Volume 3, Issue 10, October 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Oil Exploration Data Mining Image Processing using Various Artificial Neural Network Algorithms

Hempreet Kaur Sandhu, Sandeep Kaur
SBBSIET, Punjab Technical University
India

***Abstract---** The way the prices of oil are rising and the Indian currency is getting devalued there is a need for improving the oil exploration in India. This paper presents a new technique for improving oil exploration to enhance the chances of success and to reduce the cost and time involved in searching for oil reserves. The seismic prospecting collects seismic data then this seismic data is processed and finally this data is analysed with computer as the main tool predict oil and know the positions which might generate oil by analysing the properties of the rock and the layer relevant to the oil. This paper presents an improved software tool that can help in explaining the seismic data can help in finding the oil field accurately and economically using image registration to determine a transformation between two image spaces or between an image space and a physical space so that correspondent features can be matched using Neural Networks for better accuracy.*

***Key Words :** Data Mining, Artificial Neural Networks, Oil Exploration*

1.0 Introduction

Almost everyone has been impacted by the rise in petroleum prices. From the household economy to the viability of large industrial houses, all are dependant on petroleum prices. The need for oil exploration in India cannot be overemphasized. Many new techniques have been evolved in the area of oil exploration to enhance the chances of success and to reduce the cost and time involved in searching for oil reserves.

Oil generally exits very deep in the earth's crust. It is hard for geologists to detect the geological body containing oil directly. Drilling costs much more without knowing where the oil may exist. To overcome this problem geologist's use geophysical prospecting to explore oil in which they use instruments to make seismic wave on the surface of ground. This wave will reflect and refract back when it meets different geological layers. The sensors on the ground can receive those waves and record the relevant data. The geologists can estimate or determine where the oil might exist [1].

The first step seismic prospecting is collecting seismic data then this seismic data is processed and finally this data is explained. The first step is mainly about preparation of the outdoor exploration which has strict standards. However, it is necessary to know about the survey line. The survey line normally is a straight lineon which many observing points are arranged according to the position of the detector, besides the survey line should be perpendicular to the geological structure [1]. In second step, as computer is the main tool to process the data, the information recorded is transferred on the magnetic media into the data of sectional views. Moreover, the data is de-noised to improve the quality of the data. This data contains the underground geological information. However, what is needed is to know is the positions which might generate oil and the properties of the rock and the layer relevant to the oil. Hence it is needed to find the information about the oil from the seismic data. This is crucial to the fact whether the geologists can find the oil field quickly and accurately. In other words, explaining the seismic data needs to be transferred into the cognitive information of the explored target oil field. In this process, the data will firstly be digitally handled then the data becomes numbers of stacked sections and migrated sections ora three-dimensional data body. The data of these sections will be synthetically analyzed, computed and contrasted repeatedly at the workstation referring to the relevant geological, drilling and measuring data. Then the data can be presented in the form of figure or chart. As is generally known, drilling a well may lead to huge expenditure, especially when exploration is in the sea. Therefore, software tools can help in explaining the seismic data can help in finding the oil field accurately and economically [1].

For accurate and early detection of oil Image Data Mining Techniques are very helpful. Image registration is a procedure to determine a transformation between two image spaces or between an image space and a physical space so that correspondent features can be matched using Neural Networks for better accuracy [2].

We can consider an artificial neural network as a highly simplified model of the structure of the biological neural network. An ANN consists of interconnected processing units. The general model of a processing unit consists of a summing part followed by an output part the summing part receives N input values, weights each value, and computes a weighted sum. The weighted sum is called the activation value. The output part produces a signal from the activation value. The sign of the weight for each input determines whether the input is excitatory (positive weight) or inhibitory (negative weight). The inputs could be discrete or continuous data values, and likewise the outputs also could be discrete or continuous [2].

In an ANN several processing units are interconnected according to some topology to accomplish a pattern recognition task. Therefore, the inputs to a processing unit may come from the outputs of other processing units, or from external sources. The output of each unit may be given to several units including itself. The amount of the output of one unit received by another unit depends on the strength of the connection between the units, and it is reflected in the weight value associated with the connecting link. If there are N units in a given ANN, then at any instant of time each unit will have a unique output value. The set of the N activation values of the network defines the activation state of the network at that instant. Likewise, the set of the N output values of the network defines the output state of the network at that instant. Depending on the discrete or continuous nature of the activation or output values, the state of the network can be described by a discrete or continuous point in an N – dimensional space. Depending on the arrangement of various layers and neurons there are different architectures for ANN [2].

2.0 LITERATURE REVIEW

Ninget al in their work on oil exploration data mining image processinghave designed an oil exploration software. The software is based on the procedures of oil exploration and seismic prospecting. In the environment of Microsoft Visual Studio 2008 with the C++ programming language, the standard SEG-Y seismic data is successfully transferred into visible geological sectional view and histogram chart, which makes the geologists obtain a quick access to the idea of geophysics structures of the target field [1]. Yin et al have worked on predicting oil and gas reservoir and calculating thickness of reservoir from seismic data using neural network. Neural networks have been developed and widely used in oil and gas exploration. Combining the neural networks with traditional seismic prospecting methods, authors have put forward a new method to predict oil and gas reservoirs and calculate the thickness of the reservoirs. This method is applied in oilfields and the results are satisfactory [2].

Pitas et al have proposed Bayesian estimation in seismic migration [3] and Denham has published a paper on seismic interpretation. In seismic exploration, an acoustic energy source radiates elastic waves into the earth from the surface; receivers on the surface detect acoustic energy reflected from geological interfaces within the earth [4].

Mousa et al have presented Seismic Migration which is a Digital Filtering Process Reducing Oil Exploration Risks. To produce oil or gas, it is needed first to determine the subsurface structure. This can be done by the reflection seismology method. This geophysical technique relies on the generation of artificial seismic waves and the recording of their reflections from different geological layers. However, such acquired seismic data does not reveal an accurate image of the subsurface unless appropriate signal processing is used such as frequency and/or wavenumber filtering, multichannel and/ or multidimensional filtering, deconvolution, and so on [5] and Kamel et al have proposed seismic computations on the IBM 3090 Vector Multiprocessor [6].

Wood et al have worked on seismic signal processing. Seismic prospecting for oil and gas has undergone a digital revolution during the past decade. Most stages of the exploration process have been affected: the acquisition of data, the reduction of this data in preparation for signal processing, the design of digital filters to detect primary echoes (reflections) from buried interfaces, and the development of technology to extract from these detected signals information on the geometry and physical properties of the subsurface [7]. Mendel et al have worked on white-noise estimators for seismic data processing in oil exploration. This paper is motivated by a problem from seismic data processing in oil exploration [8]. Donno et al have proposed Seismic Velocity and Polarization Estimation for Wavefield Separation. Authors address the problem of estimating the shape parameters of seismic wavefields using linear arrays of three-component (3C) vector sensors with uncertain acquisition geometry. The goal is to separate the different seismic waves, which is of practical need for oil exploration and geophysics. Authors present a parametric model for multiple wideband polarized signals received by an array of three-component sensors with positional and rotational calibration errors, and derive the Cramer-Rao lower bounds on the performance of the model parameters for both the exact physical model and the model with uncertain acquisition geometry [9]. Guizzo et al have presented Winner: Geophysics Solving the Oil Equation. In this paper, the Kaleidoscope code is developed and used as a seismic imaging tool for oil exploration [10]. Johnson et al have described Seismic exploration in the North Sea. The use of accurate seismic data, although expensive, is a very cost-effective technique in an uncompromising environment such as the North Sea [11]

Bois et al have worked on Applications of Pattern Recognition to Oil and Gas Exploration. This paper presents an introduction to pattern recognition, a summary of previous applications in seismic processing, and several new pattern recognition approaches. Topics discussed include impluse recognition, horizon correlation, determination of the boundaries and nature of a reservoir, and two-dimensional Fourier (f-k) filtering to attenuate ground roll. To conclude, authors examine the outlook for pattern recognition and propose increased use of the theory of fuzzy sets for pattern recognition applications in oil and gas exploration [12]. Delaney et al have proposed VizSim technology helps find oil faster [13]. It can be seen from the literature survey presented above that there is definitely a need for developing tools for helping in improving the accuracy and reducing the cost involved in oil exploration and Image processing based data mining has played a major role in analyzing seismic data files for predicting oil. No one has till date tried to improve the accuracy and speed of oil exploration by trying out architecture change in Artificial Neural Networks for image processing using data mining. Thus there is a need to Enhance Oil Exploration Data Mining Image Processing using Artificial Neural Networks Architecture changes.

3.0 SIMULATION MODELLING DETAILS

Since seismic details of various parts of India are not available so seismic data files from US have been used to make this system. These files are free to download from seismic observatory database resources. These have been put there to help agencies in predicting earthquakes in those regions.

The first step was to download sample seismic data file images from USGS. The USGS is a science organization that provides impartial information on the health of our ecosystems and environment, the natural hazards that threaten us, the natural resources we rely on, the impacts of climate and land-use change, and the core science systems that help us provide timely, relevant, and useable information. The raw image data was then pre-processed in the form of SEG-Y standard IEEE format and made suitable for input to matlab. Then feature extraction was performed using the image processing toolbox of Matlab and a database was constructed in csv format for input to the neural network. Using the Neural Network toolbox of matlab ANN was constructed to improve the accuracy of oil prediction. Matlab was used to simulate ANN. Attempt was made to Enhance the Accuracy of Detection of Oil in Seismic Data File using Artificial Neural Networks. Parameters of the Seismic data were used as input to the image. Neural networks have proved themselves as proficient classifiers and are particularly well suited for addressing non-linear problems. Given the non-linear nature of real world phenomena, like Detection of Oil in Seismic Data File, neural networks is certainly a good candidate for solving the problem. The Parameters will act as inputs to a neural network and the Enhance the Accuracy of Detection of Oil in Seismic Data File will be the target. Given an input, which constitutes the measured values for the parameters of the Seismic data, the neural network is expected to identify if the accuracy has been achieved or not. This is achieved by presenting previously recorded parameters to a neural network and then tuning it to produce the desired target outputs. This process is called neural network training. The samples will be divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy. The trained neural network will be tested with the testing samples. The network response will be compared against the desired target response to build the classification matrix which will provide a comprehensive picture of a system performance.

Different architectures such as cascade-forward backpropagation networks, Elman backpropagation networks, feed-forward backpropagation networks, feed-forward input-delay backprop networks, regression neural networks, Hopfield recurrent networks, learning vector quantization networks and backpropagation networks will be employed to the Artificial Neural Network model to improve the accuracy of the neural network and increase the speed of Oil detection.

The main problem faced during simulation work was the conversion of seismic data files into format usable by matlab. The headers had to be suitable modified to be made suitable for reading into matlab workspace.

The second problem faced was due to the size of seismic data files. Sne the seismic data is collected over the years it occupies a large space in the RAM. To avoid memory overrun problem the system variable not required were cleared as soon as possible and a computer of higher memory was used.

4.0 RESULTS OF SIMULATION FOR MINIMIZATION OF ERRORS IN OIL DETECTION USING ANN

The first sets of results of the simulation work are the spectrum files obtained from the SEGY files. The figure below shows the 1st spectrum file for the area which has oil reserves.

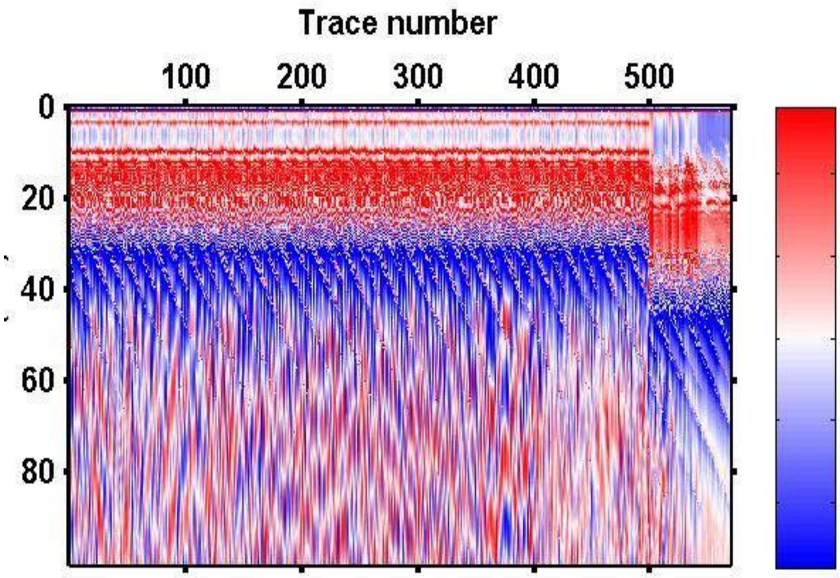


Fig 1. Spectrum file for the area which has oil reserves

The matlab function computes and displays the spectra of seismic data in interactivelyselected windows. Null values in the input data set are replaced by zeros prior to spectrum computation. Seismic data set is the input to the function and one or more cell arrays; the first element of each cell array is akeyword, the other elements are parameters. The range of frequencies to display have been chosen appropriately. If the end frequency is greater than the Nyquist frequency of the data set with the smallest sample interval, itis set to the Nyquist frequency. The amplitude spectra are normalized. Trace with fewer than padding samples are padded withzeros. This parameter is ignored if the number of samples per trace exceedpadding. Logarithmic scale (dB) has been used for amplitude spectrum. A pseudo-headerietrace_nohas been used. The window applied to the data before computing the FFT is Hamming The second sets of traces are the performance graphs for various neural network architectures. Performance was measured in terms of mean squared error, and shown in log scale. It rapidly decreased as the network was trained. Performance is shown for each of the training, validation and test sets. The version of the network that did best on the validation set will be chosen after training. The trained neural network was tested with the testing samples. This gives us a sense of how well the network will do when applied to the real seismic data. The seismic characteristics will be divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy. The trained neural network will be tested with the testing samples. The network response will be compared against the desired target response to build the classification matrix which will provide a comprehensive picture of a system performance. The figure below shows the progress of feed-forward input-delay back propagation network.

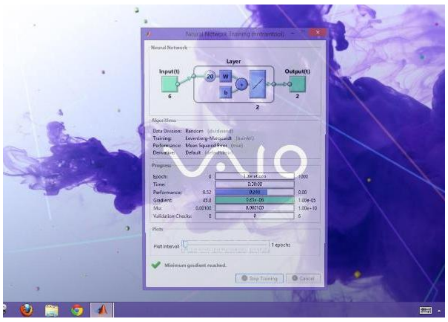


Fig 2. Progress of feed-forward input-delay back propagation network

The performance plot and the training state plot for feed-forward input-delay back propagation network are as given below. The characteristics obtained from segy file of seismic data act as inputs to a neural network and the prediction of oil is the target. Given an input, which constitutes the measured values for the parameters of the seismic data, the neural network is expected to identify if the area where this seismic observations have been made has oil or not. This is achieved by presenting previously recorded seismic parameters to a neural network and then tuning it to produce the desired target outputs. This process is called neural network training. The figure below shows the training state plot for feed-forward input-delay backpropagation network. It can be seen that the Mean Square Error falls from 10 to 1 in ½ an epoch.

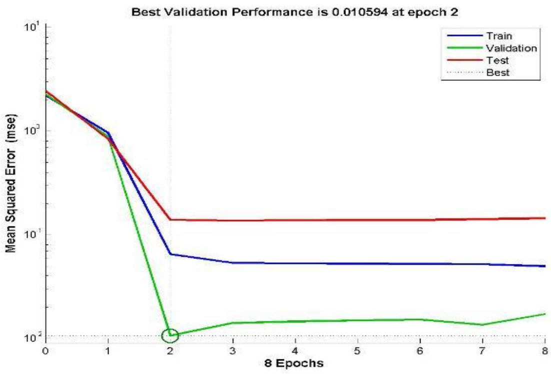


Fig 3. Mean Square Error

For a given input, the system always responds with the same output. On the other hand, a stochastic unit has a binary-valued output which is a probabilistic function of the input activity net.

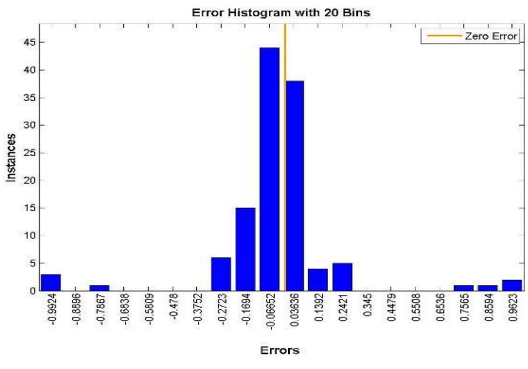


Fig 4. Histogram Distribution

Also, this allows for a natural mapping of optimal stochastic learning and retrieval methods onto neural network. Gradient descent/ascent has been employed in a simple search strategy which will allow the discovery of global minima. The gradient of learning process is shown in the figure below for layered-recurrent network. It can be seen that mu varies from 0.001 to 0.1 in 8 epoch.

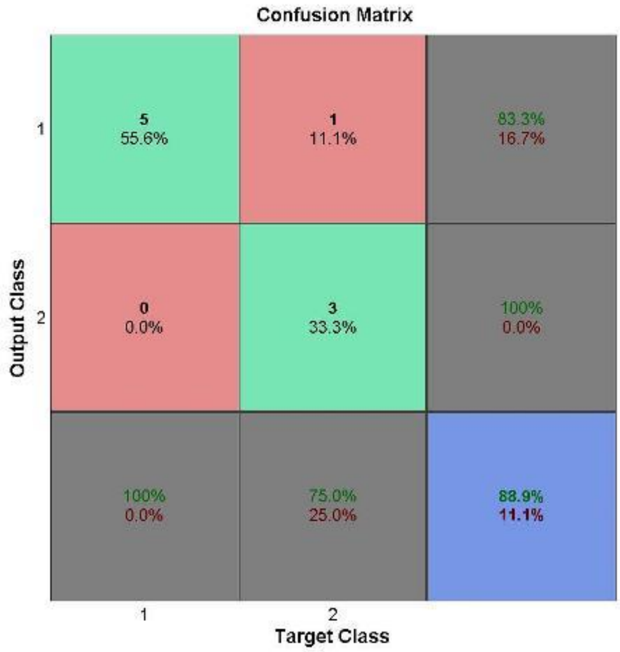


Fig 5. Confusion Matrix

Neural networks are noise tolerant. However, there is a limit to this tolerance; if there are occasional outliers far outside the range of normal values for a variable, they may bias the training. The best approach to such outliers is to identify and remove them either discarding the case, or converting the outlier into a missing value. If outliers are difficult to detect, a block error function may be used, but this outlier-tolerant training is generally less effective than the standard approach. After the training reached optimal number of epochs and was stopped the MSE was observed.

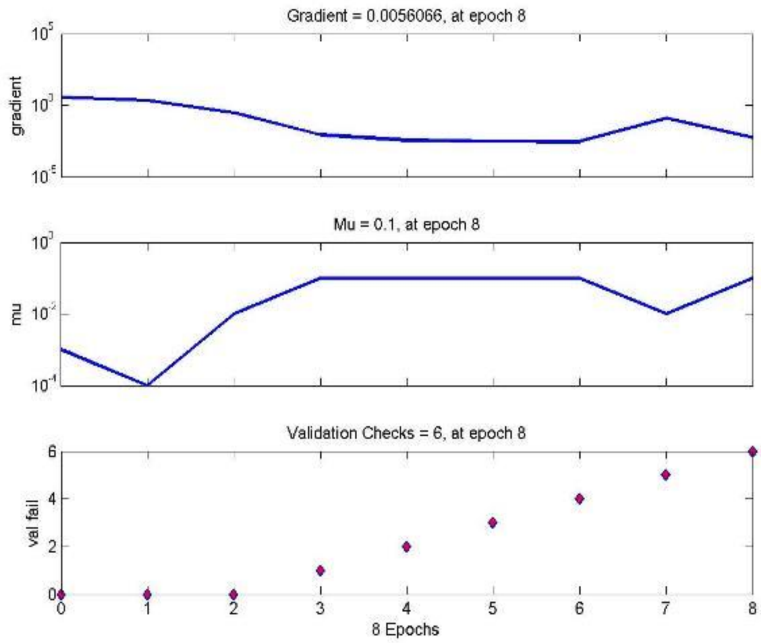
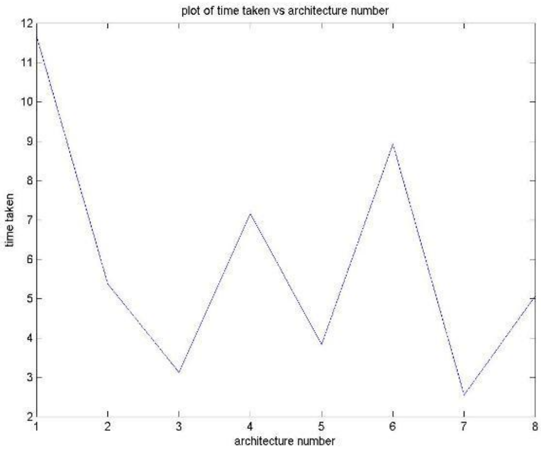


Fig 6. Gradient of Training

The figure shows the excellent correlation between the training samples and the corresponding network outputs. The correlation with the test data is good. The character of errors is best visualized by the histogram of differences of the true values and the network outputs. The figure shows the error histogram of the training data set and the testing data for layered-recurrent network. It can be seen that the instances of errors increase to 44 for -0.06652.

In the field of Artificial Neural Networks, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm for a supervised learning of ANN. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. It makes it easy to see if the system is confusing two classes or commonly mislabeling one as another. The figure shows the confusion matrix for layered-recurrent network. It is clear from the table that Percentage Correct oil detection is 55.6% and Percentage Incorrect oil detection is 11.1%

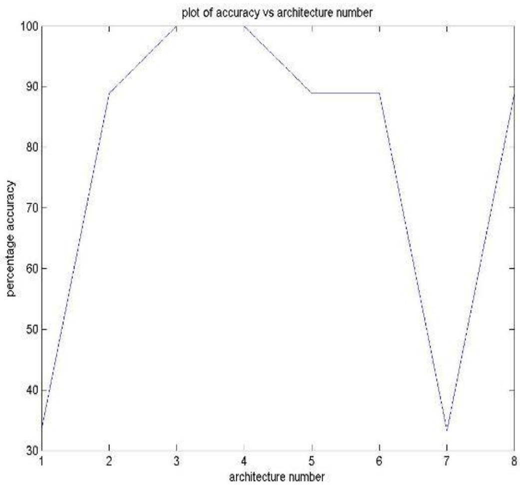
The figure below shows the variation of speed of oil detection observed for different architectures of ANN.



1. Feed-Forward Input-Delay Backprop Network
2. Layered-Recurrent Network
3. Feedforward Backpropagation Network
4. Cascade-Forward Backpropagation Network
5. Elman Backpropagation Network
6. Distributed Time Delay Neural Network
7. Feedforward Input Time-Delay Backpropagation Network
8. back propagation network

Fig 7. Comparison of Different Architectures wrt Speed

It can be concluded for the result above that the speed is maximum for Feedforward Back Propagation Network



The figure below shows the accuracy of oil detection observed for different architectures of ANN.

1. Feed-Forward Input-Delay Backprop Network
2. Layered-Recurrent Network
3. Feedforward Backpropagation Network
4. Cascade-Forward Backpropagation Network
5. Elman Backpropagation Network
6. Distributed Time Delay Neural Network
7. Feedforward Input Time-Delay Backpropagation Network
8. back propagation network

Fig 9. Comparison of Different Architectures wrt. Accuracy

It can be concluded for the result above that the Accuracy is maximum for Layered-Recurrent Network and Feed forward back propagation network.

5.0 CONCLUSION AND FUTURE SCOPE

Oil exploration is a very important for defence and industrial growth. Almost everyone has been impacted by the rise in petroleum prices. From the household economy to the viability of large industrial houses, all are dependant on petroleum prices. The need for oil exploration in India is very important. This research work will help India in the area of oil exploration to enhance the chances of success and to reduce the cost and time involved in searching for oil reserves.

It can be concluded from the simulation work that although high accuracy can be achieved in oil exploration using feedforward backpropagation network network the time taken for detection of oil is the least for feedforward backpropagation network. So it can be concluded that for oil exploration the best tool is to use feedforward backpropagation network for achieving both accuracy and speed. It can be seen from the literature survey presented above that there is definitely a need for developing tools for helping in improving the accuracy and reducing the cost involved in oil exploration and image processing based data mining has played a major role in analyzing seismic data files for predicting oil. No one has till date tried to improve the accuracy and speed of oil exploration by trying out architecture change in Artificial Neural Networks for image processing using data mining. Thus there was a need to Enhance Oil Exploration Data Mining Image Processing using Artificial Neural Networks Architecture changes.

In the present work literature survey was conducted study all aspects of image processing based data mining for analyzing seismic data files for predicting oil and software was developed for reading seismic imaging data files and extracting features from the image files. Data mining program was for handling features extracted from seismic imaging data files for analysis and the accuracy and speed of oil exploration was improved by trying out various Artificial Neural Network Algorithms like cascade-forward backpropagation networks, Elman backpropagation networks, feed-forward backpropagation networks, feed-forward input-delay backprop networks, regression neural networks, Hopfield recurrent networks, learning vector quantization networks and back propagation networks.

In the future other soft computing techniques such as Support Vector Machines (SVM), Fuzzy logics (FL), Evolutionary computation (EC), Evolutionary algorithms, Genetic algorithms, Differential evolution, Metaheuristic and Swarm Intelligence, Ant colony optimization, Particle swarm optimization etc can be used to further enhance the accuracy and speed of the oil detection algorithms.

REFERENCES

[1]. Ning Li; Xiao Yu; Qingqiang Wu; Maowen Wu; , "Oil exploration data mining image processing," Computer Science & Education (ICCSE), 2012 7th International Conference on , vol., no., pp.1115-1120, 14-17 July 2012
[2]. Yin Xing-Yao; Wu Guo-Chen; Yang Feng-Li; , "Predicting oil and gas reservoir and calculating thickness of reservoir from seismic data using neural network," Signal Processing, 1996., 3rd International Conference on , vol.2, no., pp.1601-1604 vol.2, 14-18 Oct 1996
[3]. Pitas, I.; Venetsanopoulos, A.; , "Bayesian estimation in seismic migration," Acoustics, Speech and Signal Processing, IEEE Transactions on , vol.36, no.2, pp.252-264, Feb 1988
[4]. Denham, L.R.; , "Seismic interpretation," Proceedings of the IEEE , vol.72, no.10, pp. 1255- 1265, Oct. 1984
[5]. Mousa, W.A.; , "Seismic Migration: A Digital Filtering Process Reducing Oil Exploration Risks [Applications Corner]," Signal Processing Magazine, IEEE , vol.29, no.3, pp.117-123, May 2012
[6]. Kamel, A.; Kindelan, M.; Sguazzero, P.; , "Seismic computations on the IBM 3090 Vector Multiprocessor," IBM Systems Journal , vol.27, no.4, pp.510-527, 1988
[7]. Wood, L.C.; Treitel, S.; , "Seismic signal processing," Proceedings of the IEEE , vol.63, no.4, pp. 649- 661, April 1975
[8]. Mendel, J.; , "White-noise estimators for seismic data processing in oil exploration," Automatic Control, IEEE Transactions on , vol.22, no.5, pp. 694- 706, Oct 1977
[9]. Donno, D.; Nehorai, A.; Spagnolini, U.; , "Seismic Velocity and Polarization Estimation for Wavefield Separation," Signal Processing, IEEE Transactions on , vol.56, no.10, pp.4794-4809, Oct. 2008
[10]. Guizzo, E.; , "Winner: Geophysics Solving the Oil Equation," Spectrum, IEEE , vol.45, no.1, pp.32-36, Jan. 2008
[11]. Johnson, R.; , "Seismic exploration in the North Sea," Electronics and Power , vol.21, no.20, pp.1122-1124, November 27 1975
[12]. Bois, Pierre; , "Some Applications of Pattern Recognition to Oil And Gas Exploration," Geoscience and Remote Sensing, IEEE Transactions on , vol.GE-21, no.4, pp.416-426, Oct. 1983
[13]. Delaney, B.; , "VizSim technology helps find oil faster," Computer Graphics and Applications, IEEE , vol.19, no.2, pp.10-16, Mar/Apr 1999

