

LSTM-Based Temperature Prediction for Hot-Axles of Locomotives

Can LUO, Di YANG, Jin HUANG, Yang-Dong DENG

*School of Software, Tsinghua University, Beijing, China
luocan14@mails.tsinghua.edu.cn, yangd15@mails.tsinghua.edu.cn,
huangjin@tsinghua.edu.cn, dengyd@tsinghua.edu.cn*

Abstract: The reliability of locomotives plays a central role for the smooth operation of railway systems. Hot-axle failures are one of the most commonly found problems leading to locomotive accidents. Since the operating status of the locomotive axle bearings can be distinctly reflected by the axle temperatures, online temperature monitoring has become an essential way to detect hot-axle failures. In this work, we explore the feasibility of predict the hot-axle failures by identifying the temperature from predicted nominal values. We propose a data-driven approach based on the Long Short-Term Memory (LSTM) network to predict the sensor temperature for axle bearings. The effectiveness of the prediction model was validated with operation data collected from commercial locomotives. With a prediction accuracy is within a few percent, the proposed techniques can be used as a dynamic reference for hot-axle monitoring.

1. Introduction

With the fast-growing trend of the global industrialization and urbanization, the railway transportation plays an increasingly important role in modern society since it's the most energy-efficient and environment-friendly means of mass transportation. As the major asset of the railway transportation, locomotives have become highly complex machines involving the synergy of components from many different domains. Among various components, axle bearings are safety-critical and their failures can lead to disastrous accidents. The temperatures of the axle bearings are usually considered to be one of the most important health indicators. High-speed rotation of the bearings will cause the rise of the temperatures due to the mechanical friction. As a result, the respective failures are designated as hot-axles. Figure 1 shows an example of the accident caused by hot-axle failures.

Today the temperature based detection of hot-axle is becoming popular in the railway industry. However, there is a strong demand for predicting hot-axle failures so as to further improve the safety of locomotives and optimize the maintenance cost.

Because the temperature rising can be rather abrupt after the happening of hot-axles, an early failure warning solution is to forecast the nominal temperature given the environmental and operational information and then detect the deviation of the measured data from temperature sensors. However, it is rather a difficult task for such predictions due to the complex working environmental factors of the locomotives. In fact, the sensor data for axles are determined by a large number of factors including the operation modes of locomotives, the

geographical situation of current route, the environment, and the disturbances from various sources. The hardship in building the analytical physical model enhanced such difficulties.



Figure1. An accident caused by a hot-axle failure

Due to the criticality of hot-axles, many researchers and engineers have been dedicated to develop fast and accurate systems to monitor the temperatures of axle bearings [1] [2]. The rapid development of machine learning and artificial intelligence in recent years has provided a new perspective for the temperature prediction of the hot-axle problem. Many researchers begin to use statistical learning methods to study the early warnings signals of the hot-axle based on the historical temporal data [3] [4]. Ma et al. use the stepwise regression analysis (SRA) to predict the train axle temperature to avoid possible failures [5]. Chen et al. apply the spectral based techniques in wheel-bearings defect detection problem

[6]. Brockwell et al. adapted the ARIMA model for single time series analysis, Lütkepohl uses the Vector Auto-Regression (VAR) model for multivariate time series analysis, while both got reasonable results for the state prediction problems [7] [8].

The fast-growing machine learning techniques provide feasible and powerful solutions to improve the safety of industrial equipment in the application of Prognostics and Health Management (PHM) [9]. Under such practical needs, in this paper, we propose a temperature prediction framework for the hot-axle problem based on the Long Short-Term Memory (LSTM) networks. The data driven approach avoids the modeling of the potential physical mechanisms and focus on predicting according to the relevant historical time series data. After training with the practical running data of the locomotives, the LSTM-based model can predict the temperatures at different points of the axle rapidly with acceptable accuracy. This work shows that the LSTM network is a promising method to deal with the temperature prediction of the hot-axle problem and make an important reference for avoiding the hot-axle failures, which can further support the repair and maintenance process of the locomotives.

The paper is organized as follows. Section II describes the hot-axle problem and introduces the LSTM model for the time-series prediction. Section III proposes the prediction framework for the temperature of the hot-axles of locomotives. In Section IV, we implement the experiments and evaluate the results. Section V then follows with the conclusion of the paper.

2. Problem Formulation and the LSTM Model

2.1 The Temperature Prediction Problem

The axle bearings are the main parts that bears the loads of a locomotive, and the temperature distribution is one of the most important indicators of the state of the axle bearings. Since the temperature distribution is determined by many factors, such as the ratio between axle and wheel diameters, loading of the axle, locomotive speed, friction factor, the heat capacity, etc. [10], it is difficult to get the analytical temperature distribution.

A common solution is to deploy temperature sensors at the key points of the axle bearings. There are usually eight temperature sensors on each of the six axles of a locomotive, two of which are placed on the endpoints of the axle to gather environment temperatures and the remaining six are evenly locate at the two bearings attached to the axle with each bearing has three sensors positioned on the upper left, upper, and upper right side, respectively. The deployment is shown in Figure 2. The meanings of sensors are in Table I.

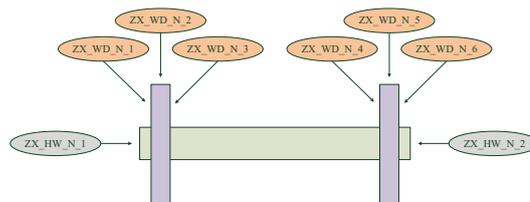


Figure2. Temperature sensors deployment on an axle bearing

In this work, we endeavor to build a LSTM model to predict the temperatures at the point 1 to 6 and compare them with the monitored temperatures by the sensors ZX_WD_N_1 to ZX_WD_N_6.

2.2 The LSTM Model

Recurrent neural networks (RNN) are feedforward neural networks with edges that span adjacent time steps introduces the time series features to the model. The recurrent edges that connect adjacent time steps may form cycles. At time step t , hidden node values $\mathbf{h}^{(t)}$ with recurrent edges receive the input from the current data point $\mathbf{x}^{(t)}$ and also from the hidden node values $\mathbf{h}^{(t-1)}$ at the previous time step. The output $\hat{\mathbf{y}}^{(t)}$ at time step t is computed according to the hidden node values $\mathbf{h}^{(t)}$. The input at the previous time step can influence the output $\hat{\mathbf{y}}^{(t)}$ at current time step by the recurrent edges. A simple RNN model is shown in Figure 3. It is because the recurrent edges that make the RNN has the ability to model the time series features. The long time steps dependence can be passed along the recurrent edges.

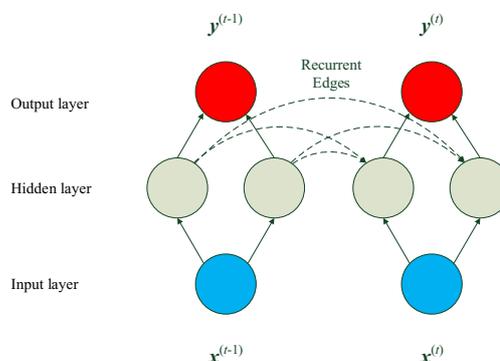


Figure3. A simple RNN model

The LSTM is one of the most successful modern RNN architectures for sequence learning tasks. The memory cell is introduced in [11] and replaces traditional nodes in the hidden layer of RNN. A memory cell is a composite unit, built from simpler nodes in a specific connectivity pattern, and the design of the unit can ensure that the gradient can pass across many time steps without vanishing or exploding.

In recent years, the LSTM networks have achieved great success dealing with sequential data in a range of different fields, including video analysis [12], information retrieval [13], natural language translation [14], and handwriting recognition [15]. Many studies have been done according to different tasks. Malhotra et

al. [16] use the stacked LSTM network to detect the anomaly points in time series. Lipton et al. [17] apply the LSTM network on the clinical time series data to solve the problem of phenotyping critical care patients. ElSaid et al. [18] predict the excess vibration events in aircraft engines with LSTM recurrent neural networks.

Although many variations have been proposed to optimize the design of the LSTM unit in view of the connectivity pattern and the activation functions since the original LSTM was introduced, all of them have explicit memory nodes for storing information for long time steps.

In this paper, we select the LSTM unit introduced by Zaremba et al. [19] for the temperature prediction of hot-axles of locomotives. This model is very successful and gain good performance on a variety of tasks, including language modeling, speech recognition, image caption generation, and machine translation. The graphical representation of the LSTM memory cell is shown in Figure 4.

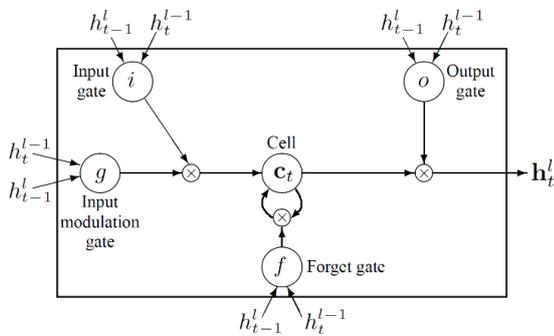


Figure 1. The graphical representation of LSTM memory cell for the temperature prediction

The $h_t^l \in R^n$ represents the state of the LSTM memory cell in the hidden layer l at time step t . Since the h_t^l is determined by h_{t-1}^l and h_{t-1}^{l-1} , this LSTM unit can be applied to deep LSTM networks with multiple hidden layers. In general, we denote the h_t^0 as the input point $x^{(t)}$ and the h_t^L as the prediction point $\hat{y}^{(t)}$, where L is the number of layers of the LSTM network.

The state transition equations of the above LSTM unit defined are as follows [19]:

$$\text{LSTM: } h_{t-1}^{l-1}, h_{t-1}^l, c_{t-1}^l \rightarrow h_t^l, c_t^l$$

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \text{tanh} \end{pmatrix} T_{2n,4n} \begin{pmatrix} D(h_{t-1}^{l-1}) \\ h_{t-1}^l \end{pmatrix}$$

$$c_t^l = f \odot c_{t-1}^l + i \odot g$$

$$h_t^l = o \odot \tanh(c_t^l)$$

Where \odot is element-wise multiplication, sigm and tanh are applied element-wisely, and D is the dropout operator.

3. LSTM-Based Temperature Prediction Framework

As shown in Figure 5, our axle temperature prediction framework is mainly composed of three parts, a data preprocessing module, a core LSTM-based prediction

module, and an assistant warning module. When the prediction module gets possible indicators for the system failure, the assistant warning module will be activated to remind the drivers for proper operations. Here we only introduce the data preprocessing and the prediction parts in details.

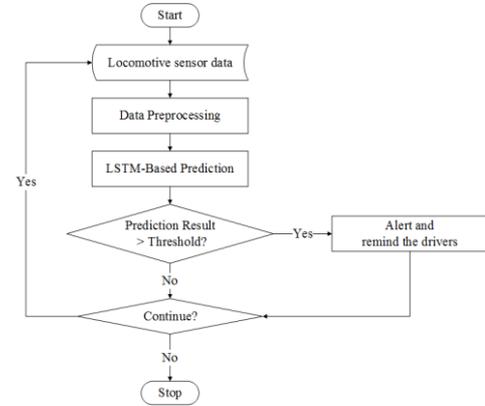


Figure 2. Overview of our prediction framework

3.1 Data Preprocessing

The temperatures at different positions of the axle are determined by many factors, e.g., the physical states of the locomotives including the running speed and the traction level, the route characteristics including the altitude and the slopes, the environmental temperature and other environmental parameters, as well as the disturbances from various sources. We select some main parameters and collect the time series data from the data sensors deployed at different positions of the locomotives. Table I lists the data sensors of relevant physical variables that we need.

Table1. Data sensors in the hot-axle prediction

| Data sensors | Description |
|--------------|--|
| ZD_LCG | The train pip pressure |
| ZD_TFG | Park cylinder pressure |
| ZD_JHG | Balance cylinder pressure |
| ZD_LLJ | The train pipe flow |
| ZD_SPEED | Locomotive's running speed |
| ZX_HW_N_1 | Left side environment temperature of axle N |
| ZX_HW_N_2 | Right side environment temperature of axle N |
| ZX_WD_N_1 | Temperature for point 1 of axle N |
| ZX_WD_N_2 | Temperature for point 2 of axle N |
| ZX_WD_N_3 | Temperature for point 3 of axle N |
| ZX_WD_N_4 | Temperature for point 4 of axle N |
| ZX_WD_N_5 | Temperature for point 5 of axle N |
| ZX_WD_N_6 | Temperature for point 6 of axle N |

The data collected by the sensors are usually with much noise, the data preprocessing step is thus needed to improve the quality of data. Here we removed the redundant data and obvious external interferences first, then a resampling process was carried out on the input

time series so that the data can be sampled at a fixed interval (10s in our experiments). During the resampling process, irregularities may present due to the multiple measurements, making the interpolation or sophisticated methods be needed to reconstruct the data. We select the linear interpolation technique since it is simple and efficient for the LSTM-based prediction in the next subsection.

3.2 LSTM Network Architecture

The input of the LSTM-based prediction model is a time series $\mathbf{x} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(T)}\}$, where the T is the input window size. Each point $\mathbf{x}^{(t)} \in R^n$ is an n -dimensional vector $\{x_1^{(t)}, x_2^{(t)}, \dots, x_n^{(t)}\}$. Here we choose n as 13 so that each element in $\mathbf{x}^{(t)}$ corresponds to one data sensor in Table I.

The output $\hat{\mathbf{y}}^{(T+k)} \in R^m$, the future temperatures to be predicted for point 1 to 6 of the axle after k time steps, where $m = 6$.

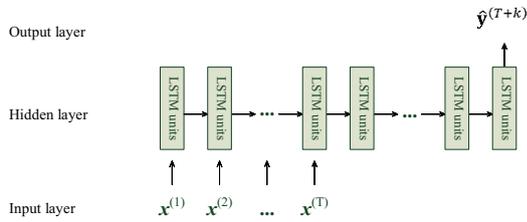


Figure 3. The architecture of the LSTM network

The LSTM network architecture is shown in Figure 6. The input layer has T input points, where each input point is 13-dimensional vector. The output layer has only one unit of a 6-dimensional vector. A hidden layer with LSTM units are connected to the input and output layer. There are 64 LSTM units in the hidden layer at each time step.

We set $k = 1, 6, 12$, so as to predict the results every 10 seconds, 1 minute, and 2 minutes, respectively. The time window size T is set to 16 empirically according to the experiment results.

4. Experiments

4.1 Experiment Design

We use Tensorflow to implement and train our LSTM networks. Tensorflow is an open-source software library developed by Google for numerical computation and is widely used in machine learning and deep learning tasks. Data flow graphs are used in Tensorflow to represent the architecture of networks. With the Tensorflow API, we can perform the computation on one or more CPUs or GPUs in the platform.

We deploy the LSTM networks on a high performance server equipped with two Intel(R) Xeon(R) E5-2609 v3 CPU and 256 GB of DDR3 main memory. Two Nvidia GTX Titan X (Pascal) GPUs are added to the motherboard as co-processor to accelerate the

computation in Tensorflow. We run the experiments on CentOS 7.0 with Tensorflow v1.0 and CUDA 8.0. The configuration of the hardware and software platform is outlined in Table II.

Table2. Hardware and software configuration

| Parameters | Description |
|--------------------|---------------------------------------|
| CPU | Intel(R) Xeon(R) E5-2609 v3 @ 1.90GHz |
| RAM | 256GB |
| GPU | NVIDIA GTX Titan X (Pascal) |
| OS | CentOS 7.0 |
| CUDA version | 8.0 |
| Tensorflow version | v1.0 |

We collect the raw sensor data from 10 locomotives of the same type running in the same route. After preprocessing, ten continuous resampling data with each has about 5500 sampling points were generated. We randomly select nine of them as the training data and the last one as the test data. We segment the continuous sampling data to coordinate with the input time window size and the time steps in the prediction and feed the segments to the LSTM networks defined in Tensorflow.

The Mean Squared Error (MSE) is selected as the cost function to train the LSTM networks as it provides a smoother optimization results. The Adam optimizer [20] was employed in training the LSTM networks as it is a simple and computationally efficient algorithm for gradient-based optimization. Additionally, the training batch size is set to be 10 and the learning rate is set to be 0.0001.

4.2 Results and Evaluation

We evaluate the max prediction error $\mathbf{e}^{(t+k)}$ on the test set where $e_i^{(t+k)} = |y_i^{(t+k)} - \hat{y}_i^{(t+k)}|, i = 1, 2, \dots, 6$. Here $y_i^{(t+k)}$ represent the true value of the temperature at the point i of the axle collect by the sensor ZD_WD_i at the time step $t + k$, while $\hat{y}_i^{(t+k)}$ represent the prediction value of the temperature at the point i of the axle output by the LSTM network according to the input time series $\{\mathbf{x}^{(t-T+1)}, \mathbf{x}^{(t-T+2)}, \dots, \mathbf{x}^{(t)}\}$

The number of iteration steps is chosen as 100000 and the corresponding training time is about 2 hours. The max prediction errors of different prediction time step on the test set are listed in Table 3

Tanle3. Prediction errors

| Axle Position | Max Prediction Errors(°C) | | |
|---------------|---------------------------|---------|----------|
| | $k = 1$ | $k = 6$ | $k = 12$ |
| Point 1 | 1.19 | 1.58 | 1.87 |
| Point 2 | 1.17 | 1.61 | 1.74 |
| Point 3 | 0.98 | 1.22 | 1.68 |
| Point 4 | 1.04 | 1.36 | 1.66 |
| Point 5 | 1.05 | 1.45 | 1.65 |
| Point 6 | 1.12 | 1.55 | 1.74 |

From the results we can see that the max prediction error does not exceed 1.2°C, 1.7°C, 1.9°C while $k = 1, 6, 12$. That means we can predict temperatures of the axle two minutes ahead and the prediction error of any points does not exceed 1.9°C.

In the practical application, the prediction values are acceptable as they serve as a reference temperature under normal operation. A significant deviation of the observed sensor data from the referenced prediction value can be considered as a potential danger and trigger a warning notification to the driver and locomotive monitoring center. Therefore, the prediction errors are acceptable and the LSTM-based prediction framework is effective in the hot-axle prediction problem.

5. Conclusions and Discussion

The temperature prediction is useful in predicting the failure of the hot-axles in the PHM system of the locomotives. This paper developed a temperature prediction framework for hot-axle problem in the locomotives with a core component of the LSTM-based model. The prediction errors are acceptable and the results provide a reference for capturing abnormal trend of the axle's temperatures. The effectiveness of the prediction model is validated on the practical running data of a locomotive. Experiment results shows the acceptable accuracy and effectiveness of the temperature prediction of the hot-axle problem in engineering practice.

Acknowledgment

The research is sponsored in part by NSFC Program (No. 61527812), National Science and Technology Major Project (No. 2016ZX01038101), MIIT IT funds (Research and application of TCN key technologies) of China, and The National Key Technology R&D Program (No. 2015BAG14 B01-02).

References

[1] R. W. Ngigi, C. Pislaru, A. Ball, and F. Gu, "Modern techniques for condition monitoring of railway vehicle dynamics", In *Journal of Physics: Conference Series*, vol. 364, no. 1, p. 012016. IOP Publishing, 2012.

- [2] S. D. Milic, "A stationary system of noncontact temperature measurement and hotbox detecting", *IEEE Transactions on Vehicular Technology* 57, no. 5 (2008): 2684-2694.
- [3] A. Amini, M. Entezami, and M. Papaalias, "Onboard detection of railway axle bearing defects using envelope analysis of high frequency acoustic emission signals", *Case Studies in Nondestructive Testing and Evaluation Volume 6, Part A*, pp. 8–16, 2016.
- [4] W. Ma, S. Tan, X. Hei, J. Zhao and G. Xie, "A Prediction Method Based on Stepwise Regression Analysis for Train Axle Temperature", *12th International Conference on Computational Intelligence and Security (CIS) IEEE Computer Society*, pp.386-390, 2016.
- [5] W. Ma, S. Tan, X. Hei, J. Zhao, and G. Xie, "A Prediction Method Based on Stepwise Regression Analysis for Train Axle Temperature", In *Computational Intelligence and Security (CIS)*, 2016 12th International Conference on, pp. 386-390. IEEE, 2016.
- [6] B. Chen, Z. Yan, and W. Chen, "Defect Detection for Wheel-Bearings with Time-Spectral Kurtosis and Entropy." *Entropy*, vol. 16, pp. 607-626. 2014.
- [7] P. J. Brockwell, and R. A. Davis, "Introduction to Time Series and Forecasting". springer, 2016.
- [8] H. Lütkepohl, "New Introduction to Multiple Time Series Analysis", Springer Science & Business Media, 2005.
- [9] J. Lee, E. Lapira, B. Bagheri and H. A. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment", *Manufacturing Letters*, vol.1, pp.38–41, 2013.
- [10] D. Chen, "Study of the Rule of Train Axle Temperature and Infrared Detection Mode of Axle Temperature", Harbin Engineering University, 2003.
- [11] S. Hochreiter, and J. Schmidhuber, "Long short-term memory", *Neural computation* 9, no. 8, 1735-1780. 1997.
- [12] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms", In *International Conference on Machine Learning*, pp. 843-852. 2015.
- [13] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward, "Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval", *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 24, no. 4, 694-707. 2016.
- [14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks", In *Advances in neural information processing systems*, pp. 3104-3112. 2014.
- [15] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, "Dropout improves recurrent neural networks for handwriting recognition", In *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th

- International Conference on, pp. 285-290. IEEE, 2014.
- [16] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series", In Proceedings, p. 89. Presses universitaires de Louvain, 2015.
- [17] Z. C. Lipton, D. C. Kale, and R. C. Wetzell, "Phenotyping of Clinical Time Series with LSTM Recurrent Neural Networks", arXiv preprint arXiv:1510.07641. 2015.
- [18] A. ElSaid, B. Wild, J. Higgins, and T. Desell, "Using LSTM recurrent neural networks to predict excess vibration events in aircraft engines", In e-Science (e-Science), 2016 IEEE 12th International Conference on, pp. 260-269. IEEE, 2016.
- [19] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization", arXiv preprint arXiv:1409.2329. 2014.
- [20] D. Kingma, and J. Ba, "Adam: A method for stochastic optimization", arXiv preprint arXiv:1412.6980 (2014).