Discussion on 04-03-2020 - Quantifying Voter Biases in Online Platforms: An Instrumental Variable Approach

The discussion started with two questions: why does the authors use two-stage least squares(2SLS) regression to estimate the causal effects of impression signals? How can we justify the instrument variable? The key to conduct 2SLS is that the dependent variable's error terms are correlated with the independent variables and the instrument variable(IV) is truly identified. Finding the IV is challenging because the IV should be correlated with the exposure and the IV should not directly affect the outcome. In addition, the IV and the outcome should not share causes. However, only the correlation between IV and exposure is empirically justifiable while the other two conditions can only be justified through augmentation. Due to the observational nature of the study, the only way to verify the IV is to use our knowledge to build a case for showing that how the IV may be reasonably assumed to meet the conditions. This is also the reason why we need to choose the IV carefully in 2SLS.

Himel mentioned the motivation of building an IV model to quantify the voter's bias. In online platforms, voter's bias is easy to neglect while the aggregate user feedback is commonly used for measuring content quality. Statistical models from prior work identify the correlation, not the causal effects of different impression signals, which motivate the present work because to formulate bias quantification we must rely on causation instead of dependent on association only.

Another question involved in the discussion is Why not using randomized experiments as finding a valid IV is hard. Prof. pointed out that we can neither conduct randomized control trials on the

actual platforms nor expect on user's reaction in a simulation. With access to the observational data, IV is a reasonable approach for estimating causal effects.

A student pointed out that the debiased platform which hides the votes and user profile would make students spend unnecessarily longer time to find a reasonable answer, especially for students trying to find debugging help with on stackoverflow. Prof. pointed out that the point is to make users focus on the content itself and it is not that the vote will be hidden forever but it will appear after you vote. So it is trying to lead people to unbiased voting.