kaggle   Search kaggle   Q   **Competitions**   **Datasets**   **Kernels**   **Discussion**   **Jobs**

RTA Freeway Travel Time Prediction

This competition requires participants to predict travel time on Sydney's M4
freeway from past travel time observations.

$10,000   ·   356 teams   ·   7 years ago

Overview   **Data**   Discussion   Leaderboard   Rules

Competition Data   Edit

⊞ RouteLengthApprox.cs...	**m4-map.pdf** 385.91 KB	⬇ Download
⊞ RTAData.csv		
⊞ RTAError.csv		
⊞ RTAHistorical.csv		
⊞ sampleEntry.csv		
◻ m4-map.pdf		

Data Description

The NSW Roads and Traffic Authority (RTA) has made available several years' worth of historical travel time data
for Sydney's M4 freeway. The data are collected from loops on the road at three minute intervals.

The file is formated as follows

	40010	40015	40020	...
1/03/10 15:01	804	209	804	248
1/03/10 15:04	892	212	801	237
1/03/10 15:07	857	214	821	243
...	849	222	834	252

The header row shows the route IDs and the first column has timestamps. The cells show the travel time in
deciseconds.

Route IDs are ordered sequentially. Routes 40010-40150 are westbound travel times and 41010-41160 are eastbound travel times. See m4-map.pdf for an approximate guide to the freeway's layout.

RTAData.csv holds the travel time dataset. Complete data is provided from March 2010 to July 2010. For the data from August 2010 to mid-November 2010, the letter "x" appears whenever a prediction is required.

29 cut-off times have been selected. After those cut-off times, predictions must be made for the next 15 minutes, 30 minutes, 45 minutes, 1 hour, 90 minutes, 2 hours, 6 hours, 12 hours, 18 hours and 24 hours. To prevent particicipants from meaningfully using the future to predict the present, data are not revealved again until 18 hours after the last forecast has been made (or 42 hours after the cut-off point).

Participants are required to provide 290 forecasts (10 for each cut-off point) for the 61 routes. The route IDs should appear in a header row, and the timestamps in the first column. The file sampleEntry.csv is an example entry, showing how entries should be formatted.

The RTA has also provided some additional historical data, RTAHistorical.csv (collected using their old system), which covers November 2008 to February 2010. The data are consistent with the newer RTA data except that:
- they don't cover weekends; and
- there are no data for routes 40092 and 41140.

Update 17 December: The file RTAError.csv shows where loop readings are suspected to be inaccurate. Loop readings can be inaccurate because the loop is behaving erratically or because the loop is not responding at all (in which case the average travel time from adjacent loop combinations have been used).

The file RTAError.csv shows the proportion of loops in a route that have failed (a route is made up of many loops). A reading of 0 means that all loops are functioning properly, a reading of 0.5 means that 50 per cent of the loops are functioning and a reading of 1 means that no loops are functioning.

The file RouteLengthApprox.csv shows approximate route lengths. It is calculated as the number of loops in the route multiplied by 500m (the approximate distance between loops).

Our Team   Terms   Privacy   Contact/Support