NETWORK ANALYSIS TECHNIQUES AS VISUALISATION TOOLS

R.P. Good & T.D. Gedeon

School of Computer Science and Engineering
The University of New South Wales
P.O. Box 1, Kensington 2033, Australia

1. ABSTRACT

Few advanced visualisation techniques have been developed to assist with the development and analysis of artificial neural networks. This is unusual, given that the attributes of neural networks make them excellent candidates for visualisation, and noting the advances in other fields that can be at least partially attributed to scientific visualisation methods and applications We present three network analysis techniques and demonstrate their effectiveness as visualisation tools within a complete back propagation support application. Such high level visualisation tools are extremely useful in neural network application development.

2. DISTINCTIVENESS MEASURE

Distinctiveness is a measure of the similarity between hidden units within an artificial neural network across its training set [1]. The distinctiveness of hidden units is determined from the unit output activation vector over the pattern presentation set. That is, for each hidden unit we construct a vector of the same dimensionality as the number of patterns in the training set, each component of the vector corresponding to the output activation of the unit. This vector represents the functionality of each hidden unit in pattern space. The recognition of similarity between vector pairs is measured by the calculation of the angle between them in pattern space.
Functional similarity is thus inversely proportional to the angle between vectors. Figure 1 shows the total sum of squares (tss) and the distinctiveness between units in a six hidden unit network. Three phases can be identified using the tss value: an initial phase; an intermediate stage (which may be continued indefinitely if the network is stuck in a local minimum or just prolonged if the network is on a plateau [2]); and a solution phase (unit functionality ceases to change).

The use of distinctiveness as an analytical tool during network training has been documented [2], and it has been shown to be an effective indicator of networks reaching a training plateau and as a post training network reduction technique. The use of this technique as a visualisation tool has previously been limited to two dimensional analysis. This work demonstrates the effectiveness of this technique as part of a real time development tool for networks with large data sets using three dimensional analysis and correlated with error vectors across the training set.

3. CAUSAL INDEX

The causal index [3, 4] is a measure of the effect that an input neuron *i* has on an output neuron *k*. It is effectively the rate of change of *k* with respect to *i* across the range of possible input values for input *i*. The causal index C_{ki} represents a relationship between the k^{th} output and the i^{th} input neurons. The sign of C_{ki} represents a positive or negative correlation between input and output neurons.

We calculate the full causal index, because we have found that a simplification made in [3, 4] which would allow a static analysis of the weight matrix does not hold for any of the networks we have studied [5, 6].

Figure 1. 2D *distinctiveness* plus total sum of squares

Figure 1. 2D Causal Index, GDP data

Figure 1. 3D Causal Index Surface, GDP data

Figure 2 is a simple plot of the causal index of each input in a 14 input, 8 hidden unit, and 1 output unit network. The data is selected economic indicators of developing countries [7], to predict their gross domestic product (GDP). Some of this work has been reported elsewhere [8].

The process of developing and interpreting causal indexes currently [5, 6] relies on the determination of various *characteristic* input sets to the network by simple arithmetic means, which can result in poor interpretations for networks with small training sets. From these developed causal indexes after training the network, it is possible to develop rule based explanations of a network? s behaviour at a symbolic level. Serendipituously, from the analysis of the convergence process for causality indexes a new technique for developing *characteristic* input patterns can be derived, however that is beyond the scope of this paper (see also [9 - 11]).

Figure 3 shows the 3D causal index surface for the same data. The lines running into the background are the increasing values of input units, and the hills and valleys represent positive and negative effects on the GDP. The most significant negative bulge on the left corresponds to low values of the percentage of the total labour force engaged in agriculture. This low value is relative to other developing countries. Thus, for median to high values of this input variable, there is relatively little effect on the GDP, however when the percentage drops below a certain value, there is an increasing effect on the GDP. This is quite plausible, and reasonable. The other two dips on the right, correspond to average food production, and total population. We will not discuss the interpretation of the graph further here with respect to the specific GDP network.

The interpretation of the 2D causal index shown in Figure 2 can be used to generate symbolic rules as mentioned above. Some of the same effect can be achieved through visualisation in a 3D surface as shown in Figure 3, in extraction of meaningful relationships between inputs and output.

The convergence of the causal index across training, had not been previously investigated, this aspect of the process has great utility as a visualisation tool.

Figure 4. 3D Causal Index Surface, C_{ki} convergence

Figure 4 shows the causal surface for a simple network (being XOR), as well as the causal convergence diagram at the bottom of the diagram. This was omitted for clarity in Figure 3.

The convergence diagram in Figure 4 shows the percentage change between causal surfaces. The causal surface shown corresponds to the rightmost edge of the convergence diagram. This is after the major peak, and represents the region after convergence.

The convergence diagram has two peaks, one corresponding to the transition between the initial and the intermediate stages, and from the intermediate to the solution phases, as shown in Figure 1. Clearly recognising the two peaks in the convergence diagram is simpler than recognising the transitions in the distinctiveness graph.

4. CORRELATION OF ACTIVATION PATTERNS AND NETWORK WEIGHTS

The information contained in neural networks is stored in a distributed fashion, the activation patterns of nodes and their corresponding weights then form the essence of stored network information. The functionality of a single weight can then be represented as a vector across the set of pattern activations and a measure of its similarity taken against other weights. This technique should have similar applications to that of distinctiveness, as a visualisation tool it is certainly of great utility, and has a major advantage over distinctiveness in that the *weight correlation* graph has a more regular appearance when the network is learning. This is illustrated in Figures 5, and 6.

Figure 1. Distinctiveness 4-2-4, error decreasing

Figure 1. Weight correlation 4-2-4, error decreasing

5. CONCLUSION

We have presented three network analysis techniques and demonstrate their effectiveness as high level visualisation tools for backpropagation neural networks.

The first of these techniques is *distinctiveness*, which has been reported in detail elsewhere as a network reduction tool [1], and to discover plateaus in the training of hidden units [2]. We demonstrated here the use of this technique in the visualisation of the relative behaviour of the hidden units in a backpropagation neural network.

The second technique of analysis using the causal index between inputs and outputs has been used previously to extract explanations [4] for neural network conclusions and the rules implemented by the trained network [5] (see also [12-15]). Here we demonstrated the significant gains possible in the use of 3D visualisation methods to display the causal index data. In this fashion we have shown how the complex analysis can be largely replaced by inspection of the causal index surface, and the convergence diagram. This has significant relevance to neural network application development in areas such as financial forecasting, where data mining is a major concern.

The third technique is the *weight correlation* graphs, which shows the relationships between units based on their differences in input functionality as determined by the weights and activations entering them. This is clearly different from the distinctiveness. The weight correlation graphs are clearer for distinguishing the behaviour of the hidden units over time, providing a smoother model of changes in functionality, for example they are less sensitive to the initial changes in weights and activation values.

REFERENCES

1 .
Gedeon, TD and Harris, D "Network Reduction Techniques," *Proc. Int. Conf. on Neural Networks Methodologies and Applications*, vol. 1, pp. 119-126, San Diego, 1991.

2 .
Gedeon, TD and Harris, D ? Hidden Units In A Plateau,? *1st Int. Conf. on Intelligent Systems*, pp. 391-395, Singapore, 1992.

3 .
Hora, N, Enbutsu, I and Baba,K "Fuzzy rule extraction from a multilayer neural net,? *Proc. IEEE*, vol. 2, pp. 461-465, 1991.

4 .
Yoda, M, Baba, K and Enbutu, I "Explicit representation of knowledge aquired from plant historical data using neural networks,? *International Joint Conference on Neural Networks*, San Diego, vol. 3, pp. 155-160, 1991.

5 .
Turner, H and Gedeon, TD ? Extracting Meaning from Neural Networks,? *Proceedings 13th Int. Conf. on AI*, vol. 1, pp. 243-252, Avignon, 1993.

6 .
Gedeon, TD and Turner, H "Explaining Student Grades Predicted by a Neural Network,? *Proceedings International Joint Conference on Neural Networks*, Nagoya, 1993.

7. UNCTAD Statistical Pocket Book, United Nations, New York, 1984.

8. Gedeon, TD and Good, RP "Interactive Modelling of a Neural Net Model of GDP,? *Proc. Int. Congress on Modelling and Simulation*, Perth, 1993.

9. Gedeon, TD and Bowden, TG "Heuristic pattern reduction,? *International Joint Conference on Neural Networks*, Beijing, pp. 449-453, November 1992.

10. Gedeon, TD and Bowden, TG ? Heuristic Pattern Reduction II,? *Proc. International Conference on Computer Science*, Invited Position Paper, Beijing, 1993.

11. Slade, P and Gedeon, TD ? Bimodal Distribution Removal,? *Proc. IWANN Int. Conf. on Neural Networks*, Barcelona, 1993.also in Mira, J, Cabestany, J and Prieto, A, New Trends in Neural Computation, pp. 249-254, Springer Verlag, Lecture Notes in Computer Science, vol. 686, 1993.

12. Klimasauskas, CC "Neural networks tell why,? *Dr. Dobbs Journal*, April 1991.

13. Sestito, S and Dillon, T ? The use of sub-symbolic methods for the automation of knowledge acquisition for expert systems,? *Proc. 11th International Conference on Artificial Intelligence*, Avignon, 1991.

14. Sestito, S and Dillon, T ? Automated knowledge acquisition of rules with continuously valued attributes,? *Proc. 12th International Conference on Artificial Intelligence*, Avignon, 1992.

15. Towell, GG and Shavlik, JW "The extraction of refined rules from knowledge-based neural networks,? *Machine Learning*, August 1991.