Transportation Mode Detection using Random Forest

Jasna Urbančič, Veljko Pejović, Dunja Mladenić

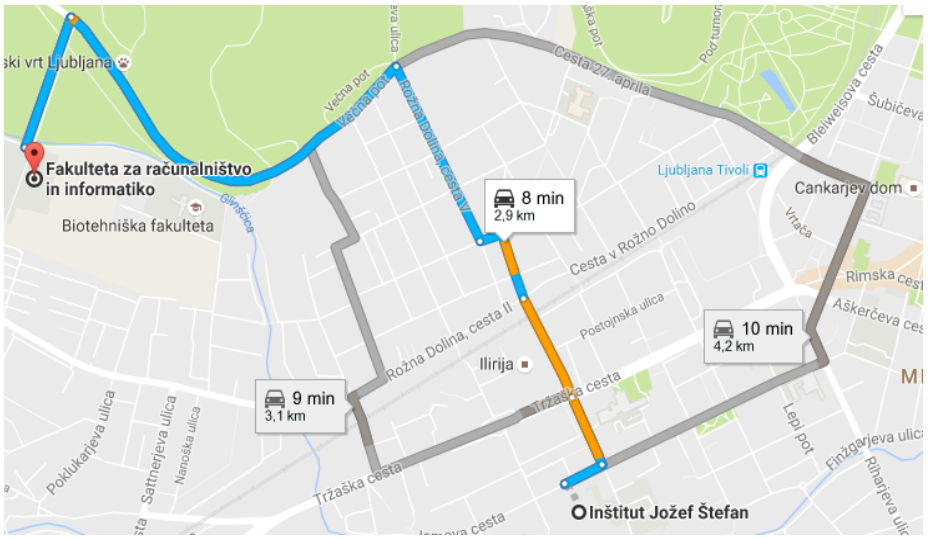
SiKDD 2018, 11 October 2018

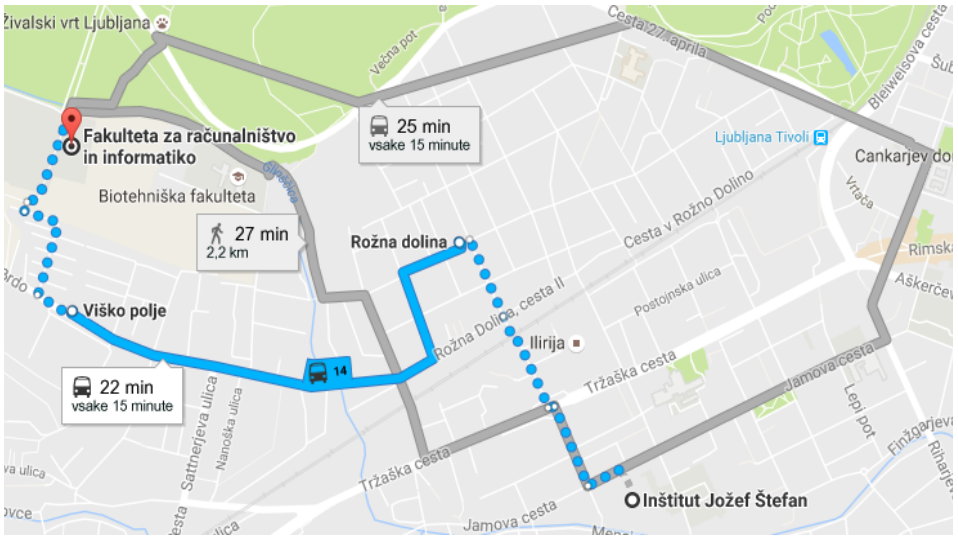
Jožef Stefan Institute

Artificial Intelligence Laboratory

Motivation

- APIs already support fine-grained classification of non-motorized forms of transportation (still, walking, running, cycling, and in vehicle)
- Not enough for tracking and routing purposes, specially in use-cases for urban environments





© Google Maps

Goals

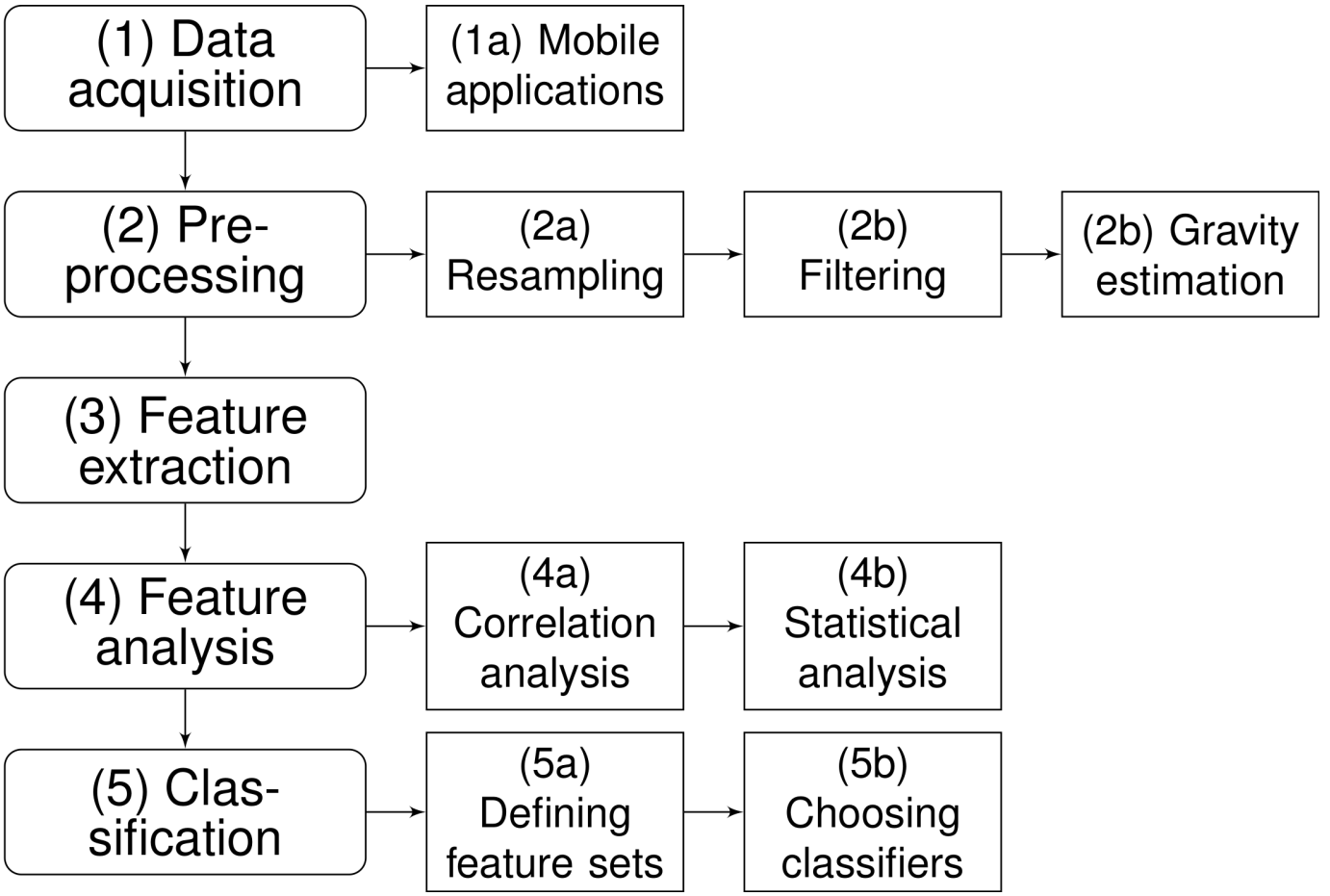
Focus on:

1. Feature extraction
2. Feature analysis

Main contribution:

Feature analysis, which revealed the impact of each feature to the classification scores

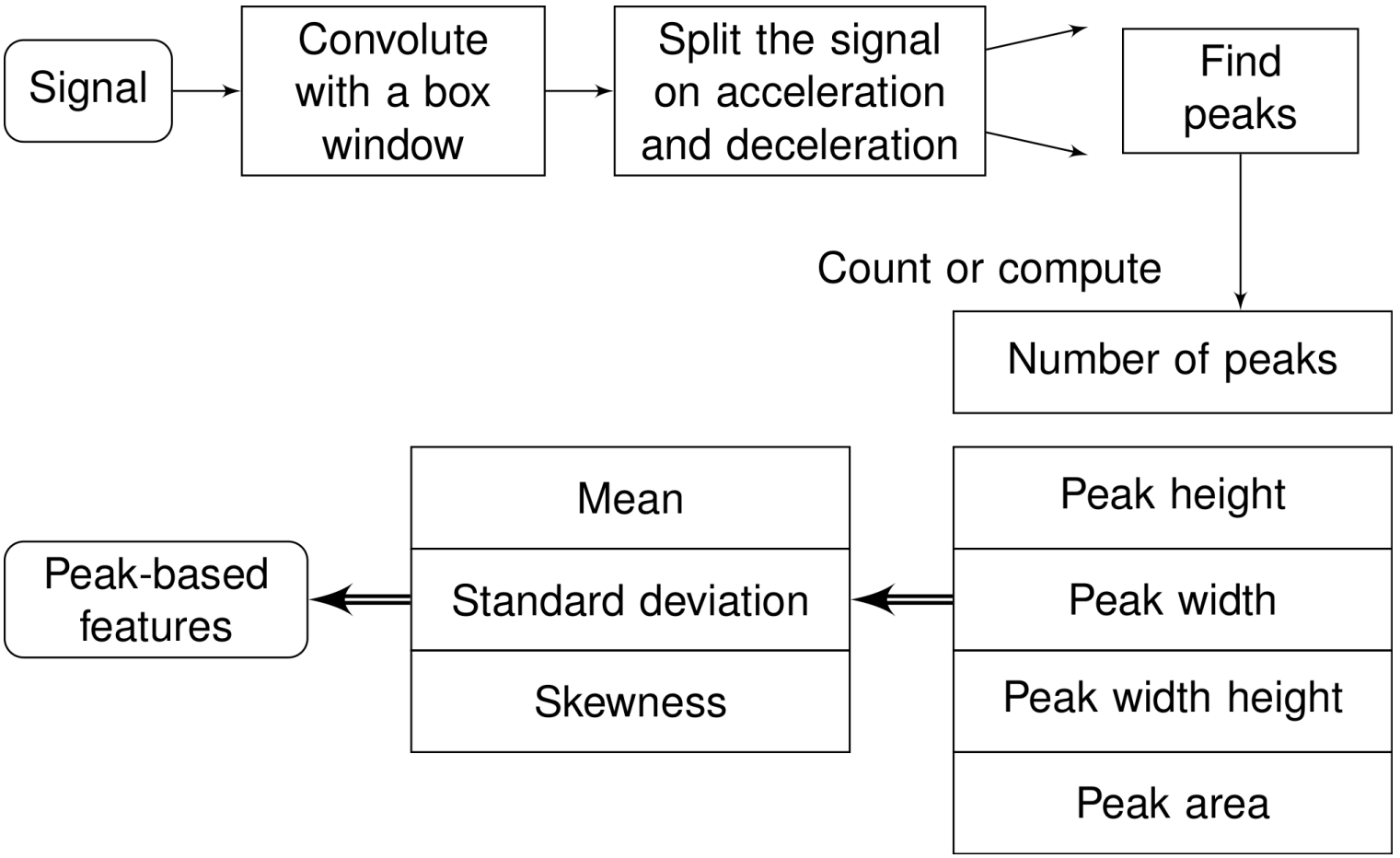
Our approach



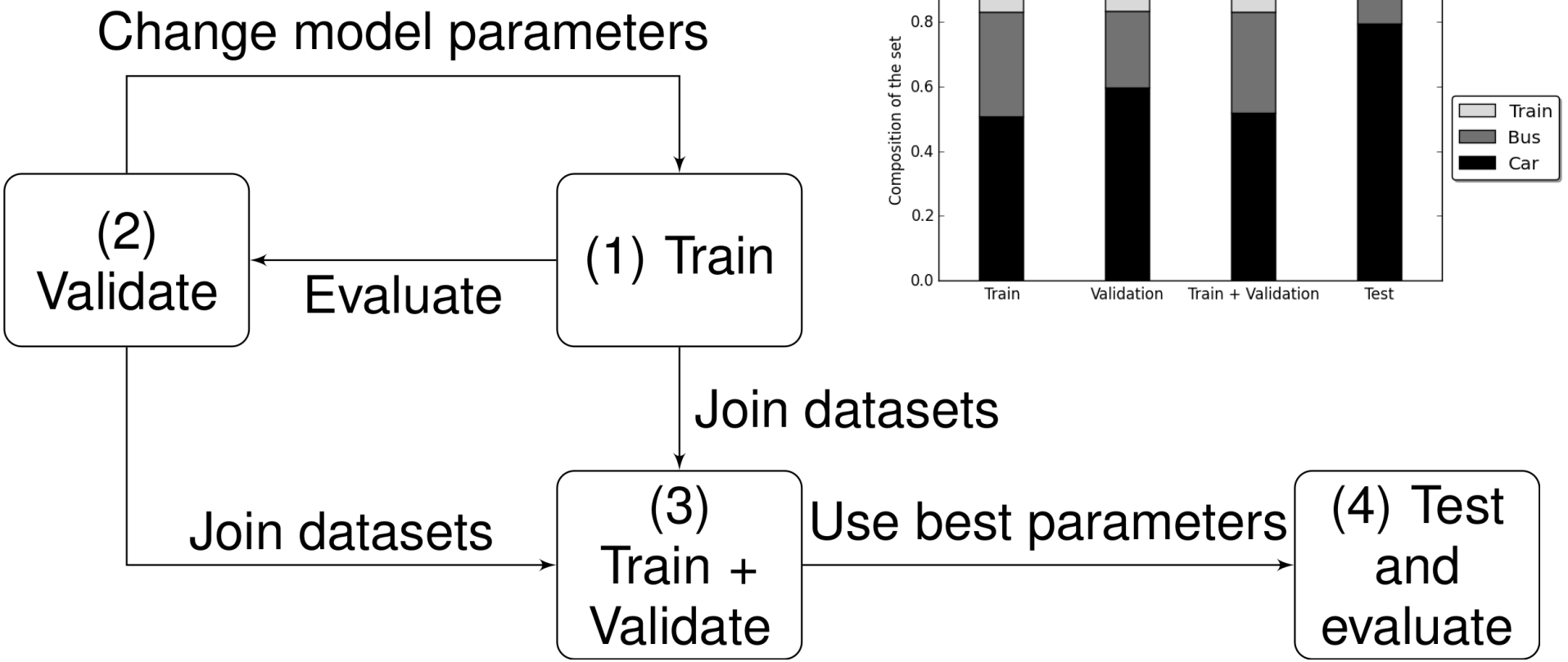
Feature domains

- **Statistical** – mean, standard deviation, skewness, percentilles, etc.
- **Time** – integral and double integral of signal over time, zero crossings, etc.
- **Frequency** – spectral energy, spectral entropy, spectrum peak position, etc.
- **Peak** – volume, height, width of a peak, etc.
- **Segment** – peak frequency, stationary duration, etc.

Extraction of peak-based features



Evaluation


Change model parameters
(2) Validate
Evaluate
(1) Train
Join datasets
Join datasets
(3) Train + Validate
Use best parameters
(4) Test and evaluate

Distribution of samples across modes
Composition of the set
1.0
0.8
0.6
0.4
0.2
0.0
Train
Validation
Train + Validation
Test
Train
Bus
Car

Feature sets

Set	Acceleration	Features	Size
D-S	Dynamic	Statistical	54
D-SF	Dynamic	Statistical, Frequency	94
D-SFP	Dynamic	Statistical, Frequency, Peak	172
H-S	Horizontal	Statistical	54
H-SF	Horizontal	Statsitical, Frequency	94
H-SFP	Horizontal	Statistical, Frequency, Peak	172
ALL			376

Results

Set	CA	RE	PR	F1
D-S	0.48	0.41	0.39	0.37
D-SF	0.60	0.41	0.41	0.39
D-SFP	0.46	0.39	0.40	0.35
H-S	**0.64**	0.40	0.43	**0.41**
H-SF	0.53	0.39	0.43	0.36
H-SFP	0.50	0.37	0.40	0.34
ALL	0.47	0.35	0.40	0.33

- Using peak features in combination with the other features sets results in the decrease in F1 score.
- F1 and CA increase when we add frequency-based features in case of dynamic acceleration and decrease in case of a similar action for horizontal acceleration.
- Smaller features generally perform better than larger

Feature selection

- Backward feature selection (elimination)
  - We remove features from the feature set
  - 28 features – more peak-based features than statistical, dynamic and horizontal acceleration appear in similar proportions
- Forward feature selection
  - We add features to the feature set
  - 10 features – mostly statictical features, followed by peak-based features, extracted from dynamic acceleration
- Only one frequency-based feature in each of these feature sets

Results

Feature set	CA	RE	PR	F1
Forward selection (10)	0.70	0.50	0.47	**0.48**
Backward elimination (28)	0.73	0.50	0.48	**0.49**

Forward selection

T/P	Car	Bus	Train
Car	**0.78**	0.27	0.05
Bus	0.51	**0.40**	0.09
Train	0.47	0.21	**0.32**

Backward elimination

T/P	Car	Bus	Train
Car	**0.83**	0.12	0.05
Bus	0.55	**0.35**	0.10
Train	0.45	0.23	**0.32**

Conclusion

- Using feature selection we are able to improve classification scores for at least 0.04, in some cases even over 0.10.
- Most of the non-car samples are still misclassified as cars.
- Peak-based features did perform well in predefined feature sets, they consistently appear among selected features in feature selection.