Computers and Electronics in Agriculture 151 (2018) 376–383

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Big data and machine learning for crop protection

Ryan H.L. Ip[a], Li-Minn Ang[a,b], Kah Phooi Seng[a,b,*], J.C. Broster[b], J.E. Pratley[b]

[a] School of Computing and Mathematics, Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW 2678, Australia
[b] Graham Centre for Agricultural Innovation (Charles Sturt University and NSW Department of Primary Industries), Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW 2678, Australia

ARTICLE INFO

Keywords:
Big data
Machine learning
Crop protection
Weed control
Herbicide resistance
Markov random field model

ABSTRACT

Crop protection is the science and practice of managing plant diseases, weeds and other pests. Weed management and control are important given that crop yield losses caused by pests and weeds are high. However, farmers face increased complexity of weed control due to evolved resistance to herbicides. This paper first presents a brief review of some significant research efforts in crop protection using Big data with the focus on weed control and management followed by some potential applications. Some machine learning techniques for Big data analytics are also reviewed. The outlook for Big data and machine learning in crop protection is very promising. The potential of using Markov random fields (MRF) which takes into account the spatial component among neighboring sites for herbicide resistance modeling of ryegrass is then explored. To the best of our knowledge, no similar work of modeling herbicide resistance using the MRF has been reported. Experiments and data analytics have been performed on data collected from farms in Australia. Results have revealed the good performance of our approach.

1. Introduction

The data-driven economy with its emphasis on developing intelligent sensing, instrumentation and machines is expected to play a transformative role in agriculture and smart farming systems. Farming systems are affected by various factors like environmental conditions, soil characteristics, water availability and harvesting practices. Other important issues which have to be mitigated for include managing plant diseases, weeds and other pests. Traditionally, these factors and issues have been managed by the farmers own expertise and experience. The emergence of new trends like the Internet-of-Things (Gubbi et al., 2013; Da Xu et al., 2014) enable farmers to take a data-driven approach to collect vast amounts of information from instrumented sensors about the status of their farms (soil, water, crops, etc.) to improve farm yield and mitigate risks from weeds, pests and diseases. In addition to data collected from traditional sensors, more advanced sensing techniques which are being increasingly deployed for smart farming systems include proximal, airborne and satellite-based sensors.

The growing popularity of sensing techniques include RGB imaging, thermal, near-infrared (NIR), hyperspectral and multispectral imaging which can be ground-based or mounted on airborne drones to capture images of the farm. These imaging sensors contribute to the large amounts of the various types of data which have to be analyzed to derive value from the collective farm information. Efficient storage and analytics solutions need to be developed to handle the data generated by these near real-time sensing and instrumentation platforms. The enormous *volume*, *variety*, and *velocity* of data generated from sensors and real-time platforms in smart farming systems lead to a problem termed as ‘Big data’ (Wolfert, 2017; Chen and Zhang, 2014). To address the issue of Big data generated from large-scale networked sensing systems, the authors (Ang and Seng, 2016) use the term ‘Big sensor data’ and give discussions for potential applications in smart cities (Ang et al., 2017). We anticipate that Big sensor data systems will play an increasingly important role in modern agricultural applications.

One of the fastest growing areas under the discipline of ‘Artificial Intelligence’ (AI) is machine learning. The field of machine learning is becoming increasing popular and offers the solution to address the challenges of Big data. A general definition of machine learning refers to a group of modeling techniques or algorithms that can learn from data and make determinations without human intervention. Machine learning techniques are typically useful in situations where large amounts of data are available and relate to the output quantities of interest. For Big data problems, machine learning provides a scalable and modular strategy for data analysis.

Crop protection is the science and practice of managing plant diseases, weeds and other pests (Oerke, 2012; Schut, 2014). This paper addresses the issue of Big data and machine learning for crop protection. In this paper, some research efforts in crop protection or weed

* Corresponding author at: School of Computing and Mathematics, Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW 2678, Australia.
E-mail address: kseng@csu.edu.au (K.P. Seng).

https://doi.org/10.1016/j.compag.2018.06.008
Received 23 November 2017; Received in revised form 3 May 2018; Accepted 3 June 2018
0168-1699/

control using Big data and machine learning are first reviewed. Various machine learning approaches including discriminative/generative and supervised/unsupervised are also reviewed. This is followed by exploring the potential of a specific machine learning technique for herbicide resistance modeling using Markov random fields (MRF) models. The MRF has been frequently used in image, texture and pattern analysis applications. Some examples include Geman and Geman (1984), Johansson (2001), Geman and Graffigne (1987) and Li (2001). In image analysis, the lattices are often regular (e.g. typically modeling pixel coordinates in an image).

There have been some attempts in modeling environmental and agricultural datasets using the auto-logistic models (Zhu et al., 2005; Gumpertz et al., 2000). For environmental datasets, in most if not all situations, the lattices considered are irregular (e.g., shires, counties, states). The irregularity of the data lattices increases the challenges for modeling environmental and agricultural datasets compared with image analysis applications. Our approach aims to model the herbicide resistance of annual ryegrass on a set of explanatory variables while taking into account the spatial autocorrelation among neighboring shires. To the best of our knowledge, no similar work of modeling herbicide resistance using the MRF in machine learning has been reported so far. The autobinomial model (Besag, 1974, 1975) is used to model MRFs where the response variable consists of count data. This model with irregular lattice has been rarely applied to applications in agriculture. Experiments and data analytics are conducted to confirm the potential of the proposed MRF approach for modeling herbicide resistance from data collected from farms in Australian shires.

The remainder of the paper is organized as follows: Section 2 presents a review of Big data applications, data analytics and machine learning techniques. The aim of this section is to introduce the reader to representative studies and applications in Big data and machine learning in crop protection, and also to discuss a taxonomy of machine learning approaches for Big data which can be applied. Section 3 continues the discussion using a particular technique (the MRF) for machine learning modeling which takes into account the spatial component and irregular lattice in the data set. Section 4 illustrates the approach with a case study for modeling herbicide resistance of ryegrass using the MRF. Results and discussions on empirical data collected from farms in Australia are presented in Section 5. Finally, some concluding remarks are given in Section 6.

2. Review of Big data and machine learning approaches in crop protection

This section gives an overview of technologies and potential applications in crop protection using Big data and machine learning approaches. The section discusses four applications in crop protection: (i) Prediction and modeling of herbicide resistance; (ii) Detection and management of invasive species and weeds; (iii) Decision support systems for crop protection; and (iv) Robotics and autonomous weed control systems. Some major components in Big data such as data acquisition, storage and analytics are also briefly discussed. This is followed by a review of some popular machine learning techniques including discriminative/generative and supervised/unsupervised learning approaches. Fig. 1 shows an overview of the crop protection applications and its links with Big data and machine learning which also gives a summary outline of this section.

2.1. Big data & machine learning approaches in crop protection

Table 1 shows a summary for representative studies and applications in Big data and machine learning for crop protection. The applications have been briefly summarized into four categories (herbicide resistance modeling, detection/management of invasive species/weeds, decision support systems (DSS) for crop protection and robotics/autonomous weed control.

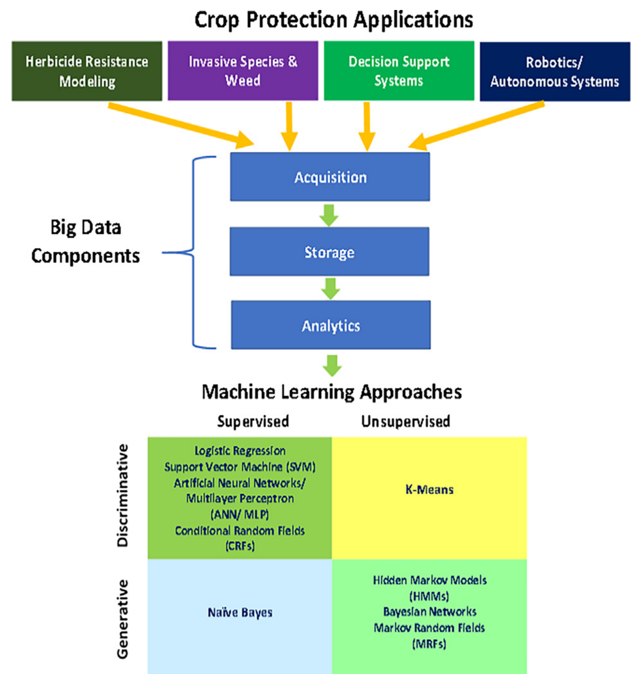

Fig. 1. Overview of potential crop protection applications and its links with Big data and machine learning approaches.

A recent review by Heap (2014) showed that the extent of herbicide resistance in agricultural weeds is increasing due to widespread and persistent use of herbicides in agriculture. In Australia significant research is undertaken to quantify the extent of herbicide resistance, especially in annual ryegrass (*Lolium rigidum*) (Boutsalis et al., 2012; Broster et al., 2011, 2012; Owen et al., 2014). Several researchers have proposed intelligent-based approaches and techniques to address the issue of herbicide modeling and prediction. An early study by Diaz et al. (2005) was to model and predict the heterogeneous distribution of wild-oat (*Avena sterilis* L.) density in terms of environmental variables. The authors used a rule-based model machine learning technique that performs a genetic search to discover the best rule set according to the classification instances of an experimental database. The best rule set using their approach was able to explain about 88% of the weed variability. The work by Evans et al. (2015), modeled the glyphosate resistance for the *Amaranthus tuberculatus* weed using classification and regression trees (CART) to identify the important relationships among 66 environmental, soil, landscape, weed community and management variables. The authors showed that Herbicide mixing was strongly linked with reduced selection for glyphosate resistance.

Machine learning approaches have also been applied towards the problem of detecting invasive species and weeds. The work by Lawrence et al. (2006) used random forest classifiers to map and detect invasive plants (leafy spurge and spotted knapweed) from aerial-based hyperspectral imagery. The aim of random forest techniques is to build multiple classification trees by repeatedly taking random subsets of the data to determine the splits in the classification trees. Using their approach, the authors reported an overall accuracy from out-of-bag data of 84% for the spotted knapweed and 86% for the leafy spurge. Schmidt and Drake (2011) used machine learning techniques to investigate the biological traits on why some plant genera are more invasive. The authors used boosted regression trees to develop classification models for each class of invasive plants. The advantage of boosting regression tress compared with conventional tree-based methods is that the boosting technique improves the analysis of large data sets containing many independent variables by combining large numbers of simple models adaptively to optimize the prediction accuracy. The authors showed that their approach could explain 24% and 29% of the variation in