Machine Learning Methods for Site-specific Input Management

Shunkei Kakimto

library(knitr)  
library(here)

## here() starts at /Users/shunkeikakimoto/OneDrive - University of Nebraska-Lincoln/ML\_VRA/GitControlled

knitr::opts\_chunk$set(  
 cache = FALSE,  
 warning = FALSE,  
 message = FALSE  
 )  
  
# opts\_knit$set(root.dir=here())

## Introduction(2 min)

### Slide 1 (11sec)

Hello, my name is Shunkei Kakimoto. I’m a graduate student in Agricultural Economics at the university of Nebraska-Lincoln. Today, I’ll be talking about a new machine learning method for site-specific input recommendation.

### Background (15sec)

The application of Machine Learning methods for site-specific economically optimal input rates, called EOIR ,have been getting more attention in recent years.

For example, Barbosa et al. (2020) applied Convolutional Neural Network methods and Krause et al. (2020) used Random Forest-based approaches.

### Research Gap (30sec)

The conventional ML methods focus on the accuracy of yield prediction, and the past studies relied on the models’ predictive power of yield for validity of their models.

However, EOIR recommendation should be based on the causal identification of input on yield, rather than yield prediction. Having good yield prediction capability does **not** necessarily mean that it is also capable of estimating EOIR well because they are **two** distinct objectives.

### New Trend of Causal Machine Learning Application (28sec)

The lack of causal interpretation from existing popular machine learning methods has been increasingly recognized in many scientific fields, and this has led to an emerging branch of machine learning methods, called causal machine learning. Causal machine learning methods focus on identifying causal impacts of an event.

In this study, we use causal forest which was developed for estimating heterogeneous causal impacts of a treatment based on observed characteristics.

### Research Questions (35sec)

Through our study, we propose two causal forest-based methods and we named them CF-stepwise and CF-base. Using those methods, we address the following two research questions.

The first question is that In terms of estimating EOIR, how do causal forest-based methods compare to other prediction-oriented ML methods?

The second question is whether the predictive ability of yield is a good indicator of the performance of EOIR estimation?

## Methods (30 sec)

To address those questions, we conducted one thousand rounds of Monte Carlo simulations under four different production scenarios, and compared the performance in estimating economically optimal nitrogen rates of causal forest -based methods to other methods.

For prediction-oriented ML methods, we investigated their yield prediction performances in contrast with their EONR estimation performances.

We use R-squared as an evaluation metric.

## Results (1:10min)

### Key Results: EONR estimation (25 sec)

Table 1 shows the mean R-squared of the estimated EONR for each ML method and modeling scenario over the 1000 simulations. CF-base showed the best accuracy. Meanwhile, the performance of CF-stepwise did not change as much as other ML methods. Part of the reason why is because the predicted treatment effects from CF-stepwise are more susceptible to prediction errors than CF-base.

### Key Results: EONR estimation and Yield Prediction

To examine the consistency between yield prediction performance and EONR estimation performance of an ML method, we counted the number of simulation rounds where a single ML method consistently showed the highest R-squared in both yield prediction and EONR estimation. Table 2 shows the result. For example, in model 1, RF attained the highest R-squared for yield prediction 363 times, and out of those simulation rounds RF attained the highest R-squared for EONR estimation 57 times. If the model with the highest yield predictive performance is likely to result in the highest EONR estimation performances as well, the two numbers should be close. However, overall there is no such a consistent relationship between these numbers.

### Key Findings (20 sec)

In conclusion, through this study, we found that our proposed CF-base method is capable of estimating site-specific EONR more accurately than other prediction-oriented ML methods, and the ability of predicting yield does not necessarily translate to good performance in estimating EONR.

Thank you for your time.

Barbosa, Alexandre, Rodrigo Trevisan, Naira Hovakimyan, and Nicolas F Martin. 2020. “Modeling Yield Response to Crop Management Using Convolutional Neural Networks.” *Computers and Electronics in Agriculture* 170: 105197.

Krause, M, Savanna Crossman, Todd DuMond, Rodman Lott, Jason Swede, Scott Arliss, Ron Robbins, Daniel Ochs, and Michael A Gore. 2020. “Random Forest Regression for Optimizing Variable Planting Rates for Corn and Soybean Using Topographical and Soil Data.”