

Can Argumentation Help AI to Understand Explanation?

Doug Walton

Explanation-aware computing needs to supplement the older model that sees an explanation as a chain of inferences with a pragmatic and communicative model that structures an explanation as a dialog exchange. The field of argumentation, originally put forward in linguistics and philosophy, is now seen as providing a core approach that has been widely adopted in artificial intelligence, including multi-agent systems (Dunn and Bench-Capon, 2006). This paper presents an example that shows how the argumentation methodology works on the concept of explanation by transforming an example of an explanation into a formal dialog structure. However, the project of extending argumentation theory to the concept of explanation is still at its very early stages, and some key problems to be solved in future research are indicated.

The most developed analyses of explanation have been carried out in the philosophy of science, where they have concentrated on scientific explanation, and adopted a positivistic approach. Early work on explanation in the philosophy of science abstracted from context by seeing an explanation inferentially, much in the same way that early expert systems saw an explanation as chaining of inferences. This approach omitted, for the most part, to see the purpose of an explanation as increasing understanding, because (a) it was directed to the analysis of explanation in the natural sciences rather than to the study of explanation in everyday reasoning, and (b) because it saw the notion of understanding as not amenable to precise logical/scientific analysis. Recent work in cognitive science, on the other hand, has postulated that the aim of an explanation is to increase understanding, and argued that explanations fail when they do not increase the understanding of the phenomenon the purport to explain. Recent research in argumentation and artificial intelligence (Dunn and Bench-Capon, 2006) is now based on such dialog structures in which two parties reason together to work towards attaining a communicative goal. Thus it is natural to look to argumentation as a framework for studying the concept of explanation in a precise and analytical way suitable for artificial applications like multi-agent systems. This paper shows how the dialog model can be successful, even though it is an early stage, and some key problems need to be solved in future research. Some suggestions for future work that could extend the model to topics of interest in AI are made.

Background

A traditional model of explanation in the philosophy of science, still widely accepted it would appear, is the so-called deductive nomological model, or DN model, most perspicuously formulated by Hempel (1965). On this model, an explanation is essentially a deductive inference from premises containing a set of laws and antecedent facts to a conclusion that is the proposition to be explained. In his outline of the generally accepted views of scientific explanation in philosophy through the second half of the twentieth century, Salmon (1989) characterized the third decade as the

adding of an inductive-statistical component that widened the scope of the DN model, but not enough to deal with deepening difficulties. The problem is that DN model, while it can be applied to some simple examples of scientific explanation, does not apply to all explanations, and is of little or no use for modeling explanations that have a communicative purpose, of the kind especially important in artificial intelligence. In the age of the Internet, seeing an explanation as a deductive or inductive inference is limited. Now it is important to see an explanation as a type of communication exchange or dialogue among agents of a kind that can have various styles and standards of success in different disciplines.

A more pragmatic approach is to think of an explanation as a communicative exchange in which a question is asked about something said to be not understood and the purpose of the response of offering an explanation is to aid the questioner's understanding. This approach has proved to be very promising in artificial intelligence (Leake, 1992). It has also been advocated and developed by Schank and his colleagues in cognitive science (Schank, 1986; Schank, Kass and Riesbeck, 1994). Another push in this direction has been provided by work that has used a dialog models to study argumentation (Walton and Krabbe, 1995). Although argument is different from explanation, both concepts can be analyzed in the dialog format. As part of a study aimed at analyzing abductive reasoning as inference to the best explanation, a dialog model was postulated in (Walton, 2004) in which explanation is seen as a transfer of understanding from one party to another in a rule-governed question-reply dialog.

The biggest problem in modeling explanation as a communicative exchange that has the structure of such a question-reply dialog is whether the notion of understanding is clear enough to be a component in building a precise dialog model of explanation. However, it can be argued that his problem may be solved by looking to work on scripts in AI, described by Schank, Kass and Riesbeck (1994, p. 77) as "frozen inference chains stored in memory". Scripts represent common knowledge about common situations and routine ways of doing things that agents share. In the usual example, called the restaurant script (Schank, Kass and Riesbeck, 1994, p. 7), a person can be taken to know when he or she