

Mating decision support system using computer neural network model in Kenyan Holstein-Friesian dairy cattle

D M Njubi*, J Wakhungu and M S Badamana

Department of Animal Production, University of Nairobi, P.O. Box 29053-00625, Nairobi Kenya

* P.O Box 30623-00100 Nairobi Kenya, 0722 206 856

davidnjubi@yahoo.com

Abstract

Knowledge discovery in databases (KDD) should provide not only accurate predictions but also comprehensible rules. In this paper, we demonstrate that the machine learning approach of rule extraction from a computer trained neural network system can successfully be applied to milk production analyses in dairy cattle. Such extracted knowledge should be useful in interpretation and understanding how the neural network (NN) model makes its decision.

Data consisting of 6095 lactation records made by cows from 76 officially milk recorded Holstein Friesian herds in the period 1988-2005 were used to extract rules using neural network. Two different methods of attribute categorization; auto-class and the domain expert were used. For automated knowledge acquisition, rule induction used Weka software while SAS was used in domain expert.

The neural nets were first trained to identify outputs for different inputs. The trained networks were then used for rule extraction. The study showed that the decision trees generated from the trained network had higher accuracy than decision trees created directly from the data. The study also indicated a need for a process to determine important inputs before using a neural net and showed that reduced input sets may produce more accurate neural nets and more compact decision trees.

The “black-box” nature of neural networks was explained by extracting rules with both the domain expert and auto-class for both the continuous and the discrete valued inputs with rule sets performing better on the ‘low’ and ‘high’ levels. It follows from these analyses that performance at the two extremes was more important than average performance. It implied that the end user was particularly concerned with identifying mating with good potential and avoid mating with poor potential animals. The decision tree showed that when the herd performance was low then the foremost limiting factor was the dam performance whereas for medium and high herd performance sire level performance was the limiting factor. Through sensitivity analysis the most important and sensible factors with respect to productivity were sire breeding value and herd performance. It was, therefore, concluded that neural network rule extraction and decision tables were powerful management tools that allow the building of advanced and user-friendly decision-support systems for mating strategy designs and their evaluation.

Keywords: dairy cattle, mating decision, rule extraction

Introduction

Although artificial neural network (ANN) has proven to be an efficient classification and regression tools, understanding the established relationships between the input and the output variables is a problem. Knowledge discovery is defined as the process of identifying valid, novel, potentially useful and ultimately understandable structure in data (Bradley et al 1998). There are several methods that explain the complex and often non-linear relationships from the input-output mapping performed by the models.

In a previous paper, it has been successfully shown that the results for the networks to be significantly better than the linear regression method (Njubi et al 2009) and this has made the neural network approach interesting enough to exploit further. At the same time, the ability to present the "knowledge" learned by the network in a more transparent notation was identified as a key property for the model to be used as a tool for decision-making.

In the study by Njubi et al 2009, the trained neural networks were used as a basis for finding a model transparent enough to enable decision-making. Several extraction methods can be used to create decision trees from the trained neural nets.

Some techniques, such as decision trees and linear regression, are regarded as transparent while others most notably neural networks (NN), are said to be opaque and must be used as black boxes. This is one of the criticisms of NN in that interpretation is difficult although their prediction has proved more effective at forecasting than the conventional linear techniques. The provision of a mechanism that can interpret the networks input/output mappings in the form of rules that human experts can verify would be very useful. This process of converting opaque models into transparent models is often called *rule extraction*.

Extracting information from trained neural network

There are various techniques for extracting information from trained neural network. The rule extraction and sensitivity analysis are some of the techniques for extracting information from trained neural network. In sensitivity analysis, one determines the effect of an input variable on the output while holding other inputs to some fixed value, usually the mean or median value derived from the data. In particular, we intend to show that by applying sensitivity analysis to neural network models, we can identify sensible factors that play important roles to total milk production.

Sensitivity analysis may be used to determine which input parameter is more important or sensible to achieve accurate output values. Sensitivity analysis has been applied in different domains (Poh et al 1998; Yao et al 1998). In data mining, we may apply sensitivity analysis to find items that are sensible to total profit. These techniques allow us not only to mine these patterns and rules that can lead to profit or actions but also to maximize profit in a dynamic environment.

In rule extraction the rules extracted are generally represented by a set of *if-then* statements that may be examined by a human expert for decision making. Neural network rule extraction is a technique that translates the decision process of a trained neural network into an equivalent decision process represented as a set of rules. The technique of rule extraction has been used to model the knowledge that the neural network has gained while training or adapting.

Classification

Accurate identification and categorization of sire and bull-dams by clients (AI technicians, farmers etc) is an important element in enhancing genetic gain. For example, a farmer may want an average performing daughter due to his/her environment conditions and this will necessitate mating the dam to an average sire. In this study we explore the use of machine learning approach called clustering for classifying independent variables affecting daughter performance. Cluster analysis is one of the most prominent methods for identifying classes amongst a group of objects, and has been used as a tool in many fields such as biology, finance and computer science. The algorithms evaluated in this study use unsupervised learning mechanism as opposed to supervised machine learning approach. Decision trees are powerful and popular tools for classification. The attractiveness of tree-based methods is due in large part to the fact that, in contrast to neural networks, decision trees represent *rules*.

In this paper, we present the results from analyzing the Kenyan-Holstein dairy cattle lactation data sets using neural network rule extraction techniques for mating decisions. The previous paper (Njubi et al 2009) dealt with predictive problem in the Kenya dairy industry.

Clarifying the neural network decisions by explanatory rules that capture the learned knowledge embedded in the networks can help the animal breeders in explaining why a particular sire/dam is classified as either low or medium or high for particular breeding objectives.

Two different methods of attribute categorisation, autoclass (Cheeseman and Stutz 1996), and the domain expert are compared. For automated knowledge acquisition-rule induction we used Weka J4.8 (Witten and Frank 2000) implementation [Weka-Waikato Environment for Knowledge Analysis available from <http://www.cs.waikato.ac.nz/ml/weka/>]. Our objective is to obtain rule sets which are both accurate and comprehensible to the farmer/inseminator as the end user of the system.

Materials and methods

The Kenyan-Holstein cattle population

This has been described in detail by Njubi et al 2009. In brief, this study uses the Kenyan Holstein-Friesian cattle data for prediction and ultimately rule extraction. A total of 6095 lactation records from 2267 Holstein-Friesian cows milk recorded with the dairy recording society of Kenya (DRSK), and made in the period 1988-2005 were available for the analysis.

Data collection and processing

Data for this study were obtained from cow files maintained at the DRSK in Nakuru, which is the organization responsible for the official milk recording in Kenya. Each record contained the following information; herd identification, individual cow identification, cows dates of birth (day-month-year), cows calving dates (day- month-year), lactation milk yield (kg), lactation length (days), parity, sire and dam. The data was preprocessed with all the inconsistency removed eg. animals without known sire and dam. For detail see Njubi et al 2009.

Neural network analyses

The back propagation training algorithm was employed to predict the production trait of daughter milk yield for first lactation. The initial dataset was divided into two subsets. One was used in the training step, the other in the testing step as described in Njubi et al (2009). There are several methods that explain the complex non-linear relationship from the input-output mapping. In this paper, rule extraction and sensitivity analyses for the Multilayer Perceptron (MLP) has been used in order to acquire knowledge about the problem.

The rule sets

Neural network has been criticized in that interpretation is difficult. The provision of a mechanism that can interpret the networks input/output mappings in the form of rules would be very useful. The following inputs were used to derive rules as one solution for understanding the networks. The Weka (Weka 2005) software was used.

The five attribute variables for the study were: Herd milk yield; dam's second lactation yield; sires's breeding value for milk yield; and daughter's first lactation milk yield which is the quantity to be predicted

To classify daughter first lactation milk yield (FLMY), we used three inputs for the neural network in Weka J4.8 implementation (Witten and Frank 2000); the average herd milk yield, dam second lactation milk yield (SLMY) and sire breeding value (BV) for milk yield.

Categorization

Our study considers one unsupervised clustering algorithm Autoclass which uses the Expectation Maximization (EM) algorithm (Weka 2005) to estimate the parameter values that best fit the data for a given number of classes. Domain expert (Pietersma et al 2001) was used to categorize the continuous variables.

Expert categorization

The output variable was split into to three classes; low, average and high, containing 25%, 50% and 25% of the records respectively in order to ease interpretation and comparison with the auto-class method. Input variables were categorized into low, average and high into varying proportions (Table 1). SAS (2003) software was used in the analyses.

Table 1. Expert and Autoclass categorization results

		Expert		Autoclass	
		Upper Bound	Frequency, %	Upper Bound	Frequency, %
Herd milk average yield	Low	3250	13	4685	36
	Average	7250	70	6625	36
	High	¥	17	¥	28
Dam SLMY	Low	2750	9	4488	36
	Average	7750	76	6759	35

	High	¥	15	¥	29
Sire BV for milk	Low	0	49	-328	22
	Average	650	40	263	47
	High	¥	11	¥	31
Daughter FLMY	Low	3754	25	4447	35
	Average	6798	50	6957	41
	High	¥	25	¥	23

¥-upper limit open ended

Auto-class categorization

This is unsupervised clustering algorithm based on Bayes theorem. User interaction was necessary to set upper limit on the number of clusters to generate. Clusters were defined in terms of mean and standard deviation of Gaussian distribution (Cheeseman and Stutz 1996). The boundary, b , between any two classes having means m_1 and m_2 ($m_2 > m_1$) and standard deviations s_1 and s_2 , respectively, used the formula: $b = m_1 + s_1(m_2 - m_1)/(s_1 + s_2)$

J4.8 Learning algorithm

The classifier Weka J4.8 (Witten and Frank 2000) implementation is a supervised learning decision tree classifier. Having constructed a complete decision tree J4.8 proceeds to extract rules.

Four different datasets were used in this study. The input and output variables were prepared differently.

- (i) Dataset A: inputs continuous values, outputs discretized by expert
- (ii) Dataset B: inputs continuous values, outputs discretized by Auto-class
- (iii) Dataset C: inputs and outputs discretized by expert
- (iv) Dataset D: inputs and outputs discretized by Auto-class

The data sets were compared through two criteria, rule set accuracy and the number of rules. Rule set accuracy was determined by calculating the average of precision and recall for each output class. Precision is defined as the percent of instances classified as true positive while recall is defined as the percent of actual positive instances correctly classified as positive.

Sensitivity analyses

The next study focused on rating the importance of variable with respect to a particular model. Sensitivity analysis does enable them to explain which inputs are more important than others. The network learning is disabled during this operation such that the network weights are not affected. The basic idea is that the inputs to the network are shifted slightly and the corresponding change in the output is reported either as a percentage or a raw difference. The activation control component generates the input data for the sensitivity analysis by temporarily increasing the input by a small value (dither). The corresponding change in output is the sensitivity data. Each input to the network was varied between its mean $\pm$ standard deviation, while all other inputs were fixed at their respective mean/median values.

Once the model is trained we used the strategy of studying the influence of input variables on the dependent variable by evaluating the changes in the error committed by the network that would result if an input were removed.

Results and discussion

Rule extraction

Although neural networks have gained popularity in many fields they have not been widely used. This is because some of critics of neural network argue that there is no explanation of the mechanism inside the models. Extracting rules from trained neural networks is one of the solutions for understanding the networks.

In this study the extraction of classification rules is made for both continuous and discretized data, with the purpose of making clear and comprehensible to the user (AI managers) how the attributes are ‘acting’ to perform the classification of each one of the daughter milk yield. The discretized data was compared to the continuous variables through the accuracy percentage and the number of rules generated. Two methods, domain expert and the Autoclass classifier were used for categorization of variables.

Categorization

Table 1 shows a summary description of the categorization of continuous attributes resulting from the expert and auto-class. There were two major choices to be made, firstly the choice between continuous versus discrete input representation and secondly; the choice between expert and auto-class discretisation of the attributes.

Input representation: continuous vs discrete

The two input representation were similar, with continuous valued inputs giving greater accuracy (Tables 2 and 3) in all the three output classes (Low, Average, and High).

Table 2. Data categorization by the expert

Output class	Dataset A		Dataset C	
	Continuous valued inputs		Discrete valued inputs	
	accuracy	#rules	accuracy	#rules
Low (25%)	74	2	87	2
Average (50%)	85	1	65	1
High (25%)	73	5	77	1

Table 3. Data categorization by Autoclass

Output class	Dataset B		Dataset D	
	Continuous valued inputs		Discrete valued inputs	
	accuracy	#rules	Accuracy	#rules

Low (35%)	67	1	78	1
Average (41%)	78	3	64	1
High (23%)	82	9	74	3

The domain expert clustering resulted in less discrete rules as compared to the continuous valued inputs. This is advantageous as far as a support system is concerned where it is semantically easier to employ category labels rather than numeric intervals. For example, a rule such as ‘if Kenyan sire breeding value is average then the daughter milk production is low’ is easier for a user (eg. Farmer) to appreciate than “if Kenyan sire breeding value is less than 230, then the daughter’s milk production is less than 4060 litres”. The other advantage of rules being discrete categories is that as absolute animal yields increase resulting from improved breeding and management there will be no change in the category boundaries.

Categorization method: expert vs auto-class

Table 1 shows the categorization of continuous variables, using the expert domain and the auto-class. The categorization was similar in the two datasets for the auto-class with discrete valued inputs slightly better. In both the expert and auto-class the discrete valued inputs, rule sets performed better on the ‘low’ and ‘high’ classes while for the continuous inputs there were no clarity in the 3 classes. The number of rules for the discrete valued inputs in both expert and auto-class were fewer. It follows from this analyses that performance at the two extremes are important than average performance for the discrete valued inputs. The farmers as one of the end users is particularly concerned with identifying mating with good potential and avoid mating with poor potential bull dams.

The domain expert decided on 3 categories for each of the input attributes and 3 categories for the output attribute while the autotclass produced the same categories for herd, dam, estimated breeding value and daughter first lactation milk yield. Since the performance of expert and automated categorization were generally the same then the use of automated tools to categorise and classify data is very important. From the foregoing it is clear that automating the categorization process in the intelligent decision support system (IDSS) is justified.

Data mining using decision trees

The ability to explain the reason for a decision is crucial in herd management. We used J4.8 implementation in Weka toolkit. The performance metrics of accuracy was 78.9%. This clearly shows that the results are promising for the application of the data mining method into the prediction of daughter first milk production. The tree in Figure 1 correctly classifies 79% of the records.

Figure 1. Daughter first Milk Yield decision tree

The result shows that when the herd performance is low then what need improved is the environment eg. management before thinking about the sire to be used on the dams. On the other hand when the herd performance is medium then sire breeding values become important. In this case where the sire breeding value is high then one has to be careful on using high performers.

If the herd environment is low then only average dams should be mated to average sires. Use of sires of high breeding value whose semen is expensive is not recommended because it will not result to high milk production and consequently decrease in dairy farm profitability.

On the other hand if the herd environment is medium to high then more attention should be focused on the sire performance. For medium herds, sires of medium to high performance can be used while for high performing herds it is recommended that higher performing sires be used on high performing dams. For high performing herds, sires of high breeding value should be used.

Sensitivity analysis

In this section, a method of calculating output sensitivities to inputs' variation from a trained ANN is discussed. The following inputs were varied: sire breeding value, herd and dam milk yield levels. The changes in the error committed by the network showed the relative importance of inputs.

Removing the sire component (EBV) variable resulted in the highest error relative to herd and the dam milk yield. This was further confirmed by simulating variation of herd production

management by 1 unit (Figure 2) resulted in 16 Kg annual increase in milk yield (translates to 0.0438 kg per day)

Figure 2. Sensitivity for Kg EBV

while varying the breeding value of sire by 1 unit translate to 44 Kg annual increase in milk yield (translates to 0.121 kg per day) (Figure 3)

Figure 3. Sensitivity for Kg Herd Milk yield

and that of dam variation by 1 Kg translate to 6 Kg annual increase in milk yield (translates to 0.0164Kg per day) (Figure 4).

Figure 4. Sensitivity for Kg Dam Milk Yield

This meant that the milk production environment had a higher influence on the daughter milk performance than either the sire breeding value or the dam milk yield. This meant that the herds' management varied to the extent that the effect of sire differences was masked. This fact is extensively referenced in the literature (Rege et al 1991, Njubi et al 1992). The impact in dairy cattle through the sire selection pathway is higher relative to that of the dam (Van Tassel and Van Vleck 1991) and this may explain why the contribution of sire is higher compared to that of the dam.

Conclusions

- Although the discoveries of interesting and previously unknown patterns are important for data mining applications, it is more important to discover actionable rules. Data mining is viewed as the process of turning the data into information, the information into action, and the action into value or profit. As shown in the previous paper (Njubi et al 2009), neural network prediction is superior to linear and therefore understanding the way neural network makes their decision is imperative. The results from this study show that the “black-box” nature of neural networks was explained by extracting rules with both the domain expert and auto-class (a Bayesian classifier) for both the continuous and the discrete valued inputs with rule sets performing better on the ‘low’ and ‘high’. It follows from this analysis that performances at the two extremes are more important than average performance. It implies that the end user is particularly concerned with identifying mating with good potential and avoid mating with poor potential animals.
- Rules like ‘when need an average producing daughter then mate average producing dam to average sire’ is more meaningful to a farmer than dealing with numerical figures. That the black-box nature of the neural network can be interpreted into useful rules which are

more semantically easier understood by a user is more meaningful. The other advantage of rules being discrete categories is that as absolute animal yields increase resulting from improved breeding and management there will be no change in the category

- Sensitivity analysis can be used for one of the interesting mining tasks, productivity. Sire was the most important factor with respect to productivity than the dam or herd management. This study concludes that neural network rule extraction and decision tables are powerful management tools that allow us to build advanced and user-friendly decision-support systems for mating evaluation.

Acknowledgements

Appreciation is expressed to the Dairy Recording Society of Kenya for providing data.

References

Bradley P S, Fayyad U M and Mangasarian O L 1998 Mathematical programming for datamining: formulation and challenges. *NFORMS Journal on Computing* 3(11): 217-238

Cheeseman P and Stutz J 1996 'Bayesian Classification (AutoClass): Theory and Results', in advances in: Knowledge discovery and Data Mining. American Association for Artificial Intelligence Menlo Park, CA, USA

Njubi D M, Rege J E O, Thorpe W, Collins-Luswet E and Nyambaka R 1992 Genetic and Environmental variation in reproductive and lactational performance of Jersey cattle in the coastal lowland semi-humid tropics . *Tropical Animal Health and Production* 24(4): 231-241

Njubi D M, Wakhungu J and Badamana M 2009 Milk Yield Prediction on Holstein-Friesian cattle using Computer Neural Networks System(submitted to Livestock Research for Rural Development.

Pietersma D, Lacroix R, Lefebvre D, Block E and Wade K M 2001 A Case-Acquisition and Decision-Support System for the Analysis of Group-Average Lactation Curves. *Journal of Dairy Science* 84: 730-739
<http://jds.fass.org/cgi/reprint/84/3/730.pdf>

Poh H L, Yao J T and Jasic T 1998 "Neural networks for the analysis and forecasting of advertising and promotion impact", *International Journal of Intelligent Systems in Accounting, Finance and Management* 7(4): 253-268
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.3036>

Rege J E O 1991 Genetic analysis of reproductive and productive performance of Friesian cattle in Kenya. II. Genetic and phenotypic trends. *Journal of Animal Breeding and Genetics* 109: 424

SAS 2003 Procedures guide for personal computers (version 9.1 edition). SAS Institute Inc, Cary, NC, USA

Van Tassel C P and Van Vleck L D 1991 Estimates of genetic selection differentials and generation intervals for four paths of selection. *Journal of Dairy Science* 74:1078–1086
<http://jds.fass.org/cgi/reprint/74/3/1078.pdf>

Weka 2005 Data Mining Software in Java <http://www.cs.waikato.ac.nz/ml/weka/>

Witten I H and Frank E 2000 Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann. Academic Press, USA

Yao J T, Teng N, Poh H L and Tan C L 1998 "Forecasting and analysis of marketing data using neural networks", Journal of Information Science and Engineering 14(4): 523-545
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.5556>

Received 16 October 2008; Accepted 31 January 2009; Published 18 April 200