MITRE

SERIES
Number 17

INTELLIGENCE AFTER NEXT

IMPROVING INTELLIGENCE ANALYSIS - ADOPTING THE WISDOM OF THE CROWD

by Stephen Stratman, Shinal Patel, and Will Scannell

Judgmental Forecasting – Needs and Opportunities

One of the core functions of the U.S. Intelligence Community (IC) is to assess the likelihood of occurrence – and, by extension, the strategic implications – of uncertain future events (e.g., foreign election outcomes, acts of political violence, shifts in foreign policy and strategic priority, etc.). Problems of interest to the IC involve what is conventionally known as judgmental forecasting, [1] in which subjective opinion plays a critical role and for which the application of strict causal model-based or algorithmic (to include statistical) forecasting approaches is often of limited utility.

A growing body of evidence suggests, however, that large crowd analytic forecasting platforms offer an efficient and practical mechanism for bringing transparency and rigor to geopolitical forecasting and related forms of human analytic judgment. The theory underpinning the notion of crowd wisdom forecasting is that aggregating and summarizing the analytic forecasts of multiple, distributed individuals can lead to consensus estimates that are more accurate than those ventured by most individuals in the group.

The prospect of improved forecast accuracy within the IC is an obvious potential benefit of crowd analytic forecasting platforms, but there are additional benefits in an organizational setting, which together make the case for adoption even more compelling. Crowd forecasting platforms:

- Promote information sharing and interaction across traditionally siloed groups.
- Encourage the expression of uncertainty using precise numerical probabilities while simultaneously logging rich data about forecast trends and historical forecast accuracy.
- Are dynamically updated in near real time, making them highly complementary to static forecasts rendered in traditional written analytic reports.

Together, these properties promote realism, accountability, and institutional learning. In today's complex geo-political environment, the IC should consider adopting crowd forecasting as a way to promote information sharing and to encourage the systematic tracking of the accuracy of its analytic judgments over time.[2]

CROWD-BASED ANALYTIC FORECASTING PLATFORMS OFFER AN EFFICIENT AND PRACTICAL MECHANISM FOR BRINGING TRANSPARENCY AND RIGOR TO GEOPOLITICAL FORECASTING AND RELATED FORMS OF HUMAN ANALYTIC JUDGMENT

Can Traditional Subjective Analysis be Improved?

Encouraging the generation of well-calibrated and accurate judgmental forecasts is a persistent challenge for the IC, as is the objective measurement of forecasting accuracy.[3] Having well-calibrated forecasts means having forecasts that, over time, align well with observed relative frequencies of occurrence for events and outcomes of interest.

- For instance, observing rain on 70 out of the 100 days for which a forecast calls for a 70% chance of rain is an example of excellent long-run forecast calibration. Accuracy, in the forecasting context, is a relative concept that implies being as close as possible, in one's estimates, to an actual (true) observed outcome or value. A 70% forecast of rain would be considered more accurate than a 50% forecast of rain, in the event that rain was, indeed, observed. Formally and mathematically, there are many ways of objectively scoring forecasts for accuracy and calibration.[4]

The analytic ideal of accurate and well-calibrated forecasts notwithstanding, social science research has demonstrated that humans are imperfect judging machines, susceptible to bias that can lead subjective judgment to deviate from the objective ideal, sometimes in systematic and foreseeable ways.[5] Perhaps counter-intuitively, this tendency to succumb to systematic bias in judgment has been shown to occur even when judgments are rendered by domain experts.[6]

What's more, quantifying forecasting success is difficult in real-world contexts for at least two reasons. First, forecasts often are expressed using imprecise and ambiguous language rather than precise numerical estimates, leaving room for variable interpretation of intent. Second, even dispassionate and well-meaning observers routinely disagree about whether and when a forecasted outcome of interest actually occurred, particularly when forecasts are ventured on complex and nuanced geopolitical topics. Together, these conditions make it hard to reliably "keep score" of forecasting success and easy for forecasters to retrospectively reframe failures as successes and big misses as near misses.

There are exceptions, of course. Weather forecasting and sports betting are two real-world domains in which forecast accuracy and calibration are routinely calculated and tracked. What these domains have in common are local conventions of:

- Expressing forecasts as numerical probabilities
- Focusing on outcomes (e.g., wins vs. losses or rain event vs. non-rain event, etc.) that are relatively unambiguous and uncontroversial
- Centrally aggregating historical forecast and outcome data

In the IC, as in most domains where judgmental forecasting is routinely employed, forecasts are expressed using descriptive language rather than precise numerical estimates (e.g., X is 'likely' or 'unlikely'), leaving room for disagreement in interpretation. In addition, operational definitions for and status of focal outcomes can be vague and elusive, even with – and sometimes particularly with – the benefit of hindsight. Finally, there is no systematically ingrained convention of centrally aggregating historical forecasts and their associated outcomes to facilitate the systematic tracking of success.[7]

Applied Crowd Wisdom for Decision Support

The idea of leveraging crowd wisdom to solve real-world decision support problems is not new. It has a long history of application, especially if one views free markets as a variation on the theme of aggregate crowd judgment. In this instance, consensus beliefs about the value of a security are encoded in the market prices that arise from the equilibrium established between buying and selling pressures.[8]

The ability of markets to function as efficient, dynamic aggregators of distributed information has led to various efforts at using so-called information markets – also known as prediction markets, futures markets, betting markets, and decision markets – as a means of efficiently aggregating and summarizing distributed information about the likelihood of uncertain events.

- While implementation details can vary considerably, the basic premise of information markets is that participants buy shares in possible outcomes of an event such as an election outcome, using either real or virtual money. Prices in information markets are typically constrained such that they trade within a fixed range from 0 to 100%. Once the results of an event are known, the market pays out 100% per share for realized outcomes, but nothing otherwise.
- This model incentivizes market participants to buy shares of outcomes that they believe are more likely to occur and to sell (or short sell) outcomes they believe are less likely to occur, with the convenient side effect that fluctuating market prices can be interpreted as consensus market assessments of the probability of the outcome in question coming to pass.[9]

Information markets have been fielded in numerous real-world contexts. The Iowa Electronic Markets (IEM) were an early example. Launched by a group of economists at the University of Iowa in 1988, the IEM allows academic traders to use real money to buy and sell contracts on election outcomes. Over time, the IEM has proven to be highly competitive with – and frequently more accurate than – traditional opinion polls in forecasting candidates' shares of the final vote in presidential elections, without the need to systematically assemble representative groups of likely voters.[10] Information markets have subsequently been fielded in numerous government and corporate settings, as a means of aggregating information in a decision support context.[11]

Although market-based aggregation models have drawn much attention, a market-based incentive structure is not required to take advantage of the wisdom of crowds in a judgmental forecasting setting.

- Research informed, in part, by a series of large-scale forecasting competitions run by the U.S. Intelligence Advanced Research Projects Activity (IARPA)[12] proved that weighted aggregations of directly elicited probabilistic estimates compete favorably (from an accuracy standpoint) with market-based aggregation models when weighting schemes take into account individual participant historical accuracy and the temporal currency of the aggregated estimates.[13] This point warrants emphasis because when leveraging crowd wisdom for organizational decision support, there is some risk of encountering objections to the concept of forecasters betting on, and in some sense "profiting" from (even if only with virtual currency), geopolitical events.[14]
- Some of the insights from IARPA's large-scale forecasting competitions are distilled in Philip Tetlock and Dan Gardner's 2015 book, *Superforecasting: The Art and Science of Prediction*.[15] Tetlock led one of several forecasting teams who participated in IARPA's Aggregative Contingent Estimation (ACE) Program.[16] Tetlock and Gardner argue that when aggregating crowd judgments, aggregate accuracy can be improved by upweighting the inputs of forecasters who demonstrate consistently higher-than-average forecasting accuracy. They also note that accuracy improvements can, in some cases, be gained through algorithmic manipulation (e.g., extremization) of consensus estimates and by providing forecasters with basic training on best practices for reducing cognitive biases. Finally, they advocate for teaming forecasters to foster interactions and discussions, which can have a positive impact on both engagement and accuracy.

INSTITUTIONAL CROWD WISDOM SOFTWARE SOLUTIONS ARE READILY AVAILABLE IN THE MARKETPLACE, THE EVIDENCE OF THEIR UTILITY AND VALUE IS PERSUASIVE, AND THE COST OF THEIR ADOPTION NON-PROHIBITIVE

Bringing the "Crowd" to Bear in a Structured Forecasting Environment

As stated earlier, aggregating and summarizing the opinions of multiple, distributed individuals under the right circumstances can lead to consensus estimates that are more accurate than those ventured by most individuals in the group. Crowd-based analytic forecasting tends to work well (i.e., can lead to improved forecasting accuracy) when:

- Individuals whose estimates are being aggregated operate largely independently of one another
- Individuals have access to and can collectively bring to bear diverse sources of relevant information
- Opinions of the group are not centrally shaped or directed
- There is a mechanism in place through which individual estimates can be reliably aggregated and summarized
- Forecasts can be updated in the face of new information

While discussions of the benefits of crowd wisdom in a judgmental analytic forecasting context emphasize the prospect of improved accuracy, it is not only accuracy that merits attention.

Systematically surveying the judgments of members of a group via a centralized crowd forecasting platform also:

- Facilitates transparency and information sharing
- Avoids ambiguity in the expression of uncertainty by expressing estimates as numerical probabilities
- Allows for near-real-time dynamic updating and tracking of evolving trends
- Naturally enables and encourages the explicit calculation and tracking of both individual and aggregate historical accuracy and calibration – arguably, prerequisites for organizational accountability, learning, and improvement.

Open Research Questions and Issues

There are issues related to crowd analytic forecasting that any organization should consider when determining its use as an institutional judgment and decision toolkit.

- Before one can systematically distill consensus judgments from a group, one must assemble and mobilize that group. In our experiences in fielding and evaluating crowd wisdom platforms,[17] participant recruitment and retention has consistently emerged as the single greatest challenge to the successful deployment of crowd wisdom forecasting. This challenge is particularly acute in institutional settings where organizational support is limited or ambivalent.
- Sound theoretical and empirical evidence of utility does not guarantee that the outputs of crowd wisdom platforms and approaches will be well received, or even well tolerated, by the decision makers who, in principle, stand to benefit most from their properties. Such reaction is a matter of institutional culture and practice, not a technical or theoretical constraint.
- The types of forecasting problems that are fielded in crowd aggregation platforms can tend toward narrow and artificial formulations that prioritize unambiguous and near-term resolution. This can lead to a tension between forecasting problems that are formally convenient and those that are operationally relevant. Finding ways to address this tension in crowd forecasting scenarios (i.e., to find ways to tackle complex and operationally relevant forecasting problems while still providing participants with adequate feedback and incentives) remains an open challenge.
- In short, while the general case for crowd wisdom rests on a firm theoretical and empirical footing, some of the implementation details are still being actively debated by the research community. Recurring areas of active research include:
  - Determining how to select the best (i.e., most net advantageous, from an accuracy standpoint) individual participants for crowd aggregation systems
  - Determining the optimal means of aggregating, weighting, and summarizing individual inputs
  - Exploring mechanisms for addressing abstract or open-ended forecasting problems that can't be unambiguously resolved in the near-term
  - Overcoming institutional barriers to adoption

Moving Forward

We view the primary barriers to crowd wisdom platform adoption in the IC as being neither technical nor theoretical, but rather cultural in nature (i.e., related to institutional norms and traditions). Institutional crowd wisdom software solutions are readily available in the marketplace, the evidence of their utility and value is persuasive, and the cost of their adoption non-prohibitive. We therefore endorse the incorporation of crowd wisdom forecasting platforms into the IC's overall judgment and decision-making toolkit.

References

1. See, e.g., M. Lawrence, P. Goodwin, M. O'Connor, and D. Önkal, "Judgmental forecasting: A review of progress over the last 25years," International Journal of Forecasting, vol. 22, no. 3, pp. 493–518, Jan. 2006 and W. E. Walker et al., "Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support," Integrated assessment, vol. 4, no. 1, pp. 5–17, 2003

2. For more in-depth discussion of issues and options, see Perry World House, Keeping Score: A New Approach to Geopolitical Forecasting, 2021, available online at https://global.upenn.edu/perryworldhouse/keeping-score-new-approach-geopolitical-forecasting.

3. Office of the Director of National Intelligence, Intelligence Community Directive 203, Analytic Tradecraft, 2015, https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf

4. See, e.g., B. Casati et al., "Forecast verification: current status and future directions," Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, vol. 15, no. 1, pp. 3–18, 2008.

5. See, e.g., T. Gilovich, D. W. Griffin, and D. Kahneman, Eds., Heuristics and biases: the psychology of intuitive judgment. Cambridge, U.K. ; New York: Cambridge University Press, 2002 and https://www.simplypsychology.org/cognitive-bias.html

6. P. E. Tetlock, Expert political judgment: how good is it? How can we know? Princeton, N.J: Princeton University Press, 2005.

7. B. J. Stastny and P. Lehner, "Comparative evaluation of the forecast accuracy of analysis reports and a prediction market," Judgment and Decision Making, vol. 13, no. 2, pp. 202–211, 2018.

8. See, e.g., E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," The Journal of Finance, vol. 25, no. 2, pp. 383–417, 1970.

9. See, e.g., R. Hanson, "Decision markets for policy advice," in Promoting the General Welfare: New Perspectives on Government Performance, 2006, pp. 151–173.

10. See https://iem.uiowa.edu/iem J. See, e.g., Berg, R. Forsythe, and T. Rietz, "What Makes Markets Predict Well? Evidence from the Iowa Electronic Markets," in Understanding Strategic Interaction: Essays in Honor of Reinhard Selten, W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, and E. van Damme, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 444–463.

11. See https://www.lancaster.ac.uk/lums/news/how-prediction-markets-could-improve-climate-risk-policies-and-investment-decisions

12. See https://www.iarpa.gov/research-programs/ace and https://www.iarpa.gov/research-programs/forest

13. See, e.g., P. Atanasov et al., "Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls," Management Science, vol. 63, no. 3, pp. 691–706, 2017.

14. See, e.g., R. Hanson, "The Policy Analysis Market (A Thwarted Experiment in the Use of Prediction Markets for Public Policy)," Innovations: Technology, Governance, Globalization, vol. 2, no. 3, pp. 73–88, 2007.

15. P. E. Tetlock and D. Gardner, Superforecasting: the art and science of prediction, First edition. New York: Crown Publishers, 2015.

16. https://goodjudgment.com/resources/the-superforecasters-track-record/the-first-championship-season/

17. See, e.g., Stastny & Lehner (2018, cited above) and E.C. Merkle and R. Hartman "Weighted Brier score decompositions for topically heterogenous forecasting tournaments." Judgment and Decision Making, vol. 13, no. 2, pp. 185-201, 2018.

.

Authors

Stephen Stratman is a Lead Social Scientist in MITRE's Social and Behavioral Sciences Department. His work focuses on helping organizations to select, adopt, and make optimal use of evidence-based methods and tools for judgment and decision support.

Shinal Patel is a Senior Systems Engineer at MITRE and has operated on a wide variety of projects ranging from intelligence analysis to emerging biomedical research. Formerly, Shinal was a chemistry professional with experience in driving public health transformation in the clinical industry and for the Commonwealth of Virginia.

Will Scannell is the Department Manager of the Strategy and Policy Department within MITRE's National Security Sector. He has led development of operational concepts and implementation plans to meet changing strategic, operational, and tactical priorities and presenting strategies to inform policy deliberations at the most senior levels of the government.

Intelligence After Next

MITRE strives to stimulate thought, dialogue, and action for national security leaders developing the plans, policy, and programs to guide the nation. This series of original papers is focused on the issues, policies, capabilities, and concerns of the Intelligence Community's workforce as it prepares for the future. Our intent is to share our unique insights and perspectives surrounding a significant national security concern, a persistent or emerging threat, or to detail the integrated solutions and enabling technologies needed to ensure the success of the Intelligence Community.

About MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

MITRE