Professor Qianwen Wang (University of Minnesota) - "Interpreting and Steering AI Explanations with Interactive Visualizations"

There are many different relationships between AI and humans. The speaker states that there is one relationship that is very important to her which is, “How humans use and apply AI techniques to solve big tasks.” I really liked the slide which stated that even if we made no further progress in regard to AI, simply using existing AI on all possible problems would be a “game changer for most industries.” To me this really hits on the promise of AI as it relates to changing our society (hopefully for the better).

However, many people are anxious about AI. Dr. Wang looks at the medical field and notes that AI software has delivered inaccurate information and this can lead to trust issues for both the patient and the physician. These trust issues, of course, can extend to other fields besides medicine. How do we overcome the challenges of AI delivering faulty information and the general users distrust of AI?

Dr. Wang acknowledges that this can be hard to achieve and demonstrates this by describing how AI may be used in drug development. What AI needs to consider during this development phase is different from what AI needs to consider during application of said drug to patients. Her premise is that to effectively use AI the gap between these two sides must be bridged and that the way to do this is through interactive visualization. A model of communication between two people is shown where they have a discussion about a patient’s diagnosis. It is proposed that this model of communication should be used when a person has a discussion with AI software.

The model has three parts which are receive, interpret, and feedback. For receiving there are different types of explanations including receive and counterfactual. This part of the model, receive, seems to be well defined but the speaker posits that receiving explanations does not guarantee proper insights. That is, what is received may not be interpreted properly. So, there is an issue around “how to select and present a visual explanation that can lead to proper insights.” (speaker’s slide) Even for something so simple as presenting a graph neural network there are choices in how to represent the explanation, neighbor nodes, subgraph, or paths. Some user studies follow which demonstrate some different alternatives for formatting explanations.

The main thrust of this is that Dr. Wang is trying to determine if we can format AI explanations that will allow users to steer AI as well as interpret results from AI. To accomplish this Dr. Wang uses anchors. An anchor is two similar testers? captures? clusters? (I had a really hard time trying to decipher what work our presenter was using here, which is unfortunate. I listened to the video over and over and just could not make out the word. As I watch the video I think it is clusters.) from the same dataset. This leads into polyphony which is an interactive AI tool that assists in generating anchors by generating anchor recommendations.

Unfortunately, knowledge is a bit more complicated than classes. Sometimes there are no clear clusters in which case how would we use anchors? Knowledge, essentially, is represented by concepts. This is the case when there are no clusters where what we really want to do is to extract concepts from the AI model.

In the future Dr. Wang’s research will include safeguards for human interaction. In my opinion this is really important since AI is definitely imperfect. A second area is multimodality, an area which will improve AI comprehension as this will highlight patterns which would be missed by a single modality. Finally, she is planning to explore how AI might be used by science in general. For example, will it be able to generate and validate hypotheses?