Forecasting Annotation Guidelines

1 Introduction

This data comes from the Geopolitical Forecasting Competition (GFC2) where The Intelligence Advanced Research Projects Activity (IARPA) posted questions about future geopolitical events, such as "Will France's President Emmanuel Macron experience a significant leadership disruption between 3 April 2019 and 29 November 2019?". 537 individuals took part in this competition, answering a subset of the questions, assigning a probability to each possible outcome, and a rationale explaining their reasoning behind it. The participants ("forecasters"), answered questions set in the future, having no way of finding out the correct answer, making their best guess and detailing their reasoning. The forecasters could answer each question as many times as they wished, meaning part of the forecasts have updates to their prediction, which the forecaster has made over time.

This data and as well as data from similar competitions has provided very interesting datasets which have been used for research into improving probability judgements of real world events and, among other things, has found that this is a task that some people are very good at. These individuals have been dubbed "superforecasters", due to their ability to consistently make forecasts which are more accurate than the general public's or those made by experts in the area. [1]

Now that we know all of the outcomes of the geopolitical events the forecasters were making judgements on, we can explore if there are any patterns in the way people reason and argue and the likeliness to come to a more accurate prediction. We have chosen a random sample of the forecasts to be annotated using IAT, to explore any potential correlations between how people argue and their ability to make an accurate forecast. Unlike in previous annotation environments, we have automatically generated and provided part of the analysis, with the remainder to be completed by the annotator.

2 Question and Answers

Each part begins with a question posed by the Hybrid Forecasting Competition (HFC), including multiple choice answer options. This is followed by the

[1] If you're interested in more information about the superforecasting work, look into "The Good Judgement Project" and "Superforecasting" by Phillip E. Tetlock.

Forecaster, identified by unique id, providing a probability score for each of the options, and a rationale for their prediction. Example (1) shows this below.

(1) Part ID: 1999

a. HFC: Will France's President Emmanuel Macron experience a significant leadership disruption between 3 April 2019 and 29 November 2019?

b. HFC: Yes or No

c. c70e1394f75bf4b3c2c61082b5fb2a1d75026ab1: Yes, 0.05 probability

d. c70e1394f75bf4b3c2c61082b5fb2a1d75026ab1: No, 0.95 probability

e. c70e1394f75bf4b3c2c61082b5fb2a1d75026ab1: I believe with the yellow-vest protests put on by the working class who have been hit with low wages and heightened gasoline taxes are a sign of the people's abilities, but I do not believe this is something that will take Macron out of office. Macron has attempted to meet voters demands and is trying to rectify the situation by being more involved and less out of touch. Macron is hosting or involved in political discussions, debates and meetings to address the unrest in his country. I do think he's made plenty of mistakes that caused a lot of damage (albeit some being done by the protestors), but I believe he is regaining 'popularity' enough to maintain office.

Part (1-a) of the text will be referred to as the **question**.

Parts (1-b) of the text will be referred to as the **hypotheses**, each outlined option being a hypothesis with its own propositional content.

Parts (1-c) and (1-d) will be referred to as the **forecasts**, each forecast also has its own propositional content.

The remaining part (1-e) of the text is the **rationale**, where the forecaster explains their forecast.

2.1 Annotating the Question

The question and answer parts of the analysis are already annotated for you as shown in Figure 1. The first part of the question ((1-a)) *"Will France's President Emmanuel Macron experience a significant leadership disruption between 3 April 2019 and 29 November 2019?"* is annotated as Pure Questioning. The question is reconstructed by the annotator, *Will France's President Emmanuel Macron experience a significant leadership disruption between 3 April 2019 and 29 November 2019*, is changed to *France's President Emmanuel Macron will*

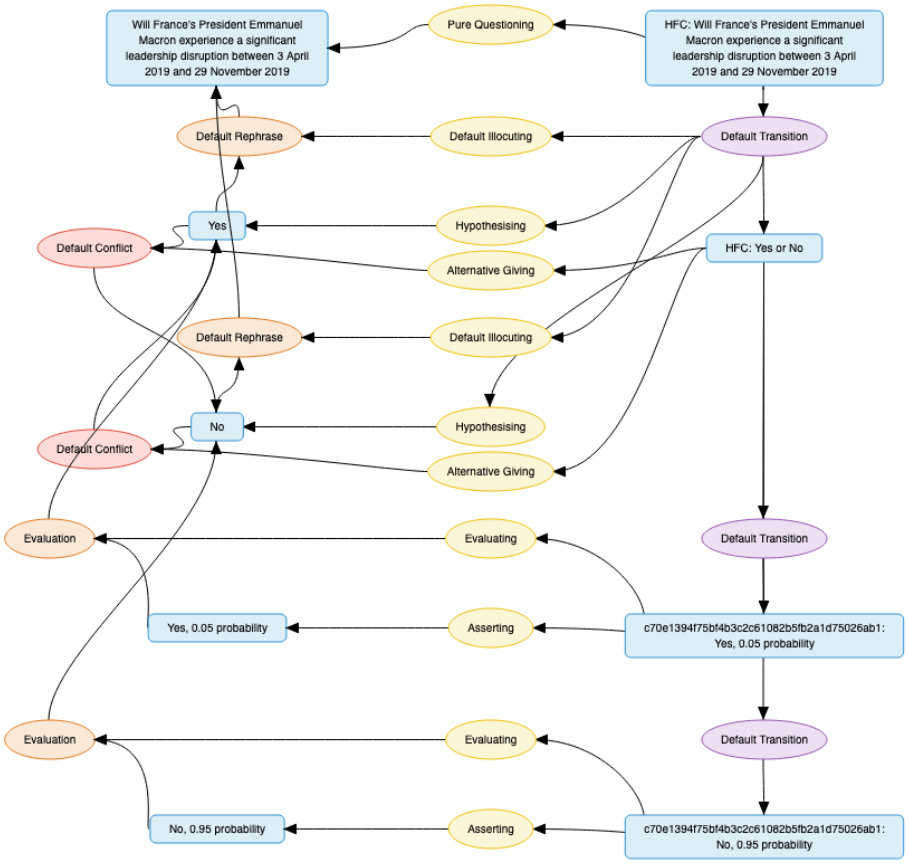

Figure 1: Part 1999 - Provided Annotation

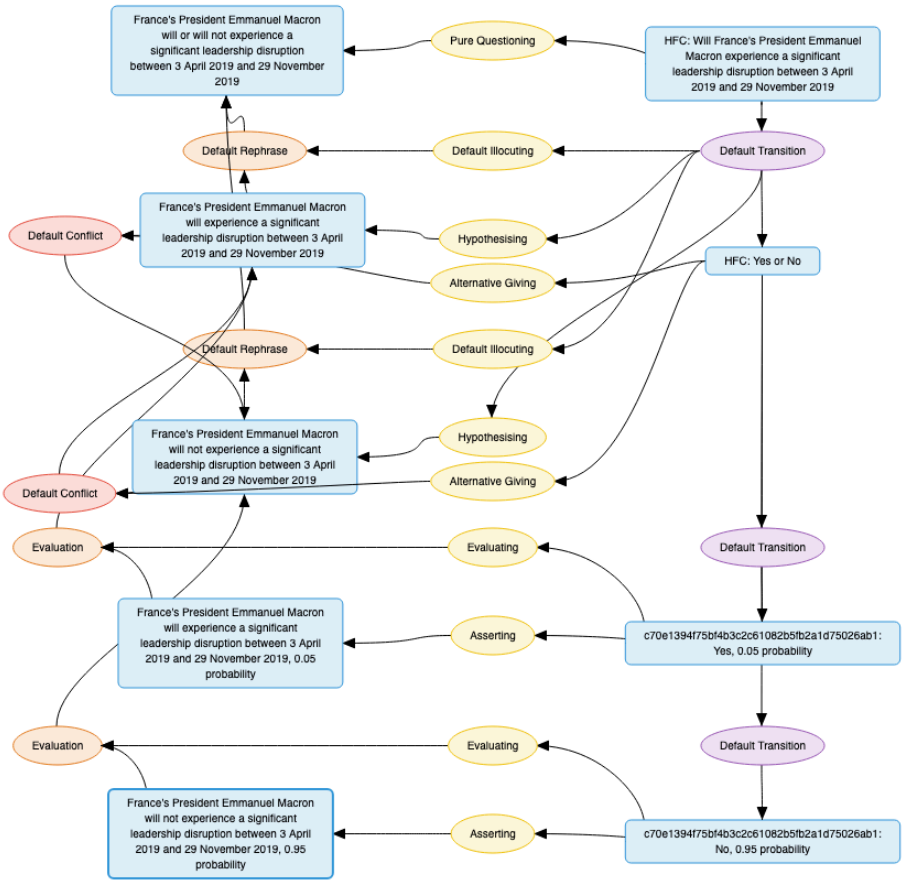

Figure 2: Part 1999 - Reconstruction

or will not experience a significant leadership disruption between 3 April 2019 and 29 November 2019. The following part of the question ((1-b)) lists the multiple choice options available as answers, in this case *Yes or No*, and this is annotated as Hypothesising, as the HFC are setting out all of the possible hypotheses. Each of the propositions also has an MA anchored in 'Default Illocuting' to model that it is a possible answer to the question. All of the options should be part of one locution (L-node), but be separate propositions (I-nodes). Each proposition needs to be reconstructed, as with usual IAT analysis. In this case, the annotator changes *Yes* to *France's President Emmanuel Macron will experience a significant leadership disruption between 3 April 2019 and 29 November 2019* as shown in Figure 2.

Each of the propositions from part (1-b) should have a CA (Conflict Node) to each of the other propositions coming from that same locution. Each proposition is a hypothesis which is an alternative to each other hypothesis, therefore each

have a CA, anchored with YA 'Alternative Giving'.

2.2 Annotating the Answer

Parts (1-c) and (1-d) are the forecasters giving their forecasts for each of the multiple-choice options, as a percentage. Each forecast should be annotated as Asserting and have a locution and proposition. Each of the forecasts needs to be reconstructed by the annotator, in this example *Yes, 0.05 probability* is reconstructed to *France's President Emmanuel Macron will experience a significant leadership disruption between 3 April 2019 and 29 November 2019, 0.05 probability.*

Each forecast made by the forecaster is an evaluation of the hypothesis, this is annotated as an MA (Rephrase) of type Evaluation. This is anchored from the locution, with YA Evaluating. The MA Evaluation relation comes from the proposition of the forecast, to the corresponding multiple choice hypothesis above, set out by the HFC. Any following forecasts made as updates, are annotated in the same way, with an MA Evaluation to the hypothesis at the top ((1-b)), it is not a rephrase of the previous prediction.

3 Rationale

After the question and answer section, which is always in the same format and already completed for you, apart from reconstruction, the next section is not done already and should be annotated according to the usual IAT guidelines (section (1-e)).

To see a full annotation of the example, including the rationale, open the map in ova (map ID: 27759). Note that some propositions will often support both forecasts, as is the case in this example. The node *"this is not something that will take Macron out of office"* supports both *''France's President Emmanuel Macron will experience a significant leadership disruption between 3 April 2019 and 29 November 2019, 0.05 probability"* and *"France's President Emmanuel Macron will not experience a significant leadership disruption between 3 April 2019 and 29 November 2019, 0.95 probability"*, as "this is not something that will take Macron out of office" both supports that there is a very low probability of experiencing disruption, and that there is a high probability of not experiencing disruption. Figure 3 shows this more closely.

More than one forecast/evaluation being supported or attacked is not always the case. Whether a proposition supports one, a subset, or all forecasts is dependant on the particular example. If there is not a large difference between the probabilities and it seems that the forecaster is arguing for one forecast over the other, then follow example 32262 in Figure 4, and have the proposition support only one forecast. Otherwise, if there is a large difference between the propositions and it seems like the forecaster is arguing for a point of view that is represented in both forecasts, follow the above example of part 1999 (Figure

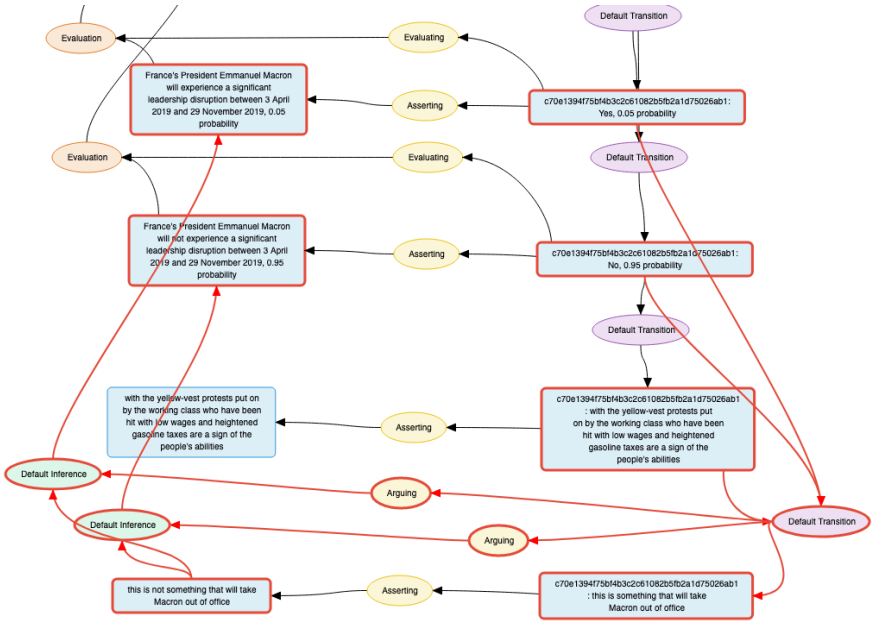

Figure 3: Part 1999 - Support for both forecasts

3) and have the proposition support both forecasts. If in doubt, assume the proposition supports more than one forecast.

4 More Examples

To take a closer look at any of the examples, the map IDs which can be used to open the map in OVA are: **28256, 28257, 28258, 28259**

4.1 More Multiple Choice Options

All of the questions are in a similar multiple choice format, some will be binary yes/no and some will have up to five different options. There is no difference in how these should be annotated, all follow the same format and have been added already, to be checked over and reconstructed. To see an example, take a look at map ID 27785.

4.2 Updates

Some of the examples have a second set of forecasts and rationale, when the forecaster has updated their predictions. In this case, all sets of predictions are added already and can be moved lower or higher as needed to slot in the analysis

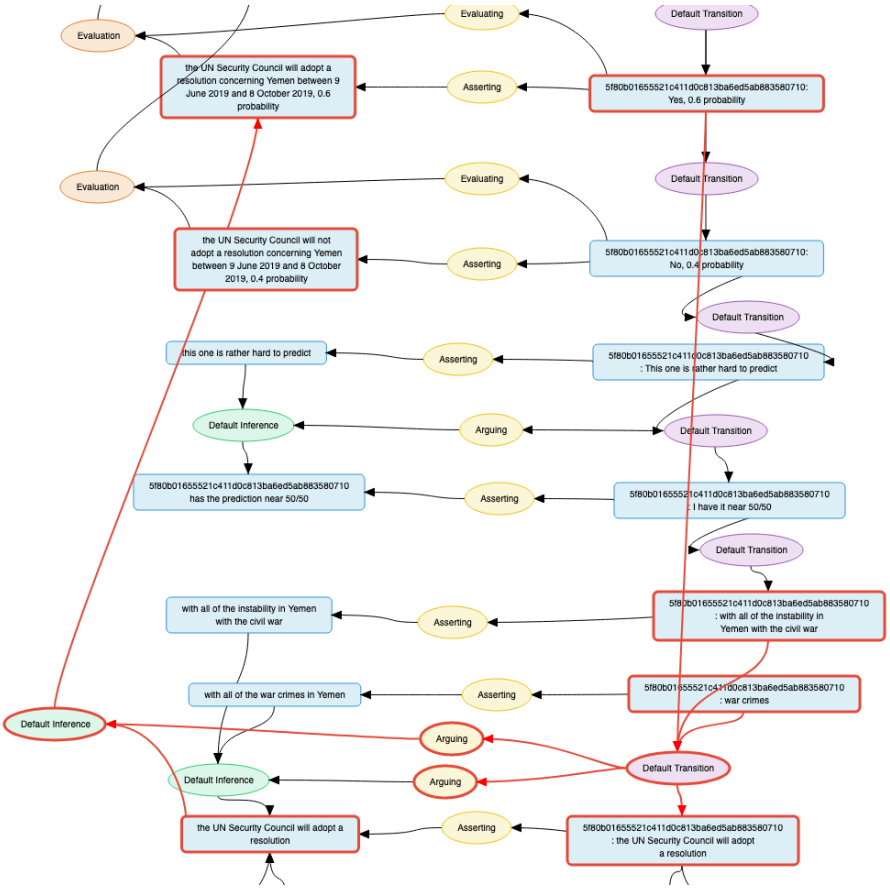

Figure 4: Part 32262 - Support for One Forecast

of the rationale in between, so that the map follows the order of the text. To see an example, take a look at map ID 27786.

5 Annotation Checklist

1. Check Question text is correctly highlighted in the left-hand side text in OVA, and annotated as Pure Questioning
2. Check multiple choice options are highlighted correctly in the left-hand side text in OVA, all in separate propositions and annotated as Hypothesising, with CAs between them.
3. Reconstruct propositions of the question and the multiple-choice options
4. Check forecasts are highlighted correctly in the left-hand side text in OVA, annotated as Asserting, and have the MA Evaluation implemented correctly to the hypothesis above.
5. Reconstruct propositions of forecasts.
6. The remaining text of the rationale is to be annotated using IAT and connected up to the propositional content of the forecasts.