Assignment #9

Sanittawan Nikki Tan

12/14/2018

Referee Report

Paper: Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning

Author: Adam Bonica

Field: Political Science

Summary

Bonica's paper, *Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning*, should be published by the American Journal of Political Science due to the following reasons. Firstly, the paper addressed a clear research question with strong supporting arguments, data, evidence and sound methodology. Secondly, the paper convincingly justified its contribution to the existing literature in the study of ideology. Finally, although there are some writing issues such as insufficient elaboration and the use of complicated language, these issues are minor. In addition, the paper may lead to many interesting studies in other fields of political science such as understanding how countries voted in the United Nations.

This referee report assesses the quality of the paper titled *Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning* submitted by Adam Bonica to the American Journal of Political Science ("AJPS"). It is evaluated based on four criteria which are (i) the research question and the author's answer (ii) the methodology used to answer such question (iii) the author's overall contribution to the body of the existing literature (iv) writing issues such as writing style and typographical errors. The report concludes with a discussion on potential ways to extend the research question and/or the methodology.

Research Question and Answer

Bonica adequately defined the research question. He examined how well supervised machine learning models forecast voting behavior of legislator by using unconventional data sources like campaign contribution prior to entering legislative offices in comparison with conventional and alternative roll-call measures such as Nokken-Poole, Bailey Scores, and Common-Space CFScores (Bonica 2018a, 830-831, 847). By voting behavior, the author means how a legislator would be rated by the DW-NOMINATE, which is a roll-call voting-based scores for political ideology developed by Keith Poole and Howard Rosenthal. The author found that fundraising data during the nonincumbent period coupled with random forest model performed as accurately as using the first two years voting records of legislators in predicting political ideological points. In addition, feature analysis also provides insights to what variables are important components of the prediction, an area which has been a "black box" in the study of ideology (Bonica 2018a, 840).

From my perspective, the author compellingly answered the question and convincingly supported his arguments with sufficient data. The replication materials are also provided via Dataverse. The campaign contribution data that he used in this study seems to be reliable although it is the same dataset that he collected and made available for other research purposes (DIME). Some parts of the article may be unclear or not well-elaborated. I will discuss writing-related issues in the writing section of this report.

Methodology

Although the paper is about a new methodology for studying political ideology, we may be able to assess the paper based on whether or not the proposed methodology is appropriate for the study since there are a variety of supervised learning models. I believe the author adequately supported his choices of random forest and support vector regression models because he argued that DW-NOMINATE score is "measured along a continuous dimension" which convinced him to use regression-based models (Bonica 2018a, 835). This suggests that he did not randomly choose a model that would yield predetermined results. Not only did he test the performance of three supervised learning models, but he also reported in-sample and out-of-sample predictive power. I think that the author substantially increases the credibility of his finding through cross-validation. What I really like about this paper is the use of campaign contribution. Bonica was right that nearly all candidates engaged in fundraising when they ran for offices, so the campaign contribution data is suitable for linking both incumbent and nonincumbent candidates for predicting voting behavior.

Contribution to the Literature

I think Bonica did very well in discussing the existing work in this topic and how his work contributed to the existing literature. At the beginning of the article, he gave ample background on the study of spatial maps of preferences which is center to the study of political ideology, connected it to DW-NOMINATE which is used in his article, and surveyed recent work by Tausanovitch and Warshaw which pointed out the fact that several alternative measures of political ideology do not perform very well among within-party comparison, one of the main shortcomings of these measures (Bonica 2018a, 831). This discussion nicely led to his proposal of using supervised machine learning models which not only do they perform well in distinguishing members of the same party, but they also accurately predict voting behavior of candidates who lack voting records.

In terms of citation, McCarty et al. (2006) is the only reference that was not explicitly mentioned in the paper. However, the inclusion of this work seems reasonable because the author referred to numerous work by either McCarty, Poole or Rosenthal. It appears to me that Bonica's work may be influenced by this book although it was not directly cited in the text. I believe that Bonica included most of the seminal work in this field. I do not have any particular suggestion.

Writing Issues

In general, the paper is well-structured. The author begins by laying out the current challenges in measuring political ideology which include the problems associated with the holographic interpretation of ideology. The first problem is that the interpretation and the ideological methods utilized in scaling models are conflicting. While the scaling models begin with revealed preferences on issues which are mapped on a scale which has a lower number of dimensions, the interpretation starts in the opposite direction (Bonica 2018a, 832). Secondly, the scaling methods are sensitive to how frequently an issue is voted in the Congress. This places more weight on popular issues than issues that are voted on once or have not been voted on at all (Bonica 2018a, 832). Finally, the current method of scaling ideology reveals very little about why certain issues are bundled or how they are related to ideological space (Bonica 2018a, 832). To illustrate the stated problems, Bonica supported his argument with a necessary chart in figure 1 (Bonica 2018a, 833). The writing in the problem statement part is clear but too complicated at times.

After having established the problems, he then moved on to proposing his method of predicting political ideology which is the use of supervised machine learning models by comparing the performances of random forest, support vector regression, and supervised CFscores with roll-call measures and alternative measures. The target variable is the DW-NOMINATE scores. Although the author provided an adequate and well-

constructed table 1 that details fit statistics, errors, and number of observations (Bonica 2018a, 838), I personally find that the description of both support vector regression and random forests (page 835 and 836 respectively) does not provide sufficient details for me to replicate the work. In particular, it is unclear from the model training section how the author constructed the labeled data set. It would be useful for the audience of AJPS to understand the research procedure in enough details that they do not have to refer to the supporting materials. On a separate matter, the supporting materials which include 37 files can be overwhelming at first; however, the author provided a clear well-written instruction on replication the description of each file which is helpful (Bonica 2018c).

The feature analysis section of the paper is unquestioningly one of the major areas of contribution. The section was allocated appropriate space and is well-supported by tables and charts (Bonica 2018a, 841 - 844). I think that table 3 can be shortened to include only top 10 to 15 organizational donors because the continued portion of the table (page 843) appears to add little value to the information available in page 842 and author's point on the types of donors that matter most to the ideological mapping (Bonica 2018a, 840). The full table could have been added to the appendices.

Finally, it seems that the author did not quite explain how his methodology will aid or improve the theories on representation and legislative behavior. It is clear that he contributed to the scenario when voting records are lacking and that his method performed significantly better than other existing methods in distinguishing within-party legislators. However, the author could have elaborated on how contribution in these two areas is linked to the two theories, especially within-party differentiation. Furthermore, the author mentioned that the method proposed in the paper has a wider application to other studies, but the paper falls short on giving examples of those applications.

Regarding typographical errors, I did not detect any misspellings. For the most part, the author's style of writing is clear, concise, and straight to the point. Nevertheless, some sentences are more complicated and wordy than they ought to. For example, the sentence "There are reasons to doubt such an interpretation." could have been rewritten as "The interpretation is doubtful for several reasons." (Bonica 2018a, 832). Some sentences have repetitive words. For instance, the word "widely" is used twice in a single sentence: "As the most widely used measure in the literature, DW-NOMINATE is widely viewed as a benchmark measure of ideology." (Bonica 2018a, 830). Despite these minor flaws, the paper is considered well-written in my perspective.

Further Suggestion

Although I am not equipped to propose an extension of the methods used in this paper due to my limited machine learning knowledge, for the time being, I would like to propose an extension of the research question. It seems to me that the methods could very well be applied to the study of the United Nations roll-call votes. Specifically, how similar and dissimilar voting behavior of countries within the same block, such as African countries, are as well as voting behavior prediction. Firstly, we can adopt the supervised machine learning method proposed by Bonica to predict the future voting behavior of newly admitted countries or to place countries with minimal voting records on an ideological scale. Secondly, it would be interesting to uncover any voting patterns among countries that are treaty-allies and those that are not. Interesting questions would be, for example, is there any distinguishable pattern of how the United States and Japan voted at the UN on human rights and security issues.

Based on my research, scholars such as Erik Voeten and Michael Bailey have worked on UN roll-call votes and countries' ideal points (Bailey et al. 2015; Voeten 2009). However, using techniques that Bonica proposed may contribute to the existing literature by increasing the precision of predicted voting behavior. The data on UN roll-call votes are also available online through UNBISNET, the United Nations database, Voeten et al.'s compilation from 1946 to 2017 (2009), and the U.S. Department of State.

References

Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2015. "Estimating Dynamic State Preferences from United Nations Voting Data." Journal of Conflict Resolution 61 (2): 430–56. https://doi.org/10.1177/0022002715595700.

Bonica, Adam. 2018a. Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning. American Journal of Political Science, 62: 830-848. doi:10.1111/ajps.12376

———. 2018b. "Replication Data for: Inferring Roll Call Scores from Campaign Contributions Using Supervised Machine Learning." Edited by Adam Bonica. Harvard Dataverse. https://doi.org/10.7910/DVN/7EWDRF.

———. 2018c. "codebook_bonica_dwdime_ajps.pdf." Replication Data for: Inferring Roll Call Scores from Campaign Contributions Using Supervised Machine Learning. https://doi.org/10.7910/DVN/7EWDRF/CFETUO, Harvard Dataverse, V1.

McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 2006. Polarized America: The Dance of Ideology and Unequal Riches. Cambridge, MA: MIT Press.

Voeten, Erik, Anton Strezhnev, and Michael Bailey. 2009. "United Nations General Assembly Voting Data." Edited by Erik Voeten. Harvard Dataverse. https://hdl.handle.net/1902.1/12379.