

Estimating Conditional Average Treatment Effect (CATE) using Meta-Learners in Educational Settings

Lalita Awasthi

12, 2024

Abstract

This thesis explores different meta-learning methods for estimating the Conditional Average Treatment Effect (CATE) in educational settings. Utilizing a dataset derived from the National Study of Learning Mindsets, which examines the effectiveness of growth mindset interventions among students, this study compares the performance of various meta-learners such as the T-Learner, S-Learner, X-Learner, and R-Learner. The dataset emulates an observational study, consisting of approximately 10,000 students across 76 schools, and includes a range of covariates at both the student and school levels. The research focuses on identifying heterogeneous treatment effects across student and school demographics and explores interaction effects between these factors.

Contents

1	Introduction	4
1.1	Background	4
1.2	Problem Statement	4
1.3	Objectives	4
1.4	Thesis Structure	5
2	Literature Review	6
2.1	Causal Inference in Observational Studies	6
2.2	Meta-Learners for CATE Estimation	6
2.3	Applications of CATE in Educational Settings	6
3	Methodology	7
3.1	Dataset	7
3.2	Meta-Learners	7
3.2.1	T-Learner	8
3.2.2	S-Learner	8
3.2.3	X-Learner	8
3.2.4	R-Learner	8
3.3	Evaluation Metrics	8
4	Results	9
4.1	Comparison of Meta-Learners	9
4.2	Heterogeneous Treatment Effects Across Demographics	9
4.3	Interaction Effects	9
4.4	Treatment Effect Moderation by Prior Achievement	9
5	Discussion and Conclusion	10
5.1	Summary of Findings	10
5.2	Implications for Educational Interventions	10
5.3	Limitations and Future Work	10

List of Figures

List of Tables

3.1	Covariates in the dataset	7
-----	-------------------------------------	---

Chapter 1

Introduction

1.1 Background

The effectiveness of educational interventions often varies across different student populations and school contexts. One such intervention, the growth mindset intervention, is designed to shift students' beliefs about the malleability of intelligence. The original National Study of Learning Mindsets assessed the impact of this intervention through a randomized controlled trial across U.S. public high schools. However, understanding the heterogeneity of treatment effects in an observational context requires robust statistical methods, particularly when estimating the Conditional Average Treatment Effect (CATE).

1.2 Problem Statement

This thesis aims to estimate the Conditional Average Treatment Effect (CATE) of the growth mindset intervention using different meta-learners. The dataset is structured to emulate an observational study based on covariates such as student demographics and school characteristics.

1.3 Objectives

- To compare different meta-learners (T-Learner, S-Learner, X-Learner, R-Learner) in estimating CATE.
- To identify heterogeneous treatment effects across student-level and school-level variables.
- To investigate the interaction effects between student and school covariates on the effectiveness of the intervention.

- To explore how prior student achievement moderates the effect of the intervention.

1.4 Thesis Structure

- **Chapter 1: Introduction** - Overview of the research, problem statement, and objectives.
- **Chapter 2: Literature Review** - Existing work on CATE estimation and meta-learners.
- **Chapter 3: Methodology** - Dataset, meta-learners, and methods of analysis.
- **Chapter 4: Results** - Analysis and findings from meta-learners.
- **Chapter 5: Discussion and Conclusion** - Interpretation of results, implications, and future work.

Chapter 2

Literature Review

2.1 Causal Inference in Observational Studies

Overview of causal inference techniques in educational interventions and observational studies.

2.2 Meta-Learners for CATE Estimation

- **T-Learner** - Description and common use cases.
- **S-Learner** - Strengths and limitations.
- **X-Learner** - Known for its robustness in observational data.
- **R-Learner** - Methodology and adaptability.

2.3 Applications of CATE in Educational Settings

Overview of studies applying CATE estimation to educational interventions, focusing on heterogeneous effects across demographics.

Chapter 3

Methodology

3.1 Dataset

This study uses a dataset emulating the characteristics of the National Study of Learning Mindsets. It consists of 10,000 students from 76 schools, with a continuous outcome variable measuring student achievement (\mathbf{Y}), a binary treatment variable indicating the intervention (\mathbf{Z}), and ten covariates.

Variable	Description
S3	Student’s expectations for future success (proxy for prior achievement)
C1	Student race/ethnicity (categorical)
C2	Student-identified gender (categorical)
C3	First-generation college student status (binary)
XC	Urbanicity of the school (categorical)
X1	Average fixed mindset of students before the intervention
X2	School achievement level
X3	Racial/ethnic minority composition of the school
X4	Poverty concentration in the school
X5	Total student population in the school

Table 3.1: Covariates in the dataset

3.2 Meta-Learners

Detailed description of the meta-learners used in this thesis.

3.2.1 T-Learner

3.2.2 S-Learner

3.2.3 X-Learner

3.2.4 R-Learner

3.3 Evaluation Metrics

- Mean Squared Error (MSE)
- Bias and Variance Analysis
- Confidence Intervals for CATE Estimates

Chapter 4

Results

4.1 Comparison of Meta-Learners

Analysis of the performance of T-Learner, S-Learner, X-Learner, and R-Learner in estimating CATE.

4.2 Heterogeneous Treatment Effects Across Demographics

Examine CATE estimates for different student demographics (e.g., race, gender) and school-level covariates (e.g., urbanicity, poverty concentration).

4.3 Interaction Effects

Explore interaction effects between student-level and school-level covariates.

4.4 Treatment Effect Moderation by Prior Achievement

Analyze how prior student achievement moderates the effectiveness of the growth mindset intervention.

Chapter 5

Discussion and Conclusion

5.1 Summary of Findings

Provide an overview of the key findings from the results.

5.2 Implications for Educational Interventions

Discuss the implications of CATE estimation and heterogeneous treatment effects for future educational interventions.

5.3 Limitations and Future Work

- Limitations of observational data in causal inference.
- Suggestions for future research, including more advanced methods or alternative datasets.

Bibliography