A Survey on Policy Learning

MATH 818.01 Midterm Survey

Wonjun Choi

October 15, 2020

Abstract

This paper surveys growing literature on policy learning in the interdiscipline area of economis, statistics, and computer science. Policy learning incorporates statistical decision making, AI/ML algorithms, and potential oucome model in economics literatue to find optimal policy assignment rule. Policy learning predicts expected outcome of a policy given practical restrictions such as budget constraint, fairness, or political reasons. Considerations rasied in the economics literature, applications of AI/ML algorithms, and mathematical foundation of the results are briefly introduced in turn. A suggestion for the final project is also presented in the final section.

1 Introduction

Suppose you are a dean of the department and you are to allocate each graduate student to put an elephant into a refrigerator.

2 Economic Modeling of the Allocation Problem

Before we begin, it is worth to stop and introduce typical approaches in economics for policy evaluation. Introduced notations would ease our conversation throught this paper.

Policy evaluation has been one of the most important topic in the field of economics. As the government implements various economic policies, the outcome of the interventions need to be analyzied and has been studied in the framework of *potential outcome*, which was frequently refered as *Rubin's causal model.*

Potential Outcome Framework

Consider we are giving aspirins to patients and observing their body temperatures. After treatments are assigned to each patient, we can only observe only one side of the outcome; the temperature with or without taking aspirin. If one takes her aspirin, we cannot know what the temperature would have been without taking it, and vice versa. This is the fundamental problem of *treatment effect* analysis.

What would be the effect of an aspirin on body temperature? It would be the difference of boty temperature between taking aspirin and not. With D equals 1 if the treatment is applied and 0 otherwise, let the potential outcome $Y(1)$ be the outcome with the treatment and $Y(0)$ be the outcome without the treatment. Now we can denote the treatment effect τ as

$$\tau = Y(1) - Y(0).$$

τ cannot be obtained without further assumptions since one of the outcome is not observed. In economics/statistics literature, the unobserved outcome is called *counterfactual.* I refer [Imbens and Rubin, 2015], among many others, for detailed explanations and issues in treatment effect analysis.

Regret Function and Optimal Policy

One practical concern in designing policy could be 'How should we assign (limited) treatments for the best outcome?'. [Manski, 2004] introduced a framework for this analysis to economists based on *statistical decision rule* of statistics literature([Wald, 1950]).

The concern of statistical treatement rule is that how tow assign a treatment D to the population based on their covariates(features) $X \in \mathcal{X}$ to maximize utilitarian welfare U[1]. Let's call this assignment rule(function) as *policy* $\pi : \mathcal{X} \to \{0, 1\}$ and the collection of possible policies as $\Pi = \{\pi : \text{some restirctions on } \pi\}$.

It is not difficult to find this kind of problem in reality. For example, consider a Youtube's recommendation algorithm (that helps you awake all night). For a given video A, the algorithm decides whether to recommend this video to you or not based on your characteristics[2]. Now let's say Youtube has decided to recommend this video to total 100 people among its users. Then Youtube would want to find an algorithm(policy) to maximize its total viewership(utilitarian welfare) among possible policies.

How can we measure the succesfulness of a given policy? In what sense, the optimal policy can be regarded as the best? As already mentioned, we are maximizing the utilitarian welfare so the best policy could be thought as

$$\pi_{opt} = \arg\max_{\pi \in \Pi} U(\pi)$$

where $U(\pi)$ is a total welfare of population when the policy π is implemented. By defining a *regert function* $R(\pi)$ as a difference between the welfare with the policy π and the best possible outcome,

$$R(\pi) = U(\pi_{opt}) - U(\pi),$$

we can measure the successfulness of the policy π by comparing $R(\pi)$. Now our objective becomes clear: finding a policy funtion π that minimizes the functional $R(\pi)$. For a formal treatment about this setup, refer [Manski, 2004], [Kitagawa and Tetenov, 2018], [Athey and Wager, 2017], and references therein.

[1] If U is stochastic, consider the mean $E(U)$.

[2] Of course, if we set a policy as $\pi : X \to \{A, B, C, D, ...\}$, we can consider a more complexed decision makings.

Theoretical efforts has been made to bound a reget function. With a properly decided policy, we can find a uniform bound of the regret function. Notice that if a regret function degenerate fast enough, we might be willing to adopt that that policy as a solution to our decision making problem. For some results regarding a regret bound and the minimax decision criteria, refer [Stoye, 2009], [Hirano and Porter, 2009] in addtion to the references in the previous paragraph.

3 Policy Learning Embedding AI/ML Algorithms

The similar context can be found in computer science literature; *multi-armed bandit* and their close cousins. A decision of treatment rule(or policy) is becoming more of importance as many applications of the problem now have access to informative, lucrative data. Moreover, *online* experiment allows us to collect new experimental data in our specific need with less cost.

However, it is also important to study *offline* version of the problem: where historical data is available, while collecting new data is not. Still, real world experiment is not that easy in many cases, and sometimes they are not available for political, ethic reasons. This survey also focuses on this offline setups.

[Dudik et al., 2011].

AtheyWager2018

Frequentist point of view in economics literature.

4 Mathematical/Statistical Foundations

Statistical Decision Making.

As we compare policies in a policy class Π, the things we have in our consideration depend on the complexity of Π. Statistical literature offers some valuable tools to control the complexity of a class. VC-dimensions and entropy integrals.

5 Empirical Example

6 Discussion

7 Conclusion

Conclusion

Final project

References

[Athey and Wager, 2017] Athey, S. and Wager, S. (2017). Policy Learning with Observational Data. *arXiv*.

[Dudik et al., 2011] Dudik, Langford, and Li (2011). Doubly Robust Policy Evaluation and Learning.

[Hirano and Porter, 2009] Hirano, K. and Porter, J. R. (2009). Asymptotics for Statistical Treatment Rules. *Econometrica*, 77(5):1683–1701.

[Imbens and Rubin, 2015] Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

[Kitagawa and Tetenov, 2018] Kitagawa, T. and Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, 86(2):591–616.

[Manski, 2004] Manski, C. F. (2004). Statistical Treatment Rules for Heterogeneous Populations. *Econometrica*, 72(4):1221–1246.

[Stoye, 2009] Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 151(1):70–81.

[Wald, 1950] Wald, A. (1950). Statistical decision functions.