A generic data-driven sequential clustering algorithm determining activity skeletons

Wim Ectors, Bruno Kochan, Luk Knapen, Davy Janssens, Tom Bellemans

ANT 2016

universiteit hasselt | imob
TRANSPORTATION RESEARCH INSTITUTE

Table of contents

- Motivation and introduction
- Methodology
  - Data description
  - Sequential clustering algorithm
  - Sensitivity analysis
- Results
- Discussion and conclusion

Motivation and introduction

- ABM: need for transportation as derived demand from people's activity patterns
  - Mandatory (inflexible) activities scheduled before more flexible activities
  - Conventional mandatory activities: work & education
- HTS Flanders, Belgium (OVG):
  - Only 45% contains a 'mandatory' activity
  - No structure in other 55%?

→ Data-driven approach to reveal the real basic structure of individuals' schedules: skeleton schedule

Methodology – Data description

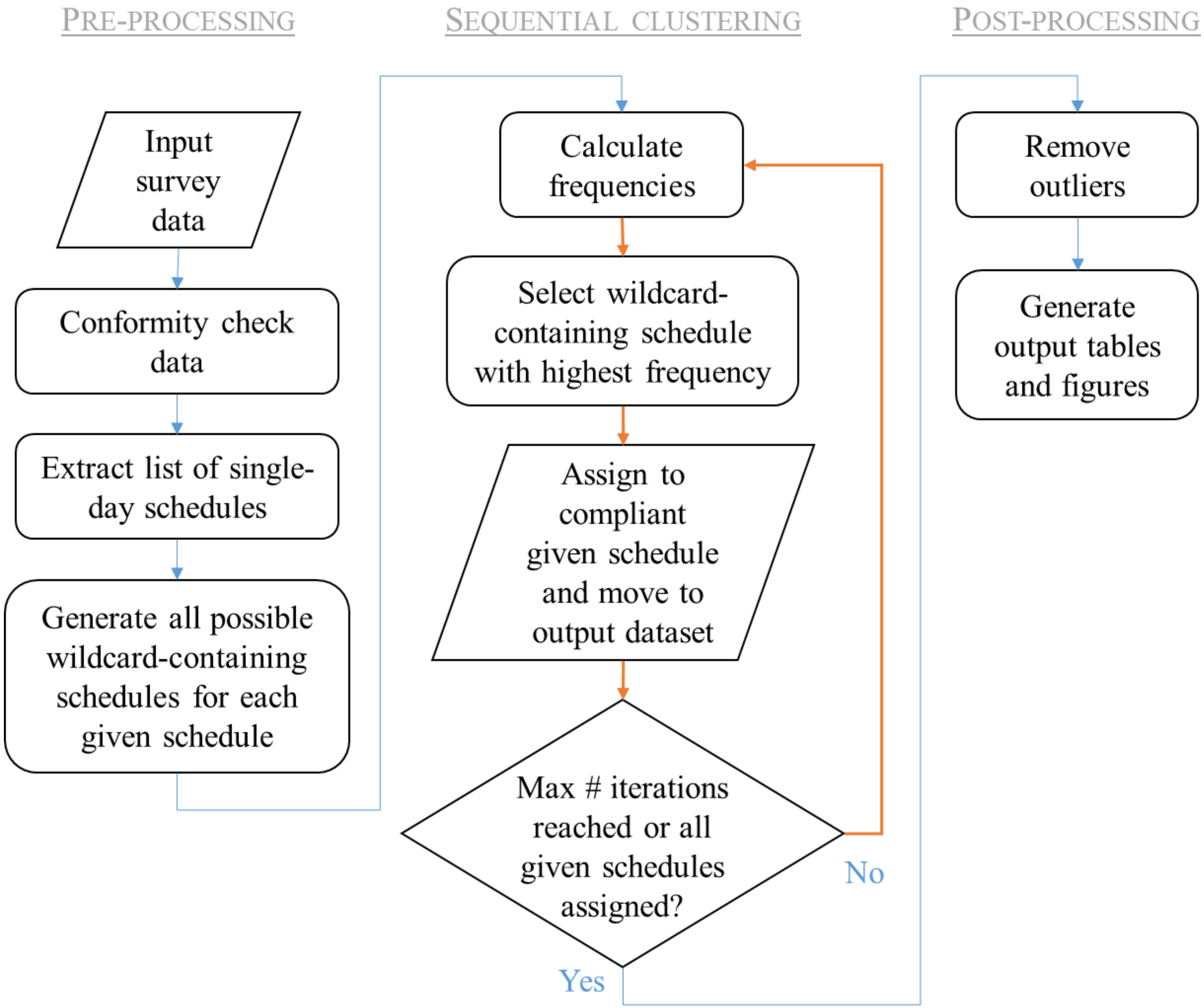
- HTS of Flanders, Belgium
  - Single-day, including weekends
  - Only out-of-home activities
  - 17,300 individuals
    - 13,200 at least one trip
  - Weights
  - 14 (of 2600 different) most frequent day-long schedules: 45% of observations (each other pattern <1%)
    - 55% more complex behavior → skeleton schedules??
- Pre-processing
  - Consecutive activities merged

Methodology – Sequential clustering algorithm

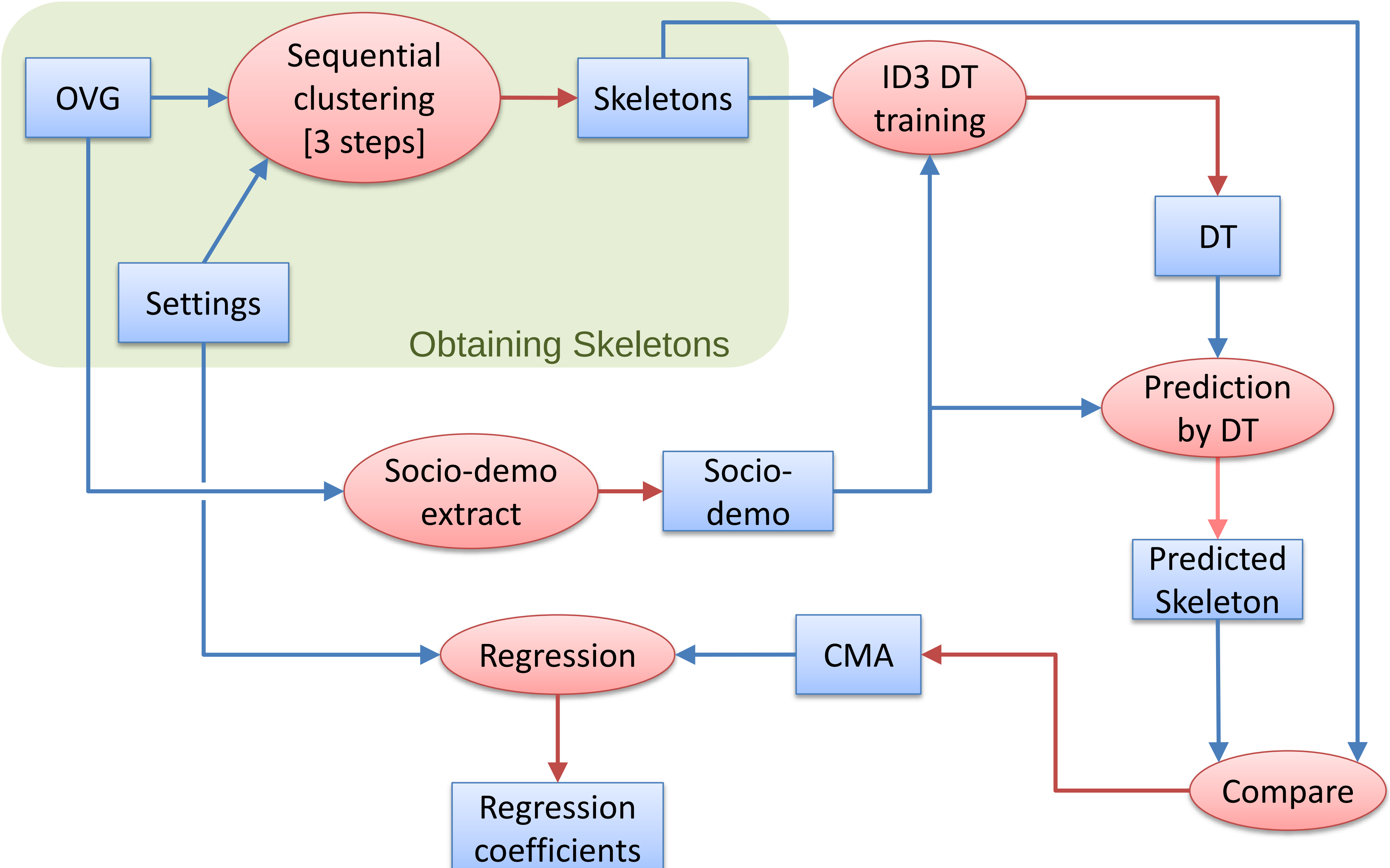
- Main idea:
  - Find common activity patterns in otherwise highly heterogeneous activity schedules
  - H-S-H-S-H ↔ H-S-H-R-H ↔ H-S-H-Se-H ?
  - ➔ H-S-H-X-H ?
  - Optimization of location X ?

H	Home
S	Shopping
R	Recreation
Se	Services
X	'Wildcard'

Methodology – Sequential clustering algorithm



Methodology – Overview of the research

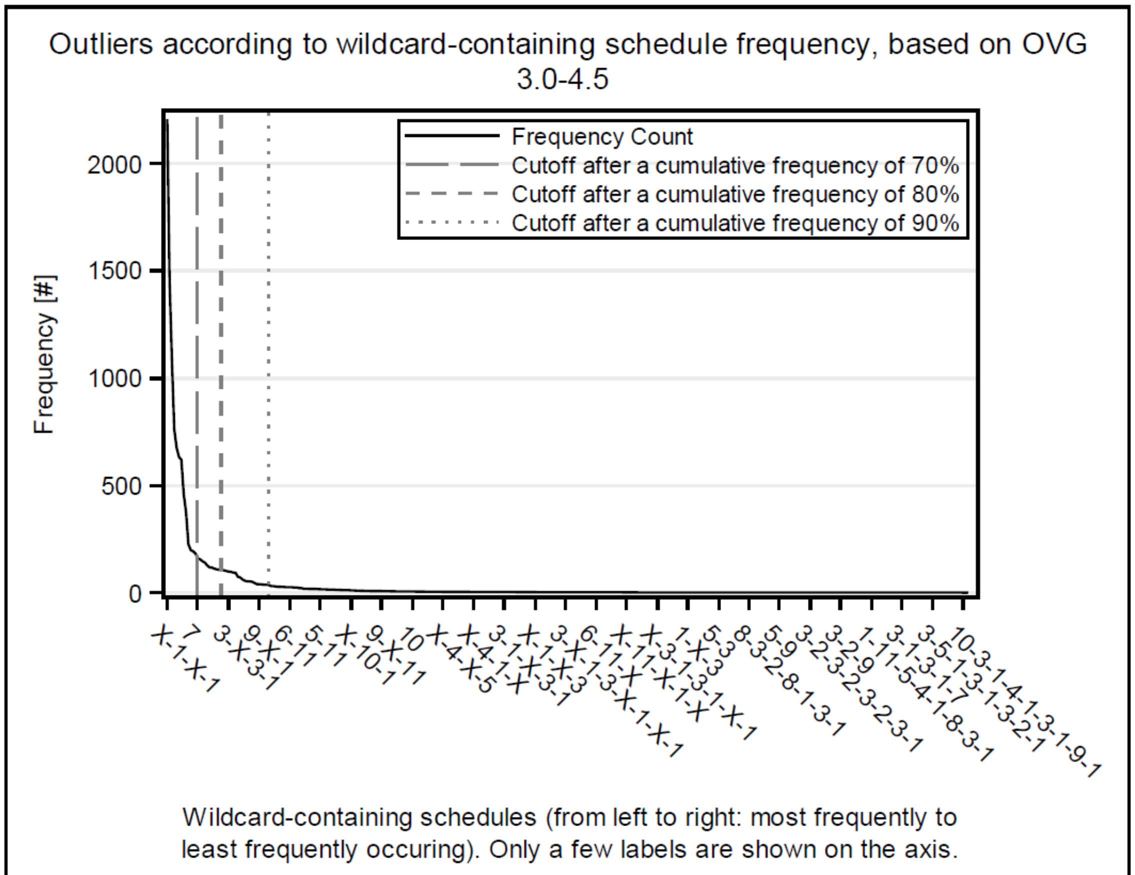


Methodology – Sequential clustering algorithm

- Pre-processing
  - Cleaning
    - ⑤ Remove schedules with >x activities?
  - ∀ schedules: find all possible wildcard-containing schedules according to settings:
    - ① Minimum # activities not replaced by X ?
    - ② H cannot become X ?
    - ③ W cannot become X ?
    - ④ Merge consecutive X ?
    - $N = \sum_{r=s}^{n} \frac{n!}{r!\,(n-r)!}$
- Sequential clustering
  - determine the largest groups of unique wildcard-containing patterns

Methodology – Sequential clustering algorithm

- Post-processing
  - Exclude *odd patterns* (“outliers”)
  - ⑥ Cutoff after cum. freq. of x %

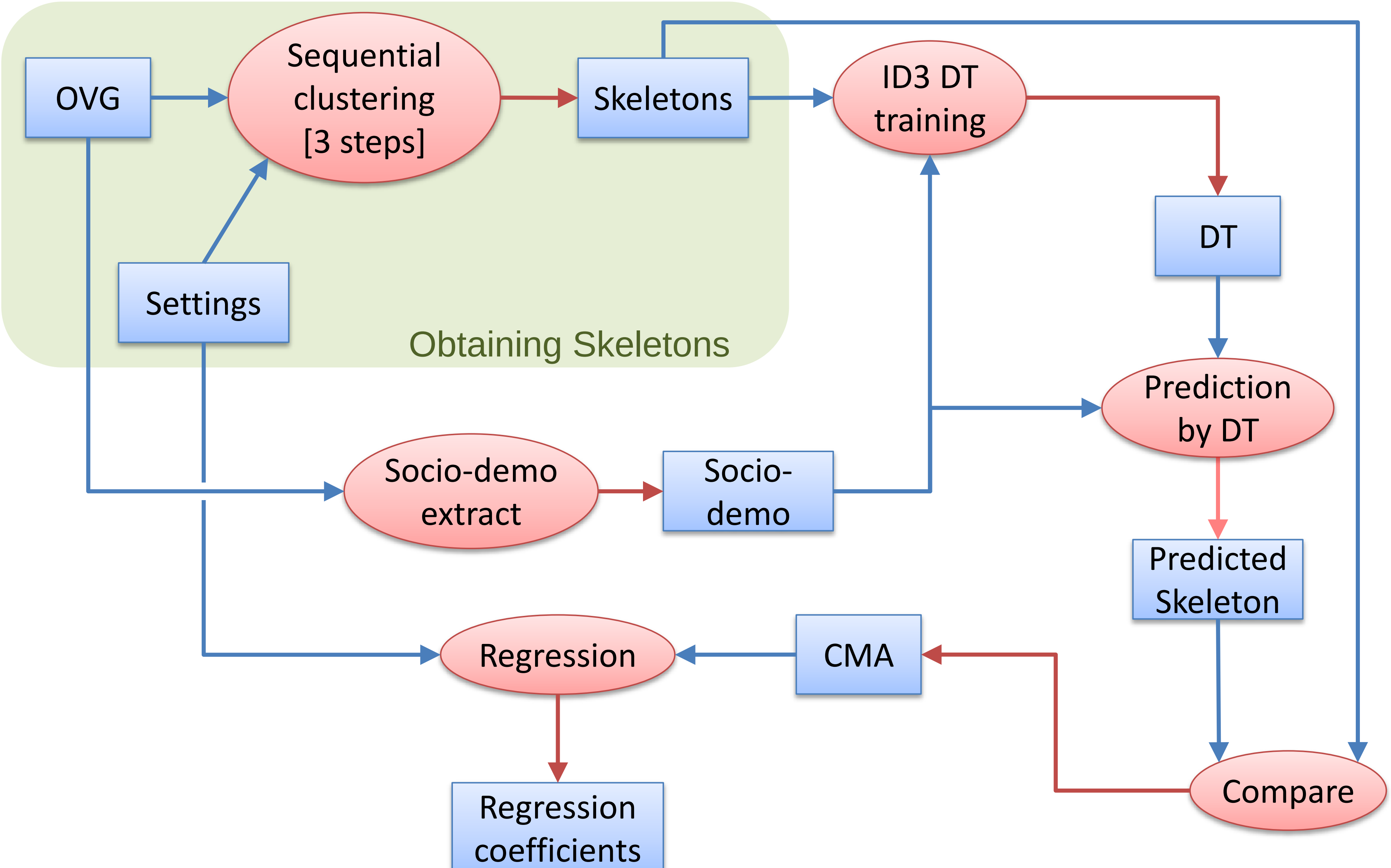


universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

Methodology – Sensitivity analysis

- Effect of ①, ②, ③, ④, ⑤, ⑥ …?
- Ultimate goal: predictions
  - Use DTs as in ABMs such as FEATHERS, ALBATROSS
- Two stages
  1. Generate many sets of skeletons with different setting combinations
     - 2520 sets were generated
  2. Use ID3 algorithm to train DT and estimate accuracy of skeleton classification
     - ⑦ minimum number of records in a leaf ?
     - ± 44,000 DTs fitted
     - Training (75%) and test set (25%) CMAs

Methodology – Overview of the research

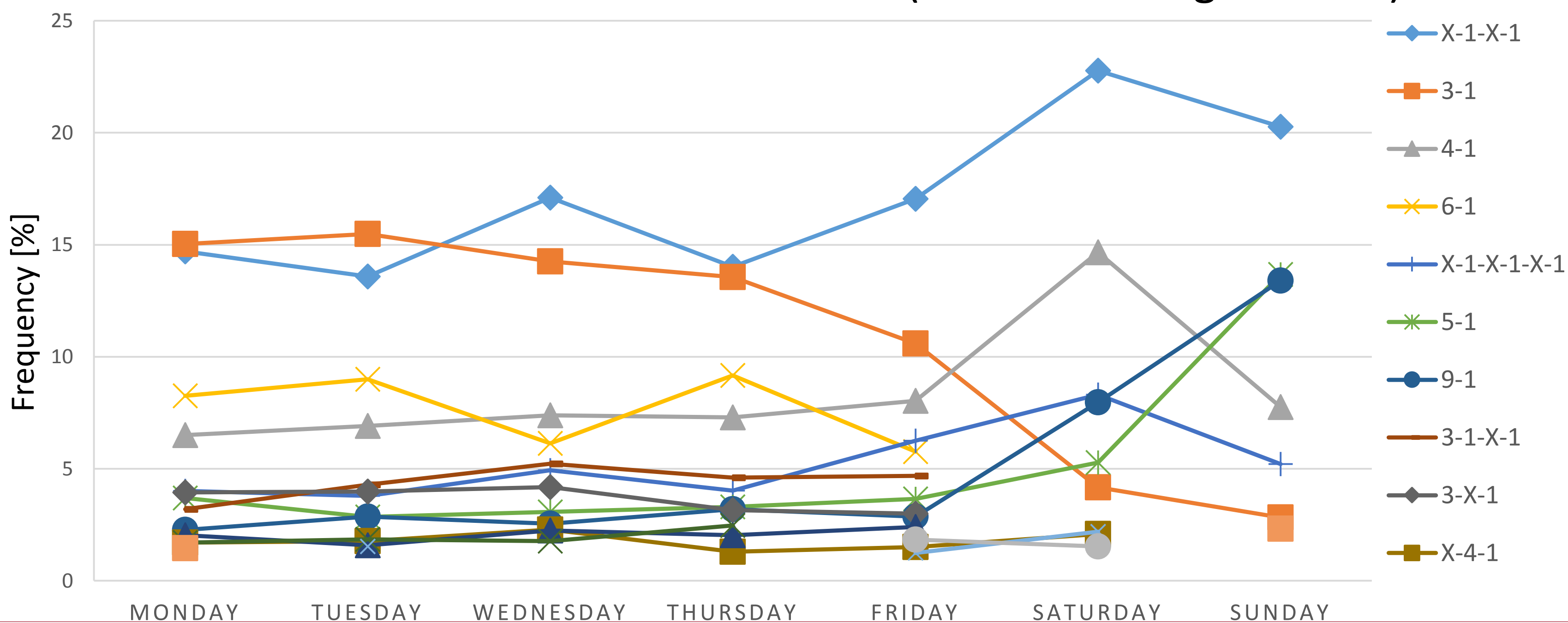


Methodology – Sensitivity analysis

- Influence of ①, ②, ③, ④, ⑤, ⑥, ⑦ on classification accuracy?
- Analyzed in regression model (adj. R^2 0.82)
  - Minimum # activities not replaced by X: inversely correlated
  - Cutoff after cum. freq. of x %: inversely correlated
  - Remove schedules with >x activities from input dataset: Marginal effect on CMA
  - H cannot become X: marginal negative effect
- 'Practical optimum' set of settings yields test set CMA of 32% (↔ null model accuracy 13.3%)

Results

- Two runs
  1. ① Minimum # activities not replaced by X = 3
     - 733 skeletons from 2,600 schedules
  2. ① Minimum # activities not replaced by X = 2
     - 341 skeletons from 2,600 schedules
     - 14 skeletons = 70 % of all records (↔ 45% in original data)



Discussion and conclusion

- Only single-day data is limitation
- Temporal component not accounted for
- Number of trips affected by merging of consecutive X

Yet:

- Activity-distribution in X quite complex; common travel behavior extracted
- Algorithm universal and simple
- Data driven
- Compatible with current ABM approaches

Thank you

Wim Ectors*
Bruno Kochan
Luk Knapen
Davy Janssens
Tom Bellemans

**wim.ectors@uhasselt.be*