MERCURY PROGRAM

Motivation

- This publication constitutes a Broad Agency Announcement (BAA) and sets forth research areas of interest in the area of event forecasting using classified foreign Signals Intelligence (SIGINT) data.
- The Mercury Program seeks to develop methods for continuous, automated analysis of foreign SIGINT in order to **anticipate and/or detect terrorist activity, military actions, political crises, and disease outbreaks.**

Motivation

- Past research has shown that publicly available data can accurately forecast societal events such as civil unrest and disease outbreaks.

    - For example, DOD's Integrated Crisis Early Warning System (ICEWS) and IARPA's Open Source Indicators (OSI) have developed methods to forecast societal events using structural data, news feeds, blogs, web search queries, and other publicly available data.

- However, in many cases, relevant data have **significant lag times**, **lack accuracy, or are classified**. There has been little research to examine whether classified data from foreign Signals Intelligence (SIGINT) can be used to forecast events with **high accuracy and lead-time**.

Expectation

- Performers must address the following technical challenges in order to meet the objectives of Mercury

    - Development of methods to **automatically sample SIGINT data to detect group-level behavior changes** in anticipation of, and response to, events of interest.

    - Development of SIGINT data **extraction techniques that focus on volume, rather than depth**, by identifying shallow features of SIGINT data (keywords, geotags, timestamps, etc.) that correlate with group behavior.

    - Development of **multivariate time-series models** robust to non-stationary, heteroscedastic, and heterogeneous data to **reveal patterns that precede events**.

    - Development of innovative **uses of statistical methods** to fuse combinations of time series for **generating probabilistic warnings of events**.

    - Development of capabilities that account for **geographical variations in dialect**, **communication preferences, rate, technology adoption, and political openness.**

Expectation

- Mercury is **not focused on advanced natural language processing (NLP)**, or **on the development of other advanced tools for analyzing a single signal**. As applicable, known NLP approaches should be sufficient for extracting features of interest and **Mercury methods should focus on aggregating multiple, noisy signals indicative of significant events.**

Evaluation

- Lead Time = the number of days between the time the performer delivered the warning to IARPA and the GSR timestamp for the event (not necessarily the time of the event). This number will be calculated by rounding the difference between the timestamps to the nearest day, so that a warning needs to be sent at least 12 hours before the GSR timestamp to have a lead time of 1 day. A warning sent less than 12 hours before the GSR timestamp will have a lead time equal to zero.
- Probability Score (quadratic score) = $1-(o-p)^2$ where p is the probability assigned by the performer to the event, and o is ground truth: 1 if the event occurred, 0 if the event didn't occur within 30 days of the expected date.
- Precision = the number of warnings matched to GSR events divided by the number of warnings sent.
- Recall = the number of GSR events matched to warnings divided by the number of GSR events.
- Warning quality measures how well the event descriptions in the performer's warning correspond to the GSR event details. Each event description will consist of four fields: the population involved in the event, the type of event, the date of the event, and the location of the event. For each of these fields, IARPA will calculate the match between the performer's warning and the GSR. These component scores will be then combined into an overall quality score on a scale from 0.0 to 4.0, where the component weights will vary according to the event type.

Milestones

Table 2: Mercury Program Milestones

Metric	Month 15	Month 27	Month 35
Mean Lead Time	2 days	5 days	7 days
Mean Probability Score	0.75	0.80	0.85
Mean Quality Score	3.0	3.25	3.5
Recall	0.60	0.70	0.80
Precision	0.60	0.70	0.80