Predictive Mapping with Self-Organising Maps: A Synthetic Study

Angela Carter-McAulans, Peter G. Lelièvre, Colin G. Farquharson
Department of Earth Sciences, University of New England

Introduction

Geological mapping in remote areas or in areas with limited roadway exposure can be challenging and expensive. The development of predictive geoscientific maps from remote sensing data can enable an unprecedented precision and rapid flood of spatial data distribution on an area by area where these challenges can be mitigated. Generation of geological maps can be accomplished through the application of computer algorithms designed to cluster and discriminate multivariate datasets. In this project self-organising map algorithms (SOMs) are being investigated as a possible means of multivariate dataset analysis of remote sensing data for geological discrimination. Here we present a synthetic study illustrating the capabilities of SOMs.

Self-Organising Maps Algorithms

SOMs are a class of unsupervised neural network algorithms which use a statistical approach to cluster multivariate datasets. SOMs were developed by Kohonen (1982), and are designed to reduce the dimensionality of data by forming the data in a 2D-computational space (i.e. a neural map). This topology of algorithm has been applied to a number of datasets, in particular medical and geophysical (Brandes et al., 2011), and remote sensing data (Liu, 2008). The implementation of SOMs used in this project, SOMs, was developed specifically for use with remote sensing data from Australian CSIRO.

Synthetic Model and data

The work flow investigated in this poster is summarised in Fig. 1. The first step was to develop a synthetic model (Fig. 2). This model consists of six geological units. Each unit has the form of a simple geometric structure, assigned a general depth from surface, and assigned five properties (p1, p2, p3, p4, and p5) associated respectively with the data types (a, b, c, d, and e) below. The form, depth, and properties of the units are summarised in Table 1. The synthetic objects of geochemical and geological properties are not constant and have intrinsic variability. To replicate this variability the properties for each unit were treated as the mean of a normal distribution and the value for that property at any point was allowed to vary in Gaussian form for a given standard deviation.

Spatial Resolution and Depth Sensitivity

A standard deviation of 0.1 was used for the examples discussed here. The relationship between a property and its associated data (variability and spatial resolution of each of the five datasets created in this project were treated on the basis of a specific type of geophysical data) and its associated physical property, as only one dataset can be said to be analogous to a type of geophysical property. The relationship between data property and depth sensitivity is of fundamental importance: a simple prediction neural network is an artificial, non-physical, system. It is only possible to replicate the original model rather well. Further work needs to be done to investigate the effects of increased intra-unit property variability before this method can be extended to testing with real data.

Conclusion

The results of these tests can be validated by plotting the data colored based on their cluster. In Fig. 7 the input data and results of the final test (where all datasets meet the spatial resolution and depth sensitivity criteria) are presented showing that much of the mis-clustering is occurring on the edges of the units.

Table 1: A summary of the characteristics of each of the units in the model for this project.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Rock Type</th>
<th>Properties</th>
<th>Depth (m)</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cluster 1</td>
<td>p1, p2, p5</td>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Cluster 2</td>
<td>p1, p2, p3</td>
<td>2000</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Cluster 3</td>
<td>p1, p4</td>
<td>3000</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Cluster 4</td>
<td>p1, p5</td>
<td>4000</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Cluster 5</td>
<td>p2, p4</td>
<td>5000</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>Cluster 6</td>
<td>p3, p5</td>
<td>6000</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 1: Work flow for this synthetic study showing the processing of data points from model building through analysis of test results.

Figure 2: Synthetic model constructed for this project.

Figure 3: Distribution of the five properties with added variability, (a) p1, (b) p2, (c) p3, (d) p4, and (e) p5, and their associated basic datasets (a), (b), (c), (d), and (e) respectively.

Figure 4: The Q values after each of the 83 test training runs plotted on the overlay of the neural map being trained.

Figure 5: Topographic errors for each of the 83 test training runs plotted with respect for the numbers of rows and columns in their respective neural maps. This figure also contains a map of nets with real data.

Figure 6: Accuracy of clustering for all data points for each of the 16 tests.

Figure 7: The input data (top five panels), with observation location indicated by black dots, and final clustered results (bottom panel) for the final test where all datasets are consistent with their spatial resolution and depth sensitivity characteristics.

References

Acknowledgements

No work would be possible without Professor Exelrod's generous offer to fund this research and OCR to provide the necessary funds for this project. A particular thanks to Dean Beesley at UNE for being so patient with all the man-hours spent on the OCR work.