Research Statement
Dhanya Sridhar

My research is on machine learning and causality, developing algorithms that answer causal questions from data. I focus on methods that allow social scientists, economists and policy experts to study human behavior from large-scale data. **My research vision is to create machine learning methods that draw causal conclusions about human behavior and help researchers make better policy decisions.**

Data abound on social media platforms, e-commerce sites, online forums, and news media sites. These data are snapshots into what people think, who they interact with, and what they like. We can make scientific progress if we could exploit such datasets to answer causal questions about human behavior. Do users on social media influence the choices of their peers? How does a social media user's writing choices affect the responses to their posts? Can the wording of a news article cause people's opinions to change?

Although datasets about people show promise for answering causal questions, there are challenges. First, we must identify which causal questions can be answered with the data we have. In datasets about human behaviors, causal patterns are obscured by non-causal correlations which may be impossible to disentangle. Second, even if estimating a causal effect is possible, we require computational methods that reliably recover answers from finite data. Finally, to use causal inferences for decision making, statistical methods need to be robust and come with guarantees about validity.

I develop modern machine learning tools to address these fundamental challenges. My work formalizes the causal questions that be answered, develops efficient estimation algorithms using both probabilistic and black box machine learning, and provides results about the validity of causal conclusions. My technical contributions include methods that answer causal questions from text, social networks and biological data. While I focus on applications to the social sciences, the methods I develop are broadly useful across scientific domains.

1 Previous Research

My work on causality and machine learning has two broad parts: extracting causal patterns from text and social network data, and developing new models when standard machine learning assumptions are violated.

Causal inference with machine learning. The goal of causal inference is to discover whether manipulating one variable (treatment) changes another (outcome), invariant to all other changes. Establishing a causal relationship between the treatment and outcome is difficult because it requires ruling out all other sources of correlation between the two.

One focus of my research is identifying the causal questions we can answer from text data and developing methods for effect estimation. For example, consider comments from Reddit.com and their popularity. We want to study the effect of treatments on comment popularity. My work formalizes causal inference in two settings: when treatments are properties of text such as the sentiment, or non-text attributes such as whether or not the user is a frequent poster on the site. When treatments are properties of the high dimensional text, we typically only have noisy measurements of these properties, e.g., by predicting sentiment using a separately trained classifier. In this setting, specifying the causal effect of interest is difficult, since the concept of manipulating text properties is inherently ambiguous. In [2], I formalized causal effects of text properties and proposed an estimation procedure that exploits noisy measurements of properties from text. I proved that the bias due to measurement error can be bounded, making this estimation procedure practical.

The next challenge for causal inference is eliminating non-causal correlations due to variables called confounders. Text documents such as comments encode confounders such as topic, variables that affect both the treatment (e.g., posting frequency) and outcome (e.g., popularity). Instead of working with high-dimensional text, it is common to replace the text with a low-dimensional numerical representation, called an embedding. To exploit this idea for causal inference, in [11], I introduced causally sufficient text representation learning, a method for adapting any language model to produce embeddings that contain sufficient information about confounders. Both black box embeddings and probabilistic models benefited from the causally sufficient learning approach, yielding more accurate causal inferences than their standard counterparts.

My work has also studied the causal questions that we can ask about social networks. For example, a user on a social media site may share a news article one day; the next day, their friend does the same. Did the user influence her friend? From observational data, it can be impossible to tell whether this shared behavior is due to the user influencing her friend, or because the users have similar tastes in articles, which played a role in their friendship in the first place. The latter effect is called homophily, a process by which similar people selectively become connected in a network. The problem of disentangling influence, a causal quantity, and homophily is made more challenging because the traits of users that played a role in homophily are unobserved. In [5], I proposed a probabilistic matrix factorization method for estimating causal influence from observational data. I formalized the insight that the network structure and people's high-dimensional behaviors, such as which articles they share, provide indirect evidence of people's latent traits and preferences, which can be estimated using probabilistic factor models. In a different vein, I also developed probabilistic machine learning methods to discover causal patterns from biological networks. [7].

Probabilistic models for relational and heterogeneous data . To make causal inferences, we require statistical methods that can estimate relationships between variables from data. However, standard machine learning assumptions are often not appropriate for the data we observe about people. For example, on social media and online forums, users engage in conversations with other users. Conversations are best viewed as a graph, where reply interactions induce dependencies across the latent properties of users (and their posts). We refer to these types of dependent data as relational. Relational data are not independent samples from a distribution, an assumption underpinning modern machine learning.

A key thread of my research is on developing machine learning methods that are suitable for relational data. In [9], I developed probabilistic models of online debate conversations based on Markov random fields, undirected graphical models that can capture dependencies between variables. The probabilistic framework formalized the insight that modeling disagreement between a post and its reply together with the posts' content helps with text understanding, since disagreeing authors tend to have differing views. Separately, this approach was also adapted for analyzing biological data [8], demonstrating its flexibility. In [6], I formalized causal inference for relational data, finding the causal effects that one user has on another during a conversation. The work addressed foundational challenge such as obtaining statistical samples from conversational graphs and measuring noisy outcomes relevant to conversation success from high-dimensional text.

A second line of my research has focused on capturing heterogeneity in text data. Statistical patterns in text data typically exhibit heterogeneity, where the information relevant for a prediction task varies across different contexts. For example, given product reviews and their associated ratings, different aspects of the text will be predictive of the rating for different unobserved categories of products. In [4], I developed heterogeneous supervised topic models, a new probabilistic approach for jointly uncovering the topics in text data and discovering how these topics modulate prediction from text. I adapted modern optimization techniques to fit these models, applying the autoencoding variational Bayes framework. This method outperformed state-of-the-art neural methods such as BERT for certain settings, while being much more interpretable.

2 Research Vision

The amount of data we have on people is growing as digital platforms play an increasing role in our lives. The content that we view and share online affects political campaigns, beliefs about social issues, and public health interventions. At the same time, these data grow complex: people post, interact with their friends, and engage with everything from news articles to political organizations, creating multiple sources of information about users that are all interdependent. My research will bridge the gap between data on people and the causal inference methods needed to answer scientific questions with them.

Improving decision-making with causality. Causal inference provides an opportunity to improve people's lives. By extracting causal conclusions from data, we can identify actions that lead to better societal outcomes. My work on the effects of text and networks [11, 5, 6, 2] can shed light on how to be more persuasive or how to curb the spread of misinformation in a network due to influence. I have also developed fair decision making algorithms based on causality [10], and improved causal effect estimation for genetics [1]. However, there are several challenges for adapting causal inference in service of decision-making. I am broadly interested in addressing these challenges. One area of interest is developing methods that find heterogeneous effects across subgroups of people to enable personalized decisions. I am also interested in combining economic models and causality to develop algorithmic decisions that are equitable and fair.

Novel machine learning methods for datasets on people. The datasets we use to make causal inferences about human behavior are complex. They involve text, networks between people, and interactions between people and organizations such as news media, political figures, or e-commerce sites. To draw causal conclusions from these data with machine learning, we require new methods that combine these different sources of information while being interpretable and robust. My work on representation learning and probabilistic modeling [2, 6, 11, 4, 9, 5] provides a framework for analyzing text and networks, but several open challenges remain. I am interested in three broad areas for improving machine learning methods: learning from multiple modalities of data, marrying neural and probabilistic approaches, and making algorithms robust against spurious conclusions

Collaborations with applied researchers. Digital platforms have both beneficial and malicious consequences for the functioning of society. On one hand, they have already been implicated in spreading misinformation and sowing distrust in public policy. On the other hand, digital interactions effectively mobilize voters and convey important public health information. As digital platforms increasingly affect our lives, it becomes urgent to work with researchers that are at the forefront of these policy questions. My collaboration with researchers on a large-scale digital experiment revealed that peer-to-peer texting has large effects on voter turnout [3]. My research vision is to develop causal inference methods that facilitate answers to real social science questions about human behavior on digital platforms. I envision several collaborations with applied researchers to conduct large-scale studies and experiments on digital platforms. I am especially driven by questions such as: what strategies are effective in curbing the spread of misinformation; how can we write more persuasively to convince readers about useful information; can we measure people's opinions and reactions to policy changes?

Broader Impact. I strive to conduct technical research that meets interdisciplinary needs. Thus, I place an emphasis on communicating the role of causality to a wider audience. I am co-organizing the first workshop on causality and natural language processing, aimed at bringing new practitioners into the field of causal inference. I also co-organized the Data Science Institute seminar series, bridging a broad machine learning audience and applied researchers. I am committed to having a diverse group of students and collaborators. To do work that will benefit people and reduce harms, it is important to have a wide range of perspectives.

References

[1] Hartford, J., Veitch, V., Sridhar, D., & Leyton-Brown, K. (2020). Valid Causal Inference with (Some) Invalid Instruments. arXiv preprint arXiv:2006.11386. Under submission at AISTATS.

[2] Pryzant, R., Card, D., Jurafsky, D., Veitch, V., & Sridhar, D. (2020). Causal Effects of Linguistic Properties. arXiv preprint arXiv:2010.12919. Under submission at NAACL.

[3] Schein, A., Vafa, K., Sridhar, D., Veitch, V., Quinn, J., Moffet, J., ... & Green, D. P. (2020). A Digital Field Experiment Reveals Large Effects of Friend-to-Friend Texting on Voter Turnout. Available at SSRN 3696179. Under submission at WWW.

[4] Sridhar, D., Daume III, H., & Blei, D.M. (2020). Heterogeneous Supervised Topic Models for Text Prediction. Work in progress.

[5] Sridhar, D., De Bacco, C., & Blei, D.M. (2020). Estimating Social Influence from Observational Data. Under submission at WWW.

[6] Sridhar, D., & Getoor, L. (2019). Estimating the Causal Effects of Tone in Online Debates. In IJCAI.

[7] Sridhar, D., Pujara, J., & Getoor, L. (2018). Scalable Probabilistic Causal Structure Discovery. In IJCAI.

[8] Sridhar, D., Fakhraei, S., & Getoor, L. (2016). A probabilistic approach for collective similarity-based drug–drug interaction prediction. Bioinformatics, 32(20), 3175-3182.

[9] Sridhar, D., Foulds, J.R., Huang, B., Getoor, L., & Walker, M. (2015). Joint Models of Disagreement and Stance in Online Debates. In ACL.

[10] Wang, Y., Sridhar, D., & Blei, D. M. (2019). Algorithmic Approaches to Equal Opportunity and Affirmative Action via Counterfactual Predictions. arXiv preprint arXiv:1905.10870. Under submission at Science Advances.

[11] Veitch, V.*, Sridhar, D.*[1], & Blei, D.M. (2020). Adapting Text Embeddings for Causal Inference. In UAI.

[1] Equal contribution