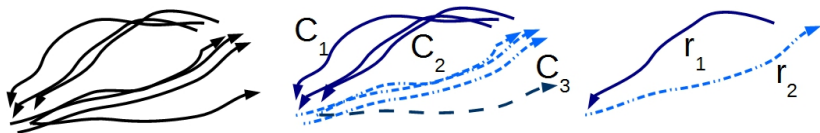


Find Your Way Back: Mobility Profile Mining with Constraints

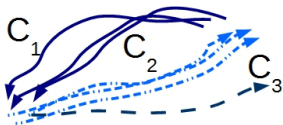
Lars Kotthoff, Mirco Nanni, Riccardo Guidotti, Barry O'Sullivan

04 September 2015

- ▷ characterise mobility of users by extracting profiles from GPS data
- ▷ identify repeated and systematic movements
- ▷ usually symmetric trips – going there and coming back
- ▷ relevant for e.g. public transportation planning, carpooling

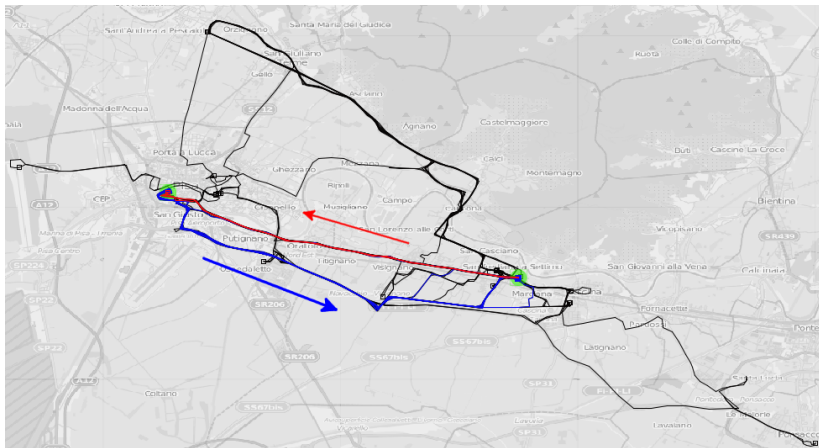


- ▷ preprocess GPS data
- ▷ cluster GPS trajectories
- ▷ find representatives for each cluster



- ▷ preprocess GPS data
- ▷ cluster GPS trajectories
- ▷ find representatives for each cluster

Application



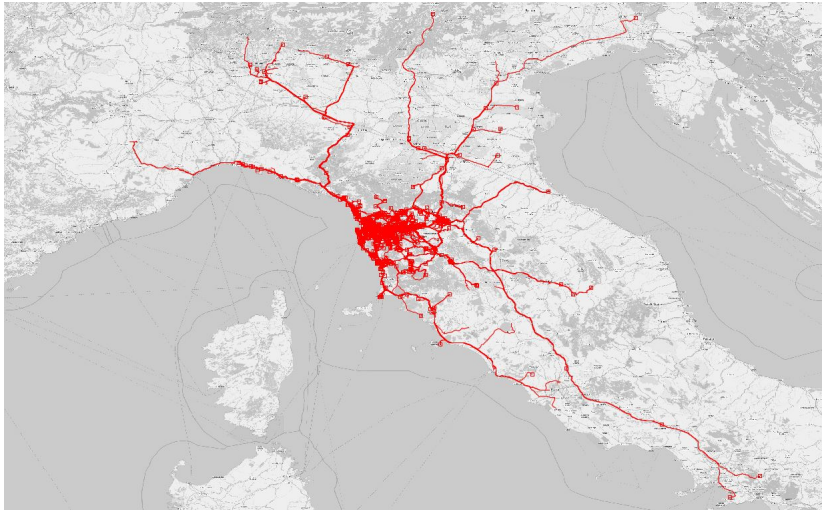
- ▷ separate signal from noise – regular from one-off trips
- ▷ standard approach: set minimum cluster size
- ▷ problem: may miss small cluster that is part of a routine
- ▷ e.g. go to work every morning, but go home via supermarket, gym, ...

- ▷ leverage symmetry of trips
- ▷ if there is a “supporting” symmetric cluster, even small clusters are not noise
- ▷ not easily done with standard data mining approaches
- ▷ easily encoded in constraint model through additional constraints

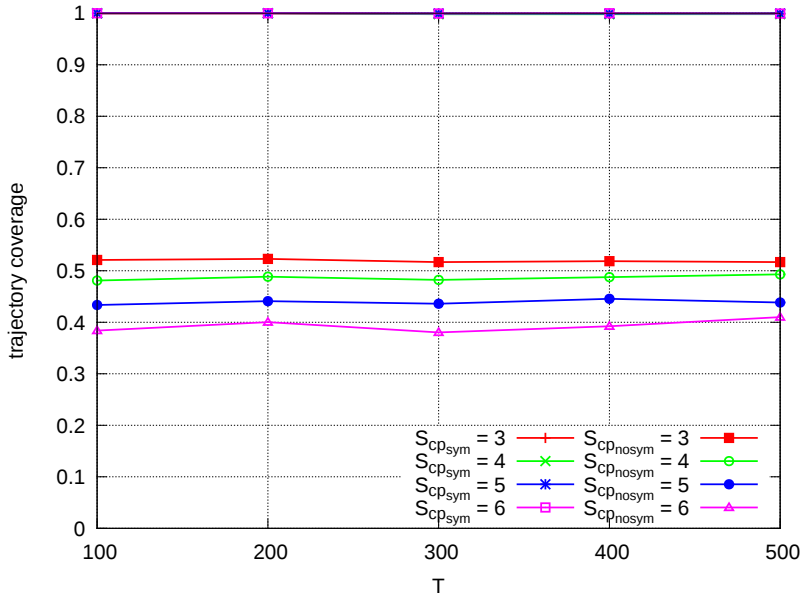
- ▷ every trajectory assigned to a cluster or a special “noise” cluster
- ▷ trajectories cannot be in the same cluster if far apart (spatially and temporally)
- ▷ trajectories are symmetric if start of one is close to the end of the other and vice versa
- ▷ minimum size for cluster which has no symmetric trajectories and smaller minimum size for cluster which does
- ▷ minimize size of “noise” cluster, i.e. cluster as many trajectories as possible
- ▷ minimize the number of distinct clusters, i.e. make clusters as large as possible

- ▷ implemented with Minion constraint solver
- ▷ cannot express complex optimisation directly
- ▷ iterate with increasing number of clusters – if too few, problem will have no solution
- ▷ force minimization of noise trajectories through value ordering

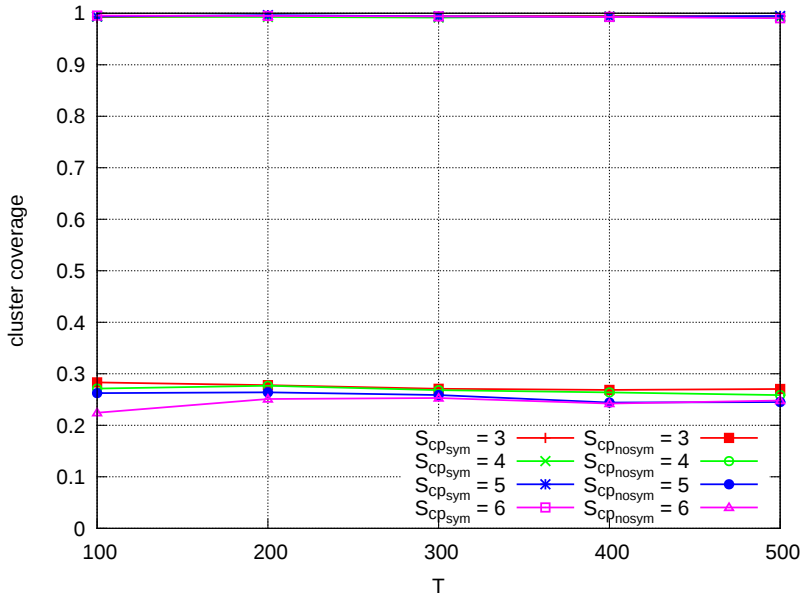
Evaluation



Results



Results





- ▷ obfuscated data available at http://www.cs.ubc.ca/~larsko/downloads/cp2015_anonymized_dataset.tar.gz
- ▷ code and obfuscated data available in virtual machine at <http://recomputation.org/node/64>
- ▷ for more information, see <http://www.recomputation.org>

- ▷ applied CP to traditional data mining task (clustering)
- ▷ achieved qualitatively significantly better results
- ▷ challenges in terms of computation time remain

