TRAM: Time-series Rail Schedule Analysis Methods

by Jeff Page (Student number 500223090)

# Introduction

SEPTA Regional Rail provides commuter rail service and wishes to maintain a high degree of reliable service. Service reliability is measured with On-Time Performance. The goal is 91% of trains arrive at their destination less than 6 minutes late. The primary question is: Does SEPTA meet their annual On-Time Performance goal? If not, can patterns of lateness be identified? The train history data provided by SEPTA will be analyzed to determine an answer.

# Literature Review

There are multiple considerations in the creation of a train schedule and why trains may not stay on schedule. The following papers summarize some of the considerations.

**Deadlock analysis, prevention and train optimal travel mechanism in single-track railway system**

Feng Li a, Jiuh-Biing Sheu b,⇑, Zi-You Gao c,

Transportation Research Part B 68 (2014) 385–414

Since train scheduling is NP-hard it may not be possible to determine the optimal solution for large scale cases. A simulation framework may be used to obtain a local optimal or approximately optimal solution. System deadlock avoidance is a very high priority constraint because a single deadlock event may quickly replicate though the entire rail network.

**Development of hybrid optimization of train schedules model for N-track rail corridors**

Hamed Pouryousef a,⇑, Pasi Lautala a,1, David Watkins b,2

Transportation Research Part C 67 (2016) 169–192

There are two objectives for efficient management of a rail line: avoiding schedule conflicts; and balancing the maximization of line use with level of service.

**A disjunctive graph model and framework for constructing new train schedules**

R.L. Burdett, E. Kozan

European Journal of Operational Research 200 (2010) 85–98

Train schedules may be made more accurate through the use of a job shop approach.

**Fuzzy optimal schedule of high speed train operation to minimize energy consumption with uncertain delays and driver’s behavioral response**

A.P. Cucala n, A.Ferna´ ndez, C.Sicre,M.Domı´nguez

Engineering Applications of Artificial Intelligence 25 (2012) 1548–1557

An economic goal of train scheduling is minimizing energy consumption by the optimal distribution of

the available slack time among all the stretches of the service taking into consideration punctuality constraints at every station. This may be accomplished by Fuzzy Mathematical Programming combined with Genetic Algorithms and Simulation.

**A Model to Simulate Delay in Train Schedule Caused by Crowded Passengers: Using a Time–Space Network**

Shigeki Toriumi1, Azuma Taguchi1, and Tetsuro Matsumoto

International Regional Science Review

2014, Vol. 37(2) 225-244

In Tokyo, commuters concentrate in commuting during short time period, traveling radial lines directing to the central district, and choosing a ‘‘better (higher speed)’’ train service. Commuters’ trip distance is much longer than that of one-time ticket users. Most of the commuters start their trips from seven to eight o’clock. During these hours, trains as well as stations are highly crowded with commuters. Almost all commuters move from the suburbs to the central district of Tokyo, which leads to concentration in train services along several radial lines. Many commuters care not so much congestion rather than trip time, and therefore when both local and express train services are available, they choose the latter first to make them crowded.

# Dataset

SEPTA published historical train data at <https://www.kaggle.com/septa/on-time-performance>. The data definition is as follows.

otp.csv

* train\_id
* direction *('N' or 'S' direction is demarcated as either Northbound or Southbound)*[*^1*](https://www.kaggle.com/forums/f/1300/septa-regional-rail/t/21259/why-are-all-trains-originating-in-thorndale-marked-as-southbound)
* origin *(See map below - you'll see 'Warminster', 'Glenside',...'Airport Terminal..')*
* next\_station *(Think of this as the station stop, at timeStamp)*
* date
* status *('On Time', '5 min', '10 min'. This is a status on train lateness. 999 is a suspended train)*
* timeStamp

# Approach

The general approach is illustrated below.

## Step 1: Obtain data set from SEPTA

As noted above, SEPTA published on time performance data on Kaggle.com.

## Step 2: Load data set into database

I used Microsoft SQL Server Integrated Services to load the data set into a Microsoft SQL Server database.

## Step 3: Generate numeric lateness attribute from descriptive attribute

Details will be published at https://github.com/jeffreypage/TRAM.

## Step 4: Analysis numeric lateness attribute

Details will be published at https://github.com/jeffreypage/TRAM.

# Results

The following table summarizes the frequency of occurrence of late time in minutes.

|  |  |
| --- | --- |
| Late time | Frequency |
| 0 | 460275 |
| 1 | 179854 |
| 2 | 148978 |
| 3 | 112485 |
| 4 | 86905 |
| 5 | 65118 |
| 6 | 49287 |
| 7 | 37649 |
| 8 | 28820 |
| 9 | 23183 |
| 10 | 18875 |
| 11 | 15474 |
| 12 | 12910 |
| 13 | 10856 |
| 14 | 9166 |
| 15 | 8351 |
| 16 | 6688 |
| 17 | 5977 |
| 18 | 5023 |
| 19 | 4428 |
| 20 | 3864 |
| 21 | 3146 |
| 22 | 2855 |
| 23 | 2759 |
| 24 | 2015 |
| 25 | 1975 |
| 26 | 1562 |
| 27 | 1314 |
| 28 | 1348 |
| 29 | 1167 |
| 30 | 1036 |
| 31 | 845 |
| 32 | 837 |
| 33 | 693 |
| 34 | 680 |
| 35 | 543 |
| 36 | 477 |
| 37 | 452 |
| 38 | 344 |
| 39 | 384 |
| 40 | 355 |
| 41 | 334 |
| 42 | 221 |
| 43 | 242 |
| 44 | 204 |
| 45 | 167 |
| 46 | 164 |
| 47 | 138 |
| 48 | 149 |
| 49 | 120 |
| 50 | 136 |
| 51 | 87 |
| 52 | 103 |
| 53 | 67 |
| 54 | 61 |
| 55 | 49 |
| 56 | 40 |
| 57 | 48 |
| 58 | 37 |
| 59 | 56 |
| 60 | 104 |
| 61 | 30 |
| 62 | 52 |
| 63 | 31 |
| 64 | 33 |
| 65 | 31 |
| 66 | 29 |
| 67 | 48 |
| 68 | 19 |
| 69 | 17 |
| 70 | 23 |
| 71 | 8 |
| 72 | 13 |
| 73 | 15 |
| 74 | 26 |
| 75 | 11 |
| 76 | 16 |
| 77 | 32 |
| 78 | 9 |
| 79 | 21 |
| 80 | 4 |
| 81 | 7 |
| 82 | 19 |
| 83 | 7 |
| 84 | 4 |
| 85 | 13 |
| 86 | 6 |
| 87 | 13 |
| 88 | 4 |
| 89 | 5 |
| 90 | 10 |
| 91 | 4 |
| 92 | 2 |
| 93 | 2 |
| 95 | 4 |
| 97 | 3 |
| 98 | 3 |
| 99 | 1 |
| 100 | 1 |
| 102 | 1 |
| 103 | 2 |
| 104 | 2 |
| 107 | 1 |
| 116 | 1 |
| 119 | 1 |
| 124 | 2 |
| 999 | 390 |
| 1440 | 1 |

A value of 0 indicates the train was on time. A value of 999 or 1440 probably indicates a catastrophic system failure resulting in the failure of the train to complete its route. This may be due to a significant reduction in service when a significant portion of the rolling stock was taken out of service due to mechanical defects.

Sample Size: 1322457 Train/station lateness measurements

Sample median: 2 [minutes late]

Sample mean: 3 [minutes late) The mean was skewed by the outliers (incomplete trains).

Sample Standard Deviation: 18.04

Confidence Interval: 3 minutes +/- 0.03 minutes

# Conclusions

The overall results indicate SEPTA generally met its goal of being within 6 minutes of the schedule. Clearly the vast majority of measured arrivals were reasonably on time. Despite that, it would be advisable for management to address the small number of cases that significantly exceeded 6 minutes in order to assure customer satisfaction.