AI Recommender System With ML for Agricultural Research

Debra P. C. Peters
USDA-ARS Jornada Experimental Range

Heather M. Savoy
USDA-ARS Jornada Experimental Range

Geovany A. Ramirez
New Mexico State University

Haitao Huang
New Mexico State University

***Abstract*—We describe an AI recommender system (RS) with machine learning to harness past user choices and large volumes of data, yet account for changes in weather and management decisions characteristic of agricultural systems. Our goal is to maximize the use of data relevant to solving agricultural problems and improve the efficiency of the scientific workforce while also improving the accuracy of estimates of the amount of food produced. Our example shows how the RS learns data analysis choices from user behavior for predicting agricultural production responses to rainfall and learns to identify classes of agroecosystem responses to alternative climate scenarios. We account for changes in relationships using spatial and temporal statistics. The RS provides a powerful approach to make use of the large amounts of data and scientific expertise in the agricultural enterprise to predict agroecosystem dynamics under changing environmental conditions.**

Predicting agroecosystem dynamics at a location requires an understanding of site history and potential as well as how the climate and management decisions may change in the future.[1] High-quality data combined with scientific and local expertise are needed as part of the agricultural enterprise to understand and relate the components of the system as part of a theoretical construct.[2] Increasingly, large amounts and streams of data are becoming available to the agricultural enterprise for these predictions, yet the tools needed to synthesize, integrate, and assimilate these data under changing climate and management decisions are limited. Machine learning (ML) is a powerful class of tools as long as the training data based on historic patterns represent current or future conditions; however, novel conditions, such as extreme events or invasive species, are typically considered outliers and dropped from the analysis. Alternative approaches, such as recommender systems (RS), based on processes rather than patterns combined with ML

Digital Object Identifier 10.1109/MITP.2020.2986125
Date of current version 21 May 2020.

U.S. Government work not protected by U.S. copyright

Authorized licensed use limited to: New Mexico State University. Downloaded on August 03,2021 at 20:24:20 UTC from IEEE Xplore. Restrictions apply.

technologies are expected to be a powerful data-intensive approach in a changing environment, such as climate.

RS WITH ML

RS use a broad range of techniques to learn from past experiences to better inform current choices (e.g., Han *et al.*).[3] In general, a RS will filter all possible options available to recommend the most suitable solution for the given situation. The filtering may use training data of previous users' interactions (collaborative filtering), data attributes (content-based filtering), or a combination. In agricultural research, a RS can assist with retaining previous data analysis choices to better inform current ones. For example, a RS could accumulate a history of ML methods used for predicting response variables such that when the next user aims to predict a response variable, the RS would provide an informed suggestion on which ML method(s) could be applied. Other data analysis choices throughout the research pipeline (e.g., which explanatory variables to use) can also be informed by the RS. By attaching a specialized RS to each choice along the pipeline, an AI system with a RS foundation can be devised to guide a user through the analysis while adapting to continuous feedback from the user.

We present recent progress toward an AI system to address data analysis challenges at the Jornada Experimental Range site near Las Cruces, NM, USA (32.5 N, 106.45 W). This AI/ML RS is an operationalized, updated version of a conceptual model.[2] The system relates to learning analysis choices for modeling agricultural production responses under alternative climate scenarios following ML classification of those scenarios. Data quality and researcher efficiency will be improved, ultimately leading to reducing uncertainty in estimates of agroecosystem dynamics. An AI/ML RS is used based on prior users' decisions and historic training data, but the RS is adaptive to future choices and novel data that have not been experienced previously; thus allowing the RS to keep learning through time as climate and management decisions change.

PREDICTION OF AGRICULTURAL PRODUCTION

Estimating agroecosystem production (ANPP) using relationships based on historic ANPP data with explanatory variables, such as precipitation (PPT), are likely to over- or underestimate ANPP under nonstationary climate.[5,6] Alternative approaches are needed that allow ANPP responses to PPT to evolve over time as climate changes, such as an increase in the frequency of wet and dry periods. We are developing an AI system to: 1) identify discrete changes in PPT patterns, such as breakpoint analysis of wet and dry periods;[7] 2) develop within-period relationships between ANPP and PPT or other explanatory variables; and 3) apply this learned behavior to other ecosystems or research sites (see Figure 1). Our RS is a hybrid system that uses collaborative filtering and content-based filtering in conjunction with ML algorithms. Users interact with the system via a web-interface attached to a repository of R scripts implementing the RS and ML methods, and to an extensive georeferenced database of agricultural, environmental, and climate data.

Our AI system was initially trained on historic data from 1980–2008 where a sequence of steps was followed to: *first* divide PPT into a drought (2000–2003), wet period (2004–2008), and no-trend period (1990–1999), *second* conduct regressions between ANPP and PPT for each period using alternative forms (linear, exponential, other nonlinear models); recommend the best model to the user, and allow the user to select the model to be used; in the case of drought and no-trend, a linear model was selected; for wet years, the RS was unable to fit a line to the points, *third* consult the user for additional information for the wet period; based on user input, conduct regressions between ANPP and time for the wet period, and recommend an exponential model as the best model to the user. *Fourth* recommend additional explanatory variables to user, such as soil texture and temperature, through access to a colocated database. All user choices (e.g., the choice of explanatory variable(s), data transformations, or resolutions of data) are recorded as feedback to the RS. As the system learns, it can recommend how to approach similar questions at different locations (e.g., other shrublands). *Fifth* system can be revisited as new climate data are added through time.

This system is currently being developed to predict patterns in ANPP for different ecosystems at the Jornada. Ultimately, it will be available to users at other research sites.

CONCLUSION

A RS with ML provides a powerful data-intensive approach with user input based on processes rather than historic patterns to handle changing environmental conditions in agroecosystems.

Authorized licensed use limited to: New Mexico State University. Downloaded on August 03,2021 at 20:24:20 UTC from IEEE Xplore. Restrictions apply.

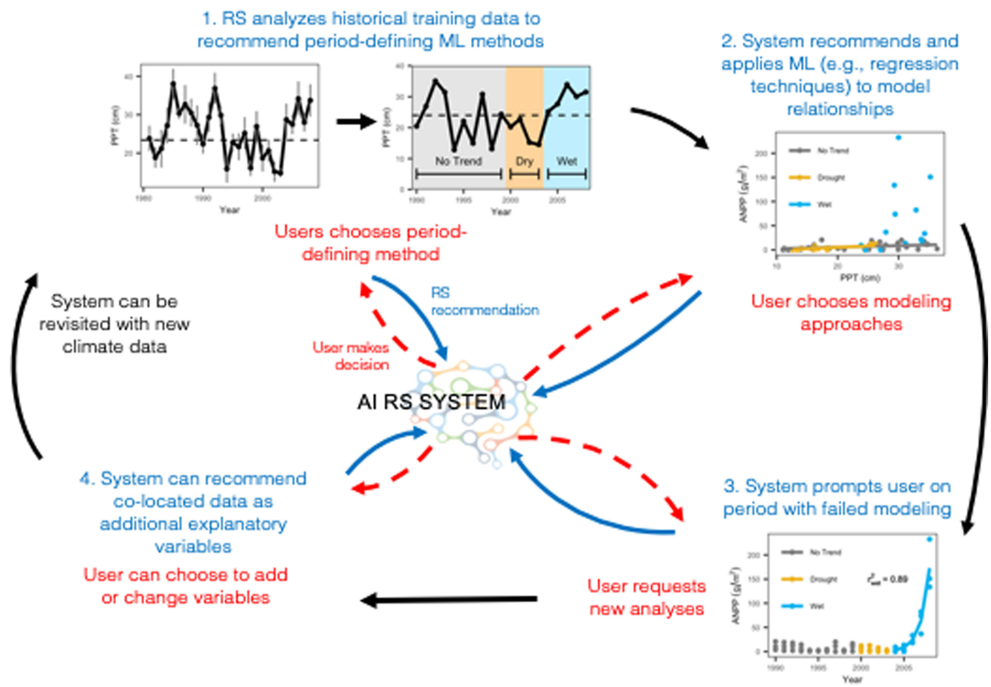

Figure 1. RS with ML to predict agricultural production under changing climate.

ACKNOWLEDGMENTS

This work was supported by the USDA-ARS CRIS Project at the Jornada Experimental Range (#6235-11210-007), and by the National Science Foundation to New Mexico State University for the Jornada Basin Long Term Ecological Research Program (DEB 18-32194).

REFERENCES

1. J. L. Hatfield and C. L. Walthall, “Meeting global food needs: Realizing the potential via genetics x environment x management interactions,” *Agronomy J.*, vol. 107, no. 4, pp. 1215–1226, 2015.
2. D. P. C. Peters, K. M. Havstad, J. Cushing, C. Tweedie, O. Fuentes, and N. Villanueva-Rosales, “Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology,” *Ecosphere*, vol. 5, no. 6, pp. 1–15, 2014.
3. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2012.
4. D. R. Vincent, N. Deepa, D. Elavarasan, K. Srinivasan, S. H. Chauhdary, and C. Iwendi, “Sensors driven AI-based agriculture recommendation model for assessing land suitability,” *Sensors*, vol. 19, pp. 1–16, 2019.
5. P. C. D. Milly *et al.*, “Stationarity is dead: Whither water management?” *Science*, vol. 319, pp. 573–574, 2008.
6. D. P. C. Peters, J. Yao, O. E. Sala, and J. P. Anderson, “Directional climate change and potential reversal of desertification in arid and semiarid ecosystems,” *Global Change Biol.*, vol. 18, no. 1, pp. 151–163, Jan. 2012.
7. B. T. Bestelmeyer *et al.*, “Analysis of abrupt transitions in ecological systems,” *Ecosphere*, vol. 2, no. 12, pp. 1–26, 2011.

Debra P.C. Peters is the Acting Chief Science Information Officer with the USDA Agricultural Research Service in Beltsville, MD, USA, and a Research Scientist with the USDA-ARS Jornada Experimental Range, Las Cruces, NM, USA. Contact her at deb.peters@usda.gov.

Heather M. Savoy is a Computational Ecologist with the USDA-ARS Jornada Experimental Range, Las Cruces, NM, USA. Contact her at heather.savoy@usda.gov.

Geovany A. Ramirez is a Data Scientist with the Jornada Basin Long-Term Ecological Research Program, New Mexico State University, Las Cruces, NM, USA. Contact him at garamirez.jrn.lter@gmail.com.

Haitao Huang is a Computer Programmer with the Jornada Basin Long-Term Ecological Research Program, New Mexico State University, Las Cruces, NM, USA. Contact him at haitaohuang@hotmail.com.

Authorized licensed use limited to: New Mexico State University. Downloaded on August 03,2021 at 20:24:20 UTC from IEEE Xplore. Restrictions apply.