Learning from Data to Optimize Control in Precision Farming

Alexander Kocian [1],* and Luca Incrocci [2]

[1] Department of Computer Science, University of Pisa, 56127 Pisa, Italy
[2] Department of Agriculture, Food and Environment, University of Pisa, 56124 Pisa, Italy;
    Luca.incrocci@unipi.it
*   Correspondence: kocian@di.unipi.it

Abstract: Precision farming is one way of many to meet a 55 percent increase in global demand for agricultural products on current agricultural land by 2050 at reduced need of fertilizers and efficient use of water resources. The catalyst for the emergence of precision farming has been satellite positioning and navigation followed by Internet-of-Things, generating vast information that can be used to optimize farming processes in real-time. Statistical tools from data mining, predictive modeling, and machine learning analyze patterns in historical data, to make predictions about future events as well as intelligent actions. This special issue presents the latest development in statistical inference, machine learning, and optimum control for precision farming.

Keywords: statistics; precision agriculture; IoT; machine learning; reinforcement learning; water; production; soil; predictive analytics

1. Introduction

The world's population is expected to be nearly 10 billion by 2050, corresponding to a 55 percent increase in global demand for agricultural production based on current trends. In 2011, according to FAO, agriculture made use of 2710 km^3 (70 percent) of all water withdrawn from aquifers, stream and lakes, but this number masks large geographical discrepancies. Middle East, Northern Africa, and Central Asia have already withdrawn most of the exploitable water with 80–90 percent of that going to agriculture. Hence, rivers and aquifers are depleted beyond sustainable levels [1]. Shifting the focus to arable land, 1.6 billion hectares are arable worldwide. The total world land area suitable for cropping is 4.4 billion corresponding to around 40 percent of world's land. However, in several regions, soil quality constraints affect more than half the cultivated land base, notably in sub-Saharan Africa, Southern America, Southeast Asia, and Northern Europe [1]. When forests are converted into farming land, the largest stores of carbon locked in those trees will be released to the atmosphere, contributing to global warming on top of today's level.

Clearly, crop production on current land needs to be increased through adopting new technologies. To increase profits reduces waste and maintains environmental quality at the same time. Farmers are supplied with decision support systems that propose the right dose/action at the right place and at the right time [2,3]. The core piece of such decision support system is an agricultural model related to either crop growth, epidemiology, or market development that optimizes a control function based on probabilistic assessment of causal relationships [4]. Satellite telemetry tracking data along with existing geo-referenced digital map as well as Internet-of-Things based sensor data act as input to the model. Automated data processing systems, often located in the cloud, train the model. The trend goes from manually trained to self-calibrating models that adapt to changes in the environment over time. Smartphone applications have become a key interface in precision agriculture between the farmer

and the cloud. These applications not only visualize the control parameters and suggest possible actions but also return the farmers' reaction (irrigation, sowing, fertilization, etc.) back to the cloud. Fully automatized actions that go beyond human level performance while minimizing resources are still subject to research.

2. Statistical Inference and Machine Learning

The key to effective experimentation in precision farming is blocking, replication, and randomization [5]. To analyze and interpret the experimental results as well as to predict upcoming data, tools from statistics are deployed. Probabilistic models approximate the complex dynamics of the underlying process using statistical assumptions on the generation of sample data. Statistics draws population inferences from data samples. Neither training nor test sets are necessary to infer the parameters. The supervised machine learns from training data to build a statistical model that can be used to make repeatable predictions. The unsupervised machine, in contrast, learns the model on its own without external training data. With the development of Internet-of-Things, machine learning applications for precision farming have been rapidly developing over the last years [6].

2.1. Low-Order Statistics

Random variables have a discrete or continuous probability distribution. Low-order statistics denote the first and second moments of a sample from the distribution. The former and the second correspond to the mean and the statistical auto- and cross-power of the random variables. Low order statistics, however, require a very large number of samples to estimate with a reasonable level of confidence. When the random variables are normal distributed, this now ranked data are often used for *ANalysis Of VAriance* (ANOVA), comparing the ratio of within group variance and between group variance, to assess systematic factors (bias) and random factors (covariance). The former has statistical influence on the data set while the latter does not. For example, there is an average weight variation within one kind of pumpkin, but there might be another average weight variation among different pumpkin varieties. The *Pearson correlation coefficient* defines as ratio of covariance to the product of individual variances measures linear correlation between two random variables. For example, the Pearson correlation between evapotranspiration and precipitation is positive over the southern/deforested but negative over the northern/forested Amazonia [7].

2.2. Regression

Multiple regression models characterize the relationship between a dependent target variable and multiple weighted independent feature variables. The weights, also known as regression coefficients, are an average functional relationship between target and features which might be linear or nonlinear. For example, an exponential regression is adequate to model the relation between tree height and leaf-area index of Prunus [8]. The least square fitting technique yields the model parameters. A probit regression, in contrast, considers binary target variables with Gaussian distributed model noise and possibly multiple weighted independent variables. The maximum likelihood technique is often used to obtain the model parameters. Voting with binary outcome is a typical application of probit regression. For example, Sevier and Lee used this method in [9] to predict the probability of Florida citrus producers adopting precision agriculture technologies. Note that regression analysis is sensitive to multicollinearity, arising whenever two or more independent variables used in a regression are strongly correlated with each other. In this case, the weights become very sensitive to small changes in the model.

2.3. Classification

Classification is a supervised learning problem like the above regression. Considering models for solving classification problems, the classical Fisher linear discriminant analysis is a standard multivariate technique both for dimension reduction and supervised classification. The data vectors

are transformed into a low dimensional subspace that maximizes separation of class centroids. In many applications, however, the linear boundaries do not adequately separate the classes. Roth and Steinhage present in [10] a nonlinear generalization of discriminant analysis that uses the kernel trick to replace dot products with an equivalent kernel function.

Sparse Kernel Machines evaluate the kernel function only at a subset of the training data points to predict a new data point, making the computation time feasible [11]. Specifically, the support vector machine (SVM) by Boser et al. in [12] discards all data points but the support vectors, once the model is trained. The determination of the model parameters is a convex optimization problem so that any local solution is also a global solution in contrast to many other algorithms. The SVM has become popular for solving problems in classification, regression, and novelty detection. For example, Jheng et al. predicted in [13] the rice yield in Taiwan by a SVM using training data from 1995–2015. The relevance vector machine (RVM) [14] is a Bayesian sparse kernel technique that provides posterior probability outputs in contrast to the SVM. At the same time, RVM based prediction models utilize dramatically less basis functions than a comparable SVM. To name an example from remote sensing, the RVM with plate spline kernel is able to spatially estimate chlorophyll from an unmanned aerial system at low computational cost [15]. Finally, we want to point out the Informative Vector Machine (IVM), constructing sparse Gaussian process classifiers by greedy forward selection with criteria based on information theoretic principles. The IVM performs similar to the SVM by only a fraction of training data. Roscher et al. use in [16] an incremental version of the IVM to classify hyperspectral image data for various agricultural crops in Italy, Europe, and Indiana, USA.

2.4. Clustering

Clustering is an unsupervised process of partitioning a set of data (or objects) in a set of meaningful sub-classes, called clusters. Clustering techniques can be categorized into (i) partitioning algorithms constructing various partitions and then evaluate the result by some criterion (k-means, k-medoids, CLARANS, ...); (ii) hierarchical algorithms creating a hierarchical decomposition of the set of objects by some criterion (AGNES, BIRCH, CURE, DIANA, ...); (iii) density-based methods that are guided by connectivity and density functions (DBSCAN, OPTICS, ...); (iv) grid-based methods that are based on a multi-level granularity structure (STING, WaveCluster, CLIQUE, ...); and (v) model-based clustering methods that find the best fit to a hypothetical model (Autoclass, Rock, EM-algorithm, ...). Massive computing power makes it possible, for example, to mine a large amount of existing crop, soil, and climatic data. Clustering the result based on districts with maximum wheat yield gives the optimal range of best temperature, worst temperature, and rain fall [17]. To scale clustering algorithms with the number of dimensions and the number of data items, attention has been drawn to a distributed approach [18]. Nevertheless, the scaling problem is still a challenge for most of the above clustering algorithms such as big data applications.

2.5. Artificial Neural Networks

An Artificial Neural Networks (ANNs) consist of many simple connected nodes dubbed neurons, each deploying a real-valued nonlinear activation function. Input neurons are activated by data from external sensors. Other neurons are activated by weighted edges from previously active neurons. Feed-forward neural networks, forming a directed acyclic graph, process the sensed data without memory. In contrast, the recurrent neural network (RNN) allows connections among neurons in the same or previous layers. They have internal memory and their graph is directed with cycles. When fed with environmental and historical dynamic information, this type of neural network is well-suited to time series forecasting [19]. In the convolutional neural network (CNN), forward and backward propagations perform convolutional operations. Usually, the edge weights are point estimates based on stochastic gradient training. Bayesian Neural Networks model the uncertainty of the estimated edge weights by interpreting them as maximum likelihood or maximum a posteriori estimates. A comprehensive state-of-the-art overview of ANN is available in [20]. Notable examples

in precision farming are the feedforward neural network by Adisa et al. in [21] for maize production prediction. In this work, the feature space is spanned by the environmental parameters, potential evapotranspiration, soil moisture, and land cultivated. Barbosa et al. deployed in [22] a CNN that predicts the spatial yield map of corn fields in Illinois, Nebraska, and Kansas, USA. Here, satellite images as well as environmental data span the feature space. In a third notable application, multi-layer (deep) CNN have been applied in [23] to detect plant leaf disease based on 54,000 (large number) training images. Finally, we want to point out the example in [24] where RNN has been used for spatio-temporal prediction of leaf area index in rubber plantation. The feature space in the experiment was spanned by the individual CCD images. The underlying theory of many neural network architectures is still in research phase.

2.6. Bayesian Time-Series Forecasting

Bayesian time-series forecasting is another promising field of research in precision farming. Within this framework, all sources of uncertainty are expressed by stochastic processes. The Bayes Theorem turns the a priori probability and the distribution of the observed data, also known as likelihood, into the posteriori distribution of the parameters for predictive inference. A partially observed state-space model such as the Hidden Markov Model (for discrete states) or the Kalman filter (for continuous states) are ideally suited to describe the dynamics of the process. A typical example in agriculture research is price prediction of crops. In [25], a Kalman filter has been deployed to predict the price time-series of rice. When the model parameters are unknown, the observation sequence and the state sequence can be used to estimate them. The linear dynamic Bayesian network developed in [26] does this by relating indicative parameters of crop development to environmental control parameters. The expectation–maximization algorithm is used to track the states in the expectation step and to learn the parameters of the Bayesian network in the maximization. At iterative convergence, the algorithm provides a time-series predictor many time instants ahead. When the dynamics is nonlinear on top of that, sequential Monte Carlo techniques often lead to accurate parameter predictions by sampling from the posterior distribution on the expenses of computational complexity. In the special case of sigmoid-type growth dynamics, a linear dynamic model leads to the exact predictor for the reciprocal (non-linear) time-series of the parameter [27].

3. Closing the Loop

Thus, far, machines have mostly be used to learn from the observations with the goal to predict future outcome given current conditions. Clearly with an increasing number of observations, the machine becomes smarter over time, but it does not have control over the environmental conditions. Currently, these are controlled by the agronomist's experience. A more efficient approach is to let agents make optimal actions subject to minimizing resources. The result is a close-loop precision farming system where the model learns from data in the forward loop and controls actuators in the backward loop, as outlined in [28]. Reinforcement learning, making smarter decisions over time, has enjoyed a great success in several domains such as computer game, medical diagnosis, and energy management. Bu and Wang build in [29] a smart agriculture IoT system based on deep reinforcement learning that decides the amount of water needed to be irrigated by analyzing the collected agricultural environment data. Though there had been great progress, the technology cannot yet achieve the human-level performance in adaptation to dynamic environments and solving complex tasks. Ergo, there is still a lot of space for research towards optimum precision farming. Table 1 lists strengths and weaknesses of common statistical models and machine learning algorithms.

Table 1. Comparison of common statistical models and machine learning algorithms.

Method	Strengths	Weaknesses
MANOVA	- Powerful test for finding truly significant factors. - Robust to Type I errors.	- Relation between independent grouping variable and dependent variables sometime ambiguous. - Computationally complex.
Multiple Regression	- Theory well understood. - Good results are obtained with relatively small data sets. - Ability to determine impact of independent variable on dependent variable.	- Missing data erroneously changes regression coefficients. - Correlation does not necessarily correspond to a causation. - Sensitive to outliers.
Deep Neural Networks	- Perform well on audio, image, text data. - Architecture can be adapted to a number of problems.	- Computationally intensive to train. - Tuning hyper-parameters needs expert knowledge.
Dynamic Bayesian Network	- Accurate prediction of temporal behavior. - Flexible adapts to environmental changes. - Underlying theory is well understood.	- Cannot handle real biological systems with feedback loops (cycles). - Initial guess of parameters is crucial for convergence.
Support Vector Machine	- Memory efficient. - Flexible (nonlinear) threshold using Kernels. - Convex optimization problem with unique solution.	- Does not scale with data dimension. - Sensitive to tuning the regularization parameters (overfitting). - Finding a proper kernel is often cumbersome.
k-means clustering	fast, simple.	Model order must be known in advance.
DBSCAN clustering	- Model-order free. - Scalable. - Estimate is unbiased.	- Sensitive to choice of hyperparameters. - Good results only for uniform densities.
Reinforcement Q-Learning	- Computes most successful rewards even when the environment is large. - Model-free. - Convergence to the optimum policy is guaranteed.	- Computationally complex. - Assumes that all of the states and all of the actions are presentable as matrix.

4. Conclusions

Precision farming for current arable land is a promising approach to meet the vast global demand for agricultural products on current land. Internet-of-Things provides vast real-time information on crop related parameters, soil, and weather that feeds machine learning algorithms for better crop productivity while protecting the environment. The ultimate goal is to maximize yield by minimizing water consumption, usage of fertilizers, and amount of arable land in an automatic fashion. Although there has been an evolution of research in this area, more knowledge is needed to close the gap between current practice and optimum precision farming.

References

1. FAO. *The State of the World's Land and Water Resources for Food and Agriculture. How Should Agriculture Produce Enough Food for the World*; FAO: Rome, Italy, 2011.

2. Gallardo, M.; Elia, A.; Thompson, R.B. Decision support systems and models for aiding irrigation and nutrient management of vegetable crops. *Agric. Water Manag.* **2020**. [CrossRef]

3. Thompson, R.B.; Incrocci, L.; van Ruijven, J.; Massa, D. Reducing contamination of water bodies from European vegetable production systems. *Agric. Water Manag.* **2020**. [CrossRef]

4. Shi, X.; An, X.; Zhao, Q.; Liu, H.; Xia, L.; Sun, X.; Guo, Y. State-of-the-Art Internet of Things in Protected Agriculture. *Sensors* **2019**, *19*, 1833. [CrossRef] [PubMed]

5. Young, J.C. Blocking, Replication, and Randomization—The key to effective experimentation. *Qual. Eng.* **1996**, *9*, 269–277. [CrossRef]

6. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]

7. Vergopolan, N.; Fisher, J.B. The impact of deforestation on the hydrological cycle in Amazonia as observed from remote sensing. *Int. J. Remote Sens.* **2016**, *37*, 5412–5430. [CrossRef]

8. Pardossi, A.; Incrocci, L.; Incrocci, G.; Tognoni, F.; Marzialetti, P. What Limits and How to Improve Water Use Efficiency in Outdoor Container Cultivation of Ornamental Nursery Stocks. *ISHS Acta Hortic.* **2009**, 73–80. [CrossRef]

9. Sevier, B.J.; Lee, W.S. Precision Farming Adoption by Florida Citrus Producers: Probit Model Analysis. In Proceedings of the ASABE Annual Meeting, Ottawa, ON, Canada, 1–4 August 2004; Paper No. 041080.

10. Roth, V.; Steinhage, V. Nonlinear Discriminant Analysis using Kernel Functions. In Proceedings of the Advances in Neural Information Processing Systems, Denver, CO, USA, 29 November–4 December 1999; pp. 568–574.

11. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer Science + Business Media LLC: Berlin/Heidelberg, Germany, 2009.

12. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the 5th ACM Annual Workshop on Computation Learning Theory (COLT'92), Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.

13. Jheng, T.Z.; Li, T.H.; Lee, C.P. Using Hybrid Support Vector Regression to Predict Agricultural Output. In Proceedings of the 27th Wireless and Optical Communication Conference (WOCC 2018), Hualien, Taiwan, 30 April–1 May 2018.

14. Tipping, M.E. The Relvance Vector Machine. *Adv. Neural Inf. Process. Syst.* **2000**, *12*, 652–658.

15. Elarab, M.; Ticlavilca, A.; Torres-Rua, A.; Maslova, I.; McKee, M. Estimating chlorophyll with thermal and broadband multispectral high resolution imagery from an unmanned aerial system using relevance vector machines for precision agriculture. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 32–42. [CrossRef]

16. Roscher, R.; Waske, B.; Forstner, W. Incremental Import Vector Machines for Classifying Hyperspectral Data. *IEEE Trans. Geosci. Remote. Sens.* **2012**, *50*, 3463–3473. [CrossRef]

17. Majumdar, J.; Naraseeyappa, S.; Ankalaki, S. Analysis of agriculture data using data mining techniques: application of big data. *J. Big Data* **2017**, *4*, 20. [CrossRef]

18. Hore, P.; Hall, L.O. Scalable Clustering: A distributed approach. In Proceedings of the IEEE International Conference on Fuzzy Systems, Budapest, Hungary, 25–29 July 2004; pp. 143–148.

19. Balducci, F.; Impedovo, D.; Pirlo, G. Machine learning applications on agricultural datasets for smart farm enhancement. *Machines* **2018**, *6*, 38. [CrossRef]

20. Schmidhuber, J. *Deep Learning in Neural Networks: An Overview*; Technical Report IDSIA-03-14; University of Lugano & SUPSI: Manno, Switzerland, 2014.

21. Adisa, O.; Botai, J.; Adeola, A.; Hassen, A.; Botai, C.; Darkey, D.; Tesfamariam, E. Application of Artificial Neural Network for Predicting Maize Production in South Africa. *Sustainability* **2019**, *11*, 1145. [CrossRef]

22. Barbosa, A.; Trevisan, R.; Hovakimyan, N.; Martin, N.F. Modeling yield response to crop management using convolutional neural networks. *Comput. Electron. Agric.* **2020**, *170*, 105197. [CrossRef]

23. Bapat, A.; Sabut, S.; Vizhi, K. Plant Leaf Disease Detection using Deep Learning. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 3599–3605.

24. Chen, B.; Wu, Z.; Wang, J.; Dong, J.; Guan, L.; Chen, J.; Yang, K.; Xie, G. Spatio-temporal prediction of leaf area index of rubber plantation using HJ-1A/1B CCD images and recurrent neural network. *ISPRS J. Photogramm. Remote. Sens.* **2015**, *102*, 148–160. [CrossRef]

25. Cenas, P.V. Forecast of Agricultural Crop Price using Time Series and Kalman Filter Method. *Asia Pac. J. Multidiscip. Res.* **2017**, *5*, 15–21.

26. Kocian, A.; Massa, D.; Cannazzaro, S.; Incrocci, L.; Lonardo, S.D.; Milazzo, P.; Chessa, S. Dynamic Bayesian Network for Crop Growth Prediction in Greenhouses. *Comput. Electron. Agric.* **2020**. [CrossRef]

27. Kocian, A.; Carmassi, G.; Fatjon, C.; Incrocci, L.; Milazzo, P.; Chessa, S. Bayesian sigmoid-type time series forecasting with missing data for greenhouse crops. *Sensors* **2020**, *20*, 3246. [CrossRef] [PubMed]

28. Burchi, G.; Chessa, S.; Gambineri, F.; Kocian, A.; Massa, D.; Milano, P.; Milazzo, P.; Rimediotti, L.; Ruggeri, A. Information Technology Controlled Greenhouse: A System Architecture. In Proceedings of the IoT Vertical and Topical Summit for Agriculture, Tuscany, Italy, 8–9 May 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]

29. Bu, F.; Wang, X. A smart agriculture IoT system based on deep reinforcement learning. *Future Gener. Comput. Syst.* **2019**, *99*, 500–507. [CrossRef]