PROJECT INFORMATION

April 2021

Identifying Topics in Patient and Doctor Conversations Using Natural Language Processing Methods

Principal investigator

Zac E. Imel, PhD

Organization

University of Utah

What was the project about?

The way doctors communicate with patients during office visits can affect the quality of care. Studying conversations between doctors and patients can help doctors improve their communication skills.

To study conversations, researchers rely on written records, or transcripts, of office visits. They read the transcripts and give each conversation topic a label. For example, topics may include smoking or pain. But labeling topics in this way may take a lot of time.

In this project, the research team created and tested a new method to make this work easier using natural language processing, or NLP. With NLP, computer programs interpret written language. NLP methods use a process called machine learning, where computer programs use data to learn how to perform different tasks with little or no human input.

What did the research team do?

First, the research team trained the NLP computer program to label topics in transcripts. The team used 279 transcripts of patient-doctor conversations from two previous studies. A researcher had already assigned labels to topics in these transcripts. Using these transcripts, the NLP computer program learned to associate specific words with specific labels. For example, the word cigarette would be labeled as smoking. The program could then label topics in other transcripts.

Next, the research team used the NLP computer program to create three types of NLP methods. Each method used a different statistical approach to label topics. The methods were called non-sequential, window-based, and sequential.

Then the research team tested how well the NLP methods worked to label topics in new transcripts. The team compared topic labels from the new NLP methods with topic labels assigned by a person. They also compared the new methods with a basic NLP method that only labeled the most common topics.

Patients provided input during the project, including how to explain complex research methods in a clear way.

What were the results?

The three new NLP methods labeled topics more accurately than the basic NLP method. The sequential NLP method labeled topics more like a person than the non-sequential NLP method.

When topics were similar, like weight and diet, the methods were less accurate.

What were the limits of the project?

How well the new NLP methods work depends on the data used to train the NLP computer program. The methods may be less accurate for topics that patients and doctors rarely talk about.

Future studies could test if adding other data to the training transcripts, like doctor's notes, improves the methods' accuracy.

How can people use the results?

Researchers can use the new NLP methods when studying conversations between patients and doctors to help doctors improve their communication skills.

*To learn more about this project, visit
www.pcori.org/lmel369.*