AI-Based Argumentation in Participatory Medicine

Nancy L. Green
Department of Computer Science
University of North Carolina Greensboro
Greensboro, NC 27402 USA
nlgreen@uncg.edu

Abstract

This paper discusses how AI models of argumentation can play a role in personalized and participatory medicine. It describes our previous research on natural language generation of argumentation for genetic counseling and a pilot study on risk visualization, and our current research on argumentation mining.

Introduction

Among argumentation theorists, there is growing interest in the role of argumentation in healthcare, e.g. a special issue of the journal *Argumentation in Context* (Rubinelli and Henkemans 2012). Artificial intelligence researchers also have been interested in applications of argumentation theory for the healthcare consumer (e.g. Bickmore and Green, 2006). AI-based argumentation can play a role in a variety of use cases of *participatory medicine* (PM). The use case that has received the most attention to date involves *persuasive* argumentation for health promotion, e.g., an intelligent dialogue system that attempts to persuade the user to adopt a healthier diet or to exercise (Grasso et al. 2000, De Boni et al. 2006, Mazotta et al. 2007), or a system using cartoon-like animation and risk visualization to persuade young adults to reduce alcohol consumption (Bisset et al. 2013). A precursor of those interactive systems was an intelligent system to generate tailored letters to persuade recipients to stop smoking (Reiter et al. 2003).

A second use case involves making the reasoning of clinicians *transparent* to healthcare clients by presenting arguments for the experts' conclusions. We assert that "transparency is necessary before a [lay] audience can fully comprehend, evaluate or challenge an argument, or re-evaluate it in light of new findings about the patient or changes in scientific knowledge" (Green et al. 2011, p. 23). Lastly, a potential new role for argumentation in PM involves *argumentation mining*, the automatic detection of argumentation structures, in the biomedical literature (Green 2014a). Although methods for argumentation mining this literature have not yet been developed and constitute a significant research challenge, those methods will enable use cases for finding and summarizing personally relevant healthcare information.

The next section describes our previous research on generation of transparent argumentation for genetic counseling. Next the paper discusses a pilot study on visual communication of risk in that domain. The last section discusses the potential of argumentation mining.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Argument Generation

To investigate methods for transparent argument generation, we implemented the GenIE Assistant, a prototype system in the domain of genetic counseling (Green et al. 2011). The GenIE Assistant generates the first draft of a genetic counseling patient letter, including arguments for the diagnosis of the patient's genetic condition and for the inheritance of the condition. The arguments are the same types of arguments that we identified in a sample of patient letters written by genetic counselors. It was assumed that a healthcare provider or electronic healthcare record would supply patient-specific data (symptoms, test results, etc.) and the clinicians' conclusions (e.g. the diagnosis and parents' genotypes). The GenIE Assistant then would synthesize a letter and a genetic counselor or other healthcare provider would review and edit the draft before sending it to the client.

The Argument Generator of the GenIE Assistant makes use of a causal knowledge base (KB) describing various single-gene autosomal genetic conditions (including their modes of inheritance and pathways from mutations to observable symptoms and test results) and a set of domain-independent, abstract argumentation operators that are defined in terms of qualitative probabilistic properties of the KB (Druzdzel and Henrion 1993). Given a specific claim, the Argument Generator creates atomic and/or compound arguments for it by instantiating the argumentation operators with information from the KB and inputs about the patient. Then a natural language generation component expresses the arguments in English.

For example, the Assistant reconstructed the clinician's argument that the source of a patient's achondroplasia (a common cause of human dwarfism) was likely to be a germline mutation carried by his father: In

summary, the patient was diagnosed as having achondroplasia due to a single G1138A mutation of the *FGFR3* gene, which must have been inherited from his mother or father. Neither parent has symptoms of achondroplasia, so neither is likely to have the mutation in every cell. However, the father is 45 years old and being over 35 years old increases the risk of having a new change in a reproductive cell, called a germline mutation. Therefore, the father could have a germline G1138A mutation that was inherited by the patient and is the cause of his condition.

In current practice, patient letters are written by genetic counselors for documentation purposes, e.g. for future care providers, and for informational, educational, and counseling purposes for clients. However, as the role of genetics in medicine increases, the potential benefit of at least partially automating this writing process is obvious. Moreover, as the PM paradigm shifts more responsibility to healthcare consumers, it is vital that they have access to the argumentation supporting the diagnosis, prognosis, and/or treatment options presented to them. Ideally, patients should be able to engage, at any time, with a system with the capacity to present and explain all points of view about their case. Ideally, the system could respond to questions and challenges from the user. Also, as new biomedical knowledge becomes available that strengthens the original arguments or that conflicts with their original assumptions, such a system could notify the patient of these changes.

Risk Visualization

Understanding genetic risk is often necessary for making an informed decision, whether to modify one's behavior to reduce risk (e.g. to lose weight to reduce risk of colon cancer) or to select among treatment options (e.g. after a diagnosis that one carries a *BRCA1* mutation). New methods of personal genomic testing can provide pre-symptomatic, probabilistic risk assessment for many common multifactorial disorders. However, the challenges of effective health risk communication are well known (Ancker and Kaufman 2007). Health communication researchers have studied various visualization techniques to improve risk communication (Ancker et al. 2006, Lipkus 2007). Ideally, an AI-based system would be able to use that type of knowledge to tailor risk graphics to a model of the user and to provide a real-time explanation of any part of the visualization that the viewer does not understand (Green 2011).

As a step towards integrating AI-based argumentation and risk visualization, we performed a pilot study to evaluate the comparative effectiveness of three graphics for depicting the risk of colon cancer of someone who has the mutation for hereditary nonpolyposis colon cancer (HNPCC). Based upon a survey of the literature, our hypothesis was that the first graphic, a line graph showing cumulative risk for HNPCC versus non-HNPCC individuals (Figure 1) would be less preferred and less effective than a display of natural frequencies (Gigerenzer 2002) (Figure 2) or a dot matrix (Figure 3).

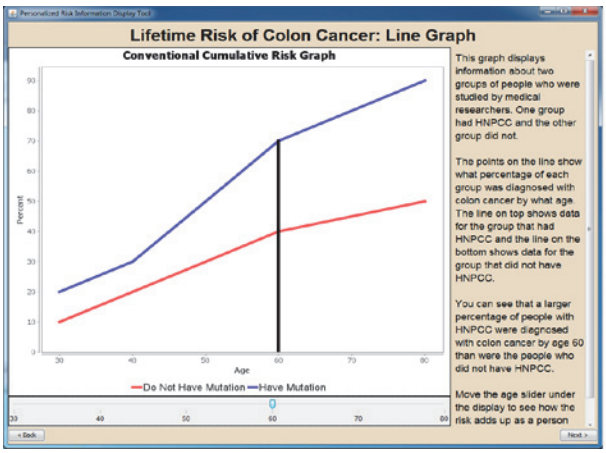

Figure 1. Line graph display.

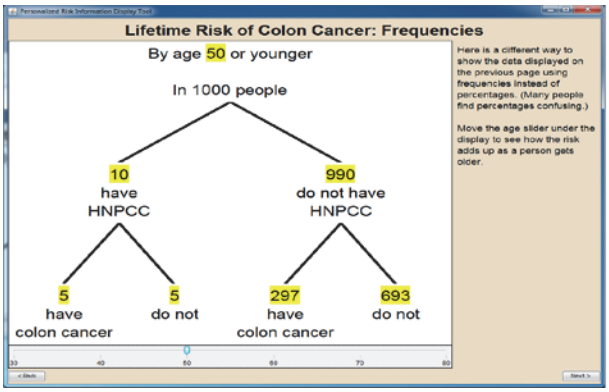

Figure 2. Natural frequencies display.

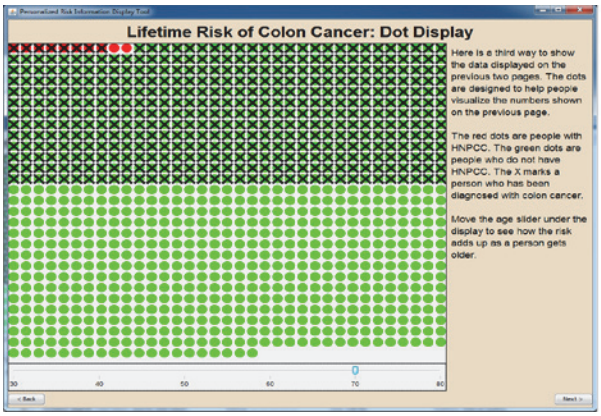

Figure 3. Dot matrix display.

The three graphics were implemented in Java, each on a separate screen. A slider at the bottom of each screen enabled the user to select the age of the patient for which the risk information was shown and to see how the risk changes over time. Eleven university student volunteers from an introductory course on use of computer application software were invited to look at the applets and complete a short survey. Contrary to our expectations, the majority of participants said that they found all three types of graphics clear and understandable, and equal numbers said that they liked the line graph (5/11) and the natural frequency graphic (5/11) best. A majority (8/11) liked the dot matrix the least. However, for both the line graph and the natural frequencies graphic, only 6/11 correctly answered the fundamental question, *Are more people with HNPCC diagnosed with colon cancer by age 70 than people who do not have HNPCC?* Only 4/11 correctly answered that question using the dot matrix. This raises the question of how to support argument transparency when the recipient may not even comprehend the supporting data, while believing that he does. (However, as the referees of this paper have pointed out, the intended meaning of this question may have been unclear. The intended meaning was: *By age 70, is a person with the HNPCC mutation more likely to be diagnosed with colon cancer than a person without the HNPCC mutation?*)

Argumentation Mining

Argumentation mining is the process of detecting argumentation structure in one or more documents. Argumentation structure consists of each atomic argument (premises and conclusion), the type of each argument, and the relationships among the atomic arguments, e.g., when one argument attacks another argument, where multiple independent arguments support the same conclusion, or where one argument supports a premise of another argument. A relatively new area of natural language processing research, it has explored techniques for mining existing corpora of non-science content, e.g. legal documents and product reviews. Our goal is to create a freely available corpus of open-access, full-text scientific articles from the biomedical genetics research literature, annotated with argumentation structures to support this type of research (Green 2014a).

As a step towards developing an annotated corpus, we have analyzed the argumentation in a representative article (Schrauwen et al. 2012) whose main claim is that a c.637+1G>T mutation of the *CABP2* gene is a cause of autosomal recessive non-syndromic hearing loss (arNSHL) (Green 2014b). The argumentation structure of this article can be summarized as: (A1) Justification for looking at a particular region of chromosome 11 for a candidate cause of arNSHL, (A2) after identifying the variants in that region of an individual affected with arNSHL, all but the c.637+1G>T variant of *CABP2* could be eliminated as candidates, (A3) the c.637+1G>T variant segregates with arNSHL in that individual's family but not in a control group, (A4) the c.637+1G>T variant segregates with arNSHL in two other families, perhaps distantly related to each other and the first family, (A5) there is a partial biochemical explanation for how the variant could lead to hearing loss, but (A6) the variant was not found in some other individuals with hearing loss. Note that A1 provides partial support for A2, A2 provides partial support for A3, and A3-A5 provide support for the argument that the variant is one of the causes of arNSHL, while A6 is not an argument against that claim.

Since many acceptable arguments used in science (as well as in law and everyday reasoning) are not necessarily deductively valid, it is important to test the *critical questions* of such arguments (Walton et al. 2008). For example, a critical question to ask of A4 is: *Is there some other factor present in the affected individuals and absent in the unaffected individuals that could account for the condition?* Another critical question of A4 that is partially addressed in A5 is: *Is there a plausible causal mechanism explaining how the c.637+1G>T variant could cause the condition?* When applying biomedical studies to particular patients, an important critical question is whether the population that was studied is representative of this patient, i.e., whether or not the study is applicable to her.

To illustrate a few of the techniques that have been investigated so far for argumentation mining, Mochales and Moens (2011) applied statistical techniques to legal documents to classify whether a sentence is part of an argument, to determine the boundaries of arguments, and to classify the premises and conclusion of arguments. Cabrio and Villata (2012) used textual entailment to detect support and attack relations between arguments in a corpus of on-line debates. A number of other researchers have investigated automatic classification of sentences in scientific documents to identify novel knowledge claims and components of a scientific investigation (e.g. Teufel 2010, Liakata et al. 2012). While previous researchers have attempted to extract useful information without the need to represent or reason about the domain, we plan to supplement those techniques with use of domain knowledge about genetics.

In the future, argumentation mining technology could provide healthcare consumers with an argumentation structure for the biomedical knowledge on which the clinician's conclusions about their case (diagnosis, prognosis, treatment options, etc.) is based. Exposing the structure of the argumentation enables the patient to evaluate the acceptability of any of the premises in a chain of arguments and to test the critical questions of the arguments.

Acknowledgments

Graduate student Caleb Brough implemented the risk visualization program, facilitated the user study, and tabulated the data. Dr. Malcolm Schug of the UNCG Department of Biology verified our interpretation of arguments in the genetics research literature.

References

Ancker, J.S. and Kaufman, D. 2007. Rethinking Health Numeracy: A Multidisciplinary Literature Review. *Journal of the American Medical Informatics Association* 14(6):713-721.

Ancker, J.S., Senathirajah, Y., Kukafka, R., and Starren, J. 2006. Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association* 13:608-618.

Bickmore, T. and Green, N. 2006. (Eds.) *Proceedings of AAAI 2006 Spring Symposium: Argumentation for Consumers of Healthcare.*

Bisset, S., Wood, S., Cox, R., Scott, D., and Cassell, J. 2013. Calculating alcohol risk in a visualization tool for promoting healthy behavior. *Patient Education and Counseling* 92(2), pp. 167-173.

Cabrio, E. and Villata, S. 2012. Generating Abstract Arguments: A Natural Language Approach. In Verheij, B., Szeider, S., and Woltran, S. (eds.) *Computational Models of Argument: Proceedings of COMMA 2012*. Amsterdam, IOS Press, 454-461.

De Boni, M., Hurling, R., Dryden, W. 2006. Argumentation through an automated rational-emotive behavior therapy system for change in exercise behavior. In *Proceedings of AAAI 2006 Spring Symposium: Argumentation for Consumers of Healthcare*, 34-8.

Druzdzel, M.J. and Henrion, M. 1993. Efficient reasoning in qualitative probabilistic networks. In *Proc. of the 11th Nat. Conf. on AI (AAAI-93)*, pp. 548-53.

Gigerenzer, G. 2002. *Reckoning with Risk: Learning to Live with Uncertainty.* London: Penguin Books.

Grasso, F., Cawsey, A., Jones, R. 2000. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies* 53, 1077-115.

Green, N.L 2011. Artificial Intelligence and Risk Communication. In *Proceedings of AAAI 2011 Spring Symposium on Artificial Intelligence and Health Communication.*

Green, N.L. 2014a. Towards creation of a corpus for argumentation mining the biomedical genetics research literature. In *Proceedings of the First Workshop on Argumentation Mining*, Baltimore, MD, June 26, 2014.

Green, N.L. 2014b. Argumentation for Scientific Claims in a Biomedical Research Article. To appear in *CEUR Proceedings of Frontiers and Connections between Argumentation Theory and Natural Language Processing*, Bertinoro, Italy, July 21-25, 2014.

Green, N., Dwight, R., Navoraphan, K., and Stadler, B. 2011. Natural language generation of biomedical argumentation for lay audiences. *Argument and Computation* 2(1), 23-50.

Liakata, M. et al. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics* 28(7).

Lipkus, I.M. 2007. Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations. *Medical Decision Making* 27: 696-713.

Mazotta, I., de Rosis, F., and Carofiglio, V.P. 2007. A user-adapted persuasion system in the healthy-eating domain. *IEEE Intelligent Systems* 22: 42-51.

Mochales, R. and Moens, M. 2011. Argumentation mining. *Artificial Intelligence and Law* 19, 1-22.

Reiter, E., Robertson, R., and Osman, L. 2003. Lessons from failure: generating tailored smoking cessation letters. *Artificial Intelligence* 144: 41-58.

Rubinelli, S. and Henkemans, A. F. 2012. Argumentation in the healthcare domain. *Journal of Argumentation in Context* 1(1), 1-3.

Schrauwen et al. 2012. A mutation in CABP2, expressed in cochlear hair cells, causes autosomal recessive hearing impairment. *The American Journal of Human Genetics* 91, 636-45.

Teufel, S. 2010. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization*. Stanford, CA, CSLI Publications.

Walton, D., Reed, C., and Macagno, F. 2008. *Argumentation Schemes*. Cambridge University Press.