

Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks

Aravindh.P¹, Devakumaran.R², Gobikannan.S³, Rajesh Kumar.K⁴

UG Scholar, Department of ECE, Adhiyamaan College of Engineering, Hosur, TN, India^{1,2,3}.

Assistant Professor, Department of ECE, Adhiyamaan College of Engineering, Hosur, TN, India⁴

ABSTRACT: Timely detection and identification of faults in railway track circuits are crucial for the safety and availability of railway networks. In this paper, the use of the long-short term memory (LSTM) recurrent neural network is proposed to accomplish these tasks based on the commonly available measurement signals. By considering the signals from multiple track circuits in a geographic area, faults are diagnosed from their spatial and temporal dependences. A generative model is used to show that the LSTM network can learn these dependences directly from the data. The network correctly classifies 99.7% of the test input sequences, with no false positive fault detections. The t-Distributed Stochastic Neighbor Embedding (tSNE) method is used to examine the resulting network, further showing that it has learned the relevant dependences in the data. From this comparison, we conclude that the LSTM network architecture is better suited for the railway track circuit fault detection.

KEYWORDS: Fault diagnosis, long-short-term memory (LSTM), recurrent neural network (RNN), track circuit, IR Sensor and Network simulation.

I. INTRODUCTION

RAILWAY networks are becoming busier, they are required to operate with increasing levels of availability and reliability [1]. To enable the safe operation of a railway network, it is crucial to detect the presence of trains in the sections of a railway track. The railway track circuit is worldwide the most commonly used component for train detection. To prevent accidents, the detection system is designed to be fail safe, meaning that in the case of a fault, the railway section is reported as occupied.

When this happens, trains are no longer allowed to enter the particular section. This avoids collisions, but leads to train delays. Moreover, in spite of the fail-safe design of the track circuit, there are situations in which the railway section can be incorrectly reported as free, which can potentially lead to dangerous situations. Therefore, to guarantee both safety and a high availability of the railway network, it is very important to prevent track circuit failures. This requires a preventive maintenance strategy to ensure that the components are repaired or replaced before a fault develops into a failure. To schedule the maintenance of the track circuits in the most efficient and effective manner, it is necessary to detect and identify the faults as soon as possible.

In this paper, we propose a neural network approach to fault diagnosis in railway track circuits. The fault diagnosis task comprises the detection of faulty behavior and the determination of the cause of that behavior. Since the railway track circuit network is a large network, it is not realistic to assume that additional monitoring devices will be installed on each track circuit. Therefore, this paper assumes only the availability of data that are currently measured in track circuits. By analyzing the measurement signals from several track circuits in a small area over time, the fault cause can be inferred from the spatial and temporal dependences [2]. In contrast to [2], in this paper, a data-based approach to fault diagnosis is considered, namely, an artificial recurrent neural network (RNN) called the long-short-term memory (LSTM) network [3]. Artificial neural networks have recently achieved state-of-the-art performance on a range of challenging pattern recognition tasks, such as image classification [4] and speech recognition [5]. Some of the advances made in these domains can be applied to fault diagnosis problems as well, which makes the use of neural networks an interesting option in this domain.

Learning the long-term temporal dependences that are characteristic of the faults in the track circuit case presents a challenge to standard neural networks. The LSTM network deals with this problem by introducing memory cells into the network architecture.

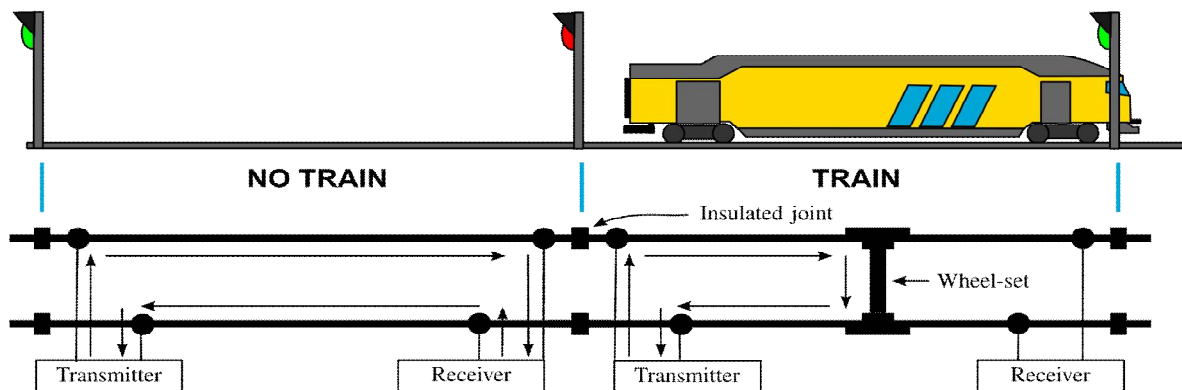


Fig 1: Current flow in a track circuit. Each track circuit detects the absence of trains in a section of a railway track. Subsequent sections are separated from other by insulated joints.

II. TRACK CIRCUITS

To enable the safe operation of a railway network, track circuits are used to detect the absence of a train in a section of railway track. Trains are only allowed to enter track sections, which the corresponding track circuit has reported to be free. A track circuit works by using the rails in a track section as conductors that connect a transmitter at one end of the section to a receiver at the other end, as shown in Fig. 1. When no train is present in the section, the transmitter will energize a relay in the receiver, which indicates that the section is free. When a train enters the section, the wheel sets of the train form a short circuit, as shown in Fig. 1. This causes the current flow through the receiver to decrease to a level, where the relay is no longer energized and the section is reported as occupied. The correct operation of a track circuit depends on the electrical current through the receiver. In the absence of a train in the section, the current must be high enough to energize the relay. Conversely, in the presence of a train, the current must be low enough, so that the relay is de-energized. To maintain the safety and availability of the railway network, it is important to detect all possible faults in the system. Moreover, to schedule preventive maintenance on the track circuits, it is important to identify the fault type and to determine the development of the fault severity over time.

A. FAULT DIAGNOSIS

Every track circuit has different electrical properties which results in different values of the high current $I_h(t)$ when no train is present, and of the low current $I_l(t)$ when a train is present. In addition, the transients between these values may be different. The current levels also depend on environmental influences and on the properties of the train passing through the section. For these reasons, it is not possible to adequately detect the presence of a fault by only considering the electrical current $I(t)$ during the passing of a single train. In this paper, we consider the current signals from several track circuits in the same geographic area, measured over a longer period of time. This makes it possible to not only detect the presence of a fault, but to also distinguish between different fault types. The reasoning behind this approach is that different faults have different spatial and temporal footprints [2].

The faults that are considered in this paper are as follows:

- 1) insulated joint defect;
- 2) conductive object (across the insulated joints);
- 3) mechanical rail defect; 4) electrical disturbance; 5) ballast degradation.

A description of these fault types, together with their spatial and temporal footprints, is given in Appendix A.

B. GENERATIVE MODEL

To enable the development, testing, and comparison of the condition monitoring methods, we have developed a generative model. This model is based on a qualitative understanding of the system and the effect of the faults considered, as well as on limited set of measurement data available from realworld track circuits. This model, together with a strategy for sampling the electrical current, is described in Appendix B.

III. NEURAL NETWORK

Artificial neural networks have achieved the state-of-the-art performance on several pattern recognition tasks. One reason for these successes is the use of a strategy called end-to-end learning. This strategy is based on moving away from hand-crafted feature detectors and manually integrating prior knowledge into the network. Instead, networks are trained to produce their end results directly from the raw input data. To use end-to-end learning, a large labeled data set is required. When this requirement is met, the benefits of a holistic learning approach tend to be larger than the benefits of explicitly using prior knowledge [20].

For the track circuit fault diagnosis case, there are currently not enough labeled data available. However, the measuring equipment that records these data has been installed. Therefore, it is reasonable to assume that at some future time, the data requirement will be met. The neural network proposed in this paper is trained and tested with synthetic data from our generative model. This enables us to analyze the opportunities of applying end-to-end learning to the track circuit fault diagnosis problem.

A. NETWORK ARCHITECTURE

The prior knowledge of the spatial and temporal fault dependences will not be explicitly integrated into the neural network. It is, however, important to give the network a structure that enables it to learn these dependences from the data.

1) *LSTM Cell*: LSTM networks are able to learn longterm time dependences by introducing specialized memory cells into the network architecture. The structure of the memory cell is shown in Fig. 2. The units a and b are the input and output units, respectively. M is the memory unit. It can remember a value through a recurrent connection with itself. The neurons denoted by g are the gate units. The input gate i determines when a new input is added to the value of the memory unit by multiplying the output of the input unit a by the output of the gate unit. In a similar way, the forget gate f determines when the value in the memory unit is kept constant and when it is reduced or reset.

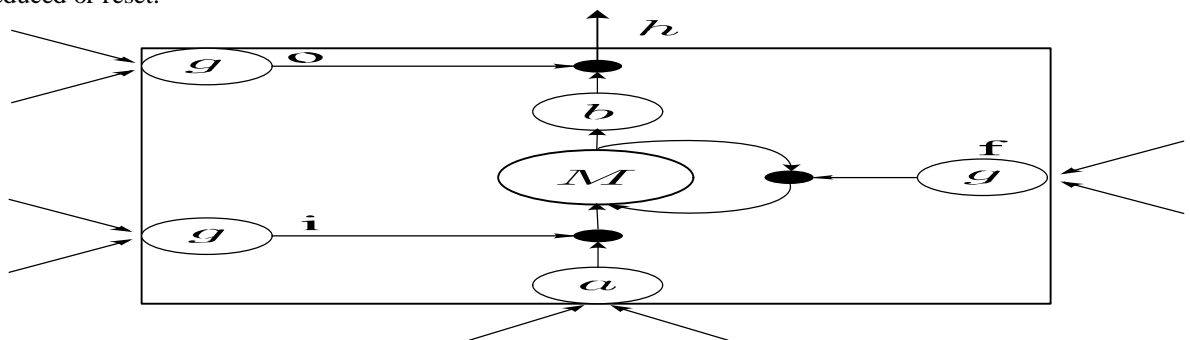


Fig. 2. Architecture of the LSTM memory cell.

Black dots: multiplication of the outputs of the gate units g by the outputs of the regular units. yield good results for this problem. Smaller networks resulted in worse performance and larger networks did not improve the performance further, while requiring significantly increased training times. In general, the ideal size of the network is based on the complexity of the problem, the amount of available training data, and the available computational resources. The inputs to each LSTM cell j in layer l consist of the inputs to the layer at that time

step $x^l(T)$, as well as the outputs of all LSTM cells in layer l at the previous time step $h^l(T-1)$. The equations that describe LSTM cell j in layer l are

$$i_j^l(T) = \text{sigm}W_{xij} x^l(T) + W_{hij} h^l(T-1) + b_{ij} \quad (1)$$

$$f_j^l(T) = \text{sigm}W_{xfjl} x^l(T) + W_{hfjl} h^l(T-1) + b_{fjl} \quad (2)$$

$$a_j^l(T) = \text{tanh}W_{xalj} x^l(T) + W_{halj} h^l(T-1) + b_{alj} \quad (3)$$

$$o_j^l(T) = \text{sigm}W_{xolj} x^l(T) + W_{holj} h^l(T-1) + b_{olj} \quad (4)$$

$$M_j^l(T) = f_j^l(T)(M_j^l(T-1)) + i_j^l(T)a_j^l(T) \quad (5)$$

$$h_j^l(T) = o_j^l(T)\text{tanh}(M_j^l(T)) \quad (6)$$

2) Inputs and Outputs: For each of the five track circuits in Fig. 3, the current magnitude is sampled four times during a train-passing event. The details of this sampling procedure are described in Appendix B. The resulting 20 current values for each train-passing event T are the inputs to the first hidden layer for that train-passing event time step:

$$x^1(T) = [I_A^1(T) \dots I_E^4(T)]$$

$$eW_h^2(T) + b_c$$

$$P(Y=c)(T) = \frac{eW_d h^1(T) + b_d}{\sum_{d=1}^6 eW_d h^1(T) + b_d} \quad (7)$$

B. NETWORK TRAINING

To train the neural network, two data sets are generated. The first one is a training data set with 21600 sequences. The second is a validation data set containing 600 sequences. For each sequence, the properties of the track circuits and the properties of the fault are stochastically determined. Each sequence has a length of 2000 train-passing events. This relates to a time period of 100 days. Note that although more trains are likely to pass through the considered sections, it is important to keep the temporal dependences from becoming too long term. Therefore, it might be necessary to limit the number of train-passing events per day that are used as network inputs. Track circuit is no longer able to function correctly. The value of 0.15 is chosen to detect the faults as early as

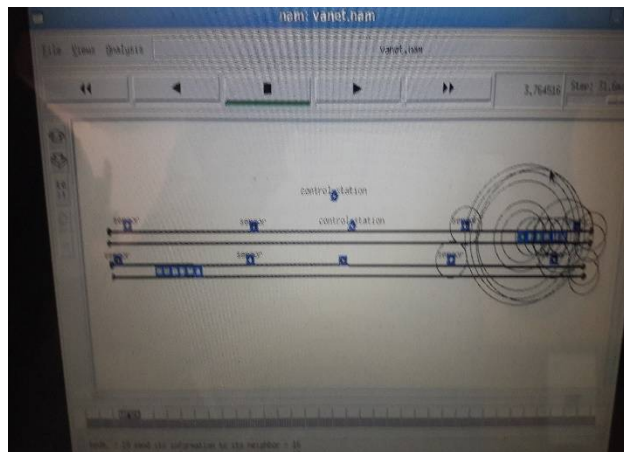
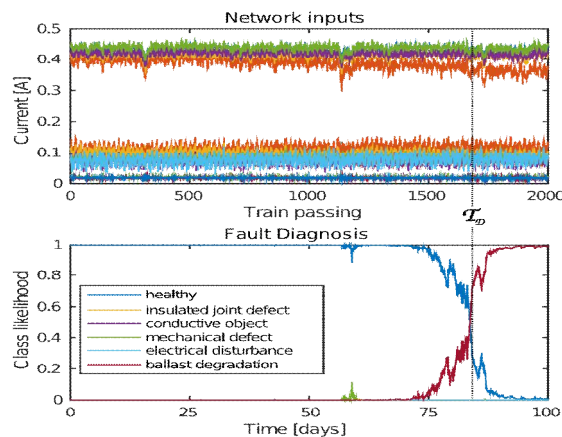
possible without having any false positive fault detections. Based on the target classifications $t(T)$, the network is trained to minimize the negative log likelihood loss function

$$l(T) = -\log(P(Y(T) = t(T))). \quad (8)$$

During the training on the training data set, the performance according to (8) on the validation data set is monitored. When this performance stops improving, the learning rate is lowered. After the training is complete, the network weights that resulted in the best performance on the validation data set are used to test the network.

IV. RESULTS AND OUTPUTS

To test the trained network, a test data set is generated containing 1500 sequences.



REFERENCES

- [1] J. Chen, C. Roberts, and P. Weston, "Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems," *Control Eng. Pract.*, vol. 16, no. 5, pp. 585–596, 2008.
- [2] K. Verbert, B. De Schutter, and R. Babuška, "Exploiting spatial and temporal dependencies to enhance fault diagnosis: Application to railway track circuits," in *Proc. Eur. Control Conf.*, Linz, Austria, Jul. 2015, pp. 3047–3052.
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [4] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun. (2015). "Deep image: Scaling up image recognition." [Online]. Available: <http://arxiv.org/abs/1501.02876>
- [5] A. Y. Hannun *et al.* (Dec. 2014). "Deep speech: Scaling up end-to-end speech recognition." [Online]. Available: <http://arxiv.org/abs/1412.5567>
- [6] L. Oukhellou, A. Debiolles, T. Deneux, and P. Akin, "Fault diagnosis in railway track circuits using Dempster–Shafer classifier fusion," *Eng. Appl. Artif. Intell.*, vol. 23, no. 1, pp. 117–128, 2010.
- [7] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998.
- [8] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.
- [9] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.
- [10] L. Deng and J. C. Platt, "Ensemble deep learning for speech recognition," in *Proc. INTERSPEECH*, 2014, pp. 191