The Workshop on Forecasting Launches a New Special Focus

[April 2021] Forecasting is a fundamental building block in science and engineering whose study draws scholars from statistics, computer science, math, operations research, and business, among other areas. Against the recent backdrop of a contested presidential election and a deadly pandemic, research in forecasting is especially timely as scientists make projections about the pandemic, pollsters analyze predictions, and citizens grapple with the concepts of uncertainty, dynamics, correlations, and distributions.

To catalyze research in forecasting, DIMACS held the Workshop on Forecasting: From Forecasts to Decisions. The event was held online March 17-19, 2021 and featured: five inspiring invited talks on topics ranging from election forecasting to predicting which social science experiments will replicate; nine contributed talks on prediction markets, peer prediction, scoring rules, wisdom of crowds, and more; and additional technical contributions presented at daily poster sessions that consistently continued past their official ending times.

The Workshop on Forecasting is the second in a series on forecasting and the first of five workshops planned for the DIMACS Special Focus on Mechanisms and Algorithms to Augment Human Decision Making. While we hope to hold the four remaining events in person, we found that there can be much to love about virtual spaces. In this article, we share a few highlights about the workshop's content and its online format.

The invited speakers, Yael Grushka-Cockayne, Harry Crane, Anna Dreber, Andrew Gelman, and Barbara Mellers, and Ville Satopää, represented a variety of backgrounds—statistics, economics, business, operations, and psychology. Many of them gave practical advice on how to improve forecasting developed over years of research and experience.

Mellers emphasized three drivers that influence forecast accuracy: (1) training forecasters to reason about probability, (2) forming teams, and (3) tracking the top two percent of forecasters and placing them into their own "superforecasting" teams. Satopää presented work with Mellers on how to improve accuracy, reduce bias, and increase information in forecasts. A key finding was that, over time, forecasts tend to gain information by correcting their bias but without reducing their level of noise.

Grushka-Cockayne presented a wonderfully easy yet highly effective method of combining correlated forecasts by inferring a single common correlation coefficient. The technique performs much better than methods that ignore correlations and also methods that try to infer a full covariance matrix, which tends to overfit.

Dreber discussed her research using prediction markets to estimate which published social science experiments (e.g., psychology and economics experiments) contain robust conclusions that stand up when the original experiments are repeated and which contain questionable or

unstable results that don't replicate reliably. Recent findings show that a worryingly large fraction of published conclusions, including well cited examples, simply don't replicate. The observation has led to urgent and sweeping measures, guided in part by a series of DARPA-funded projects, to improve how experiments are conducted, analyzed, and published. Included among them is Dreber's illuminating work.

Gelman and Crane both discussed election forecasting, including methods based on fundamental models, polls, markets, and their combinations. Gelman highlighted that election outcomes are more predictable than political polls that suffer from lack of trust and differential non-response. He noted that forecasters gain little from being right, gain attention from dramatic swings, and lose a lot from being wrong, affecting their incentives. Crane directly compared the accuracy of election models published by The Economist (co-developed by Gelman) and FiveThirtyEight and prediction markets run by PredictIt.org, using a novel market-based scoring method. He found that all the forecasts are well calibrated, especially FiveThirtyEight's. Ignoring PredictIt's fees and commissions, FiveThirtyEight and The Economist outperformed PredictIt. However, factoring in fees and commissions, PredictIt does better.

The contributed talks also sparked lively discussions. Several addressed forecasting mechanisms—including prediction markets, Good Judgment's superforecasting teams, wagering mechanisms, and peer-prediction systems—that have risen in parallel to advances in machine learning and other data-driven forecasting approaches. Workshop presenters explored methods to improve forecasts by forming teams of different types of forecasters; they compared models versus markets (and methods that combine both) for predicting uncertain outcomes; they examined peer prediction and other methods for eliciting truthful forecasts with no ground truth; and they looked at applications that included forecasting elections and clinical trials. A common theme emerged that teams of human forecasters still outperform machine-learning based forecasts in a variety of domains, including healthcare.

Adding to the talks and daily poster sessions was a panel made up of founders of startups in the rapidly evolving business of forecasting. The panel featured Pavel Atanasov, co-founder of pytho; Andreas Katsouris, Senior Vice President of PredictIt; Kelly Littlepage, co-founder of OneChronos; and Emile Servan-Schreiber, co-founder of Hypermind. The panelists gave insights into how and why their companies were formed and discussed opportunities for researchers, including open challenges and available data sets. The panel was short and barely scratched the surface of industry developments in forecasting.

Each day concluded with a one-hour poster session featuring both contributed posters and several speakers from earlier in the day either presenting related posters or simply available to answer questions and discuss the work presented. The poster sessions were well attended and

highly engaging, often lasting well past their allotted one hour, as technical and informal discussions continued. The poster sessions were truly a highlight of the event, and they were where the online platform shined.

About the online format. The workshop was hosted by Virtual Chair, which created a virtual venue on the Gather platform and ran the event. The venue had a simple linear design—a lobby

sandwiched between a lecture room to the north and a poster room to the south. During the event, attendees navigated avatars around a space resembling a 1980s video game. They could chat with people they passed or sit together at tables for a private conversation. When it was time for presentations to begin, attendees navigated to the lecture room, sat in a chair, and started Zoom to hear the presentations. During poster sessions, groups clustered around a poster to listen to the presenter and discuss as a group. Virtual Chair poster sessions in some ways work even better than physical poster sessions because everyone can see and hear. Throughout the event, virtual encounters are analogous to those at a physical conference, and if anyone has difficulty, they can visit the Help Desk staffed by Virtual Chair.

The workshop organizers—Raf Frongillo and Bo Waggoner (both of the University of Colorado) and David Pennock (DIMACS)—weighed a number of options for structuring the online program. They converged on a schedule of three hours per day (to avoid zoom fatigue) beginning at 10:00 AM EDT. The format seemed to hit a sweet spot for online meetings: two hours of talks and one of poster-centered discussions during waking hours from California to Shanghai. This format got strong reviews from participants, with one tweeting, "*The best remote experience I have had in a while*."

Nonetheless, there are opportunities lost online. Both Gelman and Crane considered the question of whether you should bet on your own election forecasts and took opposite viewpoints. Gelman felt that you should not, while Crane felt that you must. This philosophical difference left us wondering about the rich lunchtime discussion that might have ensued if the workshop had been held in person. Given the success of the Workshop on Forecasting, we may not need to wonder for too long. We are considering holding another event on the topic before the special focus ends, adding to the four workshops that are already in planning.

To receive information about these and other special focus activities, you can join the special focus mailing list.

Videos of workshop presentations are accessible from the workshop webpage and on YouTube.

Selected Links:

- Special Focus webpage: http://dimacs.rutgers.edu/programs/sf/mechanisms
- Workshop on Forecasting: http://dimacs.rutgers.edu/events/details?eID=1531