Computers & Geosciences 73 (2014) 99–107

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

A dynamic relearning neural network model for time series analysis of online marine data

Wenqing Li [a,b,*], Wenyan Wang [a,b], Xiaoyan Wang [a,b], Shixuan Liu [a,b], Liang Pei [a,b], Fadong Guo [a,b]

[a] Shandong Provincial Key Laboratory of Ocean Enviromental Monitoring Technology, China
[b] Institute of Oceanographic Instrumentation, Shandong Academy of Sciences, China

ARTICLE INFO

Article history:
Received 23 January 2014
Received in revised form
29 August 2014
Accepted 16 September 2014
Available online 28 September 2014

Keywords:
Contribution hypothesis
Evolution hypothesis
Relearning neural network
Time series
Online marine monitoring
Dynamic parameters

ABSTRACT

In order to effectively predict the time series marine data obtained from online monitoring, contribution hypothesis and evolution hypothesis are proposed. A mathematical model describing the relationship between the number of samples and the weight of relearning is developed on the basis of the two hypotheses. Conventional neural networks with static parameters are modified to have dynamic parameters, which can "evolve" repeatedly in the process of online monitoring. The algorithm flow of dynamic relearning neural networks is established, which consists of two phases, named sample training phase and dynamic computation phase. In the first phase, proper number of samples and learning weight are obtained; in the second phase, dynamic computations are carried out with conditional relearning. A linear neural network is chosen and current velocity data is selected for experiment, in all of the three chronologically selected groups, the relearning neural network outperforms the conventional static neural networks, and the mean absolute errors (MAE) of the three groups are respectively reduced 3.40 percent, 6.67 percent and 7.93 percent. Experiment results show that MAE will be reduced more and more as time goes on, which verify the contribution hypothesis and evolution hypothesis. Focusing on improving the work flow of neural networks, the proposed method could be widely applied to various types of other geographic data as well as marine monitoring data.

1. Introduction of online marine monitoring

With increasing demand for oceanographic observation, large scale multi-point marine monitoring has already appeared. Accompanied by strong wind, wave, rain, humidity, sea water corrosion and frequent passing ships, various types of marine monitoring instruments are working in harsh environments all the time, so there may be more or less error data and missing data. The quality of marine monitoring data will directly affect the subsequent data mining and data analysis, thus affecting the application activities such as marine disaster prevention, marine forecasting, marine operation, marine fisheries, and so on.

Averagely, tens of values are received from large scale multi-point marine monitoring in a minute. In order to deal with the first-hand data collected by oceanographic instruments on all kinds of platforms such as ships, buoys (Li et al., 2012) and land stations (Fig. 1), online marine monitoring system (OMMS) is established. The OMMS provides real-time data service and makes use of a variety of communication, including Very High Frequency (VHF), Code Division Multiple Access (CDMA), General Packet Radio Service (GPRS), and satellite transmission (Li et al., 2012).

Ocean current, ocean wave, water quality, wind, pressure, air temperature and CTD (conductance, water temperature, depth) are the main ocean environment parameters monitored. Each instrument on each monitoring platform obtains a time series data with uniform time intervals (10 min, 30 min, 60 min, etc), in which a number (input of the predictive model) of history values could predict the upcoming value (output of the predictive model). To avoid spending a lot of manpower and time on manual searches and evaluations, real-time diagnosis module of marine online data is designed on the online marine monitoring system to improve the efficiency and guarantee the data quality. The role of the diagnosis module is to provide a proper time series analysis method for each kind of ocean environment parameter monitored on each platform, to compare the predictive value with actual monitored value for error judgment, to correct abnormal data and to fill the missing data. The main characteristic of the diagnosis module is that it provides real-time processing for multi-source data in long-time running.

"Multi-source data" means that all kinds of the marine monitoring data come from different platforms at the same time. Each

* Corresponding author.
E-mail address: livenson@163.com (W. Li).

http://dx.doi.org/10.1016/j.cageo.2014.09.006
0098-3004/

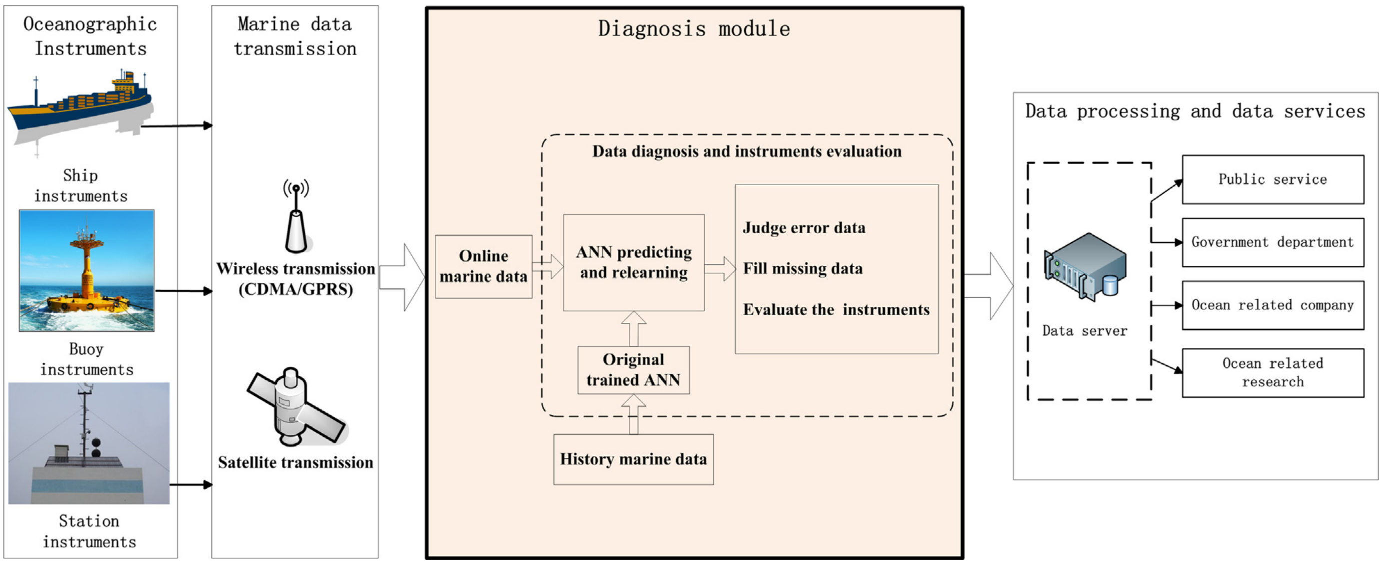

Fig. 1. Data flow of online marine monitoring system.

platform is equipped with a variety of sensors which may collect data at different time intervals, and the commonly used sensors are current sensor, wave sensor, water quality sensor, wind sensor, pressure sensor, temperature sensor, CTD sensor, etc.

“Real-time processing” means that diagnosis module must satisfy the timeliness need of large scale multi-point monitoring. Take 15 marine monitoring platforms (each equipped with 8 sensors) for example, in every 10 min, 120 online values are collected at most. That is to say, in theory the diagnosis time for each monitoring value must be controlled in 5 s on average. Nevertheless, for large scale multi-point monitoring, the number of platforms may exceed 15 and the number of sensors may exceed 8 in the near future.

“Long-time running” means that the monitoring is an on-going process. Therefore, the diagnosis module should be able to work 24 h a day, 7 days a week.

The goal of our study in this paper is to propose an efficient and useful relearning neural network model for the diagnosis of the time series marine data in the long run. For different type of marine data, the proper type of neural network and its corresponding layers, inputs should be selected.

2. Related works

Existing time series predicting methods fall into two classes: classical methods (Brooks, 2002; Box and Jenkins, 1976; Tong and Lim, 1980) based on statistical theory, and modern methods (Khashei and Bijari, 2010; Cao and Tay, 2003; Wagner and Michalewicz, 2008) based on artificial intelligence.

Of all the modern methods, artificial neural network (ANN) methods have been proved to be a good alternative to do time series forecasting. Linear neural network, BP neural network, radial basis function neural network (Qiao et al., 2014; Lee and Ko, 2009) and Elman neural network (Chandra and Zhang, 2012) are applicable to time series data analysis, the application areas involves wind power, city electricity, city water (Ghiassi et al., 2008), economy (Claveria and Torra, 2014; Yu and Huarng, 2010), biology (Panella, 2011), etc. In order to improve forecasting accuracy, many methods have been proposed to improve training speed (Zhang et al., 2007) and generalization capability (Gaxiola et al., 2014). In practical computation, these network models lack adaptability, so they could not reflect real-time changes of ocean data in the long run.

Some adaptive neural network models (Wong et al., 2010; Mirzaee, 2009) are proposed for time series forecasting and obtain good behavior and tolerance to noise, adaptive metrics or regression vector are used to get the self-adaption ability. Actually, adaptive metrics need complex computations, and the predicting results of regression vector method largely depend on training data and testing data. In order to overcome the drawbacks of static prediction models, accumulative training model and sliding window training model are proposed (Yang et al., 2005), however, the latest changes in the accumulative training data set have little impact on the model training, while the sliding window training data set may only contain recent information. An adaptive retraining mechanism (Nastac, 2010; Nastac et al., 2007) is proposed to forecast non-stationary time series on continuously changing environments, and the technique could be extended to a large period of forecasting. The disadvantage of the adaptive retraining approach lies in the fact that for each retraining phase the program may works for about one hour (Dobrescu et al., 2006), so the retraining process could not be implemented frequently, which could not meet the requirement of online marine monitoring with very frequent data.

There are many successful improvements of neural network models, but none of them are both adaptive and efficient, and they are not suitable for dealing with very frequent online data in the long run, just aiming at mining history data or online data with long intervals. A novel dynamic relearning neural network model is proposed in this study with adaptive parameters; by retraining the latest relearning sample conditionally, the parameters could be adjusted timely without complex computations.

3. Relearning neural network model

3.1. Algorithm analysis

Time series based ANN models consist of at least one hidden layer in addition to the input and output layers. The input layer contains i neurons which represent the most recent i history values already monitored (i is a variable which varies according to different data). The output layer contains one neuron from which the predicted next time value is extracted.

Due to their flexibility, ANN models lack a systematic procedure for model building (Zhang et al., 1998). Current neural network work flow mainly includes the following four steps: network