Proc. Assoc. Advmt. Anim. Breed. Genet. 22:433-436

EARLY PREDICTION OF IMPORTANT ADULT WOOL TRAITS

S. Shahinfar[1], and G. Hinch[1], J. Van Der Werf[1], L. Kahn[1]
[1] University of New England, Department of Animal Science, Armidale, NSW, Australia

SUMMARY

Both production and quality play important roles in determining the wool income received by Australian sheep producers. Enabling accurate and early prediction of wool production and quality 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), and adult Fibre Diameter (aFD). Those algorithms were evaluated in terms of prediction error and correlation between predicted and actual phenotype in a test dataset.

Multiple linear regression (MLR), Multilayer perceptron (MLP), Model Tree (MT) and Bagging (BG) were used to carry out these predictions and their performance were compared. The MLP method had the poorest performance in all three traits versus, MLR, MT, and BG had very similar performance with BG being superior in all three traits and prediction criteria, with correlation coefficients of 0.93, 0.90, 0.95 and Relative Absolute Error (RAE) of 0.34, 0.41, 0.31 for predicting aGFW, aCFW, aFD respectively.

INTRODUCTION

Phenotypic prediction of wool production of adult sheep based on their early records as yearlings has great management value for sheep producers, allowing them to base their culling decision on an accurate future prediction of wool production for each individual sheep. 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.

Various authors have identified some of the more important factors that affect wool production. For example, Masters et al., (1998) demonstrated that initial liveweight, liveweight change, and supplement choice all have effect on wool growth and staple strength in weaner sheep. Ferguson et al., (2011) reported that liveweight at joining, and liveweight change during pregnancy and lactation acted to regulate wool production of Merino ewes. They used linear prediction models based on a REML approach for predicting CFW, FD, and SS from their data. The authors did not test the model on independent test data thus preventing a generalised understanding. 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 develop and compare the performance of different ML algorithms to predict adult wool production using weather, pasture, animal health and various measurements of related phenotypes, and some related traits. Finally, the best performing model would be selected for further fine tuning and development in the form of a decision support system for industry.

MATERIALS AND METHODS

Data. Data collected over a period of more than 6 years as part of the Sheep CRC Information Nucleus Flocks were used in this study (van der Werf et al., 2010). After editing, the data set contained 7,501 records of animals that had yearling and at least one measurement of their adult GFW; 5,962 record for aCFW and 5,917 record for aFD. Data that were included as phenotypic measurements included, conformation characteristics of sheep that are related to wool production such as Body Wrinkle, health related features such as worm egg count (WEC), and pregnancy status of sheep at yearling. Weather information from each site where the flocks were managed was obtained from the Bureau of Meteorology (BOM). Addition of flock specific in the variable set can be considered as fixed effects in linear mixed model to capture the whole management and perhaps micro-climate effect that might exist in the flock. Also pasture data included predictions of pasture dry weight and digestibility of herbage mass obtained from a simulation model developed by Johnson et al. (2003) were used.

Machine Learning Algorithms. 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 of interest and then pick the best performing method and fine-tune it for use as the predictor tool. In this paper we are comparing a tree based method (MT), a gradient based method (MLP) and an ensemble method (BG) and compare them with the most common statistical method of prediction, Multiple Linear Regression.

a) **Multilayer Perceptron:** is a feedforward artificial neural network that takes a vector of real valued input 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 intercorrelated (Mitchell, 1997).

b) **Model Tree:** is a type of decision tree developed for numeric prediction. A process similar to decision trees 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 will be predicted by a linear model stored in each leaf. The MT has been used in prediction of retention pay-off in dairy cattle (Shahinfar et al., 2014). MT often provides accurate and transparent prediction of complex systems with nonlinear and intercorrelated variables.

c) **Bagging:** which stands for bootstrap aggregation, (BG), is an ensemble method in which multiple versions of a predictor will be generated on bootstrap samples of training data to finally drive an aggregated predictor. When predicting numeric values, final prediction is an average over predicted values of all models (Breiman, 1994, and Breiman, 1996). In this study we used Bagging of MT.

Variable Selection Method. In Machine learning 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 ML algorithm, nevertheless, in practice often adding irrelevant features can distract the learning algorithm and defect the prediction performance as well as increase the time needed for learning and prediction phase. Full model in this research were consist of 190, 189, and 192 variable for predicting aGFW, aCFW, and aFD respectively.

Greedy hill climbing search in forward manner was used to select a small effective subset of attributes for each trait of interest. Then the same training process was carried out with selected subset of attributes and results were compared (Table 1). The reduced models were as below:

aGFW= Sex + yYLD + yGFW + yBDWR + ytMin_6 + yDryWtAv_9
aCFW = Sex + yCFW + yGFW + yBDWR + yPregScan + yrainAv_3 + ytMin_6 + yDryWtAv_12
aFD = Sex + SireBreed + yOFDA_SpinFine + yOFDA_FDSD + yOFDA_FD + AgeDiff + yFACE + yPregScan + yCS + yDryWtAv3 + yDigA8

where "a" in prefix indicates adult time and, "y" prefix indicates yearling time. BDWR = Body Wrinkle score. tMin_6 = average of minimum temperature in the 6 months prior to first shearing. DryWtAv_9= Dry weight average per hectare in the 9 months prior to first shearing. rainAv_3= Average of Rain fall in the 3 months prior to first shearing. DryWtAv_12= Dry weight average per hectare in the 12 months prior to first shearing. AgeDiff = number of days between first and second shearing. Face= Face Cover Score. CS= Body Condition Score. DryWtAv3 = Dry weight average per hectare in next 3 month after first shearing. DigAv8= Average of Digestibility of pasture in the 8^{th} month after first shearing.

Model Evaluation. To evaluate each Model's performance in 10-fold cross validation framework, three accuracy measurements were considered, Correlation Coefficient between actual and predicted value, Root mean Squared Error (RMSE) and Relative Absolute Error (RAE). Correlation and RMSE are very well known and standard measurements for any prediction method. RAE was used in this research for two main reasons. First, it measures absolute error, which is not affected by outliers. Second, it considers the relative magnitude of the error compared with the predicted value.

$$RAE = \sum_{i=1}^{n} \frac{|\, p_i - a_i|}{|\, a_i - \bar{a}\,|}$$

Where p_i is predicted value; a_i is actual value; and $\bar{a}$ is the prediction by an arbitrary predictor, in this case the average of actual values (Witten and Frank, 2005).

RESULTS AND DISCUSSION

MT and BG always had the best performance in both cases of full (FM) and reduced model (RM) in all performance measurement criteria (Table 1). The superior accuracy of BG is due to ensemble power in which several predictor models will be aggregated to generate a high performance predictor. The superiority of BG over MT was not statistically significant and one could choose MT over BG for practical purposes, of which three are proposed herein. Firstly, the running time on MT is much less demanding than BG. Secondly, despite the black box nature of BG being ambiguous and hard to explain for users, MT is very transparent and intuitive. Thirdly, as a practical rule of thumb in ML, once a single model shows a high prediction performance, ensemble methods will not add much of accuracy. Surprisingly MLP had the poorest performance among all four methods. Perhaps in our case MLP needed much more investigation and fine tunning to deliver a reasonable performance.

In order to assess accuracy and generality of the machine learning of choice, a user should not rely on a simple comparison between two single run or even two 10-fold crossvalidation run. The problem would arise in cases that some algorithems have very close performance and some have certain advantages on others in practice. Repeated 10-fold cross validations were performed on the same partition of data for all four algorithems in use, and Tukey multiple comparison of means were performed on the mean of accuracy criteria. The results is shown in Table 1 using alphabedical superscripts. As multiple means comparison indicated, in most cases there was no significant difference between BG and MT while MLR and MLP were often associated with poorer performance in comparison. As a conclusion our method of choice was MT for early prediction of adult wool traits.

Table 1. Results of 10-fold cross validation for full and reduced models for aGFW, aCFW and aFD with multiple mean comparisons indicated in superscripts.

		Correlation		RMSE		RAE	
Trait	Method	FM*	RM*	FM	RM	FM	RM
aGFW*	MLR^{B**}	0.91^{b}	0.81^{b}	0.73^{b}	1.03^{b}	0.38^{b}	0.54^{b}
	MLP^{C}	0.87^{c}	0.81^{b}	1.03^{c}	1.21^{c}	0.58^{c}	0.68^{c}
	MT^{A}	0.93^{a}	0.92^{a}	0.66^{a}	0.70^{a}	0.35^{a}	0.37^{a}
	BG^{A}	0.93^{a}	0.92^{a}	0.64^{a}	0.69^{a}	0.34^{a}	0.36^{a}
aCFW*	MLR^{BC}	0.88^{c}	0.78^{c}	0.59^{b}	0.78^{c}	0.46^{b}	0.60^{b}
	MLP^{CD}	0.83^{d}	0.78^{c}	0.81^{c}	0.90^{d}	0.65^{c}	0.72^{c}
	MT^{AB}	0.89^{b}	0.87^{b}	0.56^{a}	0.61^{b}	0.43^{a}	0.46^{a}
	BG^{A}	0.90^{a}	0.89^{a}	0.53^{a}	0.57^{a}	0.41^{a}	0.44^{a}
aFD*	MLR^{B}	0.93^{b}	0.91^{c}	1.31^{b}	1.55^{b}	0.32^{a}	0.39^{b}
	MLP^{C}	0.88^{c}	0.88^{d}	2.00^{c}	2.06^{c}	0.54^{b}	0.56^{c}
	MT^{A}	0.94^{a}	0.93^{b}	1.26^{ab}	1.36^{a}	0.31^{a}	0.33^{a}
	BG^{A}	0.95^{a}	0.94^{a}	1.23^{a}	1.29^{a}	0.31^{a}	0.32^{a}

aGFW= Adult Greasy Fleece weight, aCFW= Adult Clean Fleece Weight, aDF= adult Fibre Diameter, *FM= Full Model, RM= Reduced Model. Correlation= correlation between actual and predicted value in test set.
** Alphabedic superscript in method column shows overall method's mean comparison.

REFERENCES

Breiman, L. 1994. Bagging predictors. Technical report. Department of Statistics, University of California Berkeley, CA.

Breiman, L. 1996. Bagging predictors. Mach. Learn. 24:123–140.

Witten, I. H., and E. Frank. 2005. Data Mining, 2nd ed. Elsevier, San Francisco, CA.

Johnson IR, Lodge GM, White RE (2003). The sustainable grazing systems pasture model: description, philosophy and application to the SGS National Experiment. Animal Production Science 43(8), 711-728

Masters DG, Mata G, Liu S. M, and Peterson AD, 1998, Innfuence of liveweight, liveweight change, and diet on wool growth, staple strength, and fibre diameter in young sheep. Aust. J. Agric. Res., 49, 269-277

Ferguson, M B, Thompson AN, Gordon DJ, Hyder MW, Kearney GA, Oldham CM, and Paganoni BL, 2011, The wool production and reproduction of Merino ewes can be predicted from changes in liveweight during pregnancy and lactation. Animal Production Science, 51, 763–775

Mitchell, T. M. 1997. Artificial neural networks. Pages 81–85 in Machine Learning. McGraw-Hill International Edition, New York, NY.

Van der Werf, J. H. J, Kinghorn, B. P, Bank, R. G, 2010, Design and role of an information nucleus in sheep breeding programs. Anim. Prod. Sci. 50: 998-1003.