Aggregative Contingent Estimation (ACE)

Jason Matheny

IARPA / Office of Incisive Analysis

19 May 2010

Disclaimer

- This presentation is provided solely for information and planning purposes.

- The Proposers' Day Conference does not constitute a formal solicitation for proposals or proposal abstracts.

- Nothing said at the Proposers' Day changes the requirements set forth in a BAA.

- BAA supersedes anything presented or said at the Proposers' Day by IARPA.

Proposers' Day Goals

- Familiarize participants with IARPA's interest in forecasting and human judgment. Please ask questions and provide feedback; this is your chance to alter the course of events.

- Foster discussion of synergistic capabilities among potential program participants, AKA teaming. Take a chance: someone might have a missing piece of your puzzle.

Schedule

- Once BAA is released, questions can only be answered in writing on the program website.
- Full proposals will be due ~45 days after BAA is published.

ACE Overview

- Goal: Develop and test tools to provide accurate, timely, and continuous probabilistic forecasts and early warning of global events, by aggregating the judgments of many widely-dispersed analysts

- Key technical challenges: Develop and test methods that elicit forecasts from users, aggregate these forecasts using data about the users and their judgments, and communicate these forecasts to a wide variety of other users

Background

- The intelligence community (IC) is often asked to provide decision-makers with forecasts of global events ("estimates") and consequences of possible actions ("contingent estimates")

- Examples: National Intelligence Estimates (NIEs) (http://www.dni.gov/nic/NIC_specialproducts.html)

Background

- IC forecasts are typically qualitative
- IC depends on human judgments for most forecasts, typically made by individual analysts or small groups of analysts
- Disagreement among analysts is generally not quantified

Current Research

- Individuals and small groups are consistently less accurate than methods that mathematically aggregate many independent forecasts
- Unweighted opinion pools and conventional prediction markets are aggregation methods that have proven difficult to beat (according to the published literature); but in some cases accuracy has been increased by weighting judgments on cognitive style, past performance, peer-prediction, or risk aversion
- No study has optimized aggregation methods by exploiting large sets of detailed data about the forecasters and their judgments
- There has been virtually no evaluation of methods for conditional forecasts

Program Structure

Period	Length (years)	Goals
Base	1	Develop elicitation, aggregation, and communication methods; identify correlates of forecasting accuracy; build tools v1
Option 1	1	Refine and test technical approaches; open source testing begins on unconditional forecasts; build tools v2
Option 2	1	Refine and test technical approaches; open source testing continues on unconditional forecasts, begins on conditional forecasts; build tools v3
Option 3	1	Refine and test technical approaches; open source testing continues; build tools v4
Option 4	1	Refine and test technical approaches on IC systems and problems; build tools v5

Elicitation Methods

Develop elicitation methods to anonymously elicit judgments:

- Whether an event will or will not occur
- When an event will occur
- The magnitude of an event
- All of the above, conditioned on another set of events or actions
- The confidence or likelihood a user assigns to his or her judgment
- The user's rationale for his or her judgment, as well as links to background information or evidence, expressed in no more than a couple of lines of text
- The user's updated judgments and rationale

Elicitation Methods

Challenges:

- Some users will be unaccustomed to providing probabilistic judgments
- There has been virtually no research on methods to elicit conditional forecasts
- Elicitation should require a minimum of time and effort from users
- Training time for users will be limited, and all training must be delivered within the software
- Rewards for participation, accuracy, and reasoning must be non-monetary and be internal to the software or community of users

Aggregation Methods

Develop aggregation methods that:

- Aggregate forecasts from multiple individuals into a single forecast

- Generate forecasts that are substantially more accurate than both the unweighted opinion pool and conventional prediction market forecasts

- Are robust across diverse populations of users and forecasting problems

- Use data that the IC has, or could have, about participating users

- Continuously improve accuracy over time, by employing machine learning or other technical approaches

Aggregation Methods

Challenges:

- Users will be highly diverse in academic and professional education, experience, and other characteristics

- No information on user performance prior to participating in this program will be available

- Characteristics of users on any given forecast will not be known a priori, because participation is voluntary

- Forecasting problems are highly diverse, including: topic area, probability, conditionality, and verifiability

Communication Methods

Develop communication methods that effectively and accurately communicate forecasting results, including (but not limited to):

- The aggregate forecast
- The distribution of individual judgments
- Trends in judgments over time
- Measures of disagreement among users
- Segmentation and cluster analysis of judgments
- Analysis linking rationale text to clusters and outliers
- Where possible, visualizations of the above data, that are intuitively clear to non-quantitative users

Communication Methods

Challenges:

- Most users will be unaccustomed to data of the types listed above

- Some users will have little quantitative training

- Training time for users will be limited, and all training must be delivered in-software

- Users' understanding of probabilities can vary depending on the format in which the probabilities are communicated, the context in which the probabilities are presented, and users' numeracy

- In some cases, users may interpret forecasts where they have existing opinions

Team Composition

- Given the combination of technical challenges, we anticipate teams will possess expertise in:
  - Psychology, economics, or decision science
  - Statistics or econometrics
  - Computer science
  - Data visualization or information design
  - Software rapid prototype development

Out of Scope

IARPA is not interested in approaches that:

- – Rely on monetary incentives
- – Encourage strategic misreporting of judgments
- – Require more than two minutes of time per elicitation per user

Testing: Elicitation & Aggregation

- Before month 12, performers will choose their own forecasting problems
- After month 12, sets of forecasting problems will be continuously provided to performers by the Program Manager
- Initially, problem sets will only include unconditional events; beginning at month 24, problem sets will include both unconditional and conditional events
  - These may include real-world political, military, economic, science and technology, social, demographic, and public health events, on 1 to 12+ month time horizons
- Performers' software will elicit and aggregate forecasts
- Performers will continuously provide IARPA with aggregate forecasts for all of the assigned events
- Forecast accuracy will be continuously evaluated as events are observed; methods should allow continuous updating, so mean accuracy will be calculated for each forecast from event posting to expiration

Testing: Elicitation & Aggregation

- Metrics:
  - Mean quadratic score (MQS) against observed discrete events
  - Mean square percentage error (MSPE) against observed continuous events
- Performance will be assessed against program milestones at 12, 24, 36, 48 months
- Milestones:
  - Percent difference over an unweighted linear opinion pool (ULinOP)

Testing: Elicitation & Aggregation

Year:	1	2	3	4	5
Study population	Recruited by performer				IC analysts
Forecasting problems	Open source, chosen by performer	Open source, unconditional, chosen by IARPA	Open source, unconditional & conditional, chosen by IARPA		Open source & classified, unconditional & conditional, chosen by IARPA
Metrics	MQS, MSPE				
Year-end milestone (difference over a ULinOP)	20%	35%	50%	65%	65+%

Testing: Communication

- Metrics:
  - Accuracy of users' interpretations of probabilities
  - Accuracy of users' ordering of events by likelihood
  - Correct segmentation and cluster analysis by users
  - Users' assessments of tools' clarity and utility
- Performers will be assessed against one another, and against program milestones at 12, 24, 36 months
- Milestones:
  - At month 12, performers' methods will be self-tested
  - At months 24 and 36, performers' methods will be tested on representative users by a government test & evaluation team

Testing: Communication

Year:	1	2	3
Study population and protocol	Chosen by performer	Chosen by IARPA	Chosen by IARPA
Probability / event types	Unconditional	Unconditional	Unconditional & conditional
Percentage of users who correctly:			
Interpret probabilities	85	95	95
Order events by likelihood	85	95	95
Perform segmentation & cluster analysis	80	90	90

Eligibility Information

- Other Government Agencies, Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers (UARCs), and any other similar type of organization that has a special relationship with the Government, that gives them access to privileged and/or proprietary information or access to Government equipment or real property, are not eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities.

- Non-US organizations and individuals may be able to participate.

  – Must comply with Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc., as appropriate

  – Specific guidance for non-US participation will be provided in the BAA

Proposal Guidance

- Your proposal should include a full discussion of the technical approach that will be used to meet the program goals.

- Programmatic issues that should be addressed in the proposal:

  - Your team's current technical capabilities

  - A teaming plan along with the roles and responsibilities of each member of the research team.

  - End-of-period milestones are set, but it is expected that other intermediate milestones that are on the critical path of the proposed approach will be offered.

  - A schedule of all milestones, including a clearly charted description of the various risk mitigation strategies that will be undertaken to achieve program goals.

Proposal Guidance (contd.)

- Download ALL materials posted to the FedBizOpps announcement (BAA, instructions, templates, etc.)
- Periodically check for amendments and other information that may be posted prior to the proposal due date
- Read FAQs posted to the web site
- Ensure submission requirements are followed:
  - Deadlines
  - Do not exceed page limits
  - Use all provided templates (see Appendix)
  - Include all required responses (OCI paperwork, Academic Acknowledgement letters, etc.)
  - No unnecessarily elaborate brochures or marketing material
  - Failure to follow the submission procedures may result in the submission not being evaluated

Proposal Evaluation Criteria

- Overall Scientific and Technical Merit

- Effectiveness of Proposed Work Plan

- Relevance to IARPA Mission and Program Goals

- Relevant Experience and Expertise

- Cost Realism

Teaming

- Because of the many challenges presented by this program, both depth and diversity will benefit your team
  - Throughput : consider all that you will need to do, all the ideas you will need to test
    - Make sure you have enough people and expertise to do the job
    - Make sure you have sufficient resources to follow the critical path while still exploring alternatives
  - Completeness: teams should not lack any capability necessary for success, e.g. should not rely on enabling technology to be developed elsewhere
  - Tightly-knit teams
    - Clear, strong, management, single point of contact
    - No loose confederations
    - Each team member should be contributing significantly to the program goals. Explain why each member is important. If you didn't have them, what wouldn't get done?
    - No teaming for teaming's sake
- Remember, you may be very accomplished, but can you do it all?

Additional Information

- dni-iarpa-baa-10-05@ugov.gov for additional questions

- ACE BAA will be posted on the FedBizOpps website (www.fedbizopps.gov)

- Q&As will appear after the BAA

Questions?