Train Delay Prediction using Machine Learning

2 authors, including:

Lokaiah Pullagura
Jain University
29 PUBLICATIONS   **123** CITATIONS

Train Delay Prediction using Machine Learning

Lokaiah Pullagura, Jeevaa Katiravan

Abstract: Indian Railways operates both long distance and suburban passenger trains and freight services daily in the country. Trains get delayed frequently due to several reasons such as, severe weather conditions such as fog, traffic, signal failure, derailing of trains, accidents, etc, and this delay is propagated from station to station. If we can predict this in advance - it would be of great help for the commuters to plan their journey either for an earlier departure or postpone, and also lets railways to take measures to avoid delays further. In this paper, we used decision tree, a machine learning method used for predicting train delays, and Recurrent Neural Networks distinguished with various fixtures. For predicting train delays, Recurrent Neural networks with 2 layers and 22 neurons per each layer gave best results with an average error of 122 seconds.

Keywords : Train Delay(TD), Machine Learning, prediction, decision tree, RNN.

I. INTRODUCTION

Railways functions by administering trains running in their entire network. It is a typical task of managing certain aspects such as timetable, equipment, etc. To avoid collision of trains, a timetable showing schedule of arrival and departure is designed. The Indian railways uses different gauges depending on the rail traffic. The Broad Gauge is used in areas with high traffic whereas Meter gauge is used in areas with less traffic and the Narrow Gauge is used in hill regions. Train delay is a major issue in IRs. The Railway Network is an interconnected group of stations connected with links. The traffic in the Indian Railway Network(IRN) is accumulated from station to station. Giving importance to network and traffic, main stations in the route are recognized. This helped to pinpoint some of the most likely nodes of blockage in the grid. For trains functioning, a span between two trains passing through the same station is to be maintained. If a train delays at a particular station, the subsequent trains stopping at same station may also be delayed. The first station is the origin for delay of train. The delay at starting point normally leads future delays at successive stations. Train Delays (TDs) occur due to several reasons thereby leading to deviation in the stipulated schedule, thereby overwhelm railway functioning and they cause service disorder and sometimes catapult in extreme cases.

II. LITERATURE SURVEY

Several factors influence train delay. [1-2] discusses factors causing train delays. [3] identifies factors and minimizes the main reasons of train delay Trains are delayed due to bad weather[4], [5] when trains travel in the same track,[6] derailments due to wheel failures and improve railway safety.[7] predicts failures in wheels of train before they hamper functioning. [8] assesses extreme weather impact on freight trains in various countries. [9] identifies and examines the progress of delays in the rail road. [10] proposes a neural network approach for predicting flight delays in an airport.[11] Suggests algorithm with a single hidden layer which dynamically takes the input with weights and ascertain the output with weights. [12] Uses Machine learning approach for alarm prediction and failure prediction. [13] presents highly accurate neural network model to predict passenger train delays in Iranian Railways. [14] captures conection between delays and characteristics of a railways. [15] Uses Machine Learning algorithms and tools for predicting train delay. [16] proposes technologies for forecasting delay of trains. [17] proposes a fast learning algorithm based on Extreme Learning Machine for predicting train delays the that totally exploits the short in-memory large-scale data processing technologies.

The remaining part of this work is organized as follows. In section 3, Data collection process is explained. In section 4 Train delay prediction model is described in detail. In section 5, experimental results are presented with and without boosting techniques as well as results of Recurring Neural Networks when tested by changing the number of hidden layers and nodes per layers.

III. DATA

Our primary focus was to predict train delays, therefore various types of data needed to be collected to make the predictions possible. The primary sources of data are websites like www.runningstatus.in and www.etrain.info/in have information about trains between stations, real-time arrival/departure information. Data collected from different origins may be in different formats. Previous weather conditions data was also collected. Data is collected on departure times for 09 major stations from Vijayawada to Chennai Central for the three-month period i.e. from Jan 2018 to March 2018. Each Train's departure was mixed with the weather data. The combined data was split into two parts: 80% was used as training data and the rest as testing data.
Data is collected for the following parameters:Station, Train_number, Date, Past_Delay, Designated_Arrival, Real_Arrival, Time of delay.

There are around 72 trains travel in the route of which 12 trains are regular. Data collected for three-month is sorted according to the train arrival. Indian Railways Website releases the delay of trains at each station. For collecting the delay information from railways websites, a Web Crawler is developed. It is observed several trains running in the route are delayed out of which 23.7% involved super fast trains.

Revised Manuscript Received on December 15, 2019.

Lokaiah Pullagura, Research Scholar, Dept. of CSE, Sri Satya Sai University of Technology & Medical Sciences. Sehore, Madhya Pradesh, India. Email: lokaiah75@gmail.com

Dr. Jeevaa Katiravan, Associate Professor, IT department, Velammal Engineering College, Chennai, Tamil Nadu.

1312

Fig.1 Route Map of Vijayawada to Chennai Central stations

IV. TRAIN DELAY PREDICTION MODEL

To build effective train delay predictor, it is required to model attributes like type of train, the period of its journey, the duration of its journey, area of its journey Delay Times, Train Maximum Speed, Route Maximum Speed, Speed Homogeneity, Route Departures, Vertex Departures/Arrivals, Edge Length, Percentage Double Track, etc.

Rail road is divided up into a number of paths, which meet at some locations. These paths having traffic cover a successive rail roads and goes through several railway nodes. Railway network is considered to be a graph which is a collection of edges and vertices as shown in fig:
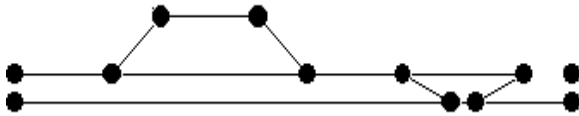

Fig:2 Illustration of a railway graph data model

A train passes between stations. Each station contains data about its respective routes or intersections. Assuming the premise that delay is a result of infrastructure,

All information about Train Departures which is downloaded was stored in a MongoDB database. Python programming language is used to develop a model as it has rich set of libraries.
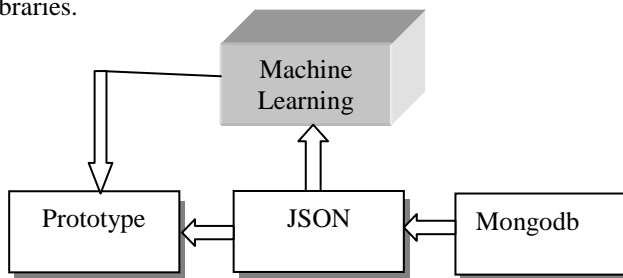

Fig: 3 Prototype showing the paths taken by the data Test for Predicting the Delay of Train.

For different machine learning methods, different tests were performed for approximating the accurate result. The average delay is the difference of average time i.e.

AverageDelay(Ad) = Tpd – Tad

where, T_{pd} is the predicted delay for a departure and T_{ad} is the actual delay.

V. EXPERIMENTAL RESULTS

Decision tree was tested with variation in maximum depth setting as shown in table 4.1. The maximum depth 4 gave the accurate result with average errors 166. The tests were run once for each setting.

Table I: Average errors for various tree depths without using boosting.

Maximum depth	Average error(s)
1	178
2	181
3	201
4	166
5	168

A. Decision tree with AdaBoost

Decision tree using AdaBoost was added with 5 estimators. The result is shown in table 4.2. The Maximum depth that gave the most accurate result was 5. The average errors of decision tree with and without AdaBoost are compared in figure 4.1. For each testing the tests were performed once.

Table II: Average error for different tree depths when using AdaBoost with 5 estimators.

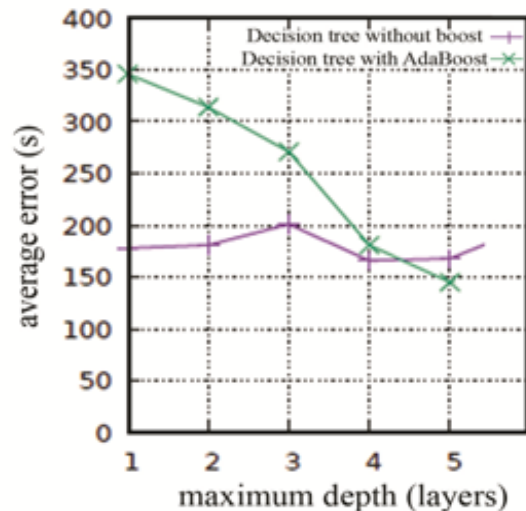| Maximum depth	Average error (s)
1	346
2	314
3	271
4	181
5	145

Fig:3 Two graphs showing the average prediction errors in seconds while using decision tree with and without AdaBoost.

B. Recurrent Neural networks(RNN)

RNNs are typical neural networks where the output of previous layer are fed as input to the current layer. The main and most important feature of RNN is **Hidden state**, which remembers some information about a sequence.
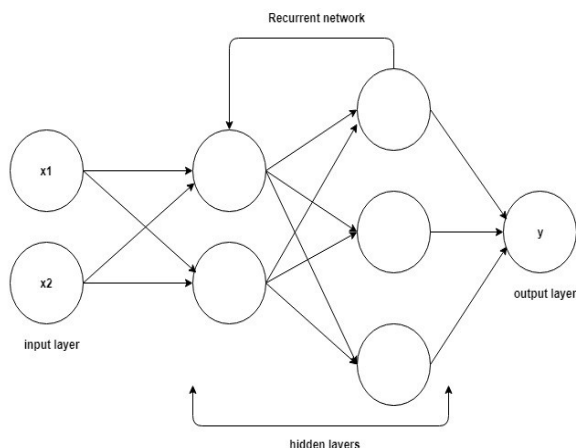
Fig:4 Recurrent Neural Networks with two hidden layers

RNN was tested by changing the number of hidden layers and nodes per layers. This is shown in table 4.3 and figure 4.2. The most accurate configuration was 2 hidden layers with 22 nodes per layer. For each testing the tests were performed once.

Table III: Average error for different layer configurations in neural network.

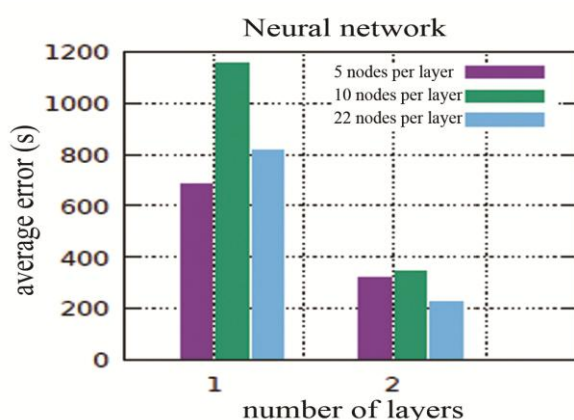| Hidden layers	Nodes per layer	Average error (s)
1	5	685
1	10	1156
1	22	816
2	5	322
2	10	346
2	22	227

Fig:5 The average errors for different RNN configurations.

VI. CONCLUSION

In this paper, one of the important goals is to predict delays by choosing the accurate method in machine learning. From results it is understood that Recurrent Neural Network is the most accurate method. However, the study of data considered was very short period. Another goal of this paper is to predict train delays using machine learning. Results obtained are vague. This is because duration of time was very short. A Delay prediction model developed displays statistics about train delays.

In a further study, the work done in this paper can be extended to collect data for a one or more years thereby giving machine to find the determinants and remove the noise data. It can be extended to RNN with more hidden layers. In this paper we have used two machine learning methods and a boosting technique. More investigation can be done for other machine learning methods and boosting technique.

REFERENCES

1. Magadagela K, Nel H, Marnewick A. Identification of delay factors that affect high dwell times of freight trains. In: IEEE Technology & Engineering Management Conference (TEMSCON), Santa Clara, CA, USA, 2017, 179–84.
2. Olsson NOE, Haugland H. Influencing factors on train punctuality—results from some Norwegian studies. Transp Policy 2004;11:387–397.
3. Jana Lalinská, Jozef Gašparík1 and Denis Šipuš: "Factors Affecting the Delay of Passenger Trains " LOGI – Scientific Journal on Transport and Logistics Vol. 8 No. 1 2017 DOI: 10.1515/logi-2017-0009: 74-81.
4. Pu Wang and Qing-peng Zhang "Train delay analysis and prediction based on big data fusion" Transportation Safety and Environment, 2019, Vol. 00, No. 00 1–10.
5. Carey, M, Kwiecinski, A. Stochastic approximation to the effects of headways on knock-on delays of trains. Transportation Research 2007; **28**: 251– 267.
6. Preventing Train Derailment (Salient Systems Inc., 2012.
7. Chunsheng Yang, Sylvain Létourneau: "Learning to predict train wheel failures" Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining Pages 516-525.
8. Johanna Ludvigsen, Ronny Klæboe: "Extreme weather impacts on freight railways in Europe" Nat Hazards (2014) 70:767–787.
9. Fabrizio Cerreto , Otto Anker Nielsen , Steven Harrod , Bo Friis Nielsen. "Causal Analysis of Railway Running Delays". Paper presented at 11th World Congress on Railway Research (WCRR 2016), Milan, Italy.
10. Yang, Chunsheng & Létourneau, Sylvain. (2005). Learning to Predict Train Wheel Failures. 516-525. 10.1145/1081870.1081929.
11. Guang-Bin Huang ; Qin-Yu Zhu ; Chee-Kheong Siew Extreme learning machine: a new learning scheme of feedforward neural networks. 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541).
12. Hongfei Li , Dhaivat Parikh, Qing He, Buyue Qian, Zhiguo Li, Dongping Fang, Arun Hampapur "Improving rail network velocity: A machine learning approach to predictive maintenance". H. Li et al. / Transportation Research Part C 45 (2014) 17–26.
13. Masoud Yaghini Mohammad M. Khoshraftar Masoud Seyedabadi "Railway passenger train delay prediction via neural network model" J. Adv. Transp. 2013;47:355–368.
14. Nikola Marković, SanjinMilinković , Konstantin S.Tikhonov, PaulSchonfeld : "Analyzing passenger train arrival delays with support vector regression".
15. Emanuele Fumeo, Luca Oneto, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, Davida Anguita "Big Data Analytics for Train Delay Prediction: A Case Study in the Italian Railway Network".
16. Luca Oneto, Emanuele Fumeo, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, Davide Anguita "Train delay prediction systems: a big data analytics perspective." Big Data 2018;11:54–64.
17. Luca Oneto, Emanuele FumeoGiorgio ClericoRenzo, CanepaFederico, PapaCarlo, DambraNadia. Mazzino Davide Anguita "Delay Prediction System for Large-Scale Railway Networks Based on Big Data Analytics".

Retrieval Number: A2088109119/2019©BEIESP
DOI: 10.35940/ijeat.A2088.129219

1314

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

AUTHORS PROFILE

Lokaiah Pullagura, Research Scholar, Department of CSE, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India. He has published several papers in International journals. His research area of interest is in Machine Learning..

Dr. Jeevaa Katiravan, He is currently Associate Professor in IT Department. He has rich experience in Research/Teaching and published several papers in International and National journals. He has guided several research scholars in the area of Information Security and Machine Learning. He is a Doctoral Committee Member at Anna University, Sathyabama University and VIT University and acted as External member For Ph.D Viva at Anna University and St. Peter's University.