Justification of Automated Decision-Making: Medical Explanation or Medical Argument?

Ravi D. Shankar, MS, and Mark A. Musen, MD, PhD
Stanford Medical Informatics, Stanford University School of Medicine,
Stanford, CA 94305-5479

People use arguments to justify their causes. Computer systems use explanations to justify their claims. The WOZ explanation framework adopts an argumentation methodology for generating explanations for a clinical decision-support system. The central component of the framework is the explanation strategy that decides what pieces of information should be presented to justify a claim. The strategy uses Toulmin's argument structure to define the pieces of information and to orchestrate the presentation of such information. WOZ uses explicit models that abstract the core aspects of the framework such as the explanation strategy. We present the use of arguments, the modeling of explanations and the explanation process in WOZ. WOZ has made use of the wealth of naturally occurring arguments, and our arguments appear to result in convincing explanations.

1 WE PROPOSE TO ARGUE

In human conversation, arguments are used to support beliefs and claims by providing evidences. This is obvious when lawyers argue their clients' cases, when writers argue their beliefs, and when physicians argue their diagnoses. The whole notion of explanation in a knowledge-based system can be thought of as a way of argument between the user and the system. Explanations essentially provide justifications for the conclusions of the knowledge-based system. The system is making a claim, and when questioned by the user, it presents different levels of information to support the claim. This kind of argumentative dialog can play a crucial role in generating persuasive explanations.

We are developing a multi-agent framework called WOZ [1] that provides explanations using arguments. WOZ explains the claims of EON [2], a knowledge-based system architecture that provides physicians with decision-support in protocol-based care. EON consists of a set of problem-solving components that use explicit models of the medical domain knowledge and the clinical protocol knowledge. WOZ, like EON, uses explicit models that, for the purposes of explanation, abstract the explanation strategy and the agent architecture. The explanation strategy defines what information should be used to justify a claim. The strategy model uses an argument structure proposed by Toulmin in his theory of reasoning [3]. Toulmin's structure allows the recipient of the argument to identify the different elements needed to support a claim. The focus of this paper is our explanation modeling based on argumentation.

We briefly describe the underlying knowledge-based system, EON, and the explanation framework, WOZ. Next, we discuss the general role of arguments and how they relate to explanations. We introduce Toulmin's argument structure and its use in previous explanation frameworks. We then describe our explanation models and their use in the explanation process. Finally, we discuss the merits of modeling explanations as arguments, and show how the models might be extended to make explanations more persuasive.

2 THE GROUNDS – EON AND WOZ

A clinical protocol such as the HIV/AIDS protocol CBCT-P001 has a set of eligibility criteria that clinicians use to decide whether a patient should be treated in accordance with that protocol. One of the decision-support systems that we have built for use within the EON architecture determines a patient's eligibility for a given protocol [4]. The system contains a set of general components that work together to automate the eligibility-determination task (see Figure 1, shaded area). The database-mediator component performs temporal database

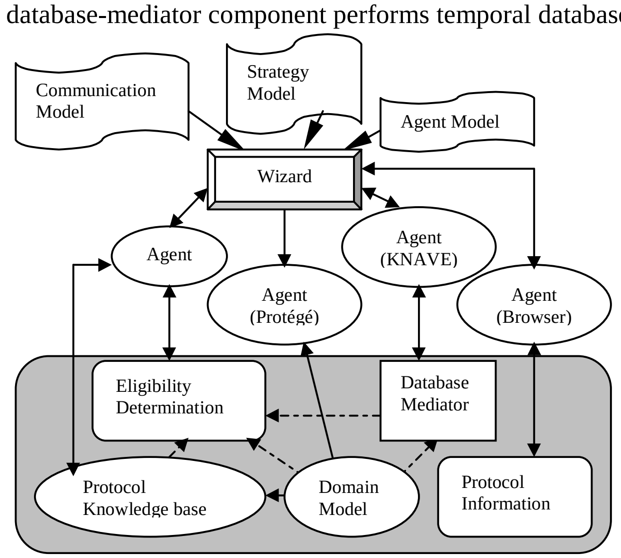

Figure 1: EON and WOZ. The EON components are shown inside the gray box. These components interact with each other to provide decision support. WOZ consists of a set of visualization agents, one agent per component. Each agent has a set of graphical user interfaces. The Wizard controls the agent–agent interactions. It takes in as input the user model, the agent model, and the explanation strategy. The Wizard parses an explanatory query, consults the explanation strategy to decide what information should be presented as an explanation, and directs the appropriate agents to present that information.

management and data abstractions on patient data. The domain-model component provides the domain-specific terms and relations used by the system. The protocol knowledge base component provides structured protocol knowledge used by the eligibility-determination component. The protocol-information component displays for the end user texts of protocol documents.

WOZ employs a collection of cooperating visualization agents (see Figure 1). The visualization agents are closely associated with the components of EON. The agents handle the higher-level aspects of explanation, such as presentation of information and interaction with the end user. The components provide the lower-level data that drive the explanation, such as the raw patient data and the reasoning knowledge. An explanation engine, called the Wizard, uses the explanation strategy model to decide which agents should present what information in response to a user query. The visualization agents, the Wizard, the explanation strategy, the agent interactions, and the components, viewed globally, constitute the explanation of the whole system.

3 HOW WE ARGUE TO EXPLAIN

The core aspect of our explanation framework is the explanation strategy that defines what constitutes an explanation for a claim. It identifies the different components of an explanation such as the claim itself, the medical evidences that support the claim, the strength of the claim, the evidences that do not support the claim, and the patient data that were used in making the claim. This essence of an explanation strategy mirrors the notion of an argument put forth when supporting a claim.

3.1 About Arguments

An addresser states a claim, and then uses arguments to increase the belief of the addressee in the claim. These arguments include presenting evidences related to the claim, generally using a structured format. Arguments are widely used by lawyers and writers. A lawyer first presents a claim in her opening statement at a trial. Then, to prove the claim, she describes the physical evidences that lead to the claim. She then highlights the laws that apply to the case, and presents experts to support her claim. She may back up these justifications with reference to historical data or cases. Writers use similar structured arguments when stating their points of view. Physicians have used arguments to present their clinical diagnoses, and to support them with patient data, medical knowledge, and other related cases. Given this role of arguments in human–human interactions, we can see how arguments can be synonymous with explanations in human–system interactions.

3.2 Toulmin's Argument Structure and Explanations

We have modeled the explanation strategy as structured arguments. We have used Toulmin's argument structure

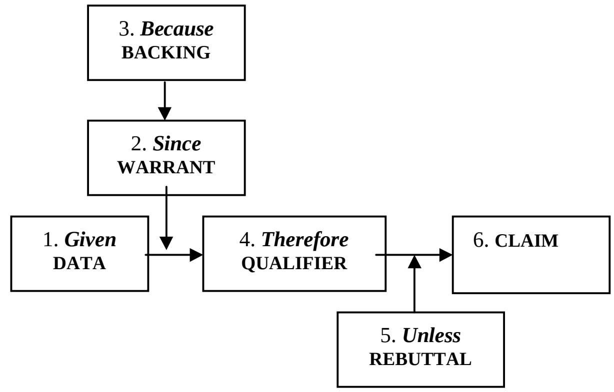

Figure 2: Toulmin's Argument Structure. The structure reads given Data, therefore Claim, since Warrant, because Backing, unless Rebuttal. The elements of the structure and the relationship among them can be used to generate explanations.

to identify, organize and present information related to explanation strategy. Toulmin, a philosopher, has developed a pragmatic method of reasoning [3]. Central to his logic is his six-element argument structure (see Figure 2) by which claims can be argued, regardless of the context that surrounds the argument:

1. ***Claim*** – the assertion or conclusion put forward for general acceptance.
2. ***Data*** – the particular facts about a situation on which a claim is made.
3. ***Warrant*** – the knowledge that justifies the leap from data to a claim.
4. ***Backing*** – the general body of information or experience that validate the warrant.
5. ***Qualifier*** – the phrase that shows the confidence with which the claim is supported to be true.
6. ***Rebuttal*** – the anomaly or exception for which the claim would not be true.

Wick [5] points out how early research in explanation has, without stated intent, evolved to eventually engulf the whole of Toulmin's argument structure. Ye [6] uses Toulmin's argument structure in an experimental investigation of the value of explanation in expert systems for auditing. Ramberg [7] describes a multiple-explanation construction model that constructs explanations for an expert system in the domain of protein purification.

3.3 Our Explanation Models

The explanation space of EON comprises the patient data (e.g., laboratory parameters), the medical domain knowledge (e.g., AIDS domain concepts), the clinical protocol knowledge (e.g., eligibility criteria for protocol CBCT-P001), the reasoning knowledge (information on the reasoning process), and other relevant external information (e.g., clinical protocol text). The explanation process in the WOZ framework includes (1) identifying the distinct elements of the explanation space that are required to satisfy user's explanatory query, (2) obtaining

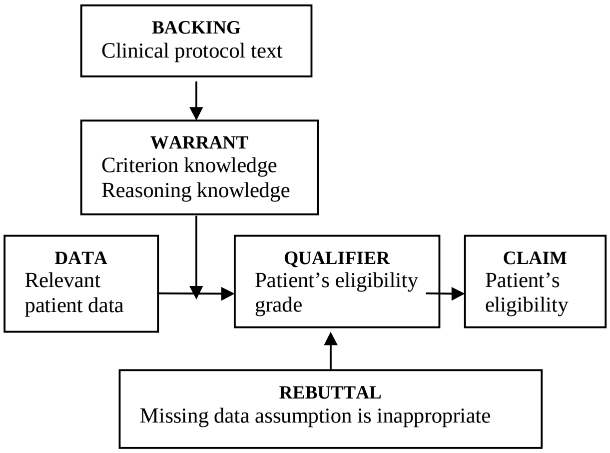

Figure 3. A Meta-argument for a class of claims. The structure illustrated here is for the class *patient's eligibility*. The meta-rgument is specified explicitly and stored in the explanation strategy knowledgebase. Note that, in this example, all the elements of the Toulmin structure have been specified. Typically, the evidences available to support a claim drives what elements get filled.

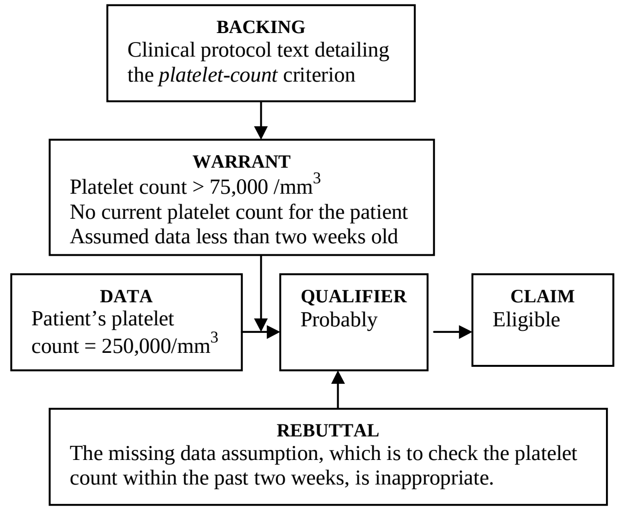

Figure 4. A Concrete argument for a specific claim. The structure illustrated here is for the claim *patient's eligibility for platelet-count criterion*. A justification for the claim can be made using the elements of this concrete argument structure.

the required information from the appropriate agents, and (3) presenting the explanation in a coherent manner. To aid in this explanation process, we have designed three declarative models, the Strategy Model, the Communication Model and the Agent Model. These explanation models abstract the core elements of the explanation process: the explanation strategy, the agent services, and the terminology used in the strategy and by the agents. We built the explanation models using Protégé, a software system that is used to create knowledge models. We also generated the corresponding knowledge-acquisition tools using Protégé.

3.3.1 The Strategy Model

We modeled the explanation strategy as *arguments*, using Toulmin's argument structure. The elements of the argument structure for a claim identify the information needed for explanation of that claim. Appropriate WOZ explanation agents provide the needed information, and the relationships among the elements of the argument structure can be used to present the explanation consistently and clearly. We define two types of arguments in our explanation strategy, (1) the meta-argument and (2) the concrete argument.

A meta-argument conceptualizes arguments for a *class of claims*. We state explicitly the elements of the meta-argument structure with abstract descriptions of appropriate pieces of information in EON's explanation space. We illustrate a meta-argument for explaining patient eligibility score in Figure 3. The meta-argument simply reads: the criterion knowledge and the reasoning knowledge support the leap from patient data to the graded eligibility score; the clinical protocol text provides the basis for the criterion knowledge and the reasoning knowledge; all this support for the claim is true unless the missing data assumption, if any, is inappropriate.

A concrete argument defines an argument for a *specific claim* in a class of claims. It follows that a concrete argument is an instance of a meta-argument. We specified a meta-argument for explaining patient eligibility score (see Figure 3). From this meta-argument, we can derive a concrete argument for explaining patient's eligibility score for a particular criterion such as the *platelet-count criterion* (see Figure 4). The abstract descriptions of the pieces of information in the meta-argument are substituted with the actual information related to the computation of this criterion's eligibility score. With this concrete argument, an explanation to support the criterion's eligibility score can be generated.

Meta-arguments and concrete arguments have a class–instance relationship. The explanation strategy knowledge essentially consists a set of meta-arguments, one for each class of claims the system makes. Since the explanation strategy is modeled explicitly, we can acquire the meta-arguments using a knowledge acquisition tool. The concrete arguments are derived from these meta-arguments at runtime during the explanation process.

3.3.2 The Communication Model

There are concepts and terms that are used when developers construct the meta-arguments and when WOZ agents communicate with one another (see Figure 1). The communication model defines the following concepts:

- Action verbs that are mainly used in agent–agent interaction (*explain* and *show*).

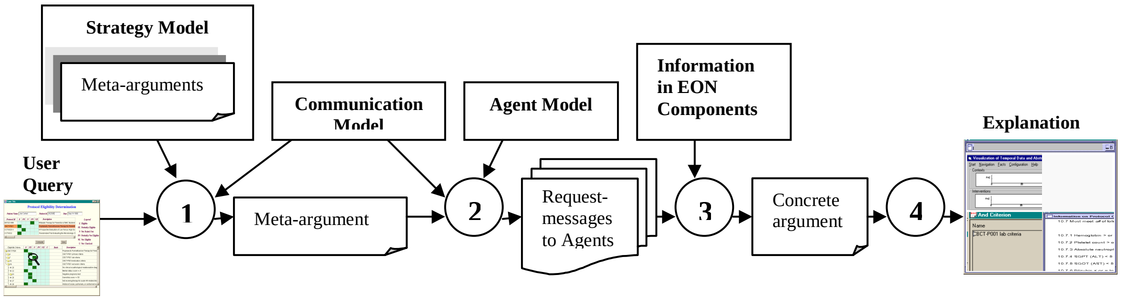

Figure 5. The WOZ Explanation Process. The user submits query and recieves explanation using GUIs. The numbered circles refer to the following action points: 1) Wizard selects appropriate meta-argument; 2) Wizard asks appropriate agents to provide information specified in the meta-argument; 3) WOZ Agents provide information; 4) Wizard generates explanation..

- Concepts related to the problem domain (*protocol identifier*, *criterion identifier*, and *eligibility score* in the patient eligibility determination domain).
- Concepts and terms related to the particular area of medicine such as breast cancer or HIV/AIDS to which the knowledge-based system is applied (*clinical findings* and *medical interventions*).

3.3.3 The Agent Model

The visualization agent in the WOZ architecture provides information in the EON component that it is representing (see Figure 1). The agent presents this information using visual media such as text, graphics and video. The agent model encapsulates the characteristics of the agents such as the agent's identifier, and the services the agent provides. The definition of a service includes the required inputs, the type of information it provides and the presentation medium. An example instantiation of the Agent Model is information on an agent identified as *KNAVE* (see Figure 1) that presents *patient data* graphically taking *patient identifier* as input.

3.4 A Dialog

To explain a specific claim, at runtime, WOZ selects the appropriate meta-argument in the strategy knowledge base, identifies the explanation information, obtains the information from appropriate agents, derives the corresponding concrete argument, and generates the explanation by organizing the presentation of the agents (see Figure 5). We demonstrate this explanation process with the help of a visual dialog between WOZ and a user of EON. The user submits queries and recieves explanations using direct manipulation via graphical user interfaces. This dialog is based on the examples we used previously to illustrate the meta-argument and the concrete argument.

The system has determined the patient's elgibility for protocol CBCT-P001, and displays the eligibilty scores of the protocol overall and of the protocol's individual criteria. Now the user is interested in the details of a specific criterion of the protocol.

User: Submits a request to display the patient's data used to compute the eligibility score for the *platelet-count* criterion.

The Wizard *Agent (see Figure 1) recieves the details of the request. It recognizes that the eligibility score of the* platelet-count *criterion belongs to the class of eligibility scores. It selects the meta-argument of the eligibility scores (see Figure 3) from the strategy knowledge base. It then consults the agent knowledge base and requests the appropriate agents to provide the information pieces identified in the meta-argument. These agents fill in the realtime values for the* platelet-count *criterion to get the concrete argument for the criterion (see Figure 4). The* data *element is filled by the* KNAVE *agent and the Wizard then uses that element to answer the user.*

WOZ: Displays the KNAVE agent's presentation that the patient had a platelet count = 250,000/mm^3.

User: Requests to explain the eligibility score for the *platelet-count* criterion.

The warrant *element of the concrete argument for the* platelet-count *criterion is filled by the* Protégé *Agent. WOZ taps into the* warrant *element to answer the user.*

WOZ: Displays Protégé Agent's presentations that the eligibility criterion is "*platelet count should be > 75,000/mm^3*", and the missing data assumption is "*assume that the platelet count is valid if less than two weeks old, and that the score is* probably eligible *if within the range*."

User: Requests more information on the eligibility criterion.

The backing *element of the concrete argument for the* platelet-count *criterion is filled by the* Browser *Agent. WOZ directs the user to the information in the* backing *element.*

WOZ: Displays Browser Agent's presentation of the protocol text that describes the *platelet-count* criterion. This presentation may also indicate the intentions behind the *platelet-count* criterion.

This visual dialog demonstrates how the concrete argument structure aids WOZ in deciding what information to present when.

4 DISCUSSION

We have demonstrated how claims made by knowledge-based systems such as EON can be justified using arguments. We used Toulmin's argument structure to integrate and present explanation information from varied sources namely the EON components. Acquisition of explanation strategies for particular domains includes identifying the classes of claims the system need to make and defining one meta-argument for each class. We believe that the number of classes is within reasonable limits for a component-based system such as EON. For example, the EON patient eligibility determination system requires only one class of claim (i.e., the patient's eligibility score).

The dialog we discussed earlier could very well have been between a physician and a patient. Horton [8] and Dickinson [9] propose that physicians should use Toulmin's argument structure to organize medical evidences supporting the physicians' diagnoses or treatment plans. These proposals further strengthen the appropriateness of using an argumentative approach to provide medical explanations.

There are explicit relationships among the elements of Toulmin's argument structure. We can use this relationship when presenting explanations. There can be other kinds of relationships in the argument structure, however, that may have to be defined and considered in the presentation. For example, when there is more than one item in any element of the argument, say the *Warrant*, in what order do we present those items? How do we expose the relationships among these items? One method can be to employ the Rhetorical Structure Theory (RST) [10] that has been developed mainly for text analysis and text generation. RST maintains that in most coherent discourse, consecutive discourse elements are related by a small set of rhetorical relations that are defined by the theory. Some of the relations that are defined are *Condition*, *Elaboration* and *Sequence*. A large number of natural language generation systems rely on the rhetorical relations defined in RST. We can use RST by describing various rhetorical relations among the different items of the warrant. By providing such connectives, we can strengthen the cohesiveness of the multiple but related items, thereby enhancing the clarity in the presentation.

There is always a question of how to tailor explanations to the user to enhance the user's acceptance of the claims. One approach is to extend the argument structure by defining multiple subarguments. We can envision this structure as multiple argument structures having the same *Claim*, *Data*, and *Modifier* but possibly different *Warrant*, *Backing* and *Rebuttal*. A superimposition of these structures will result in an argument structure with multiple subarguments. For the same *claim* and *data*, we can then create subarguments each providing a different flavor of the same argument. When providing explanation to a user, a suitable subargument can be employed resulting in tailored explanations. Naturally, this design presupposes the existence of a User Model that abstracts the user profile and preferences.

5 CONCLUSION

We showed how medical explanations could be expressed as medical arguments. Our explanation strategy uses a widely recognized argumentation structure, and can mirror naturally occurring arguments. Thus, our approach can generate convincing medical arguments (i.e., medical explanations).

Acknowledgements

This work has been supported, in part, by grant LM05708 from the National Library of Medicine.

References

1. Shankar RD, Tu SW, Musen MA. A declarative explanation framework that uses a collection of visualization agents, *Journal of the American Medical Informatics Association*, 1998; symposium supplement, 602–606.
2. Musen MA, Tu SW, Das AK, Shahar Y. EON: A component-based approach to automation of protocol-directed therapy, *Journal of the American Medical Informatics Association*, 1996; 3(6), 367–388.
3. Toulmin S. The uses of argument, *Cambridge University Press*, Cambridge MA, 1958.
4. Tu SW, Kemper CA, Lane NM, Carlson RW, Musen MA. A methodology for determining patients' eligibility for clinical trials, *Methods of Information in Medicine*, 1993; 32(4), 317–325.
5. Wick MR. Expert system explanation in retrospect: a case study in the evolution of expert system explanation, *Journal of Systems and Software*, 1992; 19(2), 159–169.
6. Ye LR. The value of explanation in expert systems for auditing: An experimental investigation, *Expert Systems with Applications*, 1995; 6(4), 543–556.
7. Ramberg R. Construing and testing explanations in a complex domain, *Computers in Human Behavior*, 1996; 12(1), 29–48.
8. Horton R. The grammar of interpretive medicine, *Canadian Medical Association Journal*, 1998;158,245–249.
9. Dickinson, HD. Evidence-based decision-making: an argumentative approach, *International Journal of Medical Informatics*, 1998; 51, 71–81.
10. Mann WC, Thompson S. Rhetorical structure theory: A theory of text organizations, *Information Sciences Institute Technical Report Number RS-87-190*, University of Southern California, Marina del Rey, CA, 1987.