Research Statement

Assadullah Mohammadi

1. Overview

As we know machine learning has been very successful in a difference of applications, there has been less responsibility on the with using this technology. Humans are involved in all steps of the machine learning process: developers building prediction models, and users interpreting the predictions, making decisions, annotators labeling the data and many other facilities. **The core objective of my research is to make machine learning more useful for humans by better understanding human interactions and building systems with rich interactions.** To that end, I develop novel machine learning models, derive efficient inference algorithms, design user interface, and conduct user studies. I consider a number of application domains: bio-medical text analysis, citizen science, and credibility of information.

2. Prior work

Better understanding and better use of annotators

Large-scale and high-quality data is critical for machine learning. We know a good dataset requires a large amount of human work to clean, extract, label, and verify the data. Crowd sourcing has emerged as a mechanism to distribute work over the Internet at a low cost, which led to the creation of many popular datasets. However, the human interactions are not well-understood and often limited to very simple micro tasks, which are then aggregated (e.g., by majority vote). My research addresses this in several aspects. I have studied many research papers, to finding techniques for developing models to better evaluate annotators. Having good estimates of annotator performance is for task routing. My key idea is to transfer these annotator performance estimates within groups (of similar annotators), between labeling tasks (when the data is available), and between data classes. Applying this method to citizen science data (where the human annotators volunteer to contribute to science), I have found significant improvement, especially from transfer within groups. I have planned to build learning models that account for each individual annotation, instead of just learning from the aggregated annotations. I am considering sequential annotations, which are popular in natural language application.

2.2 Being transparent and interactive for users

Machine learning models are often built to maximize prediction performance (such as accuracy on a test dataset), and then shipped to end users as a black box. This becomes an issue in applications that are mission-critical, have serious consequences, or require trust by end users. One of those applications is the prediction of the credibility of information based on relevant evidence. In this direction, I am building a prototype system that takes a textual claim as input from users (e.g. ‘Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language’), then retrieves relevant articles from many web sources (‘No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart’ by gizmodo.com). The system then predicts whether 1 each article supports, denies or is neutral about the claim. It then

combines evidence from all articles and the predicted reputation of each source in order to produce a final prediction on the credibility of the claim (whether the claim is true or not). From a machine learning perspective, predicting credibility is a classification problem, where previous work has primarily optimized prediction performance. I instead take a more user-centered approach in designing a system where users can observe how it arrives at the prediction and interact with that prediction. For example, users can change the reputation of a web source, or the stance (support/deny) of an article, and see how the final prediction changes. This interaction allows users to make sense of how the system works and inject their knowledge to get a more personalized prediction. Results from a user study suggest that this interactive feature helps users make better predictions. Although this research direction focuses on one specific application, it illustrates more general patterns of how end users interact with machine learning systems. I found that users are usually receptive to using machine learning in a potentially contentious area.

3. Future directions

Looking forward, I will continue my research in machine learning with human interaction. As machine learning is being deployed in human-facing applications, I expect this area to both be fruitful and have a large impact on society. My general research plan is to adapt techniques in human-computer interaction (HCI) to machine learning problems in both understanding human interaction and building interactive systems. There are several concrete directions I will pursue in the near future. Software tools for developing transparent and interactive machine learning systems. Software tools are important in machine learning development for enabling developers to quickly specify models while leaving some details of inference and learning to be automatically handled. For example, a computation graph library (Tensorflow , PyTorch , Theano) provides an automatically-derived gradient for each variable, enabling gradient-based learning (such as gradient descent). I am interested in extending these existing software tools to support developers in building transparent machine learning systems.

My extension will enable the binding of machine learning components (variables, neural layers) to User Interface (UI) elements (sliders, graphs). The results will be a UI where end users can interact with the internals of the machine learning system. This can also be used to diagnose failure or elicit user knowledge for better predictions. Besides providing the tools for developers, I expect the creation of these tools will be a step toward distilling the general principles of the interaction between humans and machine learning systems. Volunteered crowdsourcing for social good. Crowdsourced annotations in machine learning is mostly paid work, although there have been many notable successes in volunteered annotations from citizen science. I am interested in machine learning systems that benefit society, which are compelling at attracting public participation, not only in annotating data but also in steering the machine learning systems to be more accountable and fair. For example, the crowd may examine a system's predictions to identify potential biases. Presenting and interacting with uncertainty. Since any prediction is uncertain, the representation of uncertainty has been a well-studied area in machine learning, especially in the probabilistic approach. However, less work has been done on the human interaction side. Uncertainty is often presented statically (e.g., percentage of confidence in the predicted class), and with little explanation. I am interested in developing a more interactive representation of

uncertainty, enabling users to have a better sense of where that uncertainty is coming from, which can potentially lead to better human decisions. Credibility of Information. The above research directions can be realized in the application of predicting the credibility of information. I am interested in extending my previous work in this application using principled interaction tools, with mechanisms for end users to make meaningful contributions, and clear explanations of uncertainty. My interest is not limited to the news domain, but is in general knowledge, for example in Wikipedia or scientific research. While the problem of information overload is affecting more people, I see a real need for a technological solution for assisting humans in sense-making of information.

Note: This research statement is only a draft format; it is flexible to change by my supervisor suggestions.