APPLICATION RESEARCH OF TSUNAMI MONITORING AND FORECASTING USING DATA MINING TECHNIQUES

Mao Chongyuan[1] and Fu Qiang[2]

[1]*Undergraduates, Department of Oceanography, Ocean University of China, Qingdao, China*

[2]*Ph.D Candidate, School of Architecture, Tsinghua University, Beijing, China*

Email:mchy_2005@hotmail.com, fuqiangfio@163.com

ABSTRACT:

85 percent of the tsunami triggered by earthquakes, it's difficult to detect and of great danger. Since the earthquake-triggered tsunami in the open sea off spreads quickly, it is difficult to carry out effective monitoring and forecasting efficiently and timely. However, the path and the impact of the tsunami indeed follow certain rules which are not evident. Thanks to the use of data mining techniques, it is possible to analyze the factors which can cause the tsunami, or to evaluate the impact of the factors, and also to find the relation between these factors. This paper focuses on using appropriate data mining algorithms such as artificial neural networks, principal component analysis and cluster analysis to analyze related data to find the generating mechanism of earthquake tsunami and then to create the earthquake generating model, and also to analyze to create quick impact evaluating model of tsunami to coastal areas, utilizing the spatial distribution data of Chinese inshore seabed structure, topography and faulting, the historical data and dynamic data of the seismic source depth, seismic source location, seismic waves, and earthquake grades, the data of satellite based sea shock wave monitoring, data buoys, coastal tide level real-time monitoring, and the geo-data of coastal zone terrain and tied-protection facilities.

KEYWORDS: Data mining, Tsunami monitoring, Tsunami forecasting

According to statistics of the tsunami data, the tsunami has occurred more than 260 times on Earth since 18th century, which has caused about 300,000 deaths along the Pacific coast since 8th century. Table1 shows the great tsunami in recent 100 years. Since each tsunami can cause great waves which make massive people death and property loss, it is very necessary to find good way to forecast the possible tsunami.

Table1 Great tsunami in recent 100 years

DATE	TSUNAMI SOURCE LOCATION		WAVE HEIGHT(M)	DAMAGE	DEATHS
	COUNTRY	NAME			
07 April 1934	NORWAY	TAFJORD	64		40
13 September 1935	NORWAY	LOEN	74		73

This paper is supported by Inshore Marine Comprehensive Survey and Evaluation of Special Program administrated by China State Oceanic Administration. Program number:908-03-01-13.

27 October 1936	USA	LITUYA BAY, AK	149.35	roughly correspondin g to less than $1 million	
10 July 1958	USA	SE. ALASKA, AK	525	$0.1 million	2
23 March 1960	JAPAN	SANRIKU	159		
28 March 1964	USA	PRINCE WILLIAM SOUND, AK	67.1	$124 million	124
18 May 1980	USA	WASHINGTO N	250		

1. TSUNAMI

Cause of tsunami

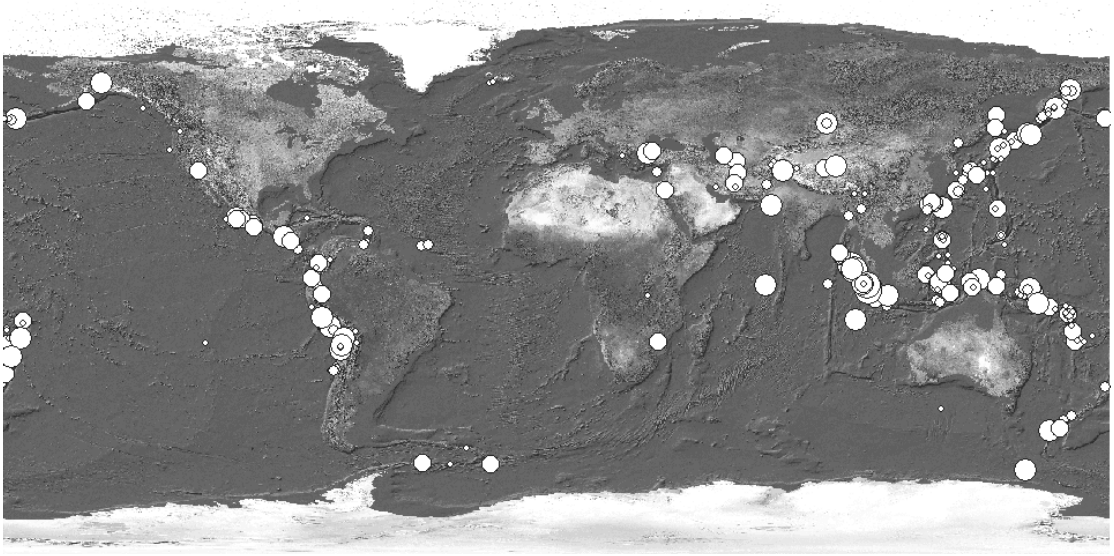
This paper only discusses the tsunami generated by earthquakes that occurred under the seabed and coast, and the tsunamis generated by nuclear explosions, volcano, and hurricane are not included. Tsunami is caused by under seabed earthquake or coastal earthquake with great magnitude (usually with magnitude above 6.7). However, whether tsunami occurs also depends on the type of faults in the place where earthquake happens. The place where tsunami occurs has the faults with type of slip faults. The more the vertical displacement of slip faults moves and the faster speed the movement has, the bigger the tsunami is. In addition, the fact that affects the generation of tsunami also includes the depth of the source of earthquake, and usually when an earthquake with magnitude above 6.5 happens at a depth inner 25 kilometers or an earthquake with magnitude above 7.5 happens at a depth inner 40 kilometers, the earthquake can trigger tsunami. Moreover, the water depth of the place where earthquake happens can play an important part in deciding whether an earthquake may generate tsunami, for example, when earthquake happens, if the place where earthquake happens has water depth about 200 meters, this earthquake can generate tsunami which is destructive, and if the water depth is more than 1000 meters, a hazardous tsunami may happen.

Type of Tsunami

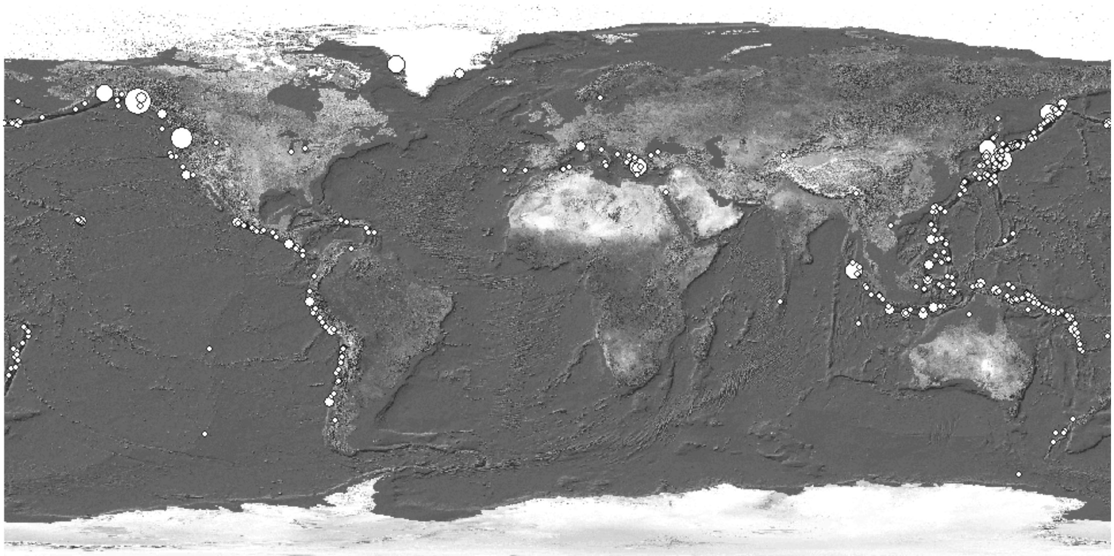
The type of tsunami can be classified by the type of vertical displacement made by the earthquake: type of uplift, type of down dropping and type of impact. The phenomenon on the coast before the tsunami of type of uplift comes is the abnormal high tidal wave first and abnormal low tidal wave later. And the phenomenon on the coast before the tsunami of type of uplift comes is the abnormal low tidal wave first and abnormal high tidal wave later. And the phenomenon on the coast before the tsunami of type of impact comes on the coast is the big wave caused by the coastal landslides caused by the earthquake. Also, the type of tsunami can be classified by the distance between the coast and the place where the earthquake happens: type of local and type of ocean wide. It is relative easy to forecast the tsunami of type of ocean wide because this type of tsunami often takes a lot of time to travel thousands of kilometers before reaching the land. On the other hand, it is very difficult to forecast the local tsunami because it reaches the land too soon for people to react.

Geological Distribution

Tsunami is triggered by marine earthquake, and the distribution of the places where tsunami happened is highly relative to the distribution of the places where earthquake happened. Picture 1 shows the distribution of earthquake with magnitude above 7.0 since the year 1950. and Picture 2 shows the distribution of tsunami since the year 1950.



Picture 1 Distribution of the earthquake with magnitude above 7.0 since 1950



Picture 2 Distribution of the tsunami since 1950

2. DATAMINING TECHNOLOGY

2.1 Artificial Neural Network

Artificial neural network (ANN) which simulates human neural system by a huge mount of information processing unit is a kind of simple human brain mathematic modal. ANN can complete study, recognization, memorization like human do, by non-linear dynamic system constructed by information processing unit named neuron.

2.1.1 Learning process of artificial neural network

The learning process of artificial neural network is to train the linking weights of the network by the training dataset which includes input data and output data in order to make the output trained data error smaller than pre-set value. Diagram 1 shows the learning process of artificial neural network.

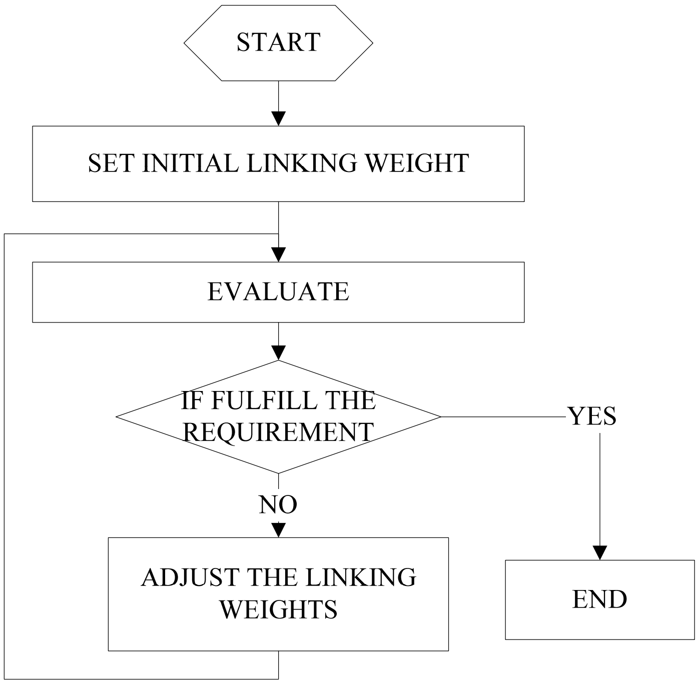


Diagram 1 Learning process of ANN

2.1.2 BP neural network

BP network is a kind of artificial neural network which uses back propagation algorithm. Because of the good ability of self-learning and self-adaptation, BP network has been widely used. And BP network can also be used in the identification of tsunami and the forecasting of tsunami.

2.1.2.1 Structure of BP neural network

The modal of BP network has three kinds of layers: Input layer, Hidden layer and Output layer. All the neurons in different layers are completely connected, and there is no connection between the neurons which are in the same layer. R.H.Nielson has certified that any continuous function in a close area can be approached by a BP network with one hidden layer. Therefore, a BP network with three layers can make the mapping from n-dimension to m-dimension. Diagram 2 shows the architecture of a 3-layer BP network.

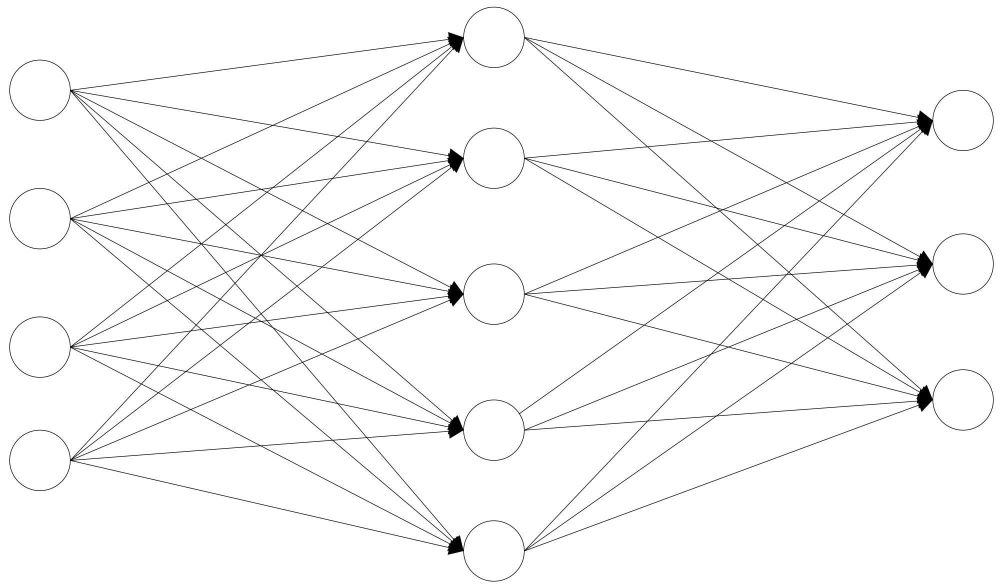


Diagram 2 Architecture of a 3-layer BP network

In a BP network, the neurons in hidden layer and output layer usually use sigmoid function, and the sigmoid function is $f(x)=\dfrac{1}{1+\exp(-x)}$. For the output of neuron j in hidden layer,

$$V_j = f(\sum_{k=0}^{K-1} W_{jk}\xi_k - \theta_j)\text{, j=0, 1,… ,J-1.} \qquad (2.1.1)$$

For the output of neuron i in output layer,

$$O_i = f(\sum_{J=0}^{J-1} W_{IJ}V_J - \theta_I)\text{, i=0,1,…,I-1.} \qquad (2.1.2)$$

θ_j in function (1),(2) means the inner range of the j th neuron in hidden layer, θ_i in function (1),(2) means the inner range of the i th neuron in output layer.

Use iterative method to process until go to the output layer. The method of adjusting linking weight follows the Delta rule.

$$W_{ij}(n+1) = W_{ij}(n) + \eta\delta_j Out_j \qquad (2.1.3)$$

$W_{ij}(n)$ is linking weight form neuron i to neuron j after n times training, Out_j is the output of neuron j, η is learning rate. δ_j is the different value of neuron j. The network will be well trained if the error of the output is smaller than a pre-set value after many times training.

2.2 Principle Component Analysis

Principle component analysis (PCA) is a kind of statistical methods which uses less comprehensive indexes to substitute for more indexes. The goal of PCA is to make a mapping from more dimensions to fewer dimensions, and maintain some construction in state of more dimensions at the same time.

2.2.1 Computing process of principle component analysis

(1) The process of standardization is for the data which is complex. And the process of standardization can deduce the negative effect to data analysis made by dimension and the order of magnitude of the data. However, this kind of process will bring some negative effect to the analytical result.

(2) Get Covariance of the matrix of samples of data or related matrix of standardized matrix X^*.

$$R = \frac{X^{*T} X^*}{n - 1} \tag{2.2.1}$$

(3) Get the non-negative real number solution according to $|R - \lambda I_p| = 0$ or $|V - \lambda I_p| = 0$.

(4) Calculate the contribution rate and accumulated contribution rate. If the accumulated contribution rate of the fore principle components has reached 85%, then these principle components can well represent the whole dataset.

(5) The relation between the k th principle component Z_k and initial indexes is:

$$Z_k = \beta_{k1} X_1 + \beta_{k2} X_2 + \cdots + \beta_{kp} X_p \tag{2.2.2}$$

And $\beta_k = (\beta_{k1}, \beta_{k2}, \cdots \beta_{kp})^T$ is the unit eigenvector which belongs to eigenvalue λ_k, X_j is the sample data or the index after the standardization of sample data.

2.3 Cluster Analysis

Cluster analysis is to catalog a certain amount of data by the likeness of the characteristic of data itself. Generally speaking, there are two ways to define the likeness: first, every data can be defined as a point in a space with m dimensions, and the likeness of two pieces of data is defined by the distance of the two points in the space; the other, the likeness is defined by some similar coefficient getting from the characteristics of the sample data. Cluster analyses classify sample data by the likeness of the sample data.

3. APPLICATION OF DATAMING TECHNOLOGY IN TSUNAMI FORECASTING

Data mining technology can be used to identify, recognize and forecast tsunami according to the own properties of tsunami. Diagram3 shows the system structure which can be used to forecast tsunami.

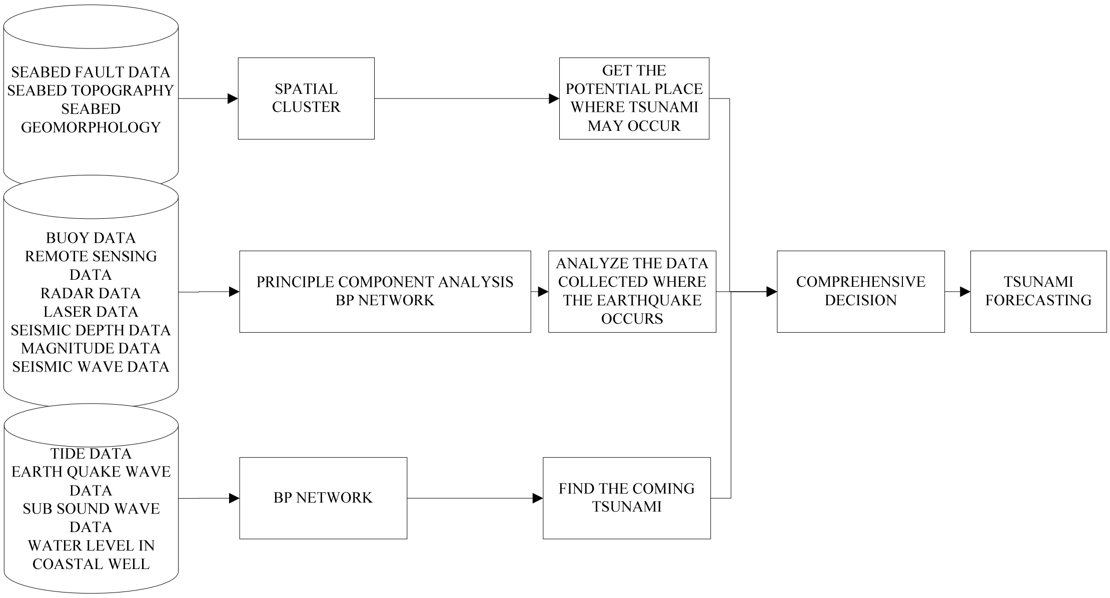


Diagram 3 System structure of tsunami forecasting

3.1 Classification of the places where tsunami occur most likely

The monitoring data can be categorized by processing the factor data as seabed fault information, seabed topography information, seabed geomorphology information, water depth information using spatial cluster technology. And adjust the result of categorization according to the historical tsunami data in order to find the kind of data with high possibility of tsunami.

3.2 Real-time analysis of far-shore monitoring data

Far-shore monitoring data means the data get from the sensors located near the place where earthquake occurs. This kind of data includes buoy data, remote sensing data, radar data, laser data, depth data, and magnitude data and so on. All the data mentioned above may become abnormal after the earthquake. Artificial neural network can be trained by related historical data to identify abnormal monitoring data.

3.2 Real-time analysis of coastal monitoring data

Coastal monitoring data means the data get from the sensors near the sea. This kind of data includes tidal data, earthquake wave data, sub sound wave data, and water level data in coastal well. As mentioned in the former part of this paper, there is abnormal high tide or abnormal low tide. There may be other abnormal phenomenon. Identifying the abnormal phenomenon on the coast can forecast the coming tsunami.

REFERENCES

http://www.ngdc.noaa.gov/nndc/struts/
Satish Kumar. (2006), Neural Networks, The McGraw-Hill Companies, Inc.

Mehmed Kantardzic. (2002), Data Mining Concepts, Models, Methods, and Algorithms, IEEE Press.