Overview

- Why do we want to predict bus journeys (#passengers)?
  - Design **new route** and know whether its profitable
  - Existing route, change a factor (e.g. frequency)
  - Create input for a wider network
- How will we do it?
  - Direct method for bus corridors
  - Decide about data & factors
    - Especially catchment area
  - Analyse and visualise factors
  - Use typical **Machine Learning techniques**
  - Get an idea about accuracy
- What did we find?

**Abstract**

The catchment area along a bus route is key in predicting bus journeys. In particular, the aggregated number of households within the catchment area are used in the prediction model. The model uses other factors, such as head-way, day-of-week and others. The focus of this study was to classify types of catchment areas and analyse the impact of varying their sizes on the quality of predicting the number of bus passengers. Machine Learning techniques: Random Forest, Neural Networks and C5.0 Decision Trees, were compared regarding solution quality of predictions. The study discusses the sensitivity of catchment area size variations. Bus routes in the county Surrey in the United Kingdom were used to test the quality of the methods. The findings show that the quality of predicting bus journeys depends on the size of the catchment area.
The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi

Background
▪ University of Surrey
▪ Surrey County Council
▪ Bus operators

Knowledge Transfer Partnership
9935

www.BusAnalytics.uk

Content
▪ Background
▪ Data & Factors
▪ Methods
▪ Results & Insights
Factors – simple model

**Passengers:** Average number of journeys during period

**Average Deviation** – deviation from schedule in seconds averaged according to the model defined time segment

**Average Headway** – actual headway averaged according to the model defined time segment

**Average Fare** – average of all above-zero payments

**Delivery Points** – postal domestic delivery points within $x$ meters (e.g. 800m) of the route

**Number of Hospitals** – number of hospitals within 800 meters of the route

**Number of Train-Stations** – number of rail stations within 800 meters of the route

**Day Group** – day classification as weekday and weekend

**Peak / off-peak** – Time of the day classification as peak (07:00-17:59) and off-peak (rest)

---

**Topology**

Legend:

- Bus routes ... blue lines
- Bus stops ... blue symbols
- Households (delivery points) ... red dots
- Railways ... black lines

---

The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi

OR61, Kent, September, 2019
Data (anonymised)

~ 100 to 200 routes / operator from several bus providers
~ 140k aggregated records for the modelling per operator
~ 2 years (on a daily basis)
~ several million passenger journeys
  peak: ~90%;
~ avg. headway: ~55 min
~ avg. deviation: ~4 min
~ avg. fare: ~£3.60

Bus corridors
Dashboard
To visualise route information quickly
Filter per route(s)
Select dates
Forecast #journeys
Identify most important routes

The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi

Content
- Background
- Data & Factors
- Methods
- Results & Insights
Methods

Machine Learning techniques + model that uses the factors

- **Decision Trees**
  - `treetc50::ct5.0(x=train,factor(train_expected),trials=1)`

- **Random Forest**
  - `rf<-randomForest(formula= passengers ~ ., data=training_data, ntree=1500, mtry=ncol(training_data)-13)`
  - `test_predicted_forest<-predict(rf, testing_data,type="response")`

- **Neural Network**

- **Input:** previously mentioned factors + varying catchment corridors

---

ANN

R – Neural network library

Feed-forward ANN

**Neuralnet**

`(passengers ~.,training_data, hidden=3, linear.output = F)`

A more systematic exploration of ANN is on its way
Content

- Background
- Data & Factors
- Methods
- Results & Insights

Correlation Matrix
Predicted versus actual

Bus service operator x
800m corridors

The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi

OR61, Kent, September, 2019

Bus service operator y

The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi

OR61, Kent, September, 2019
Catchment area

Quality comparison

#passenger error:
- |predicted – actual| / actual
- Average for each catchment distance

Max error limited to 50%
- ~140k records originally

The prediction error is lowest at a corridor distance of 400m (~5min), 2km (~24min) is similar and 1200m (~15min) is also good.
- Is the difference in quality gain neglectable?
- Why? Bus stop distance!? Why isn’t it gravitational?

12 minutes to cover 1km (or 10 to 20min)
Typical speeds: 3km/h to 6km/h;
Google Maps speed: 5km/h

Conclusion & TODOs

Initial predictions look promising

Catchment corridor prediction errors are lowest at 400m (with 0.5 max error filter)

TODOs
- Check outliers
- Rigorous statistical evaluation
- Examine, filter & investigate each factors prediction results
- Decide whether different models should be examined
- ...
Appendix

Factors for second model
Neural networks to explore

Data & Factors – detailed*

Prediction

**Passengers:** Average annual number of passengers boarding during weekdays in the hours between 7am and 6pm.

*Acknowledgement: original KTP team*
Data & Factors

Attractions in bus stop’s catchment area

- **GP**: Number of general medical service points divided by catchment area’s population size.
- **Hospital**: Number of hospitals divided by catchment area’s population size.
- **Jobs**: Average ratio of jobs per resident. Number of jobs available for LSOA area and number of residents available for postcode units.
- **School**: Total number of pupils in schools of catchment area divided by catchment area’s population size.

Attractiveness of route

- **Headway**: Average headway of departing buses between 7am and 6pm on weekdays.
- **In-vehicle time**: Ratio of estimated travel time between the terminating stops of a route for going by bus vs car.
- **Jobs along route**: Total number of jobs in route’s catchment area excluding stops less than 1km away from bus stop of interest.
- **Population along route**: Total number of people living in route’s catchment area excluding stops less than 1km away from bus stop of interest.
- **Walking distance**: Average walking distance between address points and bus stop measured by the $l_p$ norm (Brimberg, et al., 1995).
The impact of catchment areas in predicting bus journeys
Wolfgang Garn, Christopher Turner, George Kireulishvili & Vasiliki Panagi