ARTICLE TOOLS

[printer icon] [Print this article](#)

[info icon]

[Indexing metadata](#)

[citation icon] [How to cite item](#)

[envelope icon] Email this article (Login required)

[envelope icon] Email the author (Login required)

[OPEN JOURNAL SYSTEMS](#)

[Journal Help](#)

[SUBMITTING A MANUSCRIPT](#)

USER

Username ⬚

Password ⬚

☐ Remember me

Login

LANGUAGE

English
[Español](#)

INFORMATION

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

JOURNAL CONTENT

Search

Search

Browse

- By Issue
- By Author
- By Title
- Other Journals

Predictive ability of machine learning methods for massive crop yield prediction

Alberto Gonzalez-Sanchez, Juan Frausto-Solis, Waldo Ojeda-Bustamante

Abstract

An important issue for agricultural planning purposes is the accurate yield estimation for the numerous crops involved in the planning. Machine learning (ML) is an essential approach for achieving practical and effective solutions for this problem. Many comparisons of ML methods for yield prediction have been made, seeking for the most accurate technique. Generally, the number of evaluated crops and techniques is too low and does not provide enough information for agricultural planning purposes. This paper compares the predictive accuracy of ML and linear regression techniques for crop yield prediction in ten crop datasets. Multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression and k-nearest neighbor methods were ranked. Four accuracy metrics were used to validate the models: the root mean square error (RMS), root relative square error (RRSE), normalized mean absolute error (MAE), and correlation factor (R). Real data of an irrigation zone of Mexico were used for building the models. Models were tested with samples of two consecutive years. The results show that M5-Prime and k-nearest neighbor techniques obtain the lowest average RMSE errors (5.14 and 4.91), the lowest RRSE errors (79.46% and 79.78%), the lowest average MAE errors (18.12% and 19.42%), and the highest average correlation factors (0.41 and 0.42). Since M5-Prime achieves the largest number of crop yield models with the lowest errors, it is a very

suitable tool for massive crop yield prediction in agricultural planning.

Full Text:
PDF

References

Breiman L, 2001. Statistical modeling: the two cultures (with discussion). Statist Sci 16: 199-231. http://dx.doi.org/10.1214/ss/1009213726

Breiman L, Friedman JH, Olshen RA, Stone CJ, 1984. Classification and regression trees. Wadsworth, Belmont, CA, USA.

Brisson N, Marry B, Ripoche D, Jeuffory MH, Ruget F, Nicoullaud B, Gate P, Devienne BF, Antonioletti R, Durr C et al., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balance. 1. Theory and parameterization applied to wheat and corn. Agronomie 18: 311-346. http://dx.doi.org/10.1051/agro:19980501

Dixon BL, Hollinger SE, Garcia P, Tirupattur V, 1994. Estimating corn yield response models to predict impacts of climate change. J Agr Resour Econ 19(1): 58-68.

Drummond ST, Sudduth KA, Joshi A, Birrel SJ, Kitchen NR, 2003. Statistical and neural methods for site-specific yield prediction. T ASABE 46 (1): 5-14.

Fortin JG, Anctil F, Parent L, Bolinder MA, 2011. Site-specific early season potato yield forecast by neural network in Eastern Canada. Precis Agr 12(6): 905-923.

http://dx.doi.org/10.1007/s11119-011-9233-6

Frausto-Solís J, Gonzalez-Sanchez A, Larre M., 2009. A new method for optimal cropping pattern. Proc. 8th Mex Int Conf on Artificial Intelligence, pp: 566-577.

Goudriaan J, van Laar H, 1994. Modelling potential crop growth processes. Kluwer Acad. Publ., Dordrecht, the Netherlands. http://dx.doi.org/10.1007/978-94-011-0750-1

Hair JF Jr, Anderson RE, Tatham RL, 1987. Multivariate data analysis, 2nd edition. MacMillan Publ. Co., Han J, Kamber M, 2006. Data mining: concepts and techniques, 2nd ed. Morgan Kaufmann Publ.

Hand D, Mannila H, Smyth P, 2001. Principles of data mining. MIT Press.

Irmak A, Jones JW, Batchelor WD, Irmak S, Boote KJ, Paz JO, 2006. Artificial neural network model as a data analysis tool in precision farming. T ASABE 49(6): 2027-2037. http://dx.doi.org/10.13031/2013.22264

Jaikla R, Auephanwiriyakul S, Jintrawet A, 2008. Rice yield prediction using a support vector regression method. ECTI-CON 5th Int Conf, Vol. 2, pp: 29-32.

Jamieson PD, Semenov MA, Brooking IR, Francis GS, 1998a. Sirius: a mechanistic model of wheat response to environmental variation. Eur J Agron 8: 161-179. http://dx.doi.org/10.1016/S1161-0301(98)00020-3

Jamieson PD, Porter JR, Goudriaan J, Ritchie JT, van Keulen H, Stol W, 1998b. A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. Field Crops Res 55: 23-44. http://dx.doi.org/10.1016/S0378-4290(97)00060-9

Jones CA, Kiniry JR, 1986. CERES-Maize: A simulation model of maize growth and development. Texas A&M Univ. Press, College Station, Texas, USA. 194 pp.

Kohavi R, 1995. Wrappers for performance enhancement and oblivious decision graphs. Doctoral dissertation, Stanford Univ., Comp. Sci. Dept.

Liu J, Goering CE, Tian L, 2001. Neural network for setting target corn yields. T ASAE 44(3): 705-713.

Marinković B, Crnobarac J, Brdar S, Antić B, Jaćimović G, Crnojević V, 2009. Data mining approach for predictive modeling of agricultural yield data. Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment (BioSense09), Novi Sad, Serbia, October, pp: 1-5.

McQueen RJ, Garner SR, Nevill-Manning CG, Witten IH, 1995. Applying machine learning to agricultural data. Comput Electron Agr 12(4): 275-293. http://dx.doi.org/10.1016/0168-1699(95)98601-9

Ojeda-Bustamante W, González-Camacho JM, Sifuentes-Ibarra E, Isidro E, Rendón-Pimentel L, 2007. Using spatial information systems to improve water management in Mexico. Agr Water Manage 89: 81-88. http://dx.doi.org/10.1016/j.agwat.2006.11.002

Porter JR, 1993. AFRCWHEAT2: a model of the growth and development of wheat incorporating responses to water and nitrogen. Eur J Agron 2: 69-82.

Quinlan JR, 1992. Learning with continuous classes. Proc. AI'92, 5th Aust. Joint Conf. on Artificial Intelligence (Adams & Sterling, eds.), World Scientific, Singapore, pp: 343-348.

Roel A, Plant RE, 2004. Factors underlying yield variability in two California rice fields. Agron J 96: 1481-1494. http://dx.doi.org/10.2134/agronj2004.1481

Rojas R, 1996. Neural networks - A systematic introduction. Springer-Verlag, Berlin, NY.

Rumelhart DE, Hinton GE, Williams RJ, 1986. Learning internal representations by error propagation. In:

Parallel distributed processing: explorations in the microstructure of cognition, (Rumelhart DE, McClelland JA, eds), vol. 1, chapter 8. The MIT Press, Cambridge, MA (USA). pp: 418-362.

Ruß G, 2009. Data mining of agricultural yield data: a comparison of regression models. Proc. 9th Indust. Conf. on Advances in Data Mining-Applications and Theoretical Aspects, July 20-22, Leipzig, Germany. http://dx.doi.org/10.1007/978-3-642-03067-3_3

Ruß G, Kruse R, 2010. Feature selection for wheat yield prediction. In: Research and development in intelligent systems XXVI (Bramer M et al., eds.), Springer - Verlag, London. http://dx.doi.org/10.1007/978-1-84882-983-1_36

Safa B, Khalili A, Teshnehlab M, Liaghat A, 2004. Artificial neural networks application to predict wheat yield using climatic data. Proc. 20th Int. Conf. on IIPS, Jan. 10-15, Iranian Meteorological Organization, pp: 1-39.

Schlenker W, Roberts MJ, 2006. Estimating the impact of climate change on crop yields: The importance of non-linear temperature effects. Discussion Papers 0607-01, Columbia University, Dept. Economics.

Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK, 2000. Improvements to the SMO algorithm for SVM regression. IEEE Transactions on Neural Networks 11(5): 1188-1193. http://dx.doi.org/10.1109/72.870050

Smola A, Schölkopf B, 2004. A tutorial on support vector regression. Stat Comput 14(3): 199-222. http://dx.doi.org/10.1023/B:STCO.0000035301.49549.

Spitters CJT, van Keulen H, van Kraailingen DWG, 1988. A simple but universal crop growth simulation model, SUCROS87. In: Simulation and systems management in crop protection (Rabbinge R, Van Laar H & Ward S, eds). PUDOC, Wageningen. pp: 87-98.

Sudduth KA, Drummond ST, Birrell SJ, Kitchen NR, 1996. Analysis of spatial factors influencing crop yield. Proc. 3rd Int. Conf. on Precision Agriculture (Robert PC, Rust RH, & Larson WE, eds.) ASA-CSSA-SSSA, Madison, WI, USA, pp: 129-140.

Sudduth K, Fraisse C, Drummond S, Kitchen N, 1998. Integrating spatial data collection, modeling and analysis for precision agriculture. First Int. Conf. on Geospatial Information in Agriculture and Forestry, vol. 2, pp: 166-173.

Uysal I, Altay HG, 1999. An overview of regression techniques for knowledge discovery. Knowl Eng Rev 14: 319-340.
http://dx.doi.org/10.1017/S026988899900404X

Vapnik V, Lerner A, 1963. Pattern recognition using generalized portrait method. Automat Remote Contr 24: 774-780.

Vapnik V, Golowich S, Smola A, 1997. Support vector method for function approximation, regression estimation, and signal processing. In: Advances in neural information processing systems (Mozer M, Jordan M, & Petsche T, eds), MIT Press, Cambridge, MA, USA, pp: 281-287.

Varcoe VJ, 1990. A note on the computer simulation of crop growth in agricultural land evaluation. Soil Use Manage 6(3): 157-160.
http://dx.doi.org/10.1111/j.1475-2743.1990.tb00826.x

Wang Y, Witten I, 1997. Inducing model trees for continuous classes. Proc. 9th Eur. Conf. Machine Learning (van Someren M & Widmer G, eds), pp: 128-137.

Wasserman L, 2004. All of statistics. A concise course in statistical inference. Springer.

Wilkerson GG, Jones JW, Boote KJ, Ingram KT, Mishoe

JW, 1983. Modeling soybean growth for crop management. T ASAE 26: 63-73.
http://dx.doi.org/10.13031/2013.33877

Witten IH, Frank E, Trigg L, Hall M, Holmes G, Cunningham SJ, 1999. Weka: Practical machine learning tools and techniques with Java implementations. ICONIP/ANZIIS/ANNES'99 Int. Workshop (Kasabov H & Ko K, eds), Dunedin.

Yang Y, 2008. Consistency of cross validation for comparing regression procedures. Ann Stat 35 (6): 2450-2473.
http://dx.doi.org/10.1214/009053607000000514

Zhang L, Zhang J, Kyei-Boahen S, Zhang M, 2010 Simulation and prediction of soybean growth and development under field conditions. Am-Euras J Agr Environ Sci 7(4): 374-385.