Optimizing copious activity type classes based on classification accuracy and entropy retention

Wim Ectors¹, Sofie Reumers¹, Won Do Lee², Bruno Kochan¹, Davy Janssens¹, Tom Bellemans¹, Geert Wets¹
¹Hasselt University, Transportation Research Institute (IMOB) - ‡Manchester Metropolitan University, Crime and Well-being Big Data Centre

Abbreviations
ATC: Activity Type Class
HTS: Household Travel Survey

Intro
(Big) travel data: * large amounts, real-time, temporally and spatially referenced data
* personal and activity-travel info are lacking!

Methodology

Data
1. Seoul HTS (2010): 11 ATCs; ~76,800 individuals; temporal variables
   To confirm correct convergence (cfr. Ectors et al. 2017) and benchmark performance gains
2. USA NHTS (2009): 37 ATCs; ~908,903 individuals; temporal variables
   To the authors’ knowledge the richest activity encoding in a HTS case scenario
   The ‘ultimate’ challenge to optimize ATCs (using the local search algorithm) because of the combinatorial challenge (3.82-10^30 different sets of ATCs exist)
   Popular data set: optimal set of ATCs useful info

Data was split in train set (75%) and test set (25%) to train and evaluate the ATC classifier

Results

Converting to the same optimum for all 10 runs
The optimum is the same set of ATCs as in Ectors et al. (2017)
Instead of 10 hours of computation time (115,975 classifiers were trained in 20 parallel threads on a server using the method of Ectors et al. (2017)) only a few minutes were needed with the local search algorithm

Conclusion remarks
2 out of 15 converged to a ‘suboptimal’ set of ATCs because:
* too few iterations before the stopping criterion is fulfilled
* these runs were stuck in a local optimum
The latter can be rejected since some runs did find successfully converge to the optimum encountered the suboptimal set of ATCs during their iterations, meaning that there was a direct path that could lead to the same optimum. By chance (i.e. too few iterations before stopping criterion was fulfilled) such a path was not found for 2 out of 15 runs.
The grouped ATCs have similar temporal properties (usually not a long duration; could occur at any time of the day; likely to be chained together).

Contact information: Wim Ectors, Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium. E-mail: wim.ectors@uhasselt.be