Preferred citation style for this presentation

Analysing intrapersonal variability of travel behaviour using the sequence alignment method

Robert Schlich

IVT
ETH
Zürich

September 2001
Intrapersonal Variability

Inter-personal Variability

One single day

Person 1
Person 2
Person 3
Person 4
Person 5

mean

Intra-personal Variability

Person 1

Day of week
Trips and activities: different characteristics

Activity attributes:
- Duration
- Location
- Activity type

At home → Working → Leisure → At home
Trips and activities: different characteristics

Activity attributes:
- Duration
- Location
- Activity type

Trip attributes:
- Mode choice
- Trip distance
- Trip duration
- Route choice
Theory of sequence alignment I

Measuring differences between two strings $s [s_1, s_2, \ldots, s_n]$ and $g [g_1, g_2, \ldots, g_n]$

$$d(s,g) = \sum_{i=1}^{n} f(x)$$

$f(x) = 1$ if $s_i \neq g_i$

$f(x) = 0$ if $s_i = g_i$

*Example:*

$s=$ABCDE

$g=AFBCDE$

$d(s,g)=4$

→ Problem of recognising sequential order or duration
Theory of sequence alignment II: Levenshtein

Similarity as total amount of effort to equalise \( s[s_1, s_2, \ldots, s_n] \) and \( g[g_1, g_2, \ldots, g_n] \)

Four basic operation:

- Identity: \( w_e(s_i, g_i) = 0 \)
- Insertion: \( w_i(\emptyset, g_i) = 1 \)
- Deletion: \( w_d(s_i, \emptyset) = 1 \)
- Substitution: \( w_s(s_i, g_i) = w_d(s_i, g_i) + w_i(s_i, g_i) = 2 \)

Definition Levenshtein Distance:
Smallest sum of operation weighting values required to change \( s[s_1, s_2, \ldots, s_n] \) into \( g[g_1, g_2, \ldots, g_n] \):
Theory of sequence alignment III: Trajectories

- Different possibilities to equal two strings
- Combination of operations are called trajectories

*Example*

$s=$CAMBRIDGE
$g=$CAMPING

1) substitute $s_4(B:P)$, $s_5(R:I)$, $s_6(I:N)$, $s_7(D:G)$ delete $s_8(G)$, $s_9(E)$
   $d=10$

2) substitute $s_4(B:P)$, delete $s_5(R)$, substitute $s_6(D:N)$, delete $s_8(E)$
   $d=6$
Theory of sequence alignment IV: Problems

- Qualitative and quantitative data
- Cost of operation weights
- Including duration of activities as attribute or using equal time slices?
- Different attributes
  - Sum of „unidimensional“ sequence alignments across all attribute
  - Optimum trajectory based sequence alignment (Joh et al. 1999)
Software

Dana (C.H. Joh, Universiteit Eindhoven)

- multidimensional
- restricted number of allowed elements per string
- restricted possibilities to change operation weights

ClustalG (C. Wilson, A. Harvey, and J. Thompson)

- Unidimensional
- large strings allowed
- Better possibilities to change operation weights
Dataset Mobidrive

- Reporting period: six weeks
- Travel diary, weekly send out, mailed back and checked via phone
- Cities of Karlsruhe und Halle/Germany
- 162 households, 361 persons
- ca. 52,000 trips and 15,000 days reported September - November 1999 (Pretest: May-July 1999)
- used in analysis: City of Halle (159 persons, 21,000 trips)
Results: Intrapersonal variability over 42 days

Mean daily Levenshtein distance by licence ownership and age

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Not Licenced</th>
<th>Licenced</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>18-24</td>
<td>4.5</td>
<td>2.6</td>
<td>5.6</td>
</tr>
<tr>
<td>25-34</td>
<td>4.8</td>
<td>1.2</td>
<td>5.3</td>
</tr>
<tr>
<td>35-44</td>
<td>5.9</td>
<td>1.9</td>
<td>5.0</td>
</tr>
<tr>
<td>45-54</td>
<td>4.0</td>
<td>2.0</td>
<td>4.5</td>
</tr>
<tr>
<td>55-64</td>
<td>2.9</td>
<td>1.3</td>
<td>4.6</td>
</tr>
<tr>
<td>65 and more</td>
<td>2.5</td>
<td>1.2</td>
<td>4.7</td>
</tr>
<tr>
<td>All</td>
<td>4.0</td>
<td>1.7</td>
<td>4.9</td>
</tr>
</tbody>
</table>
Results: Mean distance from different type of day
Cluster analysis

Legend

+ Cluster 6

Cluster 5

Cluster 4

Cluster 3

Cluster 2

Cluster 1

IntrAPERSONAL variability

Mean distance/trip [km]
### Behavioural characteristics of the clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Mean trip distance [km]</th>
<th>Mean number of trip per day</th>
<th>Mean Levenshtein distance</th>
<th>Cluster size [n]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>6.1</td>
<td>6.0</td>
<td>4.6</td>
<td>42</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>4.0</td>
<td>24.6</td>
<td>3.3</td>
<td>9</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>4.7</td>
<td>5.9</td>
<td>3.4</td>
<td>43</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>8.1</td>
<td>7.6</td>
<td>6.0</td>
<td>9</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>2.4</td>
<td>4.5</td>
<td>2.5</td>
<td>30</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>3.1</td>
<td>12.3</td>
<td>2.6</td>
<td>25</td>
</tr>
<tr>
<td>Overall</td>
<td>4.5</td>
<td>7.8</td>
<td>3.6</td>
<td>158</td>
</tr>
</tbody>
</table>
Sociodemographic characteristics of the clusters

<table>
<thead>
<tr>
<th></th>
<th>Cluster</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Sum</td>
</tr>
<tr>
<td>Mean Age [years]</td>
<td></td>
<td>38.5</td>
<td>41.9</td>
<td>32.7</td>
<td>38.4</td>
<td>43.0</td>
<td>44.4</td>
<td>38.9</td>
</tr>
<tr>
<td>Proportion of females</td>
<td></td>
<td>52.4</td>
<td>22.2</td>
<td>51.2</td>
<td>44.4</td>
<td>53.3</td>
<td>56.0</td>
<td>50.6</td>
</tr>
<tr>
<td>Proportion of parents</td>
<td></td>
<td>47.6</td>
<td>55.6</td>
<td>9.3</td>
<td>55.6</td>
<td>23.3</td>
<td>48.0</td>
<td>33.5</td>
</tr>
<tr>
<td>Proportion of people</td>
<td></td>
<td>26.2</td>
<td>11.1</td>
<td>60.5</td>
<td>11.1</td>
<td>60.0</td>
<td>28.0</td>
<td>40.5</td>
</tr>
<tr>
<td>without licences [%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of fulltime</td>
<td></td>
<td>52.4</td>
<td>88.9</td>
<td>23.3</td>
<td>33.3</td>
<td>26.7</td>
<td>60.0</td>
<td>41.8</td>
</tr>
<tr>
<td>employed people [%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

Day-to-day variability measured with multidimensional sequence alignment:
• Sociodemographic characteristics as expected: highest variability for persons between 25 and 45 years with driving licence
• High variability between weekend days and week days

Clusters based on trip distance, number of trips per day and day to day variability
• 6 cluster solution
• Good differentiation in terms of travel characteristics
• Reasonable differences for the sociodemographic characteristics
Outlook

Further research: Sequence alignment
• Check for other operation costs (deletion, substitution, insertion)
• Check for other weights (Consideration of meaning of different attributes, e.g. mode choice)
• Consideration of duration of activities („long form“)

Further research: Travel behaviour
• Classification based on systematic and comprehensive description of travel behaviour
• Relevance for transport policy
Dataset: fatigue