Appears in *Proc. 1995 IEEE Int. Conf. Neural Networks*, pp. 37–42.

Colour Bake Inspection System Using Hybrid Artificial Neural Networks

Jeffrey C.H. Yeh[†], Leonard G.C. Hamey[†], Tas Westcott[‡] and Samuel K.Y. Sung[†]

[†]Department of Computing
Macquarie University,
NSW 2109, Australia
Email: len@mpce.mq.edu.au

[‡]Arnotts's Research Centre
Arnott's Biscuits Ltd.
Homebush, NSW 2140, Australia

Abstract

The bake level of biscuits is of significant value to biscuit manufacturers as it determines the taste, texture and appearance of the products. Previous research explored and revealed the feasibility of biscuit bake inspection using feed forward neural networks (FFNN) with a back propagation learning algorithm and monochrome images. A second study revealed the existence of a curve in colour space, called a *baking curve*, along which the bake colour changes during the baking process. Combining these results, we proposed an automated bake inspection system with artificial neural networks that utilises colour instead of monochrome images. In this paper, we present the implementation of the inspection system with a hybrid neural network of self–organising maps and FFNNs. The system was tested and its grading performance on biscuit bake levels was evaluated and compared to that of a trained human inspector. We found that the proposed colour system with a hybrid neural network performed significantly better than the human inspector.

1. Introduction

Quality control is of major importance in the food industry. Consequently, the inspection of the biscuit bake level occupies a major role in the biscuit manufacture process as it affects the taste, texture and appearance of the products. Currently, trained human inspectors are used to determine the bake level of biscuits. However, human classifications are prone to both short and long term variations due to the subjectivity of human perception and this results in unstable product quality. An effective alternative is therefore desirable.

In recent years, automated machine inspection systems have been employed in many quality control applications and have many advantages over their human counterpart in terms of economy and performance [1,2,3]. Classification techniques such as statistical pattern recognition [3,21], machine learning methods, expert systems [21], fuzzy logic [23] and artificial neural networks (ANN) [2,21] have been used by these systems. Applications of ANNs in the automated machine inspection systems have resulted in some encouraging results [4,22]

In our previous study, we found that an automated machine inspection system using a feed forward neural network (FFNN) with back propagation (BP) learning algorithm can determine the bake level of biscuits from their monochrome images. The developed monochrome system performed comparably with the trained human inspector [5]. In a separate study of biscuit bake colour, the biscuits' pixels were plotted in a RGB colour cube with red, green and blue (RGB) axes and a colour curve called the *baking curve* was observed. The biscuit colour during the baking process was observed to be changing along its baking curve. This indicated that biscuit colour development during baking process does not change in the intensity level only, but also in the colour saturation levels. A monochrome system that uses only the intensity dimension may not be optimal since the colour saturation information is left out. Conversely, a system that uses colour images can classify the biscuits on the basis of all the relevant characteristics.

In this paper, we present the implementation of the colour inspection system using a hybrid artificial neural network and colour images that we had proposed in earlier publications [8, 20]. This paper describes the colour bake inspection system and presents results of a study comparing the system's performance with that of a trained human inspector.

2. Proposed Colour Inspection System

The number of FFNN weight connections affects the number of training samples required for generalisation [7] and the speed of training. Both the colour and intensity raw images are large, each ranging from 200x100 to 500x500 pixels. The raw images are therefore unsuitable as the direct input to the FFNN, since the large number of pixels results in a large number of weight connections. The raw data can be preprocessed into meaningful histograms without loss of important features required by the neural network classifiers. The histograms are scaled to the input dimension of the required FFNN topology. Monochrome images are readi-

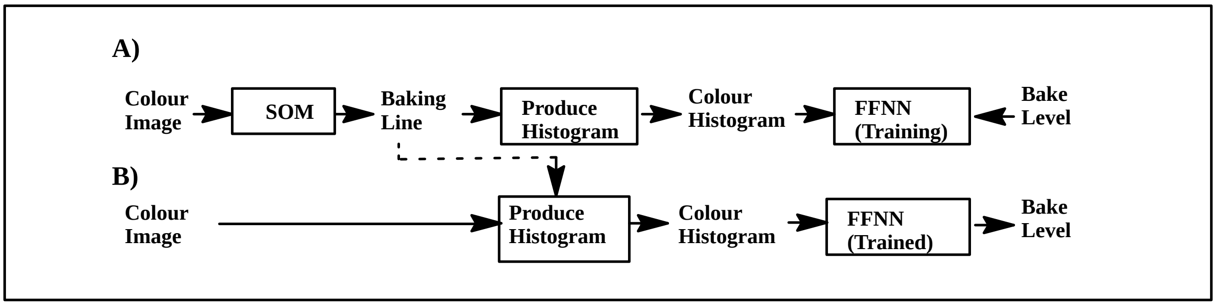

Figure 1 : Proposed system for bake inspection using hybrid neural networks and colour images. A) shows the process during training, and B) shows the trained system in action.

ly converted into histograms by counting the frequency of the pixel intensity levels in a grey scale [5]. Histograms of colour images are more complicated as a colour image contains levels of the three colour bands (RGB), and so is three times the size of the monochrome images. A naive method to produce the colour histogram is to produce three intensity histograms for the three bands, but this may not capture the shape information as described by the baking curve in the 3D space. A better method is to count the frequency of colour pixels along the baking curve, since the curve, and so the resultant colour histograms, describes the relationship between the baking duration and the change in the bake colour.

Hence in an earlier paper [8], we proposed a system [fig 1] that uses the Self–Organising Map (SOM) to extract a *baking line* from a set of colour images. The histogram of each colour image in the set can be constructed along this baking line. The colour histograms of all the biscuits are used as the input data for a FFNN for training and testing. New biscuits can then be directly histogrammed along the existing baking line and graded by the trained FFNN.

3. Data Collection and Preprocessing

298[1] samples of Milk Coffee biscuits, made up of an equal distribution of under, correct and over bake level, were collected and digitised as in [8]. The images were then calibrated to remove the effect of illumination variations occurring during the imaging process. They were also randomly divided into 5 groups of 60 with each group containing an equal proportion of the three bake levels. A trained human inspector classified the biscuits in each group according to a bake level scale of 1 to 7. The human classification experiment was repeated 10 times for each group to determine the short term variation in human classification. Since there were variations in the bake scale used by the human in the different experiments for a group, we sorted the raw scores for each experiment and assigned new *normalised scores* from the range of 0 to 1 according to the rank position of a biscuit in an experiment. The normalisation ensures that the scales used in different experiments are the same and so the differences in the normalised scores of a biscuit are comparable. For each sample, the target bake level used to train the FFNNs was the average of the normalised scores for 10 experiments. FFNN performance was measured by root mean square (RMS) error. To measure the human performance, we first computed the standard deviation of the normalised scores for each sample. The human performance measure was taken as the RMS of these standard deviations. This enables human performance to be compared with the performance of the hybrid neural networks in section 7.

4. Baking Line by Self–Organising Map

A SOM is characterised by its ability to create the topological feature map that models the probability distribution function of its training samples [9,10]. This allows a Nx1 dimension, or a string–like SOM with N nodes to model itself along the baking curve according to the pixel distribution. A baking line is formed when the SOM nodes are connected. The baking line preserves the shape of the baking curve, capturing the essential nature of bake colour development. A colour histogram constructed from a colour image along the baking line stores this essential information and can be easily used as the input to a FFNN. The drawback of the SOM, on the other hand, is that the two ends of the baking curve, which are important for classifying extreme bake levels, are not reached by a SOM of insuffcient nodes. This is because there are less training samples of extreme bake levels and so less distribution of pixels in those two ends of the curve. The biscuits used in our research had even bake level across their surface so that the pixels from a biscuit sample are fairly similar, producing equal distribution of pixels in a colour cube. We also ensured the samples to have an equal distribution from the three bake levels (under–bake, correct–bake, and over–bake) so that the whole baking curve had a re-

---

[1] 300 samples were collected originally, but two samples were damaged leaving 298 samples. Since the sample set was large, the two missing data did not have significant effect on the balance of the sample number of the three bake classes.

asonably equal distribution of pixels throughout. The even distribution of pixels enables the baking line to cover most of the baking curve.

We cropped a section of 20x10 pixels from the centre of each biscuit to represent the evenly baked milk–coffee biscuit. The cropped pixels from all 298 samples were then shuffled randomly before each pixel was used as a training sample to the SOMs. The software that we used for generating and training SOM was SOM_PAK [11]. Each SOM has a 3x1 input dimension for the RGB values of each pixel, and a Nx1 output dimension. Each SOM was trained in a two stage process. A bubble neighbourhood function with a radius of 0.8*N and a learning rate of 0.05 were used at the first training stage while at the second training stage, the radius was reduced to 0.25*N and the learning rate was reduced to 0.02 for finer training.

Output dimensions of 5, 10, 15, ..., 40 were investigated to determine the optimal SOM topology. The optimal SOM topology was determined based on two criteria in this research. Firstly, it must cover as much length of the baking curve as possible for the baking line to reasonably represent the baking curve. Secondly, the SOM must also preserve the smooth shape observed in the original baking curve. We found that a SOM with insufficient nodes does not cover the ends of the baking curve, since the pixel distribution is less dense there. However, a SOM with excess nodes exhibited an oscillating behaviour within the baking curve as it positioned its nodes to cover the width of the sausage–like baking curve. Therefore, the optimal SOM topology covers the maximum length of the baking curve without exhibiting the oscillating behaviour.

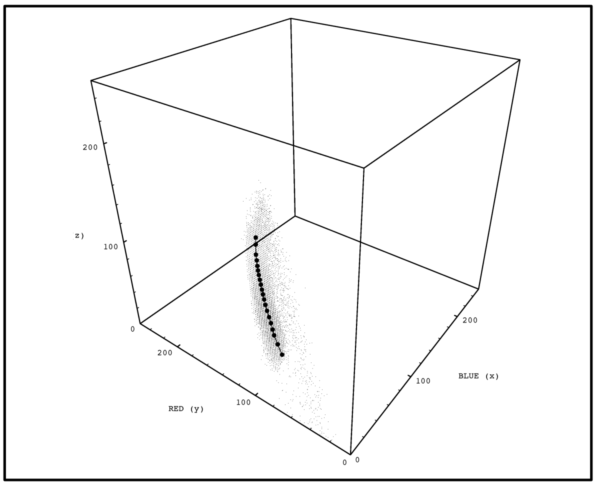

Figure 2 : Connecting the SOM nodes creates a baking line. The underlying baking curve was constructed by projecting the 298 milk–coffee samples used for training the SOM.

The effect of different training iterations was also studied. We tested 1, 5, 10 ... 40 training iterations through the whole sample set for the first training stage. The training iterations for the second training stage was 10 times that for the first training stage. An under–trained SOM does not position its nodes correctly while over–training wastes computation resources without changing the weights [10].

We found that the optimal SOM for Milk Coffee biscuits had 20 output nodes and was trained for 20 iterations through the sample set. It covered most of the baking curve's length while still preserving the smooth curvature of baking curve. Figure 2 shows the baking line constructed by connecting the optimal SOM nodes. Also shown in the background of the figure is the baking curve, in the same RGB colour cube, created by projecting the 298 biscuit samples used for training the SOM.

5. Histogram Of Colour Image

The histogram of the colour image of each training sample was constructed along the baking line created by the SOM. Each pixel of a colour image, when projected onto a colour cube, can be viewed as a digital signal. Sampling along the baking line at a rate lower than twice the rate of the original data signals (or *down–sampling*) may result in an aliasing effect. This can be overcome by filtering the data with a Gaussian filter to remove the high frequency signals that cause the aliasing effects [12]. In addition, it is useful to discard points that are too far from the baking line as they are likely to represent non–biscuit or noise–affected pixels. For this purpose, a Gaussian function, based on the distance of each pixel from the baking line, was used to weight the pixels' contribution to the histogram. The parameter σ_y controlled the spread of this filter and thus the sensitivity of the system to outliers.

We set σ_y to be the standard deviation of all the pixels' distances from the baking line. The baking line was interpolated with sufficient extra points to prevent aliasing, and each interpolating point along the baking line was a bucket in the resultant histogram. For each of the 298 samples, the pixels of the colour image were weighted at each interpolating point on the basis of the distance between the pixels and the interpolating points. The weighted values were then accumulated at each interpolating point to produce a colour histogram. 298 colour histograms were produced, each representing one biscuit. The histograms produced were further reduced in dimension by Gaussian filters to suit the input dimensions of different FFNN topologies. The histogram allows the colour data to be independent of its original dimension without the loss of essential colour information.

6. Grading by Back Propagation Neural Network

Although new guidelines for constructing artificial neural networks are emerging [7,13,14], it is still considered by many as more of an art than science [15,16]. We addressed the issue of selecting a suitable neural network architecture with an empirical approach by testing different topologies and determined the best network topology based on their performance. Although it was suggested in [7] that a neural network's generalisation of 90% accuracy can only be achieved with a number of training samples at least 10 times the number of the weights in the network, we have found that the cross–validation technique as described in [5,17] can also achieve a good level of generalisation without restriction on the number of weight connections or the number of training samples. Cross–validation has the advantage of terminating the training automatically, preventing over–training.

We investigated 5 topologies including 20–1, 40–1, 8–2–1, 9–1 and 5–1. The 8–2–1 topology produced the optimal performance in the monochrome system [5] and so it was a logical candidate for the colour system. Other topologies were selected to represent a range of input dimensions. The smaller topologies (5–1 and 9–1) satisfy the criteria set out in [7] that a neural network's generalisation of 90% accuracy can be achieved with a number of training samples at least 10 times the number of the weights in the network. A performance comparison between these smaller topologies and the remaining larger topologies reveals the generalisation ability of cross–validation.

The 298 samples which were collected and used in the SOM previously were randomly divided into three sets with equal numbers[2]. One set was used for training the network. A second set was used for the cross–validation process during training. The test performance of the trained network was obtained from the third set which was not excluded from the network training. To ensure a balanced result, six training experiments were performed for each topology with the three sets permutated as training, cross–validation and test samples and the averaged error from the six experiments were reported. In this research, we used the software package Aspirin simulator V6.0 [18] with our own enhancements for cross–validation.

All training experiments had a learning threshold of 0.00005, a learning rate of 0.05, a momentum of 0.95 and an error signal random bias with variance 0.05 [18]. During each training experiment, cross–validation error (RMS) was evaluated after every 100 training epochs. In order to ensure that a global minimum in generalisation error was reached, training was terminated after cross–validation's error did not improve for 300 successive evaluations. The weights of the network at the optimal cross–validation error were saved as the final weights. The trained networks were then tested on the test set for their true and unbiased generalisation performances on unseen data.

Table 1 shows the performance of the trained neural networks on the cross–validation and test sample groups for the five topologies tested. The reported performance for each topology is the average of six RMS errors by permutating the three sample sets for training, cross–validating and testing. The errors on cross–validation samples are used for selecting the optimal topology, while the test errors provide an unbiased performance statistic of the trained networks on unseen data.

Topology	Cross–Validation RMS Error	Test RMS Error
5–1	0.0474	0.0478
9–1	0.0453	0.0457
20–1	0.0451	0.0455
40–1	0.0453	0.0458
8–2–1	0.0416	0.0424
Human Inspector		0.0731

Table 1 : Averaged performance of trained neural networks on cross–validation and test sample groups for each of the topologies tested. Also shown here is the human's performance for comparison.

We observed from table 1 that there are no significant difference between the networks' performances on the cross–validation set and the test sets. Therefore stopping training at the lowest cross–validation error can allow the neural network to generalise on unseen data. Furthermore, the table also shows that the networks with more weights than one tenth of the number of training samples perform as well as those with less weighs. These prove empirically that the cross–validation technique preserves generalisation even when there are more weights than allowed by the number of training samples. The same empirical conclusion was also drawn by [24] which was however criticised in [25] for using correlated training and cross–validation samples in the experiments. The samples in this research were all collected randomly and then randomly divided into sample sets. This ensures independence of the samples from each other and the validity of the results. Therefore based on the experimental observations, we claim that training by cross–validation can prevent over–training and can preserve a neural network's generalisation, regardless of the size of the neural network.

We performed a t–test on each pair of topologies in table 1 to test the null hypothesis that the average RMS errors of two topologies on the cross–validation set are equal. Table 2 shows the p–values from the t–tests on different pairs of the topologies.

It is clear from table 2 that at 2% significance level, the five topologies are divided into three groups of three different average cross–validation RMS errors. 5–1 the

---

[2] The three groups consist of 100, 99 and 99 samples.

first group; 9–1, 20–1 and 40–1 the second group; and 8–2–1 as the third group. The significant difference between the first group and the second group, or the 5–1 topology and the other N–1 topologies, suggests that the grading performance can be significantly reduced when the input dimension is too small. This is because that the important grading criteria is hidden too deeply in the input and is obscured by the other properties [19]. These results indicate that in general, a suitable topology for the colour inspection system must have at least an input dimension of nine, the next higher input dimension tested.

Topology	5–1	9–1	20–1	40–1	8–2–1
5–1		0.016	0.001	0.002	0.000
9–1	0.016		0.075	0.032	0.003
20–1	0.001	0.075		0.561	0.000
40–1	0.002	0.032	0.561		0.000
8–2–1	0.000	0.003	0.000	0.000	

Table 2 : P–values from the t–tests on different pairs of the topologies in table 1. The null hypothesis is that the average RMS errors of two topologies on the cross–validation set are equal.

The t–tests also indicate that the average performance of ANN with a hidden layer performs significantly better than with no hidden layers. This is shown in the last row of the table 2, where the average RMS error of 8–2–1 topology is significantly different from the other N–1 topologies even at 0.1% significance level. This suggests that the grading of Milk Coffee biscuit is a non-linear problem requiring hidden nodes. Therefore the optimal ANN of all the topologies tested for the grading task is the 8–2–1 topology. Its performance on the test samples is a good indication of the performance of the optimal network on unseen data and this can be compared to the human inspector's performance on unseen data. Our previous work on monochrome images of a different biscuit type tested topologies with up to five hidden nodes and also found that the 8–2–1 is the optimal topology [5].

7. Performance Comparison

Also shown in table 1 is the performance of the trained human inspector in classifying the same set of 298 milk–coffee biscuits. The RMS error was calculated from the normalised and averaged scores of the 298 biscuits as described in section 3. The human inspector was well trained for grading Milk Coffee biscuits. The biscuit samples used by the ANN experiments were presented to the inspector during the experiment only. This ensures that the performances of the networks and the human inspector on unseen data are comparable.

Since the target values of the test samples used in the ANN experiments were derived from the grades given by the human inspector the ANN's performance on the test sets includes a component of human error. Therefore the *true* RMS errors of the ANNs are actually over–estimated in table 1. When comparing the performances in that table, it should be noted that the true performance of the ANNs are at least as good as the measured performances from the table.

It is clear from the result that all the neural networks perform significantly better than the human inspector. The optimal ANN (8–2–1) is at least 41% better than the human inspector with other ANNs performing at least 34% to 37% better. Since the biscuits were classified by the human inspector over a period of two days, the result indicates that in the short term, the trained networks are more consistent in the classification task than their human counterpart. The long term performance of the neural networks will also exhibit a consistency not found in human perception. These results suggest that an automated bake inspection system using hybrid neural networks and colour images is an effective alternative to human inspectors for the bake inspection task. Performance of the colour system on other more complicated biscuits and a comparison between inspection systems using colour and monochrome images are currently under investigation. This will determine the applicability of the inspection system across a spectrum of biscuit types and the advantages of using colour images in the system.

8. Conclusion

Biscuit bake grading using human inspectors is not consistent either in short or long term performance. We proposed an automated bake inspection system using hybrid artificial neural networks and colour images. The proposed system was implemented and it performed better than the human inspector. We also found that the cross–validation technique can prevent over–training and preserve generalisation of an ANN, even when the number of weights are larger than that allowed by the number of their training samples. Further investigations are being carried out on the performance of the system on other types of biscuits and the improvement of using colour over monochrome images.

9. Acknowledgment

We would like to thank Arnott's Biscuits Ltd., in particular Anne Watson, for the continuing support in this research project, especially with regards to providing data, technical facilities and financial support.

10. References

[1] Bob Sperber. Prime Time For Machine Vision, *Food Processing*, vol 53, No 10, pp19, 21, 24–25, October, 1992.

[2] R.M. Hodgson. Digital Image Processing – A Developing Technology For Enhancing Productivity, In *Proceedings of Control*, 1992.

[3] L.G.C. Hamey, A.J. Watson, and C.T. Westcott. Machine Inspection of Biscuit Bake, in Proceedings of Digital Image Computing Techniques and Applications. (K.K. Fung and A.Ginige, eds), pp.124–129, Australian Pattern Recognition Society, 1993.

[4] Mike Donlin, Jeffrey Child. Is Neural Computing The Key To Artificial Intelligence?, Computer Design, vol 31, No 10, pp87–104, Oct, 1992.

[5] J.C.H. Yeh, L.G.C. Hamey. Biscuit Bake Assessment by an Artificial Neural Network, In *Proceedings of The Fifth Australian Conference On Neural Networks*, pp266–269, 1994.

[6] S.K.Y. Sung. *A Study Of Baking Curve*. B.Sc. (Hons) Thesis, Department of Computing, Macquarie University, Sydney, Australia, 1993.

[7] E.B. Baum, D. Haussler. What Size Net Gives Valid Generalization?, *Neural Computation*, vol 1, pp151–160, 1989.

[8] T. RayChaudhuri, J.C.H. Yeh, L.G.C. Hamey, C.T. Westcott. Baked Product Classification with the Use of a Self–Organising Map, In *Proceedings of the Sixth Australian Conference on Neural Networks*, pp152–155, 1995.

[9] M. Caudill. A Little Knowledge Is A Dangerous Thing, *AI Expert*, pp16–22, June, 1993.

[10] A. Hiotis. Inside A Self–Organizing Map, *AI Expert*, pp38–43, April, 1993.

[11] T. Kohonen, J. Hynninen, J. Kanga, J. Laaksonen. *SOM_PAK The Self–Organizing Map Program Package Version 3.1*, Laboratory of Computer ad Information Science, Helsinki University of Technology, Finland, 1995.

[12] J.D. Foley, A.v. Dam, S.K. Feiner, J.F. Hughes. *Computer Graphics Principles and Practice*, Addison–Wesley Publication Company, Sydney, pp617–647, 1992.

[13] C. Cortes, L.D. Jackel, S.A. Solla, V. Vapnik, J.S. Denker. Learning Curves: Asymptotic Values And Rate Of Convergence, NIPS, vol 5/6 pp327–334.

[14] F. Tamburini, R. Davoli. *An Algorithmic Method To Buid Good Training Sets For Neural–Network Classifiers*, Technical Report UBLCS–94–18, Laboratory For Computer Science, University Of Bologna, Italy, 1994.

[15] S. Schocken, G. Ariav. Neural Networks For Decision Support: Problems And Opportunities, *Decision Support Systems*, vol 11, No 5, pp393–414, June, 1994.

[16] J. Sietsma. *A Computational Overview Of Artificial Neural Networks*, Technical Report No. 13/91. Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia, 1991.

[17] J.C.H. Yeh. Baking Inspection Using Artificial Neural Networks. B.Sc. (Hons) Thesis, Department of Computing, Macquarie University, Sydney, Australia, 1993.

[18] R.R. Leighton. The Aspirin/MIGRAINES Neural Network Software: User's Manual. Technical Report MP–91W00050, The MITRE Corporation, 1992.

[19] B. Müller, J. Reinhardt. *Physics of Neural Networks, Neural Networks An Introduction*. Springer–Verlag, Berlin, 1990.

[20] C.T. Westcott and L.G.C. Hamey, ”Data Recognition System,” Patent Specification Australia, Arnott's Biscuits Limited, 1994.

[21] S.M. Weiss, C.A. Kulikowski. Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems, Morgan Kaufmann Publishers Inc, San Mateo, 1991.

[22] A.D. Whittaker, BS. Park, J.D. McCauley, Y. Huang. ”Ultrasonic Signal Classification for Beef Quality Grading Through Neural Networks,” in Proceedings of the Automated Agriculture for the 21th Century Symposium, American Society of Agricultural Engineers, Chicago, IL, 1991

[23] K. Unklesbay, J. Keller, N. Unklesbay, D. Subhangkasen. ”Determination of Doneness of Beef Steaks Using Fuzzy Pattern Recognition,” Journal of Food Engineering, vol 8, pp79–90, Elsevier Science Publishers Ltd., England, 1988.

[24] N. Morgan, H. Bourlard. ”Generalization and Parameter Estimation in Feedforward Nets: Some Experiments,” Advances in Neural Information Processing Systems, 2, pp630–637, Morgan Kaufmann Publishers, 1990.

[25] W.S. Sarle. ”Stopped Training and Other Remedies for Overfitting,” in Proceedings of the 27th Symposium on the Interface, 1995.