linguist, I have to admit that I would like to have seen a linguistic description of how a human dialogue can be characterized.

The second chapter introduces rule-based systems. It provides a detailed and complete historical overview of the work in the field and shows well how the field has evolved. In particular, the diagram and the explanation about the dialogue system architecture were very helpful to understand such systems and, again, plenty of examples illustrate this chapter and make it easy to read. However, if the book were to be shortened—it is actually about 180 pages instead of the 50 to 150 that are usual in the series Synthesis Lectures on Human Language Technologies—I believe it should be in this chapter, giving slightly fewer details about historical dialogue systems.

From the second to the third chapter there is a very smooth transition: Thanks to the comprehensible introduction to the modular dialogue system architecture in Chapter 2, it is easy to understand how this framework can be adapted to become a statistical system. Moreover, the text explains clearly how reinforcement learning can be used for dialogue management and, again, everything is nicely illustrated with clear examples.

The fourth chapter discusses how Conversational AI can be evaluated and how training and evaluation data for systems can be collected. I particularly found the comparison between human evaluation (e.g., by using Amazon Mechanical Turk workers) and automated metrics very interesting. However, I would have liked to read a discussion about the ethical issues that can be at stake when collecting large amounts of human data on crowd-sourcing platforms. But, that marginal comment put aside, the chapter is very complete, and also provides concise descriptions of how all the subcomponents of dialogue systems can be evaluated.

The fifth chapter presents end-to-end neural dialogue systems. The reader can get a very good understanding about the difference between this type of system and a modular system (be it rule-based or data-driven). Moreover, I found the explanations about technical topics such as word embeddings and recurrent neural networks rather successful: They were easy to read and the technical mechanisms used in these architectures become clear. Throughout the book, but especially in this chapter, the advantages and the disadvantages of different types of system architectures are well explained. The up-to-date bibliographic references are impressive and will, in my opinion, be a good overview for more advanced readers as well. This is also true for the enumeration of available corpora for training and evaluation data.

The last chapter discusses a large number of challenges and future directions for the research on dialogue systems, for example: multi-modality, the problem of data sparseness, the handling of discourse phenomena, and ethical issues involved with Conversational AI. Although I find all the topics interesting, their large variety makes Chapter 6 very eclectic and one cannot shake the impression that it serves as a catch-all chapter for subjects that have not been addressed elsewhere in the book. I think that it should be possible to introduce a number of these discussions earlier in the book. For example, I believe that problems with handling discourse and dialogue phenomena, such as anaphora, could be addressed as the different types of systems are presented and maybe even discussed in Chapter 4 (about evaluation). The same would be true for ethical issues. For example, the discussion about how most bots having female voices could be seen as sexist (because the bot has an assisting function) could be introduced at the same time as speech generation in Chapter 2; gender-specific biases that result

from biased training data could be discussed after the introduction about the corpora used to train Conversational AI (in Chapter 5). In addition, there were two small ethics topics that I missed in this book. On the one hand, the protection of customer data and privacy issues: People provide personal data through dialogue with the system and some dialogue systems of the "talking speaker type" such as Alexa and Google Home are present in people's homes and may be exposed to sensitive information. On the other hand, the question of whether it is always ethical to refer people to a bot, instead of letting them speak to a real human. I think that if these discussions could be addressed throughout the book, Chapter 6 could just paint a clear vision of the future development of dialogue systems.

In conclusion, McTear's book provides a very clear overview of different types of dialogue systems, from the very beginning of the field to the most up-to-date research, and is very well illustrated with examples, which makes it an accessible reading for students and non-experts (provided that they have knowledge about AI or NLP). I highly recommend this book to people in search of a comprehensive overview on the topic.

Olga Seminck is a Research Engineer at the Centre National de la Recherche Scientifique in France. She works on various topics in computational linguistics going from anaphora resolution to commonsense reasoning and is specialized in the production, quantitative analysis, and evaluation of language resources. Her e-mail address is olga.seminck@cnrs.fr.