

Chapter #: Sequence Analysis of Time Use Patterns

Authors

Giacomo Vagni

g.vagni@essex.ac.uk, Department of Sociology,
University of Essex, Colchester CO4 3SQ, UK

Benjamin Cornwell

btc49@cornell.edu, Department of Sociology,
Cornell University, Ithaca, NY 14853

Miriam Siglreitmaier

miriam.siglreitmaier.2@hu-berlin.de
Institute of Social Sciences, Humboldt-Universität, 10099 Berlin, Germany

Keywords: Quantitative Methods, Sequence Analysis, Clustering

Abstract: Limiting time use research to the average amount of time spent doing certain activities hides the heterogeneity in the use of time. Sequence analysis goes beyond this limitation to analyze how much time, but the order and timing in which actions unfold. This type of analysis provides greater insight into the social routines, priorities, and norms of everyday behaviours. Two main categories of statistical techniques are used to examine aspects of sequences: those comparing pairs of sequences to identify patterns and decompose sequence variations; and those analyzing individual sequences to understand their internal structure. This chapter primarily discusses methods for uncovering patterns, with a brief overview of techniques for analyzing individual sequences. The presentation is illustrated with a simulated time use dataset covering 20 one-hour episodes that include six primary activities: paid and unpaid work, eating, leisure, personal care, and others.

Introduction

The way people use their time in everyday settings is both a cause and a consequence of various social phenomena, from patterns of social stratification to children's development (Fiorini and Keane 2014; Vagni 2020). Aggregate social phenomena are ultimately rooted in people's daily actions, whether conscious or unconscious, constrained or free (e.g. Gershuny, 2000). As such, time use is a fundamental link between macro-sociological forces and micro-sociological behaviour. Time-use research has significantly contributed to our understanding of the domestic division of unpaid work (Sullivan 2000), working hours (Robinson and Godbey 2010), sleep (Lamote de Grignon Pérez et al. 2019) and leisure time (Aguilar and Hurst 2007), among other important social phenomena.

However, time-use research has generally been limited to the aggregate analysis of the average time individuals spend doing certain activities. While the average time spent doing different activities is an important part of the study of daily life, averages hide the heterogeneity in the use of time. Sequence analysis moves us beyond an analysis of aggregate or overall time spent on specific activities to an analysis of patterns and structure of time use—that is, not just how much time is spent on certain things, but the *order* and the *timing* in which social actions unfold (Abbott, 1995). This type of analysis provides greater insight into social routines, priorities, and norms of everyday behaviours (e.g. Giddens, 1986; Parsons, 1951; Zerubavel, 1985).

Time as sequence

Understanding the use of time as a *sequence* of activities is relatively straightforward. This is how time unfolds in daily life. One activity is followed by another, which in turn is followed by another, and so on. Likewise, sequence analysis deals with states or events unfolding over time (Cornwell, 2015). With regard to time use, the *states* are generally specific activities, such as sleeping, working, cleaning, or watching TV. One can also layer on this other time use information, such as *with whom* the person shared an activity, where the activity occurred, or additional contextual information. The timeframe of a time diary data usually is 24 hours. However, one can also find diary data collected for two days for the same individuals (generally one weekday and one weekend). Less common are weekly or monthly time-use diaries (Sorokin and Berger 1939).

The units of analysis for 24-hour diaries typically vary from one-minute time intervals to 30-minute time intervals. Five- and ten-minute intervals are common units for time diaries. Concretely, this means that respondents report their activities every few minutes. For 10-

minute interval diaries, for example, the sequence of activities is made of 144 states ($144 \times 10 = 1440$ minutes or 24 hours). With a small loss of information, diaries with different intervals can be harmonized into a common time interval (Gershuny 2000).

Sequence analysis has generally been employed as a descriptive technique aimed at analyzing and depicting how events unfold over time in a holistic way. This requires understating what whole sequences of events are composed of, rather than just studying, for example, in the aggregate, the amount of time spent on given types of activities (Billari 2001). Time-use diaries fit this epistemological framework because a day is an indivisible sociological unit but can still be broken down into a series of actions (Sorokin and Merton 1937). Days are thus understood in their entirety, as a series of stochastically non-independent units of action. What a given person does at 9 am, for example, is understood to be related, in some way, to what they did at 8 am. Sequence analysis does not assume that this association is causal. One may employ a relational framework that suggests that what happens at a one-time point directly affects what happens at the next (e.g. Butts, 2008) or a different model that assumes that the two actions are related due to endogeneity or selection. Regardless, it makes sense to take the day as the whole unit of analysis.

Sequence analysis: aims and methods

Sequence analysis involves various statistical techniques used to examine the various aspects and components of sequences. These methods can be divided into two main categories: those that compare pairs of sequences to identify patterns and decompose sequence variations; and those that analyze individual sequences to understand their internal structure. This chapter will primarily discuss methods for uncovering patterns, with a brief overview of techniques for analyzing individual sequences.

Individual sequence measures

A sequence in the context of time use is an ordered list of daily activities. This list can be represented in a variety of formats, such as the State Sequences format (STS) and State Permanence Sequence format (SPS). These two formats are the most commonly used to express sequence and are displayed in Table #.1.

Table #.1: Common sequence formats

	State sequences (STS)	State permanence sequence (SPS)
x1	S-S-S-D-D-W-W-W-L-L	(S,3)-(D,2)-(W,3)-(L,2)
x2	S-S-S-D-D-L-L-D-D-S	(S,3)-(D,2)-(L,2)-(D,2)-(S,1)
x3	S-S-W-W-W-W-W-L-L	(S,2)-(W,6)-(L,2)

The State Sequence (STS) format is the most basic way of representing sequences. It displays each state for each time interval, as seen in Table #.1. The example given is S-S-S-D-D-W-W-W-L-L, which stands for Sleep (S), Domestic work (D), Paid work (W) and Leisure (L). This sequence indicates that this person experienced three episodes of sleep, followed by two episodes of domestic work, three episodes of paid work, and finally, two episodes of leisure. The SPS format compresses the STS by indicating its state and duration in parentheses. From the individual sequences, several indicators can be derived (Gabadinho et al. 2011; Ritschard 2021), such as the length of the sequence, the number of spells, the mean spell, the number of states (activity) visited, the number of transitions, the entropy and the complexity index. These indicators provide valuable insight into each sequence of activities. Table #.2 shows these indicators for the three sequences in Table #.1.

Table #.2: Individual sequence indicators

	Length	Visited	Visit %	Transition	Nb Spell	Mean Spell	Entropy	Complexity
x1	10	4	1	3	4	2.5	0.98	0.57
x2	10	3	0.75	4	5	2	0.76	0.58
x3	10	3	0.75	2	3	3.33	0.68	0.39

All three sequences have a length of 10. Sequence 1 visited all four possible states (100% of the possible states), reflecting a greater diversity of experience compared to the other two sequences. Sequence 1 shared more activities in common with the other two sequences than they did with each other. Sequence 3 only experienced two transitions, from sleep to work (S>W) and from work to leisure (W>L), while sequence 2 experienced the most transitions (4). The spell duration for the sequence 3 is higher, spending on average 3.3 episodes in each state or one-third of the time in each episode. ‘Spell’ refers to the consecutive time spent in a state. This contrasts with Sequence 2, which spent less time in each state and experienced more

transitions. This difference is captured by longitudinal entropy and complexity measures, which both reflect the uncertainty and diversity of individual sequences. Entropy is 0 when a person spends time in only one activity (e.g. S-S-S-S) and is maximal (1 for normalized entropy) when a person spends the same time in every possible activity (e.g. S-W-L-D). Entropy is highest for Sequence 1, as they visited all states and spent about the same time in each of them (the mean spell is 2.5 over 10, which is 25% for four activities).

The complexity index builds on the entropy measures but also considers the number of transitions. We see that Sequence 1 and Sequence 2 appear more complex than Sequence 3, which confirms our intuition. Other measures of individual sequences have been proposed that may be of interest to time-use researchers, such as the Turbulence index (Elzinga and Liefbroer 2007) and measures of Precarity, Insecurity, and more (Ritschard, 2021).

Patterns of Time use

Uncovering hidden patterns of activities is a central aim of sequence analysis and is generally done in the following way. An algorithm searches for dissimilarity (or similarity) between sequences of activities. This algorithm returns a distance score between each pair of sequences. This is called a ‘pairwise distance matrix.’ A clustering algorithm can then be applied to this matrix, with the aim of grouping similar sequences to identify homogenous clusters of actors who engage in similar activity sequences. Empirically, this results in a set of typologies, which can then be treated as a multinomial variable representing different ‘types of days’. For instance, one cluster might summarize a 9-to-5 working day, while another might summarize a night-time work shift (Lesnard 2008). Various statistical analyses are then performed using this categorical typological variable (Vagni 2020).

Measuring dissimilarity between time sequences

Calculating the degree of similarity or dissimilarity between time-use sequences is a fundamental task of sequence analysis. Two sequences with the same activities in the same exact order have a distance of 0, meaning they are equivalent. However, the degree of dissimilarity between sequences can be attributed to a wide range of factors, such as no common activities, varying orders of activities, differences in spells (i.e. time spent successively in a state), discrepancies in timing, etc. Different sequence analysis algorithms can be used to quantify which of these dimensions should be emphasized. Over the years, a great variety of dissimilarity measures have been proposed to capture these different aspects (Studer and Ritschard 2016). The family of distances can be divided into three broad groups:

those based on the number of shared attributes, those based on the distribution of states, and the Optimal Matching ‘edit’ distance.

State Distribution and Common Attributes

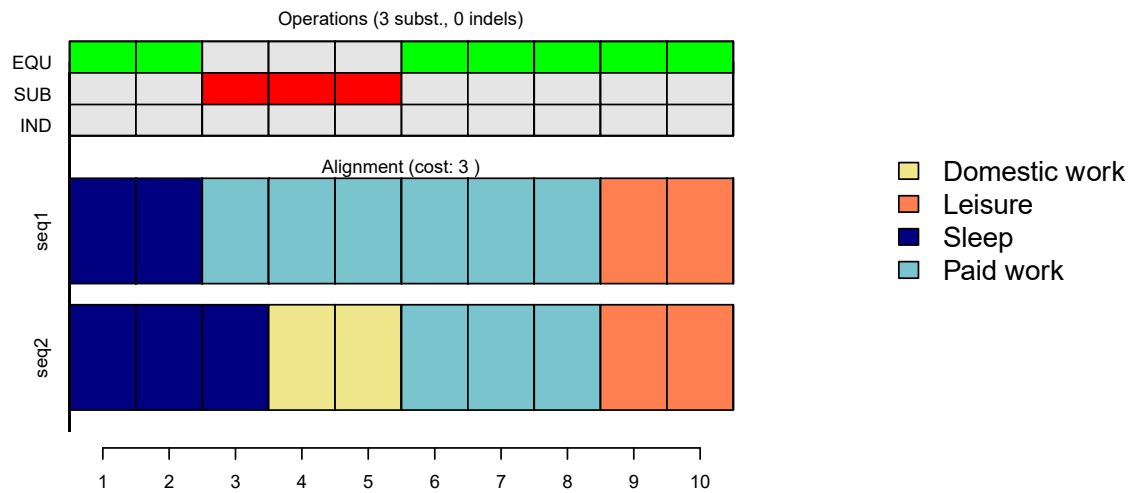
The Hamming Distance is a simple measure of dissimilarity between two sequences that takes into account the mismatched states (activities) at each time point of the day (Lesnard 2010). This distance metric is suitable for time diaries as they are of equal lengths, and it is sensitive to the timing of activities. Other basic distances, such as the Chi-square distance and the Euclidean distance (based on common time spent in different activities) capture well differences in duration (Studer and Ritschard 2016). If ordering is a key factor, then distances based on subsequences, such as the Longest Common Subsequence or the Number of Matching Subsequences (Elzinga 2003), are more appropriate. Variants of these distances have been developed to include sensitivity to spells and durations (Elzinga and Studer 2015).

Optimal Matching

The Optimal Matching approach is the most widely used distance algorithm for computing pairwise dissimilarities (Abbott 1995; Abbott and Forrest 1986; Abbott and Tsay 2000). It computes the minimal ‘cost’ required for transforming a sequence into another one (i.e. the distance, dissimilarity, or difference between sequences). Two main operations are used to compute the ‘distance’ between sequences: insertion/deletion (indel) and substitution. Indel refers to adding or subtracting a sequence state, while substitution refers to replacing one state with another at a certain cost. Optimal Matching allows researchers to set up their respective cost parameters based on theoretical principles.

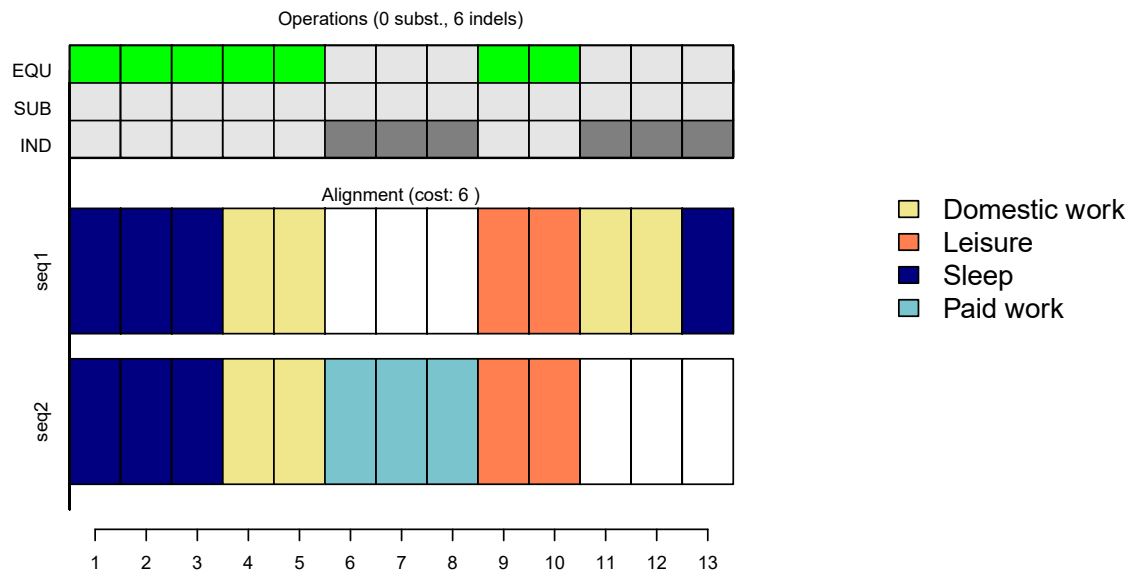
Figure #.1 illustrates a simple Optimal Matching procedure comparing sequence 1 and sequence 3 discussed previously (S,3)-(D,2)-(W,3)-(L,2) and (S,2)-(W,6)-(L,2). In this case, the optimal solution was to substitute the states in positions 3, 4, and 5 to transform one sequence into the other. Because we set the substitution cost to 1, the total alignment cost is simply 3.

Figure #.1: Optimal Matching with substitution cost set to 1. Total optimal alignment cost of 3.



Substitution costs are generally set to twice the cost of indel operations. Figure #.2 shows the alignment of two sequences: (S,3)-(D,2)-(W,3)-(L,2) and (S,3)-(D,2)-(L,2)-(D,2)-(S,1), with substitution cost set to 2 and indel to 1. The optimal alignment here used indel in positions 3, 4, 5, and in positions 11, 12, 13, which sums up to an alignment cost of 6.

Figure #.2: Optimal Matching with substitution cost set to 2 and Indel to 1. Total optimal alignment cost of 6.



In this example, substituting an activity with another, and deleting/inserting an activity, cost the same. Substituting work for sleep costs the same as substituting work for leisure. However, theoretically, this might not always be justified. Generally, researchers will use costs based on theories and assumptions about what the algorithm should consider as ‘close’ or ‘distant’. Therefore, two researchers might find different results based on how they set up costs. Thus, researchers must be clear and open about the costs they have chosen.

Strategies for setting costs

A simple yet powerful way to set up costs based on theoretical principles is to use a Gower distance to calculate substitution costs based on *attributes* (Studer and Ritschard 2016). When applied to time use, researchers can use several important dimensions, such as work, leisure, and biological necessity, to score different activities. This method ensures the triangle inequality. A strategy based on the Chi-square distance can be used to take into account future states that activities have in common (Rousset et al. 2012). For example, if two states have a ‘common future’ of commuting hours later, this could indicate proximity between them. Additionally, strategies have been proposed for indel operations (deleting or inserting a state). Generally, deleting or inserting rare states or activities should be an expensive operation. For instance, the inverse of a state proportion can be set as the indel cost (Studer and Ritschard 2016). If sleep constitutes one-third of daily time and care constitutes 20% of daily time, then

their respective indel cost will be 3 and 5. Their substitution costs can then be set as their sum, so substituting sleep for care would have a cost of 8. Other strategies for contextualizing indel and substitution costs have been proposed, such as ‘localized OM’ (Halpin 2010), OM sensitive to spell lengths, and others (Studer and Ritschard 2016).

One popular method designed with time-use diaries in mind is the so-called ‘Dynamic Hamming’ (Lesnard 2010). This distance only involves substitution costs that are time-varying. Indel operations can be problematic as they distort one of the most fundamental aspects of time use: synchronicity. The general idea of Dynamic Hamming is to take into account the context of the timing of the activities by updating the substitution cost depending on how many people are transitioning to different activities at different moments of time. The strength of the transitions between activities reveals something about the connection between them. The transition rate between two states, for example working (W) and non-working (O) at a particular moment, reveals whether they are part of a similar underlying trajectory or not. The higher the transition between two states, the lower the substitution cost, and inversely the lower the transition, the higher the cost. The Dynamic Hamming method uses transition matrices that vary according to different moments of the day. Consequently, it allows for different costs to be assigned to the same activity done at different times, emphasizing that timing is key for the social meaning of an activity (Lesnard, 2010, pp. 400-401).

Clustering

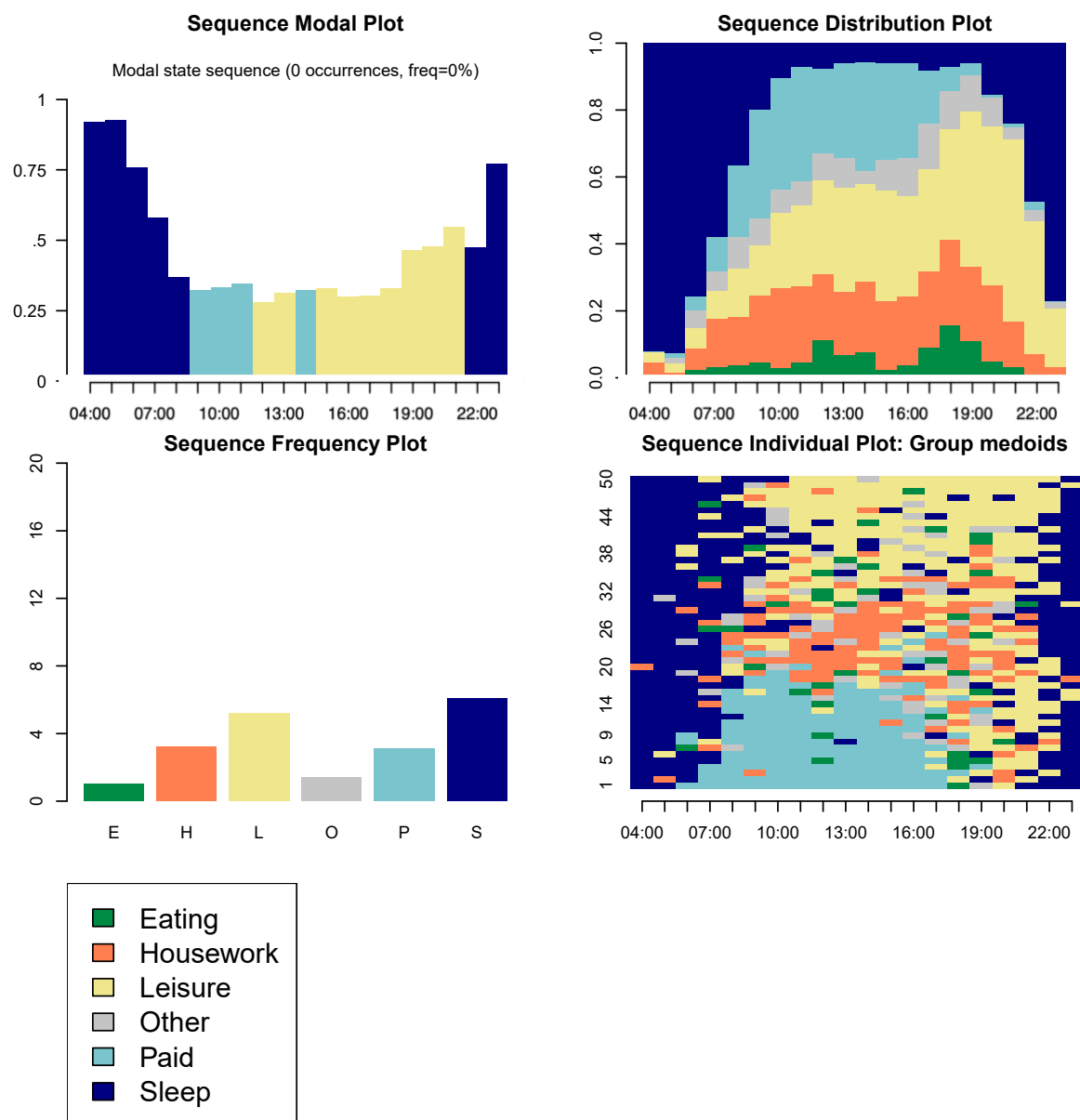
Once the dissimilarity matrix has been calculated, researchers have several options for analyzing the pairwise dissimilarity matrix, such as discrepancy analysis, multidimensional scaling, regression trees, etc. (Gabadinho et al. 2011; Studer et al. 2011). The most popular way of analyzing pairwise distances is by far clustering (Kaufman and Rousseeuw 1990). Clustering is a type of ‘unsupervised learning’ using algorithms aimed at grouping data into homogenous clusters. It is used to identify underlying structures and patterns in the data. A common method is Ward’s criterion, a type of agglomerative hierarchical clustering that minimizes the within-cluster sum of squares (Tan et al. 2005). Another common method is the Partition around Medoids (PAM) (Studer 2013).

Empirical Illustration

We illustrate our presentation with a simulated time use dataset that covers a period of 20 hours, or 20 episodes of one-hour episodes, starting at 4 am. This dataset includes six primary activities: Paid Work, Unpaid Work (domestic work and childcare), Eating, Leisure, Personal Care (sleeping, grooming) and Other (travel, etc.).

We start by presenting the most common ways of visualizing sequences in Figure #.3.

Figure #.3: Sequence Modal Plot, Sequence Distribution Plot, Sequence Frequency Plot and Sequence Individual Plot



In the top left panel, we have the *Sequence Modal Plot*, which simply displays the modal (most common) activity for each time interval. The modal sequence starts with personal care (sleeping, etc.), then paid work, followed by leisure and finally personal care again. However, it is important to note that this modal sequence only represents a portion of the most common activities in the data, as the modal activities of paid work and leisure cover only around 30-50% of people engaged in these activities at each time point.

The *Sequence Distribution Plot* in the top right corner shows the proportion of people engaged in different activities at each time interval from 4:00 am to midnight the next day. At 4:00 am, for example, the majority of people are sleeping (96%). This chart gives us a general overview of the activity distribution throughout the day.

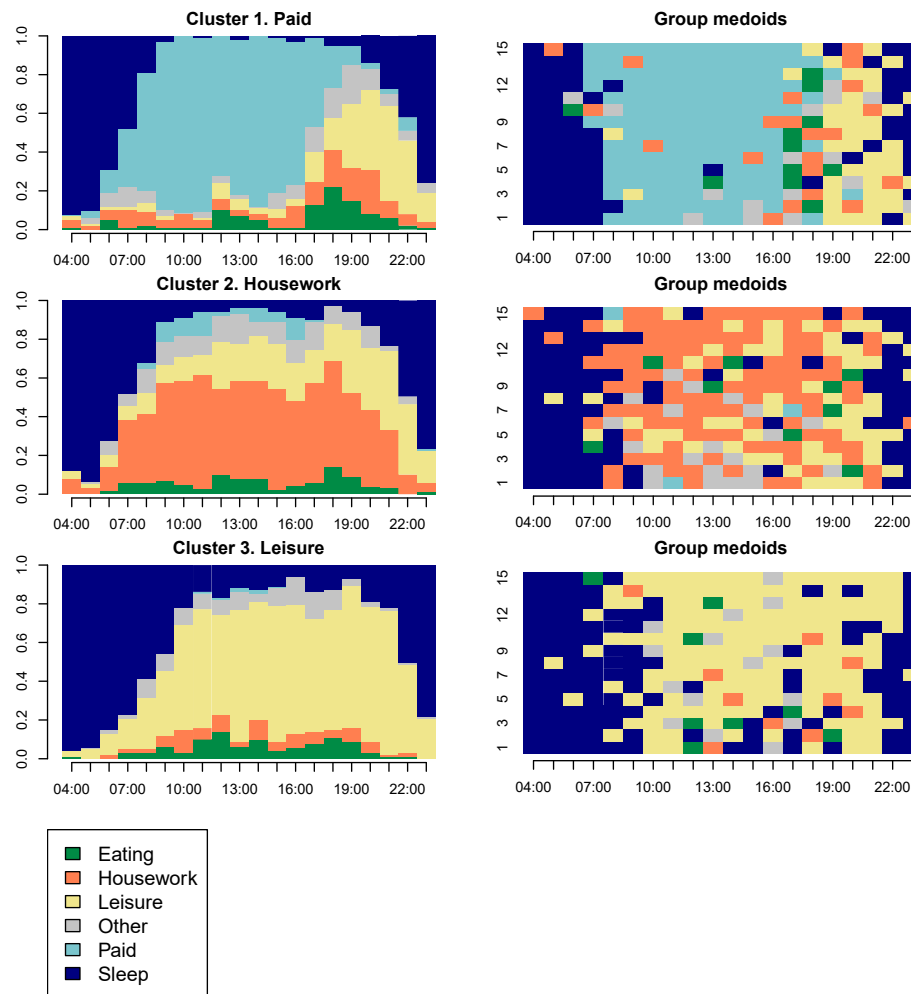
On the bottom left panel, a plot shows the average number of episodes spent in different activities. It is a simple *Sequence Frequency Plot*. On the bottom right panel, we have the *Individual Sequence Plot*. This plot shows how people's lives in the data unfolded throughout the day. This plot does not aggregate sequences but instead shows the actual observed sequences. Along the y-axis, we have the identifier of the person's sequence. By looking at this plot, we can see how a person's day panned out, such as the sequence at the top of the plot for someone who got up around 10:00 am and enjoyed leisure for the rest of the day, or the sequence at the bottom of the plot for someone who woke up early and worked until late.

Patterns of time use

To uncover underlying patterns of time use in the data, we use Optimal Matching with indel operations set to the log inverse of state frequencies (*indelslog*). We then use the clustering algorithm, *Partition around Medoids* (PAM), to partition the sequences into three main clusters.

Figure #.4 shows the sequence distribution plot and the sequence individual plot for the three clusters. We extracted 15 'medoids' for each cluster, which are the most representative sequences in each group, using the *Relative Frequency Sequence Plots* procedure (Fasang and Liao 2014). Looking at both plots enables us to understand better the main pattern and the heterogeneity present within each cluster.

Figure #.4: Sequence Distribution Plot and 15 most representative sequences for each cluster

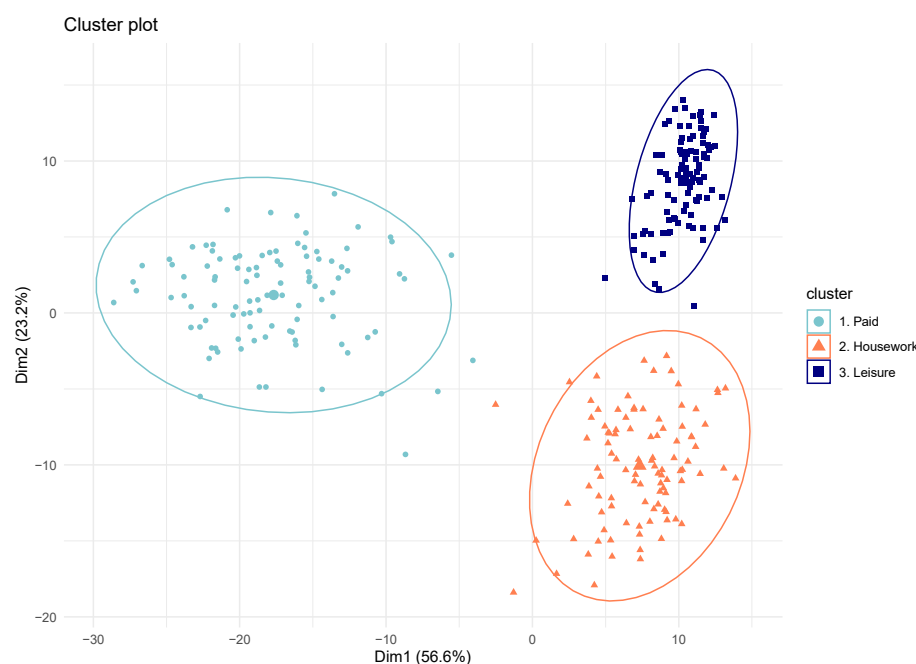


The three clusters reflect patterns of paid work, domestic work and leisure. The first cluster of sequences reflects a typical 8-to-5 paid workday. The medoids (the most central sequences) in this cluster show a consistent pattern of paid work with various evening activities such as domestic work or leisure. The second cluster displays a pattern of domestic work, with the medoids being fuzzier, suggesting more variations in this cluster. To explore sequence variance and variations more formally, discrepancy methods can be used (Studer et al., 2011). The third cluster reflects a pattern of leisure activities. We can see that in this leisure cluster, people are waking up later, around 10:00 am, compared to the two other clusters.

While clustering is useful for understanding and identifying the main patterns in the data, it can hide important within-group variations. A technique complementary to clustering that is often used is Multidimensional Scaling (MDS). In the context of sequence analysis, MDS

transforms the dissimilarity matrix into a dimensional space (generally of two or three dimensions), preserving the distances between sequences as much as possible (Kruskal and Wish 1978). In Figure #.5, we visualize the first two dimensions of this dimensional reduction. We see that the clusters are relatively well separated and are placed along substantially interpretable axes. The first dimension opposes Paid Work to Leisure and Domestic work, while the second dimension captures the opposition between Leisure and Domestic work.

Figure #.5: Principal Component Analysis of the dissimilarity matrix with ellipse around the clusters' centre



Conclusions

Sequence analysis is a powerful technique for understanding the underlying structure of time-use data. These data are highly complex, given the large number of time points and activities that make up an individual's record. Sequence analysis is an effective method of data reduction, enabling researchers to uncover patterns in complex longitudinal data. Optimal Matching (Abbott and Tsay 2000) and non-alignment techniques (Brzinsky-Fay and Kohler 2010; Elzinga 2003) are two useful applications of sequence analysis that can be applied to time use data to discover hidden patterns. As demonstrated with the empirical example, these approaches can effectively uncover commonalities across seemingly complex and detailed sequences of social action. Through the use of sequence analysis, researchers can gain a more comprehensive understanding of time use data and the structure of daily life.

Furthermore, sequence analysis is increasingly more easily utilized with user-friendly programs, such as Stata and R. Sequence analysis packages within these programs, such as SQ (Kohler et al. 2020) or SADI (Halpin 2017) in Stata, and TraMinerR (Gabadinho et al., 2011) in R, make it relatively easy for analysts to uncover general patterns in time use data, such as typical sleeping/working/eating patterns. These patterns can then be used as dependent or independent variables in other types of analyses, such as regression analysis.

There is a vibrant future for the application of sequence analysis approaches to time-use data (Vagni and Cornwell 2018). Not only are sequence-analytic techniques becoming more widely used, but time-use data are becoming more widely available. This includes both subjectively reported time-diary-type data and passively collected electronic data, such as time-stamped activity and locational data from smartphones. Sequence analysts continue to develop and refine algorithms for comparing everyday activity sequences, and thus enhance our ability to identify patterns of social action. These improvements in data collection, coupled with the greater availability of computer programs for analyzing such data, promise major advances in our understanding of the socially structured nature of everyday life.

References

- Abbott, A. (1995), 'Sequence Analysis: New Methods for Old Ideas', *Annual Review of Sociology*, **21**, 93–113.
- Abbott, A. and J. Forrest (1986), 'Optimal matching methods for historical sequences', *The Journal of Interdisciplinary History*, **16** (3), 471–94.
- Abbott, A. and A. Tsay (2000), 'Sequence analysis and optimal matching methods in sociology: Review and prospect', *Sociological Methods & Research*, **29** (1), 3–33.
- Aguiar, M. and E. Hurst (2007), 'Measuring Trends in Leisure: The Allocation of Time Over Five Decades', *The Quarterly Journal of Economics*, **122** (3), 969–1006.
- Billari, F. C. (2001), 'Sequence analysis in demographic research', *Canadian Studies in Population*, 439–58.
- Brzinsky-Fay, C. and U. Kohler (2010), 'New developments in sequence analysis'.
- Elzinga, C. H. (2003), 'Sequence Similarity: A Nonaligning Technique', *Sociological Methods & Research*, **32** (1), 3–29.
- Elzinga, C. H. and A. C. Liefbroer (2007), 'De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis', *European Journal of Population / Revue Européenne de Démographie*, **23** (3), 225–50.
- Elzinga, C. H. and M. Studer (2015), 'Spell sequences, state proximities, and distance metrics', *Sociological Methods & Research*, **44** (1), 3–47.
- Fasang, A. E. and T. F. Liao (2014), 'Visualizing sequences in the social sciences: Relative frequency sequence plots', *Sociological Methods & Research*, **43** (4), 643–76.
- Fiorini, M. and M. P. Keane (2014), 'How the Allocation of Children's Time Affects Cognitive and Noncognitive Development', *Journal of Labor Economics*, **32** (4), 787–836.
- Gabadinho, A., G. Ritschard, N. S. Mueller and M. Studer (2011), 'Analyzing and Visualizing State Sequences in R with TraMineR', *Journal of Statistical Software*, **40** (4), 1–37.
- Gershuny, J. (2000), *Changing Times: Work and Leisure in Postindustrial Society*, Oxford University Press.
- Giddens, A. (1986), *The Constitution of Society: Outline of the Theory of Structuration*, New Ed edition, Cambridge Cambridgeshire: Polity Press.
- Halpin, B. (2010), 'Optimal matching analysis and life-course data: The importance of duration', *Sociological Methods & Research*, **38** (3), 365–88.
- Halpin, B. (2017), 'SADI: Sequence analysis tools for Stata', *The Stata Journal*, **17** (3), 546–72.
- Kaufman, L. and P. Rousseeuw (1990), *Finding Groups in Data. An Introduction to Cluster Analysis.*, New York: Wiley.

- Kohler, U., M. Luniak and C. Brzinsky-Fay (2020), *SQ: Stata Module for Sequence Analysis*.
- Kruskal, J. B. and M. Wish (1978), *Multidimensional Scaling*, vol. 11, Sage.
- Lamote de Grignon Pérez, J., J. Gershuny, R. Foster and M. De Vos (2019), ‘Sleep differences in the UK between 1974 and 2015: Insights from detailed time diaries’, *Journal of Sleep Research*, **28** (1), e12753.
- Lesnard, L. (2008), ‘Off-Scheduling within Dual-Earner Couples: An Unequal and Negative Externality for Family Time’, *American Journal of Sociology*, **114** (2), 447–90.
- Lesnard, L. (2010), ‘Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns’, *Sociological Methods & Research*, **38** (3), 389–419.
- Parsons, T. S. (1951), *The Social System*, New York: Free Press.
- Ritschard, G. (2021), ‘Measuring the nature of individual sequences’, *Sociological Methods & Research*, 00491241211036156.
- Robinson, J. and G. Godbey (2010), *Time for Life: The Surprising Ways Americans Use Their Time*, University Park: Penn State Press.
- Rousset, P., J.-F. Giret and Y. Grelet (2012), ‘Typologies de parcours et dynamique longitudinale’, *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, **114** (1), 5–34.
- Sorokin, P. A. and C. Q. Berger (1939), *Time-Budgets of Human Behavior.*, Cambridge: Harvard University Press.
- Sorokin, P. A. and R. K. Merton (1937), ‘Social Time: A Methodological and Functional Analysis’, *American Journal of Sociology*, **42** (5), 615–29.
- Studer, M. (2013), *WeightedCluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R*.
- Studer, M. and G. Ritschard (2016), ‘What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures’, *Journal of the Royal Statistical Society: Series A*, **179** (2), 481–511.
- Studer, M., G. Ritschard, A. Gabadinho and N. S. Müller (2011), ‘Discrepancy analysis of state sequences’, *Sociological Methods & Research*, **40** (3), 471–510.
- Sullivan, O. (2000), ‘The Division of Domestic Labour: Twenty Years of Change?’, *Sociology*, **34** (3), 437–56.
- Tan, P.-N., M. Steinbach and V. Kumar (2005), *Introduction to Data Mining*, Pearson College Div.
- Vagni, G. (2020), ‘The social stratification of time use patterns’, *The British Journal of Sociology*, **71** (4), 658–79.

Vagni, G. and B. Cornwell (2018), 'Patterns of everyday activities across social contexts', *Proceedings of the National Academy of Sciences*, **115** (24), 6183–8.

Zerubavel, E. (1985), *Hidden Rhythms: Schedules and Calendars in Social Life*, University of California Press.

