Architecture for the Modeling and Analysis of Rapid Transit

A-MART

Minor Appendix Q
Regression Analysis.

David Claypool
Mahesh Balakrishna
Kimberly Baumgartner
Yimin Zhang

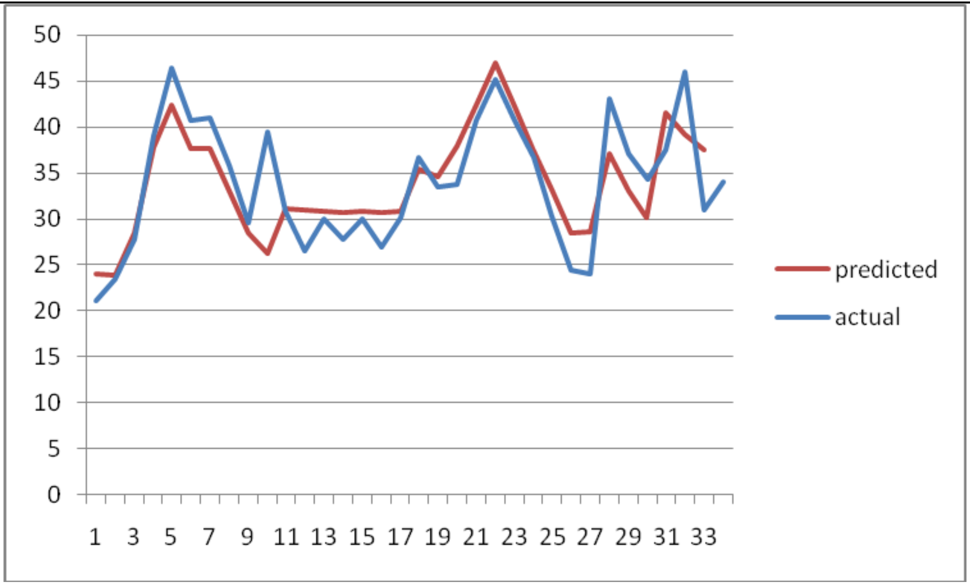
SYST 798/OR 680

The first task of our analysis was to determine which parameters were correlated to Platform crowding. The idea behind this was to use regression. If we could successfully find a regression that would cover the majority of the variance in station platform levels, than we could do 2 things. First off, we could determine which parameters contributed to station crowding and conduct further analysis of those parameters to discover patterns and optimize performance of those parameters to ultimately reduce station crowding. Secondly, the regression would yield a method to predict station crowding based on the status of those parameters deemed as dependent.

In order to accomplish the regression, we first had to acquire necessary data for the parameters for which we wanted to explore and transform them into a usable form for the regression. To begin, we set up a few test runs using the Rosslyn model. The inputs included set pedestrain passenger flow rates into the station and train arrivals with passengers remaining on the train, exiting the train and leaving the station and exiting the train to wait for a different train. The outputs collected from the model run included queue lengths at various points in the staion, average system time, average wait time at the queues, and of course the number of riders on the platform at any given time. After the first run we began to assimilate the data and realized that much of it was not in the same form and that we had to transform some of the parameters to make use of them. Some of our parameters were measured at specific points in time such as queue lengths and number of personnel on the platform, each of which were measured at every minute interval. However others events did not happen on specific 60 second intervals such as train arrivals. In order to establish each degree of freedom we decided to take averages of the output data over periods of time. For instance, the average of each parameter over a 30 min period would mean the average passenger flow rate, the average train interarrival time, the average number of open seats on trains arriving etc. Each 30 minute average of the parameters would account for 1 degree of freedom. We also saw from the raw outputs that at many times the queue lengths at individual station elements such as turnstyles was 0 or 1, so we eliminated these as being correlated to platform crowding under normal conditions. What we were left with was train interravial time, passenger flows into the station, and the open capcity (number of open seats) on arriving trains. Each of these parameters are inputs to the model, which allowed us to perform our simulations under various conditions and measure system performance in terms of number of passengers on the platform. We then took the average of these parameters and the average of the number of personnel on the platform at any given time in half horu intervals.

Our first real analysis included running the simulation over a virtual “metro day” at Rosslyn. This included using the actual train schedules for Rosslyn and simulated passengers flows throughout the day to account for increases during Rush hours etc.

From this first attempt to find a regression that fit, our best fit had an R-squared value of .62 which means that the regression only accounted for 62% of the variance in platform crowding from those parameters. The below graph shows the initial results of the regression. The Blue line represents the actual platform crowding numbers while the red line is the predicted numbers based on the regression analysis.



As you can see, the predicted follows the actuals rather closely accept for one spike. This spike happens to occur at the end of a rushour. If we go back and look at our input data, we see that this is at the end of the morning rushour when the train schedule switches from approximately a 5 minute inter-arrival time to a 12 minute inter-arrival time. During this time, the station is still dealing with residual rushour riders waiting for trains, which led us to the hypothesis that there were two different relationships that accounted for the platform crowding and that the relationship changes when there are more passengers arriving at the station than can be accomodated. To test this theory we re-engineered the test to incorporate multiple hour long runs under various conditions. Below is a summary of the runs completed.

Interarrival of trains	Ped Flow	Open seats on trains	# Pax on Platform
5	4	200	12.852
5	4	150	12.992
5	4	100	13.164
5	4	50	13.459
5	8	200	22.226
5	8	150	22.672
5	8	100	22.95
5	8	50	22.885
5	16	200	42.197
5	16	150	42.223
5	16	100	42.77
5	16	50	42.918
10	4	200	21.787
10	4	150	22.623
10	4	100	23.164
10	4	50	23.344

10	8	200	40.746
10	8	150	40.443
10	8	100	40.508
10	8	50	41.164
10	16	200	78.967
10	16	150	79.049
10	16	100	81.393
10	16	50	228.694

The first column includes inter-arrival of trains at 5 minutes (rush hour) and 10 minutes (non-rush hour), the second is the pedestrian arrival rate per minute into the station. The third column are the number of open seats on each arriving train. These three parameters were the inputs into the model. The fourth column is the average number of riders on the platform for each hour-long run of the model. The raw regression of these outputs yielded only a .55 R-Square. However this was expected. If you look at the last 2 rows, you will see that for the 10 minute inter-arrival time with high (16/min) pedestrian arrival rate and decreasing availale seats, that the passenger crowding jumps up. This is because at approximately 125 open seats per arriving train, the station experiences saturation – in that more passengers are arriving than can be taken away, therefore over an hours time, the total crowding steadily increases rather than behaving in a normal pattern with an stable average. By eliminating these two rows and perfroming regression analysis we get an R-square of .92. Below is the graph showing the results of the predicted verses actual platform crowding.