X. F. TANG, Y. B. LONG

ISSN 0543-5846
METABK 63(3-4) 403-406 (2024)
UDC – UDK 621.822/621-233:621.822.7.004.64.001.4:534.63=111

INTEGRATION OF GRADIENT LEAST MEAN SQUARES IN BIDIRECTIONAL LONG SHORT-TERM (LSTM) MEMORY NETWORKS FOR METALLURGICAL BEARING BALL FAULT DIAGNOSIS

Received – Primljeno: 2024-01-08
Accepted – Prihvaćeno: 2024-02-28
Original Scientific Paper – Izvorni znanstveni rad

This paper introduces a novel diagnostic approach for bearing ball failures: a synergistic implementation of a bidirectional Long Short-Term Memory (LSTM) network, empowered by Gradient Minimum Mean Square. This method leverages deep analysis of operational data from bearings, enabling the precise identification of incipient bearing ball failures at early stages, thus markedly improving prediction accuracy. Our empirical results underscore the superior performance of this composite methodology in accurately detecting a spectrum of five mechanical bearing ball failure types, achieving a substantial enhancement in diagnostic precision.

Keywords: bearing, ball fault detection, mechanical vibration, Bi-LSTM, optimization algorithm

INTRODUCTION

In the metallurgical sector, the paramount challenge to safety stems predominantly from equipment malfunctions and human operational errors. A critical aspect of this is the failure of bearing balls, which are integral to the stability and efficiency of entire mechanical systems [1]. It is noteworthy that over half of the failures in rotating machinery have a direct correlation with bearing malfunctions. Rolling bearing failures can induce intense vibrations in equipment, leading to operational shutdowns, production stops, and even human casualties. Early detection of such faults, often subtle and intricate in their initial stages, poses a significant challenge. The burgeoning interest in bearing ball fault detection and diagnosis in recent years has foregrounded vibration signal analysis as a pivotal and effective diagnostic tool. The correlation between the extent of bearing wear and the amplitude of vibration signals is a crucial indicator of system performance. The ability to detect these vibrations without disrupting production processes presents significant cost-saving opportunities. Bearing ball vibration analysis thus plays a vital role in fault detection and the broader context of mechanical health monitoring.

The deep learning paradigm, especially in the context of bearing ball safety monitoring, has unveiled considerable potential. Deep learning algorithms facilitate the extraction of pivotal features from extensive and complex operational datasets, enabling the precise prediction of early-stage bearing ball failures. This methodology not only improves fault detection accuracy but also substantially minimizes the downtime and maintenance costs associated with such failures, offering robust and effective technological support for the metallurgical industry's safe operation.

This study conceptualizes the challenge of bearing ball fault detection in metallurgical plants as a problem of time series analysis. It leverages a Bidirectional Long Short-Term Memory (BiLSTM) network as the principal network to capture the temporal sequential features

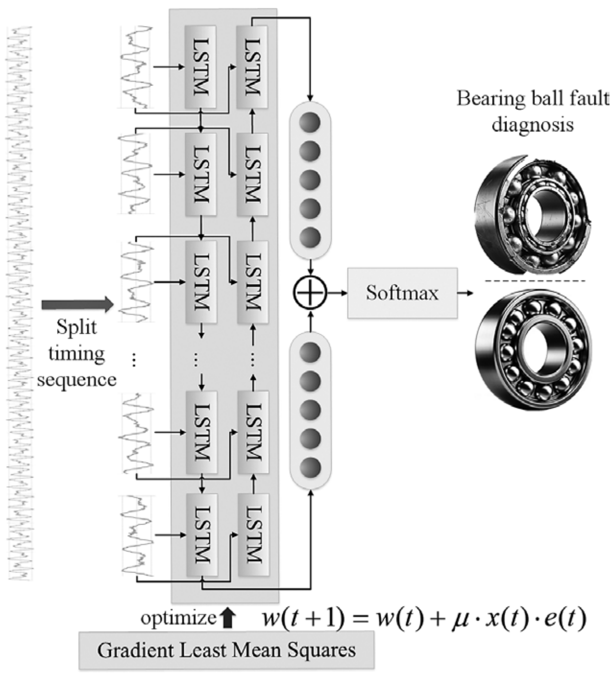

Figure 1 The structure of GLMS-BiLSTM

X. F. Tang, Y. B. Long, School of Computer Science and Software Engineering, University of Science and Technology Liaoning, China. Corresponding author: X. F. Tang (tangxiaofei@163.com)

in bearing fault scenarios. This model is further refined through the integration of one advanced machine learning optimization algorithms, which optimize the weight parameters and the optimal number of units in the sequential neural network. To validate the efficacy of the GLMS-BiLSTM model in the classification of bearing faults, an extensive evaluation is conducted on a specialized bearing fault detection dataset.

GLMS-BiLSTM MODEL

In this paper, i introduce a cutting-edge solution for detecting bearing ball faults in metallurgical facilities, embodied in a BiLSTM that synergistically incorporates GLMS, as depicted in Figure 1. This novel approach is adept at conducting an in-depth analysis of vibrational data from bearings, thereby enabling a more precise detection of even the most subtle fault signals. The integration of the GLMS algorithm significantly refines the learning process of the network, thereby boosting the efficiency and accuracy of fault prediction.

This innovative methodology not only improves the accuracy in fault detection but also serves as a powerful technological pillar for ensuring safety within the metallurgical industry. It marks a significant advancement in enhancing both the efficiency of production processes and the reliability of equipment. This approach exemplifies the effective integration of advanced machine learning techniques with practical engineering applications, setting a new standard in the field of mechanical fault diagnosis and offering a valuable tool for the maintenance and operational optimization in heavy industries.

In managing vibration data from bearings, the GLMS-BiLSTM model employs a BiLSTM framework, adept at encapsulating the nuances of time series data. This proficiency is rooted in the model's complex architecture, which is depicted in Figure 2. Central to this architecture are the input gates, forget gates, output gates, and cell states. Each of these components plays a pivotal role in the model's functionality:

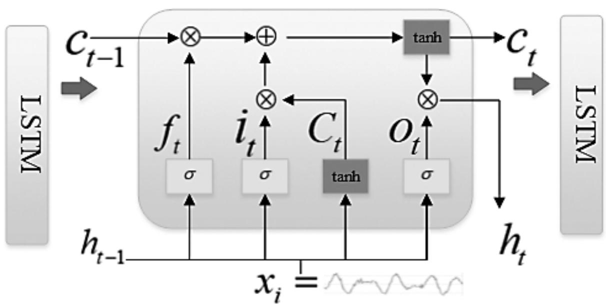

Figure 2 The network structure of LSTM

The GLMS-BiLSTM model initially employs its input gate to determine which pieces of bearing fault information need to be updated. This gate comprises two parts: a sigmoid layer that decides which values require updating, and a tanh layer that creates a new candidate value vector, which may be added to the cell state.

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \tag{1}$$

$$\tilde{C}_t = \tanh(w_{xC}x_t + w_{hC}h_{t-1} + b_C) \tag{2}$$

Next, the GLMS-BiLSTM model uses its forget gate to decide which information to discard from the cell state. This gate examines h_{t-1} and x_t through a sigmoid layer, outputting a number between 0 and 1 for each number in the cell state C_{t-1}. A value of 1 signifies 'complete retention,' while 0 indicates 'complete discard.'

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \tag{3}$$

The model then updates the cell state by combining the information from the input and forget gates, filtering, and storing important information.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Finally, the output gate of the GLMS-BiLSTM model determines the value of the next hidden state. This hidden state contains information from previous time steps and is also used for prediction.

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where σ denotes the sigmoid function, tanh represents the hyperbolic tangent function, and * signifies element-wise multiplication. W and b respectively represent weight matrices and bias vectors, x_t is the input at the current time step, h_t is the hidden state, and C_t is the cell state. Through this mechanism, GLMS-BiLSTM effectively captures both long-term and short-term dependencies in time series data, making it apt for sequential data processing tasks in bearing roller vibration data analysis.

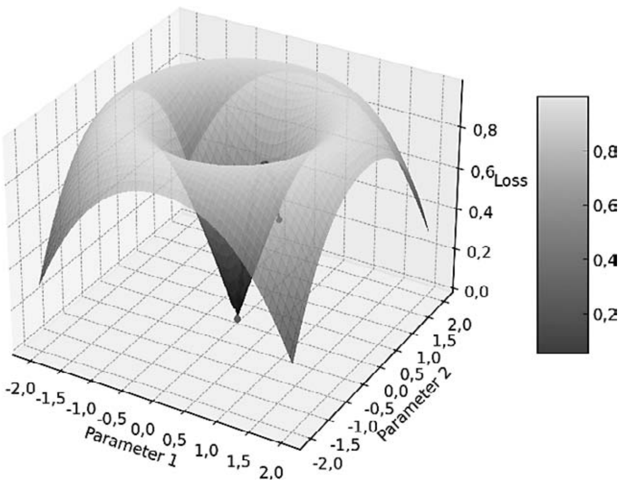

Figure 3 Gradient Least Mean Squares optimization weight process

The GLMS method, with its high robustness to noisy data, combined with the unique advantages of BiLSTM in capturing long-term dependencies in data, presents a novel solution for processing time series data with noise. This approach not only enhances the network's adaptive capability in dynamic environments but also improves the precision and stability of the predictive model. As illustrated in Figure 3, the GLMS algorithm optimizes the weights and biases of the Bi-LSTM network by iteratively minimizing the mean square of prediction errors. It employs the logic of gradient descent optimization to compute the gradient of the prediction error, subsequently adjusting network parameters based on this calculation. This effectively reduces error margins and enhances the overall predictive capacity of the model.

To elaborate, the optimization of the BiLSTM through GLMS in the GLMS-BiLSTM model typically involves two critical steps:

The first step involves computing the error between the current prediction and the actual value. This can be mathematically represented as:

$$e(t) = d(t) - w^{\mathrm{T}}(t) \cdot x(t) \tag{7}$$

where $e(t)$ is the error at time t, $d(t)$ is the actual value, $w(t)$ is the weight vector, and $x(t)$ is the input feature vector.

The second step updates the weights based on the gradient of the error, as shown in:

$$w(t+1) = w(t) + \mu \cdot x(t) \cdot e(t) \tag{8}$$

where μ is the learning rate.

These steps underpin the GLMS-BiLSTM's optimization strategy, which is pivotal in accurately calibrating the Bi-LSTM network for efficient and precise fault detection in mechanical systems. By continuously adjusting the weights in response to the calculated error gradients, the model effectively minimizes prediction errors, enhancing its reliability and accuracy in fault diagnosis. This optimization process not only reflects the sophistication of the GLMS-BiLSTM model but also underscores its suitability for complex, real-world applications in mechanical fault detection.

(a) ball with damage (b) ball without damage

Figure 4 Ball data sample

Roll ball data analysis

The dataset for the bearing ball faults, exemplified in Figure 4, is modeled after the bearing ball dataset from the Case Western Reserve University collection. We have meticulously gathered fault data from the fan end bearings, specifically those with a fault diameter of approximately 0,18 millimeters, under a sampling rate of 2kHz. The formula used for calculating the characteristic frequency of these bearing ball faults is delineated in Equation (9).

$$f_{rb} = \frac{N}{2}\frac{D}{d}[1-(\frac{d}{D}\cos\alpha)^2] \tag{9}$$

where f_{rb} represents the bearing rotation ball failure frequency, n is the number of rolling elements, d is the diameter of rolling elements, D is the diameter of bearing pitch, α is the contact Angle of rolling elements, and N is the bearing rotation frequency. Fault data of 1,2,3,4 HP motor loads and a non-fault vibration signal are used respectively, as shown in Figure 5.

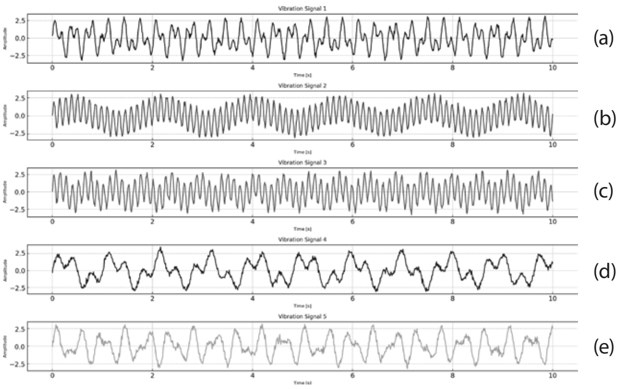

Figure 5 Vibration signal of bearing ball

where (a) is 1 HP motor loads, (b) is a non-fault vibration signal, (c) is 2 HP motor loads, (d) is 3 HP motor loads, (e) is 4 HP motor loads.

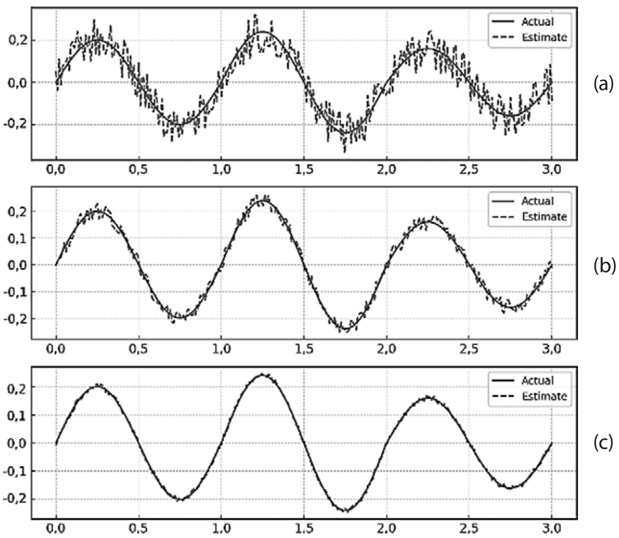

Figure 6 The ablation experiment affects the change, a represents the use of Bi-LSTM for fault bearing rotation diagnosis, b represents that GLMS-BiLSTM does not use GLMS optimization, and c represents the diagnostic performance of GLMS-BiLSTM.

In our study, we employ accuracy (Acc) as the metric to evaluate the performance of fault diagnosis models. We conducted a comparative analysis across three different models: a support vector machines (SVM) model [2], an LSTM model [3], and a Bi-LSTM model [4]. The comparative results of fault diagnosis using these models are presented in Table 1. The formula for calculating Accuracy (Acc) is as follows:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{10}$$

The experimental results, as shown in Table 1, demonstrate that the proposed GLMS-BiLSTM model achieved an average accuracy of 95,65 %, surpassing all baseline fault diagnosis methods.

Table 1 **Diagnostic results of different models**

Model	Acc(%)
SVM	85,34
LSTM	90,42
Bi-LSTM	93,18
GLMS-BiLSTM	95,65

In addition, we performed ablation studies on the GLMS-BiLSTM model to investigate its performance without the use of GLMS, as shown in Figure 6.

CONCLUSION

This paper introduces a novel method for diagnosing and detecting bearing ball faults in metallurgical applications, based on the GLMS-BiLSTM network model. This model leverages machine learning to bilaterally optimize its parameters, effectively extracting the temporal characteristics of bearing ball faults. The paper conducts comparative and ablation experiments across five different fault categories, affirming the model's effectiveness. This approach offers a new direction for enhancing safety in metallurgical plant operations.

REFERENCES

[1] Y. Hou, J. Wang, Z. Chen, et al. Diagnosisformer: An efficient rolling bearing fault diagnosis method based on improved Transformer [J]. Engineering Applications of Artificial Intelligence, (2023), 124.

[2] Z. Xu, X. Li, H. Lin, et al. Fault Diagnosis of Rolling Bearing Based on Modified Deep Metric Learning Method [J]. Shock and Vibration, (2021).

[3] Z. Lu, Y. Qin, X. Cheng, et al. Bearing Fault Diagnosis Method of Bearing Based on LSTM Auto-Encoder[C]. 5th International Conference on Electrical Engineering and Information Technologies for Rail Transportation, (2022), 582-591.

[4] T. Wang, R. Qin, H. Meng, et al. Frequency Domain Feature Extraction and Long Short-Term Memory for Rolling Bearing Fault Diagnosis[C]. 2022 International Conference on Machine Learning, Control, and Robotics, (2022), 72-77.

Note: The responsible for English language is the lector from University.