

MSC 68T10

THE TECHNIQUE OF INCREASE THE EFFICIENCY OF LEARNING NEURAL KOHONEN MAPS FOR RECOGNITION OF PERTURBATIONS GEOACOUSTIC EMISSION

A.V. Shadrin

Institute of Cosmophysical Researches and Radio Wave Propagation Far-Eastern Branch,
Russian Academy of Sciences, 684034, Kamchatskiy Kray, Paratunka, Mirnaya st., 7,
Russia

E-mail: SugerMas@yandex.ru

This work is dedicated to technique of training Kohonen maps on the example of geoacoustical signal in the subrange 1500-6000 Hz. Describes the parameters of learning the Kohonen maps to classify anomalies in geoacoustical signal on different types.

Key words: geoacoustical emission, geoacoustic signal, disturbance, neural Kohonen maps, learning

Introduction

Scientists from many countries have studied natural catastrophic phenomena for many years. One of the main aims of these researches is the forecast of such events, which would help to reduce human and economic losses. Science predicts hurricanes, floods and other disasters, and only earthquakes are unpredictable, killing people at home where they feel the most safe [1]. Though earthquakes occur suddenly, it has been scientifically proved that it takes some finite time to accumulate energy in a source for rock breaking [2],[3].

One of the perspective directions of investigations to detect anomalies preceding the earthquakes is registration and analysis of geoacoustic emission disturbances. This phenomenon is determined by deformation processes from the sources of future earthquakes. Investigations of geoacoustic emission are carried out by hydrophones oriented according to the cardinal points and installed in small reservoirs in Kamchatka [4], [5]. The frequency range under study is from 0.1 Hz to 10 kHz (the subranges are: 0.1–10 Hz, 10–50 Hz, 50–200 Hz, 200–700 Hz, 700–1500 Hz, 1500–6000 Hz, 6000–10000 Hz). Signals are stored for every 4 seconds on a PC hard disc. Data analysis showed that for the period 2001-2003, 34 from 74 earthquakes with the magnitude $M > 4$, occurred

Shadrin Alexander Vitalyevich – Junior Researcher of Lab. Acoustic Research, Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS.
©Shadrin A.V., 2014.

at the epicentral distance up to 250 km from observation points, were preceded by a high increase of geoacoustic emission level in kilohertz range within a day [4], [5].

It was ascertained that emission disturbance amplitude depends on earthquake magnitude and epicenter location. Besides the disturbances of geoacoustic nature, the systems register signals determined by bad weather conditions, by precipitation and strong wind, first of all. The frequency range of such effects is also from hundreds of hertz to units of kilohertz, and it is close to the range of disturbances before earthquakes. For the more detailed investigation of geoacoustic emission behavior before seismic events and for its recognition at the background of weather anomalies, there was a necessity to classify signals into the main types: rain, wind, anomalies of deformation nature. Due to the peculiarities of signal registration, application of standard frequency methods for processing is not effective. For this reason, the author applied one of neural network variants, Kohonen maps, the bases of which are data classification and clusterization.

The paper considers a method for Kohonen maps training on the example of a geoacoustic signal, and describes the techniques to solve the problems which arose during the training. To test the method, the subrange of 1500-6000 Hz was taken, since all the major events which need to be classified appear there the most clearly. Network training was carried out on the signal for 2007. The criteria, determining the choice, were the following: starting the operation of the digital system for data registration in this year, and high activity of the signal.

Choice of initial parameters for network training

At the first stage, the training sample was formed from elementary effects of anomalies of different nature with the duration of 80 sec (20 stored for 4 s samples). Then the network was trained by MATLAB program. There are no recommendations how to train Kohonen maps, thus, default training parameters suggested by the program were used. The dimensionality of the network was chosen to be two-dimensional.

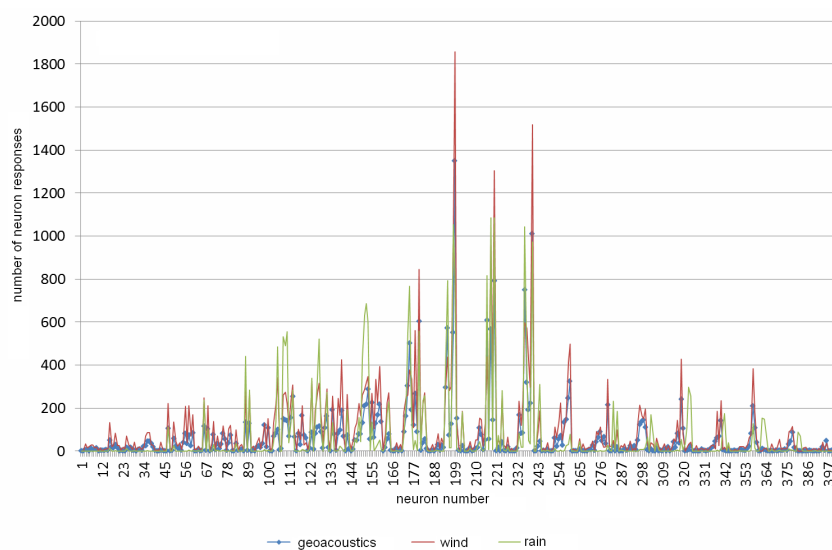


Fig. 1. Neuron distribution of events applying the recommended parameters for network training

After network training, neuron distribution of events was studied, which showed that the network distributed events evenly but could not classify anomalies of different types by separate neurons. It is clear that the same number of responses for signals of different types fall on the same neurons (Fig. 1).

Method for improvement of training efficiency of Kohonen maps

After having analyzed the neural network functioning, a method was developed allowing us to train the network and to separate the signal into different classes.

Due to the large dimensionality of the training sample, the following algorithm for data choice was used. Signal mathematical expectation (ME), its root-mean square deviation (RSD), and their relation to each other were calculated. Only signals which had a unique relation of ME and RSD fitted the training sample. In connection with the fact that it is impossible to determine the definite dependence of network neuron on the type of input signal, every neuron of the network was presented in the form of a ratio of anomaly distribution falling on it. It is considered on the example of estimation of disturbance distributions. Five responses of deformation nature anomalies, 1 response from wind, and 0 responses from rain fall on neuron 1. Processing the result, the estimation is be presented in the following form: 1 anomaly of deformation nature, 0 anomalies from wind, and 0 anomalies from rain. If event distribution on a neuron is even (5 responses of deformation nature anomalies, 5 from wind, and 5 from rain), the estimation is 0.3 0.3 0.3. If two events have the same number of responses, then 0.5 0.5 0 or 0 0.5 0.5 or 0.5 0 0.5. It is necessary to choose the network parameters so, that after the network training, the number of neurons with the result 0.3 0.3 0.3 is minimal, and with 0 0 1, 0 1 0 and 1 0 0 is maximal.

During the network training, it is very important to choose the appropriate number of training epochs that prevents from the lack of training and retraining of the map. There is a set of tools in MATLAB system allowing us to monitor the process of network training. One of them is the graphic representation of arrangement of network weight coefficients in vector space of the training sample during network training. Unfortunately, the tools allow us to see how the trained network covers the input data in space only on two first coordinates. The network frequently covers the whole data cloud unevenly, it is vary clear and noticeable. The training epoch, for which weight coefficients «covered» the input data, is taken as the datum point. Fig. 2 shows the uneven change of the arrangement of map weight coefficients.

After the network training, the total number of neuron responses for deformation nature anomalies is estimated. The network is trained with the step increase of epoch number relative to the datum point so long as the maximal number of neuron responses reaches the maximal level and stops growing. If we continue to increase the number of epochs further, the total number of responses will begin to decrease, and the network becomes retrained, as the consequence.

There may be a situation, when a network with the same number of responses but with different number of training epochs may be obtained. To evaluate the quality of the network, methods for estimation of neural networks, suggested in [6]-[8], may be used. It is qualitative (average quantization error) and quantitative (topographic error) assessment. The qualitative assessment shows the capability of a neural network to reveal the hidden structure and to cluster the data. Such assessment may be applied as

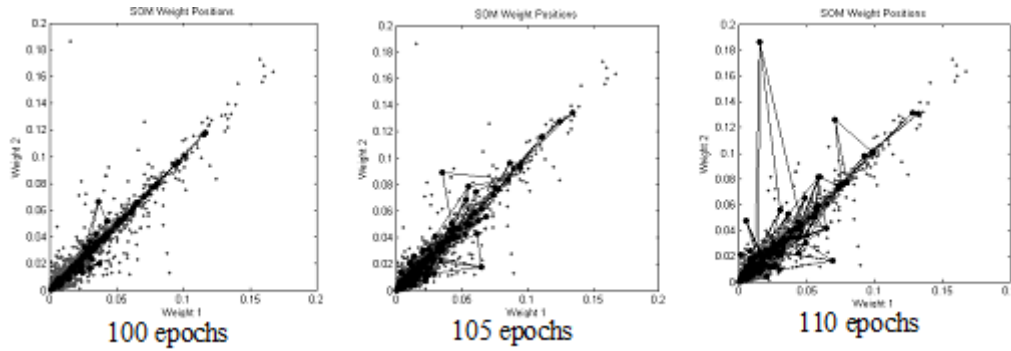


Fig. 2. Uneven change of arrangement of map weight coefficients at different stages of training. Small dots indicate the training sample; large dots, connected by lines, indicate weight coefficients

a map resolution measure. The average quantization error is estimated according to the formula:

$$eq = \frac{1}{N} \sum_{i=1}^N d(X, W_w),$$

where N is the total number of input vectors, participating in the assessment of the map; X is the current vector from the input sample; W_w is the vector of neuron winner weights on the current input effect.

The quantitative assessment determines the uninterrupted reflections of input vectors on the map space [6]-[8]. It measures the proportion of all data vectors, for which the first and the second neuron winners are non-adjacent. The less this error is, the better the map keeps its topology. The topographic error is estimated as follows

$$et = \frac{1}{N} \sum_{i=1}^N u(X),$$

where X is the current vector from the input sample; $u(X)$ is the function, possessing the value 0 if the first and the second neuron winners of the network are adjacent, and 1 in the contrary case.

Training results

After training the network with best characteristics, neuron distribution of events was re-investigated. It showed that the network could classify the larger part of anomalies of different types by separate neurons (Fig. 3).

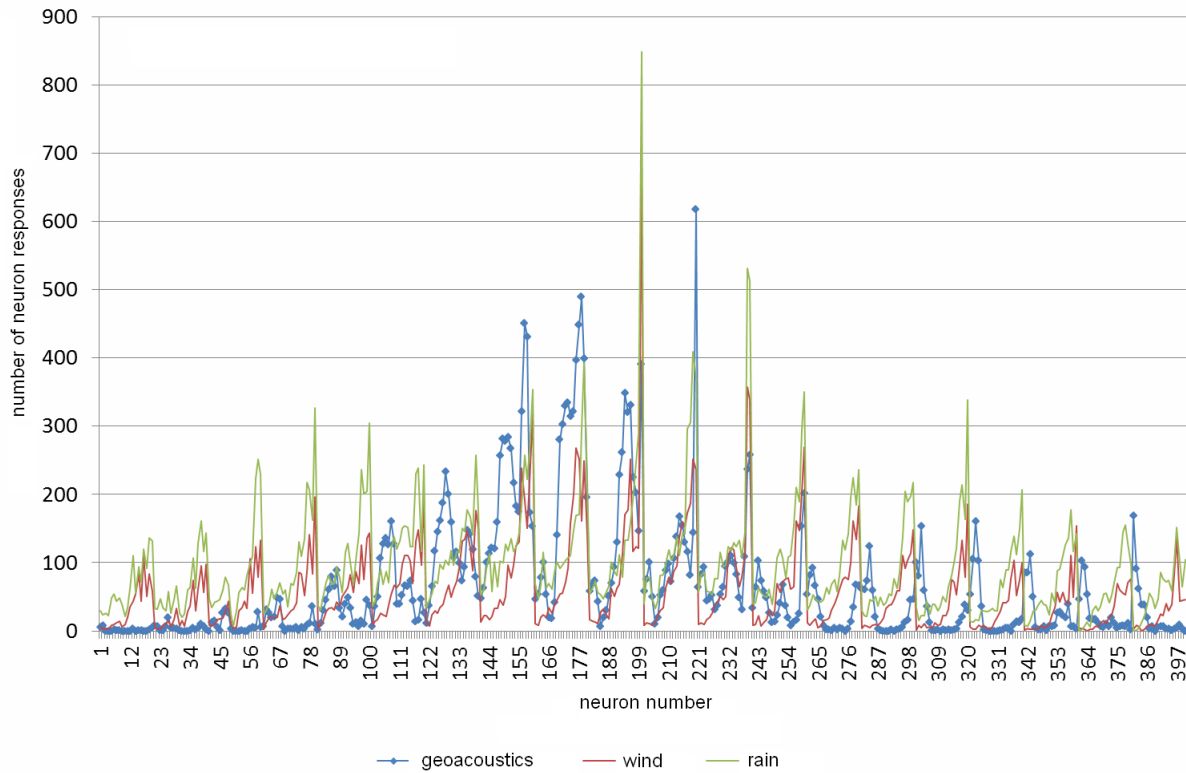


Fig. 3. Neuron distribution of events applying the best parameters of network training

Having chosen the network with best characteristics, the signal for 2007 was analyzed. The obtained assessment of the network was compared with earthquake catalogue [9]. After the comparison, the accuracy of network classification was estimated (Table)

Table

Percentage of signals for 2007 classified correctly

	February	March	April	May	June	July	August	September	October	November	December
Deformation nature anomalies	35	10	4	7	1	4	3	3	4	12	17
Weather anomalies	92	94	96	97	99	98	98	98	98	96	91

As it is clear from the table that the network handled well weather anomalies, and poorly the deformation nature anomalies. It is due to the fact that a great many of anomalies are similar in properties with the signal determined by weather anomalies. For this reason, to improve the recognition results, it is necessary to make additional study on the improvement of network training. Moreover, many experiments showed that it is necessary to perform complex analysis of all the signal subranges to achieve more exact results.

Conclusions

Thus, the developed method allows us to apply Kohonen maps to classify weather anomalies in geoaoustic signal. To improve the accuracy of recognition of deformation nature anomalies, additional investigations are required.

References

1. M. Rodkin. Prognoz nepredskazuemyh katastrof [Forecast unpredictable disasters]. *Vokrug sveta – Round the world*, 2008, no. 6, pp. 88-100.
2. A. G. Sobolev, A. V. Ponomarev. Fizika zemletryasenij i predvestniki [Physics and forerunners of earthquakes]. Moscow, Nauka Publ., 2003. 270 p.
3. Dobrovol'skij I.P., Zubkov S.I., Myachkin V.I. *Ob ocenke razmerov zony proyavleniya predvestnikov zemletryasenij. Modelirovanie predvestnikov zemletryasenij* [On the estimation of the size of the zone display of earthquake precursors. Simulation of earthquake precursors]. Moscow, Nauka Publ., 1980. pp. 7–44.
4. A. V. Kupcov, Yu. V. Marapulec, B. M. Shevcov. Analiz izmenenij geoakusticheskoy 'emissii v processe podgotovki sil'nyh zemletryasenij na Kamchatke []. *Issledovano v Rossii – Investigated in Russia*, 2004, vol. 262, pp. 2809-2818. URL: <http://zhurnal.ape.relarn.ru/articles/2004/262.pdf>.
5. Kupcov A. V., Larionov I.A., Shevcov B.M. Osobennosti geoakusticheskoy 'emissii pri podgotovke kamchatskih zemletryasenij [Features geoacoustic emission during preparation Kamchatka earthquakes]. *Vulkanologiya i seismologiya – Volcanology and Seismology*, 2005, no. 5, pp. 45-59.
6. J. Schatzmann. Using Self-Organizing Maps to Visualise Clusters and Trends in Multidimensional Datasets BEng thesis, Imperial College. June 19. 2003. URL: <http://mi.eng.cam.ac.uk/~js532/papers/schatzmann03soms.pdf>
7. Vesanto J. Data Exploration Process Based on the Self-Organizing Map, *Acta Polytechnica Scandinavica. Mathematics and Computing Series*, 2002, no. 115, pp. 96.
8. Arsuaga Uriarte, F. Diaz Martin. Topology Preservation in SOM. *PWASET*, 2006, vol. 15, pp. 187-191.
9. Mischenko M.A. Statisticheskij analiz vozmuschenij geoakusticheskoy 'emissii, predshestvuyuschih sil'nym zemletryaseniyam na Kamchatke [Statistical analysis of the perturbation geoacoustical emission prior to strong earthquakes in Kamchatka]. *Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of the Kamchatka Regional Association «Education-Scientific Center». Physical & Mathematical Sciences*, 2011, vol. 2, no. 1, pp. 56–64.

Original article submitted: 15.11.2014