ISBN: 978-981-18-7950-0
2023 the 13th International Workshop on Computer Science and Engineering (WCSE 2023)
doi: 10.18178/wcse.2023.06.020

Multi-temporal Processing Quality Prediction Based on Graph Neural Networks and Transfer Learning

Bin Yi[1], Jun Tang[1], Wenqiang Lin[1], Xiaohua Gao[1], Bing Zhou[1], Junjun Fang[2], Yuqi Xi[2], Wenqi Li[1+]

[1] Technology Center, China Tobacco Yunnan Industrial Co., Ltd.

[2]Affiliation Yuxi Cigarette Factory, Hongta Tobacco (Group) Co., Ltd.

Abstract. The model abstracts the complex influence relationships between parameters as graph data, uses graph neural networks to calculate the spatial information between parameters, and uses long and short term memory networks to model the complex temporal dependencies of workshop processing quality index sequences. Experimental results show that the model was achieve absolute performance improvements of 0.011, 0.001 and 2.35% compared to time series analysis methods.

Keywords: quality prediction; graph neural networks; deep neural networks; recurrent neural networks; deep learning

1. Introduction

With the continuous upgrading of a range of advanced technologies such as the Internet of Things, cloud computing and the Internet and their widespread use in the manufacturing sector, manufacturing companies have begun to transform into an intelligent manufacturing model. This new way of working, with digitalisation, networking and intelligence as its main features, enables the integration and automation of production processes [1], which has become an inevitable trend. As a core component of the transformation of manufacturing companies into intelligent manufacturing models, the workshop is an essential part of improving production efficiency and product quality [2]. The production process data derived from the workshop is a true reflection of the workshop manufacturing process and is also a fundamental element for optimising quality control in modern enterprises. Therefore, in the data-driven operation concept, the focus on mining meaningful information and mining key information from the shop floor production process data to guide the shop floor operation optimisation has attracted widespread attention from academia and industry [3].

At present, the production process on the shop floor is mainly based on setting up checkpoints at each production stage to ensure the manufacturing quality of the end processes. However, this approach brings problems, such as increased product time costs, higher staff skill requirements and greater difficulty in equipment inspection. Therefore, studying the impact of different process parameters on product quality in smart shops and designing accurate prediction models for product quality play an important role in improving shop floor productivity and supporting intelligent analysis and regulation in the shop floor [3].

Typically, industrial data is collected by industrial sensors at fixed time intervals, so that production data at adjacent moments in time form interdependent relationships. For the challenges in industrial data prediction, existing prediction methods have been studied mainly from a time-series prediction perspective. Initially, support vector machines showed good performance in time series data prediction, leading to some applications in areas such as industrial data prediction [4]. Later, with the development of neural networks, recurrent neural network models in the field of natural language processing were considered to be more accurate in solving time series prediction problems. In order to solve the problem of engineering system life prediction, the literature [5] proposed to use LSTM techniques to extract inter-frame information from aero-

[+] Corresponding author. Tel.: + 86 13708672966
E-mail address: 9466559@qq.com

engine health monitoring data and use it to fit aero-engine physical degradation data, and this method accurately predicted the engine This method accurately predicts the remaining service life of the engine. To address the problem of difficult temperature prediction in electric arc furnaces during ferronickel production, the paper [6] constructed a multivariate time series prediction model for 16 thermocouples in electric arc furnaces based on gated cyclic units to support the management and maintenance of special equipment in complex working environments. The literature [7] uses a dynamic time delay extractor (DTDR) to process industrial data and then feeds it into an attention-based LSTM for training and prediction to achieve prediction of high-dimensional, multivariate industrial time series data. In the literature [8], a bidirectional LSTM network is used to model industrial sensing data for the non-linear and dynamic nature of industrial production process data, and the model shows good performance in industrial data prediction.

The aforementioned studies have mainly been carried out from the perspective of time series forecasting, using temporal sequential order to predict future changes, however, the control relationships between parameters [9] have been ignored by most studies on industrial data forecasting, i.e. information provided by process parameters has not been included in the forecasting task, and related studies have shown that shop floor control relationships exhibit obvious characteristics of graph structured data [10], and in recent years, as graph Neural networks have received wide academic attention due to their better performance in processing graph data, and their concept was proposed with 2005 [11], Scarselli [12] and Micheli [13] et al. performed feature mapping and node aggregation on the data to elaborate the original graph neural network. bruna [14] et al. introduced convolutional operators into graph neural networks and proposed that weights on the graph shared convolutional neural network GCN, which improves the computational ability of graph neural networks in coping with complex graphs.

To address the complex temporal and spatial dependencies of workshop data that are difficult to analyse by traditional data analysis methods, this paper proposes a multi-temporal processing quality prediction model G_BiLSTM based on graph neural networks and migration learning, which uses graph neural networks to calculate the spatial information between parameters, and uses bidirectional long and short term memory networks to model the complex temporal dependencies of workshop processing quality index sequences, for which the proposed The model is designed with relevant pre-training and migration learning mechanisms to form an accurate prediction of shop floor process quality driven by both temporal and spatial features.

2. Model Introduction

2.1. Problem Definition

In the time dimension, this paper mainly considers the time dependence of the quality parameters themselves, i.e. their own time series feature extraction. In the spatial dimension, this paper considers the information obtained from the process parameters controlling the quality parameters according to the physical constraints, i.e. the results of the graph calculation based on the process constraints.

For the prediction with step size S, define $\bar{X}$ as the time series input to the model, corresponding to the time series data of the quality parameters to be predicted. At the same time, the process parameters and quality parameters of the workshop are constructed as a node set V, their influence relationships are constructed as a set E according to the process document, and the weight set W corresponding to E is constructed to form a workshop relationship graph $G = (V, E, W)$, initializing all elements within W to 1. Subsequently, W is updated by graph computation, defining $\hat{X}$ as the node feature of V, corresponding to the data of the process parameters to be predicted.

Therefore, the shop floor machining quality prediction task is defined as follows: the time series data X containing temporal features are input to the shop floor machining quality temporal prediction model $f(\cdot)$ to obtain the temporal features $\bar{y} = f(\bar{X})$, the process parameter data $\hat{X}$ and the process graph data $G = (V, E, W)$ are input to the graph neural network model $g(\cdot)$ to obtain the spatial features $\hat{y} = g(\hat{X})$,finally, $\bar{y}$ and $\hat{y}$ are input to the fully connected layer $\tilde{f}(\cdot)$ to achieve feature fusion and realize the prediction of process quality data Y in the future period, meanwhile, the migration learning strategy

$Tr(\cdot)$ is constructed to realize the prediction of $f(\cdot)$, $g(\cdot)$ and $\tilde{f}(\cdot)$ updates to improve the adaptability of the proposed model to new data.

2.2. Framework Analysis

The framework of multi-temporal machining process quality prediction based on graph neural network and migration learning is shown in Figure 1. influence relationship 1 is converted into the form of adjacency matrix [15], process parameter data 1 is used as node features and input to graph neural network to get spatial features, machining quality data 1 is used as input to BiLSTM to get temporal features, and the two types of features are spliced to get prediction results.1 To improve the model In order to improve the adaptability of the model, when the model accuracy is insufficient and needs to be updated, the migration learning strategy is set off to achieve the prediction of machining quality for the new data, i.e. influence relationship 2, process parameter data 2 and machining quality data 2. In the meantime, the details of the overall prediction model and the design of the migration strategy will be discussed in subsequent sections to ensure the accuracy of the prediction.

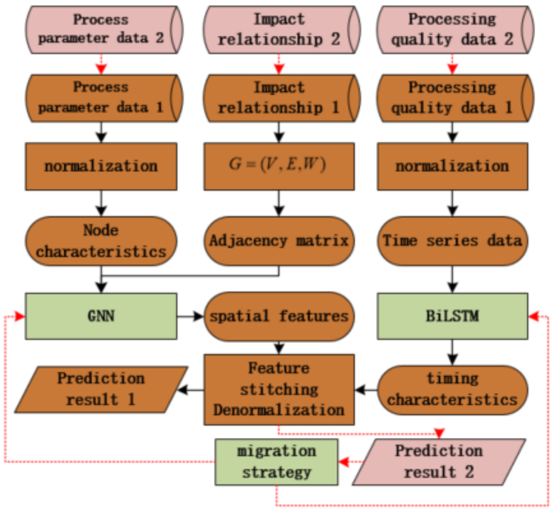


Fig 1: Framework of shop floor processing quality prediction model

2.3. Model Design

Considering the association of workshop data in time and space, the G_BiLSTM model mainly consists of two parts: graph calculation and time series calculation, which implement the calculation of workshop spatial characteristics and the calculation of the processing quality's own time series information in the time domain respectively.

(1) Graph calculation. The relationship graph $G = (V, E, W)$ of the workshop is encoded, and if node i affects node j, then there exists an edge e_{ji} to construct the adjacency matrix A. The elements a_{ij} within A are calculated as shown in equation (1).

$$a_{ij} = \begin{cases} 1, \exists e_{ji}, i \neq j \\ 0, oterwise \end{cases} \tag{1}$$

In the adjacency matrix A, the encoding of the elements corresponds to the encoding of the nodes, the positions of the elements and the corresponding values contain information about the edges, providing direction for the calculation of the parameters, while the graph neural network as shown in equation (2) is introduced.

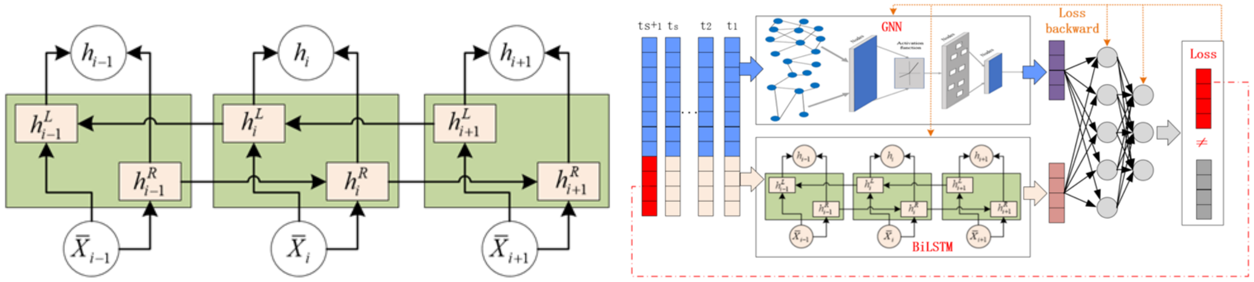
$$H' = \sigma(A \cdot W \times H) \tag{2}$$

where H is the machining quality node feature $\hat{X}$, and H' is the information after the graph calculation $\hat{y}$. After the calculation of equation (2), the process parameters are calculated for the machining quality related data $\hat{y}$ according to the connections in the relational graph G.

(2) Time series calculations. Considering that the time dependence of the quality parameter data is mainly its own, for the calculation with time step S, each time the data with dimension $M \times S$ is input to the

time series prediction model $f(\cdot)$, the data at the $S+1$st moment is output, where M is the number of quality parameters.

As shown in Fig. 2(a), BiLSTM is a two-way cyclic structure combining forward and backward in LSTM (Long Short-Term Memory) passing information from front to back and back to front respectively, i.e. for any moment i, its output is composed of two parts, h_i^L and h_i^R. h_i^L is the computation result of the left-to-right calculation approach at moment i and h_i^R is the result of the right-to-left computation at moment i. Thus, the output of BiLSTM is twice as dimensional as that of LSTM, but it can better mine the characteristics of the time series in time and predict the results more accurately than LSTM, which follows the computation as shown in equation (3).



(a) BiLSTM (b) G_BiLSTM model design

Fig2: Model structure

$$
\begin{aligned}
h_i^L &= \sigma\left(W_x^L \cdot X_t + h_{i-1}^L \cdot W_h^L + b_h^L\right) \\
h_i^R &= \sigma\left(W_x^R \cdot X_t + h_{i-1}^R \cdot W_h^R + b_h^R\right) \\
\bar{y}_i &= \sigma\left(W_y^L \cdot h_i^L + W_y^R \cdot h_i^R + b_y\right)
\end{aligned}
\tag{3}
$$

where σ is the activation function, W is the weight bias term, and $\bar{y}_i$ is the final output.

(3) Feature stitching and output. Let the output dimension of the graph computation layer be D_g, and the output dimension of the BiLSTM output layer be D_t, for the spatial feature $\hat{y}$ and the temporal feature $\bar{y}$, their output dimensions are $M \times D_g$ and $M \times D_t$ respectively, and the stitching operation is performed on these two features in the second dimension to obtain the features with output dimension $M \times [D_g + D_t]$, on which the hidden layer of dimension $[D_g + D_t] \times D_h$ and the output layer of $D_h \times D_o$ are superimposed, where D_h and D_o are the output dimensions of the hidden layer and the output layer respectively, and finally, the output feature dimension of G_BiLSTM changes from $M \times [D_g + D_t]$ in the hidden layer to $M \times D_h$, and in the output layer from $M \times D_h$ to $M \times D_o$. The predicted value y with the true value Y for the output dimension of $M \times D_o$ is computed as Loss by equation (4), and the weights of each network layer are updated in a back-propagation manner.

$$
\mathrm{Loss} = |Y - y|^2 \tag{4}
$$

2.4. Migration Mechanism of G_BiLSTM Model

In the manufacturing workshop processing, often accompanied by process adjustments, from the perspective of the process, process adjustments are divided into two main categories, one does not affect the process, such as the adjustment of process standards, such adjustments are reflected in the data as changes in values, while the original processing logic is unchanged, the other category of process adjustments is the adjustment of the production process, such adjustments change the original workshop processing logic, such adjustments are reflected in the data as changes in the relationship between the impact of parameters.

Therefore, for the G_BiLSTM model proposed in this paper, the first type of process adjustment does not involve a change in the influence relationship between parameters, and the timing of production is not altered, only the prediction accuracy of the model may be reduced, therefore, the model parameters before the process adjustment can be used as the starting point for the training of the model after the process adjustment, and while avoiding wasting arithmetic power the new high-precision prediction model can be obtained quickly. On the other hand, the second type of process adjustment changes the control relationship between the production parameters and the graph data input G in the GNN is therefore changed, whereas in the case of time series prediction, if the quality parameters are not adjusted, the temporal nature of the quality parameter data is also not changed and the parameters of the original BiLSTM model part can still be migrated, but the parameters in the GNN need to be retrained, whereas in the case where the quality parameters In the case that the quality parameters are adjusted, the parameters of the BiLSTM model part also have to be retrained, so the whole G_BiLSTM update strategy is shown in Figure 3.

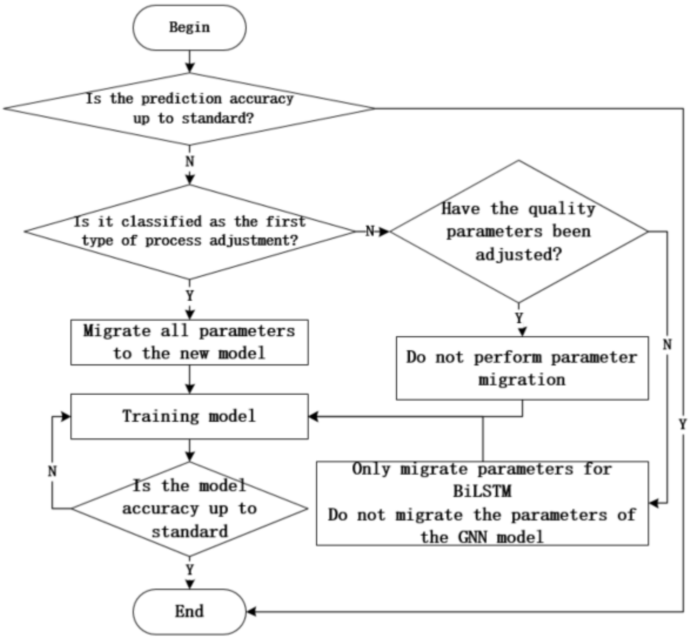


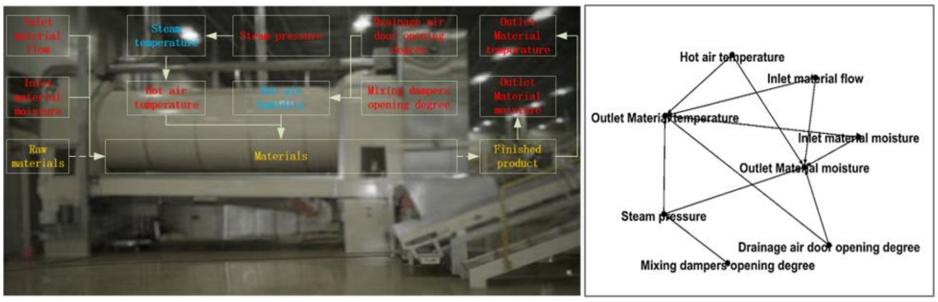
Fig. 3: Migration strategy of G_BiLSTM model

3. Experimental Analysis

In order to verify the validity of the proposed method, the production line of a Chinese company was used as the research object. This study used the production data collected in October 2022 as the data set, and after data pre-processing, a total of 11,877 rows of data were obtained, containing eight parameters, namely inlet material moisture, inlet material flow rate, hot air temperature, steam pressure, discharge damper opening, mixing damper opening, outlet material temperature and outlet material moisture. To verify the accuracy of the model, the data set was divided into a training set and a test set in the ratio of 7:3. Considering the time-series nature of the shop floor data, GRU [16], LSTM [17] and RNN [18] networks were selected as the control set and these networks were used to learn the time series features of the process parameters and the quality parameters to be predicted and to achieve the time series prediction of the two quality parameters.

The process flow of the line is shown in Figure 4(a). The plant sensors need to measure the moisture and flow rate of the inlet material, which are important indicators for monitoring the state of the material. During the production process, the material is mainly affected by the hot air temperature and moisture, which in turn are controlled by other parameters, so the hot air temperature and moisture as well as other parameters need to be collected and analysed. Ultimately, the quality of the shop floor process can be assessed by monitoring the exit material temperature and moisture of the finished product, which is one of the important indicators for process quality evaluation.

In this paper, the model code is written in Python and a neural network model is constructed using the PyTorch framework. The GRU, LSTM and RNN networks are implemented based on the torch.nn function library in the PyTorch library, respectively, and the GLSTM network is built using the Networkx library with the PyTorch library. The Intel i7-10750 processor was used for computing during the training of the models.



（a）Process relationship of workshop　(b) Process relationship graph data of silk workshop

Fig 4 Process relationship in the silk making workshop

In actual production, some parameters may not be available for acquisition, such as steam temperature and hot air temperature. As these two parameters are not available, they are not involved in the calculation. In the case of process relationship diagram data as shown in Figure 5(b), setting $step = 8$, the G_BiLSTM model will use the shop floor time series data from the first 8 time points to predict the shop floor quality indicator data from the next 1 time point. This approach is often referred to as time series forecasting and it uses past data to predict future trends and patterns.

In this paper, the Optuna [19] optimisation framework was used to perform 100 parameter search for optimisers, learning rates and the number of fully connected layer cells for four models with the objective of maximising the fitted value R. In the parameter search process, the optimizer optimization range was set to [RMSprop, Adam, SGD], the learning rate optimization range was set to [1e-02, 1e-03, 1e-04, 1e-05, 1e-06], and the number of fully connected layer cells optimization range was set to [4,8,16,32,64,128]. In the end, the best combination of parameters for the four models was obtained through parameter search. This process resulted in a better combination of parameters for the model, thus improving its performance and accuracy.

In this paper, three metrics were used to assess the validity of the models [20], namely the mean absolute error, the mean square error, and the goodness of fit R2. For an accurate assessment, 100 calculations were performed on four different models and their average results were taken for comparison. The comparison on the test set resulted in the predictive performance of the four models shown in Figure 5. In addition, the specific values of these metrics were summarized in Table 1 to better present the evaluation results of the models.

Table 1: Comparison of the error of the comparison experiments

	G_BiLSTM	GRU	LSTM	RNN
MAE（rate of water content）	6.15E-03	9.86E-03	1.35E-02	9.66E-03
MAE（temperature）	2.45E-02	4.21E-02	4.90E-02	5.35E-02
MSE（rate of water content）	5.84E-05	1.33E-04	2.42E-04	1.48E-04
MSE（temperature）	1.09E-03	2.54E-03	3.20E-03	4.31E-03
R2（rate of water content）	9.88E-01	9.72E-01	9.48E-01	9.68E-01
R2（temperature）	9.76E-01	9.45E-01	9.31E-01	9.07E-01

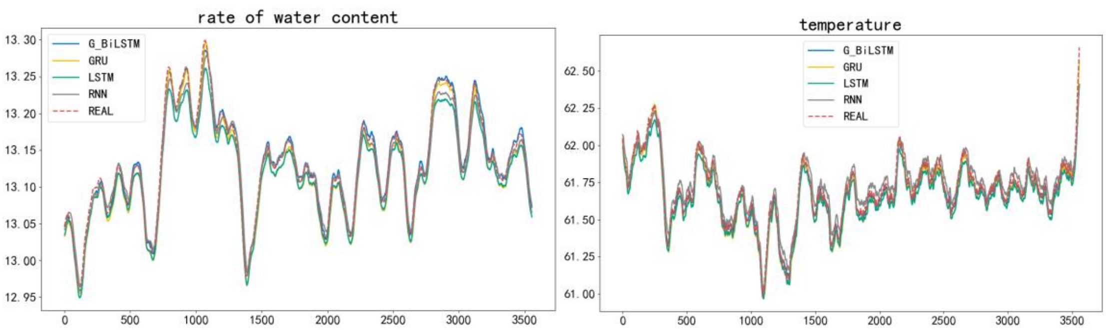


Fig. 5: Graph of predicted results of the comparison experiment

To analyze the contribution of GNN and BiLSTM in the G_BiLSTM model, two independent models were obtained by removing the feature splicing part: GNN and BiLSTM. where GNN uses only spatial information for quality prediction, while BiLSTM uses only temporal information for quality prediction, and the parameters of the models are the same as those of the G_BiLSTM model in Table 1. The average results of 100 calculations for the three models are shown in Figure 6, and the corresponding evaluation results are shown in Table 2.

The results show that G_BiLSTM outperforms the GRU, LSTM and RNN models in the three metrics of MAE, MSE and goodness-of-fit. Compared with the best-performing GRU model, G_BiLSTM showed absolute improvements of 0.011, 0.001, and 2.35% in the three metrics of MAE, MSE, and R2, respectively. The results of the ablation experiments show that the G_BiLSTM model has a relative improvement of 0.012, 0.001 and 5.60% in MAE, MSE and R2 metrics, respectively, relative to the BiLSTM model due to the inclusion of the spatial features calculated by GNN. In summary, G_BiLSTM is a better forecasting algorithm with wide application prospects in dealing with time series forecasting problems.

Table 2 Comparison of errors in ablation experiments

	GLSTM	GNN	BiLSTM
MAE（rate of water content）	**6.14E-03**	1.11E-02	1.07E-02
MAE（temperature）	**2.44E-02**	6.05E-02	4.29E-02
MSE（rate of water content）	**5.84E-05**	1.93E-04	1.45E-04
MSE（temperature）	**1.09E-03**	4.89E-03	2.69E-03
R^2（rate of water content）	**9.88E-01**	9.59E-01	9.69E-01
R^2（temperature）	**9.77E-01**	8.94E-01	9.42E-01

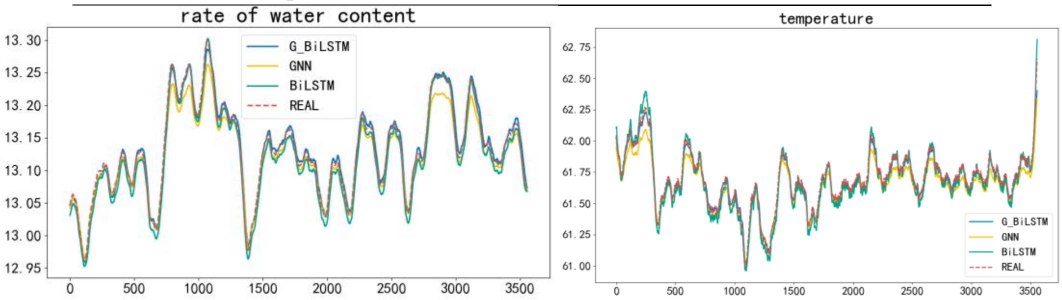


Fig. 6: Graph of predicted results of ablation experiments

The experimental results show that G_BiLSTM can more completely consider the complex spatio-temporal dependencies of shop floor data by modeling the complex relationships of shop floor metrics in space and time using GNN. For example, it is able to capture the time-series relationship of process parameters on machining quality under physical constraints. Compared with deep time series models, G_BiLSTM can better cope with the complexity of shop floor data and improve the accuracy and reliability of prediction. Therefore, G_BiLSTM has important research and application value, especially in scenarios when shop floor data needs to be considered from multiple dimensions and perspectives.

4. Conclusion

In this paper, a multi-temporal machining process quality prediction model called G_BiLSTM is proposed to address the problem that it is difficult to analyze the complex dependencies of shop floor data in time and space by traditional data analysis methods. The model is based on graphical neural networks and migration learning methods, using graphical neural networks in computing spatial information among parameters and bi-directional long- and short-term memory networks in modeling the complex dependencies of shop floor machining quality index sequences in time. In addition, relevant pre-training and migration learning mechanisms are designed to form accurate predictions of shop floor machining quality driven by both temporal and spatial features. Compared with traditional methods, G_BiLSTM can better handle the spatio-temporal complexity of shop floor data and improve the accuracy and reliability of prediction. Therefore, the model has important research and application value, especially in the scenario when shop floor data need to be considered from multiple perspectives.

5. Acknowledgements

This research was supported by Key Scientific and Technological Projects of Yunnan China Tobacco Industry Co., Ltd (grant NO. 2022GY02)

6. Reference:

[1] Yao Xifan, et al. From leagile manufacturing to long-tail production in Industry 4.0 for upgrading manufacturing [J]. Computer Integrated Manufacturing Systems,2018,24(10):2377-2387.

[2] Chen Liang, and Wu Mengting. Intelligent workshop quality data integration and visual analysis platform design [J]. Computer Integrated Manufacturing Systems,2021,27(06):1641-1649.

[3] Zhang Jie, et al. Big-data-driven operational analysis and decision-making methodology in intelligent workshop [J]. Computer Integrated Manufacturing Systems, 2016,22(05):1220-1228.

[4] Sapankevych N I, Sankar R. Time series prediction using support vector machines: a survey[J]. IEEE computational intelligence magazine, 2009, 4(2): 24-38.

[5] Wu Y, Yuan M, Dong S, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks[J]. Neurocomputing, 2018, 275: 167-179

[6] Xiang S, Qin Y, Zhu C, et al. LSTM networks based on attention ordered neurons for gear remaining life prediction[J]. ISA transactions, 2020, 106: 343-354.

[7] Li J, et al. DTDR–ALSTM: Extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models[J]. Knowledge-Based Systems, 2021, 211: 106508.

[8] Xie C, Yao R, Liu Z, et al. Process performance prediction based on spatial and temporal feature extraction through bidirectional LSTM[J]. Computer Aided Chemical Engineering , 2022, 49: 1615-1620.

[9] Sun L, Ji Y, Zhu X, et al. Process knowledge-based random forest regression for model predictive control on a nonlinear production process with multiple working conditions[J]. Advanced Engineering Informatics, 2022, 52: 101561.

[10] Yin Chao, et al. Modeling and analysis of production logistics network in discrete manufacturing workshop based on complex network theory [J]. Computer Integrated Manufacturing Systems,2020,26(08):2157-2169.

[11] Gori M , Monfardini G , Scarselli F . A new model for learning in graph domains[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.

[12] Scarselli F , Gori M , Tsoi A C , et al. The Graph Neural Network Model[J]. IEEE Transactions on Neural Networks, 2009, 20(1):61.

[13] Micheli A . Neural Network for Graphs: A Contextual Constructive Approach[J]. IEEE Transactions on Neural Networks, 2009, 20(3):498-511.

[14] Bruna J , Zaremba W , Szlam A , et al. Spectral Networks and Locally Connected Networks on Graphs[J]. Computer Science, 2013.

[15] Ghorbani M, Li X, Zangi S, et al. On the eigenvalue and energy of extended adjacency matrix[J]. Applied Mathematics and Computation, 2021, 397: 125939.

[16] Cho K , Merrienboer B V , Bahdanau D , et al. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches[J]. Computer Science, 2014.

[17] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7): 1235-1270.

[18] Lipton Z C, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning[J]. arXiv preprint arXiv:1506.00019, 2015.

[19] Akiba T, Sano S, Yanase T, et al. Optuna: A next-generation hyperparameter optimization framework[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 2623-2631.

[20] De Myttenaere A, Golden B, Le Grand B, et al. Mean absolute percentage error for regression models[J]. Neurocomputing, 2016, 192: 38-48.