Project Title:

Segmenting Shared Medical Decision Making Communication Sequence Data

Faculty Advisors

Dr. Nan Kong, Assistant Professor, Weldon School of Biomedical Engineering, Affiliated Faculty, Regenstrief Center for Healthcare Engineering, Purdue University

Dr. Cleveland Shields, Associate Professor, Department of Human Development and Family Studies, Purdue University

Dr. Xuegong Zhang, Professor, Department of Automation, Deputy Director, MOE Key Laboratory of Bioinformatics, Tsinghua University
Introduction

Patient with advanced cancer experience serious problems in pain management, understanding their prognosis, and consideration of their preferences in care. The level of such problems is considered as a function of communication and shared decision-making. The existing literature on understanding this function has focused on casual studies using linear regression analysis on survey data. However, prediction performance of these regression models developed is typically less satisfactory. Little work has applied alternative techniques or mined the detailed communication data with the attempt to improve the prediction accuracy. More importantly, it remains unclear to how to influence physicians during the consultation sessions to improve the patient satisfaction. In this project, we explore the use of unsupervised learning techniques, in particular, sequence clustering, to investigate detailed communication sequence data.

Detailed Description

In this project, we will use contextual data extracted from audio-recordings of physician visits with unannounced standard patients (SPs), including visits to oncologists and primary care physicians. SP portrayals were calibrated through hours of extensive training to achieve role accuracy in terms of both verbal content and emotional value of the role. The SPs were well-received by physicians and their staff. Decision-making sections of the audio-recordings were extracted and coded for topic and behavior by supervised undergraduate students.

This summer internship consists of two phases. In the first phase, which will be conducted in Beijing under the supervision of Dr. Zhang, the students are expected to conduct self-learning and familiarize sequence clustering methods, especially model-based methods. In the second phase, which will be conducted in West Lafayette, IN, under the supervision of Dr. Kong, the students are expected to implement the sequence clustering methods, especially the EM method.

Student Skill Requirements and Background Preference

Two students, one from Tsinghua and one from Purdue, will be paired. The students are expected to not only work closely in the research but also help each other in administrative issues that facilitate the exchange. The students are expected to have experience in C/C+ or Matlab coding. The students with some experience in statistical analysis software, e.g., R or SAS, are preferable.

Timeline

In Tsinghua (5 weeks, June 1 – July 3)

The students are expected to perform the tasks under the supervision of Dr. Zhang. This phase is about self-learning and conceptual design.

Transition (1 week, July 6 – July 10)

In Purdue (5 weeks, July 13 – August 14)
The students are expected to perform the tasks in the second phase under the supervision of Dr. Kong and Dr. Shields. This phase is about actual algorithm implementation and performance comparison. The students are also expected to explore the data mining methods for natural language processing.

**Important Issues**

- The selected student can apply for travel grants of several entities on campus, including study abroad. Our goal is to cover your airfare, visa application fee, and some of your living expenses in Beijing.
- Research credits may be registered in your home department, depending on your home department’s policy.
- You are strongly encouraged to seek help from the Tsinghua student you will be paired with and/or other Tsinghua students you are acquainted with.