Talking About Painful Subjects: Flexibility and Constraints in Patient Interviews

John Moore MD and Henry Lieberman PhD

Abstract

Increasing understanding of how to categorize patient symptoms for efficient diagnosis has led to structured patient interviews and diagnostic flowcharts that can provide diagnostic accuracy and save valuable physician time. But the rigidity of predefined questions and controlled vocabulary for answers can leave patients feeling over-constrained, like the doctor (or computer system) is not really listening to them. In addition, not hearing the patient's own words can lead to the physician overlooking subtle details that are diagnostically relevant. How can we reconcile the need for patients to express themselves with the doctor's need to understand the patient's experience in medically appropriate terms?

We present **I'm Listening**, a system for automatically conducting patient pre-visit interviews. It does not replace a human doctor, but can be used before an office visit to elicit complaint details. This information can be used to triage care and prepare patients for visits with educational materials and appropriate tests, making better use of both doctor and patient time. It uses an on-screen avatar and natural language processing to (partially) understand the patient's response. Key is a Commonsense reasoning system that lets patients express themselves in unconstrained natural language, even using metaphor, and that maps the language to medically relevant categories. For example, if a patient describes his or her pain like, "someone sticking in a knife and then turning it", the system could categorize it as sharp, intense, and localized.

Introduction

Jane is a 63 year old woman who is having trouble with her vision. She describes her chief complaint to the doctor's receptionist over the phone as, "I see floating things in my vision that aren't really there". Before seeing the doctor, she is given a new computerized medical questionnaire to fill out. It takes a standardized new-patient medical history, complete with disease and hospitalization history. Since her complaint is visual, she is given a vision-specific questionnaire. Among the questions, it asks, "Do you see flashes of light in your vision?" She answers, "No".

When Dr. James sees Jane, he looks over the summary of her answers. In a case like this, he suspects posterior vitreous detachment (the vitreous jelly in the back of the eye condenses over time and separates from the retina causing opacities commonly known as "floaters"). Since she answered "No" to the direct question about flashes of light, his suspicion of a complication such as a retinal tear or detachment is low. He asks a series of questions and examines her. In his exam, he sees clear evidence of posterior vitreous detachment. His exam is cursory since he sees this all of the time without complications. Just at the end of his exam, however, he sees a small retinal tear. He is relieved that he did not miss this important finding, but he is confused. He decides to revisit the topic with Jane. "Are you SURE that you're not seeing flashes of light in your vision?". "Yes, I'm sure." she says, "Um... well... I am seeing something, they're not flashes, but they're more like squiggly little lines, kind of like those you see at Fourth of July fireworks, and they come and go."

Dr. James realizes that her answer should have been a "Yes". But she wasn't lying or mistaken, it's just a matter of how the questions and answers are interpreted. Patients sometimes don't understand the vocabulary of medical questionnaires, and when they are forced to choose an answer, the most appropriate choice is not always clear. Questionnaires put words into people's mouths. Human experience is a complicated thing; we don't have enough words in our language to capture all the subtleties of how people experience their bodies.

Our goal is to improve the use of online medical questionnaires, diagnostic flowcharts, and automated medical history and diagnostic systems. We want to provide more flexibility in the way computers interpret what patients say and how the computer responds to them. We aim to reconcile the desire of people to express themselves in their own terms with the need to categorize responses and follow predefined diagnostic procedures (and know when to depart from them).

Our key tools are a mixed-initiative natural language understanding system that can interpret patient responses and a Commonsense reasoning system that has a broad understanding of topics of everyday life. The latter is essential because it provides the ability for the patient to express themselves in metaphorical language ("like Fourth of July fireworks") that can be mapped, either by a system or by a doctor, to medically relevant vocabulary and categories.

For example, a patient can type, "I have been having stabbing pain in my right foot for five days that gets worse when I walk." The system will identify the medical problem as "pain", location as "right foot", duration as "five days", and aggravating factor as "when I walk". In addition, it is able to reason that “stabbing” pain should likely be categorized as “sharp” since a knife can be used for stabbing and knives are sharp. The system then takes a conversational approach to the confirmation of each of the attributes that were determined. The advantage of this technique is that it not only obtains crucial information for the physician in a patient-friendly manner, but it also becomes a learning tool for the patient. It is essentially a dress rehearsal for the patient before visiting the office. Patients can learn what kinds of questions to expect every time that they have a complaint so that they can express themselves more clearly and have their problems addressed more appropriately.

Background

Physicians as Interviewers

Traditionally, physicians consider themselves to be excellent patient interviewers. There is significant research, however, suggesting that this is not true today. They are generally not good listeners. During a standard encounter, studies have shown that physicians interrupt patients in less than 24 seconds.[1,2] In addition, they are not good at explaining medical findings in that they often use terminology that is not well understood by patients.[3] Finally, they are not good at being thorough and performing exhaustive questioning. A study of primary care physicians shows that they asked only 59% of essential history items.[4] In addition, it has also been shown that up to 54% of patient problems are not elicited by physicians.[5] It should be clarified that much of the fault here lies on the constraints that have been placed on physicians and not on their skills. They have limited time for visits and are expected to pick up problems that are not associated with the main reason for a visit. This suggests that we need to design good assistive technologies for physicians, and there is good evidence that data collection is an area where physicians can use help.

Computer-Based Medical Questionnaires

A great deal of research has been done, starting in the 1960s, to prove the benefit of computer-administered medical interviews and questionnaires.[6,7] Bachman[8] provides an excellent review of the literature. As opposed to physicians, computer systems are very good at patiently listening, using patient-appropriate language, and being thorough. In fact, it has been shown in numerous studies that patients report sensitive health information more reliably to computers. These studies have included alcohol use[9], drug use[10], sexual activity[11], suicide attempts[12], and domestic violence[13] among others. Patients have also reported that they appreciate interacting with computer systems because they do not feel rushed and do not feel like they are being judged.

Although it is true that computer-based questioning systems are well received by patients and that they outperform physicians in eliciting thorough histories, their adoption is currently very poor. Statistics are not available, but with only 13% of doctors in this country using electronic medical records and only 4% of them using more advanced systems[14], it is clear that well less than 1% of them are using computer-based questionnaires. In addition, even though patient responses to individual encounters have been positive, little research has been done on optimizing user experience by using more complex interaction techniques so that long-term interactions can be sustained.

Major Flaws in Current Medical Questioning Systems

1. Systems are not Completely Automated

Current systems, such as the one outlined in the introduction, require a human to briefly interview the patient in order to choose the appropriate questionnaire for the patient's chief complaint. There are a number of significant problems with this paradigm. First of all, it involves the communication of sensitive health information to someone other than the physician. Despite new laws for health privacy such as the Health Information Portability and Accountability Act (HIPAA)[15], there is still significant risk for breach of confidentiality. Secondly, it typically requires that the patient come into the office to speak to a staff member and complete the survey. This defeats the purpose of automated medical questioning in that the potential for efficient triaging and staging of diagnostic laboratory tests before the visit is removed. In addition, the opportunity for the patient to complete lengthy interviews from the comfort of home is lost.

2. Focus on Efficiency Leads to Unfriendly Experience

Current systems are designed for maximum efficiency with text-based interfaces with rigidly structured question sequences. Patients may find them novel and interesting on the first encounter, but long-term studies have not been conducted to determine patient satisfaction. Discussions with providers using such systems, however, suggest that patients find these systems tedious and impersonal on multiple encounters.

Tim Bickmore is an alumnus of the Media Lab who is now the lead of the Relational Agents Group at Northeastern University. In his doctoral thesis, he designed a relational anthropomorphic agent that used empathetic remarks and affective facial and body gestures that were appropriate to the state of the relationship in the context of exercise behavior modification. He found that subjects chose to continue interacting with relational agent over a non-relational anthropomorphic agent. In addition, they rated the bond component of the therapeutic alliance inventory higher for their relationship with the relational agent, suggesting that they had a better working relationship with this empathetic agent.[16]

There are many people in the field of human computer interaction who argue for interfaces that are maximally efficient over those that are chatty and friendly. It is certainly true that in many scenarios, users appreciate an interface that is quick and minimal, but it is hypothesized that this type of interface will not be sustainable in the area of automated medical questioning. Since medical questioning is inherently a conversational situation, there is likely significant benefit that will be obtained from enabling the interface with human-like conversational techniques.

3. Rigid Categorization and Controlled Vocabulary Limit Patient Expression

Current systems rely solely on forced-choice responses and do not capture the richness of patient input. This can be dangerous because important details can be lost as portrayed in the patient scenario. In addition, it can be frustrating for patients because they feel that they are not being listened to appropriately. The system always forces a choice that typically does not match exactly what the patient want to express.

4. Value of the Interaction is not Proven to the Patient

Patients spend a great deal of time answering questions offered by current systems, but the benefits from spending this time are not clearly demonstrated to them. If patients do not feel that the time that they spend interacting with a technology is fruitful for them, it is likely that they will resent it. In addition, they will likely resent the physician for making them speak to a computer and limiting face-to-face time.

I'm Listening

I'm Listening is a system for automatically conducting patient pre-visit interviews. It aims to take a dramatically different approach than current computer-based medical questionnaires in order improve patient experience and the usefulness of the data obtained. The goal is for the system to parallel the conversational approach of a physician as closely as possible. Physicians are trained to develop rapport in their interviews with patients while at the same time collecting categorized information that is crucial in diagnosis and treatment planning. They are able to make the patient feel attended to even though they must drive the majority of the interview in order to get the required information.

On-Screen Avatar

I'm Listening currently communicates with the patient through an on-screen avatar using computer-generated speech through a text-to-speech engine. The avatars, Laura and Karen, were designed by Tim Bickmore of the Relational Agents Group at Northeastern University. They are capable of changing her proximity and facial expression in order to match the emotional content of their speech.

Mixed-Initiative Natural Language Understanding System

Physicians are typically trained to take an approach with patients starting with open-ended questions and then drilling-down with more constrained questions. **I'm Listening** takes this same mixed-initiative approach to eliciting patient chief complaints. The chief complaint is the term used to describe the patient's main reason for a visit to the physician. The agent starts with the open-ended question. "What is the main reason for this visit? By that I mean, what is the one problem that is bothering you the most?" The patient is able to enter a complaint that is one sentence or shorter. The system then uses a context-specific natural language understanding algorithm to process the complaint. The algorithm uses tagging, chunking, and parsing rules specific for chief complaints and is capable of identifying the core medical complaint as well as onset, duration, frequency, location, radiation, intensity, character, and aggravating and alleviating factors. For example, a patient can type, "I have been having pain in my right foot for five days that gets worse when I walk." The system will identify the medical problem as "pain", location as "right foot", duration as "five days", and aggravating factor as "when I walk". The system is robust in that the patient could have also typed, "My right foot has been hurting for five days when I walk" or "For five days, there has been pain in my right foot when I walk" or even non-grammatical sentences such as "right foot pain five days when I walk."

Of course the natural language processing algorithms will make mistakes. To address this problem, the system takes a conversational approach to confirming each of the conclusions that it makes. It starts with the most crucial conclusion, which is the core medical problem coupled with the location if appropriate, and progresses to other groups of conclusions. The agent will ask, "It sounds like your problem is right foot pain, is that correct?" If any component of the conclusion is wrong, the system will elicit from the patient which component so that only that one is addressed. It might initially appear to some that such an approach is inefficient and could be tedious. In fact, however, this type of conversation is very similar to the way that a physician conducts an interview. Confirmations not only assure that details were heard correctly, but they also serve the purpose of making the patient feel as if the physician (or computer) is actually listening. If a conclusion is wrong, the system apologizes for the error and elicits the detail in a more constrained frame. It is hypothesized that this apology will help to minimize the user's perception of error rate since there is some evidence that affective support and apology are valuable in minimizing user frustration.[17]

There are a number of advantages of using this natural language approach to eliciting patient complaints. First of all, it allows for the collection of not only the patient's free text complaint, but also the categorized attributes of the complaint, which are useful for the physician. In addition, it performs this whole process in an automated fashion, so that it can be carried out in the comfort of the patient's home without requiring sensitive patient information to be communicated to any third party. It also gives the patient a feeling of being listed to, so that, when the system progresses to more forced-choice questions, these are more acceptable to the patient. Finally, it becomes a learning tool for the patient. The system always asks the patient: "When did the problem start?", "How long has it lasted?", "Is there anything that makes it worse?", "Is there anything that makes it better?", etc. Eventually the patient will start to add these details to the complaint so that the system will not have to ask. The patient starts to understand the attributes of complaints that are important in communication with the physician. Instead of saying, "Aunt Sally thinks that I have pneumonia," they will start saying, "I have been having difficulty breathing for five days that is associated with a productive cough and mild fever."

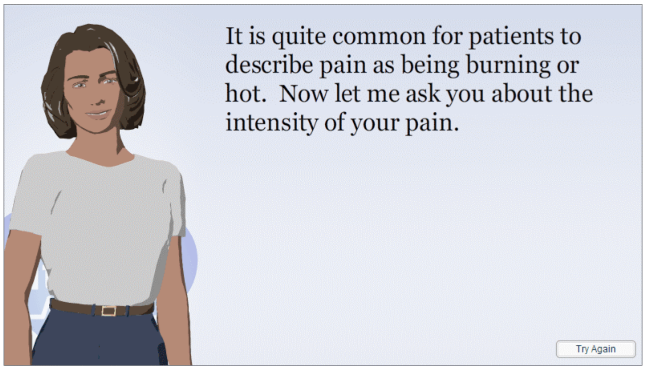
Commonsense Reasoning System

New techniques have been developed to allow open-ended responses from patients to be mapped to system-established choices. The goal is to let patients express themselves and to capture the richness of patient input but, at the same time, to allow the decision tree to proceed appropriately. The approach being used borrows techniques from a new trend in artificial intelligence called Commonsense computing.

The Open Mind Common Sense (OMCS) project is a distributed solution for the collection of common sense knowledge.[18] It enables the general public to enter common sense information through a web interface with 18 different semi-structured frames such as "_________ is a kind of _________" or "_________ is used for _________". In this way, knowledge is entered using natural language, but in a manner that is more reliably machine-interpretable. ConceptNet is then a semantic network representation of the knowledge in OMCS that can be used to explore concepts using advanced techniques including spreading activation.[19, 20] It also allows the translation of the machine-interpretable information back into natural language. Finally, AnalogySpace uses principal component analysis to allow users to infer new common-sense knowledge and to compare concepts.[21]

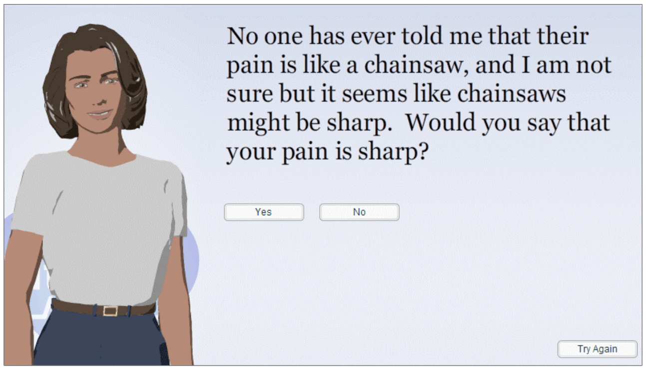
Our Commonsense reasoning system for mapping open-ended patient input to system-established choices uses both ConceptNet and AnalogySpace. The preliminary implementation allows patients to express complaints involving pain. Pain was chosen because it is the reason for over 70 million medical visits per year and because the language used to describe pain complains is some of the most complex. Rather than making the patient choose from a list the word that best characterizes his or her pain, the system gives the patient the opportunity to free-text a description of the pain. The system then stems each of the words in the input and progresses through a series of four main algorithms to attempt to categorize the input.

1. If the stemmed input is equivalent to the stemmed version of any of the pain categories, then that category is assigned to the patient's pain and no further analysis is needed. For example, If the patient says that the pain "burns", "burning" is already one of the pain categories, so no further analysis is needed.



2. Otherwise it is determined if the stemmed input is already a node in ConceptNet, indicating that there is information concerning it in the OMCS database. If so, then it is determined if any of the predefined pain categories is a property of that node with the "HasProperty" relationship. The patient is prompted for confirmation. For example, if the input is "a knife", ConceptNet knows that "a knife" is "sharp." The system presents the association in natural language as shown below.

3. Otherwise AnalogySpace is used to evaluate the possibility that the stemmed input has a "HasProperty" relationship with each of the predefined pain categories. The scores from each evaluation are compared to determine the relationship that is most likely. The patient is prompted for confirmation. For example, if the patient inputs that his or her pain is like a "chainsaw", there might not be a direct relationship between chainsaw and any of the categories of pain. ConceptNet does know that a "chainsaw" is used to cut and that a knife is used to cut and that a knife if sharp. AnalogySpace therefore rates the possibility that a chainsaw is sharp with a relatively high score. Again, the results a conveyed to the user through natural language with the suggestion that the system thinks that the relationship is possible, but that its confidence is lower.



4. If the second and third algorithms fail, then the patient is prompted to choose directly from one or two of the predefined pain categories. If the patient thinks that the chosen categories describe the initial input well, then those categories are added as "HasProperty" relations to the input in the OMCS database.

Although this Commonsense reasoning system has been initially developed to deal with pain complaints, it can now be generalized to deal with any free-text patient input that needs to be mapped to pre-defined categories. The advantage of this approach is that, as more patients interact with the system, the system becomes more intelligent. Once one patient confirms a relationship that did not previously exist, the system will be able to relay the possibility of that relationship to the next patient with a similar input. In addition, the more patients who match a given input with a given category, the higher the confidence the system will have in that relationship and this can be expressed in the language that it uses. Quickly the system will not need to use general information to make associations, but will be able to use context-specific associations that are created from the many users interacting with it.

Just as with the natural language understanding system, not only does the physician receive the categorized version of the patient input, but also the patient's own words describing each aspect of a problem. In this way, the system is flexible yet provides a constrained output.

Related Work

As discussed in the background section, there has been a significant amount of work done in the area of computer-based medical questioning. This work has focused on highly-structured, forced-choice, non-conversational decision trees as opposed to our work that aims to allow conversational structure with more open-ended patient prompts while still capturing structured data that satisfies the goals of a decision tree. Outside of the field of medicine, there are many examples of natural language based conversational systems. These systems typically use statistical inference to drive program flow in cases where input is open ended. The drawbacks of statistical techniques are that the reasoning behind an inference cannot be determined and that there is no opportunity for reasoning outside of previously encountered examples. Our approach of mapping open ended responses to categorical outputs using Commonsense inference is novel in this respect because the semantic relationships between concepts can be used to construct meaningful conversational moves and new responses can potentially be mapped to appropriate outputs through the use of other semantic relationships that are known in the Open Mind Commonsense Database.

Conclusion

Doctors are beginning to benefit from the assistance of computers, but they are going to need even more help in the future to keep up with a tremendous influx of data and increasing need from patients. This does not mean, however, that we need to force patients to interact with unfriendly and inflexible robots in order to get the structured data that is needed. **I'm Listening** is a system for automatically conducting patient pre-visit interviews that presents a number of important advances including a mixed-initiative natural language understanding system and a commonsense reasoning system. These advances are key components that will allow for conversational systems that can mimic important aspects of doctor-patient communication so as to make the interaction friendly and flexible for patients while at the same time collecting structured and categorized that is usable by the computer and useful to the physician.

Future Work

Immediate work on this project involves generalizing the Commonsense reasoning system so that it can deal with a broad range of medical complaints. In addition, testing the system on a large number of patient complaints will be crucial in determining its initial performance and its learning abilities.

The work highlighted here is part of a larger effort by the New Media Medicine group at the MIT Media Laboratory to enable radical new collaborations between doctors, patients, and communities. The goal is to develop technologies that allow patients to take a more active role in their care such that they become equal partners with their medical providers. One project that is being developed in parallel is a multimodal (speech and touch) workstation for doctor-patient shared decision making. A goal is to integrate the agent from **I'm Listening** into this interface so that there can be three way conversations between the doctor, patient, and technology. This will likely have profound effects on the psychological relationships that patients will have with the agent and their evaluations of the value of such a system. Another project being developed by Ian Eslick of the New Media Medicine group is a collective intelligence system for disease communities. The agent-based conversation techniques will also be adapted to allow for more patient-friendly data input in this system. It will be a recurring theme of the group's work to design technologies that not only improve the health of patients but also engage them and make them feel better about their healthcare experience.

Acknowledgements

We would like to thank Zuzana Trnovcova for her hard work in the development of the mixed-initiative natural language understanding system. We would also like to thank Ryan Young for his contribution to the initial prototype of I'm Listening and Ken Arnold for his assistance with ConceptNet and AnalogySpace.

References

1. Marvel MK, et al. Soliciting the patient's agenda: have we improved? *JAMA*. 1999; 281: 283-287.
2. Rhoades DR, et al. Speaking and interruptions during primary care office visits. *Fam Med*. 2001; 33: 528-532.
3. DiMatteo MR. *The Psychology of Health, Illness, and Medical Care: An Individual Perspective*. Pacific Grove, Calif: Brooks/Cole; 1991.
4. Ramsey PG, et al. History-taking and preventive medicine skills among primary care physicians: an assessment using standardized patients. *Am J Med.* 1998; 104: 152-158.
5. Stewart MA et al. The doctor/patient relationship and its effect upon outcome. *J R Coll Gen Pract*. 1979; 29: 77-81.
6. Slack WV, Van Cura LJ. Patient reaction to computerbased medical interviewing. *Comput Biomed Res.* 1968; 1: 527-31.
7. Mayne JG, Weksel W, Sholtz PN. Toward automating the medical history. *Mayo Clin Proc.* 1968; 43: 1-25.
8. Bachman JW. The Patient-Computer Interview: A Neglected Tool That Can Aid the Clinician. *Mayo Clin Proc*. 2003; 78: 67-78.
9. Lucas RW, et al. Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: a comparison. *Br J Psychiatry*. 1977; 131: 160-167.
10. Paperny DM, et al. Computer-assisted detection and intervention in adolescent high-risk health behaviors. *J Pediatr.* 1990; 116: 456-462.
11. Millstein SG, Irwin CE Jr. Acceptability of computeracquired sexual histories in adolescent girls. *J Pediatr.* 1983; 103: 815-819.
12. Greist JH, et al. A computer interview for suicide-risk prediction. *Am J Psychiatry.* 1973; 130: 1327-1332.
13. Lapham SC, Kring MK, Skipper B. Prenatal behavioral risk screening by computer in a health maintenance organization-based prenatal care clinic. *Am J Obstet Gynecol.* 1991; 165: 506-514.
14. DesRoches CM, et al. Electronic health records in ambulatory care--a national survey of physicians. *N Engl J Med*. 2008 Jul 3; 359(1): 50-60.
15. Department of Health and Human Services. "Standards for Privacy of Individually Identifiable Health Information." Federal Register. 2002: 67: 53181-53273.
16. Bickmore T, Gruber A, Picard R. Establishing the Computer-Patient Working Alliance in Automated Health Behavior Change Interventions. *Patient Education and Counseling*. 2005; 59(1): 21-30.
17. Klein J, Moon Y, Picard RW. This Computer Responds to User Frustration. Conference on Human Factors in Computing Systems, CHI '99 extended abstracts on Human factors in computing systems. 1999: 242-243.
18. Singh P, Lin T, Mueller ET, Lim G, Perkins T, Zhu WL. Open Mind Common Sense: Knowledge acquisition from the general public. Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems. Irvine, CA. 2002.
19. Lieberman H, Liu H, Singh P, Barry B. Beating some common sense into interactive applications. AI Magazine 2004.
20. Havasi C, Speer R, Alonso, J. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. Proceedings of Recent Advances in Natural Languge Processing. 2007.
21. Speer R, Havasi C, Lieberman H. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. Proceedings of AAAI. July 2008.