electronic ***International Journal of Time Use Research***
2009, Vol. 6, No. 1, 1-25.

Exploring time diaries using semi-automated activity pattern extraction

Katerina Vrotsou, Kajsa Ellegård and Matthew Cooper

Katerina Vrotsou
Dept. of Science and Technology
Linköping University, Campus Norrköping
SE-601 74 Norrköping, Sweden
e-mail: katerina.vrotsou@liu.se

Kajsa Ellegård
Dept. of Technology and Social Change
Linköping University
SE-5 81 83 Linköping, Sweden
e-mail: kajsa.ellegard@liu.se

Matthew Cooper
Dept. of Science and Technology
Linköping University, Campus Norrköping
SE-601 74 Norrköping, Sweden
e-mail: matt.cooper@liu.se

Abstract

Identifying patterns of activities in time diaries in order to understand the variety of daily life in terms of combinations of activities performed by individuals in different groups is of interest in time use research. So far, activity patterns have mostly been identified by visually inspecting representations of activity data or by using sequence comparison methods, such as sequence alignment, in order to cluster similar data and then extract representative patterns from these clusters. Both these methods are sensitive to data size, pure visual methods become too cluttered and sequence comparison methods become too time consuming. Furthermore, the patterns identified by both methods represent mostly general trends of activity in a population, while detail and unexpected features hidden in the data are often never revealed. We have implemented an algorithm that searches the time diaries and automatically extracts all activity patterns meeting user-defined criteria of what constitutes a valid pattern of interest for the user's research question. Amongst the many criteria which can be applied are a time window containing the pattern, minimum and maximum occurrences of the pattern, and number of people that perform it. The extracted activity patterns can then be interactively filtered, visualized and analyzed to reveal interesting insights. Exploration of the results of each pattern search may result in new hypotheses which can be subsequently explored by altering the search criteria. To demonstrate the value of the presented approach we consider and discuss sequential activity patterns at a population level, from a single day perspective.

JEL-Codes: C69, D13, R29

Keywords: Time-geography, diaries, everyday life, activity patterns, visualization, data mining, sequential pattern mining

A preliminary version of this paper was presented at the International Association of Time Use Research (IATUR) Conference 2007 in Washington, DC, USA.

1 Introduction

Individualization is one dominant characteristic of modernity (Giddens, 1991; Castells, 2003) but, still, most people find themselves meshed into social and material contexts that restrict their opportunities to fulfil their own personal wants. The individuals feel restricted by circumstances out of their control and unable to reach goals they have set up for long and short term projects. In the popular debate lack of time is blamed for such shortcomings. Better knowledge about how people spend their time might provide ways to understand why there is not enough time. Time use studies have a great potential in this respect due to the richness of the collected diary data: a diary not only tells what people do, where they are located, who they are together with, but also when they do what they do, for how long they do it and, not least, in what context of other activities they do it.

The richness of the diary data collected in time use surveys, however, is usually not fully utilized in their analysis. The diaries are frequently used to produce statistics on how much time individuals spend on various kinds of everyday activities (Eurostat, 2004). Comparisons between sexes, ages and family types are made and, in countries where time use surveys are performed repeatedly, changes over time are scrutinized. The important results from time use studies provide knowledge about the overall time use of average individuals in a society and about similarities and differences between groups. There is, however, much more to be found in this collected data, not least how people mesh their activities together in households and workplaces.

What activities an individual performs, and consequently what activities appear in the diary, are a result of an allocation process, during which the individuals' ambitions to perform activities of importance for reaching a personal goal are moulded by social rules, conventions, law, other personal goals and not least the restricted accessibility of material circumstances and location (Hägerstrand, 1970a). The outcome of this allocation process, meaning what activities the individual actually performs in the course of the day, often does not correspond exactly to the individuals' ambitions. Power over how time is used by individuals is introduced as soon as activities that concern division of labour in the household or in the workplace are set on the agenda. Since power exerted by one individual in a household, for example the power of a small child in immediate need of care, influences the activities performed by other household members (parents or siblings). The child's needs alter the order, sequence and pattern of activities performed by others in the household. The meshing of activities is hard to examine by the most common methods in time use studies since the appearance of activities that are related to each other in sequential order is seldom considered in the analysis of time use data. The complexity of the task seems overwhelming. The challenge is to look at the diary data in time use surveys from different angles.

The main contribution of this paper is the development of an interactive semi-automated activity pattern extraction algorithm implemented within the application developed for visualiz-

ing time use data called VISUAL-TimePAcTS[1] (Ellegård and Vrotsou, 2006). The underlying idea is that activity sequences within the empirical activity data, may give clues to research questions and hypotheses that are not identified when the order of activities is not taken into consideration. The goal is to assist and simplify the study of more complex activity combinations of everyday life. The algorithm is applicable to individual, household, group and population levels and can be used for finding arguments for policy development, for example on gender policy, as well as for individuals' own reflections upon their everyday life and what could be done to improve the personal well-being. The properties of the pattern extraction algorithm make it possible to dig deeper into the constitution of identified activity patterns, for example by changing the criteria for the pattern extraction in order to test variations within the identified pattern. Doing so also gives rise to research questions and allows the further investigation of the validity of these questions, as we will demonstrate later in the paper.

The paper is arranged as follows: in Section 2 an overview of some related work is given, Section 3 is a short description of the visualization tool and the representation that this work is based on. Section 4 describes the algorithm in detail, Section 5 presents an analysis scenario, and finally, conclusions are presented in Section 6.

2 Related work

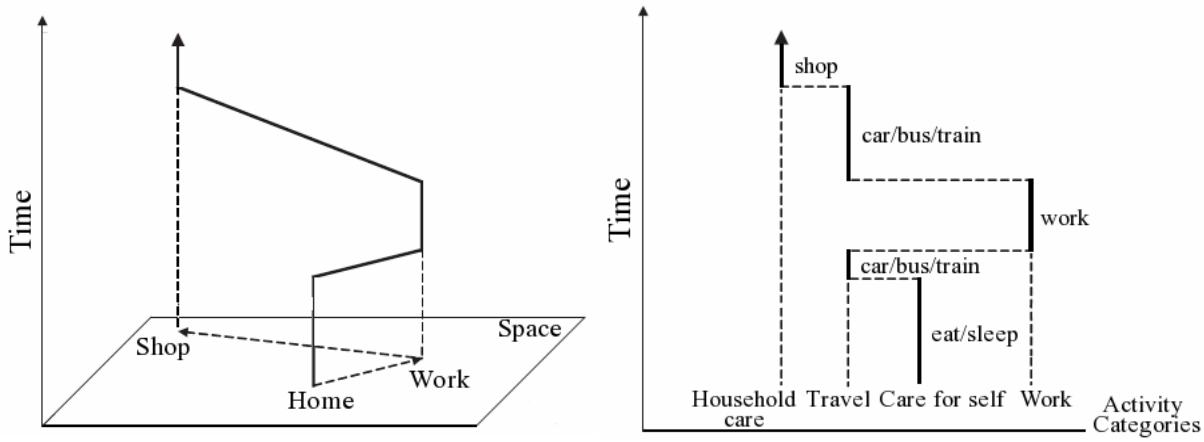
Identifying and studying patterns of activity and similarities/trends of these patterns within and between individuals' daily activity schedules is a subject of interest to the time use research community. There have been several approaches to perform studies of this kind, both visual and algorithmic. In this section we will consider research performed in different areas concerned with the identification of activity patterns.

The time geographical framework (Hägerstrand, 1970b) is an early example of using visual representation in the study of human behaviour and is considered an intuitive approach to represent and analyse similarities between individuals in space and time. This conceptual framework considers populations as groups of socially and geographically interrelated individuals and not as indistinct aggregate masses. Each individual is unique and their actions are defined and constrained by location in time and space, by socio-economic rules and conventions and by past experiences and knowledge. Time is a continuously changing variable that constrains the individuals' possibilities in the future, as an individual can be at only one place at a time and perform a limited number of activities at each time point (Lenntorp, 1976). An individual's movement in space and time can therefore be represented by a single continuous trajectory called a "space-time path" (Figure 1a). Several individuals' paths can be drawn within a single representation, the "space-time cube", revealing places in space and time where such paths meet, so called "bundles", and rendering the identification of patterns of

[1] The abbreviation VISUAL-TimePAcTS stands for VISUALization, Time, Place, Activity, Technologies used and Social companionship.

actions within populations possible. There are many studies that have used time geographical representations for the analysis of activity patterns, some examples follow. Kraak (2003) implemented the space-time cube in an interactive visualization environment. Kwan (1999, 2000) and Kwan and Lee (2004) have made extensive use of time geographical representations within a GIS environment to reveal human activity patterns. Huisman and Forer (1998, 2005) created a model for representing and analysing potential activity paths and action volumes in a GIS environment. A GIS data model was presented by Yu (2006) for analysing spatio-temporal patterns and interactions of human activities.

Figure 1
The "space-time path" (a) and the "activity path" (b)



(a) The "space-time path" representation of an individual's movement in space over time

(b) The "activity path" is an extension of the "space-time path" and is used to represent an individual's performed activities over time

Source: 1(a) Image based on Hägerstrand (1970b); 1(b) Image based on Ellegård (1999).

The original time geographical concept of the space-time path is mainly concerned with the spatial movement of an individual over time while the activities performed by the individual – if considered at all – are implicitly derived from the places visited during this time-space movement (Lenntorp, 1976). The activities an individual performs over time, however, can be visually described in a way that resembles their spatial movement over time. Activities, like the movements, take time to perform, they have a start time and a duration and occur sequentially. The original time geographical concept was therefore extended to also consider everyday life activities (Ellegård, 1999) which are also represented by a single continuous vertical trajectory in this case called the "activity path" (Figure 1b). This representation of activity diaries was incorporated into a visualization environment in order to facilitate the interactive exploration of these diaries (Ellegård and Cooper, 2004) resulting in the visual analysis tool VISUAL-TimePAcTS (Ellegård and Vrotsou, 2006). Using this representation individuals' activity paths can be compared and patterns of activity retrieved through purely visual methods. Trends can be spotted in the total representation and also a sequence of activities can be defined and highlighted revealing the distribution of this predefined pattern across the represented population. The drawback of this approach, however, is that it limits the activity com-

bination options to those that the researcher using the tool has in mind. There are also examples of visual approaches to the identification and study of activity patterns that do not use time geographical representations. Kwan (2000), for example, has used line representations and activity duration patterns over a geographical map to display activity patterns, and Zhao et al. (2008) have used representations such as 3D rods over a geographical map, and 3D activity ringmaps to display trends of daily activity.

A popular algorithmic method for the identification of activity patterns in social science, in general, and in time use research in particular, is sequence alignment (also known as optimal matching). Sequence alignment was first introduced to the social sciences by Abbott and Forrest (1986) and to activity pattern analysis by Wilson (1998). According to the sequence alignment method, which was originally developed for protein and DNA sequences (Kruskal, 1983), the similarity of two sequences can be determined by the number of operations needed to transform one sequence into the other. The operations used are insertion, deletion and substitution and each operation carries a cost. The smallest sum of these costs defines the degree of similarity between the sequences. Aligning all sequences in a set pair-wise and calculating their similarity score results in a similarity score matrix for the whole set which can then be used as input into clustering algorithms in order to classify the sequences into groups. Each of these groups can then be analysed and characteristic activity patterns identified within each. There has been a lot of research concerning the use of sequence alignment in the social sciences, Abbott and Tsay (2000) present a thorough review. Concentrating on travel and activity pattern analysis: Wilson (1998, 2001, 2006, 2008) has shown many applications and refinements to the identification of similar patterns within populations, as has Joh et al. (2001a, 2001b) and Lesnard (2006) among others. Schlich (2001) has, instead, applied sequence alignment to study variation in travel patterns within individuals' daily sequences in a population. Joh et al. (2002) introduced the incorporation of other attributes (such as location, duration, and start time among others), apart from the activity itself, in the similarity computation of sequences. They suggest a multidimensional alignment approach, and a heuristic method for its calculation, in order to reduce the search space. Wilson (2008) proposed the inclusion of geographical coordinates in the alignment process and hence the weighting of the costs calculation with a geographical distance.

There are a number of issues concerning the application of sequence alignment in activity time diaries. The greatest, which is an issue generally, is how to assign costs for the different operations since it may result in very different similarity matrices and hence classifications. Substituting activity "walking" with "running" may deserve a lower cost than substituting it with "eating", for example. Furthermore, since each alignment gives a single similarity score depending on the number of operations, two day sequences that include the exact same subsequence but at different times of day, which intuitively signifies a similarity between them, may receive the same score as two completely dissimilar sequences. Finally, choosing to include or discard duration in the alignment process can also alter the resulting classification. If duration of events is discarded then all events or sequences of events are considered equal

regardless of duration. A person, for example, performing a "care for others" activity for 5 minutes (perhaps helping a child dress) and then working the rest of the day will be ranked as identical with a person taking care of others the whole day and then working for an hour, even though their activity behaviour is actually very different. At the same time and for the same reasons, including duration can also have a negative effect on the results. Duration can be included by breaking the day up into intervals of a certain fixed time, and events are in turn broken up into several repetitions of themselves. If for example the day is broken up into 30 minute intervals, an event lasting 2 hours is represented by repeating the fixed time event 4 times in the daily sequence. Repetition of the same event several times can conceal otherwise apparent similarities between sequences and also depending on the time interval size short activities can be lost and small variations in the sequences disappear.

Less researched is the use of pattern mining methods in the social science field. The extraction of new knowledge, in the form of interesting relationships and patterns, from large databases is the central objective in the area of data mining. When the data analysed has a sequential nature, meaning that the data consist of ordered items, then the process is referred to as sequential mining (Han and Kamber, 2000). Defining interestingness in the context of pattern extraction is a complex and subjective matter. Most often frequency of occurrence is used as a representative measure, the process is then called frequent pattern mining. Frequent pattern mining was introduced by Agrawal et al. (1993) for the discovery of patterns in transaction databases, so called 'market basket analysis', and the apriori algorithm was introduced. The technique was later extended to consider also sequential data (Agrawal and Srikant, 1995) and refined in 1996 (Srikant and Agrawal, 1996). There has been extensive research on frequent pattern mining since its introduction, using different approaches. A thorough review of the current status of the discipline can be found in Han et al. (2007). In this paper we concentrate on the apriori approach, since it's the one we have based our work on, and refer the interested reader to Han et al. (2007) for further details on other methods. According to the apriori principle *a sequence of events is frequent only if all of its subsequences are frequent.* In order to identify frequent event sequences in the data, candidate sequences are then created stepwise by increasing them one element per iteration and these candidates are then identified in the database and filtered based on pre-specified constraints.

The nature of the time use diary data that we deal with here is similar to that of the sequential transaction data. A performed activity is a performed event in time. An individual performs several activities during a day in a certain order, these make up different activity sequences. The ordering of each of these sequences, their frequency of occurrence and the manner of their repetition within a population are of interest to the time use researcher as they may reveal interesting categorizations or characteristics within this population. The researcher should be able to define the attributes that these sequences must have in order to make them reveal interesting patterns to study. Hence, the apriori principle for mining frequent sequences can be used for the extraction of the patterns but the possibility should exist to include other criteria than just the frequency of their occurrence.

In this paper we have combined sequential mining, visualization and interaction techniques to allow the extraction of activity sequences from diary data. To do this we have adapted the apriori algorithm (Agrawal and Srikant, 1995) to our data and introduced interaction to its computation in order to allow the user to define interestingness through constraints that define the characteristics of the activity sequences and are not limited to frequency of occurrence. The fact that the user can control and restrict the sequence extraction is what makes the process semi-automatic.

3 Representation and data in VISUAL-TimePAcTS

The research work presented in this paper is developed as a feature in the visual activity-analysis tool VISUAL-TimePAcTS (Ellegård and Vrotsou, 2006), a visualization application for interactively studying activity diaries of individuals, groups and whole populations.

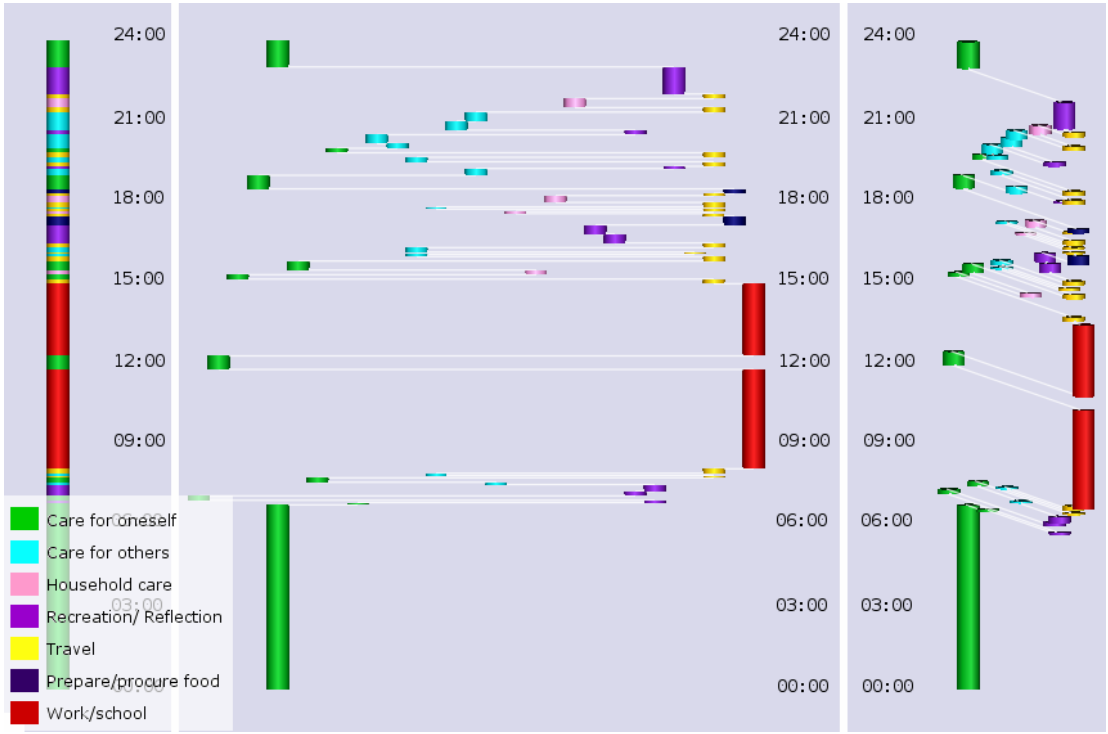
The central representation used within VISUAL-TimePAcTS is the activity path inspired by the time geographical conceptual approach (Hägerstrand, 1970b) as described in section 2. The activities in the collected diaries are classified into a hierarchical scheme of about 600 numerical codes with 5 levels of detail, with respect to the description of the activities, and grouped into 7 main activity categories (care for oneself, care for others, household care, reflection/recreation, transportation, procure and prepare food, and gainful employment or education). Each level of detail, n, is broken down into more detailed descriptions at level $n-1$ so level 5 is the most general level while level 1 is the most detailed. The seven generalized main activity categories (Ellegård, 1999, 2006) are each represented by a unique colour in VISUAL-TimePAcTS and consequently activities in all subcategories of the same main category have the same colour in the representation.

An individual's activity path in VISUAL-TimePAcTS can be rotated and studied from various angles. Seen from the front only the general division of activities into the seven main categories can be detected (Figure 3a) since sequences of activities within the same main activity category are not revealed (they all have the same colour). But if the same activity path is rotated the observer can see the breakdown of the seven main activity categories into more detailed subcategories of activities (Figure 3b, 3c). At a quick glance, the activity path seen from the front view (Figure 3a) may resemble a bar chart holding information about the time spent by the individual on each activity category (see, for example, Eurostat (2004)). There are, however, great differences since traditional time budgets represent an average individual. Important information is, therefore, hidden, such as the time of day when activities are performed, their duration and the number of times activities occur in the course of the day. This kind of sequence related information is constantly available to the viewer of the activity path in VISUAL-TimePAcTS and is important for detecting activity patterns.

The use of activity paths in the study of everyday life is useful as it also allows the study of two or more individuals simultaneously while, at the same time, preserving the uniqueness of

each individual. Drawing the activity paths of a group of individuals side by side in a box-like configuration (Figure 4), using the front view (Figure 3a), gives the researcher the opportunity to access information about the character and actual timing of the activities of whole populations in a single representation.

Figure 3
Visualization examples of the activity path of an individual in VISUAL-TimePAcTS



(a) front (b) side view (c) rotated view

Time is shown on the y-axis and colours represent the 7 activity categories. (a) shows the front view, where the general division of the activities can be detected at main category level. (b) shows the path in side view, revealing the breakdown into more detailed activity descriptions. (c) shows a slightly rotated view of the activity path in 3D.

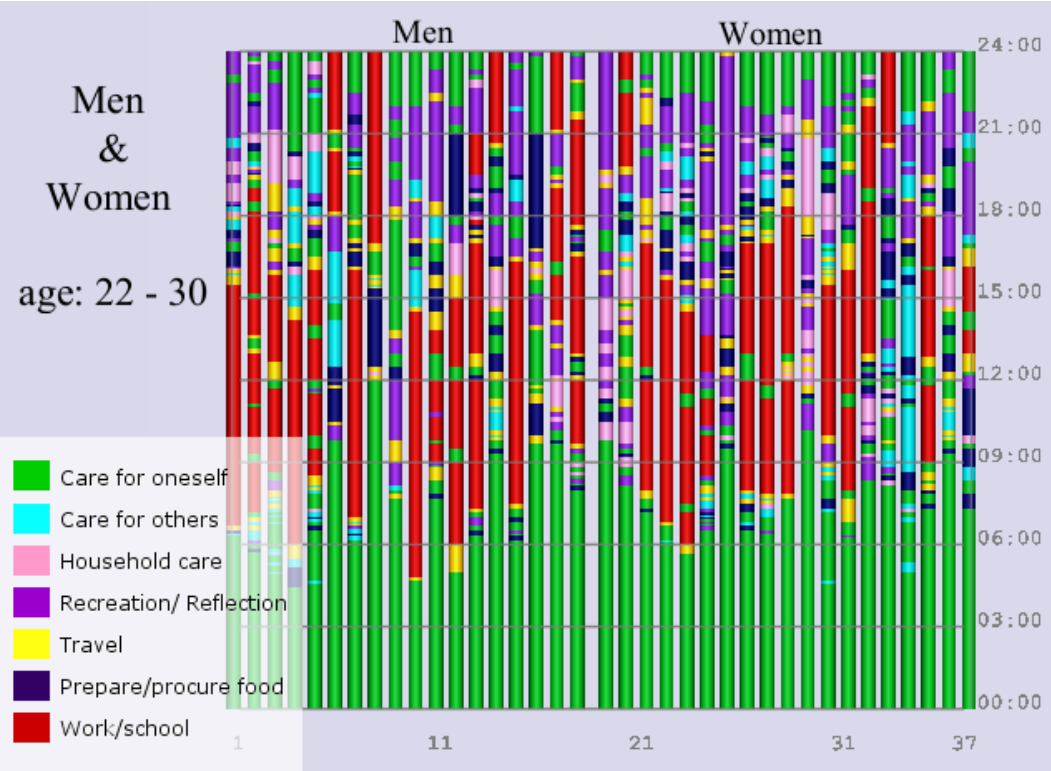
Source: Produced using VISUAL-TimePAcTS.

The diary data used in this work is a subset of time diaries collected in a pilot study by Statistics Sweden (*SCB*, www.scb.se) in 1996. A survey consisting of 179 households, in which 463 household members (aged 10 years and older) have filled in time diaries for one weekday and one weekend day. The subset we have chosen in this study includes individuals aged 20 to 65 years, 283 individuals in total (147 women and 136 men). Further, we have chosen to analyse weekdays and leave the analysis of weekend days for now. The sample might be regarded as relatively small, but since our aim is to demonstrate the algorithm and discuss research questions generated by using it, this is of minor importance.

In order to use the pattern extraction algorithm of VISUAL-TimePAcTS the diary data should be in the form of activities having a start time and a duration and occurring sequentially over a 24 hour period. Even though the coding scheme currently used in the pattern extraction dif-

fers from the schemes traditionally used in time diary surveys[2], adjustments can easily be made to incorporate these into the application.

Figure 4
Front view visualization of a weekday of a group of individuals aged 22-30 in VISUAL-TimePAcTS



Time is shown on the y-axis, individuals are ordered by sex and age from left to right on the x-axis. Colours represent the 7 activity categories.
Source: Produced using VISUAL-TimePAcTS.

4 Activity pattern extraction

An automatic pattern extraction algorithm can assist the time use researcher in two ways. First, it can allow the researcher more time to analyse the resulting activity patterns of a population, and second, such an algorithm could open up the possibility of new discoveries. The researcher may come across activity patterns that were unexpected and gain new insight about

[2] This categorization scheme differs in some ways from other schemes and the main difference is that what commonly is called "domestic work" (for example in the time use surveys used in the harmonized European scheme, Eurostat (2004)) in our scheme is divided into three main categories, namely "care for others", "household care" (comprising activities for care for buildings, maintenance, cleaning, and care for other things and belongings) and "procure and prepare food". When looking for activity sequences by extracting activity patterns in VISUAL-TimePAcTS, it is important that the main activity categories are not so general and broad that they hide variations (Ellegård, 2006).

the time use of populations. This has been our motivation for attempting to use sequential pattern mining in time use research.

4.1 Definition of activity patterns

As mentioned previously, the order in which individuals perform their daily activities is significant. Therefore, studying how identical sequences of activities are spread across the diaries of a population gives insight and reveals similarities in the way that people live their lives. Activity patterns are defined as the constellations that emerge from the way activity sequences are distributed in the diary data. We separate between activity patterns at the individual and the population level.

The same activity sequence distributed across the diary day or days of a single individual is defined as an *individual activity pattern*. These are most useful when studying repetitive behaviour of a single individual over a longer period of time. The same activity sequence distributed over single day diaries of a whole population reveals a *collective activity pattern*. Collective activity patterns are more appropriate when studying similarities and differences either between the individuals within a single group or between different groups. The choice of type of activity pattern to study depends, of course, on the research question.

4.2 Algorithm description

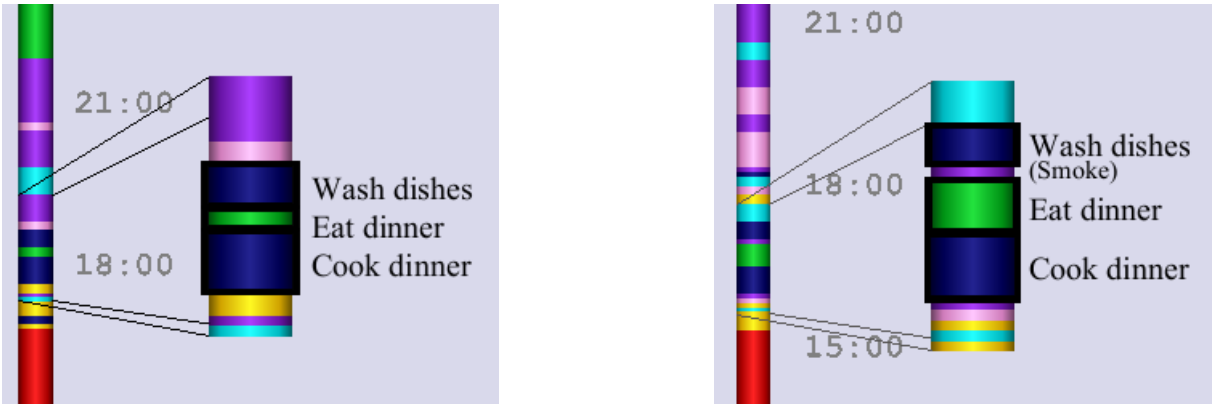
Activity diaries are considered as events occurring over time in a certain order: sequences of events. A sequence of two (*double*), three (*triple*), four (*quadruple*) or any number of n activities will also be referred to as an *n-tuple,* a *tuple* of n, or simply a *tuple*. The goal with the algorithm is to extract interesting *n-tuples* from the diaries, meaning *n-tuples* whose distribution constitutes interesting activity patterns. What is classified as interesting is defined by the researcher using the algorithm by allowing them to set constraints on the algorithm that determine the attributes of the identified activity patterns.

An *n-tuple* can be integrated in an individual's diary in two ways. Activities can succeed each other directly, leaving no gap in between ($gap = 0$) or other activities, that are not part of the *n-tuple*, can interrupt the *tuple* activities creating a gap between them ($gap > 0$). This can be seen in Figure 5 where the *3-tuple* "cook dinner → eat dinner → wash dishes" has been located in two different individuals' activity paths. In Figure 5a the individual washes the dishes immediately after having finished dinner, while the individual in Figure 5b takes a pause to smoke (a one activity gap) before washing the dishes.

We have used an apriori algorithm (Agrawal and Srikant, 1995) as our starting point for the activity pattern extraction and adjusted its computation and constraints to match our diary data. We use the lower order event sequences to create higher order ones depending on the constraints that define the interesting attributes in an activity pattern. We have introduced a lot of user control over the computation of the algorithm as the main goal is not simply to find

frequently occurring activity sequences, so the user should also be able to decide on the characteristics of the extracted patterns.

Figure 5
Examples of the activity sequence (*tuple*) "cook dinner→ eat dinner→ wash dishes" integrated in different ways in two individuals' diaries



a) a zero gap match b) a gap = 1 match

Source: Produced using VISUAL-TimePAcTS.

The activity pattern extraction algorithm principally iterates over three steps (Figure 6):

(1) generation of candidate tuples
(2) location of the candidate tuples in the dataset
(3) filtering of the located candidates according to user constraints

The user constraints that can be set, which will be explained in detail later, are:

(1) a minimum and maximum tuple duration
(2) a minimum and maximum gap between adjacent activities of the tuple
(3) a minimum and maximum number of occurrences of the tuple in a pattern
(4) a time window within which the emerging activity pattern must occur
(5) a minimum and maximum number of individuals that should perform the tuple

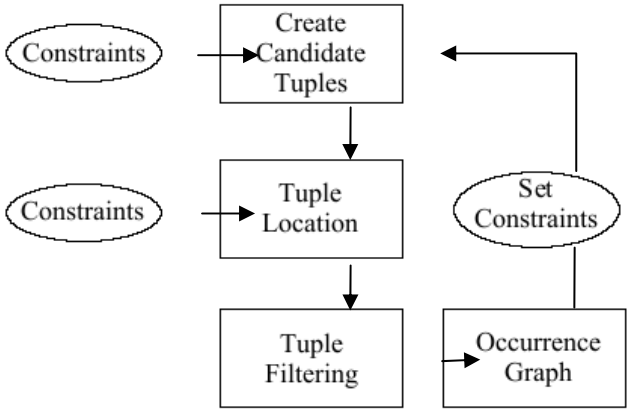
These criteria are those that we have found useful so far but the list is being extended as required. After the algorithm has run to completion the resulting extracted *n-tuples* become available to the user for visualization and interactive visual analysis of the resulting patterns. Next we will go through each step of the algorithm in more detail.

4.3 Candidate tuple generation

The first step of the activity pattern extraction algorithm is the candidate *tuple* generation. The candidate *tuples* are generated stepwise by increasing them by one activity per iteration. In the first iteration the single activities performed by the population are identified and counted and the ones that don't fit the constraints are ignored while the rest are considered the valid ones and go on to the next step of the iteration. In the second iteration the valid single activities (*1-*

tuples) are joined together to create pairs of activities (*2-tuples*). All pairs that satisfy the constraints are then the valid *2-tuples* and sent to the next step of the algorithm while the others are discarded. The iterations continue similarly, *2-tuples* are joined to create *3-tuples*, *3-tuples* are joined to create *4-tuples* etc. until no more candidate *n-tuples* can be generated that satisfy the set constraints.

Figure 6
Overview of the activity pattern extraction algorithm



In order to join two *n-tuples* they have to have *n-1* elements exactly identical and result in at most two *(n+1)-tuples*. Due to the sequential nature of the data a join operation between two *n-tuples* can be performed in exactly four ways regardless of the value of *n*: (1) the first *n-1* elements (1, ... , n-1) of both *n-tuples* are identical, (2) the last *n-1* elements (2, ... , n) of both *n-tuples* are identical, (3) elements 2,..., n of the first *n-tuple* are identical with elements 1, ... , n-1 of the second *n-tuple*, (4) elements 1, ... , n-1 of the first *n-tuple* are identical with elements 2, ... , n of the second n-tuple. Let us illustrate this by an example. If *a, b, c, d* are the activities included in two 3-tuples to be joined then the different join operations that can be applied to create the 4-tuples are (the join operation is denoted by the symbol ⋈):

$$
\begin{aligned}
(1)\quad & (\boldsymbol{a},\boldsymbol{b},c) \bowtie (\boldsymbol{a},\boldsymbol{b},d) \rightarrow (\boldsymbol{a},\boldsymbol{b},c,d) \\
& \qquad\qquad\qquad\quad \rightarrow (\boldsymbol{a},\boldsymbol{b},d,c) \\
(2)\quad & (a,\boldsymbol{b},\boldsymbol{c}) \bowtie (d,\boldsymbol{b},\boldsymbol{c}) \rightarrow (a,d,\boldsymbol{b},\boldsymbol{c}) \\
& \qquad\qquad\qquad\quad \rightarrow (d,a,\boldsymbol{b},\boldsymbol{c}) \\
(3)\quad & (a,\boldsymbol{b},\boldsymbol{c}) \bowtie (\boldsymbol{b},\boldsymbol{c},d) \rightarrow (a,\boldsymbol{b},\boldsymbol{c},d) \\
(4)\quad & (\boldsymbol{a},\boldsymbol{b},c) \bowtie (d,\boldsymbol{a},\boldsymbol{b}) \rightarrow (d,\boldsymbol{a},\boldsymbol{b},c)
\end{aligned}
$$

A candidate *(n+1)-tuple* is valid if and only if it is composed of valid sub-tuples, meaning sub-tuples that have survived the previous iterations' filtering. Because of this many generated candidates can be immediately eliminated from the process thus reducing the search space and hence the calculation time of the algorithm.

When the candidate patterns have been generated they are sent to the next step of the algorithm; the *tuple* location step.

4.4 Tuple location

The algorithm steps through the generated candidate *tuples* and matches each of them to the diary data, meaning it identifies them in the individuals' diaries. The constraints set by the user are considered during this search and the matches that don't satisfy these constraints are ignored, while the ones that do match them are considered to be the extracted *tuples*. A record is kept of the number of occurrences of each extracted *tuple*, the individuals performing them, and the *tuples*' location in the dataset. This information is saved for every iteration of the algorithm in a data structure and is then used in the study and visualization of the patterns. If no matches are found for the generated candidate *tuples* then the algorithm terminates otherwise the extracted *tuples* are filtered.

4.5 Filtering of extracted tuples

During the filtering step the extracted *tuple* matches are tested against the user specified constraints. Let us take a closer look at these constraints.

(1) The user can specify a *minimum and maximum duration* that an *n-tuple* in the activity diaries should have in order for it to be classified as an interesting activity pattern member. A user can, for example, decide that only short activity *tuples* that complete within 2 hours are interesting to study.

(2) A *minimum and maximum gap* allowed between the activities of an *n-tuple* can also be defined. This means that a user can choose the number of other activities that are allowed to interrupt two adjacent *tuple* activities. The user may want to study patterns consisting of *tuples* in which activities follow one another immediately in the individuals' days, as in figure 5a, or may regard the *tuple* in figure 5b as equally valid.

(3) The *minimum and maximum number of occurrences* of each extracted *n-tuple* can also be set by the user. The user can select to study only frequently occurring *n-tuples* for example.

(4) A *time window* deciding the time of day of occurrence for the emerging activity pattern can be specified. A user may, for example, be only interested in studying activity patterns that occur in the evening.

(5) And finally the *minimum and maximum number of people* that should be performing the extracted *n-tuple* can be set. The user for example may be interested only in patterns consisting of *n-tuples* that are performed by the majority of the population.

Some of the constraints are also applied during the candidate generation and the *tuple* location in order to speed up the process. The time window constraint, for example, is applied when initiating the algorithm and counting the single activities. There is no need to take into account activities that are outside of the specified time window as these will be eliminated in the

filtering step either way. The time window, the *tuple* duration, and the minimum and maximum gap are considered in the location step and *tuple* matches that exceed these limits are not recorded. Finally, in the filtering step all limits are tested against all the extracted *n-tuples*.

When the filtering step of an iteration has finished, a frequency graph is drawn showing the number of occurrences of the extracted *n-tuples*. The user can, at this stage, choose to define new constraints that will apply to the next iteration or continue the pattern extraction process with the same settings. If no extracted *tuples* survive the filtering then the algorithm terminates and the results are ready to be visualized, otherwise it continues to the next iteration and the generation of new higher order candidate *tuples*. The user can also choose to terminate the algorithm at any stage.

4.6 Visualization and interaction

The extracted *n-tuples* are listed, by order *n*, in the graphical user interface of VISUAL-TimePAcTS and made available to the user. The user can select, by clicking on the list with the mouse, one or more extracted *n-tuples* to be displayed in the visualization window. The extracted *tuples* are highlighted in the visualized data by being drawn in colour while surrounding activities are shown in grey. The pattern activities are coloured depending on the activity category that they belong to. Representing the sequences in this manner allows the user to interactively explore the extracted patterns in context and reveals how the activity sequences are distributed throughout the day, how different individuals perform them, and which activities are likely to interfere with and interrupt the carrying out of the larger projects which these sequences represent. An activity pattern emerges by the representation of the distribution of the *n-tuples* across the diaries in the population.

The user can switch between the default visualization and the pattern visualization, at any time, and can also switch between the different levels of the extracted patterns.

4.7 Filtering script language

The pattern extraction algorithm finds all the *tuples* in the data that match the user's criteria. This can result in large numbers of activity patterns that aren't always easy to examine. For this reason further filtering of the identified patterns has also been added to the pattern extraction feature. A scripting language has been implemented that allows the user to write commands applying logical operations on the resulting *tuple* set of a specific order, *n*, in order to narrow the results. The operators available to the user are:

(1) AND operator (&). The user can define one or more activities all of which must be present in the *tuples*. The command "work"&"lunch" (900&3), for example, will filter out all *tuples* that do not have both work and lunch activities present.

(2) OR operator (;). The user can define one or more activities at least one of which must be present in the *tuples*. The command "work";"lunch" (900;3), for example, will filter out all *tuples* that do not include either work or lunch activities.

(3) FOLLOWED BY operator (:). The user can narrow the search to patterns where certain *tuple* activities or ranges of *tuple* activities succeed one another. For example the user can search for *tuples* in which a travel activity is followed by a work activity. The command for this would be: "travel":"work" (550-649 : 900).

(4) RANGE operator (-). The user can select an activity range that the pattern activities should lie within. A single range can be decided for all elements in the *tuples*, or for each element separately. For example the user can narrow the results to *tuples* having the first element within the code range 0-100 (care for oneself activities). The command for this would be: "care for oneself": *any activity* (0-100 : *).

These different operators can be combined and create longer filtering commands to be applied. For example, the command ("lunch";"coffee"):"work" ((3;4):900) keeps only *tuples* in which the activity work is preceded by either lunch or coffee activity.

4.8 Algorithm efficiency

The algorithm and the visualization framework are implemented in C++, OpenGL and using wxWidgets for the graphical user interface. The algorithm was run on a laptop PC with a dual core 2GHz Centrino CPU and 2GB RAM, for a dataset consisting of 289 individuals performing, in total, 10,514 activities, and applying different constraints to the pattern extraction. Table 1 shows performance times for these test runs. The results show that activity patterns are extracted in interactive times for large subsets of the population, as long as constraints are set on the pattern extraction.

Table 1
Results from running the pattern extraction algorithm on a laptop PC with a dual core 2GHz Centrino CPU and 2GB RAM and applying different constraints

Example	Max. order (*n*)	Level of detail	Min. people	Max. tuple duration	Max. gap	TOTAL TIME (sec)
1	4	2	15	4 hours	0	4.03
2	5	2	15	8 hours	0	4.45
3	5	2	15	4 hours	4	12.67
4	7	2	15	8 hours	4	15.71

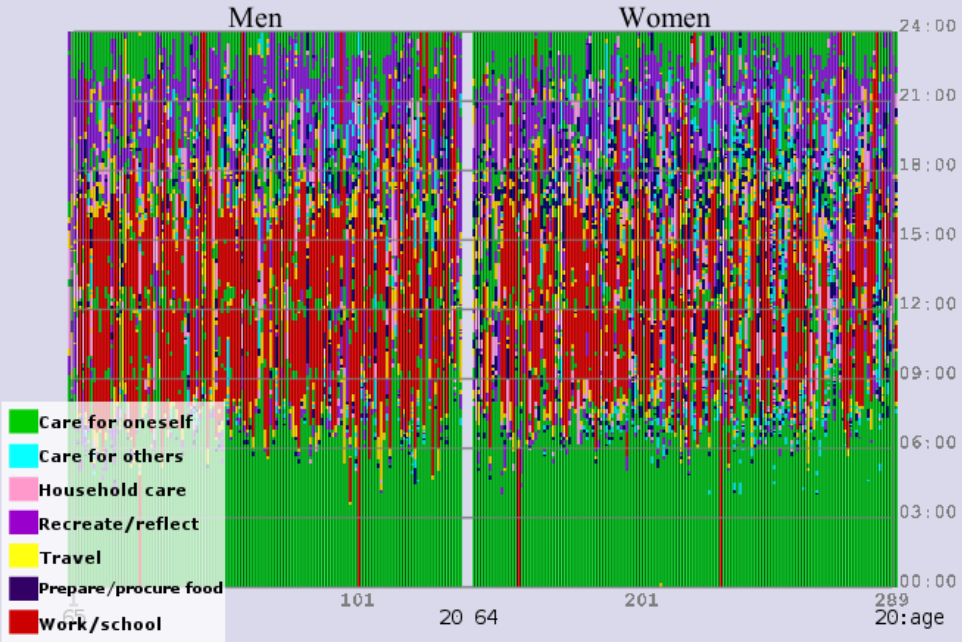
Source: Calculations computed within VISUAL-TimePAcTS.

5 Activity analysis scenario

In order to demonstrate how the pattern extraction process works in VISUAL-TimePAcTS and show how to analyse and better understand the arrangement of activity patterns we will go through an example step by step.

Individuals aged 20 to 65 in the population database are chosen to be studied on a weekday with an activity classification level of detail of 2; a quite high level of detail. Figure 7 shows the front view visualization of the activity paths of this group within VISUAL-TimePAcTS and Table 2 shows some numerical information concerning the selected group.

Figure 7
Front view visualization in VISUAL-TimePAcTS of a group of individuals aged 20 – 65



Time is shown on the y-axis and the individuals are ordered along the x-axis by age and gender. Colours represent the 7 activity categories
Source: Produced using VISUAL-TimePAcTS.

Table 2
Numerical information about the selected group of individuals

	Selected group
Age	20 – 65
No individuals	289
Women	150
Men	139
Diary entries	10514
No of unique activities	262

Source: Calculations computed within VISUAL-TimePAcTS.

For the first run of the algorithm the specified constraints were: a maximum activity sequence (*n-tuple*) duration of 10 hours, no gap between the adjacent *tuple* activities and a minimum of 15 individuals performing the activity pattern (see Table 3).

After having defined the constraints that the extracted patterns should meet, the first iteration of the algorithm can start. The unique single activities (*1-tuples*) are generated, located, and filtered according to the algorithm description in section 4. The first iteration concludes with the display of a graph showing their occurrence frequency at which point we can choose to alter the constraints that will apply to the second iteration or continue with the same ones. We choose to keep the same constraints for all iterations and continue to go through the subsequent iterations in the same manner until the algorithm terminates. Using the previously described data and constraints we extract *tuples* up to order 5, *5-tuples*.

Table 3
User specified constraints applied to the first example run of the pattern extraction algorithm

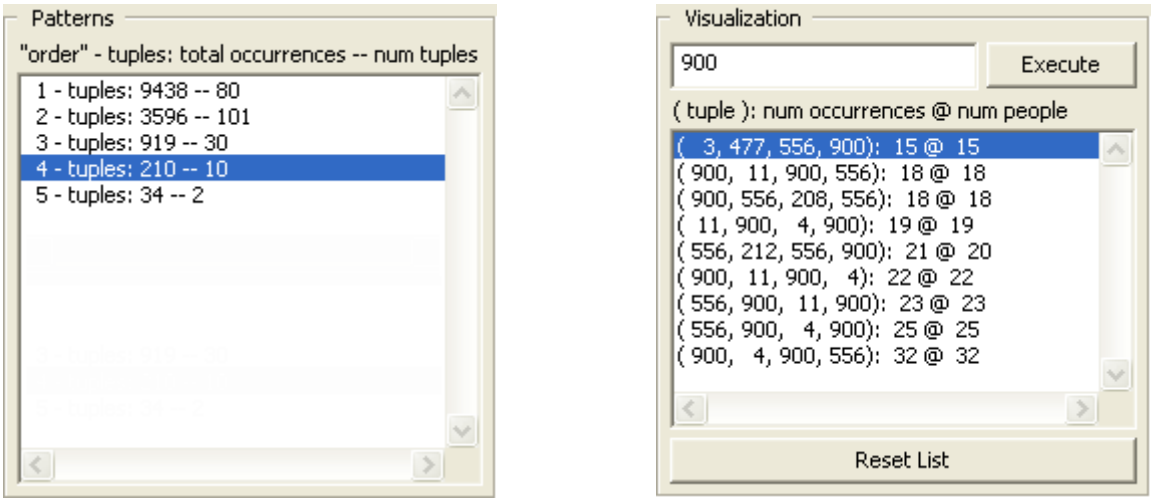
	Minimum	Maximum
Pattern duration (hours)	0	10
Time window	00:00	24:00
Activity gap	0	0
Pattern occurrences	1	no limit
No of individuals performing the pattern	15	no limit

Figure 8a shows the list of the groups of all orders *n* of the extracted *tuples*, in this case $n = 5$. By clicking, with the mouse, on an item in this list, a group of extracted *n-tuples* can be selected (in Figure 8a, for example, the *4-tuples* are selected). Upon selection, the list of all *n-tuples* that are included in this group is shown in the interface (Figure 8b shows a subset of the list of *4-tuples*). Selecting one or more distinct *n-tuples* from the list will result in their pattern being drawn in the visualization window.

We have chosen to start with the *4-tuples* (sequences of 4) in the list of extracted *tuples* (Figure 8a) and look for potentially interesting collective activity patterns containing the activity "work" (code 900). In order to do this the script language was used to filter out all *tuples* that do not include work (Figure 8b). The *4-tuples* containing "work" are 9 of the total of 10. Figure 8b shows how these are presented in the VISUAL-TimePAcTS user interface. Most of these work-related *4-tuples* are not very exciting: the majority of them are comprised of a combination of meals (here codes 3, 4, 11), travel (here code 556) and travel related activities (like dropping off or picking up somebody (codes 208, 212) on the way somewhere). However, in one of them there is one activity that stands out as it differs in nature from the rest, namely the activity "read the newspaper" (code 477). We find this deviation interesting and choose to analyse it further. The complete activity sequence that includes "work" and "read the newspaper" is: "have breakfast→ read the newspaper→ travel by car→ work" – or written in the codes: 3→477→556→900. Since breakfast is one of the activities in the chosen *4-tuple*, we can suspect that its distribution creates an activity pattern which is related to mornings. Furthermore, since the last activity in the sequence is "work" we will call this *4-tuple*

"getting ready for work". "Getting ready for work" ought to be relatively evenly spread between working men and women, at least among those who do not have to drop off children at the day care centre or school. Gender similarities and differences concerning how the morning activities are organized and performed are of interest in many respects. In households, for example, for discussing who does what kind of tasks in the morning rush and what is the division of labour, but also among policy makers for finding arguments for policy measures to provide equal opportunities for men and women to participate in the labour market.

Figure 8
Pattern extraction algorithm results as seen in VISUAL-TimePAcTS



(a) (b)

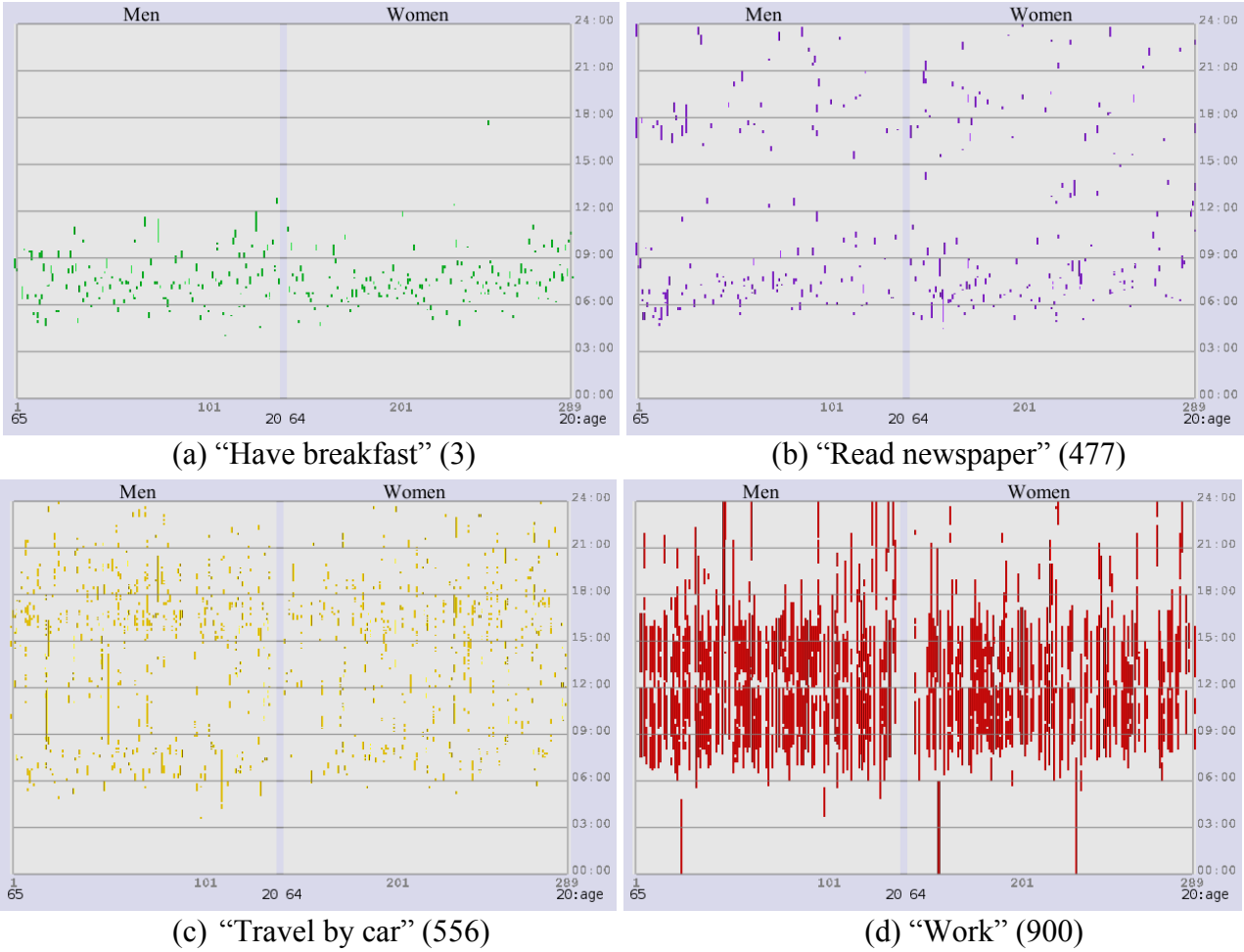
(a) List of all extracted *n-tuple* groups (*4-tuples* are selected),
(b) List of extracted *4-tuples* which include the activity "paid work" (900).
Source: Screen shot image of the VISUAL-TimePAcTS user interface.

After identifying this collective activity pattern as "getting ready for work", it may then be informative to see how often the distinct activities involved in the pattern appear among the individuals in the population, as well as examine whether there are differences between men and women. This can be done by looking at the single activities composing it. The distinct single activities making up the *4-tuple* "getting ready for work" appear frequently during the week day in the population. "Have breakfast", for example, appears in the data 258 times, "read newspaper" 287 times, "travel by car" 496 times, and "work" 947 times. These activities are quite evenly distributed among men and women, as can be seen in Figure 9, even though "travel by car" is a bit more frequent among men. From this information we can conclude that there are not very big gender differences when the activities are looked upon as single events. The next step is then to see if the result is the same when we look at the more complex (higher order) activity sequences.

The generated research question is, hence: How is the activity sequence "getting ready for work" distributed among individuals in the population and, more precisely, between men and women? The even distribution of the distinct single activities indicates that this should be the

case for the complete sequence also. To answer this question we study the visualization of the collective activity pattern created by the selected *4-tuple* (Figure 10). This collective activity pattern appears only 15 times in the population[3] and is performed by 15 individuals. It shows a great difference between men and women, with only two women performing the activity sequence as opposed to 13 men. Furthermore, we can see that it is performed primarily by men aged 35 and older. However, since each of the distinct activities of the sequence were evenly distributed between men and women in the selected population, we have to dig deeper into the data to understand why this inequality appears.

Figure 9
Visualization of the distinct single activities making up the collective activity pattern "getting ready for work": "have breakfast→ read newspaper→ travel by car→ work" (3→477→556→900) in VISUAL-TimePAcTS



(a) "Have breakfast" (3)

(b) "Read newspaper" (477)

(c) "Travel by car" (556)

(d) "Work" (900)

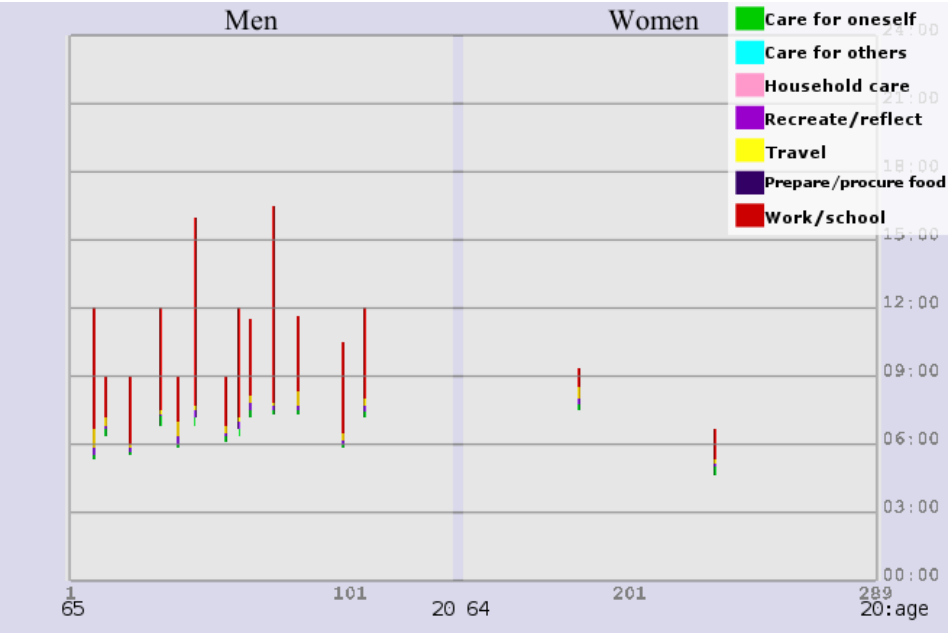
Source: Produced using VISUAL-TimePAcTS.

To do this we go back to the list of *n-tuples* and choose to look at the *3-tuples*, focusing on those consisting of activities present in the "getting ready for work" *4-tuple*. Figure 11 shows the distributions of the two activity sequences that "getting ready for work" can be broken

[3] This is also seen after the code sequence as the number 15 in Figure 9b.

into; namely the *3-tuples* "have breakfast→ read newspaper→ drive car" (3→477→556) and "read newspaper→ drive car→ work" (477→556→900). The resulting activity pattern representations (Figure 11) are somewhat surprising as they show only a slight change in the number of individuals performing the *3-tuples* and no change in the overall distribution. We already know, however, from looking at the single activities (seen in Figure 9), that women and younger men do engage, to greater extent, in all of the distinct single activities. So, we make a hypothesis that the *4-tuple* in question ("getting ready for work") is most likely performed by more individuals in the population than those extracted by the algorithm and shown in the representation. We can further assume that the *4-tuple* is probably interrupted by other activities in the majority of the individuals' diaries and therefore the strict constraints of the algorithm eliminated these individuals. In order to explore the assumed hypothesis we run the pattern extraction algorithm again with altered constraints. We permit a gap of 4, meaning that maximum 4 other activities may interrupt the adjacent activities of the *4-tuple*, as opposed to the previously set zero gap, while the rest of the constraints remain unchanged (Table 4).

Figure 10
Visualization of the *4–tuple* "have breakfast→read newspaper→travel by car→work" (3→477→556→900) in VISUAL-TimePAcTS



The constraints applied to the algorithm are: minimum of 15 people performing the *tuple*, maximum gap of zero between adjacent *tuple* activities and maximum duration 10 hours.
Source: Produced using VISUAL-TimePAcTS.

Re-analysing the data with this reduced constraint confirms our hypothesis. We find that more young men (13 additional) and women (9 additional) perform the *4-tuple* "getting ready for work", revealing a new collective activity pattern (Figure 12). 37 individuals carry out the *4-tuple*, compared to 15 when no interruptions are allowed. Further analyses can then be per-

formed to determine which are the activities that interrupt the *4-tuple* and study these in depth.

Table 4
User specified constraints applied to the second example run of the pattern extraction algorithm

	Minimum	Maximum
Pattern duration (hours)	0	10
Time window	00:00	24:00
Activity gap	0	4
Pattern occurrences	1	no limit
No of individuals performing the pattern	15	no limit

6 Conclusions

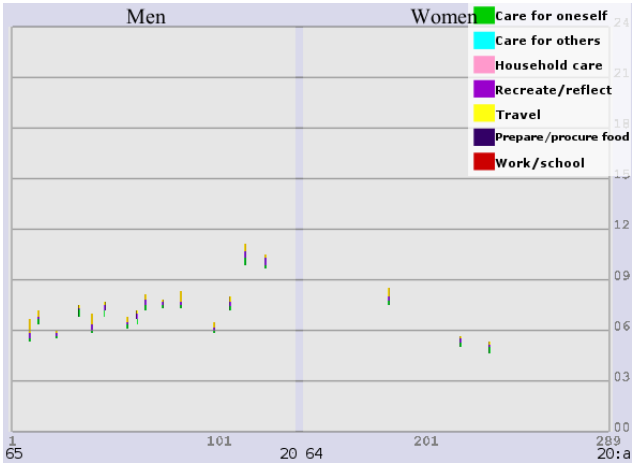
In this paper, we have presented a data mining algorithm which, combined with interaction and visualization techniques, facilitates the extraction and analysis of activity patterns from time use activity diaries. Further, we have demonstrated an example of how this analysis can proceed by going through a user scenario including identification of an interesting *tuple*, the raising of a research question, formation of a hypothesis and its verification. The goal of the pattern extraction algorithm has been to facilitate the automated identification of collective activity patterns in a population of individuals while preserving the group members' individuality when studying the identified patterns. The results from using the algorithm and analysing the extracted activity patterns appear promising with respect to this goal.

The pattern extraction algorithm should also be useful for finding answers to other methodologically and theoretically grounded research questions, for example questions relating to various activity patterns to empirically found indicators on well-being, like how health and sick leave are experienced. Activity patterns are also important in the making of a sustainable society, not least when it comes to energy used by appliances needed when activities are performed. Another interesting question is whether one specific collective activity pattern in a population or group predicts the appearance of a specific other activity pattern. Flexibility and ability to meet varying conditions and restrictions are hence important properties of methods for time use studies. This is met in the presented work by the interactive nature of the suggested pattern extraction process.

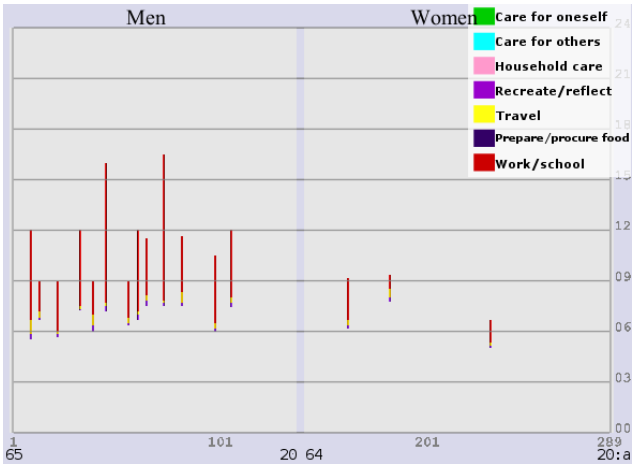
The analyst using the pattern extraction feature of VISUAL-TimePAcTS has freedom both in the extraction process of the patterns and in their analysis. The filtering script language implemented allows the analyst to narrow the results list and look at fewer at a time. The visualization of the results facilitates the understanding of the activity patterns and gives a concrete picture to use as a common ground for discussion and analysis. Using the VISUAL-

TimePAcTS pattern extraction algorithm helps researchers into time use to sort through the mass of activity data collected in diary surveys and helps to better understand combinations of activities in terms of collective and individual activity patterns. The combination of these features will help the user to extract new types of results from time use studies.

Figure 11
Visualization of the two extracted 3–*tuples* that make up the *4–tuple* "getting ready for work" (3→477→556→900) in VISUAL-TimePAcTS



(a) "have breakfast→ read newspaper→ drive car" (3→477→556)



(b) "read newspaper→ drive car→ work" (477→556→900)

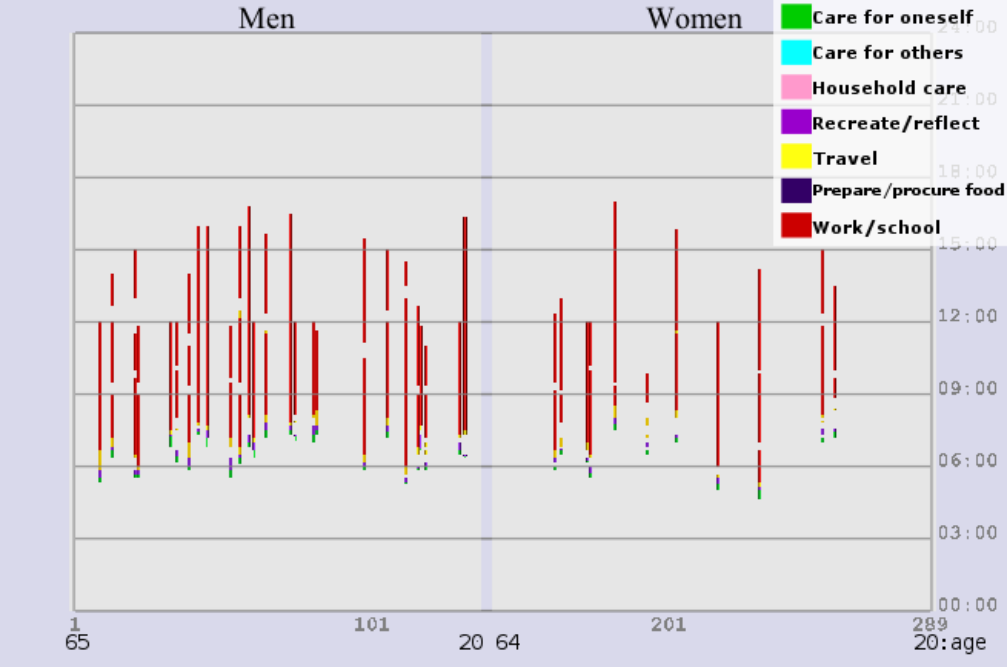
The constraints applied to the pattern extraction algorithm are: minimum of 15 people performing the *tuple*, maximum gap of zero between adjacent *tuple* activities and maximum *tuple* duration of 10 hours.
Source: Produced using VISUAL-TimePAcTS.

Future work includes the extension of the search and filtering criteria to support new users and new types of activity patterns in the data. Each new kind of task and new type of data being considered requires modifications to the search criteria and the list is becoming extensive to support the many types of user who may be interested in this type of searching.

Note

VISUAL-TimePAcTS is an application developed as part an ongoing research project and is continuously extended. A stable, distributable version of the application, including the functionality described in this paper, is currently being developed and will be available in December 2009. For further information please contact the authors.

Figure 12
Visualization of the *4–tuple* "breakfast→ read newspaper→ drive car→ work" (3→477→556→900)



The constraints applied on the pattern extraction algorithm are: minimum of 15 people performing the *tuple*, maximum gap of 4 between adjacent *tuple* activities and maximum *tuple* duration of 10 hours. 39 individuals (12 women and 27 men) display this activity pattern at the population level.
Source: Produced using VISUAL-TimePAcTS.

References

Abbott, A. and J. Forrest (1986), Optimal matching methods for historical data, in: *Journal of Interdisciplinary History*, Vol. 16, 473-496.

Abbott, A. and A. Tsay (2000), Sequence analysis and optimal matching methods in sociology – Review and prospect, in: *Sociological Methods and Research*, Vol. 29, No. 1, 3-33.

Agrawal, R., Imielinski, T. and A. Swami (1993), Mining association rules between sets of items in large databases, in: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington, DC, 207–216.

Agrawal, R. and R. Srikant (1995), Mining sequential patterns, in: *Proceedings of the Eleventh International Conference on Data Engineering*, Taipei, Taiwan, 3-14.

Castells, M. (2003), *The power of identity – The information age – Economy, society, and culture*, Vol. 2, Blackwell Publishers.

Ellegård, K. (1999), A time-geographical approach to the study of everyday life of individuals – A challenge of complexity, in: *GeoJournal*, Vol. 48, No. 3, 167-175.

Ellegård, K. and M. Cooper (2004), Complexity in daily life – 3D-visualization showing activity patterns in their contexts, in: *electronic International Journal of Time Use Research (eIJTUR)*, Vol. 1, No. 1, 37-59.

Ellegård, K. (2006), The power of categorisation in the study of everyday life, in: *Journal of Occupational Science,* Vol. 13, No. 1, 37-48.

Ellegård, K. and K. Vrotsou (2006), Capturing patterns of everyday life – Presentation of the visualization method VISUAL-TimePAcTS, in: *28th International Association for Time Use Research (IATUR) Annual Conference*, Copenhagen, Denmark.

Eurostat. (2004), *How Europeans spend their time – Everyday life of women and men, Data 1998-2002*, European Commission Theme 3 – Population and social conditions, Pocketbooks, Brussels.

Fails, J.A., Karlson, A., Shahamat, L. and B. Shneiderman (2006), A visual interface for multivariate temporal data – Finding patterns of events across multiple histories, in: *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 167-174.

Giddens, A. (1991), *Modernity and self-identity – Self and society in the late modern age*, Polity Press, Cambridge, UK.

Han, J. and M. Kamber (2000), *Data mining – Concepts and techniques (The Morgan Kaufmann Series in Data Management Systems)*, Morgan Kaufmann Publishers, San Francisco, CA.

Han, J., Cheng, H., Xin, D. and X. Yan (2007), Frequent pattern mining – Current status and future directions, in: *Data Mining and Knowledge Discovery*, Vol. 15, No. 1, 55-86.

Huisman, O. and P. Forer (1998), Computational agents and urban life spaces – A preliminary realisation of the time-geography of student lifestyles, in: *Proceedings of the Third International Conference on GeoComputation*, Bristol, U.K.

Huisman, O. and P. Forer (2005), The complexities of everyday life – Balancing practical and realistic approaches to modelling probable presence in space- time, in: *Proceedings of the 17th Annual Colloquium of the Spatial Information Research Centre (SIRC)*, University of Otago, Dunedin, New Zealand, 155-168.

Hägerstrand, T. (1970a), *Tidsanvändning och omgivningsstruktur (Time use in a structuring environment)*, Statens Offentliga Utredningar (SOU), Vol. 14, Annex 4, Allmänna Förlaget, Stockholm.

Hägerstrand, T. (1970b), What about people in regional science?, in: *Papers in Regional Science*, Vol. 24, No. 1, 6-21.

Joh, C.H., Arentze, T.A. and H.J.P. Timmermans (2001a), Multidimensional sequence alignment methods for activity-travel pattern analysis – A comparison of dynamic programming and genetic algorithms, in: *Geographical Analysis*, Vol. 33, 247-270.

Joh, C.H., Arentze, T.A. and H.J.P. Timmermans (2001b), A position-sensitive sequence alignment method illustrated for space-time activity-diary data, in: *Environment and Planning A*, Vol. 33, No. 2, 313-338.

Joh, C.H., Arentze, T.A., Hofman, F. and H.J.P. Timmermans (2002), Activity pattern similarity – A multidimensional sequence alignment method, in: *Transportation Research Part B – Methodological*, Vol. 36, No. 5, 385-403.

Kraak, M.-J. (2003), The space-time cube revisited from a geovisualization perspective, in: *Proceedings of the 21st International Cartographic Conference*, Durban, South Africa, 1988–1995.

Kruskal, J.B. (1983), An overview of sequence comparison – Time warps, string edits and macromolecules, in: *SIAM Review*, Vol. 25, No. 2, 201-237.

Kwan, M.P. (1999), Gender, the home-work link, and space-time patterns of nonemployment activities, in: *Economic Geography*, Vol. 75, No. 4, 370-394.

Kwan, M.P. (2000), Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems – A methodological exploration with a large data set, in: *Transportation Research Part C – Emerging Technologies*, Vol. 8, 185-203.

Kwan, M.P. and J. Lee (2004), Geovisualization of human activity patterns using 3D GIS – A time-geographic approach, in: Goodchild, M.F. and D.G. Janelle (eds.), *Spatially integrated social science*, Oxford University Press, 48-66.

Lenntorp, B. (1976), Paths in space-time environments - A time-geographic study of movement possibilities of individuals, in: *Lund Studies in Geography – Series B*, No. 44, Human Geography, Royal University of Lund, Department of Geography, Lund, Sweden.

Lesnard, L. (2006), Optimal matching and the social sciences, in: *28th International Association for Time Use Research (IATUR) Annual Conference*, Copenhagen, Denmark.

Schlich, R. (2001), Analysing intrapersonal variability of travel behaviour using the sequence alignment method, in: *European Transport Research Conference*, Cambridge, Great Britain.

Srikant, R. and R. Agrawal (1996), Mining sequential patterns – Generalizations and performance improvements, in: *Proceedings of the Fifth International Conference on Extending Database Technology (EDBT)*, Avignon, France.

Wilson, C. (1998), Activity pattern analysis by means of sequence-alignment methods, in: *Environment and Planning A*, Vol. 30, No. 6, 1017-1038.

Wilson, C. (2001), Activity patterns of Canadian women – Application of ClustalG alignment software, in: *Transportation Research Record*, No. 1777, 55- 67.

Wilson, C. (2006), Reliability of sequence-alignment analysis of social processes – Monte Carlo tests of ClustalG software, in: *Environment and Planning A*, Vol. 38, No. 1, 187-204.

Wilson, C. (2008), Activity patterns in space and time – Calculating representative Hägerstrand trajectories, in: *Transportation*, Vol. 35, No. 4, 485-499.

Yu, H. (2006), Spatio-temporal GIS design for exploring interactions of human activities, in: *Cartography and Geographic Information Science*, Vol. 33, No. 1, 3-19.

Zhao, J., Forer, P. and A.S. Harvey (2008), Activities, ringmaps and geovisualization of large human movement fields, in: *Information Visualization*, Vol. 7, No. 3-4, 198–209.