Bike hire schemes

Over past few years, city bicycle hire schemes have emerged as a popular solution to the last mile problem. In contrast to many other transport providers, cycle hire operators are publishing a lot of operational data about the schemes, which has enabled a large community of researchers to study the dynamics of the systems. Similarly to tradition transport providers, cycle hire operators face the following challenges

- Infrastructure planning
- Pricing
- Policy improvement

However, advances in mobile technologies have created yet another challenge, namely accurate near-term predictions for station occupancy, i.e.

- Bicycle availability at starting stations
- Parking space availability at destinations

since this information is crucial if cycle hire is considered as an option in a multi-modal end-to-end routing planner.

Can origin–destination information improve forecasts?

Yoon and Kaltenbrunner [1, 2] have shown that useful station occupancy can be made using only raw station departure and arrival time series data. In [3] we investigate whether additional origin–destination journey data has the potential to improve such forecasts. To do that, we study how extra journey information can help to predict future arrivals for small groups of neighbouring stations during rush hour.

Training models on journey data

In [3] we compare the arrival forecasts produced by two types of models

- Time-inhomogeneous Population CTMC (IPCTMC)
- Multiple linear regression with ARIMA error (LRA)

Extra journey information is incorporated through

1. Journey time distribution clustering of stations
2. Knowledge about journey destinations at departure time

Option 1 (see info box) is a natural way to fit training journey information in IPCTMC models. 2 was inspired by the observation made by Côme [4] et al., who found that most rush hour journeys are made by subscribed users. To analyse 1 and 2 we trained our models using station. The top diagram shows the RMSE forecast errors obtained for different journey data based models divided by the RMSE forecast errors of corresponding models that were trained on simple station departure, arrival time series data. Clearly, the shorter the forecast interval the more vital the extra journey data becomes. The bottom diagram shows the actual 15-min interval arrival forecasts for the 22/06/2012.

Results

Both diagrams evaluate the quality of different forecast models for the afternoon rush hour for a group of 10 stations around Waterloo station. The top diagram shows the RMSE forecast errors obtained for different journey data based models divided by the RMSE forecast errors of corresponding models that were trained on simple station departure, arrival time series data.

Conclusions

Using journey data in prediction models visibly improves 5 – 15 minute forecast quality. For longer forecast intervals, this is not the case as it is harder to predict future cluster departure rates and journey destinations. Alternative ways of improving their forecasts should thus be explored.

References


