Predicting Periodic Phenomena with Self-Organising Maps

Dekker, A.H.

P.O. Box 3925, Manuka ACT 2603, Australia
Email: dekker@acm.org

Keywords: *Self-Organising neural network, Kohonen map, periodic data, prediction*

EXTENDED ABSTRACT

In this paper, we use Self-Organising Maps (Kohonen 1989) together with a helical encoding of time. The maps can therefore learn the periodic variation in input data, which may be annual, weekly, or daily. We have implemented this technique in software, and have tested it by demonstrating how it can be used to predict missing data values.

Self-Organising Maps are a form of neural network allowing unsupervised learning. In this work, we use a 2-dimensional network. Given a set of data, the neural network learns the distribution of data points, and provides a mapping from the set of data points to the neural network. The mapping has the property that similar data points map to nearby neurons in the network.

We are particularly concerned with learning the distribution of periodic data, where the period may be a day, a week, or a year. Each point in time therefore has an associated phase angle corresponding to the position within the daily, weekly, or annual cycle. We allow for these cycles by using the helical encoding of time illustrated in Figure 1. If the time t corresponds to the data point $(d_{i1}, \ldots, d_{i\Delta})$, and a phase angle of θ, then we use the vector $(d_{i1}, \ldots, d_{i\Delta}, t, \sin\theta, \cos\theta)$ for training the Self-Organising Map.

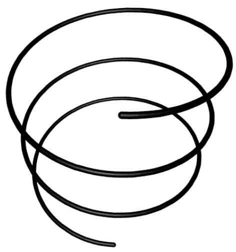

Figure 1. Helical encoding of time for three complete cycles.

By self-organising to map this extended data set, the network not only learns the distribution of data points, but also temporal trends in the data. Figure 2 shows the result of learning a sample data set.

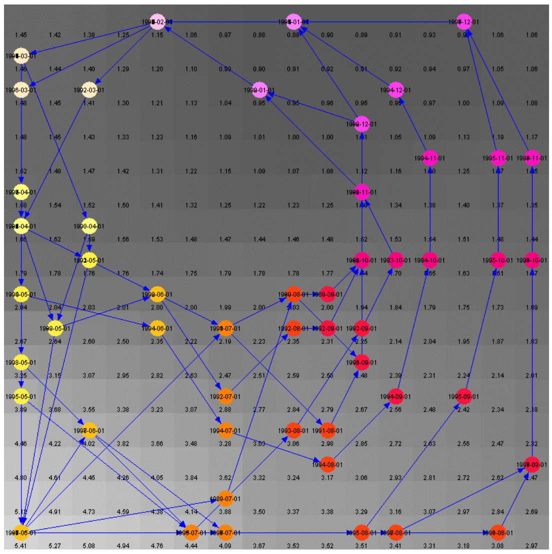

Figure 2. Self-Organising Map after learning sample data for 132 months. Light squares indicate high values, and dark squares indicate low values. Circles indicate neurons corresponding to data points, colour-coded by month. The time sequence is anticlockwise, with January at the top centre.

We can predict missing data values by interpolating within the network. This technique is stochastic, since the predicted value will depend on random aspects of the self-organisation process. However, averaging a number of predictions reduces this random element. In reporting the performance of our prediction technique, we use an adjusted error which makes allowance for the standard deviation of the stochastic prediction.

We study three data sets, one reporting the incidence of drowning in young children in Arizona between 1989–99, one measuring rainfall at a site in Thailand between 1999–2001, and one measuring the temperature in Canberra for five days in 2007. Averaged over three experiments, the adjusted error for our prediction technique was 18%, which was better than the 22% resulting from simple interpolation. This shows that, with a helical encoding of time, the generalisation capabilities of the Self-Organising Map can indeed be used to predict periodic phenomena.

1. SELF-ORGANISING MAPS

The Self-Organising Maps developed by Kohonen (1989) provide a popular form of unsupervised learning (Ritter *et al.* 1992, Hecht-Nielsen 1990). They have been applied to areas as diverse as speech recognition (Beale and Jackson 1990), pattern recognition (Kohonen 1990), ballistic movements of robot arms (Ritter and Schulten 1989), aerodynamic flow modelling (Hecht-Nielsen 1988), and image processing (Dekker 1994).

A Self-Organising Map is a δ-dimensional neural network which learns the patterns within a set of Δ-dimensional data (where $\delta \le \Delta$). This learning is robust in the face of missing data values and even the occasional erroneous data value. The choice of a 2-dimensional network, as in Figure 2, assists in visualisation of the results.

Each neuron N_i in the network holds a Δ-dimensional weight vector w_i. The learning process involves repeatedly finding the best-matching neuron N_i to a data item d_j and moving the weights of selected neurons towards d_j. Specifically, neurons in a neighbourhood of radius r around N_i are moved. The new weights w'_k of neurons in the neighbourhood are given by:

$$w'_k = \alpha\left(1 - \frac{D^2}{r^2}\right)d_j + \left(1 - \alpha\left(1 - \frac{D^2}{r^2}\right)\right)w_k\ , \qquad (1)$$

where D is the distance in the neural network between neuron N_k and neuron N_i, and α is a variable which decreases during learning. Figure 3 illustrates the process. This learning mechanism results in network where similar data items map to nearby neurons.

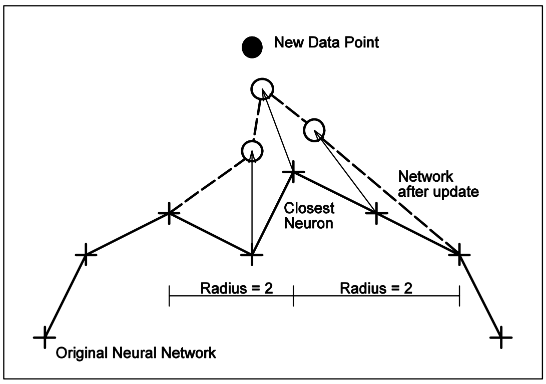

Figure 3. Self-Organising neural network update method.

For the work reported in this paper, α decreases exponentially from 1 down to 0.01 over 100 training cycles, while the neighbourhood radius r also decreases from the size of the network down to 6. This fairly high final radius ensures that the neural network learns a smoothed version of the data, which can be used to generalise the patterns found. This generalisation in turn helps to predict missing data values. The grey squares in Figure 2 indicate the weight values of the neurons in the network, and the smooth transitions in shades of grey illustrate the smoothing effect of a large final radius (the circles in Figure 2 indicate data items d_j, with each data item placed on top of its best-matching neuron N_i).

In order to obtain the best performance of the Self-Organising Map, we use the biasing mechanism developed by Desieno (Hecht-Nielsen 1990, Dekker 1994). For learning purposes, the calculation of the best-matching neuron N_i to a data item d_j is biased by $b_i = 4 - 1000 f_i$, where f_i is an estimate of the frequency at which the neuron N_i has been chosen previously. For the 16×16 network which we use, f_i is normally about 1/256, so that the bias b_i is normally close to zero, unless the neuron N_i has been chosen unusually often.

2. HELICAL ENCODING OF TIME

We are particularly concerned with data from a time series containing a periodic element, where the period may be a day, a week, or a year. The time can be expressed as a position t between the start and the end (for example, as the number of seconds since the start of the time series) and a phase angle θ. The phase angle represents the position in the daily, weekly, or annual cycle. We represent each time by a triple (t, sin θ, cos θ), although in practice we scale all three coordinates to be in the range 0…1.

The triples (t, sin θ, cos θ) correspond to points along a helix, such as that shown in Figure 1. The progression of time combines circular motion (of varying phase angle θ) with linear progress of the t component.

To the triples (t, sin θ, cos θ) we add the numerical data items. If the time (t, sin θ, cos θ) corresponds to a Δ-dimensional set of data values (d_{i1}, …, $d_{i\Delta}$), then we get our Self-Organising Map to learn the (Δ+3)-dimensional set of data values:

$$(d_{i1}, \ldots, d_{i\Delta}, t, \sin\theta, \cos\theta)$$

By self-organising to learn the patterns in this data set, the Self-Organising Map not only learns the distribution of data points, but also the temporal trends in the data.

We perform the learning using a 2-dimensional Self-Organising Map. Figure 2 shows the result of learning the data described in Section 4. Circles show the mapping of data items to the network, and arrows connect consecutive data items.

Since the learning process results in network where similar data items map to nearby neurons, consecutive data items tend to map to nearby neurons, as do data items with the same phase angle. Consequently, the neural network shown in Figure 2 illustrates a clearly visible annual cycle. Variability in the data changes during the year, being high for May to September (at the bottom of Figure 2), and low for January to March (at the top of Figure 2).

3. PREDICTION

We can use our Self-Organising Map to predict missing data values. Figure 2 shows that the sequence of data points forms a consistent trajectory around the network, and we can interpolate using this trajectory. To predict missing data values, we find the neurons in the network corresponding to:

- the data point just **before** the missing one;
- the data point just **after** the missing one;
- the data point with the same phase angle (e.g. the same month) as the missing one, but just **before** it; and
- the data point with the same phase angle (e.g. the same month) as the missing one, but just **after** it.

We then average the positions of these four neurons. Our predicted value is taken from the weights inside the neuron at the average position. We will see in the remainder of the paper that this performs better than simply averaging the four data values in the above list.

This prediction process is stochastic: the actual value produced varies randomly, depending on details of the Self-Organising Map learning process. Consequently, the prediction process should be performed several times, and an average of multiple predictions used.

For our experiments, we report an adjusted error, which takes the stochastic element into account. If the true value is x, and the prediction process produces a distribution of values with mean y and standard deviation σ, then we report the adjusted error e_{adj} as:

$$e_{\text{adj}} = \frac{|x-y|+\sigma}{x}, \tag{2}$$

Single predictions will be within this error at least 68% of the time, and the average of two predictions will be within this error more than 80% of the time. As we increase the number of predictions being averaged, the chance of the average being within e_{adj} of the true value approaches 100%.

4. FIRST EXPERIMENT

For our first test of the prediction process, we use data on the number of children aged 0–4 killed by drowning in Arizona during 1989–1999 (Arizona Department of Health Services 2005). Prediction of this kind of data can be used to target prevention campaigns and therefore reduce the frequency of these tragic events. Figure 4 illustrates the data set. The frequency of drowning is generally highest during the summer, as would be expected. In 1989, it was high also during the spring, and in 1994 the peak was delayed until autumn. However, most of the variation from year to year seems to be essentially random. The Self-Organising Map shown in Figure 2 results from learning this data set.

To test the prediction process, the data points for December 1991 and June 1995 were not included in the data set used for training, but were used to test prediction. The actual value for December 1991 was 1, and the prediction process returned a mean of 0.99 and a standard deviation of 0.07, giving an adjusted error of 8%. For June 1995, the actual value was 5, and the prediction process returned a mean of 3.87 and a standard deviation of 0.22, giving an adjusted error of 27%. Averaging over both predictions gives an overall adjusted error of 18%.

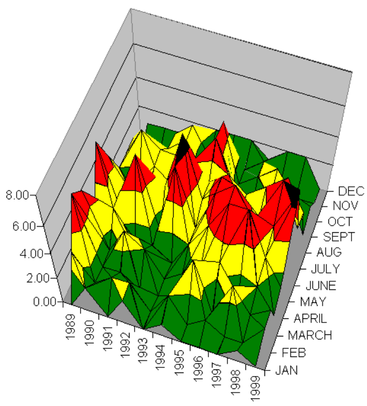

Figure 4. Number of drownings by month and year among children 0–4 years old in Arizona. Data from Arizona Dept of Health Services (2005).

For comparison, averaging the four neighbouring values (previous month, next month, same month in previous year, same month in next year) gives an average error of 33%. Averaging only the same month for the next and previous year gives an average error of 10%, but as we will see, the latter approach is not always effective.

5. SECOND EXPERIMENT

Our first experiment raises the question as to whether a richer set of data values improves prediction. To investigate this, our second data set consists of rainfall data for 1999 to 2001 at the location in northern Thailand shown in Figure 5 (Dairaku *et al*. 2000, Kuraji *et al*. 2001, Kuraji *et al*. 2003). We combined this with values of the Southern Oscillation Index (SOI) as shown in Table 1. A number of months are missing from this data set, and some low values may be erroneous. The network learning must be robust in the face of these problems.

The inclusion of Southern Oscillation Index (SOI) data provides additional explanatory power: high rainfall occurs in the rainy season, particularly when the value of the SOI is low.

The data for June 2000 was excluded from the training data set, and the Self-Organising Map was used to predict the rainfall for that month. The actual value was 181 mm, and the neural network returned an average of 167.2 with a standard deviation of 8.7, i.e. adjusted error of 12%. If the SOI data was not included in the training data set, this adjusted error rose to 15%, indicating that supplementary data in the training data set does assist the Self-Organising Map in learning the patterns in the data.

For comparison, simply averaging adjacent data values gave an error of 17% (if four values were used), or 25% (if only two values were used).

Figure 5. Location of Wat Chan rain gauge, Mae Chaem watershed, upper Chao Phraya river basin, northern Thailand. Photo from Kuraji *et al*. (2003).

Table 1. Wat Chan rainfall data (Kuraji *et al*. 2003) compared to Southern Oscillation Index (SOI) data. The June-00 data was not part of the training dataset, but was used to test the neural network prediction. Note that data for some months is missing, and some low values may be erroneous.

Month and Year	Rainfall (mm)	SOI
January-99	7	15.36
February-99	10.5	8.53
March-99	25.5	9.61
April-99	134	18.72
May-99	215.5	0.66
June-99	200.5	0.39
July-99	125.5	4.93
August-99	198.5	2.57
September-99	214	–0.11
October-99	70.5	9.06
December-99	18.5	13.2
January-00	0	4.54
February-00	27	13.2
March-00	19.5	9.39
April-00	47.5	17.25
May-00	202.5	3.97
June-00	***181***	***–5.57***
July-00	128.5	–4.16
August-00	168.5	5.1
September-00	105.5	10.09
November-00	1.5	22.23
December-00	1.5	8.27
January-01	1.5	8.51
February-01	0	12.18
March-01	78	6.33
April-01	0	1.16
May-01	267.5	–9.86
June-01	69.5	2.47
July-01	105.5	–3.7
August-01	259.5	–8.6
September-01	165	1.49
October-01	218.5	-2.18
November-01	26	8.11

6. THIRD EXPERIMENT

The final experiment used hourly temperature data from Canberra for the period 1:00 pm, 14 May 2007 to 1:00 pm, 18 May 2007 (Australian Bureau of Meteorology 2007). Barometric pressure was used as the additional variable to provide additional explanatory power.

The data for 5:00 am, 17 May and 5:00 pm, 16 May was excluded from the training data set, and the Self-Organising Map was used to predict the temperature for those times. The actual temperatures were 7.2 ºC for the chosen morning and 19.2 ºC for the chosen afternoon.

The neural network returned an average of 8.8 ºC (standard deviation 0.6 ºC) for the morning and 17.6 ºC (standard deviation 1.0 ºC) for the afternoon, i.e. adjusted errors of 31% and 14% respectively. For comparison, simply averaging adjacent data values gave errors of 23% and 7% (if four values were used), or 51% and 11% (if only two values were used).

Figure 6 shows the Self-Organising Map after learning, which provides a good visualisation of the changing temperatures over the given period. The deviation from the main cycle corresponds to the onset of rain just after midnight on the morning of 18 May.

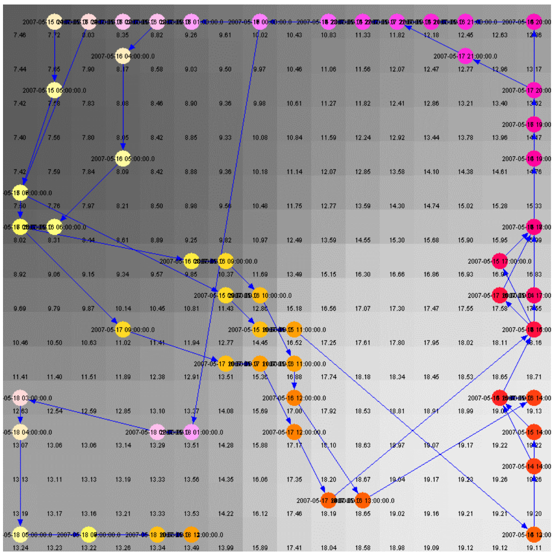

Figure 6. Self-Organising Map after learning Canberra temperature data. Light squares indicate high temperatures, and dark squares indicate low temperatures. Circles indicate neurons corresponding to data points, colour-coded by hour. The time sequence is anticlockwise, with midnight at the top centre.

Table 2. Summary of adjusted errors for prediction with the Self-Organising Map, for the three experiments reported here, compared with errors for simple interpolation.

Data Set	Self-Org. Map prediction	Simple interp. (four values)	Simple interp. (two values)
First experiment	18%	33%	10%
Second experiment	12%	17%	25%
Third experiment	23%	15%	31%
Overall Average	18%	22%	22%

7. DISCUSSION

Table 2 shows the average adjusted errors for the predictions in the three experiments reported here, compared with the errors for simple interpolation. On average, prediction with the Self-Organising Map performs better than simple interpolation, even though it is a stochastic process.

For the first experiment, with data from child drownings in Arizona, one of the simple interpolation techniques did indeed provide a better prediction than the Self-Organising Map, but the other performed much worse. For the third experiment, with temperature data from Canberra, the relative performance of the two simple interpolation techniques was reversed. One of the two interpolation techniques is based on adjacent times and the same time in adjacent years or days, while the other is based on the adjacent years or days only. Neither simple interpolation technique is consistently as good as the prediction provided by the Self-Organising Map.

The ability of the Self-Organising Map to predict missing data values results from its capacity for generalisation, together with the helical encoding of time which we have used, and the way in which the Self-Organising Map preserves topological relationships.

In addition, the Self-Organising Map with our helical encoding of time also provides an excellent way of visualising the periodic variation in the variables being studied, as in Figure 6.

In future work we intend to fine-tune the neural network to improve its predictive power.

8. REFERENCES

Arizona Department of Health Services (2005), Injury Mortality Among Children and Adolescents, Arizona, 1989-1999 (Table 23). Available electronically at www.azdhs.gov/plan/report/im/im/im99/im99.htm

Australian Bureau of Meteorology (2007), Latest Weather Observations for Canberra, accessed 17 and 18 May 2007, www.bom.gov.au/products/IDN60903/IDN60903.94926.shtml

Beale, R. and T. Jackson (1990), *Neural Computing: An Introduction*. Adam Hilger, Bristol.

Dairaku, K., K. Kuraji, M. Suzuki, N. Tangtham, W. Jirasuktaveekul, and K. Punyatrong (2000), The effect of rainfall duration and intensity on orographic rainfall enhancement in a mountainous area: A case study in the Mae Chaem watershed, Thailand. *Journal of the Japan Society of Hydrology & Water Resources*, **13** (1), pp. 57–68.

Dekker, A.H. (1994), Kohonen Neural Networks for Optimal Colour Quantization. *Network: Computation in Neural Systems*, **5**, pp. 351–367, Institute of Physics Publishing.

Hecht-Nielsen, R. (1988), Neurocomputing: picking the human brain, *IEEE Spectrum*, **25** (3), March, pp. 36–41.

Hecht-Nielsen, R. (1990), *Neurocomputing*, Addison-Wesley.

Kohonen, T. (1989), *Self-Organization and Associative Memory*, 3rd edition, Springer-Verlag, Berlin.

Kohonen, T. (1990), Pattern recognition by the self-organizing map, in Caianiello, E.R., ed., *Parallel Architectures and Neural Networks: Third Italian Workshop* (Vietri sul Mare, Salerno, 15–18 May), World Scientific, Singapore, pp. 13–18.

Kuraji, K., K. Punyatrong, and M. Suzuki (2001), Altitudial increase in rainfall in Mae Chaem watershed, Thailand, *Journal of the Meteorological Society of Japan*, **1B**, pp. 353–363.

Kuraji, K., K. Punyatrong, and I. Sirisaiyard (2003), GAME-T Rainfall Measurement in the Mae Chaem Watershed Data Archive, accessed October 2003, www.uf.a.u-tokyo.ac.jp/~kuraji/MaeChaem/

Ritter, H. and K. Schulten (1989), Extending Kohonen's self-organizing mapping algorithm to learn ballistic movements, in Eckmiller, R. and C. v.d. Malsburg, eds., *Neural Computers*, Springer-Verlag, Berlin, pp. 393–406.

Ritter, H., T. Martinetz, and K. Schulten (1992), *Neural Computation and Self-Organizing Maps: An Introduction*, Addison-Wesley.