ABSTRACT

Aiming at reducing the train accidents and enabling the railway system maintenance-free vision, this paper outlines an artificial intelligence (AI)-machine learning-based monitoring technology framework, which is capable of autonomously providing the health state awareness of the wheel-track system and direct feedbacks for adaptive positive train control (PTC).

INTRODUCTION

The year of 2018 continues to present passengers and operators with depressing fatal train derailment and collision news. Whether the train derailments and collisions were caused by mechanical wheel-rail unmatched dynamic interactions; mechanical component failures; poor maintenance; or human errors, these accidents have caused severe safety and financial concerns and detrimental impacts on the public transportation perception. Train accidents can mostly be caused by derailments in which track and equipment failure are the primary cause of accident on the main tracks [1]. Additionally, according to the U.S. Department of Transportation (USDOT) Volpe Center-Innovation Research Program Office, 70% of railroad derailments were reported in the U.S. in 2016 [2]. Monitoring the behavior of the train and railway infrastructure elements is a complex task, particularly, when it is desirable to conduct it in real time. There is a wide range of sophisticated diagnostics systems that have been used to support the health monitoring of the wheel/track systems. These systems include track, overhead line, and vehicle dynamics measurement as well as vision system, video inspection, and other capabilities such as signal processing, railroad infrastructure kinematic analysis, and environmental temperature measurement.

In fact, using those monitoring and measurement technologies, the railroad industry has been collecting a tremendous amount of data, and the concept of "big data" has become one of highly complicated research topics within the railroad community for the last decade.

Dy D. Le, Texas Tech University, 2500 Broadway, Lubbock, TX 79409, U.S.A.

Consequently, the research involved in "big data" with the railroad community has taken a big step in using the artificial intelligence (AI) and machine learning technology for the data mining and correlation tasks to potentially differentiate anomalies that may exist in complex patterns. With the advancement of the AI machine learning and the amount of data, which has already been collected, the railroad industry learns many techniques for fine tuning the data collection, processing, and reporting as well as visualization. The developed AI-machine learning technologies and their applications can potentially relieve the railroad maintainers and analysts from the inherent burden with time intensive tasks associated with engineering and analysis of a vast amount of data collected using monitoring methodologies. As a result, the health state awareness technology can now be developed to enable the automation of data fusion and analysis, minimization of required data for trending and analysis, and more effective operation and maintenance of trains and track infrastructures. With appropriate physics-based models and intelligent adaptive controls, the health state awareness technology can also provide an automation capability to achieve the maintenance-free vision for the railway systems. The maintenance-free vision for the mechanical systems has been pursued for quite a long time. The U.S. Army, for example, is currently planning to develop the next-generation helicopters with "fatigue-free" and "zero-maintenance" characteristics [3].

INTELLIGENT DIAGNOSTICS AND CONTROLS FRAMEWORK

A framework has been developed for a proof-of-concept design for a machine learning-based health state awareness system that can be potentially installed on a rolling stock for testing and demonstrating its innovation and effectiveness. The methodology hereby called "intelligent Diagnostics and Controls", or iDaC, is aimed at enabling the real-time monitoring and situation response applicable to railroad and train platforms, Figure 1.

Particularly, the conceptualized system framework is focused on the vehicle-track interactions, but can be modularly reconfigured to extract relevant data from a variety of existing monitoring systems including track, overhead line, and vehicle dynamics measurement as well as vision system, video inspection, and other sources such as signal processing, railroad infrastructure kinematic analysis, and environmental temperature measurements.

The iDac system framework is envisioned to be reconfigurable to have the following automated operational functionalities:

a. Ingest and transform real-time unsynchronized, unstructured/structured data from heterogeneous sources (e.g., sensor signals, on-board monitoring and control system, off-board inspections, and relevant information from maintenance/inspection databases) into recognized patterns.

b. Using defined machine learning-based techniques, analyze patterns to detect anomalies (e.g., abnormal rolling vibration, vehicle dynamic unstable condition, derailment and failure precursor, fatigue crack, and operational damage) and assess the train decreased performance and mechanical element integrity (e.g., extreme angle of attack, inter-axle misalignment, hunting severity, poor brake performance, and defective components).

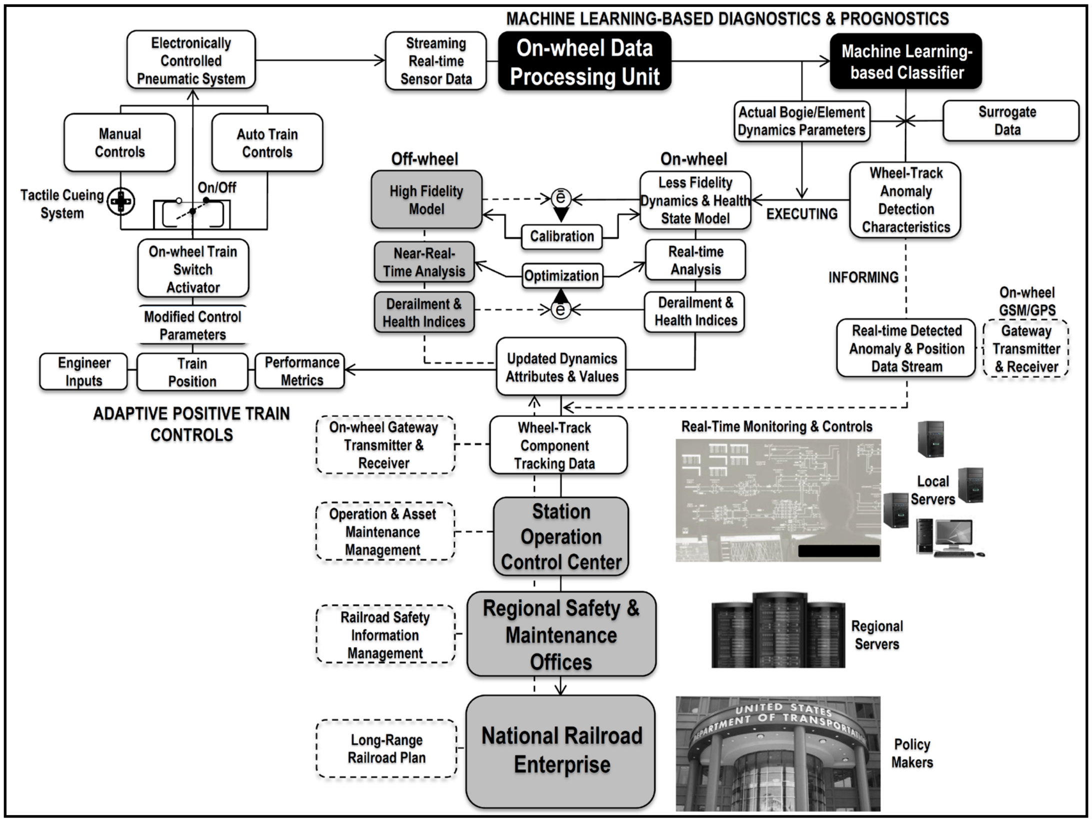

Figure 1: Overall Machine Learning-based Diagnostics and Prognostics Framework

c. Using defined machine learning-based techniques, analyze patterns to detect anomalies (e.g., abnormal rolling vibration, vehicle dynamic unstable condition, derailment and failure precursor, fatigue crack, and operational damage) and assess the train decreased performance and mechanical element integrity (e.g., extreme angle of attack, inter-axle misalignment, hunting severity, poor brake performance, and defective components).

d. Identify a set of influential operating factors (e.g., derailment coefficient, lateral/vertical contact force, wheel load unbalance ratio, damping, critical speed, and track curving radius) and develop modified control parameters (e.g., speed and braking), which are autonomously fed into the PTC feedback loop to prevent derailment and other related catastrophic failures. When the identified influential operating factors combined with or integrated into appropriate physics-models, automation strategies can be potentially developed to reduce the maintenance frequency to minimum or zero within a desirable maintenance-free interval.

e. Provide platforms and operators with options to manually or autonomously execute speed control authority accordingly to: (a) prevent derailment; decelerate or stop crack/damage growth; and mitigate other detected anomalies and (b) optimize in-situ operational performance.

MACHINE LEARNING-BASED DIAGNOSTIC AND PROGNOSTIC METHODS

Train-Rail Interaction Models and Monitoring Parameters

To enable the machine learning training scheme and comprehensive vehicle dynamics analysis for the on-board intelligent system in real-time, a coupled multi-body wheel-rail interaction physics-based model is developed to obtain the dynamic response and performance of the train interacting with the inner/outer rails. The modeling execution is automated (e.g., to receive data inputs from established on-board sensing systems) to predict or project critical wheel-rail dynamic parameters in real time. These computed dynamic parameters are autonomously compared with established threshold values, which, if exceeded, will have a high probability of causing catastrophic accidents or component failures. Additionally, the computed parameters and dynamic information can also be autonomously fused to provide appropriate health indices (HIs), which are visible on the train dashboard for the train comprehensive health state awareness in real time.

There are numerous wheel-rail dynamic models, which have been successfully developed by the railroad industry and academic community. A thorough study needs to be conducted to select appropriate sub-models, which have already been verified and validated. The selected wheel-rail interaction sub-models can then be modified and enhanced for coupling and real-time computation and analysis. Additionally, to accommodate the real-time computation, (e.g., less number of degrees of freedom, fast convergence, and high degree of accuracy), selected techniques such as cross iteration algorithm and relaxation method are needed for developing model algorithms [4].

To monitor the interaction of the train components and track and predict the vibration characteristics, a coupled multi-body wheel-rail interaction physics-based model, as mentioned above, is required. This model consists of several sub-models coupled together to characterize the lateral, vertical, and longitudinal dynamics behavior of the entire train (e.g., several train wagons and bogies). Additionally, the track and its substructures (e.g., rails, sleepers, and ballasts) are modeled as elastic beams and coupled with the train dynamic sub-models to provide a whole wheel-rail system response and behavior. The coupled wheel-rail interaction physics-based model is integrated with reduced-order mathematical equations for the real-time execution of modeling to track the dynamics of wheel-rail interactions and behavior white the train is passing through tracks.

Sub-models, which are coupled together, address not only the performance and vibration characteristics of the wheel and rail in lateral, vertical, and longitudinal directions, but also the effect of braking system. The lateral dynamic model provides the vehicle dynamic behavior in the lateral displacement, roll, and yaw directions and is used to monitor the lateral stability including hunting, rail curving effect, and derailment. The vertical dynamic model provides the dynamic performance of suspension systems and wheel-rail as a whole. The longitudinal dynamic model can be used to monitor the performance of the train braking system on the dynamics of the whole train and potential derailment as the train is running through curve tracks.

There are numerous critical dynamic parameters, produced by these models, which can be monitored while trains are running to ensure passenger safety and performance of the wheel-rail system. Critical dynamic parameters and data that may affect the safety and performance are:

a. Maximum train vibrations induced by critical speed (e.g., natural frequency), which generates the resonance and instability of the wheel-rail system. When the train speed (TS) reaches a critical speed (CS), the amplitude of the train vibration grows exponentially with time.

b. Derailment precursors can be predicted using the wheel-rail contact forces while trains are running on track. From the wheel-rail contact forces, the derailment coefficient (DC), which depends on many factors including track curve and lubrication/friction coefficient, can be computed or estimated. When the train running DC exceeds a safety limit, derailment can happen. Additionally, using the DC data and information generated by trains running on track in real time, maintenance actions can be planned for wayside lubrication on tracks to reduce the potential derailment effect.

c. For those train bogies designed with semi-active suspension controls, the damping effect can change in real time. When the damping ratio (DR) reduction exceeds a threshold value, the train can reach its critical speed and become unstable. So, the influential relationship of critical speeds of trains and suspension stiffness and damping (SD) can be monitored and/or adjusted to ensure safety and better performance.

d. The maximum and total amplitudes of acceleration for the rail and sub-structures, bogies, primary/secondary suspension, and wheels increase with the increase of the train speed, which also has great influence on the wheel-to-rail contact forces.

e. Wheel bearing temperature measurements.

Train-Rail Machine Learning-based System Functionalities

Figure 1 illustrates a framework for the machine learning-based diagnostics, prognostics, and adaptive controls designed to prevent the train catastrophic events. As the raw sensor data and information are being streamed, the envisioned machine learning-based system (e.g., hardware/software), when installed on trains, can perform the following essential tasks in real time:

a. Execute the coupled multi-body wheel-rail interaction physics-based model (low fidelity): During the execution of the modeling and analysis, critical parameters based on actual operational data, are computed, derived, and consequently, fused into health condition indicators or indices. The computation and analysis are designed to be in real time with low fidelity models, but calibrated and optimized with off-wheel modeling solutions in near real-time using on-board gateway transmitter and receiver.

b. Conduct machine-learning-based fault classifications: While the wheel-rail interaction computation is being done, the system simultaneously conducts the fault classification using the machine learning-based training models, which will be described in detail in the next section. If an anomaly were detected, the relevant wheel-track anomaly characteristics would be transmitted immediately to the train operation stakeholders including the notional Station Operation Control Center, Regional Safety and Maintenance Offices, and potentially National Railroad Enterprise. At a particular train station, a local high performing server system is also designed to receive real-time inputs from the running trains to perform the coupled multi-body wheel-rail interaction physics-based model but with significantly higher fidelity and more complicated neural network (more hidden layers). The results would be then automatically transmitted to the running train on-board server system for calibration and optimization with real-time solutions.

c. Modify influential parameters: Upon having computed and derived the health condition indices and/or fault characteristics, the on-wheel system will modify influential parameters (e.g., speed, braking, and/or other operational parameters), which have direct dynamic relationship between, for example, derailment and fault acceleration and provide direct inputs to the on-board adaptive PTC system. The notional adaptive PTC system, installed on trains, can have an on/off switch for train controls. If an "off" position is selected, upon having received modified parameters/inputs, the system can providing sensitive cueing signals and/or visual/audio warnings for operator's prompt actions. If an "on" position is selected, the adaptive PTC system can take over the controls and immediately execute the appropriate actions to reduce the speed or stop the train completely. The notional Station Operation Control Center can also perform the envisioning adaptive PTC functionality, if adaptive PTC is not available on trains.

Long Short-Term Memory and Auto-Encoder and Decoder Methods

There has been a number of machine learning methods developed for various applications including fault diagnostics and prognostics for many mechanical platforms. Taking advantage of the work, which has been substantially investigated, the iDaC framework can extend the Long Short-Term Memory (LSTM) method, Figure 2, and integrates it with Auto-Encoder and Decoder (AE/AD) models (e.g., to classify the train dynamic instability, derailment precursors, corrugated rail severity or irregularities, and other potential faults) [5-6].

LSTM-AE/AD has been used in many applications including multi-sensor prognostics for the aviation platform systems [7].

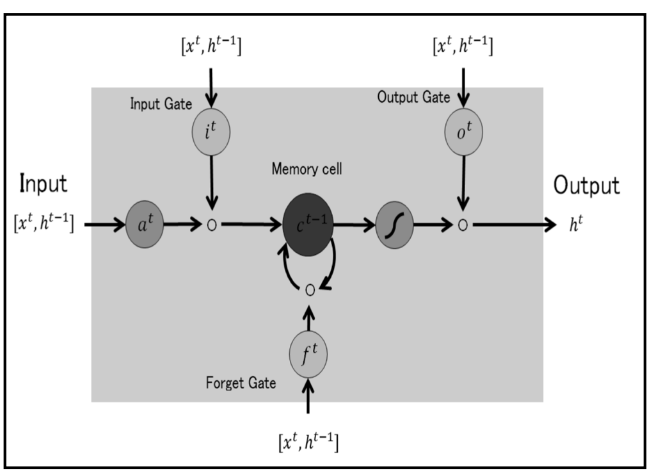

Figure 2: LSTM Cell

LSTM is a machine learning method for sequence tasks, and its cells contain information outside the normal flow of the recurrent network in a gated cell, in which information can be stored in, written to, or read from. LSTM is a suited neural network for sequential sensor data and for discovering hidden patterns from multiple operation data and fault models developed from the sensor networks installed on-board of mechanical platforms. An LSTM cell makes decisions about what to store, and when to allow reads, writes and erasures, via gates that open and close. It uses input x^t and hidden state activation h^{t-1}, and memory cell activation c^{t-1} to compute the hidden state activation c^t at time t. LSTM cell uses the Input Gate i^t, Forget Gate f^t, and Output Gate o^t, to make a decision if: (1) an input needs to be remembered, (2) a previous memory needs to be retained, and (3) when a memory content needs to be output (h^t).

The envisioned iDaC framework, with appropriate interfaces communicating to on-board data buses and other monitoring systems, is designed to extract train-rail operational parameters, modify the control variables and determine the required time for action to prevent the potential derailment accident, if a high probability of derailment exists using HIs and limited critical operation parameters or COPs (e.g., DC, CS, DR, and wheel load unbalance ratio). During operation, the on-board system extracts, processes, and feeds collected data and information to the on-board LSTM neural network for deriving time-based mean and oscillatory parameters of COPs. Using on-board lookup tables, the on-board computation with low fidelity models will perform its computation and continue its iterations with real-time data until any calculated or operational COPs exceed the limit COPs. When this condition is first reached, the associate sub-system is identified and control parameters (e.g., speed, braking movements) are calculated. Modified control parameters are then fed into the notional adaptive PTC scheme, which can provide the platform and operator options to manually or autonomously execute the train control authority accordingly to prevent a potential derailment accident.

During the operation of the trains, at a designated train operation center equipped with the high performing computing facility, the high fidelity-coupled multi-body wheel-rail interaction physics-based model and neural network with additional hidden-layers are also running for providing the model calibration and increasing the computational accuracy in real time. Solutions are then updated into lookup parameter tables. These tables, together with other relevant information including the train-rail diagnostics and prognostics data, are automatically uploaded onto the on-board system at the end of service via on-board wireless gateway transmitter and receiver. However, if anomalies are detected during operation, a set of relevant data and information for a specific faulted sub-system or low (or high) fidelity solutions can be immediately transmitted to or received from the local train operation center and enterprise server using available wireless capabilities such as Global System for Mobile Communication (GSM) and Global Position System (GPS).

CONCLUSION

Aiming at reducing the train accidents and enabling the railway system maintenance-free vision, an AI-machine learning-based health monitoring framework, which is capable of autonomously providing the health state awareness of the wheel-track system and direct feedbacks for PTC, is presented. This framework cannot be effectively designed without the integration of the AI-inspired machine learning technology and physics models. When the identified influential operating factors combined with or integrated into the appropriate physics-models, automation strategies can be potentially developed to reduce the maintenance frequency to minimum or zero within a desirable maintenance-free interval. For several decades, the research community has developed the rule and statistical learning-based AI technologies. However, there are challenges, particularly in the machine learning and training where issues with sparse data exist. Even when a vast amount of data is available, noisy data can pose a significant problem. Additionally, a realistic complex operation of the railway system can introduce an unstructured environment and hide many critical system characteristics, which might not be able to uncover unless physics-based models are integrated into an AI-inspired machine learning framework. In addition to the prior knowledge, the envisioned intelligent diagnostics and controls framework introduces the next-generation AI capability that can acquire the essential knowledge from many hidden states of complex operations and unstructured domains [8].

ACKNOWLEDGEMENTS

The author greatly appreciates the support from the Texas Tech University – Office of the Vice President for Research. Additionally, the author expresses a sincere gratitude for the encouragement from Professor Fu-kuo Chang of Stanford University and support from the Committee of the 2nd International Workshop on Structural Health Monitoring for Railway System.

REFERENCES

1. Saat, R., Barkan, C., and Liu, X., "Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates", Transportation research record: Journal of the Transportation Research Board, No 2289, Transportation Research Board of the National Academies, Washington, D.C. 2012, pp 154-163.
2. USDOT Small Business Innovation Research (SBIR) Program, Phase I Program Solicitation, '18-FR2: AI Tools for Relating Various Sensor Data to Adverse Vehicle/Track Interaction", Issue Date: January 17, 2018.
3. Le, D. and et al., "Fatigue-Free Platforms: Vision for Army Future Rotorcraft", American Helicopter Society (AHS) 70th Annual Forum Proceedings, Montreal, Canada, May 20-22, 2014.
4. Lei, X., Wu, S, & Zhang, Bin, "Dynamic Analysis of the High Speed Train and Slab Track Nonlinear Coupling System with Cross Iteration Algorithm", Journal of Nonlinear Dynamics, Volume 2016, Article ID 8356160.
5. https://deeplearning4j.org/lstm.html#long
6. Cho, K. and et al., "Learning Phrase representations using RNN Encoder-Decoder for Statistical Machine Translation", Cornell University Library.
7. Malhotra, P. and et al., "Multi-Sensor prognostics Using an Unsupervised Health Index Based on LTSM Encoder-decoder, arXiv: 1608.06154v1 [cs.LG] 22 Aug 2016.
8. Le, D., "Fabrics of Artificial Intelligence-informed Technology for Holistic Sustainment (FAITHS)", Presented to Defense Advanced Research Project Agency (DARPA), Arlington, VA, March 24, 2017