

Amplify Rootclaim: Probably a Better Method for Forecasting Unanswerable Questions

Nuño Sempere*

November 8, 2021

Motivation

In [Reciprocal Scoring: A Method for Forecasting Unanswerable Questions](#), Karger et al. describe a method to elicit predictions in situations in which resolutions are outright not possible, or very far away. They provide some preliminary evidence of its effectiveness in the form of [elaborate]. However, in the post-peer-review discussion phase in social media, Karger et al.’s method was met with a lukewarm reception from the community of forecasting practitioners, which has grown to view methods which resemble Keynesian Beauty Contests with suspicion.

In this working paper, we outline an alternative method, which briefly looks as follows:

- There is a trusted central authority capable of approximate Bayesian updates which cares about its long-term reputation, but which has limited capacity
- A larger contingent of which submit bids to that central authority, and are rewarded when their bids cause the central authority to update.

The authors have no affiliation with [Rootclaim](#), but Rootclaim might be considered a prime example of the central authority necessary for this method to work. Rootclaim produces a small number of highly detailed Bayesian analysis of current affairs, and has a strong interest in maintaining its reputation in the long-term—on the one hand because its business model is precisely based on maintaining an accurate long-term public track record, and on the other hand to prove the superiority of Bayesian methods, as every Bayesian deeply desires in her heart to do. Throughout the paper, we will thus use “Rootclaim” as a shorthand for “central authority capable of approximating Bayesian updating.”

*Quantified Uncertainty Research Institute. I am thankful to Ezra Karger et al. for allowing himself to be billed for criticism of his proposed methods.

Description of the method

In the interest of brevity, we shall outline our method by means of an example, and the example shall be the question “Will the People’s Republic of China have annexed at least half of Taiwan by 2050?”, as operationalized by [Metaculus](#).

A rough prior to reduce potential forecasting reward

Taiwan has been independent of mainland China since the 25th of October 1945, i.e., 76 years into the past. Per Laplace’s law, the chances that this will change by 2050 is $1 - (1 - \frac{1}{(2021-1945)+1})^{2050-2021} \approx 32\%$. Lets take this 32% as Rootclaim’s initial probability. Note that per the [reference class problem](#), other reference classes might have been chosen, so the point of this prior is not to be definitive, but rather to provide a starting point less arbitrary than 50% from which forecaster reward might be computed in the next steps.

Forecasters attempt to move the prior

By arguing that the prior is wrong

Forecasters might output an argument like the following:

The 32% prior is computed with respect to Taiwan, but we have more information than that by looking at how China has asserted control over other outer territories or autonomous regions. Among these one might count Hong Kong, Macau and Tibet (and perhaps the Xinjiang Uygur Autonomous Region). Although the process of asserting control is more continuous in time, if we assume for simplification purposes that it happens in any one year, this moves our Laplace prior that another external territory will be brought under control at $1 - (1 - \frac{4}{(2021-1945)+1})^{2050-2021} \approx 78\%$.

By providing evidence about current affairs

Forecasters might output an observation like the following:

China’s current bellicosity is different from the historical norm. For instance, China has been [sending record number of planes over Taiwan’s air defence zone](#). This should increase the probability of a successful resolution.

Rootclaim updates, and forecasters are rewarded

After receiving the updates such as above, Rootclaim makes an approximately Bayesian update. For instance, maybe Rootclaim averages the two possible base rates, and moves from 32% to $(32\% + 78\%)/2 = 55\%$. Then, it moves 5% to 60% on account of the increased frequency of flights over Taiwanese airspace.

Now, one way forecasters could be rewarded would be in proportion to the percentage points of the update. However, this has the problem of rewarding an update from 1% to 2% the same as an update from 50% to 51%, which seems undesirable. One parsimonious way to solve this is by denominating updates in terms of bits—which are the natural unit for strength of conviction, though we will not justify this here because acquiring intuitions for why this is the case can prove to be difficult.

For intuition, an update from $1 : 2^n$ odds to $1 : 2^m$ odds is an update of $(m - n)$ bits, whose size is of $|m - n|$ bits, where $|\cdot|$ represents out the absolute value. Continuing to our example, the original probability of 32% corresponds to around $1 : 2^1$ odds, or strength of conviction of around -1 bit. The posterior probability of 55% corresponds to around 1.22 : 1 odds, or around $2^{0.29} : 1$ odds, or +0.29 bits. The difference between these is $(0.29 - (-1)) = 1.29$ bits. An update from 55 to 60% would likewise be an update from $1 : 2^{0.29}$ to $1 : 2^{0.59}$, or 0.3 bits.

If we valued each bit of information about an invasion of Taiwan at \$1000, the first forecaster would receive \$1,290, while the second would receive \$300.

Laziness metric

More on the central authority

More on the

Evidence base

First principles, programming experience

Truth from a liar, amplification experimetns

Comparison to Karger’s method

Weird loops Less incentivized to do expensive search <https://twitter.com/LinchZhang/status/14557595861582684>

I still think this is partly addressed by “if you have large enough teams, you should expect the other team to do this stuff too” but will bring it up with Ezra!

I will be reading the paper to find out how it does not actually incentivize consensus among forecasters against the better, and private, judgements of a minority among them.

No weird loops - There is a definite connection to reality - A correct contrarian can attempt to convince the Bayesian - The Bayesian has an incentive to - A perso

The Bayesian has an incentive to

The output is much more legible

Amplify lazy rootclaim variant.

Alternative design: Attempt to predict what the central authority will predict.
Has no savings unless resolution is done probabilistically.