Self-Organizing Maps as a Tool to Analyze Movement Variability

P. Lamb1, R. Bartlett1, A. Robins2, G. Kennedy1

School of Physical Education1, Department of Computer Science2, University of Otago, Dunedin, New Zealand

Abstract

Self-Organizing Maps possess unique properties that remove redundancies in a high-dimensional input space and map that input space to a low dimensional output space, thereby showing non-linear relationships in the input data. This ability makes Self-Organizing Maps attractive for measuring inter-limb coordination patterns. The current study takes previously published data (Bartlett, Bussey, & Flyger, 2006) used to compare the reliability of different operators digitizing gait patterns with and without anatomical markers. The trained network was simulated and analyzed qualitatively using the trajectory of activated nodes for each input vector, similar to Barton, Lees, Lisboa, & Attfield (2006). Qualitative differences in map trajectories were seen between Marker and No-Marker conditions and supported the results of the original publication that, when using 2-D videography, manually digitized markers allowed accurate estimation of movement variability, whereas the No-Marker condition did not. This finding is shown by the No-Marker trajectory travelling further from the center of the network cluster than the Marker condition. Additionally, the map trajectories revealed that changes in coordination at different phases of the movement can be identified. For most trials the Marker trajectory travels closer to the centre of the map than the No-Marker condition which, as is explained by the neighbourhood function, indicates less variability. The consistency between conventional biomechanical analysis techniques and the qualitative assessment of Self-Organizing Map outputs adds to the validity of the Self-Organizing Map as an accurate measurement tool for coordination. The ability to identify changes in coordination using the map trajectories illustrates the potential for the Self-Organizing Map to show novel information about the coordination pattern.

KEYWORDS, ARTIFICIAL NEURAL NETWORKS, COORDINATION, RELIABILITY, SELF-ORGANIZING MAPS, VARIABILITY

Introduction

From dynamical systems theory, human behaviour is thought to be non-linear and self-organizing (Davids, Glazier, Araujo, & Bartlett, 2003; Kelso, 1995). Using non-linear analyses may help identify characteristics in the movement pattern that would otherwise go unnoticed. Artificial neural networks are non-linear and thus conform to the dynamical systems theoretical framework. By analyzing human movement data with artificial neural networks, the data is not forced through a linear path, as is the case with most statistical
operations. This analysis technique is intended to reduce the dimensionality of the data so it can be visualized while preserving the original topology of the data. The Kohonen Self-Organizing Map is an artificial neural network designed for pattern classification through dimensionality reduction. The Self-Organizing Map can be thought of as a layer of nodes with associated weight vectors, fed forward by a layer of inputs. Each input node is connected to all output nodes (see Figure 1). The weight vectors of the nodes in the output layer are the same dimensionality as the input. Through an iterative process the network strives to map a low dimensional representation of the input while preserving the topology of the data.

![Figure 1. The connections between input and output layers.](image)

The Self-Organizing Map uses an unsupervised learning strategy to update the weight vectors of the output nodes in an effort to match the input more closely. Weights are updated if the weight vector has the smallest Euclidean distance to the input. Euclidean distance is defined by

$$d_i = \sqrt{\sum_{i=1}^{n} (V_i - W_{ni})^2}$$  \hspace{1cm} (1)

where $V_i$ is the $i^{th}$ input vector and $W_{ni}$ is the $i^{th}$ weight vector and the $n^{th}$ node on the topological map. The output node, $W_{ni}$, whose weight vector has the smallest Euclidean distance of all nodes is declared the best matching unit and has its weights updated by the learning constant and neighbourhood function, given by

$$W_{t+1} = W_i + \alpha_i \psi_t (V_i - W_{ni})$$  \hspace{1cm} (2)

where $\alpha_t$ is the learning rate at iteration $t$ and lies between 0 and 1 and $\psi_t$ is the neighbourhood function at iteration $t$. Nodes in the output layer are updated based on their proximity to the best matching unit. Nodes located more closely to the best matching unit are updated more, relative to nodes that are further from the best matching unit - but still in its neighbourhood. Nodes in the best matching unit’s neighbourhood are thus influenced by that input using the neighbourhood function.

In this study, default values from the SOM toolbox (for MATLAB) for the learning parameters were used (Vesanto, Himberg, Alhoniemi, & Parkankangas, 2000). An important note about the learning parameters is that they start with fairly large values and decay towards zero as $t$ increases. A high initial neighbourhood radius ensures that the map spreads globally. If the radius is set too small in the initial stages, groups of data may get clustered together without having the chance for other similar inputs to enter the neighbourhood, this would result in an assortment of separated clusters not representative of the input distribution.
In a network not larger than a few hundred nodes, specifying the learning parameters (learning rate and neighbourhood function) is not crucial. It is generally accepted that initial values for the neighbourhood size be approximately one fifth to half the size of the network and the learning rate should begin roughly between 0.9 and 0.5. Variations within these ranges will not drastically affect a small network (Kohonen, 2001).

In the engineering domain, Self-Organizing Maps are often used for pattern classification applications using discrete variables (i.e. images (Hussain & Eakins, 2007), geographical regions (Ong & Abidi, 1999) and structural biological features (Hyvönen et al., 2001)). There are of course patterns that the human brain can recognize that artificial neural networks have not been able to. Yet there are also high dimensional patterns that have been recognized by artificial neural networks and not the human brain, illustrating the potential use of this method. With movement data, similar trials would be grouped into topologically similar regions on the output map. However, some researchers of human movement (i.e. clinicians, biomechanists) are more interested in classifying phases, or components, of a movement using continuous data. To accomplish this, individual time frames taken from each movement trial are used as input instead of the whole data set. This is to allow similar phases of the movement to be clustered together. An analysis of the clusters allows the researcher a glimpse of the coordination states during the movement.

A useful property of the Self-Organizing Map is the option to simulate the network with a portion of the training data. By doing so, the best matching units for all input vectors of one scoring of a trial can be traced on the trained network. Simulating the network allows the researcher to compare different trials or groups of trials. Comparing the best matching units of different trials at the same time frame can offer unique information about the difference in coordination states between the trials used for simulation. For this study a trajectory of best matching units throughout the time series was used to compare between and within conditions of the data set.

This study uses data from Bartlett, Bussey, & Flyger (2006) to a) validate Self-Organizing Maps as a tool for biomechanical analysis, b) show that Self-Organizing Maps provide the capability for something novel to be learnt about the data that cannot be realized using analyses of variance, and c) critique a new visualization technique for the Self-Organizing Map.

Methods

**The Dataset**

The data used for this study were adopted from Bartlett et al. (2006). One runner was filmed running on a motorized treadmill at 16 km/h for ten separate trials, two-dimensional videography captured movement in the sagittal plane at 50 Hz. In five of the ten trials the runner did not wear anatomical markers while in the other five trials he wore reflective markers on the right shoulder, elbow, wrist, greater trochanter, knee, lateral malleolus and the fifth metatarsal head. All trials were filtered using a fourth order Butterworth filter with a cut-off frequency set to 6 Hz based on residual analysis.

From each trial, three consecutive strides were chosen to be digitized. Four trained operators digitized all trials for one condition on alternate days to establish inter-operator reliability. The No-Marker condition trials were digitized before the Marker condition trials to prevent the operators from learning the marker positions. To establish intra-operator reliability, this process was repeated for five days. The data were time normalized to 100 frames so phases of the gait cycle could be described as a percentage of the gait cycle. The filtered coordinates
of the above mentioned anatomical landmarks were used to create joint angles which were used as input for the network.

**Training the Network**
a) Creating the Input Vector
Three time samples at constant intervals of 5% were appended to create data triplets (see Figure 2). These data triplets were used to represent the temporal nature of the pattern of the data to the network (Barton et al., 2006). Time samples are organized as rows in the SOM toolbox, new rows were created as,

\[
\begin{align*}
( & \text{Time = 1%}, [ \text{Time = 6%}, [ \text{Time = 11%} ] ) \\
( & \text{Time = 2%}, [ \text{Time = 7%}, [ \text{Time = 12%} ] ) \\
( & \text{Time = 3%}, [ \text{Time = 8%}, [ \text{Time = 13%} ] ) \\
& \ddots \\
( & \text{Time = 90%}, [ \text{Time = 95%}, [ \text{Time = 100%} ] ).
\end{align*}
\]

Figure 2. Example of time shifted at constant interval to represent the temporal nature of the pattern of the data.

The necessity for there to be ten samples in the trial following the first sample in the triplet reduced the number of input vectors per trial from 100 to 90. With four operators digitizing five trials in two conditions, on five different days the network was trained with 200 trials (20 scorings of each trial). Each trial consisted of 15 input variables (5 x 3 data triplets) and 90 time samples resulting in 270,000 data points in the dataset.

b) Initialization
Before initializing the map, the eigenvalues and corresponding eigenvectors of the training data are calculated to determine the map size. Once the number of map units is known the dimensions of the map are calculated. The ratio of the two largest eigenvalues, \( \lambda_1 \) and \( \lambda_2 \) respectively, is used to determine the dimensions of the map.

\[
\frac{n}{m} = \sqrt[2]{\frac{\lambda_1}{\lambda_2}}
\]  

(3)
For the dataset used in this study the map size was a [29, 23] hexagonal lattice. Linear initialization, which organizes the weights of the nodes linearly along the two (number of dimensions of the map) largest eigenvectors of the training data, was used to speed up training (Vesanto et al., 2000).

c) Training

In batch training, Euclidean distances to all nodes for each input vector are calculated. Instead of calculating the distances as in equation (1), the process is accelerated by expressing the learning algorithm as a matrix operation (Vesanto et al., 2000). As a result of redundancies in the data, an efficient initialization process and the batch training algorithm, the network was able to model this data set extremely quickly.

d) Visualization

Much research has been focused on visualization techniques for the low dimensional map outputs (Kohonen, 2001; Pampalk, Rauber, & Merkl, 2002; Pölzlbaier, Rauber, & Dittenbach, 2005). To show the coordination state at a given time frame, the map can be arranged into a fixed grid (grid space) with the best matching unit highlighted. Similarly, a trajectory can be added on top of the grid to show several consecutive coordination states throughout the trial (Barton, 1999; Barton et al., 2006; Bauer & Schöllhorn, 1997; Lees & Barton, 2005; Schöllhorn & Bauer, 1998). The grid trajectory is an advantageous visualization technique because qualitative changes in coordination are easily seen. Different best matching unit trajectories between trials allow for generalizations to be made about the movement, not only that there is a qualitative change in coordination, but at which phase of the movement the qualitative change occurs.

This study uses a unified distance matrix (u-matrix) for visualization of the trained network and the best matching unit trajectory. Each ‘unit’ in the u-matrix will be referred to as a cell for clarity, so that it is not confused with nodes in the output layer. Recall that each output node in the SOM has an associated weight vector. The u-matrix represents each output node (grid space) as an uncoloured (white) cell. Between each pair of nodes / white cells is a coloured cell, which represents the similarity of the weight vectors of that pair of nodes. Blue represents a pair of weight vectors which are close (in the input / weight space), red represents vectors which are far apart. Hence for each pair of adjacent nodes in the output layer we also have a coloured cell representing how similar the weight vectors of those two nodes are. This makes the u-matrix a hybrid representation of grid space and weight space (see Figure 3).
Figure 3: A 2X2 u-matrix. The colourless cells in the u-matrix represent nodes in the output layer. The distance between adjacent cells is represented by coloured distance cells. The blue distance cell represents similar weights of the neighbouring cells.

For a ‘sheet’ map topology of size \([m \times n]\), the corresponding u-matrix would be of size \([2m-1 \times 2n-1]\). Nodes which are adjacent in the output layer (grid space) are represented by cells which are adjacent in the u-matrix. The u-matrix, however, represents the proximity of the weight vectors of each pair of adjacent nodes with coloured distance cells. The coordinates of the nodes are much like a rectangular grid; however, for the hexagon-shaped cells to fit, every other row in the grid is shifted horizontally by half a unit. Visualizing the best matching unit trajectory in weight space is an important distinction to make because it allows for a more objective assessment of the change in coordination between trials.

Analysis

The trajectory of the best matching unit sequence through the time series of the trial was used to compare various trials simulated on the same trained network (Barton, 1999; Barton et al., 2006; Bauer & Schöllhorn, 1997; Lees & Barton, 2005; Perl, 2004; Schöllhorn & Bauer, 1998). The trajectories used for this study are in both grid space and weight space, implying that the distances between cells actually represent the difference between the weight vectors. Thus, the comparison of the best matching units between trials at the same time frame represents the similarity (or dissimilarity) of the movement pattern during that coordination state.

The Self-Organizing Map excels in clustering similar input vectors. On the u-matrix, clusters (blue distance cells) and borders (more brightly coloured distance cells) can be identified. The blue distance cells represent a small Euclidean distance between neighbouring weight vectors and thus are representative of clusters of similar input vectors. If there is more than one cluster in the output layer those clusters will be separated by more brightly coloured distance cells (see Figure 4), indicative of a larger Euclidean distance between the weight vectors and less similarity in the input distribution. Shifts between best matching units within clusters do not indicate a major change in coordination; however, conclusions about the variability of the dataset can be made. If the path of the best matching unit trajectory is
inconsistent between cycles of the same trial there may be high inter-cycle variability and if the trajectory between trials is inconsistent there may be high inter-trial variability. Additionally, for the data set used in this analysis, differences between trajectories of different scorings of a trial by the same operator may be indicative of intra-operator variability whereas differences between trajectories of all scorings of all trials in each condition could indicate intra-condition variability, which is the focus of this study.

Figure 4. Phases of gait cycle visualized on u-matrix

To quantify the difference in the best matching unit trajectories between all scorings of each trial the Euclidean distance between all scorings for each data triplet was summed to give a measure of similarity within each trial. This analysis technique is used to further validate visual analysis of best matching unit trajectories.

An equally important consideration when analyzing the map trajectories is how well the best matching units on the output map represent the input. Quantization error is a measure of the fit of the map to the input distribution in the form of an average distance between each input vector and its best matching unit. Quantization error between several different elements can be found. In this paper an average quantization error between the input and weight vectors of individual trials was used to characterize the variability between Marker and No-Marker conditions.

Results and Discussion

The purpose of this study was to a) replicate the findings of Bartlett et al. (2006), b) show something novel as a result of using Self-Organizing Maps and, c) critique a new visualization technique for Self-Organizing Maps.

Bartlett et al. (2006) concluded that movement variability accounted for more of the total variability in trials that were digitized with anatomical markers compared to trials digitized without anatomical markers. In the No-Marker condition, inter-operator variability for the ankle was five times higher than the movement variability while the hip was twice as high as
the movement variability. Movement variability contributed to a particularly low proportion of the total variability in the No-Marker condition for all operators digitizing the ankle and the hip. None of the operators achieved movement variability accounting for more than 25% of the total variability for ankle angle and for two out of four operators, movement variability did not account for more than 50% of total variability for hip angle. The physical errors contributing to these differences are not the same for each condition. In the Marker condition, the operator had to ascertain the centroid of a circular physical marker centred over the joint axis of rotation. This error could have been of a similar magnitude to the resolution of the image. In the No-Marker condition, the operator had to estimate the position of the joint axis of rotation without any marker to assist; in this condition, the error in so doing was, not surprisingly, much larger than the screen resolution. From this, the authors suggested that multiple operators not be used as they would add an extra source of variability; and that anatomical markers should be used when capturing 2-D videographical data for the purpose of assessing movement variability.

Upon visualizing the trained network with the u-matrix, a brightly coloured, centrally located, branching border can be seen that separates clusters of cells identified by blue distance cells. By matching the time frame of the best matching unit it was found that certain phases in the gait cycle could be identified in the u-matrix. Distinct phases identified on the u-matrix were the contact phase and swing phase (see Figure 4). The contact phase is shown between heel strike and toe off while the swing phase took up the remainder of the u-matrix.

A qualitative difference between trials of the Marker and No-Marker conditions is apparent by visual analysis. In the Marker condition the best matching unit trajectory travels closer to the brightly coloured border than the trajectory for the No-Marker condition. As a result of linear initialization the weights of the nodes are spread uniformly throughout the input data. Nodes along the edges of the map represent extreme values in the data while nodes located more closely to the centre represent more typical data. Trajectories passing more closely to the centre of the u-matrix may occupy more typical coordination patterns compared to other coordination patterns in the dataset.

Visualizing the trained network on a u-matrix allows the visualization to be in grid/weight space and thus differences between best matching units for various trials can be found using Euclidean distances between the weights. To show similarity or dissimilarity within trials of the same marker condition the Euclidean distance between all scorings of each trial is shown in Table 1. The Euclidean distances between all trials in the Marker condition is less than the Euclidean distances between all trials in the No-Marker condition. Trials one and two in the No-Marker condition show smaller Euclidean distances than the remaining trials in the No-Marker condition – although still larger than for the Marker condition trials. The difference between Euclidean distances for each condition shows that the best matching unit trajectories in the Marker condition, for each trial, are more consistent than the trajectories in the No-Marker condition. The higher Euclidean distances between all scorings of trial three, four and five of the No-Marker condition compared to the other two trials is a unique finding that was not exposed in the original study using analyses of variance. The original analysis made comparisons based on each individual angle while this study compares trials based on a composition of all angles. Grouping the variables by means of the Self-Organizing Map shows that the coordination of the variables in trials three, four and five is more variable than the other two trials in the No-Marker condition. The differences shown in Table 1 indicate that digitizing participants who are wearing anatomical joint markers increases the reliability of the procedure. This analysis is intended to provide more objectivity to the qualitative difference between trajectories.
Table 1. Euclidean distances between all 20 scorings of each trial

<table>
<thead>
<tr>
<th>Marker Condition</th>
<th>Euclidean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1</td>
</tr>
<tr>
<td></td>
<td>22.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No-Marker Condition</th>
<th>Euclidean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1</td>
</tr>
<tr>
<td></td>
<td>46.7</td>
</tr>
</tbody>
</table>

Upon analysis of the quantization errors, the mean error for the Marker condition (0.063±0.007) was significantly less than the mean error for the No-Marker condition (0.108±0.018), $p < .001$ (see Figure 5). The outside edges of the u-matrix represent the extreme values in the weights of the nodes and should, therefore, be activated less often than the more typical nodes located closer to the centre. The largest proportion of total quantization error was contributed by the weights of nodes located along the edges of the u-matrix.

![Figure 5. Mean (± SD) quantization errors for all scorings of each trial.](image)
As shown in Figure 6, the No-Marker condition trajectories travelled on or close to the edges of the u-matrix for three out of four edges on the map. From this the No-Marker condition trials can be thought of as more variable because the quantization error for each input and its best matching unit was higher than in the Marker condition. The lower quantization error in the Marker condition does not mean the trials or the cycles within the trials are less variable; rather, the scorings of these trials is less variable on a trial by trial basis.

Using Self-Organizing Maps to analyze human movement is more complicated than conventional statistical approaches; therefore, for Self-Organizing Maps to be appropriate they must not only be shown to be valid but must also have the potential to highlight something new in the data. The original publication relied on comparisons between individual angles, angular velocities or angular accelerations to establish the reliability of digitizing a participant running on a treadmill in Marker and No-Marker conditions. In the current study the analysis of best matching unit trajectories as well as quantization errors supported the original findings and also may have uncovered something new in the data. The best matching unit trajectories in the Marker condition travelled closer to the centre of the u-matrix along three of the four sides of the map while the No-Marker trajectory travelled closer to the centre along the top edge of the map. The top edge represents the phase in the gait cycle when the knee is extending for heel strike. A speculative reason for this could be that as the heel strikes the running surface the foot moves in the shoe slightly, meaning that the marker on fifth metatarsal may not be accurately representing the motion of that anatomical landmark. This is however only speculation, because of the redundancy in the data, the differences in the data are very small and difficult to give definitive meaning to.

**Conclusion**

This study supports other studies that show Self-Organizing Maps to be effective for classifying movement patterns using time series data (Barton, 1999; Barton et al., 2006; Bauer & Schöllhorn, 1997; Lees & Barton, 2005; Schöllhorn & Bauer, 1998). Output from
the Self-Organizing Maps also supported the findings of the original publication (Bartlett et al., 2006) using analyses of variance. The distance cells in the u-matrix allow for meaningful conclusions to be made about the shifts of coordination state as the Euclidean distance between the cells in the u-matrix are representative of the difference between the weight vectors of the respective nodes. The potential for quantifying the differences between movement patterns based on their map trajectories lies in using a visualization technique, such as the u-matrix, that incorporates information about weight space.

Caution must be used with artificial neural networks. The network output must be analyzed before any inferences about the underlying movement can be made; however, observations made in this study are consistent with expectations based on the nature of Self-Organizing Maps. In this study Self-Organizing Maps were shown to be effective in distinguishing between the two digitizing conditions. Since human movement data is inherently redundant an effective method for reducing the redundancies in the dataset holds great potential for movement analysis. Various unique properties of Self-Organizing Maps make them attractive for studying human movement and also make them conform with the dynamical systems theoretical framework. Further work needs to be done to make more concrete statements about the underlying coordination pattern based on best matching unit trajectories.

References


