

Heterogeneous Treatment Effects, Machine Learning, and Experiment Design

Christoph Engel

1. *What is the question?*

在實驗收集來的資料中，因為受試者往往是異質性的。所以在處理數據時，資料的變異不只來自受試者本身對措施（Treatment）的變異，也會來是有不同類型的受試者所產生的變異，本文即是提出一個新方法來處理受試者個體的差異。

2. *Why should we care about it?*

傳統中有 finite mixture model 來處理上述的問題，然而使用此模型必須提供更多的資料，以及實驗者對受試者類型的數目。而且該模型很難用來估計許多實驗產生的 panel data。

3. *What is the author's answer?*

在本文用機器學習的方法結合 frequentist statistics 來估計 panel data 中類型的數目。在假設類型是受試者本身人格的變數的條件下，也就是不同的受試者會對措施做出不同的反應。就能用本文的模型來預估類型的數目。

4. *How did the author get there?*

透過兩步驟的方法，首先將每個個體在所有外生變數上跑一次迴歸。此時的迴歸係數，再來用常見的機器學習的方法，如 CART(classification and regression tree)，來將類型空間做出分割。第二，將上述所得的分類用來區分每一個受試者的類別。