Machine Learning Methods for Site-specific Input Management

Shunkei Kakimoto[1], Taro Mieno[1] and Takashi S. T. Tanaka[2]

[1]Agricultural Economics, University of Nebraska Lincoln [2]Applied Biological Sciences, Gifu University

September 1st 2021

Background

- The application of Machine Learning(ML) methods for site-specific economically optimal input rates (EOIR) (e.g., seed, fertilizer) have been getting more attention in recent years
  - Barbosa et al. (2020) applied Convolutional Neural Network
  - Krause et al. (2020) used Random Forest-based approaches
  - Grant Gardner (2021), Wang et al. (2021), Coulibali, Cambouris, and Parent (2020), etc.

Research Gap

- The conventional ML methods focus on predicting yield well rather than causal identification of input on yield
- The past studies have used predictive ability of yield for validity of their models.

note

- EOIR estimation should be based on the change in yields associated with the change in input levels *ceteris paribus*
- Having good yield prediction capability does not necessarily mean it is also capable of estimating EOIR well

New Trend of Causal Machine Learning Application

- Causal Machine Learning (CML) methods:
  - Unlike the conventional prediction-oriented ML methods, CML focuses on identifying causal impacts of an event (in our context, an increase or decrease in input rate for example)
- **Causal Forest (CF)** (Wager and Athey 2018; Athey and Imbens 2016):
  - CF estimates heterogeneous causal impacts of a treatment (a change in the input level) based on observed characteristics (e.g., organic matter)

Research Questions

- In terms of estimating EOIR, how do CF-based methods (CF-stepwise and CF-base) compare to other prediction-oriented ML methods: Random Forest (RF), Boosted Random Forest (BRF), and Convolutional Neural Network (CNN)?
- Is the predictive ability of yield a good indicator of the performance of EOIR estimation?

Methods

- Conduct one thousand rounds of Monte Carlo simulations under four different production scenarios
- Compare the performance in estimating economically optimal nitrogen rates (EONR) of CF-based methods (CF-base and CF-stepwise) to other methods: RF, BRF, and CNN
- For RF, BRF, and CNN, contrast their EONR performances against their yield prediction performances

Key Results: EONR estimation

Table 1: Mean R-squared of EONR Estimates by ML Methods and Modeling Scenarios

Model	RF	BRF	CNN	CF-stepwise	CF-base
Model 1	0.383	0.483	0.000	0.497	0.616
Model 2	0.344	0.485	0.000	0.478	0.600
Model 3	0.296	0.449	0.000	0.428	0.567
Model 4	0.271	0.455	0.000	0.434	0.555
Avg.	0.324	0.468	0.000	0.459	0.585

Key Results: EONR estimation and Yield Prediction

Table 2: Count table about the number of simulation rounds where an identical ML method showed the highest R-squared in both EONR estimation and yield prediction in that round

Model	RF	BRF	CNN	Total
Model 1	57 (363)	372 (432)	0 (205)	429
Model 2	45 (424)	374 (411)	0 (165)	419
Model 3	6 (85)	473 (532)	0 (383)	479
Model 4	11 (101)	527 (584)	0 (315)	538

The number in ( ) indicates the number of rounds with the highest R-squared in yield prediction.

Key Findings

- The proposed CF-base method is capable of estimating site-specific EONR more accurately than other prediction-oriented ML methods
- The ability of predicting yield does not necessarily translate to good performance in estimating EONR.

References I

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113 (27): 7353–60.

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.

Coulibali, Zonlehoua, Athyna Nancy Cambouris, and Serge-Étienne Parent. 2020. "Site-Specific Machine Learning Predictive Fertilization Models for Potato Crops in Eastern Canada." *PloS One* 15 (8): e0230888.

Grant Gardner, David S Bullock, Taro Mieno. 2021. "An Economic Evaluation of Site-Specific Input Application r x Maps: Evaluation Framework and Case Study." *Precision Agriculture*.

References II

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."

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42. https://doi.org/10.1080/01621459.2017.1319839.

Wang, Xinbing, Yuxin Miao, Rui Dong, Hainie Zha, Tingting Xia, Zhichao Chen, Krzysztof Kusnierek, Guohua Mi, Hong Sun, and Minzan Li. 2021. "Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn." *European Journal of Agronomy* 123 (May 2020): 126193. https://doi.org/10.1016/j.eja.2020.126193.