Weak teachers: A machine learning approach for assisted specification of Discrete Choice Models

AUM
Symposia on Applied Urban Modelling

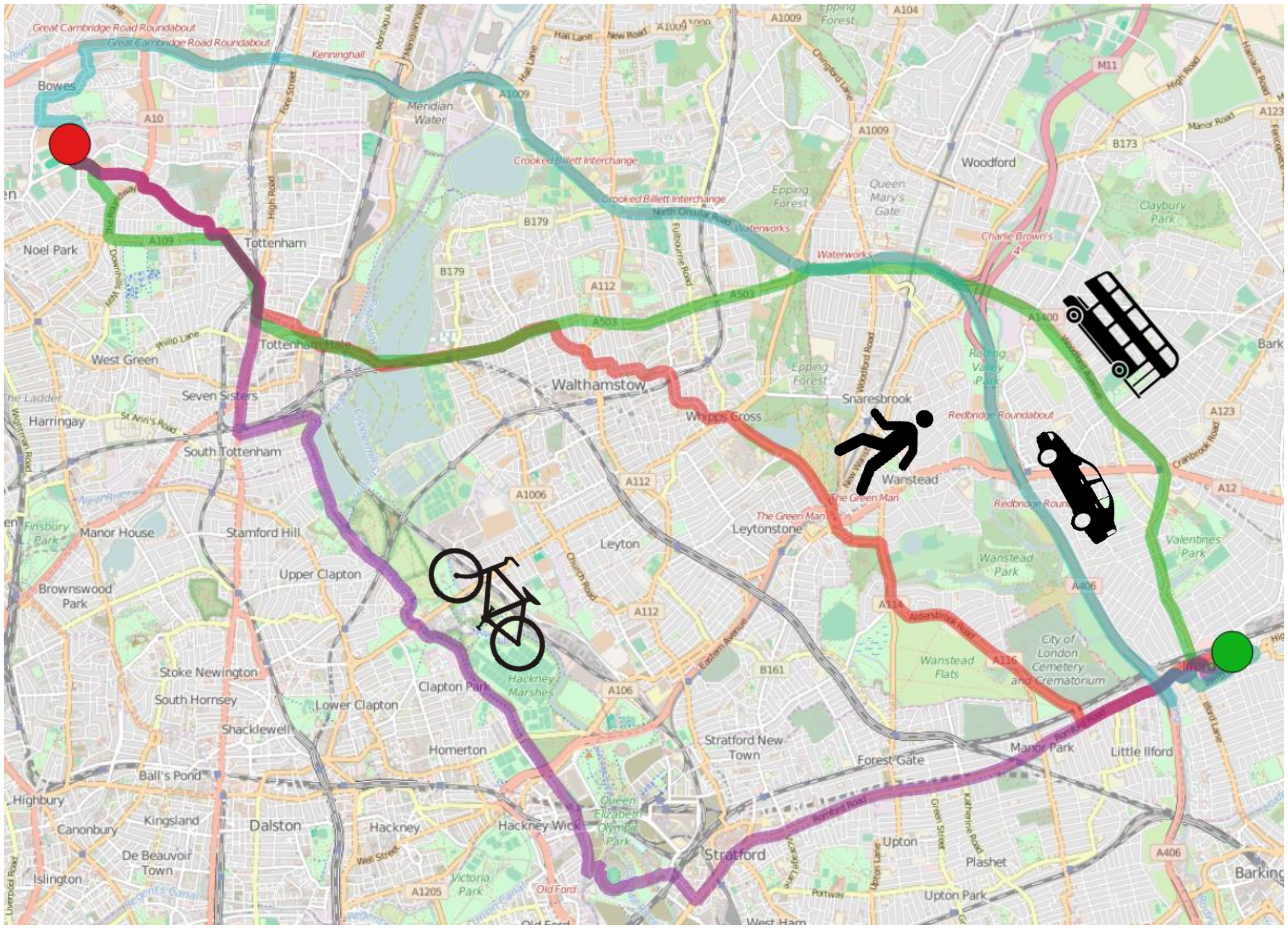
Tim Hillel, Michel Bierlaire, and Ying Jin

Transport and Mobility Laboratory TRANSP-OR
École Polytechnique Fédérale de Lausanne EPFL

TRANSP-OR

EPFL

Mode choice


Bowes
Noel Park
Tottenham
Tottenham Hale
West Green
Seven Sisters
Harringay
South Tottenham
Manor House
Stamford Hill
Upper Clapton
Brownswood Park
Stoke Newington
South Hornsey
Shacklewell
Lower Clapton
Clapton Park
Hackney Marshes
Homerton
Highbury
Canonbury
Kingsland
Dalston
Hackney
Hackney Wick
Ball's Pond
De Beauvoir Town
Islington
Victoria Park
Meridian Water
Walthamstow
Whipps Cross
Leyton
Leytonstone
Snaresbrook
Wanstead
Woodford
Claybury Park
Valentines Park
Wanstead Park
Wanstead Flats
City of London Cemetery and Crematorium
Queen Elizabeth Olympic Park
Stratford
Stratford New Town
Forest Gate
Upton
Plashet
Upton Park
Manor Park
Little Ilford
Barking

Two approaches

1. Discrete Choice Models (DCMs)
   - Aim to ***describe*** behaviour of population
     - Emphasis on *model structure*
     - Less focus on *model validation*

2. Machine Learning (ML) classifiers
   - Aim to ***predict*** unknown class for a feature vector
     - Emphasis on *model validation*
     - Less focus on *model structure*

TRANSP-OR

EPFL

DCMs

+ Highly interpretable
+ Can check for consistency with behavioural theory
- Need to manually specify utility functions

ML

- ➕ No need for manual specification
- ➕ Can automatically model non-linear interactions
- ➖ Model can not be easily interpreted
- ❓ Better predictive ability than DCMs?

Motivation

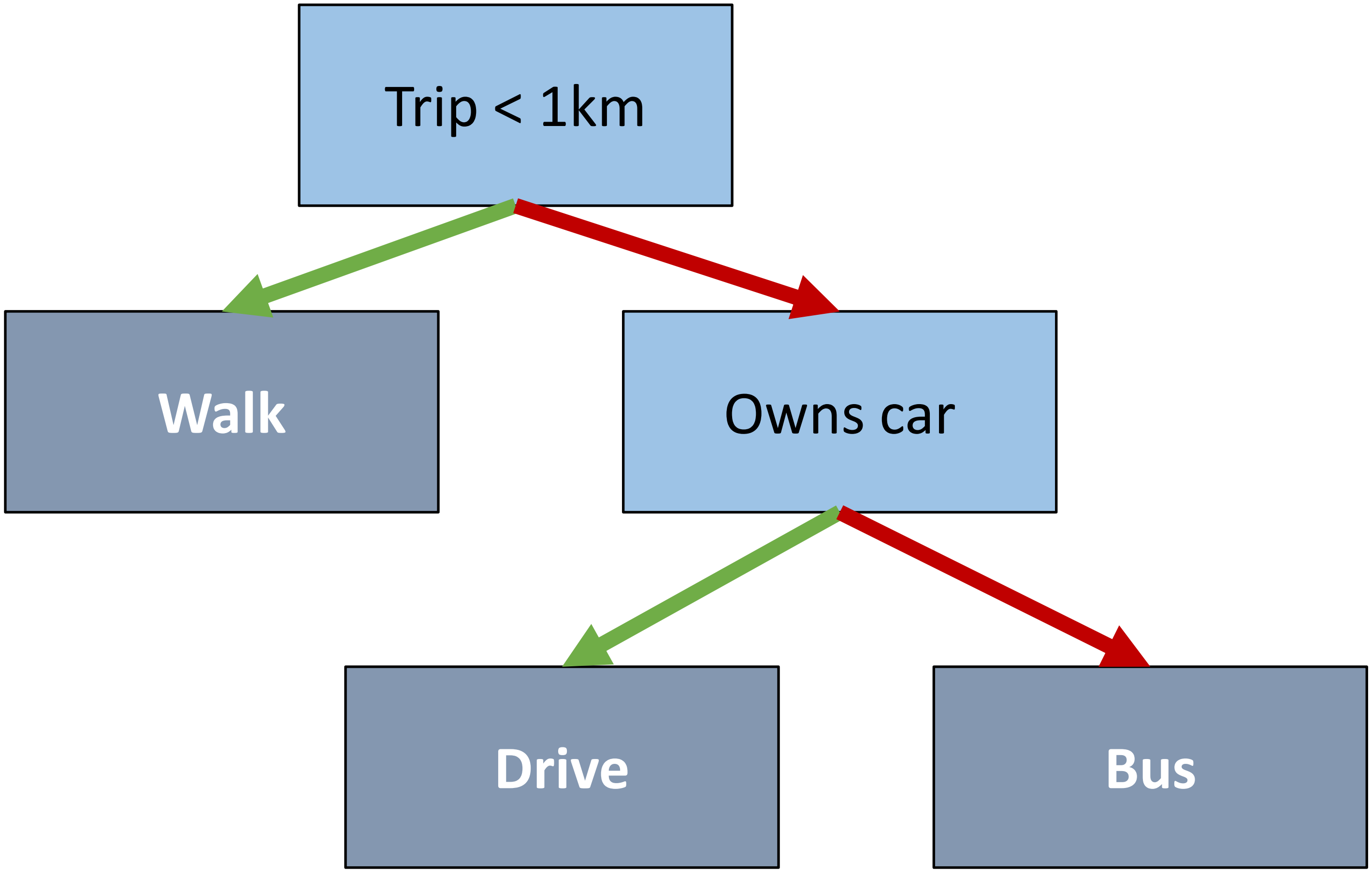
How to assist with specification of DCM utility functions…

…in order to reduce complexity of manual search…

…and improve performance of resulting classifiers?

(Use ensembles of Decision trees as weak teachers!)

Decision trees (DTs)



DTs

- Recursive hierarchical structure of *binary* splits

- To calculate split at a node:
  - For each feature:
    - Sort the data at a node over each feature
    - Calculate the **gain** in entropy for every possible split point
  - Identify split with the highest gain (across all features)
  - Repeat for new sub-nodes

- Sub-nodes model *interaction* of input features

DTs

- Splits are not sensitive to scaling/any monotonic transformations of features

- Information gain is known for each split

Ensemble learning

- Individual DTs have very high variance
  - Perform poorly in non-trivial cases
- Can be used as ***weak learners*** in ensembles of multiple DTs – exploiting wisdom of crowds
- By averaging effects of binary splits over DTs, complex non-linear relationships can be modelled
- Loss of interpretability compared to individual DT

Gradient boosting

- Fit a DT to all available data
- Subtract the predictions of the DT (multiplied by learning rate) from the data
- Repeat until stopping criteria is met

Extreme gradient boosting

- Each DT in ensemble is a *regression* tree predicting continuous value
- Passed through *softmax* (logistic) function to generate class probabilities
  - Each tree directly predicts choice probabilities

DTs as weak teachers

- Feature importances – sum gain over all splits for each features
- Feature interactions – sum gain for each hierarchical combination of n features
- Non-linear interactions of input features – investigate distribution of split values over all splits for each feature

Assisted specification approach

1. Optimise the hyper-parameters of GBDT model on (training) dataset
2. Train optimised GBDT model on the same dataset
3. Investigate structure of GDBT model, using it to inform utility specifications for DCM
4. Estimate assisted specification DCM
5. Simplify DCM by combining parameters where necessary

Methodology

- 3 years of London mode choice data
  - 2012/13-2013/14 train
  - 2014/15 test
- 100 iterations of bootstrapping to estimate model performance

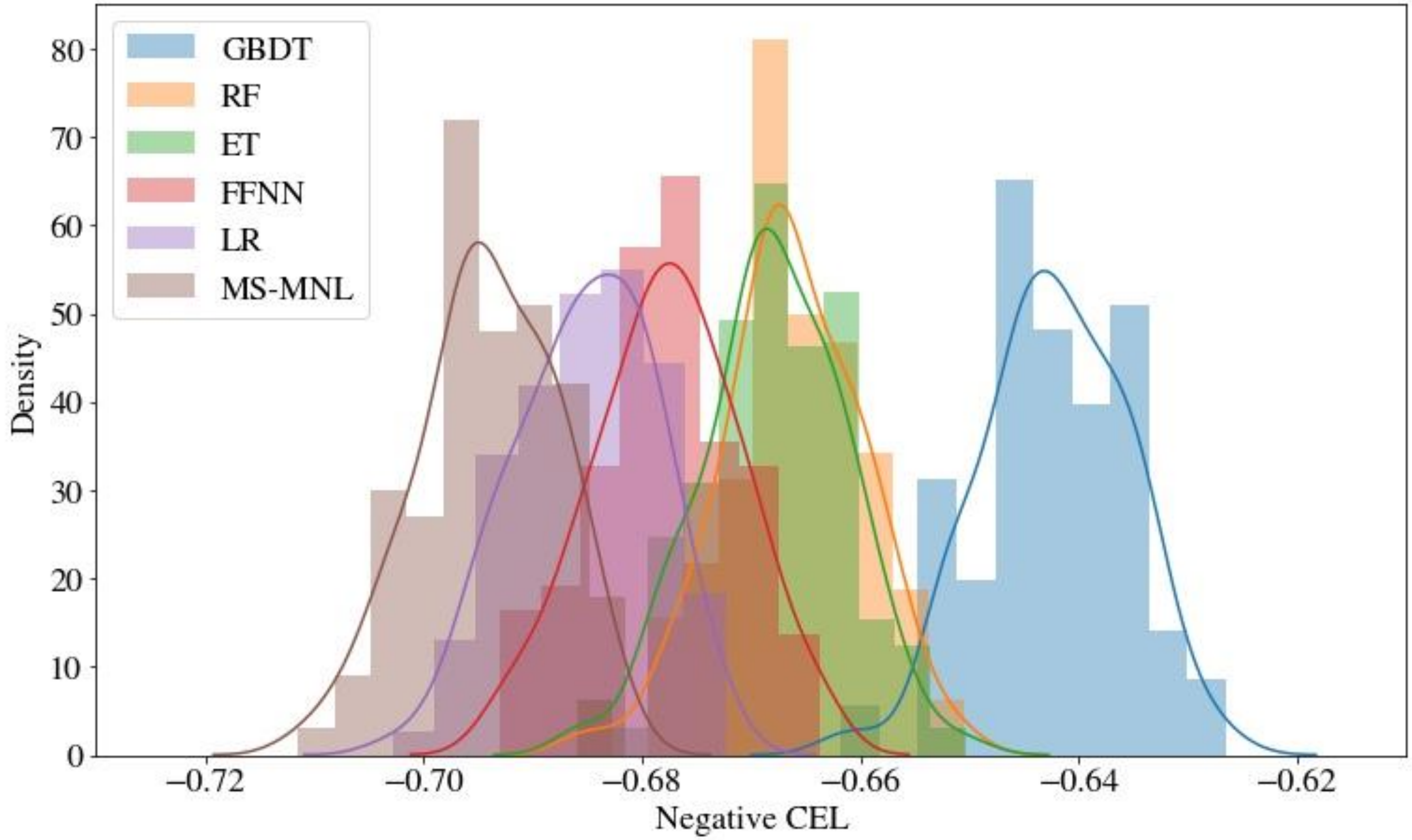
Models

- Two baseline DCMs:
  - Dummy MNL (MS-MNL)
  - ML Logistic Regression (LR)
- Four ML algorithms:
  - Gradient Boosting Decision Trees (GBDT)
  - Artificial Neural Network (ANN)
  - Random Forest (RF)
  - Extremely randomised Trees (ET)

Dummy MNL

- Estimate MNL model with all features
  - Attributes included for appropriate mode (e.g. driving duration in driving utility only)
  - No feature interactions
  - All socio-economic variables included as dummy-variables
  - A-priori bins used for age (child <18, adult 18-64, pensioner 65+) and departure time (AM peak, inter-peak, PM peak, overnight)
- Combine parameters where necessary, so that all parameters are significant
- Check parameter signs are consistent with expected values

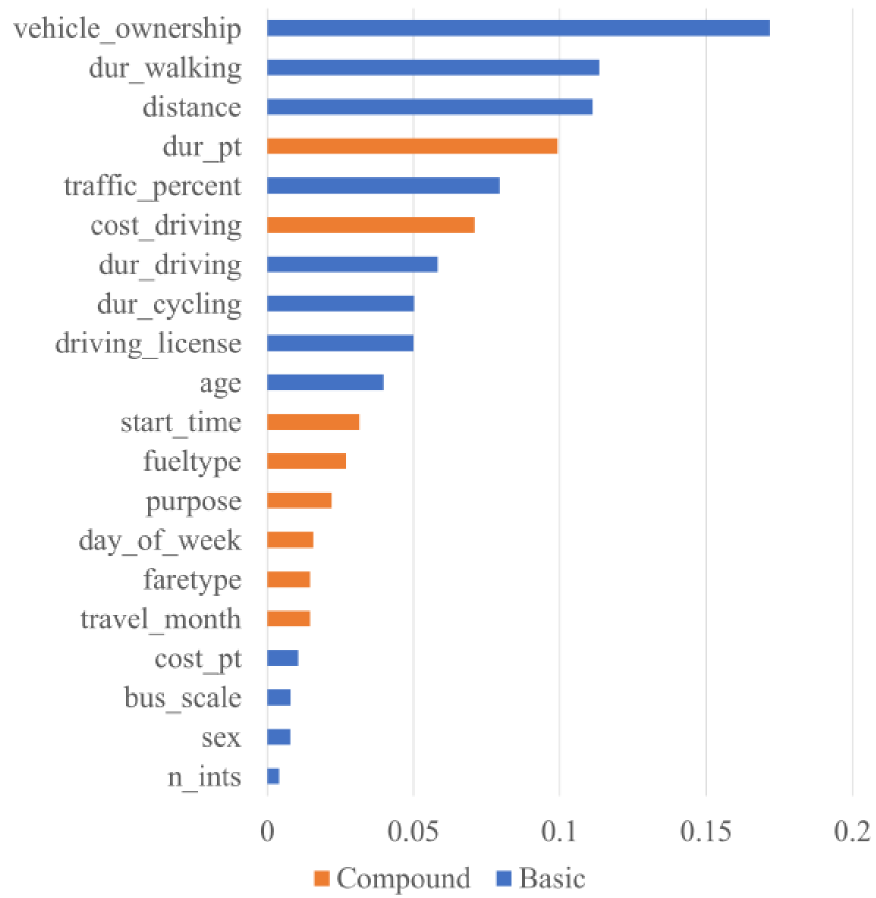
Results – benchmark models



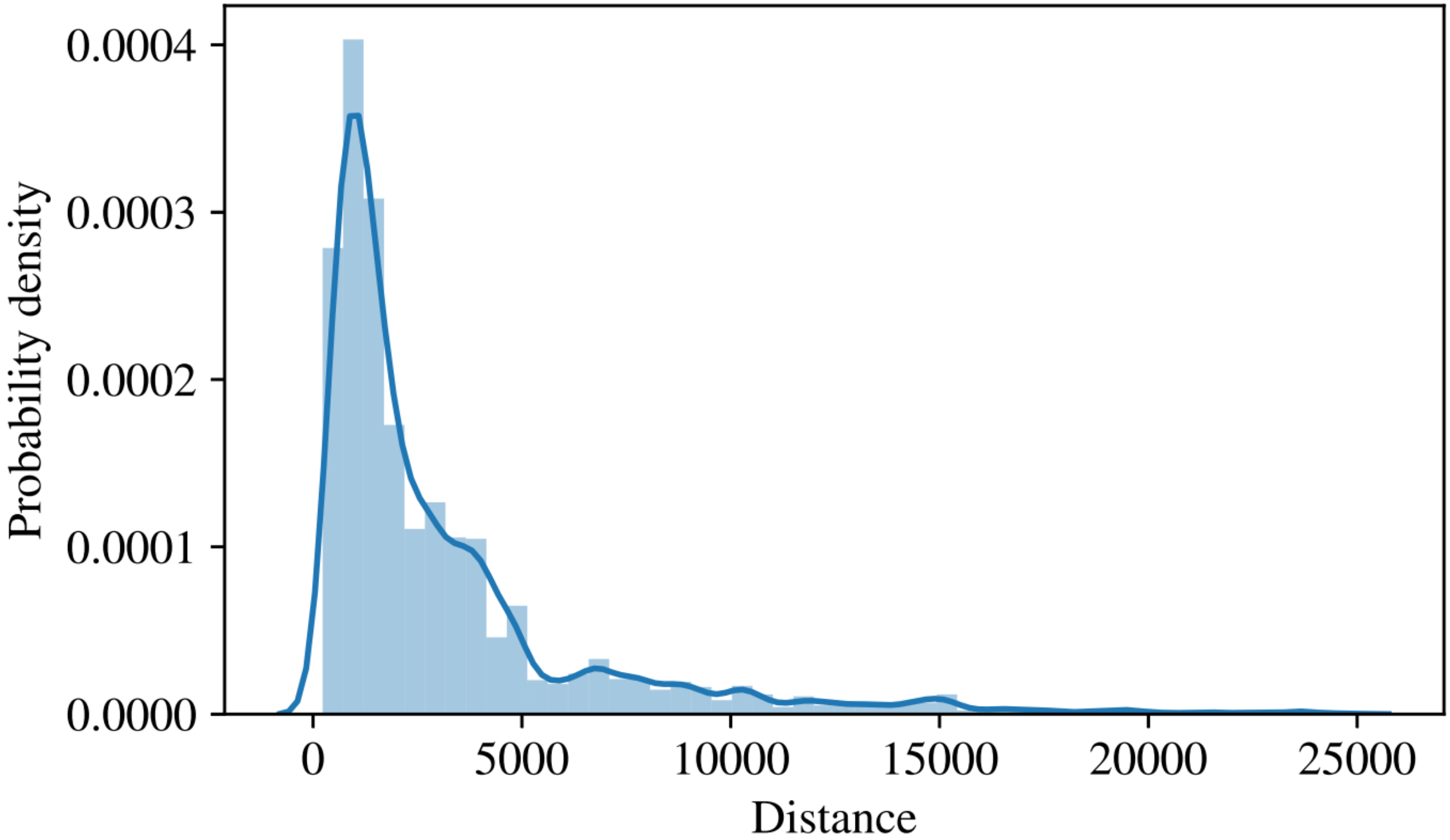
TRANSP-OR

EPFL

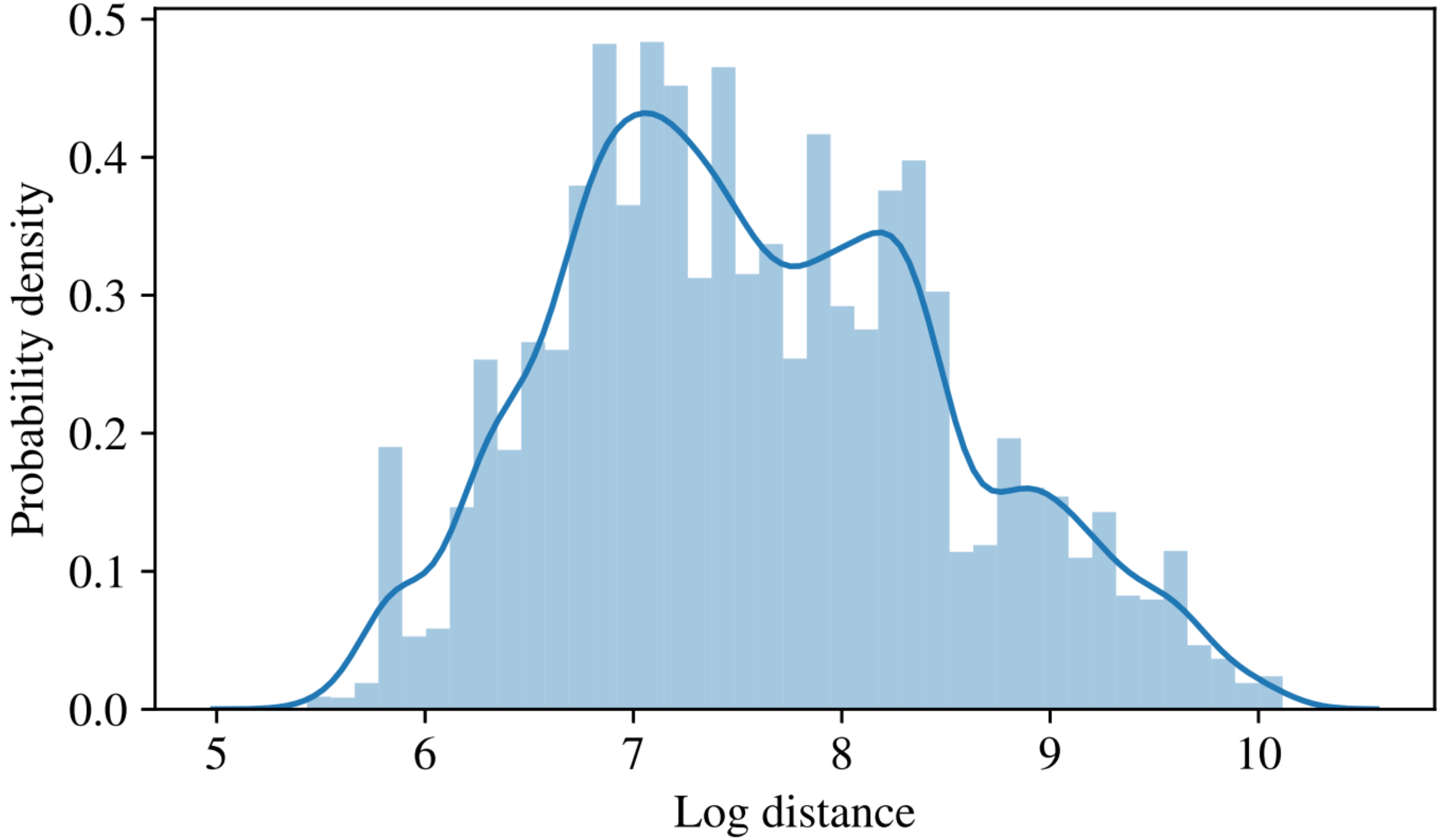
Feature importances



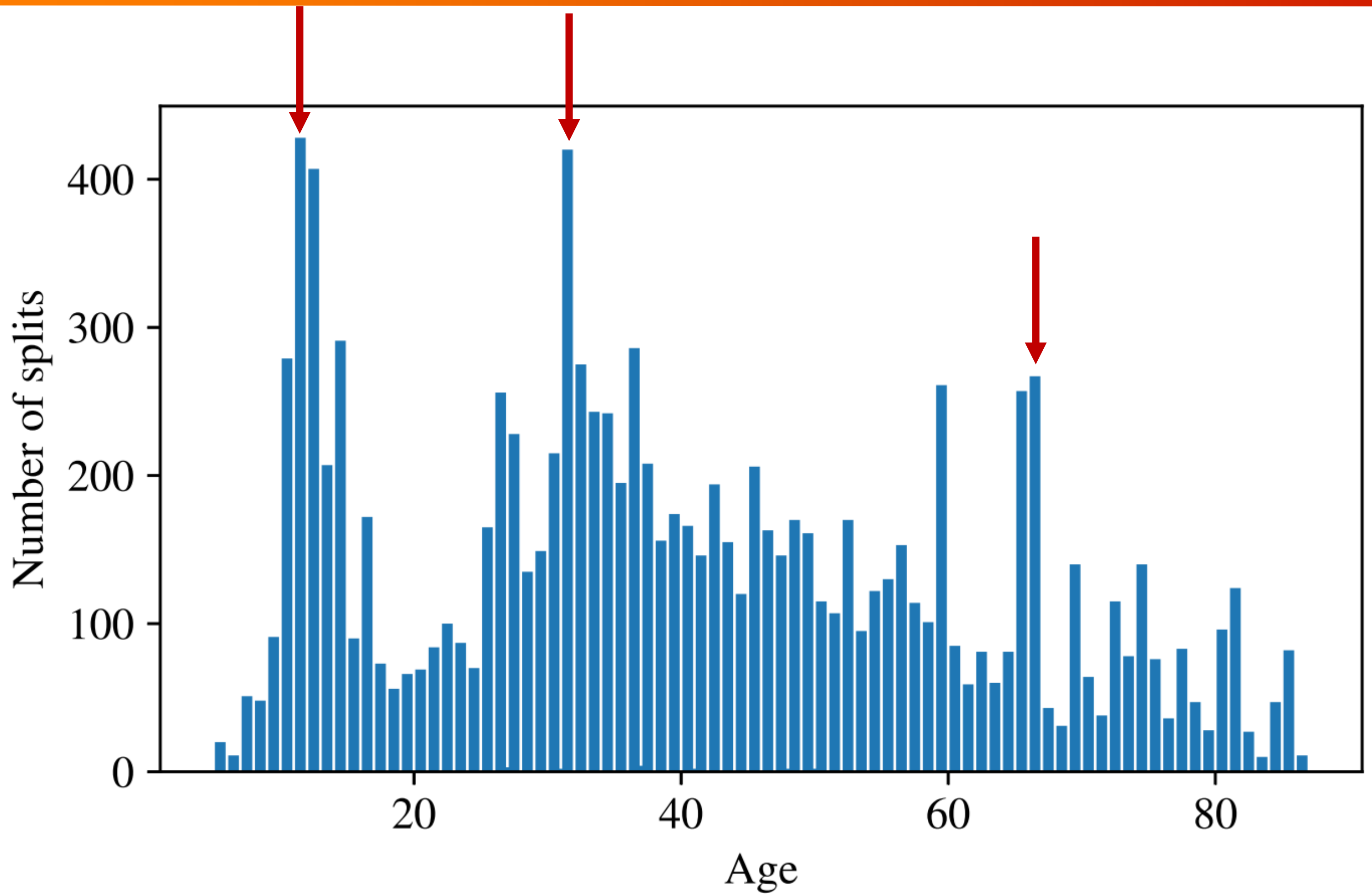
Distance split distribution



Log-distance



Age split distribution



Feature interactions

- Of 10 most important second order feature interactions, 6 include car ownership
- Most import second order is car-ownership with driving license ownership
- Most important third order feature interaction with 2 socio-economic variables contains both car-ownership and driving license ownership
- Car ownership and driving license therefore fully interacted with variables in model (6 parameters for each variable before simplification)

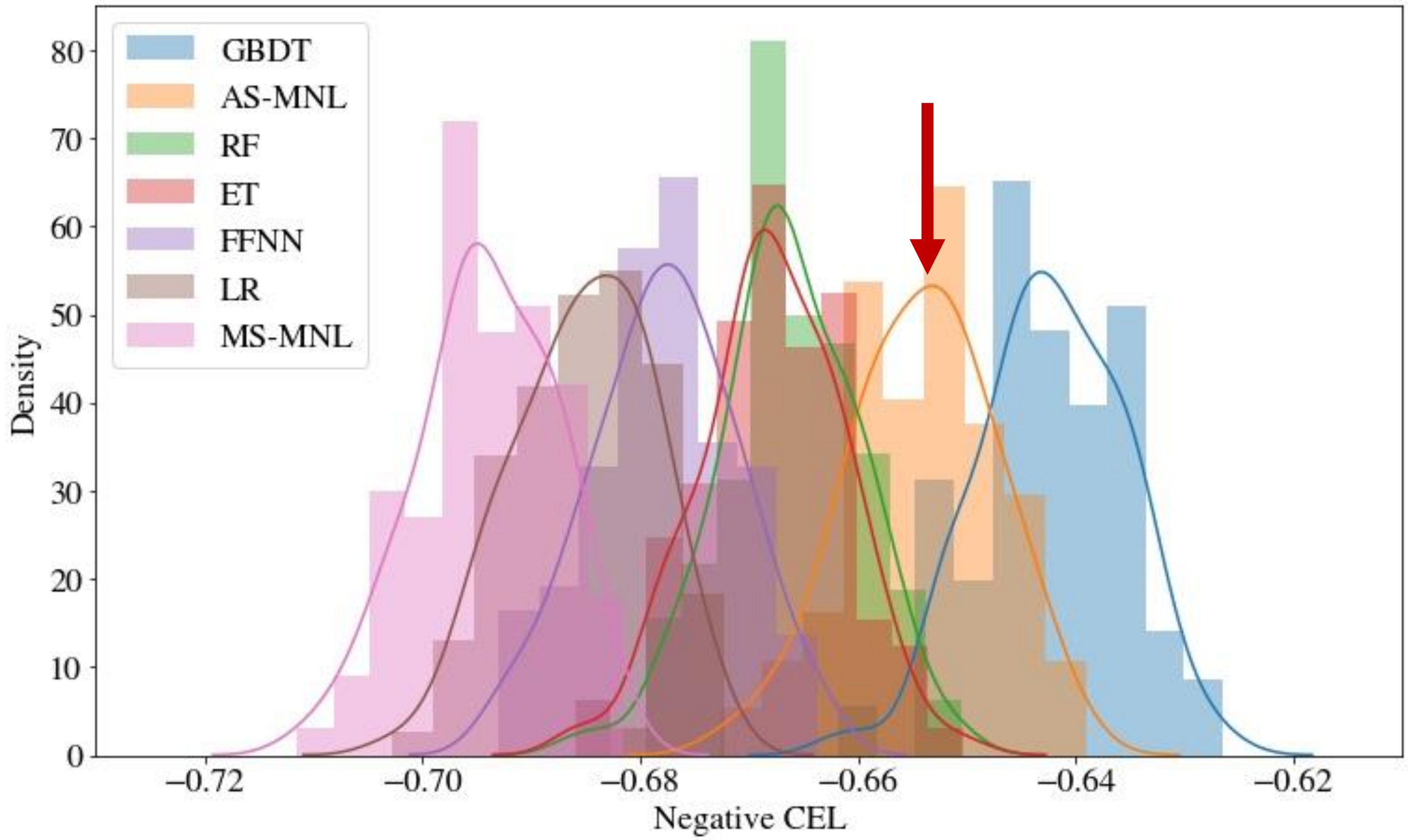
TRANSP-OR

EPFL

Final AS-MNL model

- 100 parameters
  - All significant
  - All but fuel-cost parameter for no-car ownership driving license holders have expected signs

AS-MNL results



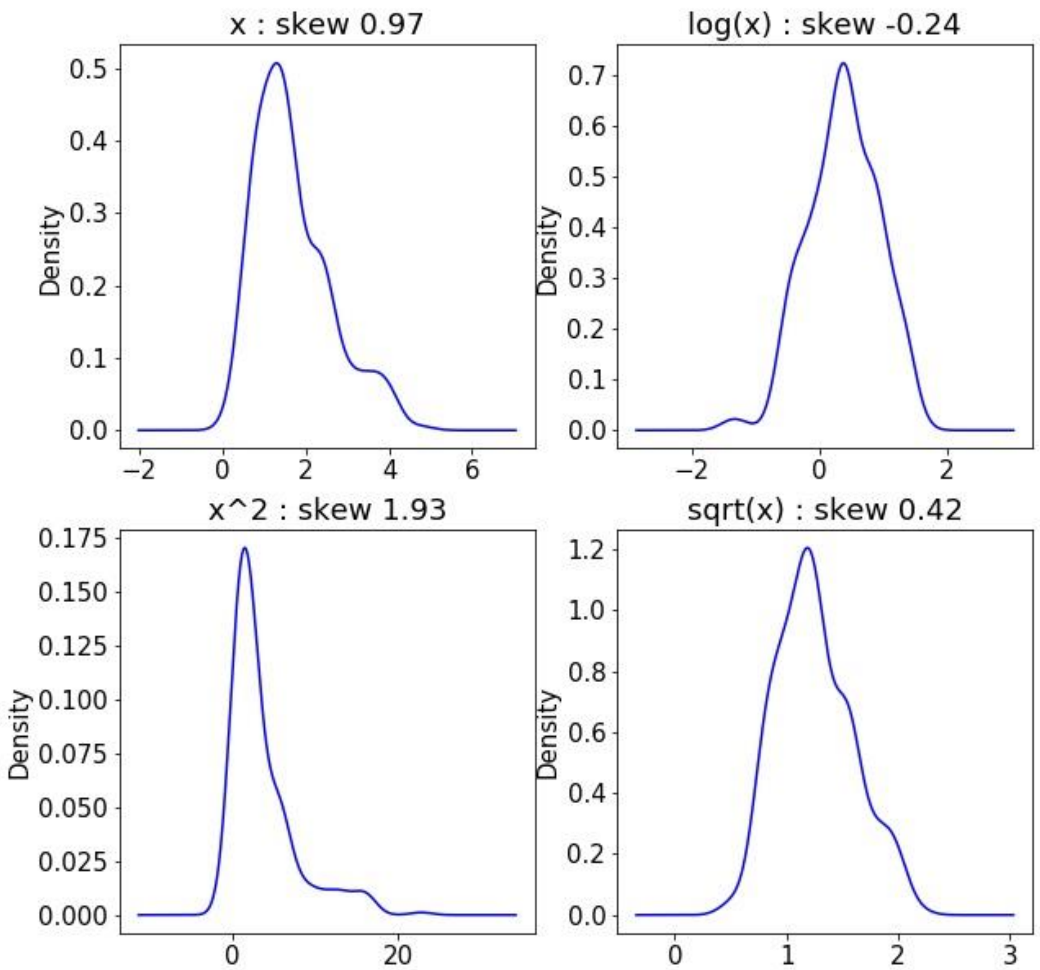
Conclusions

- Gap between ML and DCM for this problem is smaller than suggested by previous research
- AS-MNL achieves better performance than all but best ML model (GBDT)
- AS-MNL maintains interpretable linear utility specification with significant parameters

Further work

- Formalise framework into *assisted specification report*
  - User specifies alternative-specific and socio-economic variables
  - User specifies complexity/number of parameters in the model
  - Report generated suggesting which non-linear transformations/feature interactions/splines etc to include in the model
  - User can investigate suggestions using traditional model specification, retaining control over process

Parking cost for car ownership



Thank you

Hillel, Tim, Mohammed Z E B Elshafie, and Ying Jin (2018). "Recreating Passenger Mode Choice-Sets for Transport Simulation: A Case Study of London, UK". In: *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction* 171.1, pp. 29–42.

tim.hillel@epfl.ch

AUM

Symposia on Applied Urban Modelling

TRANSP-OR

EPFL