Karlsruher Institut für Technologie

Image-based activity pattern segmentation using longitudinal data of the German Mobility Panel

Sascha von Behren, Bastian Chlond (presenting)

INSTITUT FÜR VERKEHRSWESEN, FAKULTÄT FÜR BAUINGENIEUR-, GEO- UND UMWELTWISSENSCHAFTEN

Agenda

- Data
- Idea
- Challenges
- Some results
- Conclusions and lessons learned
- Next steps

Data – The German Mobility Panel

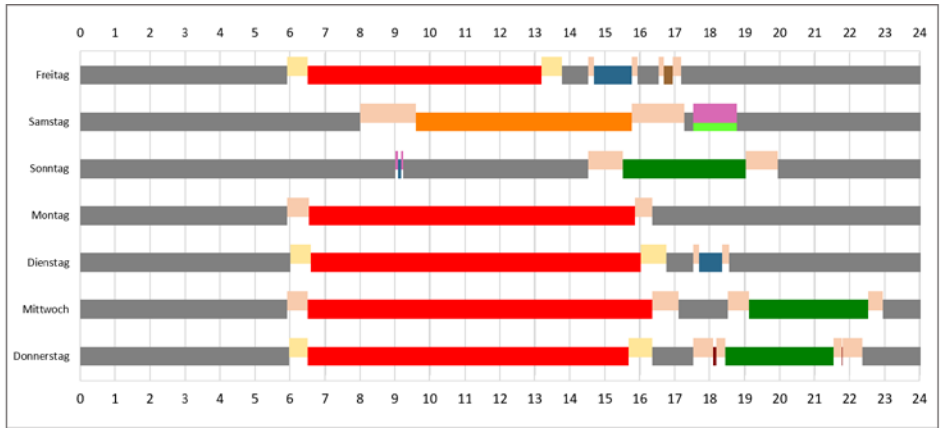
Deutsches Mobilitätspanel (MOP)

- Survey about everyday travel behaviour of the German population
- About 1.500 -1 .800 households and 3.000 - 3.600 participants per year
- Up to 3 participations per individual in three consecutive years
  - Socio-demographics of the individuals
  - Household information
  - **Travel Diaries for a period of one week**

Allows - without any other information – to assign individuals into groups (socio-demographics / activity levels)

GraDiV

- A tool for visualisizing activities and mobility for the reporting period
- Originally implemented to better understand „travel and activity patterns“
- Primary Utilization: Plausibiliziation of trip diaries



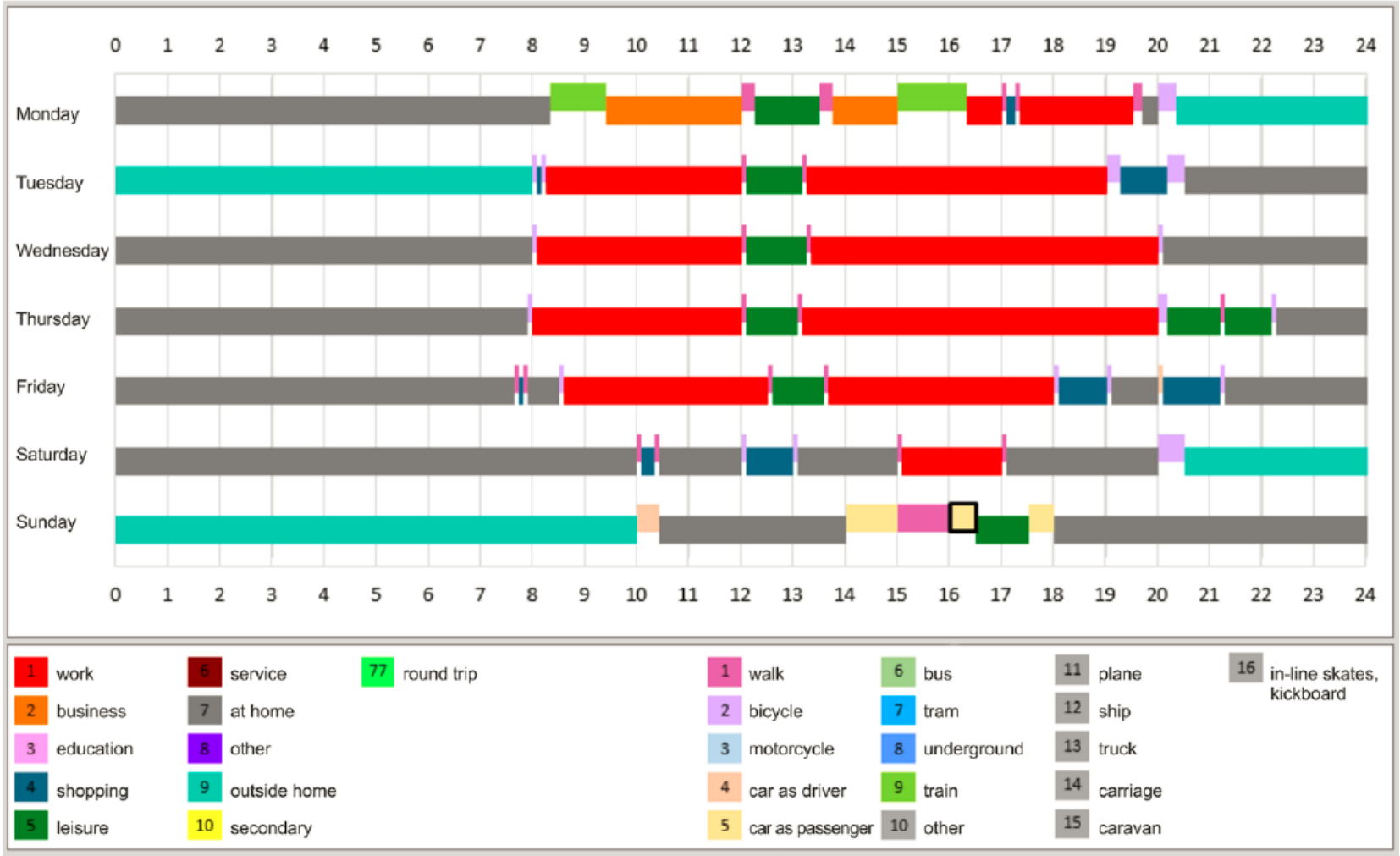
Bildquelle: GraDiV

The Gradiv Visualization

At a glance

- out-of-home activities
- trips by purpose and mode
- temporal information

We learn a lot about the multiplicity of human behaviours!

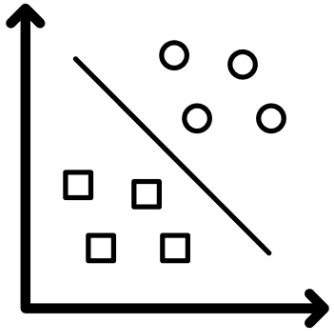


The idea: Segmentation of traveller types by an image-based approach

Karlsruher Institut für Technologie

Why a segmentation:

- To understand similarities in behaviours to provide a limited number of appropriate measures
- To ease our understanding about the complexity and multiplicity of life



Why the images:

- Utilization of information, not yet used in conventional segmentation approaches…
- Not really new insights, however the application a new methodology for an old problem…

Our Research questions:

- How can we utilize the information within the images?
- What algorithms can be applied? How much computing is necessary?
- Can this approach help to improve our understanding about behavior?

The idea: Segmentation of traveller types by an image-based approach

- **„A-priori“ versus „A-posteriori“ segmentation approaches**
  - **A priori:**
    Characterized by a definition of the expected groups and their characteristics
    based on comprehensive prior knowledge and assumptions
  - **A posteriori (Post-hoc):**
    Uses only statistical methods instead of comprehensive prior knowledge

- **Image based:** Up to now **A-posteriori** ➔ Attempt to use the information as originally in the images available
  - **Advantages**
    The available information in the image form the basis for segmentation.
  - **Drawbacks**
    Existing prior knowledge has been ignored

Image based Segmentation

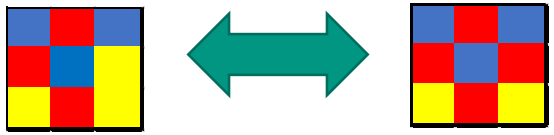
- **What does the utilization of images mean?**
  - Comparison of the behaviours activity by activity, trip by trip ➔ pixel by pixel!

- **What does similarity mean?**
  - Identical behavior ➔ each pixel is identical to the image of another?

6 identical ⇔ 3 different

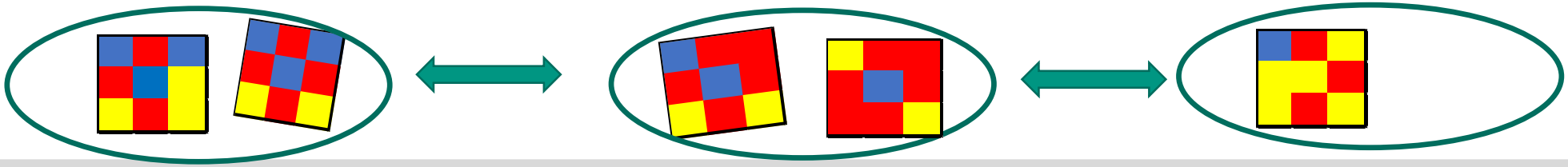
9 identical ⇔ 0 different

  - **Similarity** ➔ the difference in pixels between individuals is comparably small



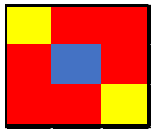
8 identical ⇔ 1 different

- **Segmentation based on clustering:**
  - As usual for clusterung approaches, the distances are small within clusters, and larger between clusters

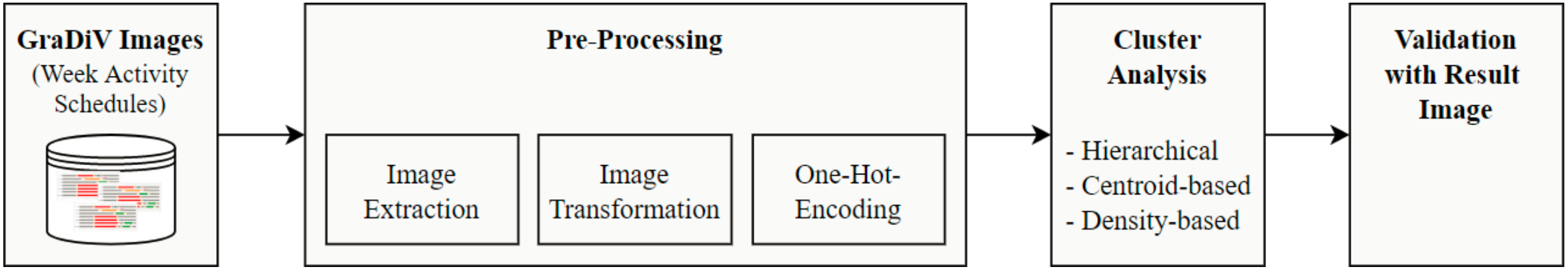


The Challenges

- The complexity increases with the number of variables in the cluster process and the dimensionality arises
- Image based approaches: Each pixel represents a “dimension”



- The images contain a lot and too much information which need to be condensed, irrelevant information must be kicked out!

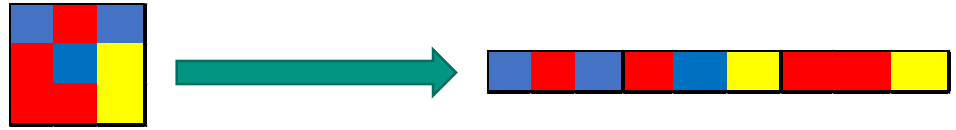


Pre-Processing ➔ Reducing dimensionality

- **Image extraction ➔ Reduction of complexity: Image for humans ≠ images for a computer**
  - Condensation of graphical information

- **Image transformation ➔ Reduction of dimensionality!**
  - Reduction of data volumes (1 line for one day is sufficient!) / reduction to 5-minute-intervals
  - Reduction of the information of colors ➔ Transformation into grey-values
  - **Reduction of the image data to a vector of 4242 pixels per week (reduction of size and colors)**

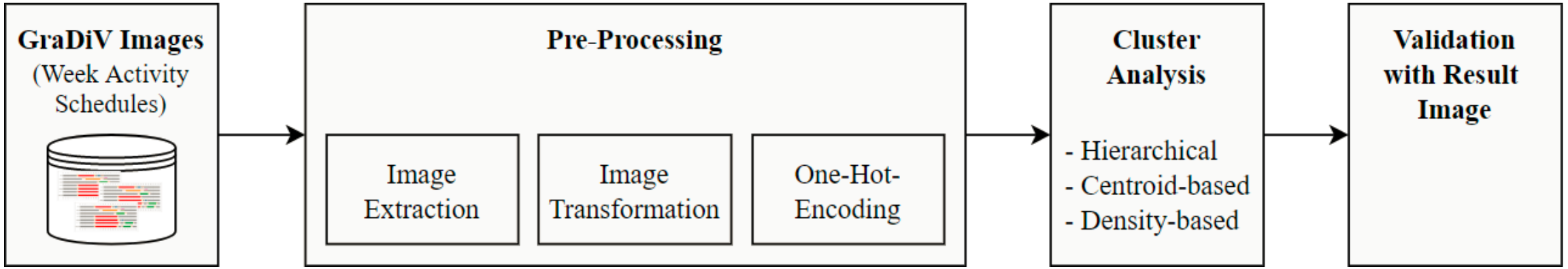
- **One-hot-encoding:**



  - **To avoid misinterpretations of numeric distances between colors ➔** color values (7) in binary form

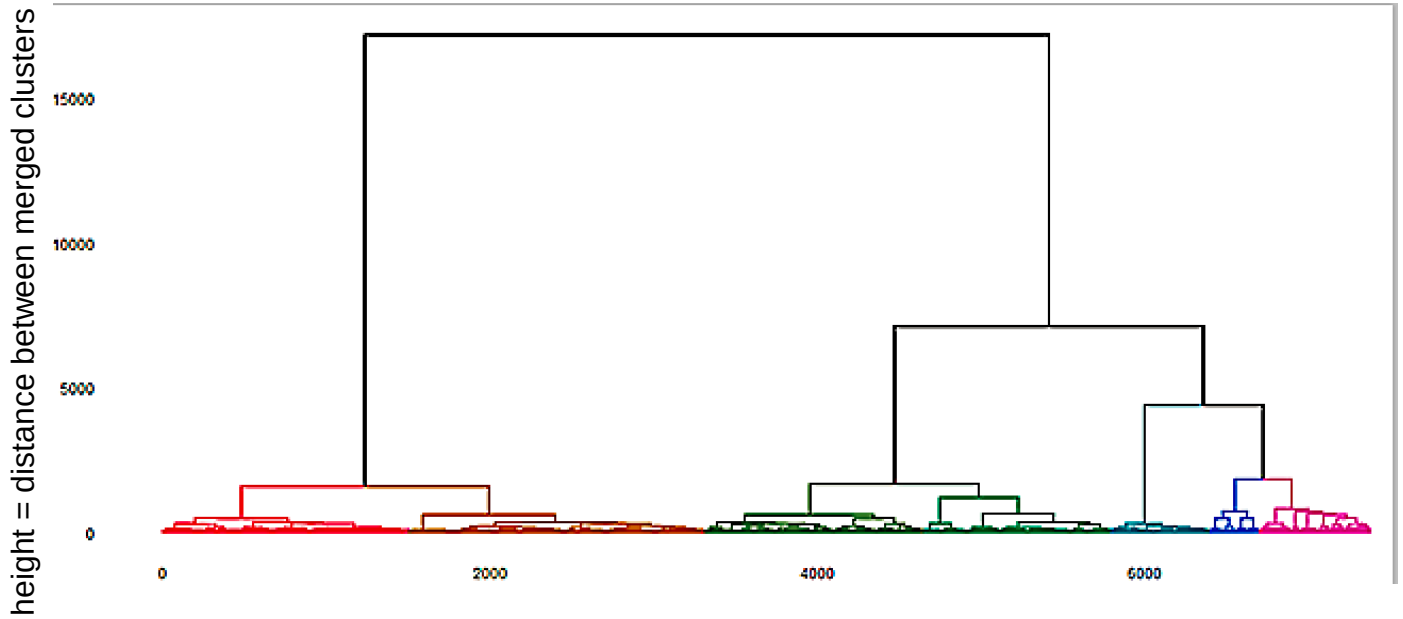
- **Outcome:**
  - **Per individual a “vector” in a binary form ➔ length = 29,694 (= “dimensions”)**

Cluster Analysis

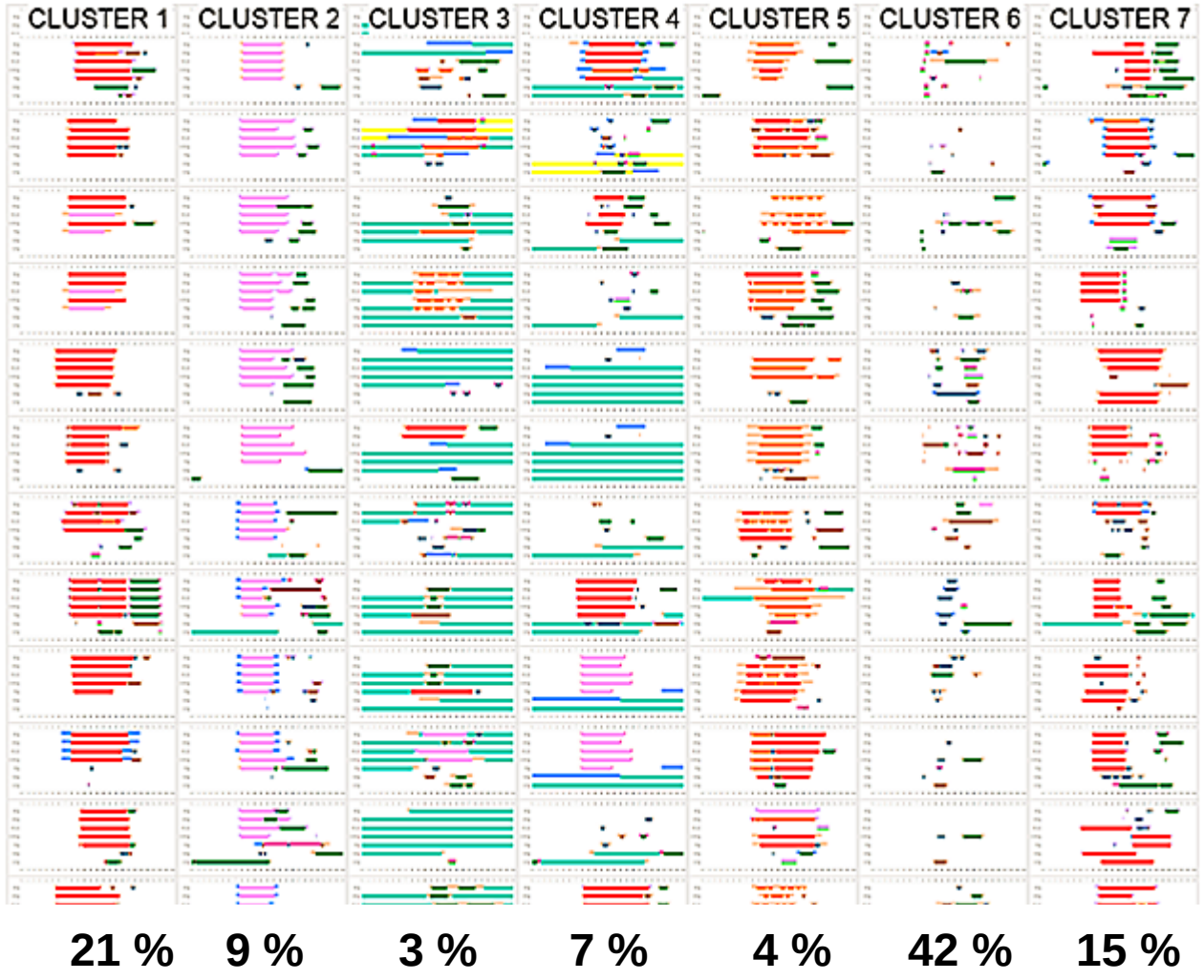


- **Clustering the information: 7,362 activity plans × 29,694 dimensions**
- **Application of agglomerative hierarchical clustering algorithms**

k-means ➔ k = 7



Results: Extraction of the result image with (altogether 7362 images of behaviors)



1	Work	6	Escorting	11	priv. purpose
2	Business Trip	7	At home	77	Roundtrip (e.g.Stroll)
3	Education	8	Others		
4	Shopping	9	Out of home (Hotel etc.)		
5	Leisure	10	2. Residency		

Cluster Names

1. Employed middle-class
2. Pupils & students
3. Out-of-home long-distance traveler
4. Weekend-actives
5. Daily performers in job & life
6. Pensioners
7. Part-time mothers

Heat-Maps of activities per cluster

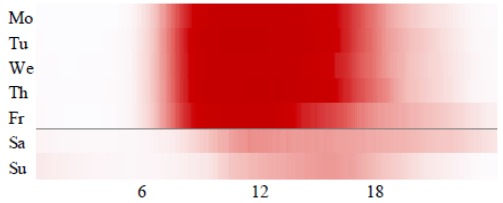
- The time use is the determining dimension:

  To what extent the activities in the clusters are distributed over the week?

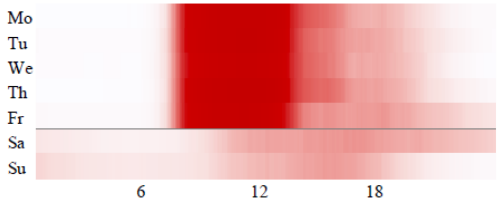
- Information about the interpersonal variation and stabilities within the clusters

- How large is the similarity in temporal terms.

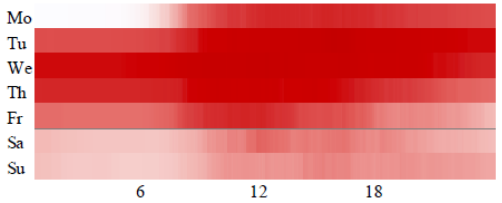
Cluster 1: Employed middle-class



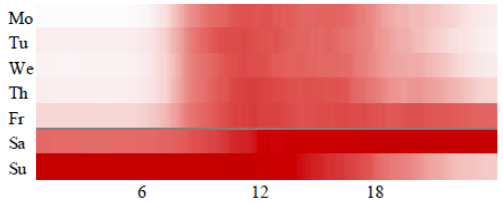
Cluster 2: Pupils & Students



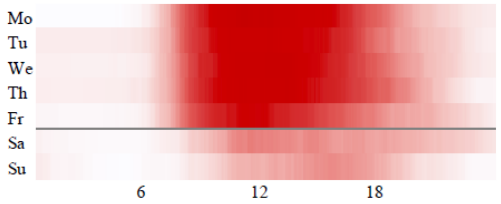
Cluster 3: Out-of-home long-distance traveler



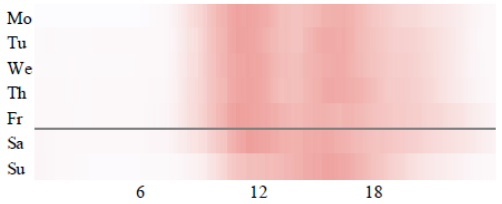
Cluster 4: Weekend-actives



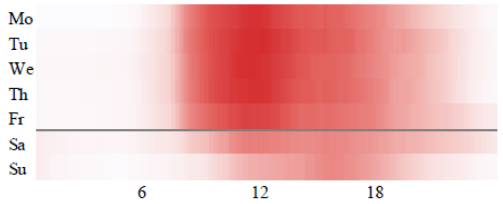
Cluster 5: Daily performers in job & life



Cluster 6: Pensioners



Cluster 7: Part-Time-Mothers



Cluster-definition

- Activity behavior is as to expect mostly explained **by socio-demographis variabels**
- The clustering of the images is time-use and activity-purpose based

Cluster Names

1. **Employed middle-class**
2. **Pupils & students**
3. **Out-of-home long-distance traveler**
4. **Weekend-actives**
5. **Daily performers in job & life**
6. **Pensioners**
7. **Part-time mothers**

Cluster #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Ø**	**CV**
Cluster name	Employed middle-class	Pupils & students	Out-of-home long-distance traveler	Weekend-actives	Daily performers in job & life	Pensioners	Part-time mothers		
% Size (of n=7,362 weekly activity schedules)	21%	9%	3%	7%	4%	42%	15%		
Sociodemographics									
Sex [1]	1.39	1.48	1.38	1.50	1.28	1.52	1.65	1.49	0.08
Age	47.70	**16.18**	45.18	50.27	49.69	**63.05**	49.52	51.86	0.28
% 10-17	0%	**74%**	6%	4%	0%	0%	0%	12%	2.25
% 18-25	4%	21%	15%	8%	3%	2%	4%	8%	0.89
% 26-35	12%	2%	11%	11%	11%	5%	9%	9%	0.42
% 36-50	36%	1%	22%	20%	33%	10%	34%	22%	0.59
% 51-60	40%	0%	26%	25%	38%	17%	38%	26%	0.55
% 61-70	8%	0%	13%	19%	12%	**32%**	13%	14%	0.72
% 70+	0%	0%	7%	13%	3%	**34%**	2%	8%	1.42
HH-size	2.49	3.77	2.63	2.39	2.67	2.12	2.61	2.70	0.21
Children within HH [2]	0.28	**0.61**	0.27	0.20	0.37	0.13	**0.35**	0.25	0.60
Income	7.36	7.73	7.50	7.28	7.60	**6.22**	7.15	6.89	0.07
Occupation									
% Full-time	**87%**	1%	55%	43%	**79%**	11%	44%	37%	0.87
% Part-time	10%	1%	8%	18%	15%	13%	**48%**	16%	0.91
% Education	2%	**95%**	18%	11%	1%	3%	3%	12%	2.92
% Others	2%	3%	19%	28%	5%	**73%**	5%	35%	0.74
Home Office [3]	0.14	0.00	**0.24**	0.16	**0.30**	0.09	0.18	0.12	0.80
Driving License	0.97	**0.22**	0.89	0.92	0.97	0.90	0.95	0.87	0.31
Car-availability [4]	**1.24**	2.25	1.36	1.36	**1.23**	1.41	1.31	1.38	0.26
Mobile Restricted	0.03	0.01	0.06	0.07	0.04	**0.18**	0.05	0.10	0.56
Geographic area [5]	2.08	2.15	2.00	**1.81**	2.09	2.00	2.03	2.02	0.05

Intra-individual cluster stability

From Cluster (year 1)	To Cluster (year 2) 1	2	3	4	5	6	7	Total	Share of Cluster
1 Employed middle-class	59%	1%	3%	8%	3%	11%	14%	100%	20%
2 Pupils & Students	4%	78%	6%	1%	0%	9%	1%	100%	7%
3 Out-of-home long-distance traveler	13%	16%	28%	6%	0%	28%	9%	100%	4%
4 Weekend-actives	19%	8%	4%	20%	5%	38%	8%	100%	8%
5 Daily performers in job & life	24%	0%	4%	15%	33%	13%	11%	100%	3%
6 Pensioners	3%	0%	3%	7%	1%	81%	5%	100%	44%
7 Part-Time-Mothers	13%	1%	1%	10%	2%	22%	51%	100%	14%

Who remains stable in terms of the cluster assignment?

- Analysis of the subsample of individuals who participated twice ➔ 2,540 transitions „unweighted“!
- Comparably stable as well as instable clusters – why?

(Self-)critical evalation

- **The approach (image based segmentation)**
  - The image based aproach is feasible, however additional information must be utilized as well!
  - Travelling and mobilty behaviour (e.g. mode use) is up to now not sufficiently represented (➔ Weighting e.g.trips more than activities)
- **The results**
  - So far – basically nothing new!
  - The behavioural data of one week does not represent the complete intra-individual variability.
  - The relevance of irregularly / with low frequencies performed activities (➔ LD-activities)
- **New insights**
  - For the improvement of the approaches
  - Improvement of the understanding of the relevance of LD-travel (➔ need for a survey-integration)

Next Steps!

- A priori approach – predefinition of groups by „human experience“

- Inclusion of not yet considered dimensions and information
  - E.g. characteristics of travelling / mode use as monomodal or multimodal behaviour

- Training an Artificial Intelligence System by predefinitions of typical representatives for groups („Labeling“)

- Assignment of the remaining individuals according to the predefined clusters, e.g. by Random Forest approach.

Thank you for your attention!

Bastian Chlond
Institute for Transport Studies
Kaiserstraße 12, 76131 Karlsruhe
+49 (721) 608-42257
Bastian.Chlond@kit.edu