Participating in Explanatory Dialogues: Interpreting and Responding to Questions in Context

Johanna D. Moore
(University of Pittsburgh)

Reviewed by
Alison Cawsey
University of Glasgow

Moore's book, based on her doctoral thesis, presents her work on the automatic gener-
ation of natural language explanations and also gives an excellent summary of previ-
ous work in the field. The book is well written, and should be accessible with limited
prior knowledge of the field. I would recommend it both to those wanting a detailed
overview of Moore's work (partly summarized by Moore and Paris [1993]) and to
those wanting a solid introduction to recent work on explanation generation with an
emphasis on dialog issues. The book should be of interest to researchers both in com-
putational linguistics (particularly in text generation and discourse structure) and in
expert systems.

Moore focuses on textual explanations given by expert and advisory systems, and
she describes in detail the explanation component of an expert system that advises the
user on how to improve their Lisp program. However, it is clear how the basic ideas
and techniques presented apply to a wide range of applications in which explanations
need to be provided, such as help systems, online documentation systems, tutorial
systems, and so on.

Moore starts from the point of view that explanation should be an essentially
interactive process, requiring a dialog between the person or system giving the expla-
nation (the advisor) and the person receiving the explanation (the advisee). Without
such a dialog the chances of the advisee obtaining the information required, and in
a form they understand, are much reduced. In order to participate effectively in such
a dialog the advisor must be able to respond appropriately to *follow-up questions* after
the initial explanation is given. Moore argues that this requires that the advisor un-
derstand the *context* of these questions, and in particular the context created by the
advisor's previous responses. To do this, the advisor must be able to reason about its
own previous responses. This theme is developed throughout the book, through the
detailed discussion of a particular implementation of the ideas.

Moore starts off with two very useful and clearly written review chapters. Some
people might choose to study the book primarily for these reviews, which might in-
deed be suitable as a basis for postgraduate-level introductory courses on explanation
in expert systems. The first chapter argues that if an expert system is to provide effec-
tive explanations, then the system itself must be designed with that in mind, to ensure
that the knowledge required for possible explanations is explicitly represented. Prior

work in this area is presented, but the main emphasis is on the Explainable Expert Systems Framework (Swartout, Paris, and Moore 1991), which is presented clearly and convincingly.

The second review chapter argues that an explanation system for a sophisticated expert system must have knowledge of how to explain. This includes knowledge of how explanations are structured (discourse structure) and how particular explanation strategies are used to achieve communicative goals. A good overview is given of early explanation systems (and their limitations), and of previous approaches to generating multisentence texts. However, the main focus in this chapter is on how communicative goals can be linked to rhetorical or coherence relations, and how that provides the basis for a principled approach to the planning of coherent multisentence texts to address particular communicative goals. The main approach presented is Rhetorical Structure Theory (RST) (Mann and Thompson 1988).

The next two chapters are primarily concerned with a detailed description of Moore's own explanation system. This uses an RST-based text planner to construct multisentence explanations, given a particular communicative goal and assumptions about the user. The text plan makes explicit the goal behind the explanation (and the subgoals behind its different parts), the way the different parts of the explanation are related (by rhetorical relations), and any assumptions about the user that are made when planning the explanation. This information allows ambiguous or vague follow-up questions to be interpreted, and context-dependent responses to be given. Although the general approach is well argued and clearly presented, some of the particular details (e.g., of the text plans and representation of the user's cognitive state) occasionally appear ad hoc. Perhaps that is the price one must pay for presenting the ideas in sufficient detail and sufficiently clearly that the reader could develop his or her own implementation.

The remainder of the book presents more-recent, current, and suggested future work: the development of a direct manipulation interface to allow the user to provide feedback and follow-up questions; a discussion of how the success of a complex explanation could be monitored as it is presented; and current work on linking text planning to surface realization, making more use of the dialog history, and developing a better text planner.

The book as a whole has been substantially reworked from the original thesis, and largely brought up to date. It is clear and accessible in its style, and although a substantial proportion is concerned with the details of Moore's particular system, this is balanced by lengthy and clear review sections, strong general arguments, and plenty of discussion of current ideas and ongoing work.

One weakness of the book is the tendency to presume that human (verbal) explanations provide the appropriate foundation for computer-generated explanations, and that making explanations more adaptive to the user will, of necessity, make them more acceptable and effective. HCI researchers may be skeptical, particularly given the notable absence of any discussion of system evaluation, and limited elicitation of user requirements. This is a weakness shared with most of us in the field, and one that I hope will be addressed in the future through more user studies comparing adaptive and nonadaptive generation systems.

Moore's work, like much in this area, balances the development of general theoretical principles for explanation generation with the development of a particular applied system. As a result, the theoretically minded might quibble with details of the representations used in the system, while the practically minded might be concerned with assumptions implicit in the approach and the lack of evaluation. However, these minor problems are inevitable in a work that crosses the boundaries between theory

and practice and that has such wide potential relevance. I would therefore warmly recommend the book as one that is thought-provoking and well argued in the ideas and principles presented, and one that presents enough practical detail to allow readers to, if they wish, develop their own explanation systems and text planners, for diverse application areas.

References

Mann, William C., and Thompson, Sandra A. (1988). "Rhetorical structure theory: Towards a functional theory of text organization." *Text*, 8(3), 243–281.

Moore, Johanna D., and Paris, Cécile L. (1993). "Planning text for advisory dialogues: Combining intentional and rhetorical information." *Computational Linguistics*, 19(4), 651–695.

Swartout, William R.; Paris, Cécile L.; and Moore, Johanna D. (1991). "Design for explainable expert systems." *IEEE Expert*, 6(3), 58–64.

Alison Cawsey is a lecturer in artificial intelligence at the University of Glasgow. Her doctoral research in explanation generation was published as *Explanation and Interaction: The Computer Generation of Explanatory Dialogues* (The MIT Press, 1993). Her current research focuses on the generation of personalized explanations in the context of computer-based patient education systems. Cawsey's address is: Department of Computer Science, University of Glasgow, 17 Lilybank Gardens, Glasgow G12 8QQ, Scotland, UK; e-mail: alison@dcs.gla.ac.uk.