Xiaolong LI, Gang PAN, Zhaohui WU, Guande QI, Shijian LI, Daqing ZHANG, Wangsheng ZHANG, Zonghui WANG

Prediction of urban human mobility using large-scale taxi traces and its applications

Summary

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.438.4632&rep=rep1&type=pdf

Goal

- Predict taxi pick-ups quantity from urban hotspots in Hangzhou (China)

Data

- GPS data of 4000 taxis in the city during 385 days
- Every 60 seconds sample of taxi:
  - id, time, gps position, speed, orientation, gps state (ok/nok), meter state (occupied/vacant)
- Derived trajectories from former data
- City was devided to pick-up hotspot
  - first they divided city to rough 10x10m blocks, filtered out blocks not frequent enough and finally, they used adaptive watershed algorithm to make blocks smaller
- Day devided into various segment with uniform duration (20 min, 1 h, 3 h...)

Used methods

Naive - demand will be same as in the same time of previous day

Bayesian networks -

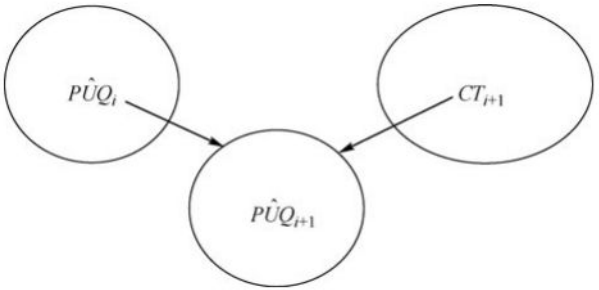
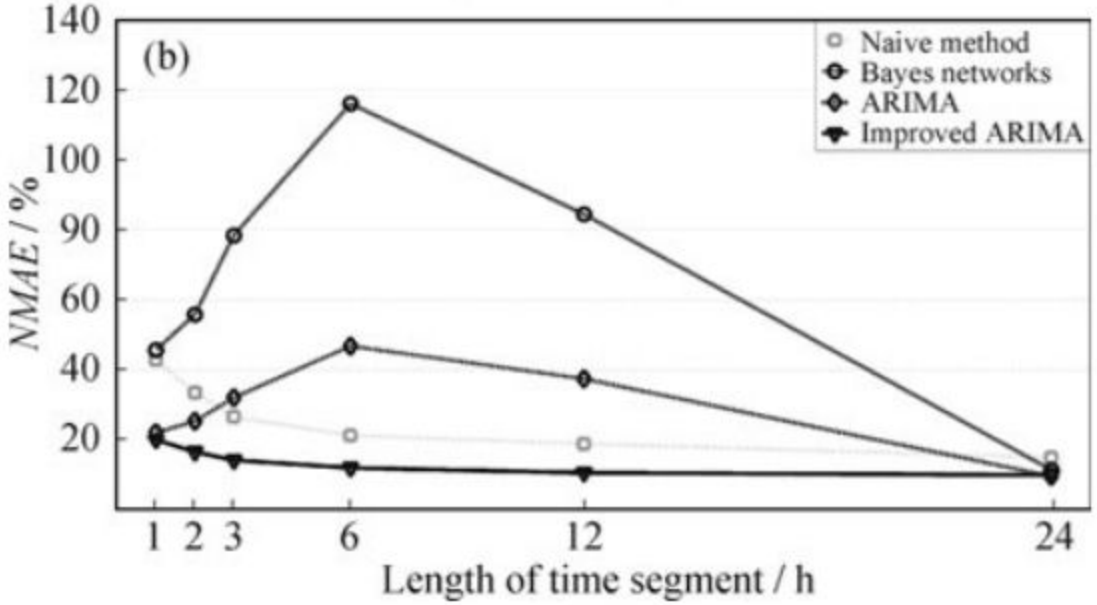


Fig. 3 Bayesian networks for prediction of PUQ

Auto-regressive integrated moving average (ARIMA) - considers only nearest historical data of hotsports

Improved ARIMA - adds periodical data from previous days to ARIMA

Results



NMAE = Normalized mean absolute error

Evaluation

They’ve deduced some interesting equations for measuring the overall time and distance needed for taxis and compared real taxi routes with routes suggested by them

Table 1 Improvement in time cost and driving distance cost by prediction guide

	Real world data	ARIMA	Improvement/%	Improved ARIMA	Improvement/%
Time/s	525.2	396.9	24.4	330.2	37.1
Distance/m	1 777.2	1 613.6	9.2	1 664.3	6.4