Assignment #9 – Referee Report

MACS 30000, Dr. Evans
Ying Sun

Article: Athey, Susan, "The Impact of Machine Learning on Economics," in Joshua Gans Ajay K. Agrawal and Avi Goldfarb, eds., The Economics of Artificial Intelligence: An Agenda, National Bureau of Economic Research

Research Question & Methodology

What's the early contribution of machine learning to economics and the impact of machine learning on the future development of economics? In order to answer this research question, the author provides a comprehensive assessment of the applications of machine learning in economics. The author adequately defines the research question in the introduction part by concluding the current achievements of machine learning in the field of economics and the predictability of its future developments.

The author compellingly answers this research question according to the following logical structure:

First of all, the author comes up with the definition of machine learning used in this article. By clarifying this narrow but practical definition of machine learning, the author emphasizes the difference between machine learning and traditional econometric approaches. Furthermore, the author divides the applications of machine learning into two main branches, "supervised and unsupervised ML" (Athey, 2018, p.3). From several perspectives, the author compares machine learning with traditional methods in empirical analysis. Initially, they have different emphases. Traditional methods focus on the estimated effects while machine learning pays more attention to the fitting level of the model. In addition, they have distinct concerns and the corresponding approaches to address these concerns. For machine learning, the key concern is the "tradeoff between expressiveness of the model and risk of over-fitting" (Athey, 2018, p.4). To relieve this concern, machine learning develops multiple techniques to strike a balance between expressiveness and over-fitting. In contrast, the key concern of empirical economics is the "causal effect estimation" (Athey, 2018, p.5). More specifically, "whether the parameter estimates from a particular sample are spurious" (Athey, 2018, p.6) and "whether the assumptions required to 'identify' a causal effect are satisfied" (Athey, 2018, p.6) are main concerns for causal models. The author further analyzes

whether machine learning can solve these two issues to some extent.

Second, the author demonstrates the application of machine learning methods in policy analysis. Due to the brilliant performance of machine learning in prediction, there have been many successful applications such as hip replacement operations, credit scoring and stop-and-frisk laws (Athey, 2018, p.7-8). However, there are also some problems that cannot be ignored, such as "the interpretability of models", "the question of fairness and nondiscrimination", "stability and robustness" and "manipulability" (Athey, 2018, p.8-9).

Third, the author makes the prediction based on the above analysis. She is optimistic about the impact of machine learning on economics in the future. New methods combining machine learning with causal inference will be created. Successful applications of ML methods in causal inference are not rare to see. One important aspect is the "estimation of average treatment effects under the unconfoundedness assumption" (Athey, 2018, p.12). For example, regularized regression is an effective way to solve potential covariates "in an environment where the outcome model is 'sparse'" (Athey, 2018, p.12). Other methods such as "residual balancing" and "double machine learning" (Athey, 2018, p. 12-13) also have amazing performances. Another vital aspect is "the estimation of heterogeneity in treatment effects" (Athey, 2018, p.13). Both "causal trees" and "causal forests" play important roles (Athey, 2018, p.14). Then, in the following exposition, the author lists other successful applications of ML in causal inference. For instance, machine learning facilitates the researches on "estimating optimal policies", "robustness and supplementary analysis", "panel data and difference-in-difference models" and "factor models and structural models" (Athey, 2018, p.16-21).

Finally, the author predicts that machine learning will bring tremendous changes to economic research.

From my perspective, the method adopted by the author to answer the research question is appropriate and sufficient. In terms of the current impact of machine learning on economics, the author provides substantial examples of machine learning methods in policy analysis to explain why machine learning has great potential in the field of economics. Based on the comprehensive and persuasive analysis, the author makes her confident prediction that machine learning will fundamentally influence the methodology in economics, especially in the combination of causal inference. In general, the author answers the research question in an organized and reasonable way except for a minor problem that there are some biases in the example selection. Relating

to the author's field of expertise, most of the examples chosen by the author focus on policy analysis and econometrics which may not sufficiently answer the research question which uses a broader term –"Economics".

Literature

In general, the author has an excellent performance in the literature review. She covers a wide rang of topics such as economics, machine learning, policy analysis and other interdisciplinary studies. Especially in the policy analysis part, the author comprehensively demonstrates that machine learning greatly promotes the policy analysis and related decision making by citing sufficient literature. These literatures help the author to express her opinion more persuasively and clearly. The authors states that some "off-the-shelf ML methods" (Athey, 2018, p.7) is important in policy analysis and decision making by citing an example of "hip replacement operation" in Kleinberg et al. (2015) and another example of "credit scoring" in Bjorkegren and Grissen (2015). Then, by citing the example of examining "stop- and- frisk" laws in Goel et al. (2016) and the example of allocating inspector resources in Glaeser et al. (2016a), the author further proves that the wider application of this type of prediction methodology in the policy context.

However, there are also some minor flaws in author's citations.

First, the author misses some necessary citations before drawing the conclusion that "there are many interesting methodological issues involved in finding models that have stable performance and are robust to changing circumstance" (Athey, 2018, p.9). Obviously, it is not a direct statement or analysis. I think it is necessary to cite some literature to specify the methodological issues emerged in changing circumstance.

Second, there are some unnecessary citations in this paper. For instance, the author cites the study in Yeomans et al. (2016) to illustrates that "there are other ways to mathematically formalize the interpretability of models" (Athey, 2018, p.9). However, this citing paper mainly focuses on the reasons and possible solutions for the lack of trust in machine recommendations (Yeomans et al., 2016). It is not closely related to the author's argument.

Grammatical, Spelling and Style Errors

There are several grammatical, spelling and style errors in the paper. On page 7, in the sentence of "where the probabiity of selecting…" (Athey, 2018, p.7), "probabiity"

should be changed to “probability”. On page 8, I think there is an inappropriate expression in “Of course, economists have long understood that simple models can also be misleading” (Athey, 2018, p.8). It is much better to change to “Of course, economists have long belief that simple models can also be misleading”. On page 18, “since we dont observe the counterfactual outcomes…” (Athey, 2018, p.18), “dont” should be corrected as “don’t”. On page 20, an “ ’ ” is missing in “users willingness” and a similar error exists in “users typical morning location” (Athey, 2018, p.20). On page 22, “This article has al discussed the first three predictions…” (Athey, 2018, p.22) should be “This article has already discussed the first three predictions…”. Still on page 22, “Thus, there is no obvious benefit from ML in terms thinking about identification issues” (Athey, 2018, p.22) misses an “of” after “terms”.

Extension

Based on the method we discussed above, we can extend the research question to a more complex interaction involved biology, machine learning and economics. We could review the early interactions among these three areas and make a corresponding prediction based on our analysis. For instance, many machine learning methods have been developed from biological concepts or models and further applied to economics area such as stock prediction. More specifically, BP neural network algorithm has important applications in stock prediction. It evolved from neurosynaptic structure in biology and then it was introduced into machine learning. The self-adaptive ability of BP neural network can fit in with the characteristics of stock market – dynamic and difficult to predict.

References

Athey, Susan, “The Impact of Machine Learning on Economics,” in Joshua Gans Ajay K. Agrawal and Avi Goldfarb, eds., The Economics of Artificial Intelligence: An Agenda, National Bureau of Economic Research

J. Kleinberg, J. Ludwig, S. Mullainathan, and Z. Obermeyer. Prediction policy problems. American Economic Review, 105(5):491–95, 2015.

S. Goel, J. M. Rao, R. Shroff, et al. Precinct or prejudice? understanding racial disparities in new york citys stop-and-frisk policy. The Annals of Applied Statistics, 10(1):365–394, 2016.

E. L. Glaeser, A. Hillis, S. D. Kominers, and M. Luca. Predictive cities crowdsourcing city government: Using tournaments to improve inspection accuracy. The American Economic Review, 106 (5):114–118, 2016a.

M. Yeomans, A. K. Shah, and J. Kleinberg. Making Sense of Recommendations. 2016. URL http://goo.gl/8BjhMN.