RECOGNITION OF A FIRST SEISMIC ARRIVAL BY A NEURAL NETWORK

F.E.Vinberg
Institute of Earthquake Prediction Theory and Mathematical Geophysics
Russian academy of science, 117997, Moscow, Russia, winberg@mitp.ru

Abstract. The algorithm of recognition of the first arrival from an earthquake designed for using with high-noise seismic signals. Special attention has been paid to operating the algorithm under the conditions of high level seismic anthropogenic noise. The algorithm is based on signal recognition by a neural network. The type and parameters of the network must be specified independently for each region of possible application. The main idea of the algorithm is fast real-time detection of a sudden change in the monitoring of a noisy seismic process. It was shown that for detection of the first arrival of an earthquake it is enough only to use a few (from 4 and up to 5) seconds. The algorithm is developed as part of a system for use in automatic early warning earthquake systems that can work in very high noise.

Strong earthquakes often entail serious human and economic losses. The chief danger posed by earthquakes consists in their suddenness. Great losses of human life usually occur because there are no devices that could warn of the beginning of an earthquake. It is therefore high time mobile automatic information systems for analysis of seismic data in increased manmade noise should be developed [Low Kong Chiew, 2006; Horiuchi, 2003; Ionescu, 2007]. These devices are to be operated in real time with a minimum of possible computational delays and should envisage rapid decision making procedures.

There are two methods to implement ultrashort warning of an earthquake that has occurred: the method based on a dense network of seismic stations and the "single sensor" method. Both of these have their own advantages and drawbacks. [Allen, 1978]

The advantages of an early warning system based on a seismic network include a high reliability of warnings and the prediction of earthquake shaking intensity at each site in the alarm area [Nakamura, 1989]. The main drawbacks of such systems include:

- complexity and high costs involved, due to the necessity of deploying a dense network of seismic stations;
- unreliable operation during later shocks in case some of the stations, power supply or communications facilities have been damaged;
- the presence of an extensive dead zone (~50 km) where no warning is possible;
- difficulties in providing relevant information to the user;
- the necessity of deploying the sensors at low-noise sites.

The advantages of the early warning systems based on the "single sensor" method:

- expensive seismic networks are not required;
- self-contained operation is possible;
- the dead zone is reduced to 20 km or still less;
- information comes directly to the end user.

However, these systems also have significant drawbacks: the warnings are less reliable compared with the network-based systems and earthquake shaking intensity is not obtained.

One essential difference of the system to be developed from those available today consists in the analysis of local seismic data in increased manmade noise (that is to say, when the noise level is comparable with or even exceeds the level of the seismic signal coming from an earthquake) and in the possibility of self-contained operation. The algorithms can be used for individual devices of personal safety and for warning the residents in seismic regions over the globe.

The real time analysis of first arrivals of earthquake waves is used to compute the time that remains until the arrival of surface waves. Estimates show that the time difference between P and surface waves will be about 30 seconds when the epicenter is about 200 km distant from the monitoring site; this is quite sufficient for evacuation of people from potentially hazardous space, insertion of moderators at nuclear
power stations, pipeline interlocking, transportation stoppage, warnings issued to rescue services and to administration by automatic transmission of SMS messages, etc.

The standard detection methods are based on the assumption that the noise is stationary on a long enough segments of the record. When a seismic record made in a megacity is to be analyzed, one has to deal with noise types having very diverse origin and characteristics. Also, the noise level is comparable with the amplitude of the signal to be detected. At the same time, the model must only incorporate independent parameters of the signal, as required by the classical mathematical statistics. It should be borne in mind, however, that interdependent parameters taken as a whole may also carry useful information. We remind the reader that the number of dynamic parameters may be very large and some of these may even be disregarded as potential candidates to be included in the model. This circumstance impedes, and occasionally makes it altogether impossible, the use of conventional methods.

The problem of detecting seismic phases due to local earthquakes observed in intensive manmade noise can be reduced to the problem of signal classification. In dealing with the problem of signal classification one has, first, to generate the vector of classification features and, secondly, to choose the appropriate classifier.

Since the form of the statistical distributions involved is unknown in our problem, the unknown signal is classified using a neural network or a threshold discriminator. Threshold discriminators can be used with the vector of classification features having the unit dimension. If the vector's dimension is greater than one, neural networks are to be used [Gutierrez, 2009]. The division into classes is to proceed in successive steps, isolating one class from the others, then the next, and so on. In the case under consideration, we would classify earthquake signals as belonging to one class and all manmade signals and seismic noise to the other [Myachkin, 1978; Mogi, 1985].

The use of neural networks as a classifier has certain indubitable advantages:

- Neural networks are processing the incoming information in parallel, simultaneously by all the neurons, which makes it possible to carry out instrumental analysis of complex signals in real time.
- Neural networks can fit any continuous function. It is thus unnecessary to assume model hypotheses beforehand and even, in a number of cases, to make assumptions as to which variables are really important
- Neural networks examine the influence of input parameters on the results. The key characteristics are identified during the training of the system to extract essential information from redundant data.
- Neural networks can solve problems using incomplete, distorted, noise-contaminated, and internally contradictory data.
- Neural networks are invariant under changes in the dimension of the feature space.
- A trained neural network can be used even by an inexperienced user.

We now list several problems that still impede a wide use of neural network technologies:

- It is not every problem that can be solved by neural networks. There is no guarantee that a neural network can be trained for a finite time. Frequently, the effort and computer time for training do not lead to the desired result.
- Training can lead to the discovery of a local error minimum, so that the best solution will not be obtained.
- Every type of neural network training procedures is subject to some limitations.
- The classification criteria are hidden when a neural network is used. It is impossible to explain how a problem is being solved and to interpret that problem in conventional analytical terms, which are commonly used to construct the theory of a phenomenon.
- A neural network yields the result with some degree of probability, hence the networks are inapplicable to those problems where absolute exactitude is required.

The neural network system of earthquake detection is operated as follows. Data are imported into the system directly from a seismometer or from any archive. Similarly to the situation with the data in the learning set, conversion to ASCII code is carried out, data are corrected for seismometer response function, and a uniform sampling rate for the seismometers is synthesized. Further, a simple threshold switch determines the beginning of the portion that requires more careful analysis. An investigation window is formed: data in a portion of a definite length are transformed, normalized, and sent to the inputs of the trained neural network. According to the classification results, an output signal is generated by the system either identifying an earthquake or concerning the presence of high-level noise.
Our algorithm is based on the continuous wavelet transform and a neural network. The algorithm is adaptive, since it incorporates individual time-dependent characteristics of the time sequence we are processing.

The identification of the origin time of an earthquake is to be performed during an average of four seconds time. [Wu 2005; Kanamori, 2005; Olivieri, 2008; Lockman, 2007; Allen, 2009]

The basic data are measured time series of seismic data recorded from a seismic sensor in real time. The sampling rate of seismic data recording is 100 Hz. This study uses a database of seismic signals consisting of more than 120 sample earthquakes and natural noise. Also, this library of manmade seismic noise was supplemented with numerous noise records we received from the St. Petersburg State University.

The architecture of a network is chosen for the application considered. The Neural Networks package in the MatLab system was chosen to simulate the structure of neural networks. The package includes twenty types of neural networks and training rules, so that an architecture suitable to the problem in hand can be selected. The Perceptron (a linear discriminator) can be used for linearly separable signals. For the case of linearly inseparable signals good results are obtained by using probabilistic radial basis networks.

The choice of the optimal network is a non-trivial task. Frequently enough, successive training steps still do not succeed in reducing the error to an acceptable size. In that case, one has to modify the configuration and even to select a network of a different type. In our case, successive trial-and-error resulted in choosing a multilayer perceptron with a hidden layer and a single neuron in the output layer. We used the method of backward error propagation (BEP) for neural network training, because the signal data volume is large and the data are redundant.

The causes and manifestations of earthquakes are different in different parts of the globe. The use of back propagation neural networks in prediction-oriented instruments requires either localization, that is, training on representative samples (patterns) as stored during previous earthquakes in the area or the generation of a "versatile" set of patterns consisting of an archive of seismograms that characterize the earthquakes in all known seismic zones. The present study is based on the latter method.

Network training begins with inputting a feature vector and computing the corresponding response. Comparison with the desired response enables one to modify the coupling weights so that the network could show a more accurate result at the next step. The synapse weights are modified with incorporation of the local gradient of the error function. The information on the network outputs is the starting point for the neurons of the preceding layers. The error function is the difference between the current network output and the ideal response. According to the method of least squares, the neural-network error function is

$$E(w) = \frac{1}{2} \sum_{j,p} (y_{jpn} - d_{jp})^2,$$

where $y_{jpn}$ is the output state of the $j$-th neuron of layer $n$ as the $p$-th learning image comes at its inputs; $d_{jp}$ is the desired output state of that neuron.

It is required to make the network output as close to the desired as possible. The synaptic coefficients are corrected as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{di}{dw_{ij}},$$

where $w_{ij}(t)$ and $w_{ij}(t+1)$ are the weights of the coupling between the $i$-th and $j$-th neurons at the current and the next training step, $\frac{di}{dw_{ij}}$ is the derivative of the error function, $\eta$ the learning speed parameter. The greater is $\eta$, the less accurate will be the next decrease in the network total error, but the smaller that parameter, the longer time will be taken by the learning process, thereby making the network more likely to be trapped in a local minimum of the error function. The simplest learning rule uses the method of steepest
descent, which is to say, the changes in the synoptic coefficients are proportional to their respective contributions into the total error. This algorithm when developed further becomes the method of gradient descent with moments incorporated. We are to add to the weight correction a value that is proportional to the change in that weight coefficient at the preceding step:

\[ w_j(t + 1) = w_j(t) + \eta \frac{dE}{dw_j} + \alpha \cdot w_j(t), \]

where \( \alpha \) is the coefficient of inertia.

\[ \alpha = \sqrt{N}, \]

where \( N \) is the number of couplings associated with a neuron.

The Neural Networks package envisages computation of learning quality. Unfortunately, the computation in question is an estimate of how far the network architecture is in agreement with the learning data and is not a measure that enables one to draw an unambiguous conclusion as to the quality of subsequent operation by the trained network. The classification quality can only be computed after network training is complete. A reserved test sample is used for the purpose, and that sample too must be representative enough.

Obviously, the configuration of the neural network should be large enough in order to provide a high reliability for earthquake detection, reducing the error of the first kind (failure to detect an earthquake that has actually occurred) and the error of the second kind (erroneous triggering, resulting in detection of a nonexistent earthquake). On the other hand, a composite neural network or a neural network with a large number of neurons may unjustifiably inflate the power consumption of the device and increase the triggering time (reducing the speed), thus considerably increasing the size of the printed circuit board and that of the entire device. The experiments we conducted in the Matlab environment showed that an increased dimensionality of the hidden layer (by more than 50 neurons) does not reduce the learning error, hence does not enhance the accuracy and reliability of predictions.
At first the feature vector is generated on an intuitive basis. It is sufficiently difficult to determine which parameters of a signal are the most important. This can be explained by, first, the fact that earthquake signals are different, even signals due to one and the same event when recorded at different stations are rather different and, secondly, by noise diversity, especially manmade noise. One may try to proceed in a straightforward manner by inputting normalized amplitudes of the signal to the 1024 neurons of the input layer. One can also supplement these with spectral values and wavelet transform coefficients [Tupitsyn, 2009; Gravirov, 2009]. The number of neurons in a network is thus dependent on the length of the analysis window, the frequency range of interest, and the presence of extra parameters. The network configuration is little by little modified, the usual changes being to vary the states of the synoptic weights so as to minimize the observed error.
precision class of the future instrument: the more precise the instrument, the greater number of patterns will be required, and the longer will be the training proper.

Conclusions

The adaptive algorithm used to detect the first arrival of an earthquake was extensively tested in real time with the algorithm parameters being adjusted during this process. When we insisted on being able to detect all events, the result was a rather large probability (about 15%) of false alarms, whereas when we wanted to remove false alarms altogether, the probability of predicting an event reduced to 92%.

True, the results as reported were achieved by training our neural network on the "versatile" set of representative samples coming from earthquakes occurring all over the globe. If we train our neural network using local data pertinent to a specific region, the error of the system must be significantly reduced.

With this approach, the creation of a self-contained system for rapid warning of an earthquake under increased levels of manmade noise becomes realistic. The algorithm could then be used in warning devices at high-risk facilities and for warning the population in seismic regions of the world.

References