

Neural networks and their rules for classification in marine geology

Alfred Ultsch¹, Dieter Korus², Achim Wehrmann³

Abstract Artificial neural networks are more and more used for classification. They are capable to process incomplete and imprecise data and to detect non-linear relations in the data. We describe the application of neural networks for classification in the area of geology. An unsupervised learning neural network system was applied to the classification of biogene sedimentation. The classification by neural networks performed better than statistical cluster analysis. In addition, a rule generation algorithm extracted rules out of the neural network, which could be used by the geological expert.

Keywords Neural network, classification, biogene sedimentation, rule generation

1. Introduction

In real world people have continuously to do with raw and subsymbolic data which is characterized by the property that one single element does not have a meaning (interpretation) of itself alone. The question is, how to transform the subsymbolic data into a symbolic form. Indications are that neural networks provide fault-tolerance and noise resistance. They adapt to unstable and largely unknown environments as well. Neural networks are able to learn structures of an input set without using a priori information.

Unsupervised learning neural networks can adapt to structures inherent in the data. They exhibit the property to produce their structure during learning by the integra-

1. Dep. of Mathematics/Informatics, University of Marburg, D-35032 Marburg, Germany
email: ultsch@informatik.uni-marburg.de

2. email: korus@informatik.uni-marburg.de
www: <http://www.mathematik.uni-marburg.de/~korus>

3. Dep. of Geology, University of Marburg, D-35032 Marburg, Germany

tion (overlay) of many case data. But they have the disadvantage that they cannot be interpreted by looking at the activity or weights of single neurons. Because of this we need tools to detect the structure in large neural networks. We developed special visualization tools for this task.

An further intrinsic property of neural networks is, however, that no high level knowledge can be identified in the trained neural network. The central problem for Integrated Knowledge Acquisition is therefore how to transform whatever a neural network has learned into a symbolic form. Therefore expertise learned by neural networks is not available in a form that is intelegible for human beings as well as for knowledge-based systems. It seems to be difficult to describe or to interpret this kind of information. In knowledge-based systems on the other hand it is easy to describe and to verify the underlying concepts.

In real world people have continuously to do with raw and subsymbolic data. Sub-symbolic data is characterized by the property that one single element does not have a meaning (interpretation) of itself alone. A single element can only be understood in a broader context, and by taking other informations into account. The question is now, how to interpret these raw and subsymbolic data, or more specific, how to transform the transform the subsymbolic data into a symbolic form. Higher abilities in Neural Networks can be observed only in large Neural Networks and as a collective behaviour of many single neurons. A first step in the transition of sub-symbolic into symbolic knowledge is the distributed representation of the raw and subsymbolic data in large Neural Networks with collective behaviour. Neural Networks with unsupervised learning can adapt to structures inherent in the data. Suitable Neural Networks exhibit the property to produce their structure during learning by the integration (overlay) of many case data. But large Neural Networks have the disadvantage that they cannot be interpreted by looking at the activity or weights of single neurons. Because of this the users of the Neural Networks need tools to detect the structure in large Neural Networks. An important way to this could be grafical visualization tools.

Most recent studies of carbonates and the biology of reef-ecosystems are concentrated to tropical environments of lower latitudes. During the last two decades, the STAR-model (Lees et al. 1972, Lees 1975) was used for a general differentiation of tropical vs. non-tropical carbonates. But the occurence of reefs and biogenic buildup's beyond the arctic circle (Freiwald 1993, Henrich et al. 1993) suggests that the controlling factors of carbonate production by marine ecosystems must be more complex as shown in the oversimplified STAR-model.

The above described unsupervised learning neural network system was applied to the classification of biogene sedimentation. The classification by neural networks performed better than statistical cluster analysis. In addition, the rule generation algorithm extracted rules out of the neural network, which could be used by the geological expert.

2. Neural Networks and Rule Generation

Artificial neural networks are capable to approximate non linear relations in data. They deal with knowledge in a subsymbolic form. In addition, incomplete and imprecise data can be processed. Neural networks learn in a massive parallel and self-organising way. Unsupervised learning neural networks like Kohonen's self organizing feature maps (Kohonen 1989) learn the structure of high-dimensional data by mapping it on low dimensional topologies, preserving the distribution and topology of the data. But large neural networks can only be interpreted with analysing tools. Looking at the learned self organizing feature map as it is one is not able to see much structure in the Neural Network, especially when processing a large amount of data with high dimensionality. In addition, automatic detection of the classification is difficult because the self organizing feature map converges to an equal distribution of the neurons on the map. We developed a visualisation method, the so called U-Matrix methods, to detect the structure of large two-dimensional Kohonen maps. It generates a three-dimensional landscape on the map, whereby valleys indicate data which belongs together and walls separate subcategories. Unlike in other classification algorithms the number of expected classes must not be known a priori. Also, subclasses of larger classes can be detected. Single neurons in deep valleys indicate possible outliers.

To prove the output of an intelligent system like neural network systems the expert of the domain wants to get an explanation of the decision. This cannot be done by neural networks because they behave like a black box. Knowledge Based Systems have the advantage that they can give an explanation of a diagnosis. But a main difficulty when dealing with knowledge based systems is the acquisition of the domain knowledge. There are several problems. It is difficult to transform the explicit and implicit knowledge of the expert of the domain, which also partly consists of own experience, in a form which is suitable for a knowledge base. The knowledge can also be inconsistent or incomplete. A second problem is that knowledge based systems are not able to learn from experience or to operate with cases not represented in the knowledge base. By integrating both paradigmas, knowledge based systems and neural networks, the disadvantages of both approaches

can be redressed (Ultsch 1992). In order to solve this problem, we have developed a rule generation algorithm, called sig*, that takes the significance of a symptom (range of a component value) into account [29]. Sig* takes a data set in the space R^n that has been classified by a self-organizing feature map and U-matrices as input and produces descriptions of the classes in the form of decision rules. In particular, the generated rules take the significance of the different structural properties of the classes into account. If some few properties account for most of the cases of a class, the rules are kept very simple. Metaphorically spoken, if a patient comes to a physician with a cut throat it would be unadvisable to run blood tests.

We are developing a hybrid system REGINA which consists of several parts. An unsupervised learning neural network maps the (preprocessed) data space onto a two-dimensional grid of neurons, whereby it preserves the distribution and topology of the input space. But only together with a visualisation module, called U-

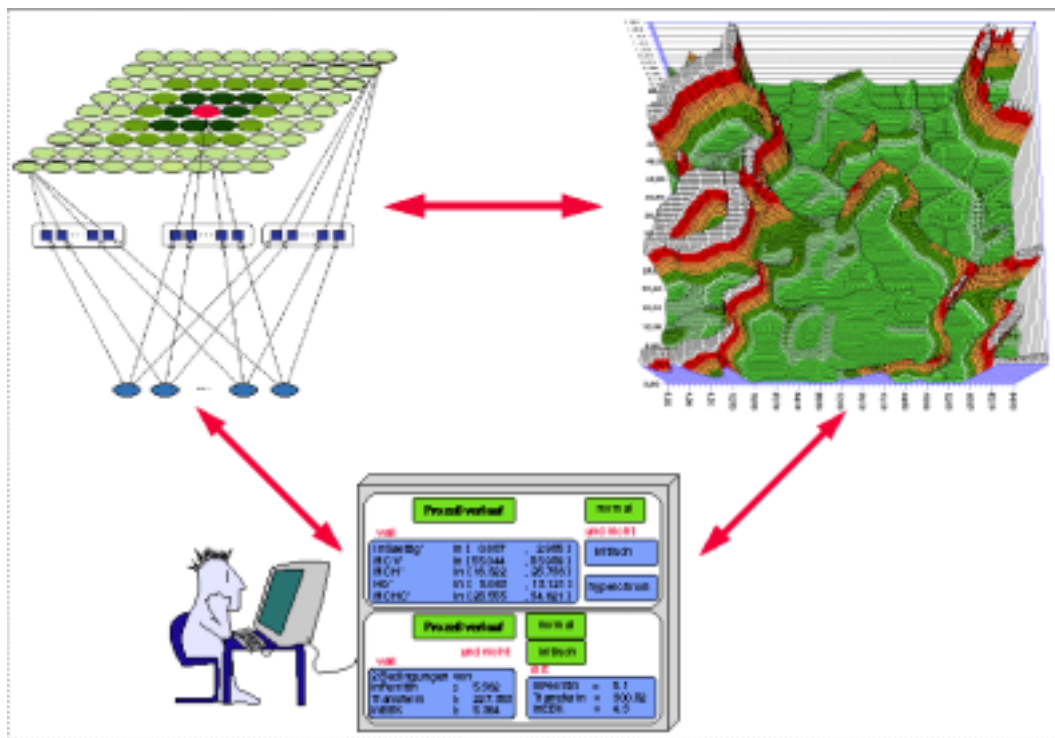


Fig. 1: The integrated system REGINA consisting of self-organizing neural network, a visualization component and an expert system.

Matrix methods, we are able to detect structure in the data and classify it. A three-dimensional coloured landscape will be generated in which walls separate distinct subclasses and subcategories are represented by valleys. The simplest *U-matrix* method is to calculate for each neuron the mean of the distances to its (at most) 8 neighbours and add this value as the height of each neuron in a third dimension. A machine learning algorithm sig* extracts rules out of the learned neural network (Ultsch et al. 1993). In distinction to other machine learning algorithms like ID3 our algorithm considers the components and only these which are relevant for the classification. This corresponds to the proceeding of a geological expert.

3. Biogene Sedimentation in Non-tropical Environments

Most recent studies of carbonates and the biology of reef-ecosystems are concentrated to tropical environments of lower latitudes. With the discovery of wide-extended carbonate deposits in higher latitudes during the last two decades, the STAR-model (Lees et al. 1972, Lees 1975) was used for a general differentiation of tropical vs. non-tropical carbonates. The model based on the annual ranges of salinity and temperature, as the mainly factors controlling the biogenic carbonate production. The occurrence of reefs and biogenic buildup's beyond the arctic circle (Freiwald 1993, Henrich et al. 1993) suggests that the controlling factors of carbonate production by marine ecosystems must be more complex as shown in the oversimplified STAR-model. In opposite to the coral-reefs of the tropical seas, the shallow water environments of higher latitude are dominated by coralline algal ecosystems and kelp-forests.

The geological investigations of the study were done in the Bay of Morlaix, a sub-tidal embayment at the rocky shore of Northern Brittany. During fieldtrips in 1991-1993 samples (46) of the superficial sediments were picked up by a research vessel.

The classification by neural networks of the different types of sediments deposited in the bay based on datas of micro-facies analysis. Micro-facies are defined as the total of all the paleontological (biological) and sedimentological criteria which can be classified in thin-sections, peels and polished slabs. In studies of recent sediments additional informations of physical environmental parameters are available.

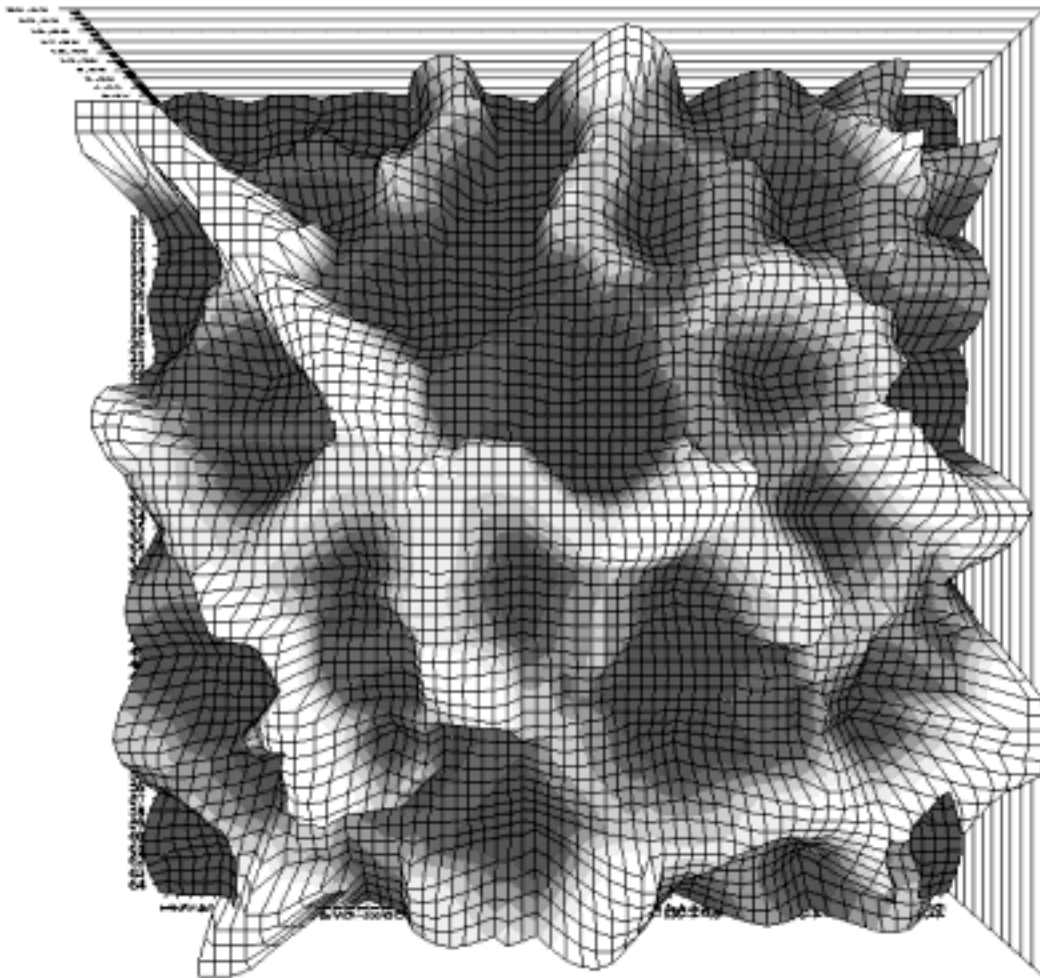


Fig. 2: U-Matrix of sedimentary facies as a three-dimensional landscape.

4. Results

Till now the classifications of sedimentary facies in geology research were done by statistical cluster-analysis. The results of these analysis are difficult to interpret in case of sedimenttypes with a high content of coarse grain particles (see Freiwald 1993). In contrast to that the analysis by neural network allows a differentiate classification of sedimentary facies in consideration of all informations of micro-

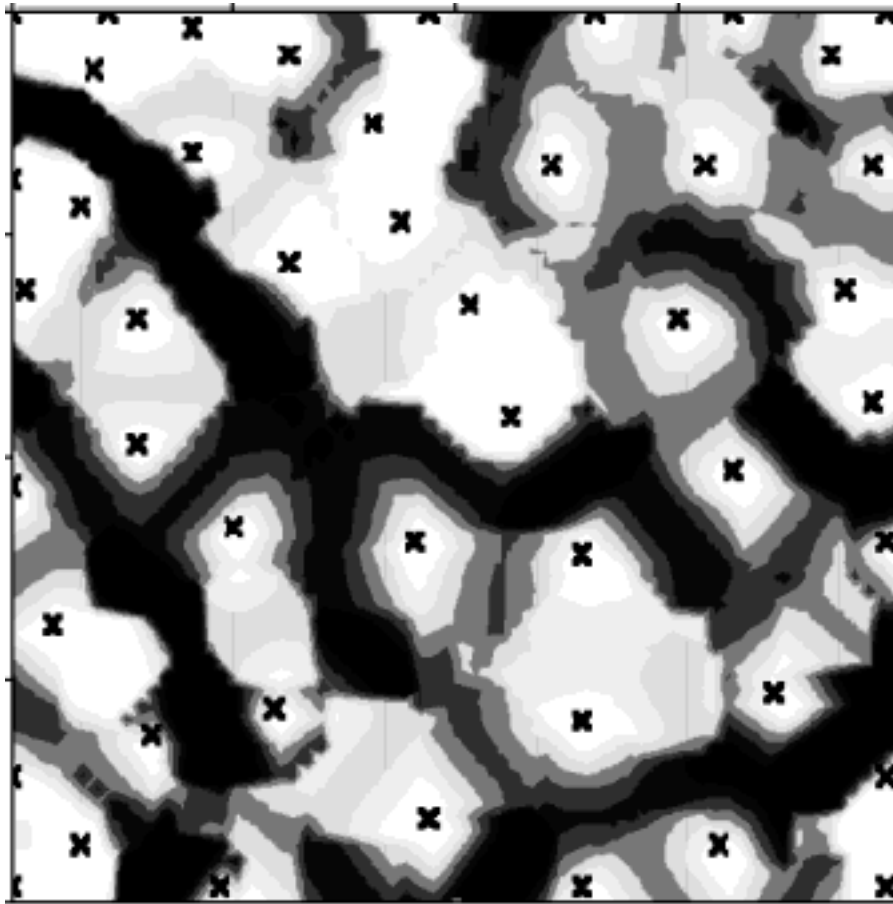


Fig. 3: Best matches of the input vectors assigned to the 2-dimensional contour map of the U-Matrix of sedimentary facies

facies-analysis. This classification gives detailed informations about source areas, direction of sediment-transport, relation of sedimentary facies to carbonate producing ecosystems and local accumulation phenomenas.

References

- Freiwald, A. (1993): Subarktische Kalkalgenriffe im Spiegel hochfrequenter Meeresspiegelschwankungen und interner biologischer Steuerungsprozesse. unpublished dissertation, Christian-Albrechts-University of Kiel.
- Henrich, R./ Reitner, J./ Wehrmann, A. (1993): Cold-water shelf carbonates and lag deposits and associated living benthic communities: Spitsbergen Bank, in: Pfannkuche, O./ Duinker, J.C./ Graf, G./ Henrich, R./ Thiel, H./ Zeitzschel, B. (Eds.), Nordatlantik 92, Reise Nr. 21.- Meteor-Berichte 93/4, Hamburg, pp. 176-180.
- Kohonen, Teuvo (1989): Self-Organization and Associative Memory, Springer-Verlag.
- Lees, A. (1975): Possible influence of salinity and temperature on modern shelf carbonate sedimentation, in: Marine Geology 19, pp. 159-198.
- Lees, A./ Buller, A.T. (1972): Modern temperate-water and warm-water shelf carbonate sediments contasted, in: Marine Geology 13, pp. M67-M73.
- Ultsch, Alfred (1992): Self-Organizing Neural Networks for Knowledge Acquisition. in: Proceedings of the European Conference on Artificial Intelligence, Wien, pp. 208-210.
- Ultsch, Alfred/ Li, Heng (1993): Automatic Acquisition of Symbolic Knowledge from Subsymbolic Neural Networks. in: Proceedings of the International Conference of Signal Processing, II, Beijing, China, pp. 1201-1204.