Modelling, Simulation and Optimization of London High Speed Train

Raksha Muddegowdana Koppalu Prakasha
Master of Science in Data Analytics
School of Computing
National College of Ireland
Dublin, Ireland
x19193181@student.ncirl.ie

***Abstract*—High-Speed Train for HS2 line is planned between London Old Oak Commons and Birmingham Birmingham Interchange station. This report documents the layout and results of the simulation and optimization experiment carried out to simulate the travel of trains in the HS2 line. The goal of the Simulation part of the experiment is to create a simulation for the London Old Oak Commons to Birmingham Interchange section of the high-speed line and induce it with situations of the temporary breakdown of electric line, weather impacts, and passenger dwell time. The goal of optimization is to maximize the train frequency per hour with the conditions that the average time should not be more than half the time of schedule between consecutive trains. Simpy which is a python discrete event simulation system is used to build the experiment.**

***Index Terms*—Simulation, Modelling, Optimization, Experiment Report**

I. Introduction

Simulation for the HS2 line running between London Old Oaks Common and Birmingham Interchange Station is designed to determine the maximum number(throughput) of trains that can be run on this line in an hour. In reality, the journey of the train is never an ideal scenario, and accordingly even in the simulation process various external factors/variability's such as strong weather condition(heavy rain, technical problems), and dwell time due to passenger boarding/movement, which impacts the speed and thereby traveling times of the train is induced. The variability of strong weather condition was induced by values generated using log-normal distribution, with a μ value of 3.378 and σ value of 0.751. The variability of dwell time caused by passenger movement is induced by values generated using python's random generator in the range (10, 20). The simulation process also helps attain the optimum number of signaling blocks between the two stations so as to aid the maximum number of train travels in an hour.

II. Assumptions

The following assumptions were taken into consideration, to carry out the simulation process:

- For the train to run at the maximum full speed, we have assumed/considered that there needs to be an at least 2 free block of signal maintained by the train, ahead of it.
- The train is allowed to enter a signal block only when there is no other train in the block and the signal for entry into that block is "green".
- The signal at the entrance of a block switches to "red" as soon as a train enters the block, and switches back to "green", 5 seconds after the tail of the train leaves the block.
- The train will accelerate as it leaves from the London Old Oak Station, which will also be the first signaling block and will begin to decelerate as it reaches Birmingham Interchange Station which is the last signaling block.
- Delays due to dwell time by passenger boarding and movement is considered as a small-scaled variability additional to the large impact variability of weather impact, which affects the future train schedules.
- For the variability of strong weather delay's induced, log-normal distribution is used, with assuming a μ value of 3.378 and σ value of 0.751 [1]
- The problem statement suggests the assumption of track consisting of k signaling blocks of equal length and a fixed schedule of n trains an hour. In our case, 14(k) signaling blocks and 10(n) trains were considered to carry out the simulation.
- The train is assumed to start its operation at 8.30 in the morning for the simulation purpose, with a breakdown of the electric line is induced to the 9 am train, which has a repair or delays time of 30 minutes or 1800 seconds.

III. Objective

A. Simulation Objective

- To build a simulation model to simulate the high-speed line train travels for "n" trains between London Old Oak Station and Birmingham Interchange Station, with "k" signaling blocks along its track.
- Inducing a temporary breakdown due to electric malfunction for the 9 am train, to carry out verification of the simulation model.

B. Optimization Objective

- For the "k" signaling blocks and "n" trains, create a schedule for trains between London Oak Station and Birmingham Interchange, with the first departure of the

train from London Oak Station is at 7 am and last one at 10 pm.
- Reporting the distribution of actual delay times using the simulation model.
- To maximize the frequency of trains per hour under, with a condition that the average delay time of trains shouldn't be greater than half the scheduled time between trains(5 min in our case).
- To determine the optimal "k" blocks of signal and "n"

IV. IMPLEMENTATION

A. Libraries

Multiple Python libraries such as Math, Matplotlib, Numpy, Pandas, Scipy, and Simpy are used to build the simulation model.

B. Dataframe Creation

Using the information given in the context of the problem state, a data frame is created with fields "FROM", "TO" "TIME", "BLOCKS" and "DISTANCE".

C. Functions for Simulation

For Simulation model I have defined 4 important functions, they are:

1) Utility Functions:

- m_to_km() : To convert meters to kilometers
- hhmmss() : To format time as HH:MM:SS
- s_to_m() : To convert time from seconds to minutes
- getTravelTimeDistance() : To find the travel time when velocity(speed) of the train is given
- getTravelTime() : To find the travel time when distance is given.
- getTravelDistance() : To calculate distance travelled, given speed and time taken by the train

2) Train CLass: Class Train() is created to handle data/information regarding the moving train and their individual train objects.

3) Network Class: Class Network() is created to handle movement of trains and signal blocks(pre) and represents the track line. It contains multiple sub functions:

- signal_control(): Function to carry out controlling of signal blocks in the line.
- new_train() : Function that handles simulation of new train objects.
- strong_weather_impact(): Function used to induce the variability(high impact) of delay due to bad weather conditions.
- dwell_time() : Function used to induce the variability(low impact) of delay due to passenger movements.

4) update_train_state(): update_train_state() function records the information of each trains travelling for every iteration and updates its current position and travelling time

5) run(): run() function is the major process of the simulation and encaptures coding logic for controlling the movement of trains. It includes the logic:

- Variability of movement of passengers/dwell_time() is added when Train leaving London Old Oak Commons Station.
- Variability of bad weather/ strong_weather_impact() is added for trains between London Old Oak to Birmingham Interchange stations, which will affect the movement/travel_time of a train and schedules of future trains too.
- Trains traveling check for 2 subsequent signals, if green, before entering the next block. If it doesn't get 2 green signal it waits for the train ahead to pass.

6) delay(): delay() function is used to induce the temporary breakdown due to electrical malfunction of the 9 am train from London to Birmingham. It takes 30 minutes to fix the problem.

7) train_simulation(): train_simulation() process is called to handle the entry of new train into Network, new run() process is called every time a new train is created.

8) main(): main() function validates the input, sets the environment for simulation and start and end time for simulation.

D. Simulation

Running Simulation for "k" number of blocks and "n" number of trains with the variability of bad weather and passenger dwell time introduces. Additionally breakdown of 9 am train because of technical issues.

The start_time and end_time of the simulation are also passed to the main() function, which starts the simulation process.

E. Simulation Results

A Dataframe is created to show the results of the simulation carried out in the above step, in our case the simulation is carried for 14 blocks of signals on the line and 10 trains.

F. Verification of Simulation Results

1) build_train_summary(): This function creates the summary to all trains simulated(Figure 1), and that entered into Network line.

A bar graph is also plotted to show the travel time(Figure 2) and delay/wait times(Figure 3) of each train in the Network line.

On verification of the simulation results,it was seen that the model was successfully able to simulate the travel of 10 trains with 14 signaling blocks, with a breakdown of 30 minutes induced to the 9 am train, with other additional variabilities. The simulation was started with a start_time of 8.30 am. The simulation result also returned an average travel time of trains on the line, which is 33:07 minutes.

G. Optimization

1) get_optimal(): get_optimal() function is used to calculate the average travel time, average delay time, and throughput for all combinations of the number of trains ranging from

	TRAIN	POSITION	TRAVEL_TIME	WAIT_TIME
0	1	15	00:33:16	65.0
1	2	15	00:33:47	90.0
2	3	15	00:32:57	39.0
3	4	15	00:33:12	59.0
4	5	15	00:33:09	51.0
5	6	15	00:33:10	56.0
6	7	15	00:33:05	56.0
7	8	15	00:33:02	52.0
8	9	15	00:32:47	30.0
9	10	15	00:32:45	34.0

Fig. 1. Simulation Results

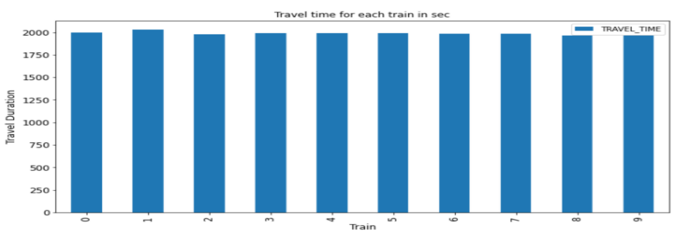

Fig. 2. Bar Map for Travel Time

(1...20) and a number of blocks ranging from (1...15) between the time 7 am - 10 pm. The data frame "optimal" records the simulation results.

2) Train Schedule: Train Schedules from London Old Oak to Birmingham Exchange with the first train departing London Oak Station at 7 am and last train departing at 10 pm is created using get_optimal() function, which generates the results for train schedule information of a data frame, "m_train".

A bar graph is plotted to show the distribution of delay times for the simulated trains. The average delay time for trains calculated is 90.05 and the minimum value of delay time is 54.01.

In the optimization carried out, our objective of delay times of train being less than half the scheduled time between trains was seen being true in all the cases of the results computed.

As per the optimization carried out using get_optimal() function, the maximum number of trains that can be scheduled

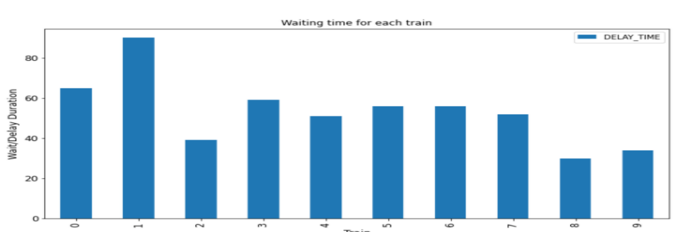

Fig. 3. Bar Graph for Dealy/Wait times

to run on the line in an hour with minimum delay time is 10 trains, with 6 signaling blocks, a throughput value of 10, traveling time of 32:38 minutes(Figure 4).

	TRAINS	BLOCKS	OVERALL_AVG_TRAVEL_TIME	OVERALL_AVG_DELAY_TIME	THROUGHPUT
131	10.0	6.0	1958.58	54.01	10.0

Fig. 4. Optimization Result

3) Travel Optimization: Travel Optimization, to reach the goal of scheduling maximum possible trains per hour online with "k" signaling blocks and delay times being less than half of the scheduled time between trains is carried out using OLS regression and Monte Carlo Optimization.

a) OLS Regression: OLS Regression is carried out(Figure 5) with Delay times of the train as a target and train and blocks as the predictors. The regression resulted in a constant coefficient of 54.0454, a block coefficient of 0.0011, and a train coefficient 0.0268, an R-squared value of 0.535, an F-statistic value of 143.0.

OLS Regression Results

Dep. Variable:	OVERALL_AVG_DELAY_TIME	R-squared:	0.535
Model:	OLS	Adj. R-squared:	0.531
Method:	Least Squares	F-statistic:	143.0
Date:	Tue, 25 Aug 2020	Prob (F-statistic):	4.37e-42
Time:	01:37:25	Log-Likelihood:	157.15
No. Observations:	252	AIC:	-308.3
Df Residuals:	249	BIC:	-297.7
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	54.0454	0.023	2352.195	0.000	54.000	54.091
BLOCKS	0.0011	0.002	0.546	0.586	-0.003	0.005
TRAINS	0.0268	0.002	16.904	0.000	0.024	0.030

Omnibus:	26.800	Durbin-Watson:	0.062
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33.283
Skew:	-0.889	Prob(JB):	5.93e-08
Kurtosis:	2.918	Cond. No.	36.6

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig. 5. OLS Regression for Optimization

The above-computed coefficients were used in a linear regression formula to find the objective function used by Monte Carlo optimization.

b) Monte Carlo Optimization: Monte Carlo Optimization is carried out using the regression results, to minimize the delay or wait times of the trains, and thereby compute maximum throughput or the maximum number of trains that can be scheduled in an hour with signaling blocks. Heatmap(Figure 6) for the optimization is computed for random seed values of 41, 30, and 20 for a value of 100.

In the first heatmap, we can see that the number of blocks is 9, and the number of trains was 10, with delay time of 54.3233. The second heatmap, the number of blocks is 4, the

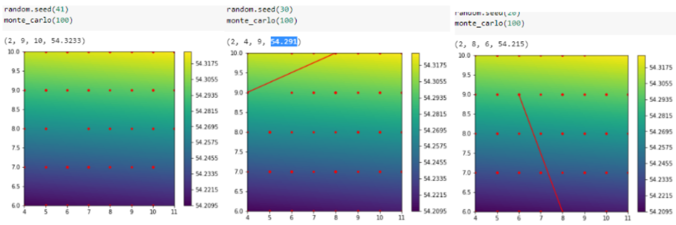

Fig. 6. HeatMaps for Monte Carlo Optimization

number of trains is 9, and the delay time is 54.291. In the third heatmap, the number of blocks is 6, the number of trains is 8, and the delay time is 54.251. Taking all three heatmaps into consideration, and keeping into account the very little difference in the delay times of the trains in the heatmaps, the optimal number of trains that can be run on the line is 10, with 9 signaling blocks and least delay was 54.251.

V. Conclusion

In this project, a simulation model was designed to run trains between the stations of London Old Oak and Birmingham Interchange, with signaling blocks. The train travel is simulated for an hour, with variabilities in travel time and delay sue to impact of bad weather, and passenger dwell time is introduced into the simulation process. Additionally, an incident of breakdown due to electric failure in the line for the 9 am train is also introduced into the model. The simulation experiment carried out concluded that 10 trains can be run in an hour between the stations, and the train in overall requires around 32-40 minutes to complete the entire journey. Optimization of the simulation model was carried out to maximize the frequency of trains per hour with a condition that the delay time of the train should not be more than half the scheduled time between the trains. And accordingly, the optimization carried out using optimal() function resulted in a maximum frequency of 10 trains, 6 signaling blocks, and a delay time of 54.01. Monte Carlo optimization was also carried out, which resulted in a maximum frequency of 10 trains, 9 signaling blocks, and a delay time of 54.251.

References

[1] P. Huang, C. Wen, J. Li, Q. Peng, Z. Li and Z. Fu, ”Statistical Analysis of Train Delay and Delay Propagation Patterns in a High-Speed Railway System,” 2019 5th International Conference on Transportation Information and Safety (ICTIS), Liverpool, United Kingdom, 2019, pp. 664-669, doi: 10.1109/ICTIS.2019.8883805.

[2] Users.iems.northwestern.edu. 2020. [online] Available at: ¡http://users.iems.northwestern.edu/ nelsonb/IEMS435/PythonSim.pdf¿ [Accessed 25 August 2020].

[3] Documentation for SimPy — SimPy 4.0.2.Dev1+g2973dbe Documentation. https://simpy.readthedocs.io/en/latest/contents.html. Accessed 25 Aug. 2020.

[4] D’Ariano, Andrea, et al. “Simulation and Optimization for Railway Operations Management.” Journal of Advanced Transportation, 2 Aug. 2018, doi:https://doi.org/10.1155/2018/4896748.