**Bayesian Aggregation of Crowd Judgments for Quantitative Fact Checking**

M. Lambardi di San Miniato, M. Battauz, R. Bellio and P. Vidoni

1 Department of Economics and Statistics, University of Udine, (e-mail: [michele.lambardi, michela.battauz, ruggero.bellio, paolo.vidoni]@uniud.it)

**Abstract:** Political fact-checking can be carried out by crowd workers, provided they are supervised by experts. We propose a Bayesian latent variable ordinal probit model for truthfulness rating data, to estimate workers’ reliability, weigh in their contributions, and surrogate expert judgments. This is a notable example of aggregation function of an implicit type. This method may be used to dynamically assign workers to new tasks, as illustrated with an analysis of PolitiFact data.

**Keywords:** Bayesian statistics, judgment aggregation, ordered probit.

1 Introduction

Fact-checking is about assessing the truthfulness of public statements to combat misinformation and improve debates. However, expert fact-checkers are few, while crowd workers are readily available but potentially biased. There is a stream of scientific research about how to surrogate expert judgements by means of workers, after some suitable calibration; see for example Roitero et al., 2021. Latent traits of statements and workers are at stake, like truthfulness and political orientation. Methods from Item Response Theory (see for example Bartholomew et al., 2011) can be adapted to this end. Here we adopt a Bayesian approach, which is suitable for the task.

We propose an ordinal probit model for quantitative fact-checking. An aggregation function is involved, which mimics expert judgments via the wisdom of crowds (Roitero et al., 2021). The truthfulness of statements, even when encoded as an ordinal variable, is often treated as numeric. This allows to summarize ratings across workers by means of a simple average, but by using a generative model there is room for improvement. We argue that, as far as the aggregation function needs not be explicit, the Bayesian inferential approach always provides one, namely, the posterior distribution of expert judgments conditional to workers’. A different proposal, with some similarities, exists in the literature (Nguyen et al., 2018).
2 Model

Let $i = 1, \ldots, n$ and $j = 1, \ldots, m$ be two indices to identify statements and workers, respectively. The truthfulness of statement $i$ is rated as $Z_i = 1, \ldots, k$ by a single expert and as $W_{ij} = 1, \ldots, k$ by worker $j \in C_i$, with $C_i$ a subset of workers that evaluate statement $i$. The aim is to predict $Z_i$ through $(W_{ij})_{j \in C_i}$.

As typical in ordinal regression models, we think of $Z_i$ as the observed discretization of a latent numeric variable $Z_i^*$, which is defined as

$$Z_i^* = \sigma \xi_i + \epsilon_i, \quad \xi_i, \epsilon_i \sim \mathcal{N}(0, 1).$$

Here, $\epsilon_i$ is a noise term, $\xi_i$ is the truthfulness of the $i$-th statement and $\sigma > 0$ is a signal strength parameter. Analogously, we think of $W_{ij}$ as the observed discretization of a latent numeric variable $W_{ij}^*$, which is defined as

$$W_{ij}^* = \alpha_j + \beta_j \xi_i + \eta_{ij}, \quad \alpha_j \sim \mathcal{N}(0, \sigma^2_{\alpha}), \quad \beta_j \sim \mathcal{N}(0, \sigma^2_{\beta}), \quad \eta_{ij} \sim \mathcal{N}(0, 1).$$

Here, $\eta_{ij}$ is a noise term, while $\alpha_j$ and $\beta_j$ are worker-specific parameters that affect their judging behavior. The worker-specific parameters $\alpha_j$ and $\beta_j$ account for correlation within workers. All the terms $\xi_i, \epsilon_i, \alpha_j, \beta_j, \eta_{ij}$ are assumed independent. Lastly, we define two sets of thresholds $(\gamma_h)_{h=0}^k$ and $(\delta_l)_{l=0}^k$ constrained as $\gamma_0 = \delta_0 = -\infty, \gamma_h < \gamma_{h+1}, \delta_l < \delta_{l+1}, \gamma_k = \delta_k = +\infty$, such that

$$\gamma_{h-1} < W_{ij}^* \leq \gamma_h \iff W_{ij} = h, \quad \delta_{l-1} < Z_i^* \leq \delta_l \iff Z_i = l.$$ Probit models are implied for $Z_i$ and $W_{ij}$. As an original proposal, parameters $\alpha_j$ and $\beta_j$ allow to represent the alignment with the experts. The model specification is then completed by assigning weakly informative priors to scale parameters and uniform priors on thresholds.

3 Example

We analyse a publicly available dataset (Roitero et al., 2020), which includes expert ratings obtained from PolitiFact. Data relate to $m = 100$ workers and $n = 62$ public statements on COVID-19. The truthfulness ratings $Z_i$ and $W_{ij}$ have $k = 6$ levels, labeled as: “pants-on-fire”, “false”, “mostly-false”, “half-true”, “mostly-true” or “true”. There were eight statements per worker and ten workers per statement, but two gold statements were rated by all the workers for control purposes. Gold statements have either $Z_i = 1$ or $Z_i = k$, while all
the other statements administered to each worker cover different $Z$ values. We estimate the model via the R interface to the Stan probabilistic programming language (Stan Development Team, 2023).

Figure 1 shows the posterior distribution of $\xi_i$, along with the thresholds $\gamma_h$. Were it only for the model on $Z_i$, the boxplots should all be similar, but the model on $W_{ij}$ complements that information, so that there can be gradients of $\xi$ within levels of $Z$.

Figure 2 summarizes the inferential results for $\alpha$ and $\beta$, which are affected by political orientation. Liberals tend to be more aligned with the truth (large $\beta_j$) and tend to give lower ratings (small $\alpha_j$). Instead, conservatives seem more gullible (large $\alpha_j$) and less aligned with the truth (small $\beta_j$). There are even two workers with negative $\beta_j$, who are detrimental on fact checking.

4 Conclusion

Our analyses support that Bayesian generative models may lead to important advances for crowd-based fact checking. Future research will focus on the usage of the model for prediction of $Z$ given $W$, and on the extension to more complex settings.

Acknowledgments  This work was supported by the Departmental Strategic Plan (PSD) of the University of Udine, Interdepartmental Project on Artificial
Intelligence (2020-2025).

References


