TRAVEL AND ACTIVITY TEMPORAL RHYTHM OVER A WEEK: RESULTS FROM THE BMW SURVEY IN BELGIUM

Tai-Yu Ma[1], Fabien Walle[2], Eric Cornelis[2], Charles Raux[1]

[1]Transport Economics Laboratory (LET), University of Lyon

[2]Transportation Research Group (TRG), University of Namur

1.INTRODUCTION

The longitudinal analysis of travel behaviour has been an important research issue in transportation research. It analyses individual's mobility behaviour over a period of time, providing useful information about day-to-day variation of individual's travel practice. Moreover, the temporal rhythm of travel demand provides useful information for transport policy decision making. The travel demand analysis based on individual's one-day travel diary may be biased due to day-to-day variation of travel behaviour. The assumption of travel demand stability over a week simplifies the dissimilarity of activity patterns (or activity chain, defined as a chain of activities conducted by an individual over a period of time); however, to what content the activity patterns vary over the day of the week and socio-demographic groups still need to be investigated. To this end, the measures of similarity of activity patterns, reflecting the contextual and sequential relationships of travel/activity characteristics is essential. Moreover, it provides similarity information to analyze temporal rhythm of activity patterns and also clustering analysis of homogeneous groups with similar activity chaining behaviour.

The progress in travel-activity behaviour analysis has developed many approaches to explore temporal rhythm of individual's activity participations. Generally, these approaches can be regrouped into three classes. The first one utilizes time-space prism to represent individual's space-time trajectories to model and simulate activity patterns. This method is originated from Hägerstrand (1970), who proposed an analytical framework to study individual's activity patterns in space-time coordinates. Numerous studies in time-space prism of activity pattern has been proposed (see review in Timmermans et al. 2002; Miller, 2003). As for the analysis of variability of time-space prisms of travel-activity patterns, Kitamura et al. (2006) applied a stochastic frontier model to analyze the variability of locations of prism vertex, aiming to investigating individual's daily schedule timeframe and to what content travel behaviour is affected by its variation. This study demonstrated the usefulness of time-space prism analysis for temporal constraints on travel behaviour.

The second class is based on duration modelling techniques, aiming to determine the effects of explanatory variables on interepisode durations of activity participations over a period of time. Schönfelder and Axhausen (2001) applied a duration model to examine the periodicity of activity participation over six-week. Similarly, Bhat et al. (2004) applied a proportional hazard model to investigate the intershopping durations based on the same data set (Mobidrive). The results provided important insights on the effects of determinants for the frequency and duration of regular and erratic shopping behaviour. Bhat et al. (2005) extended previous research by proposing a unifying multivariate hazard model to examine the interepisode activity durations. This study examined the effects of observed factors (demographic, location, computer use, and day-of-week attributes) and unobserved individual heterogeneity on interepisode durations of individual's activity participations. Ma et al. (2009) investigated individual's daily travel-activity temporal rhythm based on multistate non-homogeneous semi-Markov process. They found travel/activity episode durations depend not only on its beginning time of the day but also on the durations of travel/activity previously conducted. The temporal rhythm of activity durations conducted in different periods of the day was investigated by analyzing the profile of baseline hazards based on Cox proportional hazard model specification.

The third class for temporal rhythm analysis is based on exploratory data analysis techniques. Hanson and Hanson (1981) utilized principle component method to investigate the relationships of activity patterns and respondents' socio-demographic attributes. They proposed a set of indicators to characterize individual's complex travel-activity patterns and identified the relationship with the person and household attributes. The temporal variability of activity patterns has drawn a lot of research interests since 1970. The five-week Uppsala household travel survey in Sweden (1971) provided a well longitudinal data to investigate these questions. Hanson and Huff proposed a series of papers to examine the variability of activity patterns (1982, 1986, 1988). They developed a repetition measure based on trip to capture repetitive travel behaviour based on respondents' travel diaries. These measures involving traveler's stable (repetitive) behaviour can be utilized as a similarity measure of activity patterns. However, the sequential information in activity patterns is neglected. A detailed review of trip-based similarity measures of activity patterns can be found in Schlich and Axhausen (2003). They conducted a comparative study over a variety of trip-based similarity measures based on a six-week diary. They found that the results based on these trip-based measures are similar. They concluded that travel behaviour is more stable on work days than on weekend. They suggested also that the minimum observation period to study day-to-day variability of travel behaviour should be at least two weeks. Neglecting sequential information in comparing the similarity of activity patterns will be biased because it cannot distinguish the situations when elements (activities) with wrong positions but the same order and when same elements with different orders (Joh 2004). The sequence alignment method (SAM) is firstly applied in measuring similarity of activity patterns (Wilson, 1998). The SAM incorporates sequential information of activity patterns into the distance calculation of activity patterns. It presents a very promising way to compare activity patterns. The distance (dissimilarity) between two strings (activity patterns) is calculated as aligning costs to equalize them. This method determines a better distance measure between two one-dimensional sequences. However, when multiple attributes with different scales are considered in activity pattern comparison, the one-dimensional alignment method becomes problematic. Joh et al. (2002, 2004) proposed a multidimensional alignment approach by taking into account the dependency of attributes in calculating the pattern similarities. Applications of multi-dimensional SAM in intra-/inter- personal variability of activity patterns can be found in Schlich et al. (2004). They examined temporal and spatial variability of leisure activities based on Mobidrive dataset and compared activity patterns based on trip-based and sequence-based measures.

In this study, two similarity measures are applied to investigate intrapersonal and interpersonal variability of daily activity chains and time use distributions. The first measure for activity chain similarity is based on one-dimensional SAM. The second for individual's time-use distribution is based on Kullback-Leibler distance. This measure complete the similarity of activity chain in its temporal aspect. Based on the similarity measures of activity patterns, a clustering analysis is conducted to elucidate the relationship between activity patterns and individual's and household characteristics. The canonical discriminant analysis and canonical correlation analysis are applied to investigate the influence of explanatory variables on cluster membership and the correlations of activity chain types and individual's sociodemographic variables. Finally we conclude this study and discuss important findings and future extensions.

2. MEASURING SIMILARITY OF ACTIVITY PATTERNS BASED ON SEQUENCE ALIGNMENT METHOD AND KULLBACK-LEIBLER INFORMATION

2.1 Sequence alignment method

The similarity of activity chains (patterns) has been a widely study research issue in activity-based travel demand analysis. The comparison of activity patterns is the basis for longitudinal analysis of travel behaviour. The question concerns how different activity patterns vary over a week. In the past, most similarity indicators compare the corresponding elements between a pair of activity patterns. The distance is measured by summation of 0-1 scores resulting from pairwise comparison of sequences. The sequential (order) information between elements is neglected. As the sequential relationships reveal the dependency of travel/activity participations of individual's activity chaining behaviour, it is important to take it into account. For this issue, Wilson (1998) firstly introduced a Sequential Alignment Method (SAM) for activity pattern analysis. The SAM allows comparing the similarity between a pair of sequences by incorporating its ordering information. The method is originated from molecular biology, aiming to identifying segments of similarity reflecting some functional relationship between sequences of DNA or protein. The sequence is defined as a set of ordered elements arranged as a string. The difference (dissimilarity) between sequences is evaluated as the efforts to equalize them, which allows one takes into account the sequential information in the measurement of similarity between sequences. The SAM has gained its popularity in comparing the similarity between activity patterns recently (Schlich and

Axhausen, 2004; Shimamoto et al., 2009). Some methodological advancement in the SAM for activity pattern analysis has also been reported (Joh et al. 2002; Wilson, 2008).

The similarity of activity patterns can be measured based on a variety of attributes of each activities between two activity chains, such as activity type, transport mode, destination and activity duration etc. If we consider only one attribute, the ordered attribute values represent corresponding travel/activity characteristics in the activity chain. To measure the similarity, a distance function needs to be defined. The SAM utilized a process for which a set of operations, i.e., deletions, insertions and substitutions, are utilized to equalize the sequences. Each type of operations is associated with a cost (score) for which substitutions are generally assigned as two times of deletion/insertion costs. This is because one substitution needs to implement one deletion and one insertion. For example, to equalize the sequence [ACB] to [ABC], one needs to delete [C] before [B] and inset it at the end of the sequence [AB]. The corresponding cost is 2. More precisely, in one dimension case, the dissimilarity $\mathrm{d}(s_1, s_2)$ between sequences s_1 and s_2 can be calculated as the smallest summation of these operations costs or *alignment cost*, namely,

$$\mathrm{d}(s_1, s_2) = w_\mathrm{d}\mathrm{o}_\mathrm{d} + w_\mathrm{i}\mathrm{o}_\mathrm{i} + w_\mathrm{r}\mathrm{o}_\mathrm{r} \tag{1}$$

where w_d , w_i and w_r is the costs (weighting coefficients) of deletion, insertion and replacement, respectively. The distance (Levenshtein distance) between two sequences can be calculated by applying a dynamic programming algorithm, computing the least number of operations necessary to equalize two sequences (Table 1). As activity patterns are generally characterized by different attributes, the similarity between these patterns should take into account all important attributes, which make the calculation of similarity more complex. As mentioned by Schlich and Axhausen (2004), different attributes may depend on each others, i.e. trip duration, transport mode and destination choice are correlated. Hence, the similarity between activity chains with multiple attributes cannot be obtained by simply summing uni-dimensional alignment costs for all attributes. For this issue, Joh et al. (2002) proposed a multidimensional alignment method by taking into account the interdependence information among the attributes in activity patterns. The idea is that first a multidimensional sequence is constructed with each line representing corresponding attribute. The operation set for each attribute is identified based on the calculation of Levenshtein distance. If the operations for attributes are identical, the costs are counted once as one simultaneously aligns a bundle of elements. They further proved that the minimum cost of multidimensional alignment can be obtained efficiently by searching the combination of one-dimensional alignment results for each attribute. Joh (2004) proposed a heuristic method to calculate the multidimensional alignment costs. However, some issues concerns the choice of attributes and the categorisation of interval-scaled variable such as activity duration or trip distance still lack theatrical justification. Because different ways of selecting, weighting and categorizing attributes may generate trivial results, we applied one-dimensional alignment method to compare activity type sequences. The results is completed by another similarity measure on temporal aspect based on Kullback-Leibler distance, capturing inter- and intra-personal variability in time-use distribution.

Table 1 Dynamic algorithm for the computation of Levenshtein distance (Kohonen, 2001)

```
1. Begin: d(0,0) = 0
2. for i = 1 to length(A) do d(i,0) := d(i-1,0)+w_d D(A(i));
3. for j = 1 to length(B) do d(0,j) := d(0,j-1)+w_i I(B(j));
4. for 1 to length(A)
5. for j = 1 to length(B)
      a. m1 := d(i-1,j-1)+w_r R(A(i), B(j));
      b. m2 := d(i,j-1)+w_i I(B(j));
      c. m3 := d(i-1,j)+w_d D(A(i));
      d. d(i, j)= min(m1,m2,m3);
6. end
7. end
8. LD(A,B) = d(length(A),length(B));
9. End
```

Remarks: 1. LD(A, B) at line 8 denotes Levenshtein distance of A and B. This algorithm aims to calculate minimum cost to equalize two sequences (strings).
2. D(A(i)) denotes "*delete* ith element of A", I(B(*j*)) denotes "insert jth elements of B", R(A(i), B(j)) denotes "replace A(i) by B(j)".

2.2 Kullback-Leibler distance for measuring of the dissimilarities of time-use distributions

To compare activity patterns in time aspect, we introduce Kullback-Leibler (KL) distance (divergence) to measure the dissimilarity of individual's daily time-use distributions, extending one-dimensional alignment method in activity pattern analysis. The KL distance measures the differences between two probability distributions. The daily time-use distribution is calculated based on daily activity/trip duration over 24 hours. Each number of discrete variable represents one activity type or trip. The KL distance between two distribution p and q is defined as

$$\mathrm{D}_{\mathrm{KL}}(p,q)=\sum_i p_i \log\frac{p_i}{q_i} \qquad (2)$$

where p_i is the probability of discrete variable *i*. We define $0\log\frac{p_i}{0}=0$ and $p_i\log\frac{p_i}{0}=\infty$

One problem with KL distance in our case is that the probabilities of certain activities are zero, which generates a value of infinity. Another problem is that the KL measure is asymmetric, which is generally not desirable for our analysis. To address these issues, an alternative measure is Jensen–Shannon divergence or information radius (IRad) is defined as follows:

$$\mathrm{IRad}(p,q) = \mathrm{D_{KL}}(p, \frac{p+q}{2}) + \mathrm{D_{KL}}(q, \frac{p+q}{2}) \tag{3}$$

The information radius measures total differences of two probability distributions to their average distribution. It is symmetric with finite range from 0 (identical distributions) to $2\log 2$ (maximally different). We apply the IRad measure to investigate inter-/intra- personal variability of individual's time-use distributions. The activity type is based on the seven regrouped activity categories with additional trip categories.

3.DATA DESCRIPTION AND DESCRIPTIVE STATISTICS

The data for the analysis is based on seven-day travel diary collected in the city of Ghent in Belgium. The objective of this survey is to investigate individual's weekly activity patterns and their impact on day-to-day variation of travel behaviour. The surveyed households are randomly drawn from the population in the city of Ghent based on the stratification of household size and age of household head. The surveyed individuals are based on randomly selected individuals in the household because sampling whole household members over a week may reduce the response rate. The survey methodology is based on paper and web survey followed by phone support. Although this survey cannot collect the activity patterns of all members in the household, it still allows us to investigate individual's daily activity patterns and the determinants related to individual's socio-demographic characteristics. The collected information contains continuous trip chain information over a week (trip purposes of 12 categories, approximate address of destination, departure and arrival time of trip, travel cost, used modes and travel time) and its potential influence factors (socio-demographic characteristics and mobility practices). The survey was conducted from September to November 2008 for which totally 717 individual's full 7-day (starting from any day within a week) mobility diary were collected.

For activity purposes, 12 activity types are defined in the survey. To obtain enough samples over different types of activities, the initial twelve activity purposes are regrouped into seven categories as: 1 home (go home), 2 work (go to work), 3 school (go to school), 4 shopping (shopping for the basic needs and shopping), 5 personal business (personal business (bank, doctor etc.)), 6 social-recreation (eating, visiting families or friends, take a walk, and leisure, sport and culture activities), 7 others (drop off/pick up someone and others). The analysis of intrapersonal and interpersonal variability is conducted with respect to these regrouped activity types. The descriptive statistics of trip characteristics concerning average daily travel time, average number of trip per day, average travel time per trip (by trip purpose and by mode) and average travel distance per trip (by trip purpose and by mode) over a week is reported in Annex 1.

4. RESULTS OF ANALYSIS

4.1 Intrapersonal variability over a week

The average Levenshtein distance and information radius between all days are calculated as the average distance of all individuals based on individual's average distance of each day to the other days of the week. The results in Table 2 suggest that there is distinct difference between weekdays and weekend for both measures. It is reasonable to find that the average Levenshtein distance is higher on weekends than on weekdays, reflecting the distinction of activity chains conducted on weekdays and weekends. In general, activity patterns on weekdays are work-based. However, on weekend, social-recreation activities turn out to be major activity participations. For activity duration distributions, it is interesting to find that the average IRad measure is almost the same on weekdays. By contrast, it is higher on weekends. The result reflects individual's daily time-use distributions are more similar between weekdays than weekends.

To investigate the influence of sociodemographic characteristics on these similarity measures, the data is stratified into homogeneous classes based on two hieratical structures respectively. The hieratical structures are defined by: a. gender, employment status and household position; b. age and driving license. The results are shown in Table 3 and 4. For male fulltime workers with husband role in the household, they have the largest average Levenshtein distance, reflecting their larger intrapersonal variability than other members. By contrast, for part-time male workers, household head has largest variability. For female full-time and part-time worker with wife role in the household, she has largest intrapersonal variability compared with other roles in the household. The results suggest that female with wife role in the household ensures the major responsibility of household maintenance and then takes a more various activities than others roles. For daily time-use distribution, there are no significant difference between household head and household husband/wife over genders and employment status. The results showed in Table 3 indicate that men tend to have larger intrapersonal variability of time-use distribution than women, and full-time and part-time workers have significant variability than non-workers.

When regarding the influence of age and driving license on intrapersonal variability, the results in Table 4 reveal that respondants with driving license of age between 35-44 years have the highest variability. By contrast, for respondants without driving license, the age category of 25-34 years obtains the highest variability. For time-use distribution, the information radius in the categories of 18-24, 25-34 and 35-44 are higher.

Table 2 Average intrapersonal Levenshtein distance and Information radius by day of the week

		Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
Levenshtein distance	Mean	3.75	3.66	3.87	3.79	3.95	4.38	4.03
	S.D.	1.86	1.76	1.99	1.92	2.08	2.10	1.78
Information radius	Mean	0.17	0.17	0.17	0.17	0.18	0.22	0.23
	S.D.	0.08	0.08	0.09	0.08	0.09	0.11	0.12

Table 3 Average intrapersonal variability of daily activity patters by gender, work status and relation to the household head over a week

				Average intrapersonal Levenshtein distance		Average intrapersonal IRad	
Gender	Work status	Household position	N. Obs	Mean	SD	Mean	SD
Male	full time	head	148	4.06	1.70	0.22	0.06
		husband/wife	3	5.40	2.01	0.22	0.07
		child	30	3.32	1.34	0.23	0.07
		other	7	2.56	1.01	0.14	0.05
	part time	household head	14	4.95	1.64	0.22	0.07
		husband/wife	1	3.24	-	0.22	.
		child	4	4.00	0.66	0.19	0.11
	unemployment	household head	81	3.69	1.44	0.12	0.06
		child	58	3.03	1.55	0.18	0.07
		other	1	4.19	-	0.34	.
Female	full time	household head	46	4.06	1.43	0.20	0.06
		husband/wife	45	4.37	1.47	0.20	0.06
		child	31	3.72	1.91	0.22	0.08
		other	2	2.90	1.28	0.18	0.06
	part time	household head	22	4.89	1.63	0.20	0.07
		husband/wife	48	4.98	2.02	0.19	0.06
		child	3	3.75	0.38	0.24	0.06
	unemployment	household head	24	3.38	1.25	0.15	0.06
		husband/wife	78	3.80	1.50	0.13	0.06
		child	69	3.73	1.49	0.20	0.07
		other	2	4.19	2.15	0.24	0.08

Table 4 Average intrapersonal variability of daily activity patters by age and driving license ownership over a week

	Average intrapersonal Levenshtein distance						Average intrapersonal IRad			
	With driving license			Without driving license			With driving license		Without driving license	
	N. Obs	Mean	S.D.	N. Obs	Mean	S.D.	Mean	S.D.	Mean	S.D.
<18	75	2.90	1.11	2	2.0	1.0	0.19	0.05	0.21	0.09
18-24	81	3.98	1.73	1	4.6	.	0.22	0.09	0.19	.
25-34	131	4.11	1.59	7	4.9	1.2	0.21	0.07	0.27	0.05
35-44	148	4.48	1.80	8	4.3	1.3	0.20	0.06	0.23	0.05
45-54	83	3.94	1.73	9	3.9	1.7	0.19	0.07	0.23	0.05
55-64	92	3.80	1.53	5	4.8	1.7	0.14	0.07	0.21	0.07
>64	65	3.39	1.29	10	3.3	1.1	0.12	0.05	0.11	0.06
All	75	3.91	1.65	2	2.0	1.0	0.19	0.05	0.21	0.09

4.2 Interpersonal variability over a week

The obtained similarity at previous section can be utilized as an indicator to classify persons with similar travel diaries. Based on obtained classes of homogeneous activity chain or time-use groups, we can investigate its statistical characteristics of travel/activity participations and also the influence of explanatory variables. To this end, we identify groups of persons based on average interpersonal Levenshtein distance for activity chains and average interpersonal Information radius for time-use distributions.

For each person, the average interpersonal Levenshtein distance or Information radius are calculated as the average difference with other persons by comparing each day of the week. The resulting average similarity/dissimilarity reflects personal average differences of activity chains or time-use distributions. Based on the distance matrix between persons, we classify individuals into groups such that within-group individuals with similar feature than they are to individuals within other groups. Many clustering methods can be applied for partitioning data. One of most utilized clustering methods is hierarchical clustering method. This method needs to determine how to measure the inter-cluster distances between clusters. The clustering result is a hierarchical partition of data. The analyst then selects a small number of clusters based on predetermined choice. Another category of clustering method is a non-hierarchical method for which k-means method is widely utilized. The k-means method classifies data based on the minimisation of within-group sum of squares. The advantage of k-means approach is that it need not to define the inter-cluster distance and suitable for large-scale data set clustering.

The result of cluster solutions based on k-means method for Levenshtein distance is shown in Fig. 1. The determining of number of clusters is based on the within-group sum of squares in the data set. The choice depends then on the relationship between the percentage of variance explained and the number of clusters. The percentage of variance explained is calculated as the ratio of the between-cluster variance to the total variance. As shown in left part of Fig. 1, the within-group sum of square decrease rapidly when increasing number of clusters at its early stage and then stabilizes. The choice of number of cluster is preferred as 4. It explains 74.69% total variance in the data set. The plotting of 4-cluster solution is shown in the right part of Fig. 1 based on first and second discriminant functions. The result indicates that persons in cluster 4 have higher variance of Levenshtein distance compared with persons in other 3 clusters.

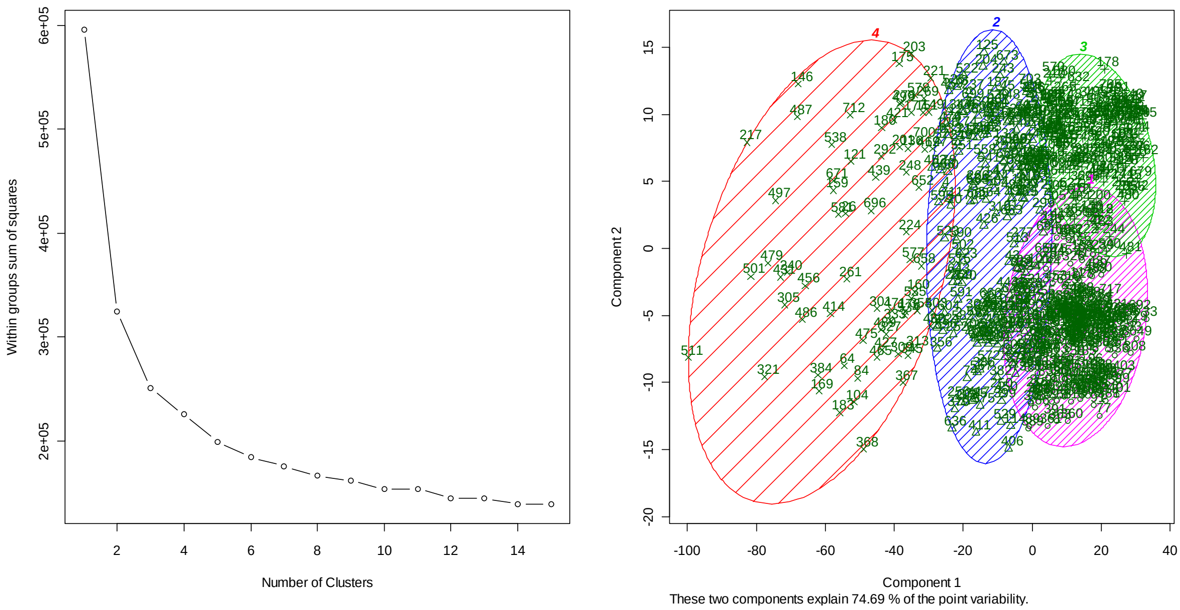

Figure 1. Cluster analysis based on k-means method for average intrapersonal Levenshtein distance. The sum of distance from points to assigned cluster centers over number of clusters (left). The plot of 4-cluster solution over first two principal components (right).

For average interpersonal Information radius, the clustering result is reported in Fig. 2. The plot of within group sum of square against the number of clusters suggests a 4 cluster solution of the data set, explaining 83.36% total variance. The plotting of data points against two first discriminant functions shows that persons in cluster 1 (red region) have larger variance compared with other groups.

The average interpersonal Levenshtein distances and information radius for each cluster are shown in Table 5. To investigate the influence of types of days on these measures with each cluster, the average interpersonal Levenshtein distances and information radius over different types of days are calculated. The results indicate that persons in cluster 4 and 2 have higher interpersonal variability of daily activity chain on weekdays and weekends. However, persons in cluster 1 and 3 have similar interpersonal variability on weekdays (2.79 and 2.68) but with more different variability on weekends (2.95 and 3.38). For time-use distribution, the average interpersonal Information radius shows that persons in cluster 2 and 4 have larger time-use variability on weekdays. For weekends two categories are distinguished: cluster 1, 2 and 4 (relatively higher, between 0.23 and 0.33), cluster 3 (relatively lower, 0.13). To understand to what content the time-use distribution differentiate over clusters, Fig. 3 reports time-use distributions of out-of-home activities over weekdays and weekends. The result indicates that on weekdays, persons in cluster 2, 3 and 4 spend more time work. The profiles of time-use distributions of clusters depend on personal and household characteristics. The sociodemographic characteristics of clusters are reported in Table 6 and Figure 4, 5, 6 and 7. Travel characteristics for clusters based on Levenshtein distance for different types of day are reported in Appendix 2.

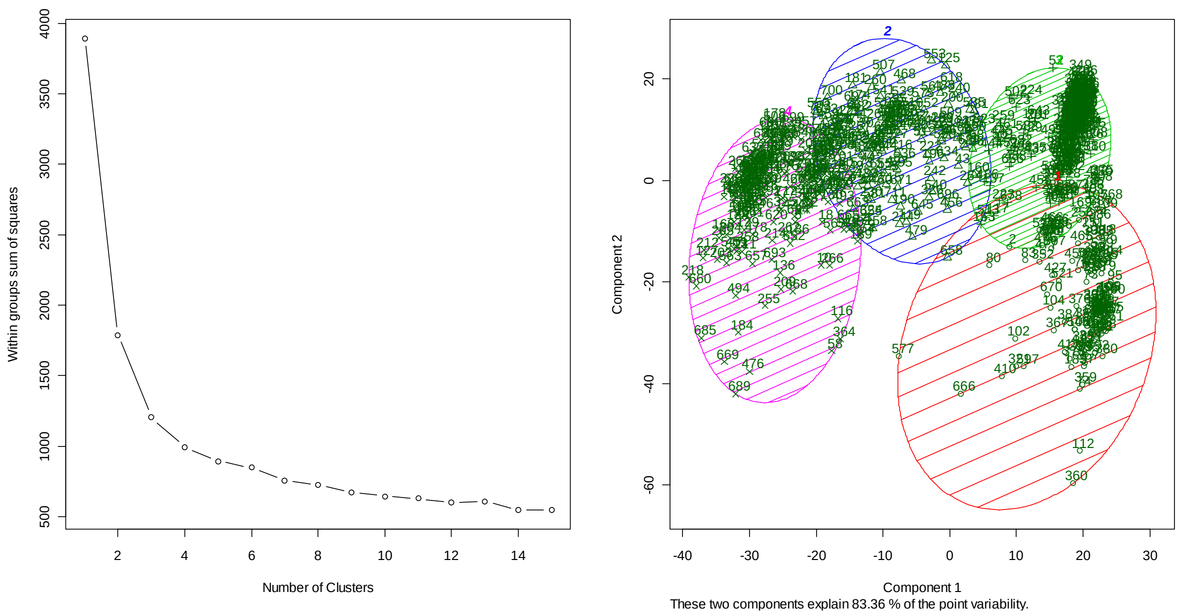

Figure 2. Cluster analysis based on k-means method for average intrapersonal Information radius. The sum of distance from points to assigned cluster center over number of clusters (Left). The plot of 4-cluster solution over first two principal components (right).

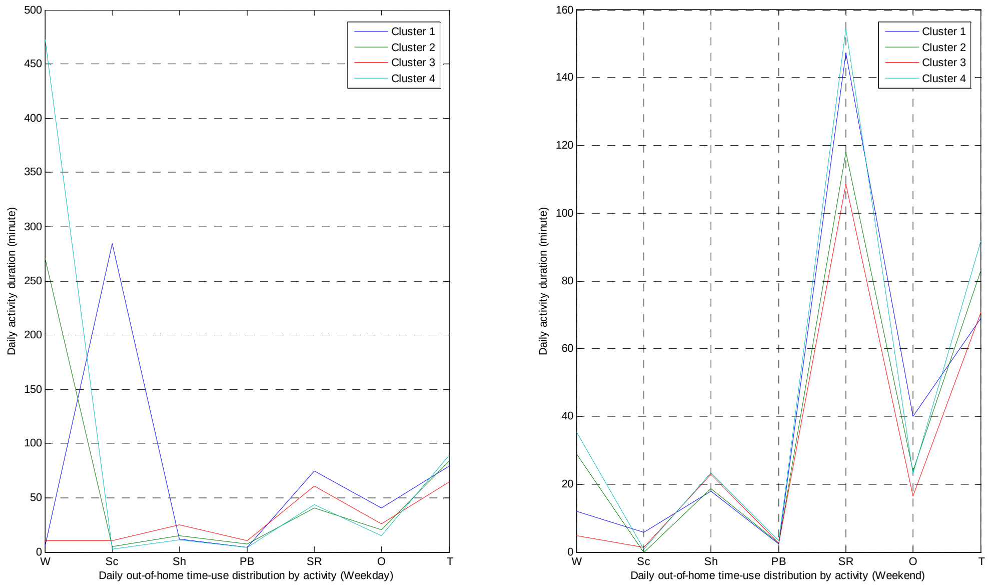

Figure 3. 4-cluster solution of average daily time-use distribution over out-of-home activities for weekdays (left) and weekends (rights)

Table 5 Average interpersonal Levenshtein distance and Information radius by cluster over weekday and weekend

		Weekday				Weekend			
	Cluster	C1	C2	C3	C4	C1	C2	C3	C4
Levenshtein distance	N. Obs	1220	1180	820	365	488	472	328	146
	Mean	2.79	4.67	2.68	6.92	2.95	5.15	3.38	7.21
	Std Dev	1.17	1.43	1.19	1.92	1.16	1.58	1.27	1.63
	Cluster	C1	C2	C3	C4	C1	C2	C3	C4
Information radius	N. Obs	605	755	1225	1000	242	302	490	400
	Mean	0.18	0.21	0.12	0.20	0.26	0.23	0.13	0.33
	Std Dev	0.09	0.07	0.07	0.08	0.09	0.07	0.08	0.09

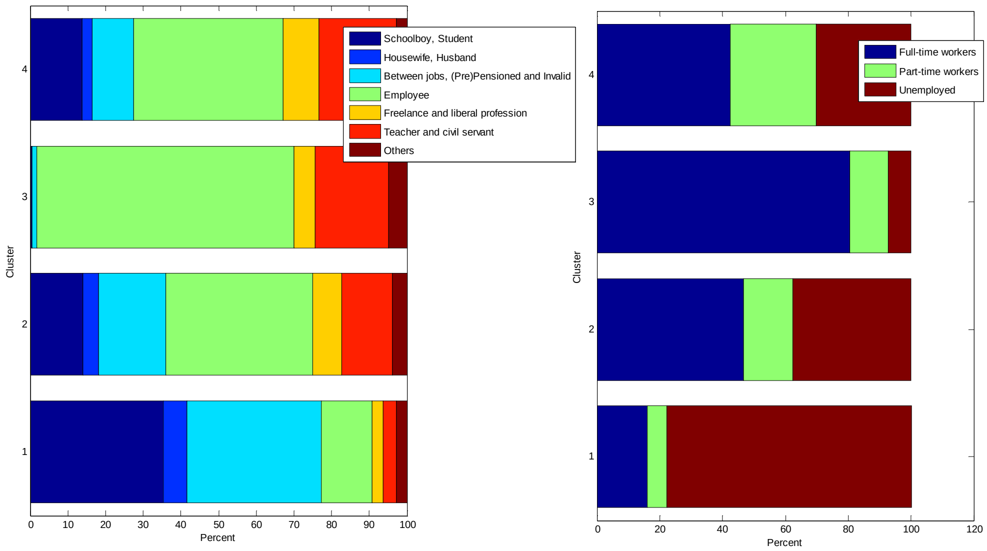

Figure 4 Profession (left) and working status (right) of 4 clusters based on Levenshtein distance

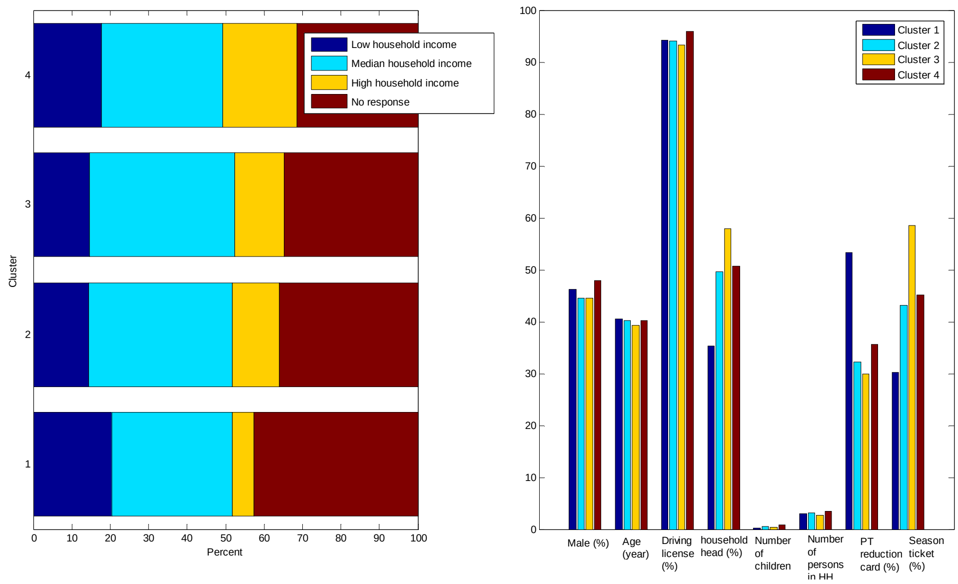

Figure 5 Household income (left) and other personal and household characteristics (right) of 4 clusters based on Levenshtein distance

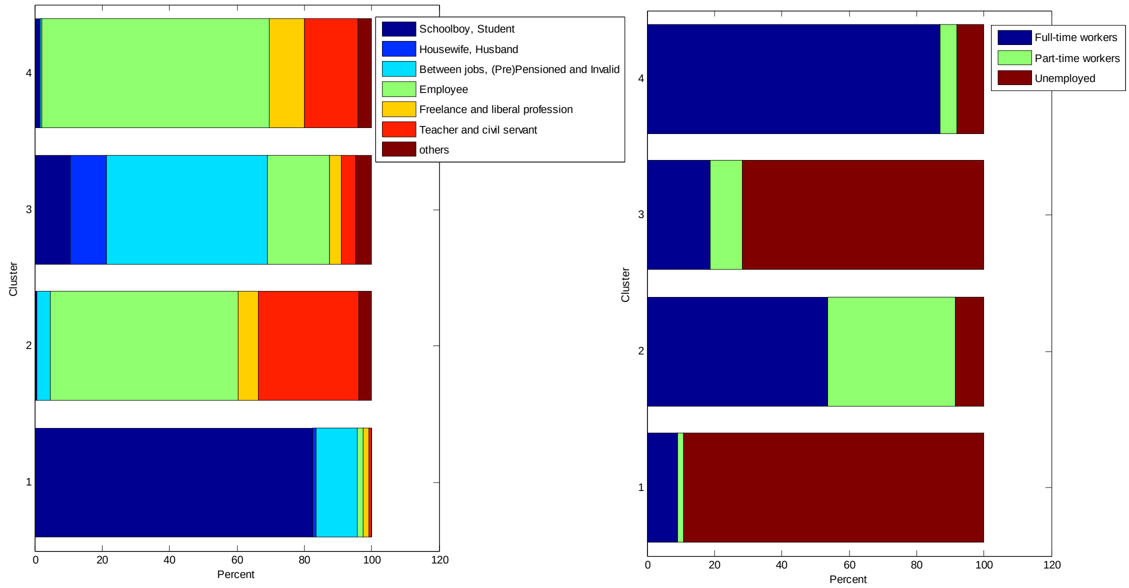

Figure 6 Profession (left) and working status (right) of 4 clusters based on Information radius

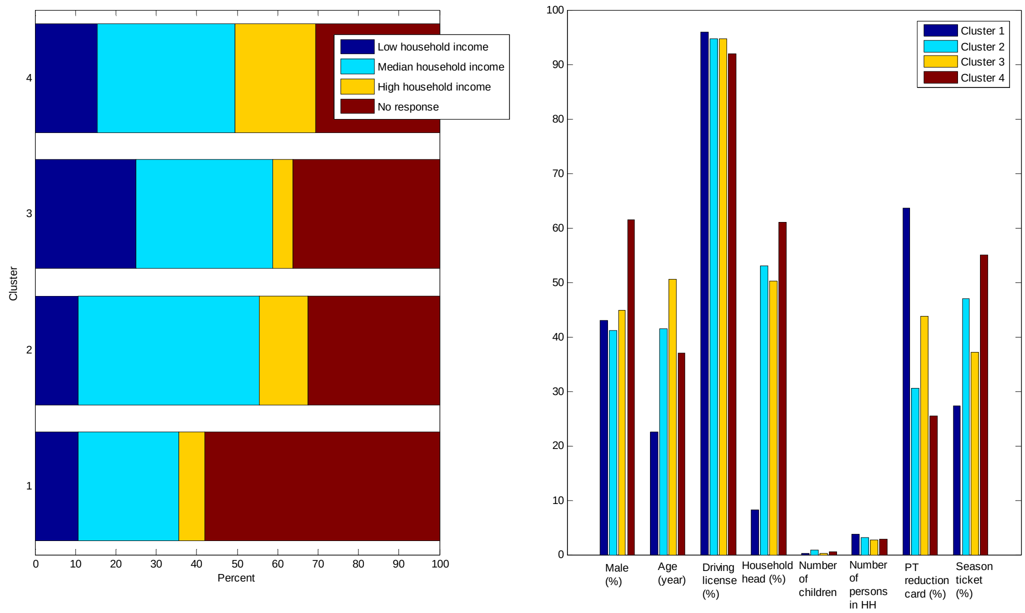

Figure 7 Household income (left) and other personal and household characteristics (right) of 4 clusters based on Information radius

Table 6 The characteristics of personal and household characteristics of 4 clusters based on average interpersonal Levenshtein distance and Information radius

	Average interpersonal Levenshtein distance				Average interpersonal Information radius			
	C1	C2	C3	C4	C1	C2	C3	C4
Male (%)	46.3	44.5	44.5	48.0	42.98	41.1	44.9	61.5
Average age (years)	40.5	40.2	39.3	40.3	22.6	41.4	50.5	37.0
Driving license (%)	94.3	94.1	93.3	95.9	95.9	94.7	94.7	92.0
Full-time workers(%)	16.0	46.6	80.5	42.5	9.09	53.6	18.8	87.0
Part-time workers (%)	6.2	15.7	12.2	27.4	1.65	37.8	9.4	5.0
Unemployed (%)	77.9	37.7	7.3	30.1	89.3	8.6	71.8	8.0
Schoolboy/Student (%)	35.3	14.0	0.6	13.7	82.6	0.7	10.6	1.5
Housewife/Husband (%)	6.2	4.2	0.0	2.7	0.83	0.0	10.6	0.0
Between jobs, (Pre)Pensioned and Invalid (%)	35.7	17.8	1.2	11.0	12.4	4.0	47.8	0.5
employee (%)	13.5	39.0	68.3	39.7	1.7	55.6	18.4	67.5
Freelance and liberal profession (%)	2.9	7.6	5.5	9.6	1.7	6.0	3.7	10.5
Teacher and civil servant (%)	3.7	13.6	19.5	20.6	0.8	29.8	4.1	16.0
Being household head (%)	35.3	49.6	57.9	50.7	8.3	53.0	50.2	61.0
Number of children less than 12 years of age	0.3	0.5	0.4	0.9	0.3	0.8	0.3	0.5
Number of persons in the household	3	3.1	2.7	3.5	3.7	3.2	2.7	2.9
Low household income (%)	20.5	14.4	14.6	17.8	10.7	10.6	24.9	15.5
Median household income (%)	31.2	37.3	37.8	31.5	24.8	45.0	33.9	34.0
High household income (%)	5.7	12.3	12.8	19.2	6.6	11.9	4.9	20.0
Public transport reduction card (%)	53.3	32.2	29.9	35.6	63.6	30.5	43.7	25.5
Season ticket (%)	30.3	43.2	58.5	45.2	27.3	47.0	37.1	55.0

Remark: Household income category: *low* if household income is between 0 and 2000 euros per month; *median* if household income is between 2001 and 4000 euros per month; high if household income is more than 4000 euros per month. Note that the percent of no answer is 35.8%, the causion should be made for the interpretation.

4.3 Canonical discriminate analysis

In this section, the discriminant analysis is applied to determine the influence of personal and household characteristics on cluster membership. For multiple clusters (group) discriminant analysis, the objective is finding a linear combination of explanatory variables best describing k separated clusters. Moreover, the discriminant function can also predict individual's allocation of clusters. The reader is referred to Khattree and Naik (2000) for detail description.

For k clusters discriminant analysis, the number of discriminant functions utilized for separating clusters equals to the smaller of k or p (number of explanatory variables). The estimated eigenvalues represent the contribution of corresponding discriminant functions for separating total clusters. The relative importance of discriminant functions contributing for the separation of clusters can be evaluated by its proportion of total eigenvalues. Based on this criterion, we can select maximal two or three eigenvalues sufficient for separating clusters. One can also utilize Wilk's Λ test for the significance of separation of clusters. This statistics is similar to F-statistics by comparing between-sum of squares to within-sum of squares of grouped data. We conduct discriminant analysis for 4 cluster solution based on Levenshtein distance and Information radius, respectively. The included personal and household variables are listed in Table 7. The stepwise discriminant method is conducted for final model specification. This method is similar to stepwise regression model specification by selecting variable step-by-step based on its statistical significance.

Table 7 List of personal and household characteristics for discriminant analysis

Variable	Definition
gender	1 if male, 0 otherwise
age	Age of the individual
license_car	1 if the individual has driving licence, 0 otherwise
student	1 if the individual is a student, 0 otherwise
full_time_job	1 if the individual has a full-time job, 0 otherwise
part_time_job	1 if the individual has a part-time job, 0 otherwise
unemployment	1 if the individual is unemployment, 0 otherwise
being_hhead	1 if the individual is the head of the household, 0 otherwise
epouse	1 if the individual is the epouse/husband of the household, 0 otherwise
n_child_12	Number of young children less than 12 years of age in the household
n_household	Number of persons in the household
hhincom_low	1 if the household income category is low, 0 otherwise
hhincom_median	1 if the household income category is median, 0 otherwise
hhincom_high	1 if the household income category is high, 0 otherwise
PT_reduction	1 if the individual has public transportation reduction card, 0 otherwise
seasonticket	1 if the individual has season ticket, 0 otherwise

4.3.1 Discriminant analysis of 4 cluster solution based on Levenshtein distance

The explanatory variables included in final specification of discriminant function based on stepwise discriminant method are: unemployment, n_child_12, full_time_job, student, seasonticket, age, PT_reduction and n_household. The results reveals that the cluster membership depends on individual's working status, household size, the number of young children and public transportation reduction cards. Gender, household income and driving licence have no significant effects. The proportion of eigenvalue for the first and second discriminant functions is 87% and 12%, respectively. It means the first discriminant function has a major contribution to the separation of clusters. The tests of significance of discriminant power for each function are showed in Table 8. It indicates that two discriminant functions that maximally separate clusters are statistical significant at 0.0001 level. The correlations between cluster membership and explanatory variables are reported on the left part of Table 9. The results reveal that for the first discriminant function, work status has most significant influence for cluster membership. For the second function, the number of children less than 12 years in the household and the number of persons in the household contribute most to cluster membership. The value of Wilks Λ-test is 0.61 with p-value <.0001 indicating that the means of all clusters are significant different. The estimated raw coefficient for discriminant functions 1 and 2 are shown on the right part of Table 9.

Table 8 Canonical correlations and explanatory contributions of discriminant functions

Discriminant function	Canonical Correlation	Eigenvalue	Proportion	Cumulative	Test of H0: the current canonical correlation and all smaller ones are zero		
					Likelihood Ratio	Approximate F Value	Pr > F
1	0.58	0.51	0.86	0.86	0.61	15.80	<.0001
2	0.27	0.08	0.13	0.99	0.92	4.25	<.0001
3	0.08	0.01	0.01	1.00	0.99	0.80	0.57

Table 9 Coefficients of discriminant functions and the correlation of cluster membership and explanatory variables

Variable	Correlation of cluster membership and explanatory variables		Raw coefficients of discriminant functions	
	Function 1	Function 2	Function 1	Function 2
unemployment	0.93	-0.19	0.11	0.67
n_child_12	-0.16	0.73	-0.47	-1.57
full_time_job	-0.83	-0.21	1.34	0.39
student	0.59	-0.11	-0.49	-0.45
seasonticket	-0.37	-0.04	0.02	0.02
age	0.04	0.04	-0.22	-0.71
PT_reduction	0.33	-0.21	0.04	0.32
n_household	0.11	0.55	1.45	-1.44

4.3.2 Discriminant analysis of 4 cluster solution based on information radius

For the clustering results based on information radius, the stepwise discriminant analysis identifies 8 variables for final specification of discriminant functions: student, unemployment, full_time_job, age, hhincom_high, epouse,n_child_12 and license_car. Similarly to previous section, individual's work status and age have significant impact on cluster membership. Note that household characteristics play important role in cluster membership of individual's time-use distributions, including the role in the household, the number of children less than 12 years of age and household income. The contributions of three discriminant functions for identifying clusters are listed in Table 10. The results indicate that function 1 and 2 contribute most for separating the clusters with proportional contribution 73% and 21%, respectively. The correlation of cluster membership and explanatory variables indicates that employment status determines most for data clustering for the first discriminant function. For the second function, age is the most important determinant. The value of Wilks

Λ-test is 0.19 with p-value <.0001 rejecting the null hypothesis that all cluster means are the same.

Table10 Canonical correlations and explanatory contributions of discriminant functions

Discriminant function	Canonical Correlation	Eigenvalue	Proportion	Cumulative	Test of H0: the current canonical correlation and all smaller ones are zero		
					Likelihood Ratio	Approximate F Value	Pr > F
1	0.81	1.92	0.73	0.73	0.19	65.98	<.0001
2	0.60	0.55	0.21	0.94	0.55	34.59	<.0001
3	0.38	0.16	0.06	1.00	0.86	19.38	<.0001

Table 11 Coefficients of discriminant functions and the correlation of cluster membership and explanatory variables

Variable	Correlation of cluster membership and explanatory variables		Raw coefficients of discriminant functions	
	Function 1	Function 2	Function 1	Function 2
student	0.88	-0.44	3.26	-1.04
unemployment	0.79	0.50	1.49	0.72
full_time_job	-0.63	-0.55	-0.05	-0.82
age	-0.32	0.83	0.01	0.04
hhincom_high	-0.17	-0.23	-0.50	-0.30
epouse	0.07	0.41	0.17	1.10
n_child_12	-0.22	-0.17	-0.04	0.04
license_car	0.05	0.03	0.34	0.70

4.4 Canonical correlation analysis of activity chain patterns

In this section, we apply canonical correlation analysis to investigate the correlation between activity chains and individual's social-demographic characteristics for two types of day of the week: weekday and weekend. The activity chain in this section is related only for activity chains with at least one out-of-home activities. The canonical correlation analysis is widely applied multivariate data analysis technique aiming to examine the association between two sets of variables. Especially when the association occurs not only between two different variable sets (activity chain and individual's social-demographic characteristics for our study), but also within the same set of variables (the activity chains pursued by an individual on weekdays or weekends is highly correlated within them). Moreover, this technique is also a data reduction technique by transforming initial correlated variables to a reduced number of independent variables (*canonical variates*) for data exploration. In canonical correlation analysis, the variates are similar to factors in component analysis, but differently they are computed based on the maximum of the correlation of two sets of variables. The number of canonical variates equals to the number of variables in smaller set. The reader is referred to Khattree and Naik (2002), Kuylen and Verhallen (1981) for detail description.

We consider activity chains conducted by individuals as “response” variables. As the types of activity chains conducted on weekdays and weekends are basically different, the canonical correlation analysis is conducted for the two types of days. Moreover, since a variety of types of activity chains presents in the data, we select a number of activity chains based on its observed frequency. Totally 18 types of activity chains are included in the analysis for each type of day, accounting for 50.54% cumulative frequencies of total observed activity chains on weekday and 67.58% on weekends. The explanatory variables are listed in Table 7, the same as discriminant analysis.

The observed frequencies of main activity chains are listed in Table 12. The result indicates that on weekdays, 2-trip pattern is the most frequent in which work, School, Shopping and social-recreation activities are main out-of-home activity purposes. For weekends, 2-trip, 3-trip and 4-trip activity chains are main patterns. Social-recreation and shopping activities are main types of out-of-home activities. Note that personal business activity does not appear in most frequent activity chain patterns on weekdays and on weekends.

Table 12 Main activity chain patterns for weekdays and weekends

Weekday				Weekend			
Activity chain	Freq.	%	Cumu. Freq. %	Activity chain	Freq.	%	Cumu. Freq. %
H-W-H	392	11.43	11.43	H-SR-H	162	12.69	12.69
H-Sc-H	169	4.93	16.36	H-Sh-H-SR	102	7.99	20.67
H-W-H-SR	131	3.82	20.18	H-Sh-H	71	5.56	26.23
H-Sh-H	128	3.73	23.91	H-SR-H-SR	70	5.48	31.71
H-SR-H	119	3.47	27.38	H-SR-SR-H	70	5.48	37.20
H-Sh-H-SR	81	2.36	29.75	H-Sh-H-Sh	51	3.99	41.19
H-Sc-H-SR	74	2.16	31.90	H-SR	40	3.13	44.32
H-W-H-Sh	71	2.07	33.97	H-O-H	38	2.98	47.30
H-O-H-O	70	2.04	36.02	H-SR-SR-SR	36	2.82	50.12
H-O-H	63	1.84	37.85	H-H-SR-H	34	2.66	52.78
H-W-SR-H	59	1.72	39.57	H-H	33	2.58	55.36
H-W-Sh-H	59	1.72	41.29	H-Sh-Sh-H	30	2.35	57.71
H-O-W-H	55	1.60	42.90	H-Sh-H-O	25	1.96	59.67
H-O-H-Sh	54	1.57	44.47	H-SR-Sh-H	23	1.80	61.47
H-W-SR-W	54	1.57	46.05	H-W-H	21	1.64	63.12
H-W-H-O	53	1.55	47.59	H-O-H-SR	20	1.57	64.68
H-W-Sh-W	51	1.49	49.08	H-O-H-O	19	1.49	66.17
H-Sh-Sh-H	50	1.46	50.54	H-H-Sh-H	18	1.41	67.58

Remark: H: home, W: work, Sc: school, Sh: shopping, PB: personal business, SR: social-recreation, O: others

4.4.1 Activity chain patterns on weekdays

As the number of canonical variates is large, a reduced number of variates is retained according to its statistical significance and their explanatory power. For weekdays, the correlation between activity chain types and sociodemographic variables is significant at 0.001 level based on Wilk's Λ test. The likelihood ratio tests for each canonical variate indicate that the associated correlations for the first 6 canonical variates are significantly different from 0 at 0.0001 level (Table 13). The proportion of variance explained by each variate is represented by the proportion of eigenvalues over all variates. Based on this criterion, three variates are sufficient (accounting for 90% of variance) to represent the variance of two variable sets. As a result, three canonical variates are selected for interpretation. The canonical correlations between two set of variables for three maximum correlated canonical variates are 0.60, 0.43 and 0.26, respectively. Similar to factor analysis, the interpretation of variates depends on the correlations (canonical loading) associated with two set of original variables (dependent and explanatory variables). The results of canonical loading for three maximum correlated variates are reported in Table 14. Similar to previous study (Colob, 1986), only absolute vaule of correlations greater than 0.20 are utilized for the interpretations, resulting in associating most of dependent variables to retained variates.

Table13 Canonical correlations between activity chain types and individual's socio-demographic variables and the proportion of variance explained by canonical variates (weekdays)

Canonical variate	Canonical correlation	Eigenvalue	Proportion	Cumulative	Test of H0: the current canonical correlation and all smaller ones are zero		
					Likelihood Ratio	Approximate F Value	Pr > F
1	0.60	0.56	0.59	0.59	0.44	11.36	<.0001
2	0.43	0.23	0.24	0.82	0.69	5.77	<.0001
3	0.26	0.07	0.08	0.90	0.85	2.96	<.0001
4	0.15	0.02	0.03	0.93	0.91	1.95	<.0001
5	0.15	0.02	0.02	0.95	0.93	1.72	<.0001
6	0.14	0.02	0.02	0.97	0.95	1.42	0.00

The interpretation of the associations between activity chain types and individual's sociodemographic variables can be drawn from the pair of variables with positive (negative) correlations retained at each side. For the first variate, activity chain types with positive/negative canonical loadings (in parenthesis) are H_W_H (0.37), H_W_H_SR (0.20) and H_Sc_H (-0.72), H_Sc_H_SR (-0.47). It indicates variate one is associated with school-based or work-based activity chain types. For the opposite explanatory variables, the canonical loadings for the variate one are: full-time job (0.58), age (0.48), being household head (0.46) (positive correlations), and student (-0.95), unemployment (-0.67), possession of public transportation reduction (-0.35) (negative correlations). This indicates, for positive correlation side, full-time worker and being household head conducted mainly two-link work chain or three-or-more-link work chain with after-work social-recreation activities. For negative correlation side, students with the possession of public transportation reduction make two-link school chain or three-or-more link school chain with after-school social-recreation activities.

For canonical variate two, the activity chain types with positive correlations (canonical loadings) are: H_W_H (0.39), H_Sc_H (0.32), H_W_H_SR (0.33), with negative correlations being H_Sh_H (-0.37), H_SR_H (-0.29), H_Sh_H_SR (-0.27), H_O_H (-0.25), H_O_H_Sh (-0.22) and others (-0.27). It indicates school- or work-based activity chain types are positively associated with variate two. The activity chain types involving shopping, social-recreation or other activities are negatively associated to this variate. For explanatory variable side, students (0.29) and full-time job (0.61) have higher positive correlations. However, age (-0.75), epouse/husband (-0.39) and unemployment (-0.60) have higher negative correlations. The pair of the above highly associated variables with positive correlations indicates the same interpretation as for variate one. For the pair of variables with negative correlations, it indicates young epouse/husband without work conducted mainly two-trip shopping, social-recreation or other type activity chain. Moreover, three-trip-or-more activity chain types involving the combination of shopping, social-recreation and other type of activities are also pursued. For variate three, dependent variables with positive correlations are H_O_H_O (0.36), H_O_W_H (0.59), H_O_H_Sh (0.24), H_W_H_O (0.29), H_W_Sh_W (0.22). For negative correlations, they are H_Sh_H (-0.27), H_SR_H (-0.23), H_W_Sh_H (-0.25). For positive correlation side, it indicates three-trip-or-more activity chain types with/without work activity involved. The negative correlation side shows mainly two-link non-work activity chain types. For explanatory variables, the number of children with age less than 12 years (0.87), the number of persons in the household (0.69) and paprt-time job (0.39) have positive correlations. Unemployment (-0.29) and low household income (-0.44) have negative correlations. The results indicate that, for the pair of variables with positive correlations, persons with part-time jobs living in the household with more young children and persons conducted three-trip-or-more activity chain types involving one or two “*other”* activity before to work or within home-based trip tour. Note that the regrouped “other” activity is composed of “drop off/pick up someone” (8.73% of total number of trips conducted by all persons on weekdays) and “other” activity (5.12% of total trips). For negative correlation side, it indicates persons without job living in low household income family make mainly two-trip activity chain with mainly shopping and social-recreation activities. Paradoxically, the activity chain type “home-work-shopping-home” is also associated with persons in unemployment situation for which the caution should be taken for its interpretation.

Table 14 Correlations between activity chain patterns, individual's sociodemographic variables and canonical variates

	Canonical variate 1	Canonical variate 2	Canonical variate 3
Dependent variable set (activity chain type)			
H_W_H	0.36	0.39	-0.19
H_Sc_H	-0.72	0.32	-0.04
H_W_H_SR_	0.20	0.33	-0.11
H_Sh_H	0.01	-0.37	-0.27
H_SR_H	-0.02	-0.29	-0.23
H_Sh_H_SR_	0.03	-0.27	-0.14
H_Sc_H_SR_	-0.47	0.18	-0.03
H_W_H_SH_	0.16	0.17	-0.13
H_O_H_O	0.04	-0.18	0.36
H_O_H	0.00	-0.25	-0.11
H_W_SR_H	0.13	0.15	-0.09
H_W_Sh_H	0.16	0.18	-0.25
H_O_W_H	0.11	0.06	0.59
H_O_H_Sh_	0.02	-0.22	0.24
H_W_SR_W_	0.15	0.18	-0.04
H_W_H_O_	0.10	0.12	0.29
H_W_Sh_W_	0.11	0.08	0.22
H_Sh_Sh_H	0.03	-0.18	-0.18
others	-0.12	-0.27	0.19
Canonical correlations between variable sets			
	0.60	0.43	0.26
Explanatory variable set			
gender	0.08	0.13	0.04
age	0.48	-0.75	-0.17
being_hhead	0.46	-0.07	0.01
license_car	-0.06	0.01	0.12
n_child_12	0.05	-0.02	0.87
n_household	-0.34	0.16	0.69
student	-0.95	0.29	-0.02
epouse	-0.01	-0.39	-0.09
full_time_job	0.58	0.61	0.02
part_time_job	0.12	-0.02	0.39
unemployment	-0.67	-0.60	-0.29
hhincom_low	0.08	-0.15	-0.44
hhincom_median	0.12	0.04	0.19
hhincom_high	0.11	0.14	0.19
PT_reduction	-0.35	-0.06	-0.09

4.4.2 Activity chain patterns on weekends

For weekend, the correlations between activity chain types and individual's sociodemographic characteristics are also affirmed at 0.0001 level based on Walk's Λ test. The likelihood ratio tests for each canonical variate indicate that only two variates are retained at 0.001 level. The cumulative proportion of explained variance of original variable sets based on the two variates is 43%. The associated canonical correlations for the two variate are 0.26 and 0.23, respectively. It indicates activity chain types on weekends are relative low correlated with individual's sociodemographic characteristics compared with that on weekdays.

The dependent structure for each variates of activity chain types and individual's sociodemographic variables are reported in Table 16. For positive correlation side, high correlated chain types are: H_Sh_H_Sh (0.31), H_O_H (0.23), H_O_H_SR (0.31), H_O_H_O (0.26). For negative correlation side, H_H_SR_H (-0.49), H_SR_H_SR (-0.40), H_H (-0.31), H_SR_SR_H (-0.21) are main associated chain types. For explanatory variables, age (0.64), being_hhead (0.30) and epouse (0.23) are positively correlated with the variate one; gender (-0.31), student (-0.55) and PT_reduction (-0.23) are negatively correlated. The pair of positive correlated variable sets indicate household head and epouse make other type activity, shopping or social-recreation activities. For negative correlation side, female students with public transportation reduction pursued in a three-or-more-link activity chain social-recreation activities. For the variate two, the positive correlated activity chain types are H_SR_H(0.41), H_SR(0.23), H_O_H(0.29), H_H_SR_H(0.23). The negative correlated chain types are: H_SR_SR_SR(-0.22), H_Sh_Sh_H(-0.24), H_W_H(-0.54) and H_H_Sh_H (-0.29). The explanatory variables with positive/negative correlations are license_car(0.25), student (0.52), unemployment(0.83) and being_hhead(-0.31), full_time_job(-0.76), hhincom_high (-0.35). The pair of associated variables with positive correlations indicates that students or individuals in unemployment with driving license conducted mainly two-link activity chain, composed of social-recreation or other type activities. The opposite pair of associated variable indicates full-time worker with household role and high household income pursued either two or three social-recreation or shopping activities or work activity on weekends.

Table 15 Canonical correlations between activity chain types and individual's socio-demographic variables and the proportion of variance explained by canonical variates (weekends)

Canonical variate	Canonical correlation	Eigenvalue	Proportion	Cumulative	Test of H0: the current canonical correlation and all smaller ones are zero		
					Likelihood Ratio	Approximate F Value	Pr > F
1	0.26	0.08	0.24	0.24	0.74	1.51	<.0001
2	0.23	0.06	0.19	0.43	0.80	1.31	<.0001

Table 16 Canonical loadings on canonical variates

	Canonical variate 1	Canonical variate 2
Dependent variable set (activity chain type)		
H_SR_H	0.01	0.41
H_Sh_H_SR	0.17	0.06
H_Sh_H	0.15	0.15
H_SR_H_SR	-0.40	0.07
H_SR_SR_H	-0.21	0.15
H_Sh_H_Sh	0.31	-0.15
H_SR	0.16	0.23
H_O_H	0.23	0.29
H_SR_SR_SR	-0.16	-0.22
H_H_SR_H	-0.49	0.23
H_H	-0.31	0.13
H_Sh_Sh_H	0.09	-0.24
H_Sh_H_O	0.14	-0.15
H_SR_Sh_H	-0.03	-0.06
H_W_H	-0.18	-0.54
H_O_H_SR	0.31	0.04
H_O_H_O	0.26	-0.03
H_H_Sh_H	-0.11	-0.29
others	0.05	-0.33
Canonical correlations between variable sets		
	0.26	0.23
Explanatory variable set		
gender	-0.31	-0.12
age	0.64	-0.05
being_hhead	0.30	-0.31
license_car	0.02	0.25
n_child_12	0.17	-0.14
n_household	0.08	0.17
student	-0.55	0.52
epouse	0.23	0.16
full_time_job	0.08	-0.76
part_time_job	-0.06	-0.11
unemployment	-0.04	0.83
hhincom_low	-0.03	-0.06
hhincom_median	0.02	0.11
hhincom_high	-0.15	-0.35
PT_reduction	-0.23	0.06

5. CONCLUSIONS

In this study, a uni-dimentional sequential alignment method is applied for analysing the interpersonal and intrapersonal variability of activity chain patterns. To investigate the time-use variability, an information radius based on Kullback-Leibler distance is applied to measure the difference of time-use distribution between individuals and between different types of days. The resulting similarity scores are then used as a basis for the classification analysis of activity chain pattern and time-use distributions. To investigate the influence of sociodemographic characterisitics on clustering membership, a discriminant analysis is applied. Moreover, the correlations between activity chain patterns on different types of days and sociodemographic characterisitics are examined based on canonical correlation analysis. The results provide useful insight on activity chain pattern and time-use distribution for different groups of persons and also the influence of related person and household characteristics.

For intrapersonal variability, the result indicates that individual's daily activity chain is more stable on weekdays. There is a distinguished difference between weekdays and weekends. We find also part-time workers have more variable activity pattern than full-time workers. Unemployment individuals have relatively low variability. Also the category of age within 35-44 years has the largest variability compared with other categories. For time-use distributions, the results indicate also larger variability on weekends than weekdays. Moreover, it provides time-use distributions for different groups of individuals based on similarity measure.

The canonical correlation analysis of activity patterns and individual's sociodemographic variables show that the two set of variables are more correlated on weekdays than on weekends. It provides also useful insight on the relationship between activity patterns and personal and household characteristics.

ACKNOWLEDGEMENTS

This research has benefited from a grant of ANR (the French Agency for Research), Project EuroCities-DATTA n° ANR-07-BLAN-0032-01

REFERENCES

1. Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfüser, G., Haupt, T. (2002). Observing the rhythms of daily life: a six-week travel diary, Transportation 29(2), 95–124.
2. Buliung, R.N., Roorda, M.J., Remmel, T.K. (2008). Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). Transportation 35(6), 697-722.

3. Castaigne, M., Cornelis, E., Toint, P., Descamps, J., Pauly, X, Walle, F. (2008). Activity report intended for the intermediate evaluation – Document G – Contract number SD/TM/03A BMW (Behaviour and mobility within the week).
4. Cox, T.F., Cox, M.A.A. (2000) Multidimensional Scaling, 2nd Edition, Chapman and Hall/CRC.
5. Golob, T.F. (1986) A nonlinear canonical correlation analysis of weekly trip chaining behaviour. Transportation research A 20(5), pp. 385-399.
6. Hanson, S., Hanson, P. (1981). The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics. Economic Geography 57, 332-347
7. Hanson, S., Huff, J.O. (1982). Assessing day-to-day variability in complex travel patterns, Transportation Research Record, 891, 18–24.
8. Hanson, S., Huff, J.O. (1986). Classification issues in the analysis of complex travel behaviour, Transportation 13(4), 273–291.
9. Hanson, S., Huff, J.O. (1988). Repetition and day-to-day variability in individual travel patterns: implication for classification. In: R. G. Golledge and H. J. P. Timmermans (Eds), Behavioural Modelling in Geography and Planning, 368–398 (London: Croom Helm).
10. Joh, C.-H., Arentze, T., Hofman, F., Timmermans, H. (2002). Activity pattern similarity: a multidimensional sequence alignment method. Transportation Research Part B 36, 385–403.
11. Joh, C.-H. (2004). Measuring and predicting adaptation in multidimensional activity-travel patterns. bouwstenen faculteit bouwkunde, 79, Dissertation an der Technischen Universität Eindhoven. Eindhoven.
12. Keuleers, B., Wets, G., Arentze, T., Timmermans, H. (2001). Association rules in identification of spatial-temporal patterns in multiday activity diary data. Transportation Research Record 1752, 32–37.
13. Kitamura, R., Yamamoto, T., Susilo, Y.O., Axhausen, K.W. (2006). How routine is a routine? An analysis of the day-to-day variability in prism vertex location. Transportation Research Part A 40, 259–279.
14. Kohonen, T. (2001). Self-Organizing Maps. 3rd ed. Springer.
15. Kuylen, A.A., Verhallen, T.M.M. (1981) The use of canonical analysis. Journal of Economic Psychology, 1, 217-237.
16. Miller, H.J. (2003). What about people in geographic information science? Computers, Environment and Urban Systems 27(5), 447-453.
17. Pas, E.I., Koppelman, F.S. (1986). An examination of the determinants of day-to-day variability in individuals' urban travel behavior. Transportation 13, 183–200.
18. Pas, E.I. (1988). Weekly travel-activity behavior. Transportation 15, 89–109.
19. Pendyala, R.M., Parashar, A., Muthyalagari, G. R. (2000). Measuring day-to-day variability in travel characteristics using GPS data. Paper presented at the 79th Annual Meeting of the Transportation Research Board, Washington, DC, USA.
20. Pendyala, R.M., Yamamoto, T., Kitamura, R., (2002). On the formation of time–space prisms to model constraints on personal activity-travel engagement. Transportation 29 (1), 73–94.
21. Khattree, R., Naik, D. (2000) Multivariate Data Reduction and Discrimination with SAS Software. SAS Institute Inc., Cary, NC, USA.
22. Schlich, R., Axhausen, K.W. (2003). Habitual travel behaviour: evidence from a six-week travel diary, Transportation 30(1), 13–36.
23. Schlich, R., Schönfelder, S., Hanson, S., Axhausen, K.W. (2004). Structures of Leisure Travel: Temporal and Spatial Variability. Transport Reviews 24(2), 219–237.
24. Schlich, R., Axhausen, K.W. (2004) Analysing interpersonal variability for homogeneous groups of travellers, *Arbeitsbericht Verkehrs- und Raumplanung*, 296, IVT. ETH Zürich, Zürich.
25. Schönfelder, S., Axhausen, K.W., (2001). Modelling the rhythms of travel using survival analysis. In: Kaspar, C., Laesser, C., Bieger, T. (Eds.), Jahrbuch 2000/2001 Schweizerische Verkehrswirtschaft, SVWG, St., 137–162.

26. Shimamoto, H., Zhang, J., Fujiwara, A. (2009). An effective algorithm to detect the similarities of activity-travel patterns. In: Proceedings of the 12th International Conference of the International Association for Travel Behaviour Research (IATBR). Jaipur, India.
27. Spissu, E., Pinjari, A., Bhat, C., Pendyala, R., Axhausen, K.W. (2009). An analysis of weekly out-of-home discretionary activity participation and time-use behavior. Transportation 36(5), 483-510.
28. Timmermans, H., Arentze, T., Joh, C.-H., (2002). Analyzing space–time behavior: new approaches to old problems. Progress in Human Geography 26 (2), 175–190.
29. Wilson, W.C. (1998) Activity pattern analysis by means of sequence alignment methods. Environment and Planning A, 30 (6) 1017-1038.
30. Wilson, C. (2008) Activity patterns in space and time: Calculating representative Hagerstrand trajectories, Transportation, 35 (4) 485–499.

Appendix 1 Day-to-day variability of daily travel time and trip characteristics

		Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Average daily travel time		75.3	74.1	74.8	79.1	77.8	81.5	80.9
Average number of trip per day		2.9	3.8	3.7	4.2	4.1	4.1	3.9
Average travel time per trip (by trip purpose)	home	23.7	18.8	19.4	18.0	18.6	18.7	17.5
	work	na	27.6	27.4	28.2	31.1	27.5	26.8
	school	na	20.8	20.8	22.2	23.3	23.0	na
	M. shopping	9.0	10.9	10.4	10.9	11.6	11.4	9.9
	shopping	16.8	17.9	24.1	19.8	23.4	23.8	28.4
	personal business	10.1	14.5	14.4	13.8	13.1	11.4	12.4
	social-recreation	33.4	24.5	23.8	22.2	20.5	25.8	29.9
Average travel time per trip (by mode)	walk	26.8	18.1	15.9	15.3	14.9	18.0	24.8
	bicycle	25.6	13.7	13.3	13.5	13.2	13.8	16.5
	moto	36.8	13.6	11.2	13.4	13.3	11.0	12.7
	car	23.2	17.2	18.3	17.0	16.7	17.7	18.5
	public transport	50.4	48.5	48.3	48.0	50.2	46.4	40.8
Average travel distance per trip (by trip purpose)	home	17.6	10.9	10.2	9.0	9.8	10.7	11.6
	work	n.a.	19.5	17.8	18.9	18.5	18.6	23.6
	school	n.a.	6.3	10.3	8.1	7.4	9.4	n.a.
	M. shopping	2.3	3.9	4.2	4.0	4.3	4.0	3.5
	shopping	17.6	9.6	8.2	7.8	5.6	10.2	9.3
	personal business	5.2	6.3	6.3	4.9	4.5	4.6	4.1
	social-recreation	15.7	9.0	10.0	9.1	8.0	11.9	15.2
Average travel distance per trip (by mode)	walk	3.8	2.0	4.1	2.1	1.4	2.0	2.2
	bicycle	6.8	3.6	3.3	3.4	3.7	3.6	4.5
	moto	25.1	8.6	5.2	7.3	7.0	5.9	8.7
	car	20.5	11.8	12.6	11.3	11.4	13.2	14.6
	public transport	27.2	29.0	26.4	27.1	29.1	25.5	21.8

Remark: M. shopping means maintenance shopping and n.a. means not applicable

Appendix 2 Travel characteristics for clusters based on Levenshtein distance for different types of day

	Weekday								Weekend							
	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distance travelled per day (Km)	25.8	40.3	44.9	54.4	46.6	51.3	51.5	55.1	28.2	44.2	49.5	63.3	46.1	62.8	61.4	72.8
Duration travelled per day	62.7	58.2	88.4	76.9	79.7	62.1	105.6	69.6	59.0	70.5	91.8	90.1	75.1	85.3	123.0	102.8
Number of trips per day	2.9	1.6	4.8	2.2	3.2	1.5	7.2	3.2	2.5	1.7	4.4	2.5	2.8	1.8	5.4	3.3
Number of returning home trips per day	1.2	0.7	1.7	1.0	1.3	0.6	2.3	1.4	1.0	0.8	1.6	1.1	1.2	0.8	1.8	1.4
Number of work trips per day	0.1	0.3	0.6	0.8	1.0	0.5	0.7	1.0	0.0	0.2	0.1	0.3	0.1	0.3	0.1	0.3
Number of school trips per day	0.3	0.5	0.2	0.5	0.0	0.1	0.2	0.6	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.2
Number of shopping trips per day	0.4	0.7	0.6	0.8	0.3	0.5	0.9	1.1	0.3	0.7	0.7	1.0	0.4	0.6	0.8	1.1
Number of personal business trips per day	0.1	0.4	0.2	0.5	0.1	0.3	0.4	0.7	0.1	0.2	0.1	0.3	0.1	0.3	0.1	0.4
Number of social-recreation trips per day	0.5	0.7	0.7	0.9	0.3	0.6	1.0	1.3	0.8	0.9	1.3	1.2	0.9	1.0	1.6	1.5
Number of other trips per day	0.3	0.6	0.7	1.1	0.2	0.6	1.6	1.7	0.2	0.5	0.5	0.9	0.2	0.5	1.0	1.4