Model building and simulation for intelligent early warning of long-distance oil & gas storage and transportation pipelines based on the probabilistic neural network

Weiwei Kong, Jia Yu, Jia Yang, Tao Tian
China Petroleum Pipeline Engineering Corporation. Langfang, 065000, China

Abstract: In order to solve the potential safety hazard of long-distance oil and gas pipelines, an intelligent early morning warning model is constructed and simulated based on probabilistic neural network in this paper. The backpropagation (BP) networks and the probabilistic neural networks (PNNs) are used to process the collected abnormal data and build the early warning model. The early warning model is simulated for its accuracy in the computer and its feasibility is verified. The intelligent early warning model is built, preparing the groundwork for the subsequent generalization and application.

1. Introduction
Long distance oil and gas pipelines is the main way to store up and transport oil and gas at current stage. This mode of transportation has its own superiorities. It can not only transport large-scale oil and gas, but also has very low transportation costs compared with other modes of transportation. However, because of the distinctiveness of its stored and transported products, serious damages would be done once an accident happens [1]. Although the development of China's long-distance oil and gas pipelines started later than that of Western countries, China's oil and gas pipelines have developed very rapidly due to the vastness of China's land area. It is expected that China will have more than 150,000 kilometers of long-distance oil and gas pipelines by the end of 2020. The transportation of long-distance oil and gas pipelines is often over long distances, and they often have problems of crossing a wide area and involving many directions. At the same time, most of the long-distance oil and gas pipelines in China are in potential safety hazards status due to the weak public awareness of the protection of long-distance oil and gas pipelines. As the domestic demand for oil and gas increase rapidly, China's oil and gas storage and transportation companies are facing the problem of explosive increase in data, and most of the oil and gas storage and transportation companies are big companies. In the context of big data, today, oil and gas storage and transportation companies have to face huge data. How to find abnormal data in it and timely warning is becoming the focus of research at this stage. This experiment aims to use probabilistic neural networks to build accurate and feasible early warning models for oil and gas storage and transportation long-distance pipelines, and provide guidance for subsequent system optimization.

2. Method

2.1 Importance of building an intelligent early warning model for long-distance oil and gas pipelines
Given the special properties of oil and gas, many problems frequently arise from the long-distance transportation of oil and gas. Research shows that major oil and gas pipeline accidents in China mainly consists of construction destruction, drilling holes for oil stealing, geological disasters. The specific
causes and proportions of the accidents are shown in Table 1 [2]. These accidents often have a certain degree of change in the surrounding environment before they occur. The intelligent early warning system can detect abnormal data indicators in advance and conduct risk assessment and analysis in advance, thereby reducing or even eliminating possible dangerous events of oil and gas pipelines.

Table 1. Causes for long-distance oil and gas pipeline accidents in China

| Cause of accident | Construction damage | Drilling holes for oil stealing | Geological disaster | Equipment material | Corrosion | Improper operation | Other |
|-------------------|---------------------|---------------------------------|--------------------|-------------------|-----------|--------------------|-------|
| Percentage        | 37%                 | 26%                             | 12%                | 10%               | 8%        | 3%                 | 4%    |

2.2 BP artificial neural network

The BP (back propagation) artificial neural network, one of the most extensively-used technologies in various fields nowadays, has relatively mature technology at present, with strong nonlinear mapping ability and flexible network structure, which is suitable for dealing with nonlinear modeling, data fitting and such issues [3]. Before the BP network is used, normally it should be continued with network training to master the ability of independent calculation. The training process is shown in Figure 1 below. In the initialized BP network, according to the number of input neurons, the number of hidden layers, and the number of output layers, the learning rate and excitation function of the entire BP network are determined. The output value of the hidden layer is determined according to the input amount, and in this process, in order to ensure the training effect, the samples have to be normalized in advance. According to the hidden layer output value, weight and threshold, the errors are calculated based on the difference between the predicted output value and the expected output value. After updating the weight and threshold according to the error, whether the iterative algorithm is over can be determined. If not, repeat the above steps.
The application of BP algorithm has fairy good control over the whole data, but at the same time, BP network also has the weakness of slow convergence speed, easy to lose sight of the global only get the local optimal solution, and difficult to determine the number of hidden layers and nodes. Shown by the research of relevant scholars, BP network is often only suitable for problem prediction in practical problems, and has no advantages in other aspects [4].

2.3 Probabilistic neural network
Probabilistic neural networks (PNNs) are built based on the Bayesian optimal classification decision theory, using the nonparametric estimation method based on the probability density function existing in statistics, and then constructs a probabilistic neural network which can calculate the nonlinear discrimination boundary [5]. As a kind of forward neural network, PNN consists of four layers: input layer, hidden layer, accumulation layer and output layer. The input layer only inputs the data, and the input data does not need to be normalized in advance like BP network, which simplifies the difficulty of operation to a certain extent. The hidden layer inputs the calculated results into the accumulation layer after calculating the data, and adds the output values of the same kind of samples in the accumulation layer, which is finally reflected in the output layer.
3. Results and discussion

3.1 Building of an intelligent early warning model for long-distance oil and gas pipelines

Currently, most oil and gas storage and transportation enterprises adopt three-level system to monitor the safety of oil and gas storage system for evaluating pipelines. The system collects the temperature, pressure, and other data of the long-distance pipelines respectively and sends the data to the data monitoring center in real time through the Internet, and then the personnel of the monitoring center will monitor them and continue real-time processing of possible early warning situations. However, the data utilization of traditional real-time monitoring equipment is low, causing waste of resources easily. Therefore, combining monitoring equipment with probabilistic neural network to achieve system optimization has become the focus of research nowadays. This paper summarizes a large number of pipeline data of oil and gas long-distance transportation and storage units together, and screens the data in accordance with certain criteria. The screening criteria are as follows: 1) remove the safety data; 2) remove the abnormal data with small fluctuation; 3) remove the irregular abnormal information, screen out the abnormal data with large fluctuation and certain regularity, and use the BP neural network for preliminary fitting.

Because the total sample of data collected is large, even after screening, there are still too many data. Bringing the data directly into the BP algorithm without processing usually lead to large errors because the sample differences is too big. Therefore, this experiment unifies the data for normalization processing, and the processing formula is as follows:

\[ \text{where } \bar{x} \text{ is the mean and } \sigma \text{ is the variance.} \]

First, this study tries to build the BP topological network structure as shown in Figure 2, in which the input layer is temperature, pressure, liquid level and medium concentration, and the output layer is set to four levels according to the actual survey results, which are security, parameter exceeding threshold, primary alarm and alarm. However, after preliminary fitting, it is found that the BP network fitting can only express the abnormal trend and association between data, and cannot find the cause of abnormal data and the root cause of problems. So next, try to use the PNN neural network algorithm for the research.

![Error back propagation (learning algorithm)](image)

Figure 2. Topology of a BP network

Compared with the BP network, the PNN is not only relatively simple for network training, but also has the advantages of strong structure and high fault tolerance. Fitting in the BP network often faces both the problem of classification rules being non-transparent and the problem of sometimes only getting local optimal solution. However, the PNN classifies the samples according to the Bayesian risk minimum criterion. As long as enough training is done, the PNN will surely get the
global optimal solution. The paper below preprocesses the data same as the BP network, and then imports them into the PNN in four indicators.

3.2 Simulation experiment of the intelligent early warning model of oil and gas storage and transportation long-distance pipeline

The long-distance pipeline was taken as the monitoring object, and the abnormal data of temperature, pressure, liquid level and medium concentration were selected for the simulation experiment. 27 consecutive data were selected, and Matlab was used to draw the above four items (Figure 3). The safe range of storage temperature was -10℃ ~40℃, the safe range of pipeline pressure was 2,000 Pa ~100,000 Pa, the safe range of medium concentration was 0~6%, and the safe range of liquid level was 1.2 m ~8 m. When three or more consecutive data were detected to distribute near the control limit, the monitoring value was determined as abnormal and should be stored in the exception database. Moreover, since the safety requirement of oil and gas storage and transportation is very high, in the case that any one of the four indicators had problems, the monitoring data were called and data mining was conducted to solve the problems as soon as possible.

![Figure 3](image_url)

Figure 3 A. Temperature Detection Diagram  B. Pressure Monitoring Diagram  C. Medium Concentration Monitoring Diagram  D. High and Low Liquid Level Monitoring Diagram

Take the storage and transportation temperature as an example. The storage and transportation temperatures were divided into four levels. The first level was [1℃,30℃), indicating a safe storage state; the second level is below -10℃, indicating a possibility of parameter overrun; the third level was [30℃,40℃), meaning that there was an initial danger of storage as in the early warning stage; the fourth level was [40℃, 60℃), representing dangerous storage conditions. This paper uses the BP non-linear algorithm for fitting, and discusses two fitting models of single hidden layer and double hidden layer in two cases. The two fitting methods adopt the collected 3000 sets of data as the sample
set, among which 2900 sets are used for the fitting system to carry out preliminary training and 100 sets are used as test samples. The fitting results are shown in Figure 4 and Figure 5.

![Figure 4](image1.jpg)

Figure 4 A. Prediction Error of Single Hidden Layer Network; B. Single Hidden Layer Network Simulation Results; C. Prediction Error Percentage of Single Hidden Layer Network
The differences of the single hidden layer and the double hidden layer in the network prediction output and expected output are small, and the double hidden layer is obviously better than the single hidden layer; the prediction error of single hidden layer is in the range of (0, 0.1), while the prediction error of double hidden layer is mostly 0, only a few errors fluctuate slightly. The single hidden layer stops learning when the fitting iteration reaches to 100 times with the minimum variance being 0.00622. The double hidden layer stops learning when the fitting iteration reaches to 72 times with the minimum variance being 0.00126. In conclusion, the fitting effect of the double hidden layer network is better than that of the single hidden layer network.

This study compared the PNN and the BP network to conclude that the PNN was suitable for the accurate classification of abnormal data in this system; meanwhile, due to the PNN’s characteristics, it could not process an excessively large size of sample data. Therefore, in this experiment, only the abnormal database was used as the sample set of the PNN. Take the storage and transportation temperature of long-distance pipelines as an example. First, the data were classified according to the above-mentioned grading rules and stored in the exception only database, from which 10 groups of samples were taken for the simulation experiment in the pre-trained PNN. The simulation results are shown in Figure 6. The simulation results indicate that the fitting effect was good and there was no error, showing that the PNN was feasible and accurate in classifying data in the intelligent early warning model of long-distance oil and gas storage and transportation.

| Sample | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|--------|----|----|----|----|----|----|----|----|----|----|
| Error percentage (%) | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
The classification results given by the PNN can be used to guide the enterprises to handle the site situations in different ways. Take the storage temperature of long-distance pipelines as an example. When the data appears in the second level, such as No. 3 and No. 6, it indicates that the long-distance pipelines need to be performed with proper insulation treatment, and the platform displays as parameters overrun at this point; when the data appears in the third level, such as No. 4, No. 8 and No. 9, it indicates that the long-distance pipelines may have leakage or combustion in some parts already, so personnel shall carry out protective measures, go to check and give an alarm, and the platform displays as early warning at this point; when the data appears in the fourth level, such as No.2, No.5 and No.10, it indicates that the internal temperature and pressure of long-distance pipelines have been too high, now an explosion may occur due to chemical reactions in the pipelines, so call the police and evacuate the crowd immediately and report to the leader for emergency measures. The other three indicators are also classified and dealt with accordingly.

4. Conclusion
This paper mainly describes the construction of early warning model for long-distance oil and gas storage and transportation detection, for which the abnormal data are screened out based on the detection of early warning indicators, and the BP network and PNN are used to fit data and carry out the simulation experiments. The experiment results corroborate that the model has high accuracy and can play a positive role in practical applications. However, this experiment is only a simulation one, and the specific problems in practice remain to be solved.

References
[1] Oliva Teles L, Fernandes M, Amorim J, et al. (2015) Video-tracking of zebrafish (Danio rerio) as a biological early warning system using two distinct artificial neural networks: Probabilistic neural network (PNN) and self-organizing map (SOM). Aquatic Toxicology 165: 241-248.
[2] Zhong F, Hu D, Ren Y, et al. (2017) Application and effect evaluation of production informatization in oil and gas fields. Natural Gas Industry 37(11): 131-139.
[3] Wang T, Li T, Xia Y, et al.(2017) Risk assessment and online forewarning of oil & gas storage and transportation facilities based on data mining. Procedia computer science , 112: 1945-1953.
[4] Abdul-Aziz B, Ghazi A, Ali A, et al. (2017) Real time intelligent alarm and surveillance system in heavy oil project-north Kuwait illustrated through case examples. Abu Dhabi International Petroleum Exhibition & Conference 11: 13-16.
[5] Liu JW, Li Q, Chen WR, et al. (2019) Research on Fault Diagnosis Method of PEMFC Water Management Based on Probabilistic Neural Network and Linear Discriminant Analysis. “Chinese Society for Electrical Engineering”, 039 (012): 3614-3621