Available online at www.sciencedirect.com
ScienceDirect
Journal of Biomedical Informatics 39 (2006) 541–555
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Automatic analysis of medical dialogue in the home hemodialysis domain: Structure induction and summarization

Ronilda C. Lacson *, Regina Barzilay, William J. Long

Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Received 25 July 2005
Available online 2 February 2006

Abstract

Spoken medical dialogue is a valuable source of information for patients and caregivers. This work presents a first step towards automatic analysis and summarization of spoken medical dialogue. We first abstract a dialogue into a sequence of semantic categories using linguistic and contextual features integrated in a supervised machine-learning framework. Our model has a classification accuracy of 73%, compared to 33% achieved by a majority baseline ($p < 0.01$). We then describe and implement a summarizer that utilizes this automatically induced structure. Our evaluation results indicate that automatically generated summaries exhibit high resemblance to summaries written by humans. In addition, task-based evaluation shows that physicians can reasonably answer questions related to patient care by looking at the automatically generated summaries alone, in contrast to the physicians' performance when they were given summaries from a naïve summarizer ($p < 0.05$). This work demonstrates the feasibility of automatically structuring and summarizing spoken medical dialogue.

Keywords: Spoken dialogue analysis; Medical discourse; Summarization; Natural language processing; Information processing

1. Introduction

Medical dialogue occurs in almost all types of patient–caregiver interaction, and forms a foundation for diagnosis, prevention, and therapeutic management. In fact, studies show that up to 80% of diagnostic assessments are based solely on the patient–caregiver interview [1]. Automatic processing of medical dialogue is desirable in multiple contexts—from clinical and educational, to financial and legal. Caregivers can use the results of this processing for informed decision making, researchers can benefit from large volumes of patient-related data currently unavailable in medical records, and health-care providers can enhance communication with patients by understanding their concerns and needs. All of these users share a common constraint: none of them wants to wade through a recording or transcript of the entire interaction.

To illustrate the difficulty of accessing medical dialogue, consider 30 s of an error-free transcript of an interaction between a dialysis patient and a nurse (see Fig. 1). This excerpt exhibits an informal, verbose style of medical dialogue—interleaved false starts (such as "**I'll pick up**, **I'll give you a box of them**"), extraneous filler words (such as "**ok**"), and non-lexical filled pauses (such as "**Umm**"). This exposition also highlights the striking lack of structure in the transcript: a request for more supplies (e.g., "**kidney**," which in this context refers to a dialyzer) switches to a question about a patient's symptom (e.g., shoulder pain) without any visible delineation customary in written text. Therefore, a critical problem for processing dialogue transcripts is to provide information about their internal structure.

This paper presents the first attempt to analyze, structure, and summarize dialogues in the medical domain. Our method operates as part of a system that analyzes telephone consultations between nurses and dialysis patients in the home hemodialysis program at Lynchburg Nephrolo-

* Corresponding author. Fax: +1 781 652 0404.
E-mail address: rclacson@mit.edu (R.C. Lacson).

1532-0464/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2005.12.009

```
(1) Umm, I'm out of kidneys
(2) Out of kidneys, ok
(3) Give me a box of them
(4) A box of them, ok, I'll pick up, I'll give you a box of them
(5) Ok
(6) And I'll leave them in the room, do you know where the coolers
    are?
(7) Yeah
(8) Ok, I'll leave them in there with your name on it
(9) Ok
(10) Ok, how's the Vioxx helping your shoulder?
(11) Oh, now I haven't actually tried to do anything, I haven't lifted
    weights for 2 weeks
```

Fig. 1. Transcribed segment of a phone dialogue.

gy, the largest such program in the United States [2]. By identifying the type of a turn—Clinical, Technical, Backchannel or Miscellaneous—we are able to render the transcript into a structured format, amenable to automatic summarization. The Clinical category represents the patient's health, the Technical category encompasses problems with operating dialysis machines, the Miscellaneous category includes mostly scheduling and social concerns, while Backchannels capture greetings and acknowledgments.

In addition, automatically processing medical dialogue has important implications for the development and evaluation of conversational systems. Current methods for developing automated dialogue systems rely on large amounts of labeled data for training (ref); human annotation of this material is an expensive and lengthy process. Our system can provide an initial annotation which can be further refined by a human, if necessary. Furthermore, for evaluation of automated dialogue systems, structure of the dialogue can be analyzed and compared to human–human dialogues. An interesting direction in analyzing the performance of automated dialogue systems is their comparison with human–human dialogues. Understanding similarities and differences in structure between human–human and machine–human dialogues can further advance the development of automated systems. Our method may also be used for mixed conversational systems, in which part of the dialogue is routed to an automated system (i.e., scheduling), as opposed to a clinical or technical query, which requires the attention of a human caregiver. Finally, our classification allows a provider to distill the portions of the dialogue that support medical reasoning and are of primary interest to clinicians. In the long run, knowing the distribution of patient requests can improve the allocation of resources, and ultimately provide better quality of health care.

Our system has two main components:

Structure induction. We present a machine-learning algorithm for classifying dialogue turns with respect to their semantic type. The algorithm's input is a transcription of spoken dialogue, where boundaries between speakers are identified, but the semantic type of the dialogue turn is unknown. The algorithm's output is a label for each utterance, identifying it as Clinical, Technical, Backchannel, and Miscellaneous. Our algorithm makes this prediction based on a shallow meaning representation encoded in lexical and contextual features. We further improve the classification accuracy by augmenting the input representation with background medical knowledge.

Summarization. We introduce a novel way to extract essential dialogue turns within our domain of spoken medical dialogue using the discourse structure just described. Our goal is to provide a caregiver with a succinct summary that preserves the content of a medical dialogue, thereby reducing the need to leaf through a massive amount of unstructured and verbose transcript.

To assess the performance of the summarizer and the contribution of structure induction, we describe a framework for evaluation of medical dialogues. Our first evaluation method follows an intrinsic methodology, commonly used in the text summarization community [3]. We compare automatically generated summaries with a ''gold stan-