The Referee Report of Athey (Economics)

Siyuan Peng

The research topic of this paper is quite straightforward and has already been mentioned in its title: “What is the impact of machine learning on economics?” In the introduction part of her paper, the author thoroughly defined her research question and gave us a quick glance at the whole structure of the paper. She not only clarified the concept of machine learning (abbreviated as ML) which she will use in the whole paper, but also introduced several themes about the application of ML and the future trend of it. Through her brief introduction, we could find that “the application of ML on economics research has already generated some impacts which raised new questions or brought new approaches. It even leads to a change of the way in which economists involving with engineering and implementation of policies. What’s more, considering the current status, it is rather easy to predict the future trend of it.”[1]

Then, the author addressed this question by offering detailed instruction of all the aforementioned concepts and themes. To start with, she defined machine learning as “a field that develops algorithms designed to be applied to datasets, with the main areas of focus being prediction (regression), classification, and clustering or grouping tasks.” [2] Followed by this specific definition, she introduced the concept of unsupervised learning, supervised learning and other categories and highlighted their potential use on economic researches. What’s more, she offered a comparison of the main difference between ML and empirical economics. Even though difference might lead to some unexpected application problems, the application of ML might help economists solve economic issues in a better way. Then, Ashey offered some concrete cases of the

[1] 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 Page 1, ‘Introduction’

[2] Same Paper, page 2, ‘What is Machine Learning and What are Early Use Cases?’

'applications of prediction methods to policy problems in economics"[3], like "hip replacement, credit scoring" and so on, which deepened our understanding of the use of ML and its advantages. Later, Ashey mentioned several strategies to "combine ML and causal inference together to make casual effect identification" and the cutting edge of the application of this new method on some problems of interest. Including "average treatment effects", "heterogeneous treatment effects and optimal policies", "estimating optimal policies using adaptive experimentation", "robustness and supplementary analysis", "panel data and difference-in-difference models" and "factor models and structural models"[4]. Last, but not least, Ashey made her own "prediction about the future impact of ML on economics". She believed that "AI and ML will profoundly transform economics"[5].

From my perspective, Ashey has compellingly answered her research question and the methods she used were appropriate. She offered plenty of concepts, ideas, cases, and skills to illustrate her idea with rigorous logic. From all of her description, even a non-professional of ML (like me) would feel convinced about her prediction of the impact and the future trend. However, if I really need to give her some advice about the methods, I might point out that the range of her research topic might be too broad to give a well-rounded conclusion. By saying that "the impact of machine learning on **Economics**", readers might expect that she will state impacts on all areas of economic researches. Nevertheless, it is an impossible mission. Considering that Ashey is an expert of econometric and marketing economics[6], the examples she used in this paper are intensively lying in microeconomics. Thus, maybe it will be better if she could give more cases of the use of ML in macroeconomics in her paper to make her methods even more sufficient.

The author indeed did a great job linking a wide range of papers to her own work. In

[3] Same Paper, page 7, 'Using Prediction Methods in Policy Analysis'
[4] Same paper, page 10-20, 'A New Literature on Machine Learning and Causal Inference'
[5] Same paper, page 21-26, 'Broader Predictions About the Impact of Machine Learning on Economics'
[6] Information from author's personal website: https://athey.people.stanford.edu/research

her paper, we not only saw a lot of examples from other ML papers and empirical economics papers, but we also found inspiring cases from public policy, statistics or even media studies. However, in her paper, I found that some of the citations are not clearly pointing out in which paper she get that conclusion or example. By citing more examples, the author wanted to make her idea more convincing. However, when providing multiple cases, it is still necessary to clearly point out the major one in which you get your main idea. For example, in page 23, when she wanted to illustrate "the ability to use predictive models to measure economic outcomes at high granularity and fidelity will change the types of questions we can ask and answer", she gave an example that "imagery from satellites or Google's street view can be used in combination with survey data to train models that can be used to produce estimates of economic outcomes at the level of the individual home". However, for this example, she cited three papers (Jean et al. (2016), Engstrom et al. (2017), Naik et al. (2014)). Even though these three papers are all related to the idea that she wanted to illustrate, she still needs to point out in which paper she got this specific example. This is an example of insufficient citation. In addition, when talking about "the privacy issue when analyzing personal data"[7], the author might get some inspiration by our Assignment 8. Both of the papers from Zimmer(2010)[8] and Sweeney(2002)[9] could be a great supplement of her argument here.

This is a rather long paper and I think it is inevitable to have some grammatical or spelling problems. In the "Introduction" part of her paper, the author said "Then, I provide an overview of the questions considered and early themes of **the the** emerging literature in econometrics and statistics combining machine learning and causal inference", in which she double typed "the". Later, still in the "Introduction" part, she said "including improved performance as well as enabling researchers", in which there

[7] Same paper, page 23-24, 'Broader Predictions About the Impact of Machine Learning on Economics'

[8] Zimmer, Michael, "But the Data is Already Public: On the Ethics of Research in Facebook," Ethics and Information Technology, 2010, 12 (4), 313-325.

[9] Sweeney, Latanya, "K-Anonymity: A Model for Protecting Privacy," International Journal on Uncertainty Fuziness and Knowledge-Based Systems, 2002, 10 (5), 557-570.

should be an “an” before “improved”. In the final paragraph of introduction, “ranging from new questions, to new approaches to collaboration” the comma between “questions” and “to” is unnecessary. In the second part “What is Machine Learning and What are Early Use Cases?”, she said “and when a human watches **the the** largest group”, double “the” again. The same part, “while they were widely studied in the self-described ML and/or “statistical learning” literatures”, literature is uncountable, thus shouldn’t be made plural. The same part, “such demand shifters can be referred to as “confounders" becaues the affect both the optimal price”, where “becaues” should be “because”. In page 9, “The models will find subtle relationships bewteen X and Y ”, where “bewteen” should be “between”. In page 13, “Their approach is similar to a “doubly-robust” method for estimating average treatment effects that proceeds by taking the average of the efficient score”, where “proceeds” should be “proceed”. Quite the same, in page 17, “The fact that the exploration benefits the future through a model of how contexts relates to outcomes changes the problem”, where “relates” should be “relate”. In page 20, “Since there are 25000 bundles when there are 5000 products, in principle each individual consumer’s utility function has 25000 parameters”, there should be a comma after “in principle”. In page 23, “with a wide range of economic variables remains stable”, where “remains” should be “remain”. Same page, “Economists are likely to refine these methods to make them more directly useful quantiatively”, where should be “quantitatively”. Even though there are minor errors in this paper, they did not quite affect our reading.

For me, I would suggest the author dig deeper into her research question. Just like what I mentioned above, the scope of her research question is too board. Thus, if she could focus on a specific area of economics, like marketing economics in which she is an expert, she may bring out more concrete and meaningful model with a more precise prediction of the ML application in this specific field. In addition, the research question raised by her is quite enlightening. ML has been used in many areas and would even affect our whole society in the future. Besides economics, we could also make a study about the application of ML in finance, public policy, business, environmental science

and so on. Take finance as an example. When I was searching for the summer internship, I found that a number of intern positions in investment bank require the intern to master at least one programming language and has been exposure to ML related courses. There was a trend that the asset management will be controlled by the well-designed program and the impact of ML to finance domain has already been shown by the layoff of traditional staff and the increasing demand of data scientists (in which I found our MACSS program competitive). Therefore, the research of ML with other subject is also meaningful.

Reference

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

Zimmer, Michael, “But the Data is Already Public: On the Ethics of Research in Facebook," Ethics and Information Technology, 2010, 12 (4), 313-325.

Sweeney, Latanya, “K-Anonymity: A Model for Protecting Privacy," International Journal on Uncertainty Fuziness and Knowledge-Based Systems, 2002, 10 (5), 557-570.