

Assessing the quality of Amazon turker and automated machine forecasts in the Hybrid Forecasting Competition

Andreas Beger and Michael D. Ward

2018-11-07

Abstract

HFC

Funding acknowledgement: This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2017-17071900005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

The Hybrid Forecasting Competition

Background on HFC

SAGE approach

Cover the basics: there is a platform that can show a chart and machine forecast; how are the machine forecasts generated.

Research design for RCT-A

Go over the conditions, and the complications that happened: turkers only in two conditions, and the data and forecaster code changes over time.

Two key questions we want to examine:

1. How accurate were Turker forecasts compared to volunteer forecasts?
2. What was the impact of the machine forecasts on human accuracy?

Design, data, and method

Key leverage point is that only some questions had a machine forecast.

Go over what the data for this are, how accuracy was assessed

Results

Basic findings

Machine forecasts were not more accurate than human forecasts.

Turker performance compared to volunteer forecasters

Machine forecast impact

Discussion/conclusion

Based on the quick check I did during the site visit, probably something like this:

- the machine forecasts were not accurate enough, although sans data and code issues they may have been
- Turkers don't forecast as well as volunteers
- seeing the machine forecasts had a slight detrimental impact on accuracy, although maybe this changed as the forecasts became more accurate later in the RCT
- seeing the charts improved performance, but I don't think it is clear why. Seeing the chart for instances where the data are hard to acquire (e.g. some data transformation are needed) did not boost accuracy more than in other series.