Improving Dairy Yield Predictions Through Combined Record Classifiers and Specialized Artificial Neural Networks

F. Salehi, R. Lacroix, and K.M. Wade

Introduction

Artificial neural networks (ANNs) are not contingent upon the particular form of data. However, their performance may be improved by optimizing network architecture, training cycles, learning parameter values, input combinations, and data preprocessing. Although in large bodies of information ANNs can unveil hidden relationships, assortment of subgroups which share a particular behaviour, and training of specialized ANNs for each sub-group may improve the results. The fact that in some applications neural networks have demonstrated impressive success in classification and prediction, offers the suggestion to devise multiple network systems comprised of both classifiers and predictors, in which the classifier would pre-categorize the data for the specialized ANNs to perform predictions. The hypothesis behind this work was that a combined system of multiple networks, each trained specially for a certain task would offer improved performance compared to only one network trained for prediction of unclassified records.

Objectives

The objective of this work was to use backpropagation networks for 1) classification of test-day records into

Since the specialized predictors were trained only for a certain category of data, their test files should only contain records from the corresponding category, and be free of misclassifications, as much as possible. Therefore, in order to develop a neural classifier for

two categories of low-yield (< 9,000 kg) and high-yield ($\geq$ 9,000 kg), and 2) prediction of milk, fat, and protein yield for each category by using specialized predictor networks. The results of the combined classifier-predictor system were then to be compared to those of a previous study which used the same data with only one network for predictions.

Experimental Procedures

Figure 1 demonstrates the flow of data through the combined classifier-specialized ANNs system. The original data set used for this research was supplied by the Québec Dairy Herd Analysis Service and consisted of Holstein test-day records, representing 32 herds, 6077 cows and 15,175 lactations. The average 305-day for milk, fat and protein yields were 7,21 8 kg, 263 kg and 227 kg, respectively.

One specialized and several semi-specialized networks were trained for prediction of milk, fat, and protein production for the two categories of low, and high-yield. The training file for each specialized ANN included all the records in the corresponding category, from a previously constructed 'balanced' data file, which had an equal number of high and low-

categorization of the records into low-yield and high-yield groups, two training procedures were used. The first, referred to as 'classifier 1', involved training with the balanced data set. The second, named as 'classifier 2', used a training set consisting of all records with milk

yield records (i.e. 4,430 in each category). This file was extracted from an original training data set of 31,263 records. The training files of the 'semi-specialized' ANNs, had an additional 5, 10, 20, and 50% of records from the other category. The test file, common to all networks, contained a mixture of data records with 26,814 low-yield and 4,449 high-yield records. The measure of network performance was the root-mean-square (RMS).

In accordance with a previous study, which used the same original data the architecture used for the predictor networks consisted of 17, 10, and 3 processing elements (PEs) in the input, hidden, and output layers, respectively. This provided grounds for comparison of the results of this study with those of the previous one. The output PEs corresponded to milk, fat, and protein annual yields. The learning parameters were maintained at their default values (learning coefficient: 0.3, momentum: 0.4, and epoch: 16), with a bipolar presentation of input data and a hyperbolic tangent transfer function.

yields at the border of cut-off value of 9,000 kg, ranging from 8,700 kg to 9,300 kg of annual milk yield. The border data (2,183 records) were extracted from the aforementioned original training data set of 31,263 records. For convenience in training and testing of the classifier, a binary

code was added to each record to replace continuous values of the yields with only its category, in both training and testing files. The classification ability of a network was measured by its classification rate which reported the percentage of correct classifications in each category, and also with respect to the total records in the test file.

In order to evaluate the performance of the combined system (Figure 1), a weighted RMS was used to consider results from the two specialized predictors, together, and was

$$RMS_c = \sqrt{\frac{RMS_1^2 * n1 + RMS_2^2 * n2}{n}}$$

calculated as follows:
where RMSc was the combined value, RMS_1 was the overall RMS of the ANN specialized in low-yield predictions using n1 number of records classified as low-yield; RMS_2 was the overall RMS of the ANN specialized in high-yield predictions using n2 number of records classified as high-yield, and n was the total number of test records.

Results

As a first step, the unclassified records were tested with all the

Table 1 allows comparison of results from predictions of classified records by specialized networks, with those from prediction of unclassified records by a non-specialized network, from a previous study, which used the same data and network characteristics. The RMS values from the combination of two specialized networks were weighted with regard to the number of records treated by each network (equation 1). It must be noted that since unclassified records were contained in one single file only, both yield categories were treated by one network. In this case the overall

predictors. The range of RMS values for the specialized to the least specialized networks, for low-yield and high-yield categories were from 440 kg to 481 kg, and 522 kg to 636 kg, respectively. These indicate that networks trained exclusively for a particular class of data offer advantageous performance for that class, compared to those trained with, and for, unclassified records. Further, the results generally suggested that the overall RMS, as a measure of network performance, should be used with prudence, when different categories are involved.

The classification results by the two classifiers showed that training with cut-off point border values, compared to training with the whole data set, improved the classification rates from 92.3% to 99.7%. This reduced the number of misclassified low-yield and high-yield records from 2,152 and 225, to 46 and 39, respectively. Considering that the number of training records in the balanced file (8,860) was four times that in the cut-off border data file (2,183), this improvement suggests that the determinant records for classification were those with milk yields at the border of cut-off point value.

The predictions for the correctly classified records of test files containing the records as classified RMS was a weighted value calculated with regard to the number of records in each category treated by the same network. The values in Table 1 show that the lowest overall errors were attained by using the classifier trained with cut-off point border values of yield, combined with the exclusively trained predictors for low-yield and high-yield categories. However, prediction errors from records classified by the classifier trained with the balanced data set were considerably higher. This is attributed to the large errors in predicting misclassified records. The

by classifier 2, closely followed the overall predictions for all the records in the files. This, was attributed to the large number of correctly classified records. Comparison of the two groups of predictors for classified records of classifier 2 revealed that specialized low-yield predictors produced smaller errors for the correctly classified records, than the high-yield predictors. However, the misclassified records were considerably better predicted by the latter. These could be explained by the fact that in the training of low-yield predictors the number of records close to 9,000 kg yield was limited, thus large errors occurred for the misclassified records (over 9,000 kg). However, training of the high-yield predictors involved a large number of records with yield close to 9,000 kg, which facilitated predictions for the unclassified records of less than 9,000 kg yield. Similar to the previous situation, for both sets of predictors, addition of irrelevant records in the training files, with regard to their specialization, caused the RMS to rise for the correctly classified records, and to decline for the misclassified records, progres-sively.

RMS values from predicting unclassified records were about 12% higher than that from the best performing combined system. These results indicate that specialized networks exclusively trained for prediction of a certain category of records provided the lowest errors for that category. Further, they indicate that in a combined system a high performing classifier plays a major role in obtaining low prediction errors. Finally, the results in Table 1 indicate that the better performing classifier-predictor system achieved substantially lower errors that the

previous study.

Impact

The results in this study indicate that a combined classifier-predictor system of artificial neural networks offers improved predictions as compared to those by a single network from unclassified data. The magnitude of this improvement compared to the previous study was about 13% reduction in the RMS values.

The most notable points from this

- Specialized neural networks trained for prediction of a certain category of records produce betteroverall results than networks trained with mixed data.
- Specialized ANNs commit larger errors for records not in their specialization than nonspecialized networks.
- For well classified records, the improved predictions of specialized networks present much greater weight than the increased errors in prediction of misclassified records.
- The ability of a classifier improves as the training concentrates on determinant records with values close to the cut-off point.

The results and conclusions of this study provide new grounds in improvement of neural network modeling. Since the methodology in the present work was not data-dependent it could be applied to any situation where pre-classification of data may seem appropriate. In future work, classifications may be further improved by developing classifier trees to refine classifications in a step-wise manner. Although, the required time and effort for data manipulation and network training and testing may be considerable, application of the above methodology may result in valuable guidelines in different fields of research.

Table 1. Weighted RMS from combination of specialized networks with classified and unclassified records.

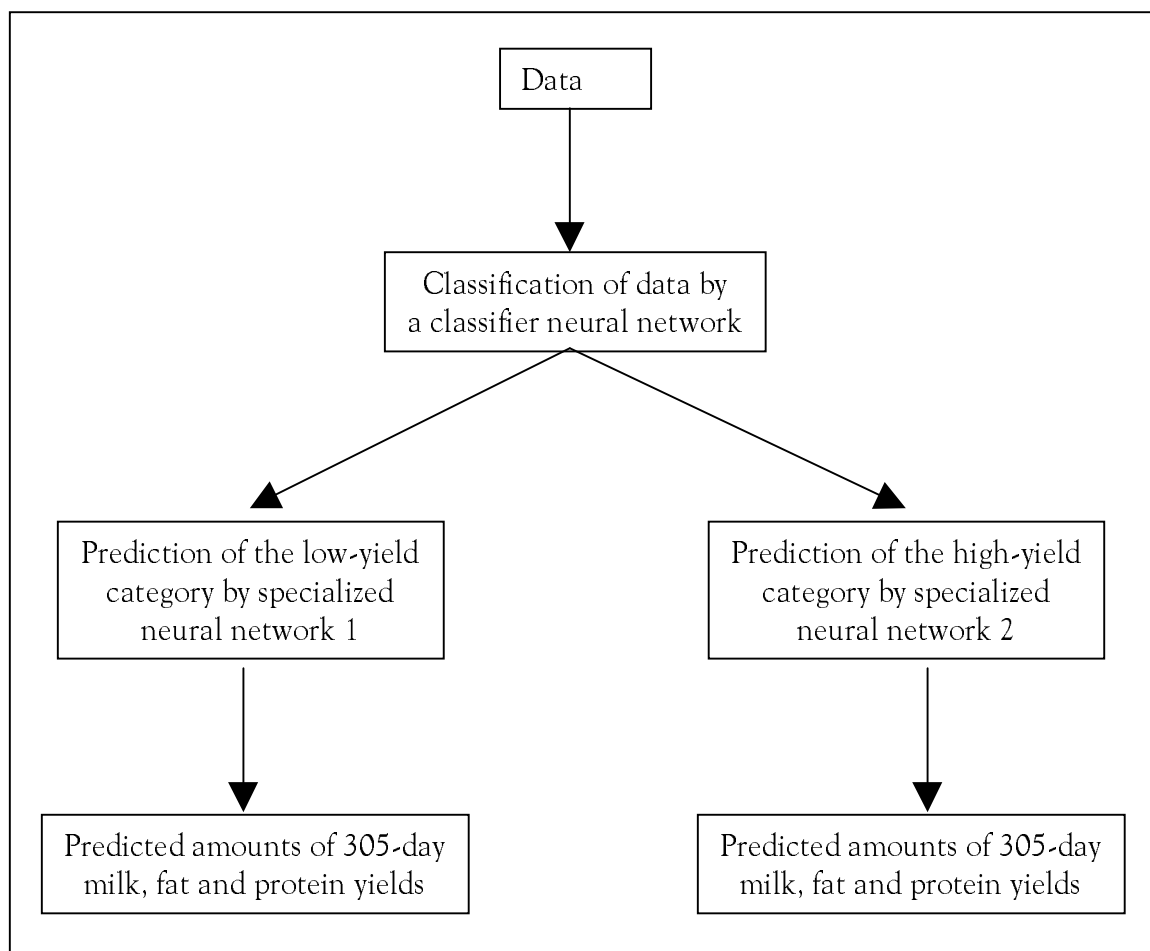| Records classified by	Combination of specialized low-yield, and high-yield predictors		
	Milk	Fat	Protein
Classifier trained with border values	455	19.9	14.8
Classifier trained with balanced data set	554	22.1	17.4
Lowest ANN RMS with unclassified records	516	21.2	16.5
Previous study	524	21.5	16.6

Figure 1.   Flowchart of record classification and dairy yield prediction