Computers and Electronics in Agriculture 148 (2018) 72–81

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Machine learning approaches for early prediction of adult wool growth and quality in Australian Merino sheep

S. Shahinfar*, L. Kahn

University of New England, Department of Animal Science, Armidale, NSW, Australia

ARTICLE INFO

Keywords:
Machine learning
Prediction
Sheep
Wool
Fibre diameter
Staple strength

ABSTRACT

Wool production and its quality play important roles in determining the total income received by Australian sheep producers. Enabling accurate and early prediction of wool production and quality traits for individual and groups of sheep can provide useful information assisting on-farm management decision-making. Robustness and high performance of modern prediction methods, namely Machine Learning (ML) algorithms, make them suitable for this purpose. In this research, flock specific environmental data and phenotypic information of yearling lambs were combined to identify the most effective algorithm to predict adult Greasy Fleece Weight (aGFW), adult Clean Fleece Weight (aCFW), adult Fibre Diameter (aFD), adult Staple Length (aSL), and adult Staple Strength (aSS). Algorithms were evaluated and compared in terms of prediction error, the correlation between predicted and actual phenotype in a test set, and for uncertainty in prediction.

Artificial Neural Networks (NN), Model Tree (MT) and Bagging (BG) were used to carry out these predictions and their performance was compared with Linear Regression (LR) as the gold standard of prediction. The NN method had the poorest performance in all five traits. MT and BG had very similar performance and for a number of practical reasons, our method of choice was MT for early prediction of adult wool traits. The correlation coefficients of MT predictions were 0.93, 0.90, 0.94, 0.81 and 0.59 with Mean Absolute Error of 0.48 kg, 0.41 kg, 0.92 μm, 6.91 mm and 6.82 N/ktex, for predicting aGFW, aCFW, aFD, aSL, and aSS respectively.

1. Introduction

Farming in the 21st century is moving towards the use of 'Data Mining' approaches on 'Big Data' produced by data analytics that make best value of precise measurements from automatic computerised devices on livestock, crops, land, and climate. This 'Precision Agriculture' can support decision making by providing accurate, timely and economically optimised forecasts for farmers. In the sheep industry, wool production and its quality contribute significantly to the profitability of the farm. In the last decades the farm gate value of wool production has decreased from over $AU 6 billion to about $AU 2.5 billion in 2010 and has recovered slightly to about $AU 3 billion in 2016 (ABS, 2017). Precision agriculture provides a means to help the wool industry to remain competitive in the global fibre market with special relevance for improvements in the efficiency of wool production systems (Doyle, 2017). One example of the use of data to improve productivity growth for the wool industry is in the prediction of adult sheep traits from their early records as yearlings as a means of enabling early selection decisions. Forecasts in general will help farmers to plan their management practices in response to variability associated with climate, pasture, disease and targeted markets. It is clear that beside genetics, many environmental factors and management practices contribute directly or indirectly in quality and quantity of wool, and predictions need to account for these effects.

Total greasy wool shorn from the sheep, referred to as greasy fleece weight (GFW), represents the combined weight (kg) of clean wool fibre, wax, suint, vegetable matter, dirt and dust and other non-fibre components. The wool fibre, clean of these impurities, known as clean fleece weight (CFW), is one of the main traits of interest for wool producers. Non-genetic factors that affect CFW and other wool traits include age, birth and rearing status, age of dam, time of shearing (Campbell et al., 2011) and nutrient supply from pasture and supplements (which in turn are affected by climate and management). For example, more than 70% of the variation in wool growth can be explained by live-weight changes (Thompson et al., 1994). Body size and body condition score also affects fleece weight and fibre diameter (Adams and Briegel, 1998).

Mean fibre diameter (FD; measured in μm) is the most important raw wool property to be measured, and typically accounts for 75–80% of the per unit value of the fleece. FD and CFW are the main indicators

* Corresponding author.
E-mail address: shahinfar@uwalumni.com (S. Shahinfar).

https://doi.org/10.1016/j.compag.2018.03.001
Received 16 September 2017; Received in revised form 25 February 2018; Accepted 2 March 2018

of wool returns from individual Merino sheep (Atkins, 1997). FD is highly heritable with high repeatability during the lifetime of the sheep (Atkins, 1997). Changes in FD throughout the year are also related to a range of non-genetic factors (see for CFW) including seasonal changes in body weight, fat depth and skin thickness (Brown and Crook, 2005). Some of these affects can remain for the life of the animal with single born progeny producing wool with lower FD than twin-born and reared lambs (Kelly et al., 2006).

Staple strength (SS) reflects the force required to break an individual staple when extended and is reported as Newtons per kilotext (Schlink, 2009) and is the second most important determinant of premiums and discounts that apply to the raw wool value (Adams et al., 1997). Penalties for low SS in Australian wool clips are significant. The major reason for low SS is the variation in FD due to seasonal changes in the availability and quality of pasture and supplements (Masters et al., 1998) with a rapid change in seasonal circumstance predisposing for a "break" of wool (Mata et al., 1999). The timing of supplementary feeding (Doyle et al., 1995) and shearing (Rogan et al., 1995) in relation to pasture availability and quality and animal requirement can help prevent reductions in SS. The length of each individual staple, staple length (SL; measured in mm) is also one of the important characteristics of wool and can contribute to the gross income from the fleece depending on the targeted market and end products.

The underlying distributions of field data from commercial farms in the sheep industry are often Gaussian, but some traits are better described as binomial, poisson, exponential, or gamma distributions. Moreover, farm data sets usually have a considerable percentage of missing values. Contrary to classic statistical methods, machine learning (ML) approaches are best positioned to accommodate big and complex data with missing values and do not rely on parametric assumptions such as normal distribution of response variables or residuals (Valletta et al., 2017; Witten and Frank, 2005). Machine learning to predict future performance has been used in dairy cattle more than other types of livestock. For example, artificial neural networks have been used for prediction of milk yield and mastitis (Yang et al., 1999), . Tree based methods also have been applied successfully for prediction of retention pay-off (Shahinfar et al., 2014a), and optimisation of reproductive management programs (Shahinfar et al., 2015).

To our knowledge, prediction models for wool production of adult sheep based on their yearling records that combine genetic, environment and management effects do not exist. The objective of this study was to identify the best performing ML algorithms -that have the least prediction error or highest correlation between predicted and actual values- for predicting adult wool production using weather, pasture, animal health and various measures of related phenotypes. Finally, the best performing model would be selected for further fine tuning and inclusion in the ASKBILL™ decision support tool (Kahn et al., 2017).

2. Material and methods

2.1. Data

Data were collected over a period of about 8 years from the Sheep CRC Information Nucleus Flocks as described by (Van der Werf et al., 2010). After quality control and exclusion of inaccurate records, the data set contained 7294 records of animals that had yearling wool records (records taken from an animal at about one year old age, ± 115 d) and at least one measurement of their adult (a) GFW that provided; 5832 records for aCFW; 5787 records for aFD; 4665 records for aSL; and 4660 records for aSS. These records were matched with: a range of other phenotypic measurements; weather information from the Australian Bureau of Meteorology (BOM); and predicted pasture dry weight and digestibility (Johnson et al., 2003). Table 1 provides a complete list of predictors for each trait and the number of records available for each trait. Missing values range from 0% to 91% (fly severity).

2.2. Machine learning algorithms

Supervised machine learning methods map a set of categorical, nominal, or continuous features (most often a combination of all of them) to their related outcome which themselves can be in any of those forms. Their biggest advantage over common linear models is their ability to learn relationships from training data and generalise it to the unseen testing set and also to overcome non-linearity and interactions among features. However, this ability needs to be carefully managed to avoid over-fitting. In order to find the best prediction model for practical use, the standard approach is to try a short list of appropriate predictive methods on the data set of interest and then pick the best performing method and fine-tune it for use as the predictor tool. In machine learning those configuration parameters that are external to the model and are not learned from data, are called hyperparameters, for example the number of hidden layers in neural networks. Hyperparameters must be optimised by cross-validation or grid search to make a balance between variance and bias in prediction, known as the variance-bias trade-off (James et al., 2013). In this study, a cross validation approach was used. Herein we are describing the performance of a tree based (MT), a gradient based (NN) and an ensemble (BG) method and the comparison of their predictive performance with linear regression (LR), for prediction of aGFW, aCFW, aFD, aSS and aSL. Weka APIs were used to implement ML methods in this study (Frank et al., 2016).

2.2.1. Artificial neural network (NN)

A feedforward artificial neural network takes a vector of real value inputs and calculates a linear combination of these inputs into a set of appropriate outputs. It is well-suited for cases in which the instance space is noisy, complex and inter-correlated (Mitchell, 1997).

2.2.2. Model tree (MT)

Model tree (MT) is a type of decision tree developed for numeric prediction. A process similar to a decision tree divide and conquer approach is used to partition the multidimensional prediction space of the problem and exploit the partitions (Quinlan, 1992). Values for test instances are predicted by a linear model stored in each prediction node. The MT often provides accurate and transparent prediction of complex systems with nonlinear and inter-correlated variables.

2.2.3. Bagging (BG)

Bagging (BG) which stands for bootstrap aggregation, is an ensemble method in which multiple versions of a predictor, as the base learner, will be generated on bootstrap samples of training data to develop an aggregated predictor. When predicting numeric values, the final prediction is an average over predicted values of all models, while for classification, the majority of vote is used (Breiman, 1996). In this study we used bagging of MT.

2.3. Variable selection methods

In ML practices, it is tempting to include as many variables as possible to the model. Although in theory having more features should increase the discriminative power of any predictive algorithm, in practice often adding irrelevant features can distract the learning algorithm and defect the prediction performance as well as increase the time needed for the learning and prediction phase. Feature selection also elucidates the optimal subset of predictors and causal drivers under the operating system of interest (Valletta et al., 2017). Possible usable features in our database consisted of 190, 189, 192, 196, and 197 features related to aGFW, aCFW, aFD, aSL, and aSS respectively. Forward greedy hill climbing search jointly with expert's knowledge were used to select a small effective subset of attributes for each trait of interest prior to the training phase. The training process was conducted with a selected subset of attributes (Table 1).