Mp-Matt: a Time Series Prediction Method with Mine Gas Sensor Data

Wang Weihan*
National Institute of Standards and Technology, Peking University, Beijing, 100871, China

Abstract. Based on the Encoder-Decoder Framework, Combined with the Pooling Preprocessing and the Multi-Attention Mechanism, This Paper Proposed the Mp-Matt Model for the Prediction of Mine Gas Time Series Data. The Model Had a Good Predictive Accuracy in the Data Set of Oxygen, Carbon Monoxide and Methane in a Mine. The Accuracy of the Best Baseline Comparison Model Had Increased by 10.30%, 22.26% and 20.02%, Respectively. This Study Compared the Effects of Each Component in the Model. The Comparison Showed That the Multi-Attention Layer Had the Most Obvious Effect. The Increase of the Metric Pooling Layer Was Lower, and the Effect of Adding the Fatt Layer Was Slightly Better Than That of Adding the Hatt Layer.

1. Introduction
In Today's World, the Development of the Internet of Things is Changing with Each Passing Day. The Basic Way of Information Acquisition in the Internet of Things is Sensing Technology and Sensing Equipment. Sensing Technology Has Become Very Important for Many Smart Applications Such as Smart Cities, Smart Transportation, Smart Industries, and More. However, Due to the Characteristics of the Data Acquired by the Sensing Device, the Limitations of Traditional Data Processing Methods, and the High Requirements for Data Processing in the Internet of Things Era, the Current Commonly Used Data Processing Methods and Integration Methods Require Further Development and Updating.

In Recent Years, Research on Artificial Intelligence Has Prospered. However, It is Still At a Relatively Early Stage in the Application Research of the Internet of Things, Especially in the Processing of Sensor Data. Many Machine Learning or AI Algorithms Still Have Some Difficulties in Simultaneous Interpreting and Processing Data in Different Environments, and Are Not Balanced in Different Sensing Applications.

Based on the Smart Industrial Scene, This Study Extracted and Processed the Gas Sensing Equipment Data in the Mine Intelligent Safety Monitoring System, and Carried out the Research of Smart Sensing Technology, and Proposed the Mp-Matt Model. The Model Study Solved the Problem of Mine Gas Time Series Data Prediction by Performing Machine/Deep Learning Processing on Sensor Data, Discovering Context, Discovering Knowledge and Analyzing from Sensor Data.

Coal is an Important Source of Energy for Human Society. The Situation of Coal Resources Stock is Complex, and Most Coal Mines Need to Be Mined Manually. The Average Mining Depth of the Coal Mine is about 500 Meters, and There Are Many Coal Mines with a Depth of More Than 1200 Meters. Mine Safety is an Important Guarantee for Coal Production. Gas Explosions and Coal Mine Fires Are the Most Frequent and Destructive Safety Accidents in the Production Process of Coal Mines, Which Will Not Only Cause Economic Losses, But Also Cause Casualties. Early Warning of Coal Mine Safety Accidents is a Powerful Safeguard Measure to Prevent Coal Mine Safety Accidents [1][2].
At Present, Safety Monitoring System Has Been Established in Most Coal Mines, and Multi-Type Monitoring Sensors Are Equipped in the Key Positions of Coal Mines. through These Sensors, the Gas Concentration in Coal Mine Can Be Monitored, and the Risk Events Can Be Predicted and Warned. These Monitoring Sensors Periodically Collect Data on Gas Concentration, Temperature, Wind Speed, Oxygen Concentration, Carbon Monoxide Concentration, Etc., and Trigger Safety Alarms Based on Changes in Monitoring Data According to the Guidance of Industry Standard Specifications.

1.1 Time Series Prediction
Time series data research is an important scientific research method in various fields such as industry, economy, finance, people's livelihood, meteorology, military, and statistics. The analysis and prediction of future conditions through time series monitoring data is an important research topic in the field of time series analysis. The data recorded by the sensors in the coal mine safety monitoring system are typical time series data. Researchers can carry out model design, parameter estimation and analysis and prediction through time series analysis and prediction methods. Therefore, the time series prediction method and the monitoring data of the gas concentration in the coal mine can be combined to realize the prediction of the gas concentration index at a certain time in the future, so as to realize the early warning of the safety accidents.

Time-series data in accordance with a time-ordered set of random variables, at equal time intervals, the actual data is recorded to reflect some objective existence [3][4]. The time series can be expressed as:

$$Y = (y_1, y_2, y_3, \ldots, y_t), t \in T$$  \hspace{1cm} (1)

Can also be written as:

$$Y = f(t)$$  \hspace{1cm} (2)

Basic Tasks of time series including classification, clustering and prediction, which is relatively complex time series forecasting. Time series data influenced by various factors, volatility and complexity was not smooth, it is difficult to solve the problem of time series analysis and forecasting through traditional statistical regression method. When constructing time series analysis and prediction methods, it is necessary to fully cope with the problem of sequence volatility and stationarity, and also evaluate the accuracy of the model output analysis and prediction results.

1.2 Prediction of Mine Gas Concentration
Carry out prediction based on key gas concentration monitoring data in coal mines is very important. According to the accurate and credible forecast data, the coal mine can trigger safety warnings in advance according to the standard specifications to deal with risks early, reduce economic losses and ensure life safety. The prediction is divided into a prediction for the next time interval and a prediction for several future time intervals. The former is a common prediction task, and the latter is a sequence prediction task[5][6].

There are multiple monitoring points in the mine, which are affected by physical factors such as gas diffusion, and the data between adjacent monitoring points is spatially related. For example, P1, P2, and P3 are three adjacent monitoring points in space, and the air flow direction is P1 \rightarrow P2 \rightarrow P3, and the gas at P1 will gradually spread to the P2 and P3 monitoring points.

The air velocity, air exchange rate, temperature and other conditions in the mine will also have a correlation effect on the gas concentration data. Therefore, when analyzing the gas concentration data, it is necessary to reflect the above factors in the model and algorithm.

1.3 Main Content of Research
The main content of this study is to predict the oxygen, carbon monoxide and methane concentration values for a given span time series by monitoring the concentration of oxygen, carbon monoxide and methane in the mine. This is a complex problem of predicting sequence data through time series sequence data. Based on the Encoder-Decoder idea, combined with the pooling preprocessing and Attention mechanism, this paper proposes a new time series prediction model--MP-mATT model.

2
Through experimental comparison, the prediction accuracy of the MP-mATT model is higher than that of the existing model algorithms. The main contributions of this research include the following three points:

2. Combining the Periodic Characteristics of the Time Series Data, the Matrix Data is Preprocessed to Enhance the Influence of Periodic Correlation on the Accuracy of the Prediction Data.

3. The Bi-Lstm Algorithm is Used in the Training Method of the Encoder Layer to Enhance the Influence of the Reverse Order Correlation on the Accuracy of the Training and Data Model.

3. Added Attention mechanism to help improve the goal orientation of the data model. Attention layer is divided into two parts, one is Attention mechanism (hATT) in Bi-LSTM training process to improve data regularity concentration; the other is Attention parameter layer (fATT), which reflects the correlation of sequence data such as wind speed, ventilation volume, temperature and location, and improves the prediction accuracy through related parameters.

The performance of MP-mATT in sequence prediction is compared with the existing main algorithms based on the three tasks of oxygen, carbon monoxide and methane concentration prediction in this study. The experimental results show that the accuracy of MP-mATT is better than other algorithm models.

4. Model

4.1 Mathematical Expression of Data

Suppose there is an array:

\[ M = (1, 2, 3, \ldots, m), m \in N \]  

(3)

Where m is the ID of the record monitoring point, which represents a certain monitoring point, and N is the range of positive integer values of the sequence of the monitoring ID. Each monitoring point produces time series data on the concentration of a certain type of gas:

\[ TS = (ts_1, ts_2, ts_3, \ldots, ts_T), t \in T \]  

(4)

Where T represents the length of the time series and \( t_2 - t_1 \) can represent the time interval between two adjacent data in the time series.

The monitored gas category can be expressed as:

\[ G = (g_1 = O_2, g_2 = CO, g_3 = CH_4) \]  

(5)

That is, the types of detection include oxygen (O2), carbon monoxide (CO), and methane (CH4). Then the data set can be expressed as:

\[ d_{mg} \in D = TS \times M \times G \]  

(6)

Suppose that at a certain time point t, the sequence data with a length of time \( T' \) is predicted, and the time series data that needs to be predicted can be expressed as:

\[ D' = \left( d_{t+1}^{mg}, d_{t+2}^{mg}, d_{t+3}^{mg}, \ldots, d_{t+T'}^{mg} \right) \]  

(7)

The time series prediction task of this study can be expressed as:

\[ MP \_ mATT(D) \rightarrow D' \]  

(8)

4.2 Model Framework

The overall frame of the MP-mATT is shown in the figure 1 below.
4.3 Data Preprocessing and Normalization

Typically, when performing time series data analysis, the data must be preprocessed to achieve a stable and analyzable state, and then analyzed by statistical data methods. Such data preprocessing is passive, and sequence mathematical details are lost during data preprocessing. In the data preprocessing, the original sequence is preserved and the ratio change processing of the data is calculated. After calculation, a new input sequence is formed, and then the whole is normalized.

4.4 Matrix Pooling Layer

In the traditional time series statistical analysis, commonly used analysis indicators include comparing the rate of change of the previous time series data (the ring ratio) and comparing the rate of change of the previous period data (year-on-year). These are two very important data analysis indicators, and also reflect the correlation between the previous time series and the previous cycle data on the current data. In order to preserve and reflect the relationship between data, this study constructs a new data matrix. Assuming that a certain period of the time series is $T_c$, construct a new data matrix:

$$d_{p_{mg}} = \begin{pmatrix} d_{mg}^{t-T_c} & d_{mg}^{t-T_c-1} \\ d_{mg}^{t-1} & d_{mg}^{t} \end{pmatrix} \in P$$  \hspace{1cm} (9)

Matrix pooling is used as the pooling layer in the model to change the data of the dataset to:

$$\text{MetricPooling}(D) \rightarrow P$$  \hspace{1cm} (10)

4.5 Bidirectional Lstm Coding Layer

Conventional time series analysis is sequential, that is, data analysis is performed along the forward direction of the series to achieve prediction of future sequences by historical sequences. For known monitoring data, there is also an intrinsic relationship between the reverse sequence data that can be searched and explored. The research shows that increasing the analysis of inverse sequence can improve the accuracy of the model and reduce the influence and interference of noise on the model. Bi-RNN and
BI-LSTM have been deeply researched and applied in the fields of speech recognition and natural language processing, and the practical application effect is obvious.

The input data is expressed as:

\[ p_t \in P \]  

(11)

In the forward (reverse) LSTM, set up two memories:

\[ h_t \in H, c_t \in C \]  

(12)

Where \( h_t \) is the hidden layer short-term memory and \( c_t \) is the hidden layer long-term memory. Set three gate functions \( g_f \) to forget the gate function, \( g_m \) is the memory gate function, and \( g_o \) is the output function. The mathematical expression of the three gate functions is as follows:

\[
g_f^t = \sigma(W_f[ h_{t-1}, p_t] + b_f) \\
g_m^t = \sigma(W_m[ h_{t-1}, p_t] + b_m) \\
g_o^t = \sigma(W_o[ h_{t-1}, p_t] + b_o)
\]  

(13)

At each time point, there are long-term memory values \( c_{t-1} \) and update values \( c_t^* \). The mathematical expression of \( c_t^* \) is as follows:

\[ c_t^* = \tanh(W_s[ h_t, h_s]) \]  

(18)

Finally, the current memory data value is jointly determined and updated by combining the forgetting gate, the update gate, the upper memory cell value, and the memory cell candidate value. The data expression is as follows:

\[
c_t = g_f^t \times c_{t-1} + g_m^t \times c_t^* \\
h_t = g_o^t \times \tanh(c_t)
\]  

(15) (16)

The forward LSTM calculation result set is \( L_f \), the reverse LSTM calculation result set is \( L_r \), and the mathematical expression of the last formed Bi-LSTM calculation result set \( L \) is:

\[ L = w_f \times L_f + w_r \times L_r \]  

(17)

4.6 Multi-Attention Mechanism Layer

The Attention layer of this study is divided into two categories.

One type is the Attention layer (hATT) enhanced for Bi-LSTM-Encoder. Since the sequence data has a stable state in which the long segment is in low fluctuation, the superimposed Attention is used in the model to improve the focus of the model on the change of the data law. Adding the hidden layer \( h_s \) in Bi-LSTM, the mathematical expression of this part is expressed as follows:

\[ s_t = v^T \tanh(W_s[ h_t, h_s]) \]  

(18)

The second type of Attention layer (fATT) is to increase the influence of data correlation such as wind speed \( f_w \), air volume \( f_v \), temperature \( f_t \), and position \( f_p \). These data are time series data, and the determination of the Attention parameter in this part is done by the Bi-RNN model. To add the hidden layer \( h_f \) to the model, the mathematical expression of this part is expressed as follows:

\[ \text{Bi-RNN}(f_w, f_v, f_t, f_p) \rightarrow h_f \]  

(19)

\[ s_f = h_f^T W_f h_f \]  

(20)

Finally, the data is processed by the softmax layer and the code is output. The mathematical expression of this part is as follows:

\[ e = \text{softmax}(f_{att}(s_v, s_f)) \]  

(21)

4.7 Model Train

The decoding layer uses the conventional LSTM method to convert the data into output sequence data. Split the data set into training and test sets.

\[ D \rightarrow \text{Train}_D + \text{Test}_D \]  

(22)

Build a training unit:

\[ (\text{train}_X, \text{train}_Y) \]  

(23)

Where \( X \) is the lag factor, which represents the influence of the pre-order \( X \) data on the result value. \( Y \) is the prediction step size, which is to predict the sequence data in the future \( Y \) time intervals. Since the model uses the Seq2Seq method, \( X \) and \( Y \) do not need to be equal. The training task is to calculate
the predicted value \( \text{pred}_Y \) through \( \text{train}_X \) and calculate the total error between \( \text{pred}_Y \) and \( \text{train}_Y \). This part of the mathematical expression is as follows:

\[
\begin{align*}
\text{Task: } & \text{train}_X \rightarrow \text{pred}_Y \quad (24) \\
\text{Loss: } & \text{RMSE}(\text{pred}_Y, \text{train}_Y) \quad (25)
\end{align*}
\]

In the training model, each training subset has a capacity of 8000 data, a lag factor of 30, a prediction step size of 3, and an optimizer using the Adam algorithm.

5. Experiment and Comparison

5.1 Data Set
The experiment used the monitoring data of a large mine in August 2017, a total of 15 monitoring equipment, the data sampling frequency is 6 seconds (the interval between adjacent data is 6 seconds), a total of about 6 million data.

5.2 Baseline Model
- ARIMA model: ARIMA model (Autoregressive Integrated Moving Average model) is a classic time series predictive analysis method.
- VAR model: Vector Auto regression (VAR) is often used to predict interconnected time series systems and to analyze the dynamic effects of random disturbances on variable systems.
- BN model: Bayesian network is a kind of probability graph model is a kind of uncertainty processing model that simulates causality in human reasoning process.
- HMM model: Hidden Markov Model (HMM) is a statistical model used to describe a Markov process with implicit unknown parameters, and then use these parameters for further analysis.
- BPNN model: BPNN (Back Propagation Neural Network) is the most basic neural network. The output is forward-propagated and the error is transmitted by Back Propagation.
- RNN model: Recurrent Neural Network, RNN can use its internal memory to process input sequences of arbitrary timing, which makes it easier to handle handwriting recognition, speech recognition, etc.
- LSTM model: Long Short-Term Memory (LSTM) is a time-cycle neural network designed to solve the long-term dependence of general RNN.
- GRU model: GRU is a very good variant of the LSTM network. It is simpler and more effective than the LSTM network, so it is also a very manifold network. Since GRU is a variant of LSTM, it can also solve the long dependency problem in RNN networks.
- Seq2Seq model: Seq2Seq is a model of the Encoder-Deocder structure. The input is a sequence and the output is a sequence. Encoder turns a variable-length input sequence into a fixed-length vector, and Decoder decodes this fixed-length vector into a variable-length output sequence.

5.3 Comparison of Experimental Results
The comparison of experimental results for various tasks is shown in the table below. It can be seen that the MP-mATT proposed in this study has a significant improvement in accuracy compared with various baseline models.

| TASK | O\(_2\) RMSE | CO RMSE | CH\(_4\) RMSE |
|------|-------------|--------|-------------|
| ARIMA | 0.7514 | 1.2021 | 5.401E-03 |
| VAR | 0.6473 | 1.1313 | 4.489E-03 |
| BN | 0.5919 | 0.9893 | 2.916E-03 |
| HMM | 0.6405 | 1.2009 | 3.513E-03 |
| BPNN | 0.8093 | 1.3227 | 3.509E-03 |
| RNN | 0.7306 | 1.2334 | 3.191E-03 |
| LSTM | 0.4854 | 0.8264 | 2.722E-03 |
In the oxygen prediction task with relatively stable concentration, the prediction accuracy of deep learning methods such as LSTM and GRU is better. On this basis, the MP-mATT model improves the accuracy by 10.30%. In the prediction tasks for carbon monoxide and methane, the Seq2Seq model performed better than other models due to the greater volatility of the monitoring data. Compared with the Seq2Seq model, the accuracy of MP-mATT increased by 22.26% and 20.02%, respectively, indicating that the attention mechanism and data preprocessing play a key role.

5.4 Mp-Matt Components Comparison
In order to further study the effectiveness of each component in the MP-mATT model, we split the parts and compare the experiments.

- mATT: remove the Metic pooling layer;
- MP-Encoder-Decoder: remove the ATT layer;
- MP-sATT: remove the external factor ATT layer;
- MP-fATT: remove the internal hidden state hATT layer.

The comparison is as follows:

| TASK   | O2       | CO       | CH4      |
|--------|----------|----------|----------|
| mATT   | 0.4541   | 0.5943   | 1.837E-03|
| MP-ED  | 0.4798   | 0.6237   | 1.910E-03|
| MP-sATT| 0.4675   | 0.6127   | 1.873E-03|
| MP-fATT| 0.4665   | 0.5920   | 1.865E-03|
| MP-mATT| 0.4284   | 0.5665   | 1.734E-03|

The results show that:
- mATT: lack of data preprocessing, the accuracy rate is affected by 4.91%-6.00%;
- MP-ED: lack of ATT layer, the accuracy rate is affected by 10.10%-12.00%;
- MP-sATT: lack of external factors ATT layer, the accuracy rate is affected by 7.99%-9.12%;
- MP-fATT: lack of internal ATT layer, the accuracy rate is affected by 4.50%-8.89%;

The results show that the increase in the multi-attention layer is most effective, and the increase in the effectiveness of the Metix pooling layer alone is lower. Increasing the effect of the fATT layer is slightly better than increasing the effect of the hATT layer.

6. Related Work

6.1 Time Series Prediction Based on Statistics
The time series analysis method began in the 1930s, and the mathematician Yule proposed the autoregressive model AR. Walker, proposed the sliding average model MA. These two models are the basic theories for future statistical analysis of time series. In 1970, American statisticians Box and Jenkins proposed the ARIMA model in 1970 and gradually became a widely used time series analysis model. Although ARIMA solved many time series prediction problems at that time, the ARIMA model has strict requirements on the stability and fluctuation regularity of time series data, and its application in non-stationary time series data prediction tasks is not very effective[7][8][9].

6.2 Time Series Prediction Based on Artificial Neural Network
Learning a non-linear process. The limitations of traditional statistical mathematics for non-linear problems limit the further development of data analysis and prediction. Machine learning and deep learning, represented by artificial neural network, have the ability of non-linear data processing and can solve the data analysis tasks of non-linear and non-stationary time series. Classical machine learning
methods include regression, artificial neural network and support vector machine. Artificial neural network model can be regarded as a black box. There is no explicit expression about the relationship between input and output. Through the training of a large number of data, it finds the inherent law. Compared with FNN, RNN is more suitable for sequential data processing. RNN can remember the intrinsic relationship of data in the sequence and influence propagation. In order to solve the problem of gradient disappearance, LSTM model and GRU model are proposed on the basis of RNN. In recent years, Seq2Seq model and Attention mechanism have been proposed to deal with the problem of prediction sequence and the problem of algorithm focusing on target. These are all important algorithmic models for sequential data analysis[10][11][12][13][14].

7. Conclusions and Prospects
This study proposes an MP-mATT model for processing time series prediction problems. MP-mATT is based on the Encoder-Decoder framework, combined with pooled pre-processing and Mupl-Attention mechanism, and performs well on the experimental dataset of coal mines. Among the three tasks of oxygen, carbon monoxide and methane prediction, the best baseline comparison model Accuracy has increased by 10.30%, 22.26% and 20.02%, respectively, and can be applied to predictive analysis research in industrial production. In this study, the effects of various components in the model were compared. The comparison showed that the multi-attention layer had the most obvious effect. The increase of the Metic pooling layer was lower, and the effect of increasing the fATT layer was slightly better than that of increasing the hATT layer.

Through this research, we can conclude that better data preprocessing, mining internal relationships, and external factors are three important directions for improving data prediction. Compared with the traditional model, the artificial intelligence model has more obvious results. In time series data analysis tasks, prediction tasks are relatively complex. In addition, artificial intelligence methods can play a better role in areas such as noise removal and correlation analysis.

References
[1] Hu X Z, Wang J K, Li X Z, et al. An Analysis of the Existing Problems of China's Coal Mine Safety Monitoring and Control System[J]. Applied Mechanics & Materials, 2014, 644-650(12):1089-1092.
[2] Ji-Ping S . Mine Safety Monitoring and Control Technology and System[J]. Coal Science and Technology, 2010.
[3] Ding H, Trajcevski G, Scheuermann P, et al. Querying and mining of time series data: experimental comparison of representations and distance measures[J]. Proceedings of the Vldb Endowment, 2008, 1(2):1542-1552.
[4] Chris Chatfield, Andreas S. Weigend. Time series prediction: Forecasting the future and understanding the past : Neil A. Gershenfeld and Andreas S. Weigend, 1994, ‘The future of time series’, in: A.S. Weigend and N.A.
[5] Gershenfeld, eds. (Addison-Wesley, Reading, MA), 1-70[J]. International Journal of Forecasting, 1994, 10(1):161-163.
[6] Thissen U , Brakel R V , Weijer A P D , et al. Using support vector machines for time series prediction[J]. Chemometrics and Intelligent Laboratory Systems, 2003, 69(1-2):35-49.
[7] Martinetz T M, Berkovich S G, Schulten K J. ‘Neural-gas' network for vector quantization and its application to time-series prediction[J]. IEEE Trans Neural Netw, 1993, 4(4):558-569.
[8] Alkan A, Yilmaz A S. Frequency domain analysis of power system transients using Welch and Yule–Walker AR methods[J]. Energy Conversion & Management, 2007, 48(7):2129-2135.
[9] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50(1):159-175.
[10] Ewinga B T, Kruseb J B, Schroederc J L, et al. Time series analysis of wind speed using VAR and the generalized impulse response technique[J]. Journal of Wind Engineering & Industrial Aerodynamics, 2007, 95(3):209-219.
[11] Yang Y, Yang H. Complex network-based time series analysis[J]. Physica A: Statistical Mechanics and its Applications, 2008, 387(5-6):1381-1386.
[12] Hsieh, William W. Nonlinear multivariate and time series analysis by neural network methods[J]. Reviews of Geophysics, 2004, 42(1):RG1003.
[13] Bowman S R, Vilnis L, Vinyals O, et al. Generating Sentences from a Continuous Space[J]. Computer Science, 2015.
[14] Saad E W, Prokhorov D V, Nd W D. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks.[J]. IEEE Transactions on Neural Networks, 1998, 9(6):1456-70.