ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 69-74

The application of neural networks to the paper-making industry

P. J. Edwards[†] A.F. Murray[†] G. Papadopoulos[†] A.R. Wallace[†] and J. Barnard[§]
[†]Dept. of Electronics and Electrical Eng., Edinburgh University
[§]Tullis Russell, Markinch, Scotland

Abstract. This paper describes the application of neural network techniques to the paper-making industry, particularly for the prediction of paper "curl". Paper curl is a common problem and can only be measured reliably off-line, after manufacture. Model development is carried out using imperfect data, typical of that collected in many manufacturing environments, and addresses issues pertinent to real-world use. Predictions then are presented in terms that are relevant to the machine operator, as a measure of paper acceptability, a direct prediction of the quality measure, and always with a measure of prediction confidence. Therefore, the techniques described in this paper are widely applicable to industry.

1. Introduction

This paper describes the application of neural network techniques to the paper-making industry, particularly for the prediction of paper "curl". By representing the task first as a classification and then as a regression problem, and also by calculating confidence intervals, we have made the tool (a neural network) fit the practical needs of the end-user. We present parameters characterising the current paper reel as inputs to a neural network and train the network to predict whether the resulting level of curl will be within a required specification (i.e. "in-specification") — a classification task. In parallel, we present these same data to another network and train it to predict the absolute level of curl, i.e. a regression task. Perhaps most importantly, we also put these two predictions in context by including confidence measures at every stage thus providing the machine operator with a powerful and insightful tool. The machine operator is then presented with a "red-light/green-light" indication of paper acceptability, a neural regression model on which the parameters can be altered to reduce curl if necessary and a clear indicator of the reliability of both diagnostics.

Paper curl is a long-standing problem that has received much interest in paper-making research, where the causes of paper curl have been studied [2], and where attempts have been made to control it using heuristic techniques [3]. Paper curl is simply the tendency of paper to depart from a flat form and is affected by a number of complex, inter-related factors. Traditionally paper curl has been controlled with limited success using variation in drier temperatures or humidity levels. However, because it may only be measured off-line after an entire roll has been produced, its control is difficult and costly. Although out-of-specification paper may be re-pulped

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 69-74

in limited quantities, bad curl is a significant problem, wasting plant time, engineering time and energy.

In this paper we describe the development of neural network models to encapsulate the non-linear processes underlying paper curl. These models can be used as a powerful tool for the reduction of paper curl, thus enhancing quality and reducing waste. In all of the work described, in terms of processing and modelling, the approach that we take can be applied to any task involving neural networks. Furthermore, they have been developed in the context of an imperfect data collection process, typical of that found in many manufacturing operations and the combination of techniques used has particular relevance to such an environment.

2. The Database

The database provided by Tullis Russell for the purpose of the work described here has a number of limitations, including missing records and measurement errors. Not least in this respect is the measurement of curl itself. Although curl is a simple quality measure, measuring curl is far from trivial. While it would seem naturally advantageous to measure curl continuously, to date this has proven to be impossible as standard techniques [3] require the paper to be dried under controlled conditions before measurement. At Tullis Russell curl is measured after individual reels have been manufactured, leading to "out-of-specification" reels being scrapped, and a retrospective adjustment made to machine settings, according to unwritten heuristic rules developed by the skilled operators. The measurement is made using a sample of paper taken from the end of a reel and by cutting a cross-shape using a template. A glancing angle light source is then used to cast a shadow due to the curling paper at the centre of the cross. After a period of a few minutes has lapsed to allow the paper to relax the shadow is measured by hand, quantised to 5mm intervals. Therefore there may be error in the measurement due to quantisation, operator error, paper misplacement, failure to allow for sufficient relaxation time, etc... Variability in the accuracy of curl and other variable measurement could lead to significant model error [5] and while we are developing an improved curl-measurement system, for the study reported in this paper, the limitations of the database are taken as an additional constraint to the modelling process.

Various parameters are measured during the manufacture of a reel of paper. These parameters were used to classify whether the current process settings and paper specification would lead to curl that was within a required specification and additionally the level of curl that would result.

3. Preprocessing and Training

To preprocess data supplied directly from the paper-making plant a number of operations were performed. Firstly the real and symbolic data fields within the database were combined into a form that could be used for neural network training. In the case of symbolic data it is important that each field, for example the grade of the paper

— one-of-three for the purpose of this task, is encoded to avoid creating an artificial "weighting" to any case. Commonly a 1-of-N code is used. However, this scheme is inefficient and we use a more concise one, where the 1-of-N coding becomes 1-of-$N-1$. Transforming the 1-of-N code geometrically to be the vertices of a hyper-tetrahedron, the codes are calculated. In this case N=3 and the transformed codes are (0.0,0.0), (1.0,0.0) and (0.5, $\cos(\pi/6)$). This reduction in dimensionality is important in that it reduces the collinearity in the input vector, especially when there are significant numbers of symbolic parameters, which will greatly simplify the requirements of the training algorithm.

The second stage of preprocessing involved selection of the principal components within the data using the Karhunen-Loéve transformation [1]. For the case of the classification task the largest nine eigenvalues (or principal components) were used, while for the regression task the largest eight were used. As the principal component transformation technique is only concerned with manipulating input data, it is unable to give any insight into how important a component will prove to be in a prediction task. This information can only be determined by extensive experimentation. Therefore for the classification and regression tasks, the optimal combination of principal components was chosen through experimentation in each case. However the transform removes any correlation between parameters and scales data to have unit variance in all dimensions, thus greatly simplifying neural network training.

To develop models for the two tasks we use Multi-layer perceptron neural networks with a sigmoidal output stage for the "in-specification" prediction and a linear output stage for the curl prediction task. The classification network contained ten hidden units and the regression network twelve. These architectures were chosen via extensive experimentation. Network optimisation was performed using Bayesian inference [4], with a Gaussian prior for the weights. We also used Mackay's evidence based framework for dynamic calculation of data noise variance and the hyperparameters defining the weight prior. This optimisation scheme also naturally allows the calculation of confidence measures that we will discuss in depth later. Underlying the Bayesian formalism we used a steepest descent optimisation scheme with incorporated line-search. For the regression task a least squares cost function was used, while for the classification task we used cross entropy.

For both the regression and the classification task multiple networks were used and combined into a committee. The committee output was simply the averaged output from the members [1]. For the purpose of the experiments described here, this form of committee proved to be the most reliable. In general, more complex methods of forming the committee weightings would perhaps prove successful. In the Bayesian framework the assumption is that the network weights are normally distributed and give rise to normally distributed outputs, where each output y_n is effectively the mean of the distribution and σ_n the associated standard deviation. In the committee therefore we assume that each network makes an approximation to the "true" distribution of outputs which has mean y_{COM} and variance, $\sigma^2_{\mathrm{COM}} = <\sigma_n^2> - <y_n>^2 + <y_n^2>$.

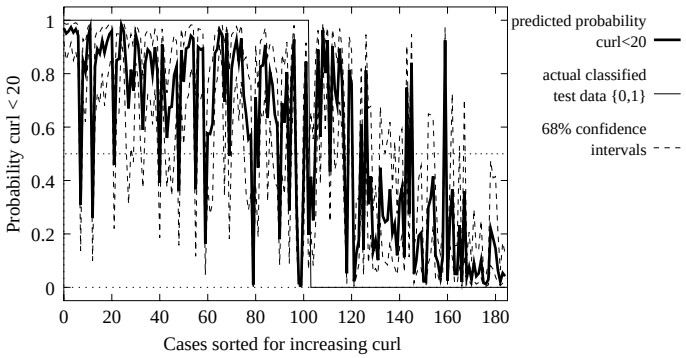

Figure 1: Graph showing variation in the predicted probability of curl < 20 for all cases in the test data set. The actual test data are sorted for increasing levels of curl and classified $\{0,1\}$, where "1" indicates "in-specification" (i.e. curl < 20).

4. Results

For the classifier experiments 40 networks were trained to classify the current paper characteristics as leading to paper "in-specification" or "out-of-specification". For the purpose of this experiment a level of curl less than 20 was used as the limit of acceptable curl. This level was chosen as typical. Different grades of paper have different acceptable levels. The results of these experiments are shown graphically in Fig. 1, depicting a classification error rate of 18.92%. The test cases are sorted for increasing levels of curl and Fig. 1 shows that the majority of the errors (where the classification boundary is set at a probability of 0.5) occur at the centre of the graph, i.e. for cases where the measured curl is 20 or near 20. The graph also shows 68% confidence intervals. Clearly the networks have been able to classify the cases to a usable degree of accuracy.

Fig. 2 shows the results for 40 networks trained to predict the absolute level of curl and tested on the same data set as above. Clearly the model has encapsulated the trend underlying these measured data, although with some imprecision. The prediction of extremely high levels of measured curl is poor. This is perhaps due to this part of the model being under-represented in the database (note that the data are densest for low curl), or that the process is different for levels of curl greater than 40. In practice, however, some accuracy in the critical $10 <$ curl < 30 region is most important and the predictor achieves adequate accuracy in that critical regime. The prediction of curl is therefore possible to limited but usable accuracy using the database provided and a neural network model. In addition to the curl prediction, Fig. 2 also shows 68% confidence intervals for those predictions.

The experiments described above have tested the networks on true test data — drawn from the same source as the training and validation data, but not used at all during training. The results allow us to judge model accuracy when tested with such data. In practice, however, the operator will be likely to want to see what effect changing a certain variable will have on the resultant curl. It is in this area that the confidence

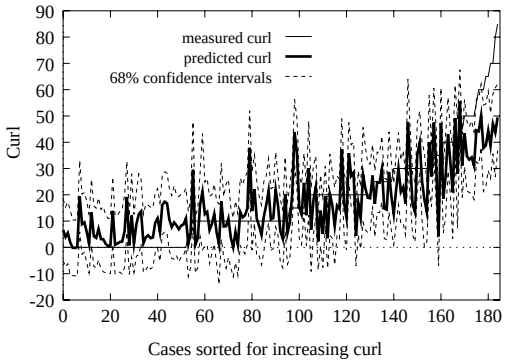

Figure 2: Graph showing the variation in the absolute value of curl as predicted by the committee of networks, as measured at Tullis Russell and contained in the test data.

measures become important as it is vital to have a measure of the validity of the model. Also, within the bounds of the training data set we can expect the model to be able to interpolate between data points given a high enough data density. However outside the bounds of the data set, extrapolating the model is more risky and this should be reflected in the confidence measures (see [1] for example). To assess the models and confidence measures, experiments were carried out by varying a single parameter while holding the others constant and noting the prediction and the associated confidence. This was done for the disparity between surface moisture on different sides of the paper as it passes through the coating machine. In addition here, rather than give the operator confidence intervals to decipher, we use our knowledge of the task and define upper and lower limit of acceptable variation, to calculate confidence as a percentage. For the regression and classification networks the limits were set to ± 10 and ± 0.2 respectively. The percentage confidence may thus be calculated by integrating the normally distributed outputs, defined by y_{COM} and σ_{COM}, between these limits.

The results of the experiment are shown in Fig. 3 for the classifier network. In addition to the prediction and confidence, the graph also shows data in the training set in that dimension. The input vector used in this experiment was taken from the test data set, where the original surface moisture value was 11.1, using the same scale as the graph. The graph clearly shows that as the parameter is adjusted there is a change in the output prediction and in addition that as the parameters exceed (whether positively or negatively) the bounds of the training data, the confidence falls. Clearly as this is a high dimensional problem it is unclear as to exactly how the whole data set relates to the single dimension shown, but the falling confidence is encouraging. In addition a physical interpretation of the results suggests that as the difference in the surface moisture increases in either direction so does the curl. While this interpretation also seems qualitatively reasonable we can tell nothing of the predictive accuracy of the results.

The variance used to calculate the confidence is due to two components, uncertainty in the data and uncertainty in the model parameters. Estimating these components using evidence maximisation, [4], as we do here only gives valid estimates when

the model is used within the bounds of the training data. In terms of further work we aim to improve and generalize our estimates of these components by, in particular, including a measure of data novelty.

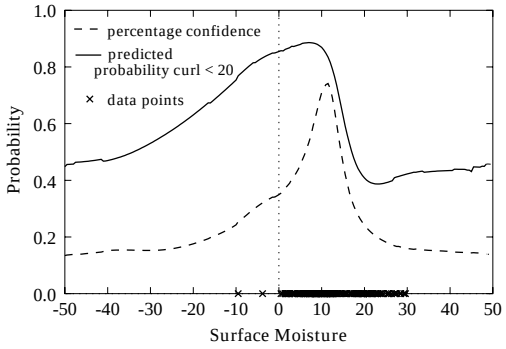

Figure 3: Graph showing variation in the probability of curl < 20 as the surface moisture varies. Also shown are data points in that dimension and percentage confidence.

5. Conclusions

Paper curl is an important quality measure in paper-making and as a problem has characteristics similar to other tasks in the manufacturing industry. Neural network techniques must be employed in the presence of erroneous and limited number of data, and always are viewed with suspicion as a "new technology". In this paper we find a solution to the prediction of paper curl and present the machine operator with as much relevant information as possible. The neural network approach thus provides a tool predicting a red/green light of paper acceptability, a direct prediction of the quality measure, and always, measures of confidence. These results are important in that they show that the combination of techniques used may be applied to complex and relevant practical problems in realistic working environments.

References

[1] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[2] L-E. Eriksson, S. Cavlin, C. Fellers, and L.Carlsson. Curl and twist of paperboard — theory and measurement. *Nordic Pulp and Paper Research Journal*, 2(2):66–70, 1987.

[3] E.T. Langevin and W. Giguere. Online curl measurement and control. *TAPPI Journal*, 77(8):105–110, 1994.

[4] D.J.C. MacKay. Bayesian framework for backpropagation networks. *Neural Computation*, 4(3):448–472, 1992.

[5] V. Tresp, S. Ahmad, and R. Neuneier. Training neural networks with deficient data. In *Proc. Neural Information Processing Systems (NIPS) Conference*, pages 128–135. Morgan Kaufmann, 1994.