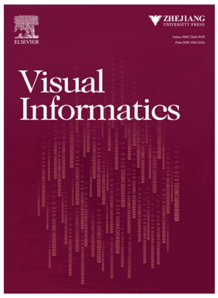
Journal Pre-proof



Exploring and visualizing temporal relations in multivariate time series

Gota Shirato, Natalia Andrienko, Gennady Andrienko

PII: S2468-502X(23)00039-6
DOI: https://doi.org/10.1016/j.visinf.2023.09.001
Reference: VISINF 184

To appear in: *Visual Informatics*

Received date : 21 June 2023
Revised date : 1 September 2023
Accepted date : 5 September 2023

Please cite this article as: G. Shirato, N. Andrienko and G. Andrienko, Exploring and visualizing temporal relations in multivariate time series. *Visual Informatics* (2023), doi: https://doi.org/10.1016/j.visinf.2023.09.001.

Exploring and visualizing temporal relations in multivariate time series

Gota Shirato[a,b,*], Natalia Andrienko[a,c], Gennady Andrienko[a,c]

[a] *Fraunhofer IAIS, Sankt Augustin, 53757, Germany*
[b] *University of Bonn, Regina-Pacis-Weg 3, Bonn, 53113, Germany*
[c] *City, University of London, Northampton Square, London, EC1V 0HB, UK*

Abstract

This paper introduces an approach to analysing multivariate time series (MVTS) data through progressive temporal abstraction of the data into patterns characterizing behavior of the studied dynamic phenomenon. The paper focuses on two core challenges: identifying basic behavior patterns of individual attributes and examining the temporal relations between these patterns across the range of attributes to derive higher-level abstractions of multi-attribute behavior. The proposed approach combines existing methods for univariate pattern extraction, computation of temporal relations according to the Allen's time interval algebra, visual displays of the temporal relations, and interactive query operations into a cohesive visual analytics workflow. The paper describes application of the approach to real-world examples of population mobility data during the COVID-19 pandemic and characteristics of episodes in a football match, illustrating its versatility and effectiveness in understanding composite patterns of interrelated attribute behaviors in MVTS data.

Keywords: temporal relations, temporal abstraction, multivariate time series, time intervals

*Corresponding author.
Email addresses: gota.shirato@iais.fraunhofer.de (Gota Shirato), natalia.andrienko@iais.fraunhofer.de (Natalia Andrienko), gennady.andrienko@iais.fraunhofer.de (Gennady Andrienko)

1. Introduction

Temporal abstraction means representing sequences of time-referenced data items as unified entities called patterns [4]. To comprehend the underlying dynamics and interrelationships among various attributes within multivariate time series (MVTS), it is crucial to uncover and explore temporal relations between patterns of individual attribute variations. Although numerous methods exist to address specific tasks in abstracting and analyzing MVTS, there is currently no overarching framework that consolidates these tasks and their corresponding methods into a comprehensive analysis workflow. Such a framework would help researchers to synergistically use different methods, leveraging the variety of existing techniques and enhancing their understanding of dynamic phenomena.

This paper presents a framework that aims to bridge this gap by supporting progressive abstraction of MVTS, from defining relevant intervals with basic behavioral patterns of individual attributes to exploring temporal relations between previously extracted patterns, which may differ in their levels of abstraction. Our primary objective is not to replace existing methods with new ones; instead, we show how to organize existing methods supporting different tasks into a cohesive visual analytics workflow [3]. We present examples of computational and visualization techniques capable to support different analysis steps, while the framework is conceptual and therefore allows the use of any appropriate methods. Thus, analysts can choose the types of patterns to search for, pick one of suitable existing methods that can detect these patterns in time series, and choose or design a time-oriented visualization technique that will show the positions of the detected patterns along the time line. To visualize relationships between the patterns, analysts may use node-link diagrams instead of matrices.

The proposed framework addresses two key problems: (a) identifying basic behavioral patterns of individual attributes, and (b) examining the temporal relations between these patterns across multiple attributes to derive higher-level abstractions. In this context, a basic pattern refers to an interpretable symbol or expression representing a sequence of values for a single variable, such as “increasing” or “decreasing” [28]. Deriving complex patterns of joint behavior of multiple variables is particularly challenging, especially when there are time lags between the starting points of individual behavior patterns. Our framework is designed to visualize temporal relations with lags, facilitating the comprehension of complex interactions among multiple attributes.

The framework focuses on three main tasks:

- T1: Defining relevant intervals with basic patterns in univariate time series
- T2: Deriving complex patterns by computing temporal relations between time intervals in multivariate time series
- T3: Exploring occurrence patterns of temporal relations through an interactive visual interface.

For T1, we apply an algorithm to define relevant time intervals from univariate time series using the geometric pattern extraction technique [30]. T2 can be fulfilled by using Allen's interval algebra [1]. T3 is supported by a visual exploration interface designed following B.Shneiderman's mantra of "Overview first, zoom and filter, then details-on-demand" [31].

We argue that these tasks serve as essential components in a comprehensive MVTS analysis workflow, demonstrating the cohesive nature of our framework. Our framework accommodates various types of patterns and temporal relationships, allowing analysts to apply existing methods for pattern detection. Example of such methods include trend-based [18] and state-based techniques [20].

The rest of this paper is structured as follows. Section 2 discusses the related work. Section 3 describes selected methods suitable for each task using the example of the mobility data during the COVID-19 pandemic. Section 4 demonstrates the effectiveness and versatility of our framework by example of another application using football (soccer) data. Section 5 discusses the concept, approaches, and answered research questions, identifies strengths and limitations, and proposes directions for future work. Finally, Section 6 concludes our work.

2. Related Work

In this section, we review the literature related to the analysis of multivariate time series, temporal abstraction, and visualization techniques for exploring temporal relations.

2.1. Multivariate Time Series Analysis

A variety of methods have been proposed for the analysis of multivariate time series (MVTS) data. These methods can be broadly categorized into statistical approaches (e.g., Granger causality [10], vector autoregression [22]), machine learning techniques (e.g., recurrent neural networks [12], Bayesian networks [26]), and matrix and tensor factorization methods [16]. While these approaches are effective in modeling and predicting various aspects of MVTS data, they often do not provide an intuitive understanding of the temporal relations between different attributes.

2.2. Temporal Abstraction

Temporal abstraction involves transforming the raw data into higher-level concepts that are easier to understand and interpret, thus creating interval-based representations from time-stamped data [29] (i.e., basic temporal abstraction [28]), and abstracting intervals into other intervals with a higher level of abstraction (i.e., complex temporal abstraction [28]). Techniques like time series segmentation [14], time periodization [2], motif discovery [27], and frequent episode mining [24] have been used to identify meaningful patterns in univariate and multivariate time series. Joint behavior of multiple variables are also derived from basic patterns by using co-occurrence [19] and simultaneity of different temporal patterns [30]. While these methods are effective in extracting temporal patterns, they do not explicitly address the problem of exploring and analyzing different types of temporal relations between the identified patterns.

2.3. Visualization Techniques for Temporal Relations

Several visualization techniques have been proposed to explore temporal relations in time series data. TimeMatrix [35] and TimeNotes [33] are examples of visualizations that present the temporal relations between events in the form of a matrix. Moreover, co-occurrences of pairs of different abstract patterns can be visualized by a network [30]. EventFlow [25] and Outflow [34] are visual analytics tools that enable users to explore temporal patterns in event sequences by providing interactive visualizations of event data. Techniques have also been introduced for visually specifying, combining, and querying complex temporal patterns [6]. Recent work supports the validation of causal relationships by showing correlations between time intervals in matrix [17] and network [7, 21]. An approach exists for investigating relationships between one or more time series within specified time frames [36]. However, these methods generally lack direct support for exploring patterns of temporal relations that include time lags.

In summary, the related work highlights a large variety of methods and techniques available for analyzing and visualizing multivariate time series data and temporal relations. Our proposed framework distinguishes itself by introducing an integrated workflow consisting of three tasks: identifying relevant intervals containing patterns, deriving complex patterns by computing temporal relations, and exploring occurrence patterns of temporal relations through an interactive visual interface . We propose a selection of methods that can be employed for each task but do not exclude the use of alternative techniques.

3. Visual analytics approach

In this section, we present a workflow composed of methods suitable for identifying time intervals with basic behavior patterns, computing temporal relations between time intervals in multivariate time series, and revealing patterns of joint behavior by visually presenting the relations between earlier extracted patterns.

3.1. Essence of the approach

The key idea of our approach is to conduct a progressive abstraction process from identifying basic patterns in univariate time series to discovering higher-level patterns formed by temporal relations between the basic patterns. This process is designed to simplify and distill complex multivariate time series data into meaningful and interpretable components that can be easily understood and analyzed. The workflow of the process is presented in Figure 1.

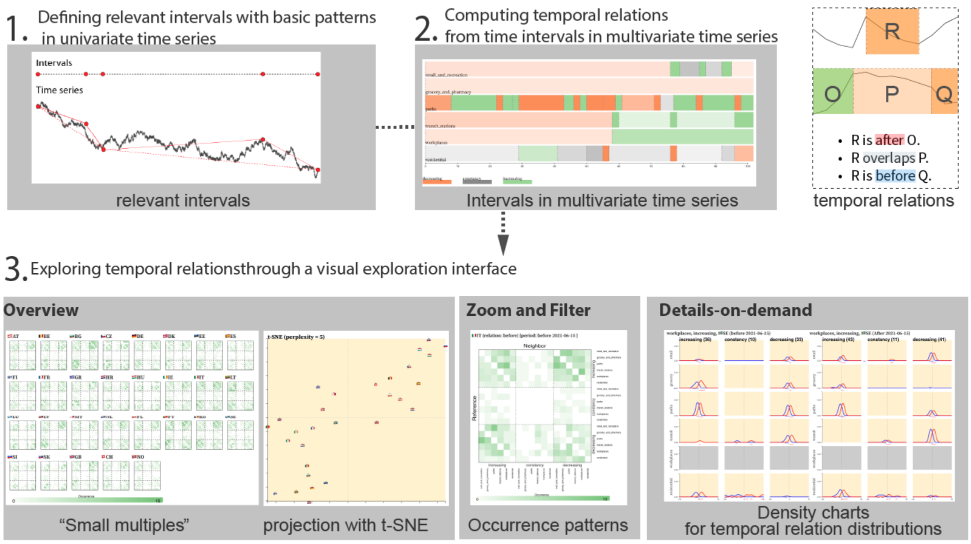


Figure 1: The workflow of the progressive abstraction process consists of three steps with increasing the level of abstraction at each step.

We demonstrate our approach by utilizing time series of continuous daily mobility indicators from Google Mobility Data [9].

3.2. Approach introduced by example

During the COVID-19 pandemic, measures were taken to curb infection spread by restricting people's movement. In mobility data, we observe temporal relations between different mobility patterns that warrant attention. When a lockdown measure is announced, for example, we can expect an `increase` pattern in coming to workplaces to occur `before` an `increase` pattern of staying at home since, many people may go to their offices to prepare for remote work. The mobility patterns may differ among different countries due to their varying policies against the pandemic. Here, our interest is to visualize temporal relations between mobility patterns within each country and to investigate the distributions of temporal relations across countries.

Data description

Following the COVID-19 outbreak in February 2020 [9], Google began publishing anonymized mobility data for six different categories of places:

`retail and recreation`, `grocery and pharmacies`, `parks`, `transit stations`, `workplaces`, and `residential` from various regions. The data consists of daily visitor counts to these categories, compared to baseline days prior to the pandemic's onset. Baseline days represent a normal value for each day of the week, calculated as the median value over a five-week period from January 3rd to February 6th, 2020. The values in the published data are presented as percentages of changes from these baseline values. For our analysis, we utilize daily time series for 29 countries across Europe, collected between the 15th of February, 2020 and the 15th of October, 2022.

T1. Defining relevant intervals containing abstract patterns

The first step in our framework involves dividing a univariate time series into a set of time intervals of varying lengths, each featuring a distinct pattern of value variation (e.g., a trend). Analysts can define the pattern types of interest and adjust thresholds for identifying them. In our example, we consider a set of basic patterns that can be represented as trends and typically labelled as `increase`, `constancy`, and `decrease`. To distinguish these trends, we employ the algorithm by Shirato et al. [30], which treats a time series as a graph of a function $V(t)$ in a Cartesian coordinate system. It determines a peak or trough pattern by identifying the largest triangle in a time interval $[t_1, t_2]$ formed by points $(t_1, V(t_1)), (t', V(t')), and (t_2, V(t_2))$, where $t_1 < t' < t_2$. In other words, the algorithm identifies a peak or trough point that forms the largest triangle with the starting and ending points of a given interval (Figure 2). If the area of the triangle is sufficiently large, the peak or trough point is taken as a break point to divide the time series into two segments each containing a simpler pattern of temporal variation that can be considered as increase, decrease, or constancy.

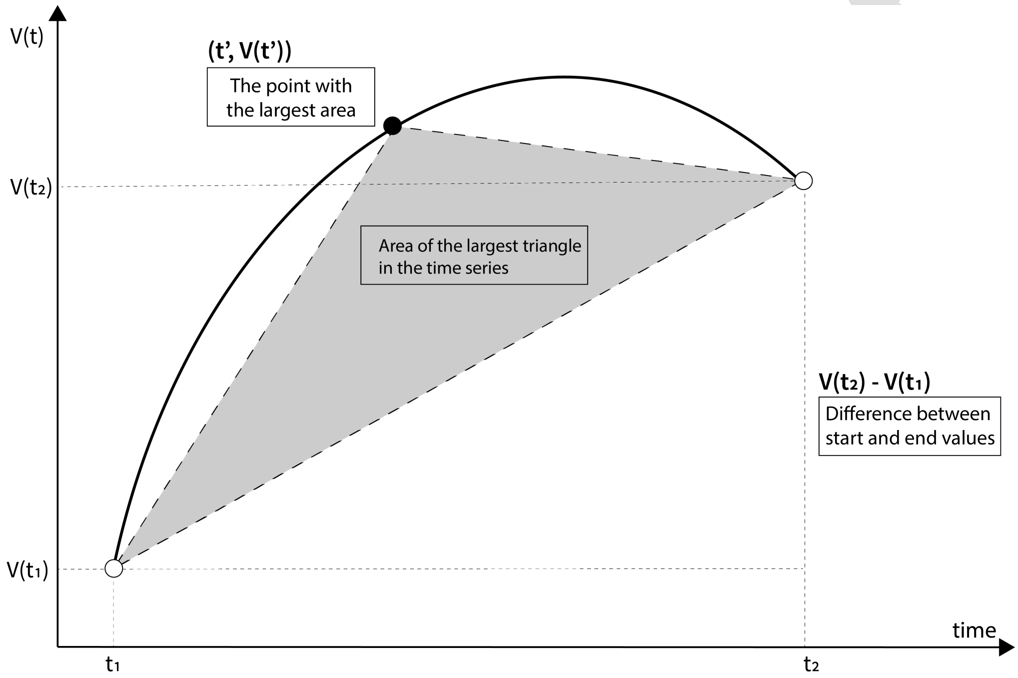


Figure 2: Illustration of the idea of the pattern detection algorithm [30]. The curve portrays the evolution of values over time as a function $V(t)$. The algorithm identifies a point $(t', V(t'))$ within an interval $[t_1, t_2]$ that forms the largest triangle with the points $(t_1, V(t_1))$ and $(t_2, V(t_2))$. The point $(t', V(t'))$ is used to divide the time series in two intervals. In this example, the first interval contains an increasing trend. Depending on a threshold for the difference between the first and last values, the attribute behavior in the second interval can be considered as a decreasing trend or as constancy.

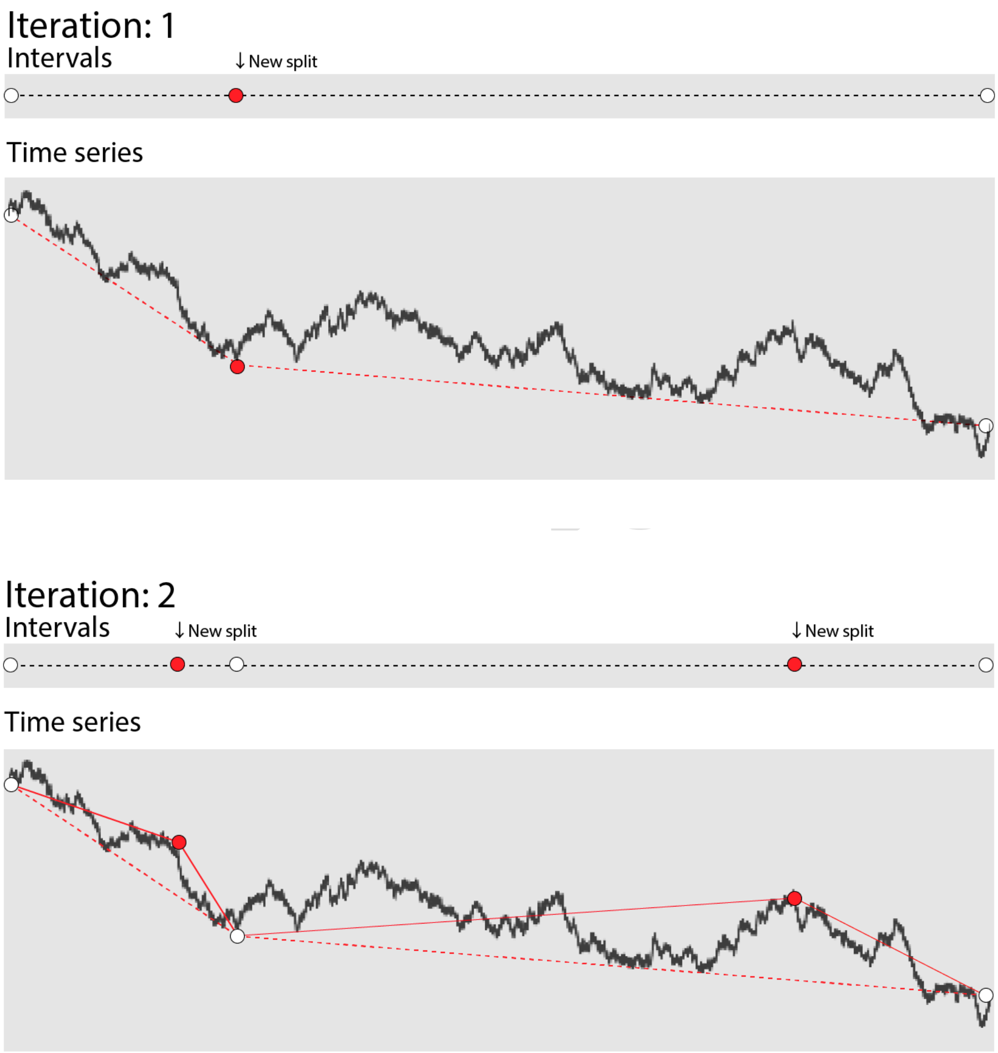


Figure 3: Dividing time intervals by iterative application of the largest triangle algorithm [30]. In the initial iteration (top), the algorithm finds the peak or trough point that forms the largest triangle in the time series, then segments the time interval at this identified time point. Subsequently, in the second iteration (bottom), the algorithm discovers peak or trough points in the time intervals obtained in the first iteration. The iterative process continues until the area of the largest triangle falls below a chosen threshold.

The operation of finding the largest triangle is applied to a time series in an iterative manner as illustrated in Figure 3. Initially it is applied to the whole time span of the time series $[t_0, t_{last}]$ (Figure 3, top). After finding the vertex of the largest triangle $(t', V(t'))$, the time step t' is used for dividing the entire time span into intervals $[t_0, t']$ and $[t', t_{last}]$. The operation is then applied to each of these two intervals, which can be, in turn, further divided into sub-intervals (Figure 3, bottom). The decision whether a given interval needs to be subdivided depends on the area of the largest triangle on this interval. A large area implies the presence of a substantial peak or trough, which is a composition of increasing and decreasing trend patterns. Hence, the interval needs to be subdivided for obtaining simpler patterns. When the triangle is small, it suggests that this fragment of the time series can be treated as a simple trend pattern with inessential noise.

With an aim to obtain a set of elementary trends, namely `increase`, `constancy`, and `decrease`, we recursively segment each time interval until its pattern is sufficiently distinct, i.e., the largest triangle within the segment is smaller than a threshold. A larger threshold allows for larger triangles, representing more significant peaks or troughs, to exist within the segment, resulting in fewer intervals as the partitioning process is less stringent. In other words, increasing the threshold value leads to a coarser segmentation of a time series and a higher level of abstraction, where segments are considered as simple trends, and internal deviations are ignored. Figure 4 demonstrates the effect of the threshold on the final division of a time series.

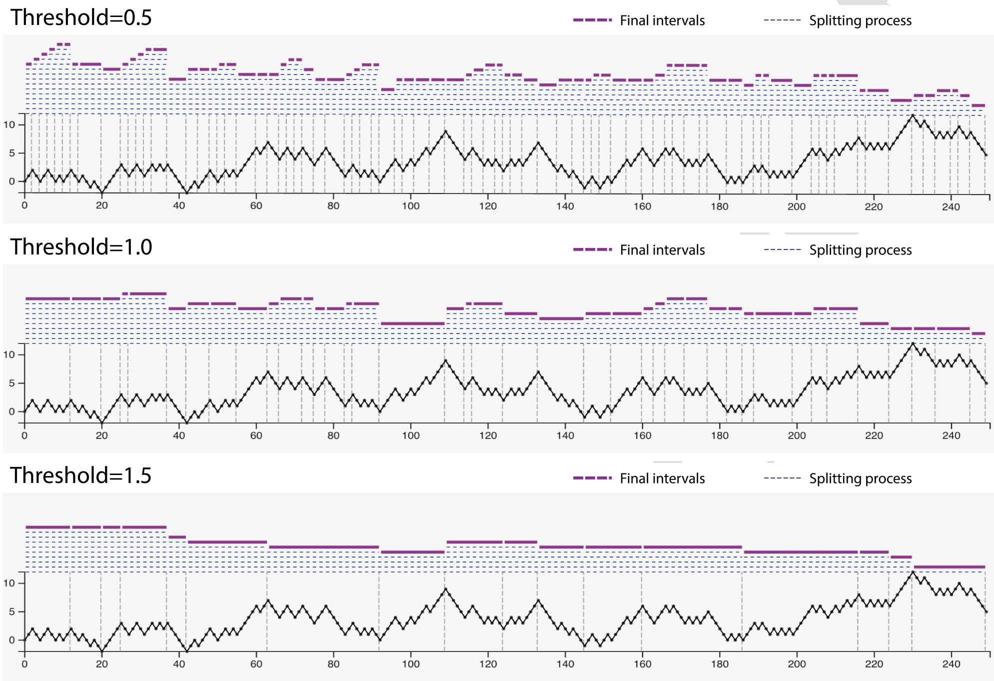


Figure 4: Division of a time series into intervals using different thresholds for the size of the largest triangle (top: threshold=0.5, middle: 1.0, bottom: 1.5). Each view consists of a raw data series (bottom) and the resulting time intervals (top). Blue dotted lines represent the recursions of the splitting process and bold purple lines represent the final intervals while grey dotted lines represent the end of each interval. A larger threshold indicates a greater tolerance for variations within the data, resulting in a coarser segmentation that represents more pronounced trends. Conversely, a smaller threshold refines the segmentation capturing subtler variation.

T2. Deriving complex patterns by computing temporal relations between time intervals in multivariate time series

After identifying the time intervals containing basic patterns in each univariate time series (Figure 5), we proceed to compute the temporal relations between the time intervals across multiple attributes by employing a subset of Allen's interval algebra consisting of the relations `before`, `after`, and `overlap`. We allow a certain margin of overlapping ω to be present in the `before` and `after` relation. Any relation where two intervals share a sufficiently long ($> \omega$) period of simultaneous existence is considered as an instance of the `overlap` relation. The threshold ω is specified as percentage of the duration of the shorter interval.

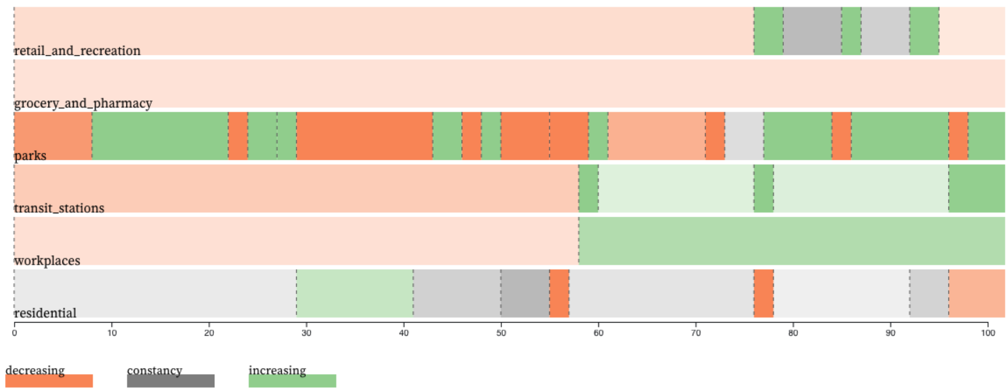


Figure 5: A segment of an univariate time series from the mobility dataset distilled into basic trend patterns. Colors denote the trend directions: orange for `decreasing`, grey for `constancy`, and green for `increasing`. Opacity signifies the change rate, i.e., the amount of change of the value divided by the interval length.

To identify relations, we employ the following algorithm. Firstly, for each interval in one time series that contains a pattern (referred to as the reference interval), the algorithm searches for its temporal neighbors in the other time series. Two intervals are considered temporal neighbors if they either overlap or if the temporal distance between the end of the earlier interval and the start of the later interval does not exceed a predefined threshold, denoted as δ. If the neighboring interval overlaps with the reference interval by more than a specified value of the threshold ω, the relation is labeled as `overlap`. Otherwise, if the neighbor starts earlier, the relation is labeled as `before`, and if the neighbor starts later, the relation is labeled as `after`. Analysts have the flexibility to adjust the parameters δ and ω based on their specific requirements. In the provided examples, we have chosen δ to be 1 day and ω to be 20% of the shorter interval's duration. Figure 6 demonstrates an example of the relation `after` between two patterns.

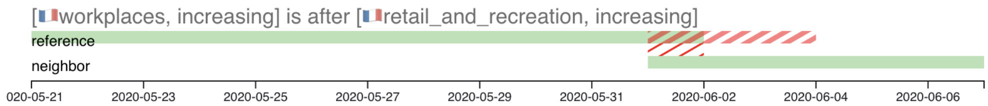


Figure 6: An example of a temporal relation between two trend patterns. The `increasing` pattern in `workplaces` (bottom) is `after` the `increasing pattern` in `retail and recreation` (top). As the two intervals overlap by less than the threshold ω (bold hash pattern in red), the relation is identified as `after` rather than `overlap`.

T3. Exploring occurrence patterns of temporal relations through an interactive visual interface.

In this task, the objective of an analyst is to understand the occurrence patterns of temporal relations by examining their frequency distributions. This task is meant to be performed separately for each type of temporal relation, i.e., `before`, `after`, or `overlap`. We shall call the relation that is currently explored the target relation. This task is structured in accordance with the Visual Information-Seeking Mantra of "overview first, zoom and filter, then details-on-demand" [31].

T3.1 Overview: Matrix visualization of occurrence patterns

To begin our exploration, we first introduce a matrix visualization that aids in understanding the occurrence patterns of temporal relationships. An example is demonstrated in Fig. 7. Each cell in the matrix represents the frequency of the target relation (`before` in Fig. 7) occurring between two patterns across different attributes. In this matrix, rows and columns are divided by dashed lines into three blocks corresponding to the increase, constancy, and decrease patterns. Within these blocks, the rows and columns correspond to the different attributes. The color intensity of a cell indicates the frequency of the target temporal relation: darker shades represent more frequent occurrences, while lighter ones signify fewer.

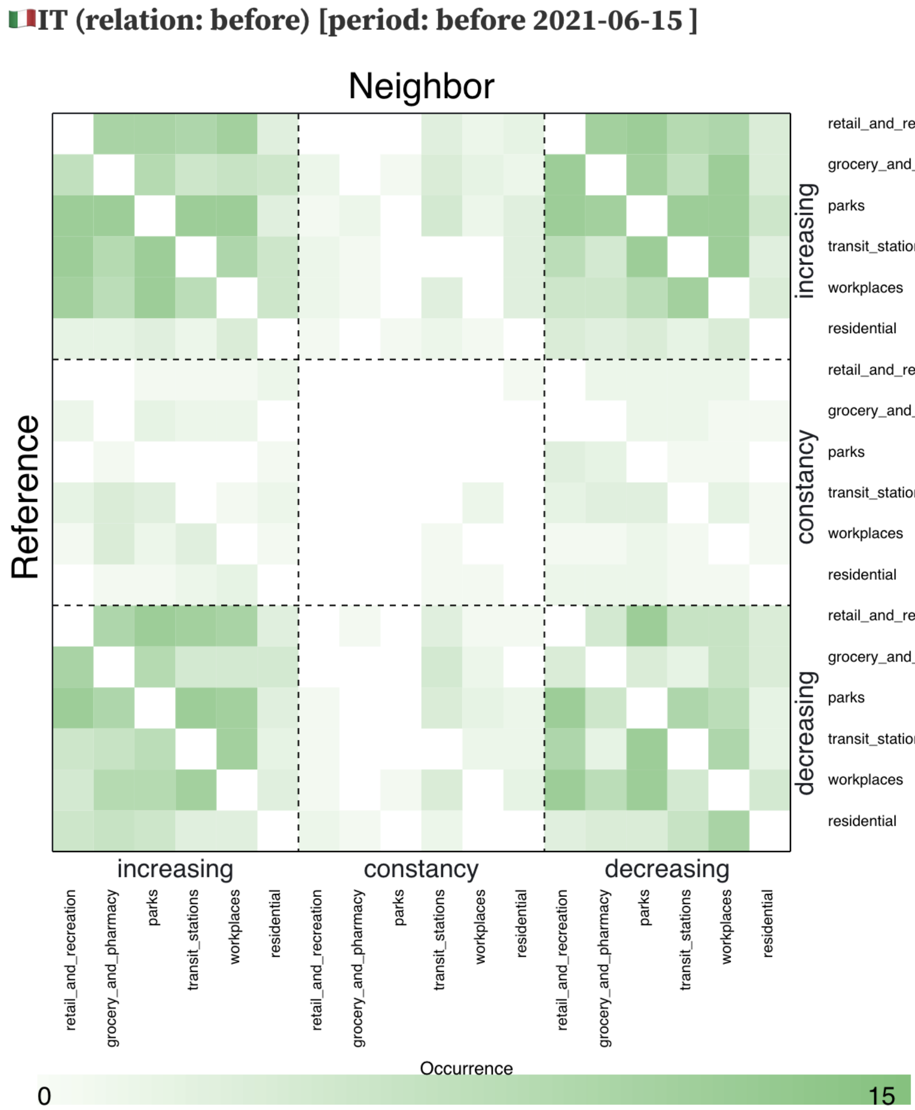


Figure 7: Relation occurrences matrix: Each cell represents the frequency of the target relation (`before` in this example) between two patterns across various attributes. The matrix rows and columns are divided into three blocks corresponding to increase, constancy, and decrease patterns. Darker shades indicate more frequent occurrences, while lighter ones signify fewer.

The investigation is done using a display with multiple matrices (Figure 8), i.e., we apply the "small multiples" technique considered by Edward

Tufte as the best design solution for a wide range of problems [32, p.67]. In accordance with the Jacques Bertin's concept of an image as "the meaningful visual form, perceptible in the minimum instant of vision" [5, p.11], each matrix can be perceived holistically as a single object. This allows for an at-a-glance comparison between matrices, without involving minute details.

The multi-matrix display in Figure 8 visualizes the occurrence patterns for the `before` relation for each country. Initial observation of these matrices reveals certain similarities, such as comparable white crossing lines in the Czech Republic, Hungary, Luxembourg, Slovenia, and Slovakia.

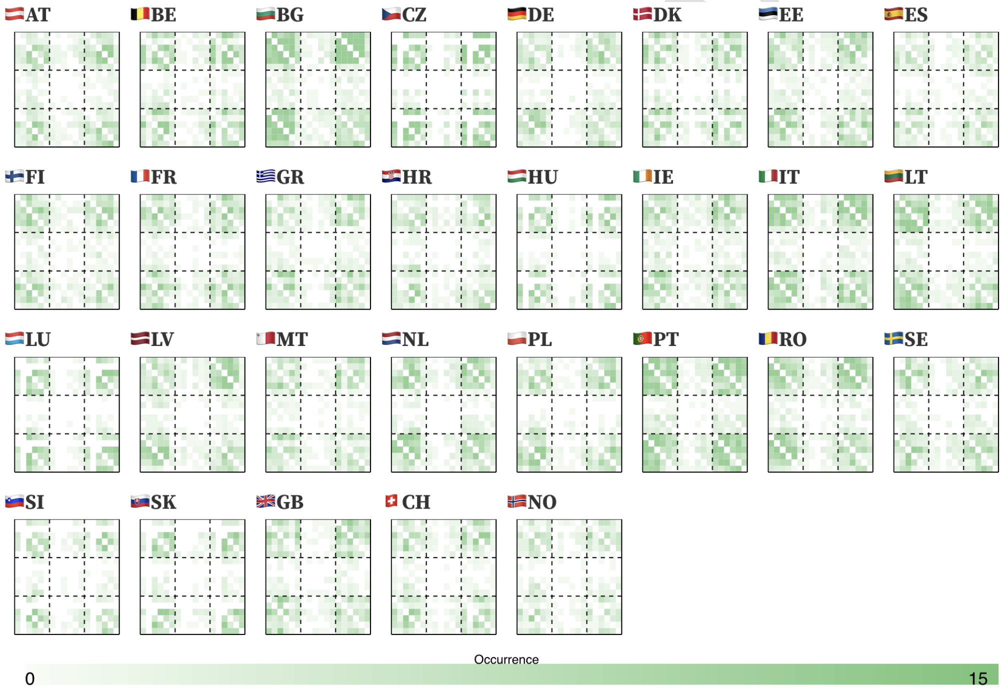


Figure 8: Grid of matrix views for 29 European countries. Each matrix within the grid represents the occurrences of the `before` relation for each pair of patterns within the respective country.

For more effective comparison, the user can set the multi-matrix view to show normalized deviations from the average (Figure 9). Within these matrices, each matrix cell displays the difference between a normalized occurrence value for a given country and the average normalized value for corresponding pair of patterns across all countries. The normalized occurrence value is computed as the ratio of the occurrence count of a specific pattern pair to the total number of occurrences of the target relation within the matrix. We can represent this concept with the following formula:

$$diff = NOV_{ref,neigh,rel,c} - ANOV_{ref,neigh,rel} \quad (1)$$

where ref and $neigh$ mean a pair of a reference interval and its neighbor, rel means a relation, and c means a country.

The Normalized Occurrence Value (or NOV) is calculated by dividing the count of occurrences for each pattern pair by the total occurrences of the target relation in the matrix:

$$NOV_{ref,neigh,rel,c} = \frac{count_{ref,neigh,rel,c}}{totalOccurrences_{rel,c}} \quad (2)$$

The Average Normalized Occurrence Value (or ANOV) is the average of the NOVs across all countries:

$$ANOV_{ref,neigh,rel} = \frac{\sum_c NOV_{ref,neigh,rel,c}}{N} \quad (3)$$

In this formula, N is the total number of countries.

Differences are shown through diverging colors, making similarities in occurrence patterns more readily observable.

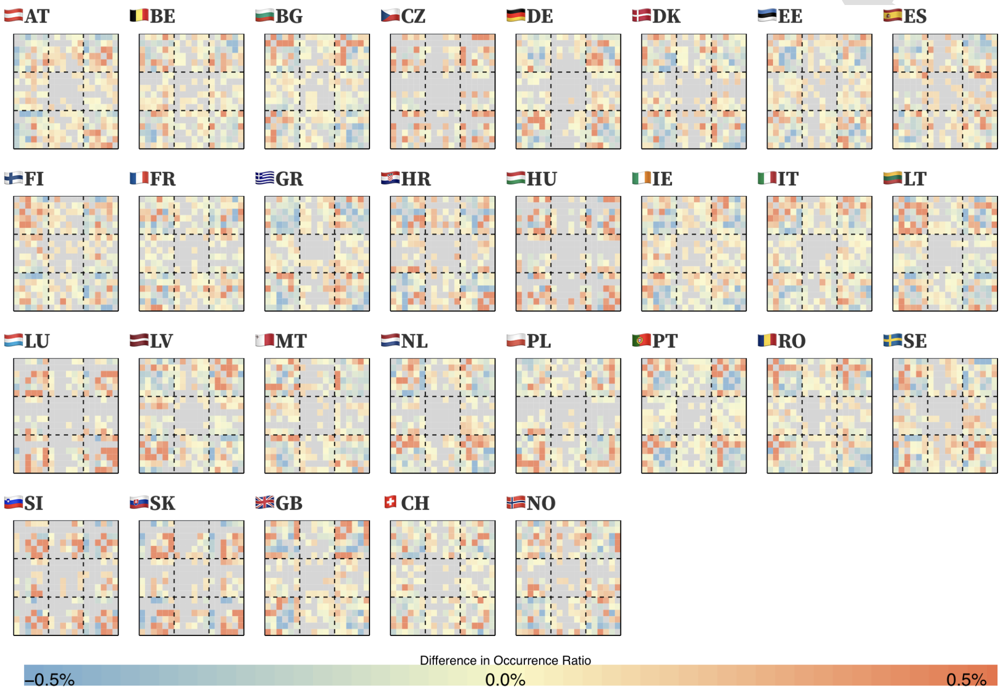


Figure 9: A multi-matrix view of the relations between mobility trends for 29 European countries. Each matrix illustrates the deviation of the normalized frequencies of the relation occurrences from the average of the normalized frequencies for the `before` relation. These differences are expressed using diverging colors.

For an overview of the similarity relationships between countries, analysts can apply dimensionality reduction to the set of matrices. We recommend using a dimensionality reduction algorithm from the class known as neighbour embedding algorithms, which give priority to preserving local neighborhoods at the cost of higher distortion of longer distances between embedded data items. Hence, highly similar items (i.e., neighbors in the multidimensional space) receive close positions in the low-dimensional projection space. One of such algorithms is t-SNE [23], which we use in our example. In Figure 10, t-SNE was utilized with two different values of the parameter `perplexity`, 5 and 25. This parameter approximately defines the number of neighbours to be considered when placing points in the projection space. Since it is not known in advance how many neighbours, in terms of similarity of the relation occurrence distribution, a country may have, it is reasonable to consider projections obtained with smaller and larger values of the perplexity parameter.

In our example, both projections exhibit clusters of countries with similar matrices, for example, the aforementioned five countries (i.e., the Czech Republic, Hungary, Luxembourg, Slovenia, and Slovakia), as well as the Baltic countries (i.e., Lithuania, Latvia, and Estonia). We do not observe significant difference between the results obtained with the two perplexity values.

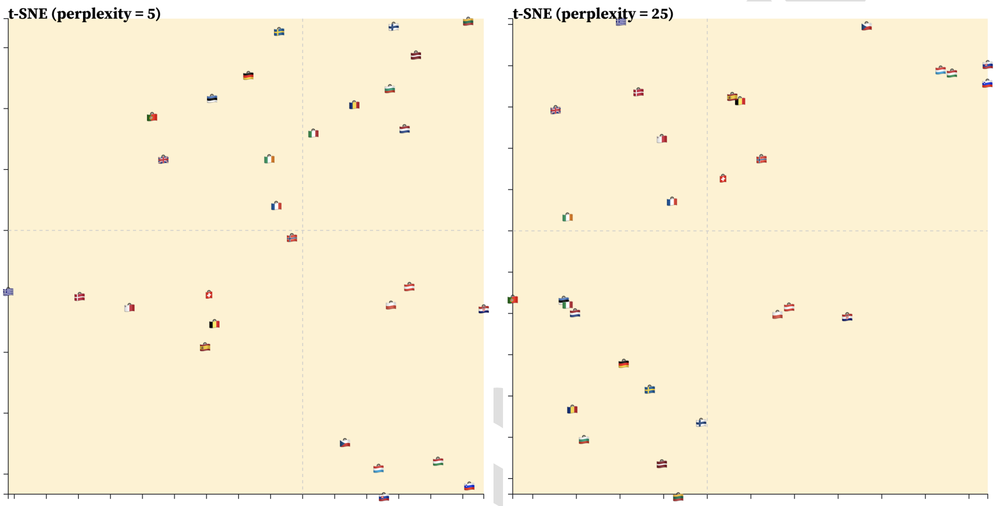


Figure 10: t-SNE projections of the set of countries obtained with perplexity values of 5 (left) and 25 (right), based on the similarity between the matrices of the normalized occurrence frequencies of the `before` relation. The projection provides a spatial representation of the similarities.

T3.2 Zoom, Filter, and details-on-demand

A. Selecting a matrix of interest and a pair of patterns

The subsequent step encompasses zooming and filtering, permitting analysts to narrow down their analysis to a particular matrix of interest (“filter”), which is shown in a larger format (“zoom”) to facilitate focusing on specific patterns and their temporal relations. Analysts can then choose a target temporal relation, and the frequency of the selected relation between each pair of patterns will be shown (“details-on-demand”). In our illustrations, the analyst chooses the `before` relation to explore the frequencies of neighboring patterns that occur prior to reference patterns.

Taking Malta as an example (Figure 11, left), the matrix visualization shows the frequencies of the relation occurrences for different pattern pairs over the period between February 15, 2020 and June 15, 2021 (i.e., in first half of the time span of the available data set, including the beginning of the pandemic). We observe that decreasing patterns of different attributes often precede increasing patterns of other attributes. For instance, we observe a large number of instances where a decreasing trend in `retail and recreation` precedes an upswing in `parks`, `transit stations`, and `workplaces` (Figure 11, left, block 1). Similarly, decreasing patterns in these three categories of places often precede an increase in `retail and recreation`. On the other hand, there are fewer instances of an increasing pattern in `retail and recreation` preceding a decrease in these place categories (Figure 11, left, block 2). Likewise, the matrix for Italy shown on the right of Figure 11 also displays similar patterns but includes another category of place, namely `grocery and pharmacy` (Figure 11, right, block 1).

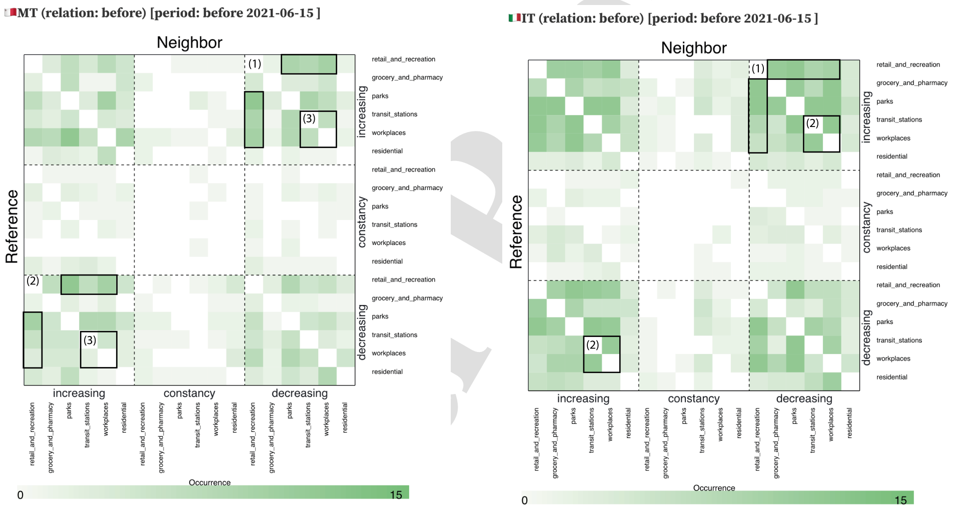


Figure 11: Matrix visualizations of the occurrence frequencies of the temporal relation `before` between trend patterns of different attributes for Malta (left) and Italy (right).

Notably, Italy has higher frequencies of the `before` relation between decreasing trends in `transit stations` and increasing trends in `workplaces` and vice versa, indicating a different pattern of mobility in Malta (Figure 11, left, block 3 and right, block 2) as compared to Italy (Figure 11).

These matrix visualizations serve to observe and compare the overall frequency distribution of temporal relations between different patterns across various contexts or regions, providing initial insights for further detailed analysis.

B. Density chart grid view for temporal relation distributions

To exhibit the distributions of temporal relations between different abstract patterns, we utilize a grid view of density charts. Differently from the matrix view, which shows information for a chosen target relation, the grid view presents information for a chosen reference pattern. For each other pattern, there is a density chart representing the distribution of the relative (with respect to the reference pattern) times of occurrence of this other pattern. The chart includes two density plots, one for the `before` relation (blue) and another for the `after` relation (red). The `overlap` relation is intentionally excluded from this visualization due to its significantly higher frequency, which decreases the visibility of the distributions of the other relations. It is worth noting that our provides the flexibility to exclude any relation, allowing for a more focused examination of the other specific relations. The density charts are arranged in a grid with rows corresponding to the attributes and columns to the different patterns, i.e., increase, constancy, and decrease.

Figure 12 illustrates the grid view. The figure includes two grids representing data from two time periods: before and after the 15th of June, 2021 (midpoint of the available data). The reference pattern is the increase of the attribute `workplaces`. In each grid view, the rows correspond to six attributes (`retail and recreation`, `grocery and pharmacies`, `parks`, `transit station`, `workplaces`, and `residential`) and columns to three trend patterns (`increase`, `constancy`, and `decrease`), resulting in a total of 18 grid cells. The cells in all but one rows contain density charts showing the distributions of the relative times of the occurrences of the neighboring patterns with respect to the reference pattern. The row of the attribute whose pattern is chosen as the reference is empty, because only relations between patterns of distinct attributes are considered in our framework.

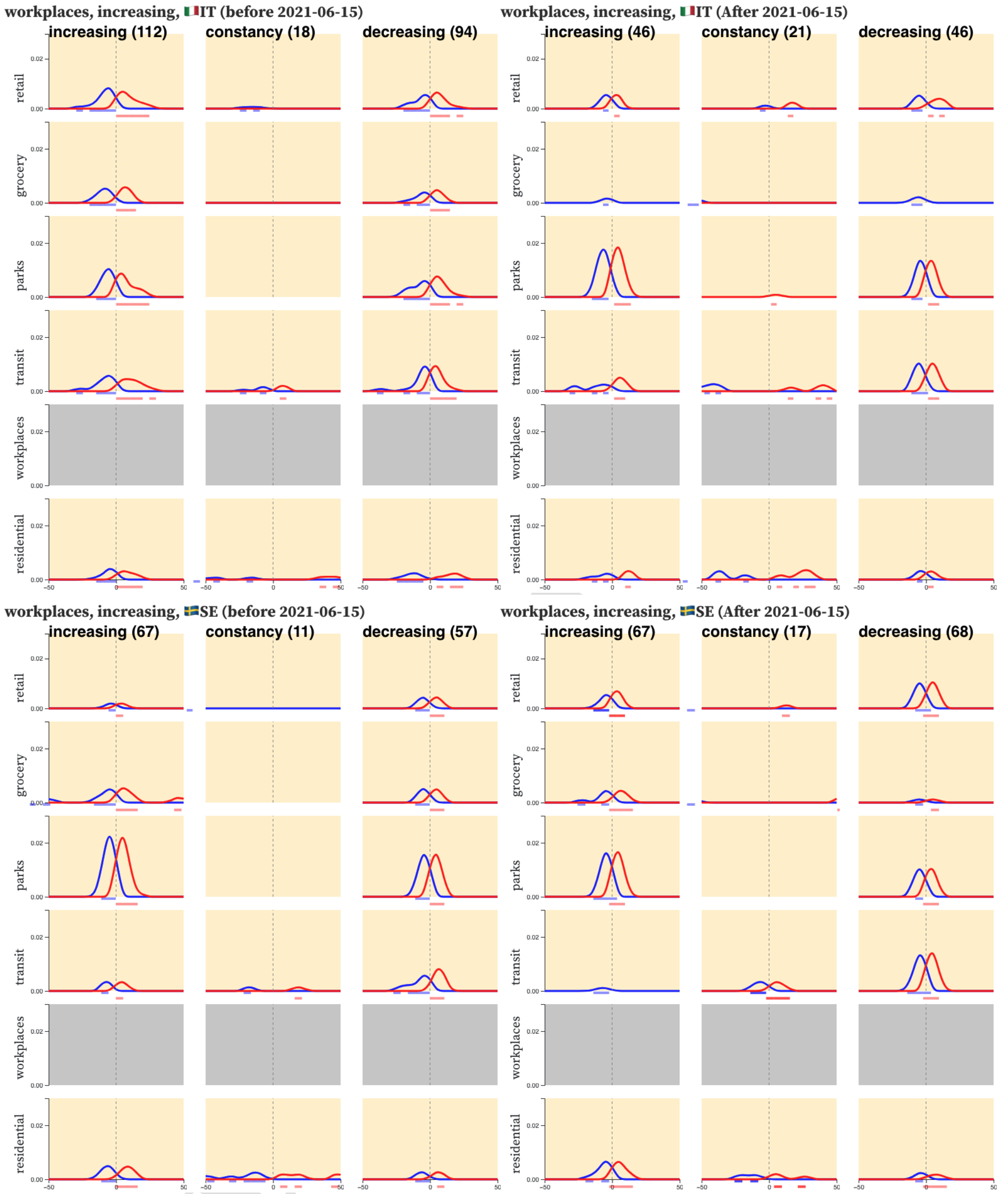


Figure 12: Grid views of density charts for the distributions of the neighbors' relative times for the `increase` pattern of `workplaces` in Italy (top) and Sweden (bottom). The grids on the left include the data from the time period before 15/06/2021, and the grids on the right show the relation distributions after this date. Colors denote different temporal relations: blue for `before` and red for `after`.

Comparison of relations of the `increase` in the visits to `workplaces` between Italy and Sweden

In this case study, we focus on the relations of the `increase` pattern in visiting `workplaces` with different trend patterns of the other attributes occurring in the temporal neighborhood of the reference pattern defined using a temporal threshold of $\delta = 1$ day. We compare the distributions of the neighboring patterns for Italy and Sweden. We segment the data into two subsets based on whether they fall before or after a chosen midpoint date June 15, 2021, for comparative analysis of two time periods.

In Italy (Figure 12, top), more occurrences of the reference pattern in neighborhoods of other patterns are noted before the midpoint date than afterwards. This suggests that the `increase` pattern of `workplaces` was more prominent at the early stages of the pandemic. Given that Italy implemented lockdown measures relatively early [13], this observation aligns with the expectation that people's mobility would have been significantly affected by these measures. In contrast, Sweden (Figure 12, bottom), known for not imposing any form of lockdown [13], exhibits fewer occurrences of the reference pattern in relation to others before the midpoint date than after it.

Upon closer inspection, we find that the same pattern before the midpoint in Italy has more relations with the `decrease` pattern of `transit station` than its counterpart in Sweden, implying different mobility patterns in these countries.

4. Case study: Team behaviours in football

In this section, we present another case study using data from professional football (or soccer) matches. Understanding collective movements is crucial for interpreting tactical behaviors in football. For example, the team that has gained the ball possession tends to extend its width while the team without possession tends to get more compact [8]. Revealing temporal relations between such kinds of trends of different attributes can enhance understanding of the data. For example, an `increase` in `average velocity` (i.e., average speed of players on both teams) `before` an `increase` in `goal distance` (i.e., distance between a team's own goal and the mean position of the outfield players on that team, excluding the goalkeeper) implies a quick attack such as a counter-attack.

Data description

We use continuous time series of the teams' collective movements computed from players' positions. We have data from two matches, labeled as BB and BN, in which the same home team, denoted as D, competes against two distinct opponents, referred to as O. One match can be divided into four subsets of game episodes distinguished by two factors: the stage of the match (either the first half or the second half) and the team possessing the ball, i.e., D or O. Each half contains around 67500 timesteps (i.e., 45 minutes given that the raw data has a sampling rate of 25 Hz). For each team and each time step, we compute `team width` (horizontal distance between the leftmost and rightmost players on a team excluding the goalkeeper), `team depth` (vertical distance between the frontmost and rearmost players on a team excluding the goalkeeper), and `goal distance` (distance between a team's own goal and the mean position of the players on that team excluding the goalkeeper). The attributes of the two teams are distinguished by the prefixes `home` and `away` in the attribute names, for example, `home width` and `away width` As there exists strong correlation between the average velocities of the players of the two teams, we compute `average velocity` on both teams (excluding the goalkeepers).

Comparative analysis of team strategies in first and second halves of the BB match

In defining temporal neighborhoods and identifying relations, we use the threshold values $\delta = 25$ frames (i.e., 1 second) and $\omega = 0.3$; see section 3.2, task T2.

Figure 13 enables a comparative study of relation occurrence patterns under eight situations, i.e., two matches, the first and second halves of each match, and different teams (D and O) in possession of the ball. Similarly to Section 3, each cell within these matrices displays the occurrence frequency of the relation `before` between the corresponding pattern pairs across all attributes.

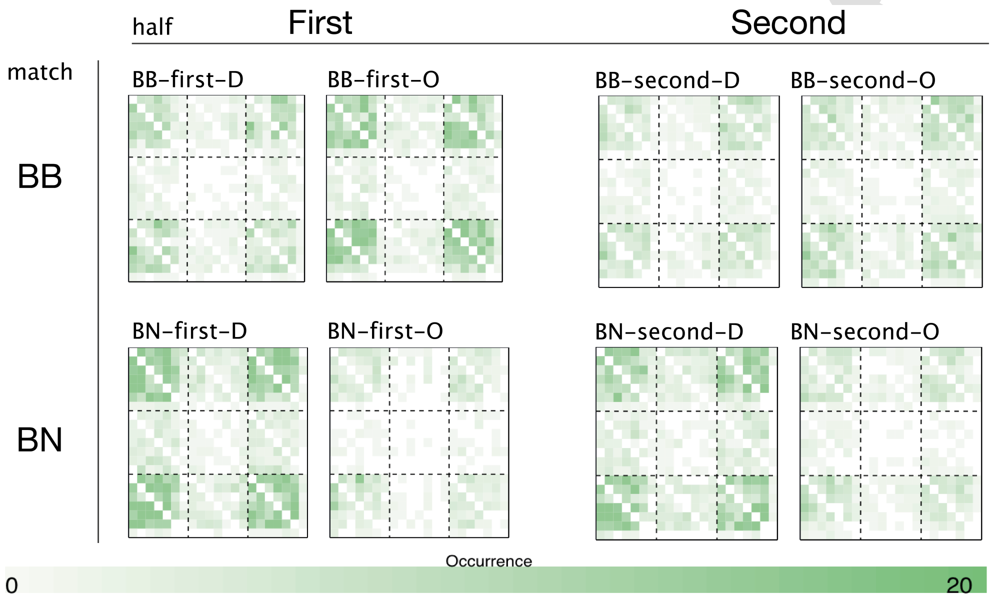


Figure 13: A 2 ×4 grid of matrix views. The matrices in the top row correspond to the first match (BB) and those in the bottom row to the second match (BN). Each matrix within the row represents the occurrences of the `before` relation for each pair of patterns in a distinct subset of game episodes: the first or the second half of a match and the team D or O in possession of the ball.

Upon initial observation, we identify remarkable similarities in the patterns of relation occurrence in the subsets of episodes in both matches when team D is in control of the ball, i.e., the matrices with labels that end with `-D`, such as BB-first-D. These patterns are also similar to the pattern of relations when team O is in possession in match BN, evident in both BN-first-O and BN-second-O. There is higher similarity between the matrices BB-first-O, BN-first-D, and BN-second-D, which indicates that the team D behaved in the match BN similarly to the behavior of their opponents in the first half of the match BB. In the first half of the match BB, team D had lower relation frequencies than team O, whereas in the second half the frequencies were nearly equal.

The apparent differences between the absolute frequencies can be explained by the differences in the total duration of the ball possession between the teams. Therefore, to reveal possible differences in team tactics, it makes sense to transform the absolute frequencies to normalized values, as in Fig. 9.

The normalized deviations from the average, as depicted in Figure 14, enhance the visibility of certain patterns of occurrences, although some similarities with the matrices in Figure 13 still persist. Specifically, a distinct difference in color (blue and yellow) for the top-left vertical line between the BN-second-D and the BN-first-D matrices can be observed. However, this difference is merely expressed by color intensity in Figure 13.

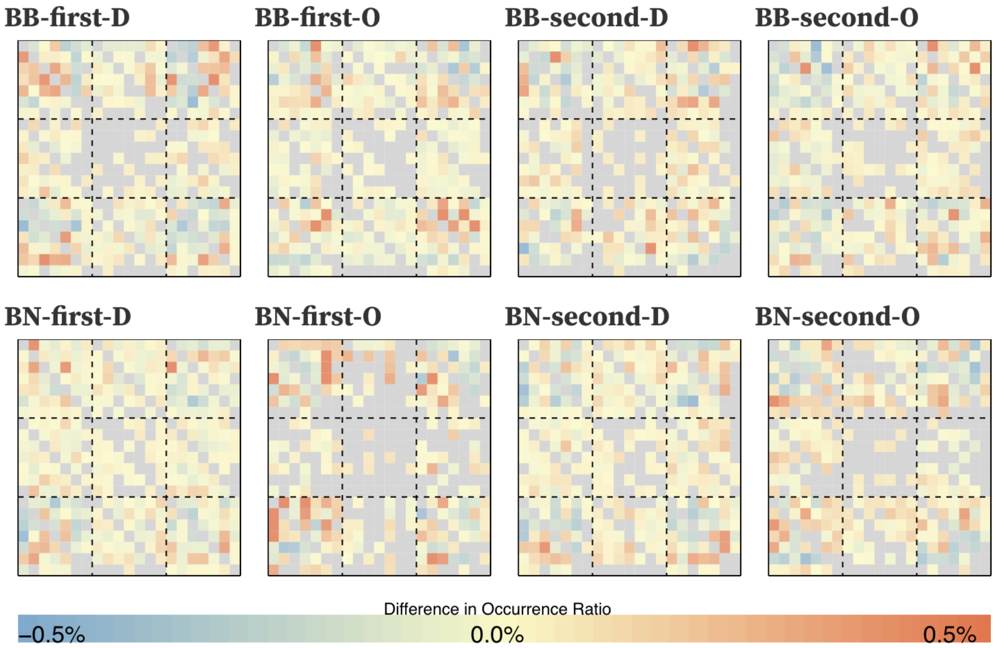


Figure 14: A 2 ×4 grid of matrix views. The matrices in the top row correspond to the first match (BB), while those in the bottom row correspond to the second match (BN). Each matrix within the row illustrates the difference between the average of normalized occurrences of the `before` relation for each pair of patterns and the corresponding normalized value. Diverging colors are used to represent these differences. Each matrix corresponds to a distinct subset of game episodes, either the first or second half of a match, and whether the team D or O in possession of the ball.

To compare two halves of one match, analysts can subtract the normalized occurrence values of the second half from those of the first half to identify which pair of patterns appears more frequently in each half. Figure 15 demonstrates the result of this operation for the ball possession of O in the match BB.

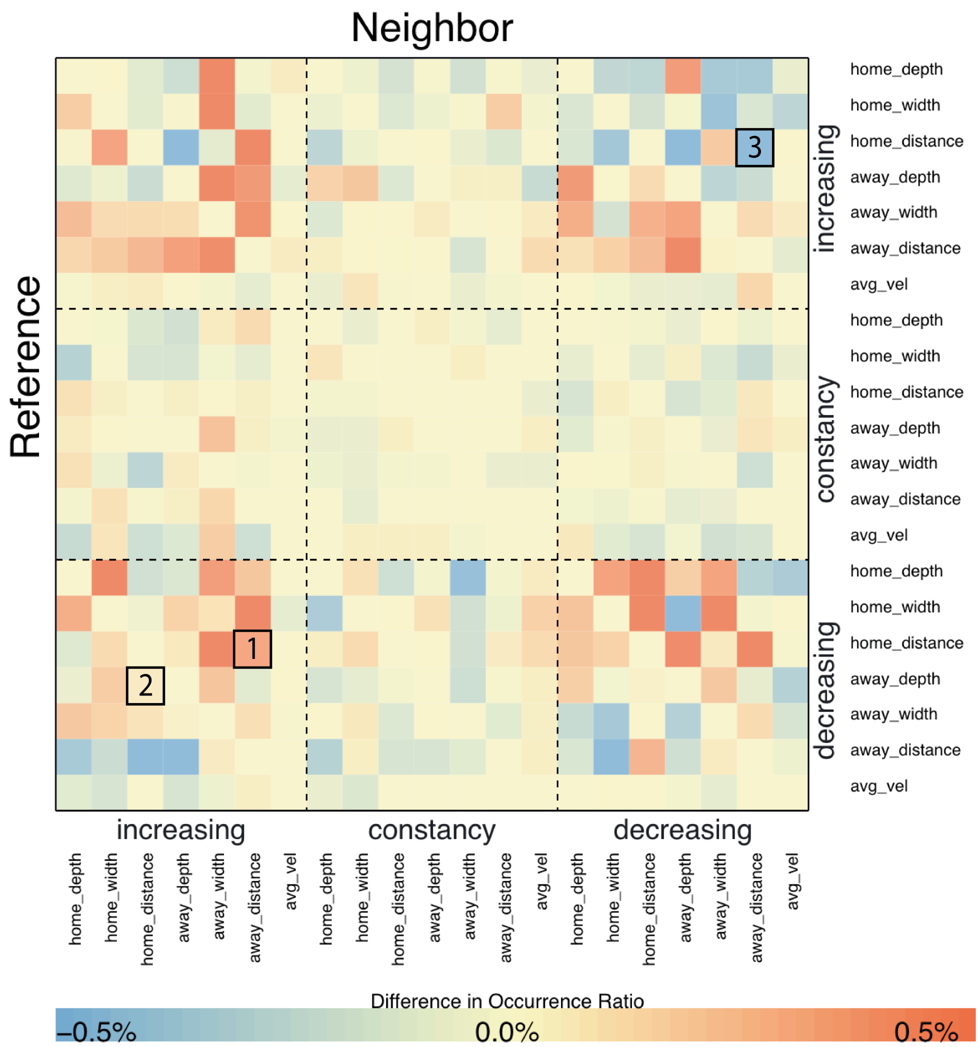


Figure 15: The matrix represents the difference in normalized occurrence values of the neighbor patterns preceding the reference patterns between the first and second halves of match BB when team O possesses the ball. The values are calculated by subtracting the occurrence frequencies in second half from those in the first half. Red indicates a higher frequency of occurrences in the first half while blue signifies more occurrences in the second half.

The `increase` pattern of `away distance` followed by the `decrease` pattern of `home distance` indicates that the away team is moving further away from their own goal (1 in Figure 15), suggesting an offensive strategy, while the home team is moving closer to their goal, suggesting a defensive strategy. This pattern is more noticeable in the first half, concurring with the match report's statement that the away team initiated a strong offensive from the start of the game [15].

Similarly, we observe the `increase` pattern of `home distance` followed by the `decrease` pattern of `away depth` (2 in Figure 15), suggesting that the home team advances before the away team becomes more compact. We also observe that the `decrease` pattern of `away distance` preceding the `increase` pattern of `home distance` is more prevalent in the second half (3 in Figure 15), suggesting that the away team adopts a more defensive posture while the home team defends more aggressively. This suggests that the home team is preparing for an aggressive strategy, potentially anticipating a turnover or looking to exploit any gaps in the away team's formation, while the away team is playing compact. The increased prevalence of this pattern in the second half suggests an offensive shift of strategy by the home team, aligning with the match report that mentioned that the home team exerting considerable pressure on the away team's defense [15].

Density charts: Changes in behaviour between first and second halves

In this section, we use grids of density plots for a more detailed investigation of the temporal relations between patterns of different attributes. To compare two halves of a football match, we juxtapose two grids representing the corresponding data. Each grid comprises seven attributes (`home depth`, `away depth`, `home width`, `away width`, `home distance`, `away distance`, and `average velocity`) and three trend patterns (`increase`, `constancy`, and `decrease`), resulting in a total of 21 grid cells.

First, we focus on the `increase` pattern in `average velocity`. We observe a generally higher number of various neighboring patterns in the first half (Figure 16, top-left) compared to the second half (Figure 16, top-right), which may be a consequence of the higher number of occurrences o the `increase` pattern of `average velocity` during the first half.

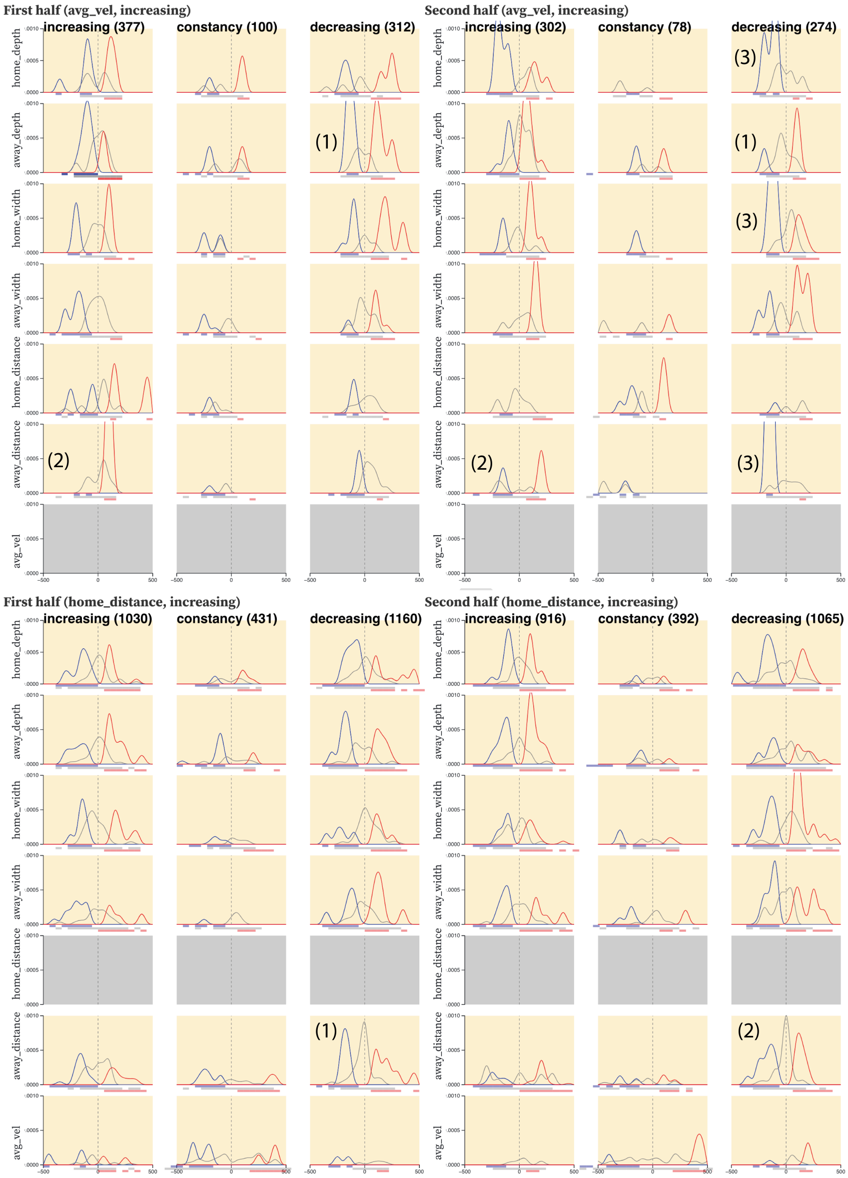


Figure 16: Comparison of two grid views from the first half (left) and second half (right) of a football game, with the reference attribute `average velocity` and pattern `increasing` (top) and with the reference attribute `home distance` and pattern `increasing` (bottom). In each grid, three density plots depict the distributions of the relative times of the neighboring patterns compared to the reference patterns, with blue representing the `before` relation, grey representing the `overlap` relation, and red representing the `after` relation.

This seems logical since players typically experience less fatigue in the first half, enabling them to change speed more frequently. We also observe less obvious differences, such as much higher frequencies of the `decrease` of `away depth` and `increase` of `away width` occurring before the `increase` of textcolorblue`average velocity` in the first half than compared to the second half (1 in Figure 16). The `increase` of `away distance` after the `increase` of `average velocity` is also observed more often in the first half (2 in Figure 16, top). These indicate that the first half included a larger number of active attacks by the away team in which they moved with increasing velocity towards the opponent's goal. To prepare the attack, the team tended to increase the width (i.e., across the pitch) while decreasing the distances between the lines, i.e., the team's depth. The second half had prominently higher frequencies of the `decrease` of `home depth`, `home width`, and `away distance` before the increase in the `average velocity` (3 in Figure 16, top). This indicates that the away team was often retreating to their goal before the increase of the `average velocity` and, at the same time, the home team was getting more compact, which usually happens in preparation to an attack of the opponents.

Next, we examine the relations of the `increase` pattern in `home distance` (Figure 16, bottom). It is apparent that this pattern has numerous relations with the `decrease` pattern of `away distance` as a neighboring pattern, particularly during transitions of ball possession, as observed earlier by Shirato et al. [30]. However, we can verify that in the first half, there are more neighboring patterns of the `decrease` of `away distance` preceding the reference pattern than those succeeding it (1 in Figure 16, bottom). In contrast, during the second half, we observe more of the same neighboring patterns following the reference pattern than those preceding it (2 in Figure 16, bottom). These observations suggest that the away team tends to move more quickly in the first half, while the reverse occurs in the second half. This implies that the home team has greater control over the match in the first half compared to the second half.

5. Discussion

In this study, we developed a framework for analysis of multivariate time series (MVTS) data. The framework includes abstraction of value sequences into instances of basic variation patterns and exploration of temporal relations between these pattern instances across different variables.

Our framework has several strengths. As its core, it offers flexibility and generality to meet a wide range of data analysis needs in MVTS.

Its flexibility emerges from the proposed approach to pattern extraction. The framework does not require to pre-define pattern duration, i.e., the length of the time interval containing a pattern. This provides flexibility for finding pattern instances of variable duration and adapting to data with diverse properties, such as sampling rate, rate of changes, and amplitudes of changes.

On the other hand, the generality of our framework manifests in its ability to handle different types of temporal patterns. While our pilot studies focused on extracting basic trend patterns for their easy interpretability, the framework is not limited to these. It is capable of extending to any other types of temporal patterns depending on the character of studied changes and the analysis goals. For example, there may be pattern types reflecting states, such as high, medium, and low values. Moreover, patterns may be composed of values of categorical attributes.

Our framework can be extended, offering additional or alternative methods for pattern definition and extraction. This accommodates the diverse needs of analysts, who may opt to sketch a pattern, use an interface like Time Searcher [11] to define the pattern, or even define composite patterns built of basic patterns like a peak followed by a trough. The system, in response, identifies patterns similar to the sketch or template provided, opening a door to more customized and insightful analysis.

Another aspect of extensibility comes in the form of temporal relations considered in the analysis. While we used a subset consisting of three relations, before, overlap, and other, other relations from the Allen's algebra of time intervals can also be considered in the analysis.

We have demonstrated the application of our framework in two distinct use cases: exploring mobility patterns during the COVID-19 pandemic and analyzing team behaviors in professional football matches. In both cases, our framework was able to provide valuable insights into the temporal relations between different patterns of attribute variation.

In the COVID-19 mobility data case, our framework was able to identify and visualize temporal relations between different mobility patterns within each country and to investigate the distributions of temporal relations across countries. This analysis revealed interesting patterns, such as the increase in workplace mobility preceding an increase in residential mobility, likely due to people preparing for remote work during lockdowns. Moreover, the framework was able to highlight differences in mobility patterns between countries, reflecting the varying policies against the pandemic.

In the football data case, our framework was able to identify and visualize temporal relations between different team behaviors. For example, it was able to detect a change in team behavior between the first and second halves of a match, which aligned with the match report.

Despite its strengths, our framework also has some limitations. It requires a number of thresholds and parameters to adjust (e.g., ω, δ, and perplexity), which could potentially confuse analysts and require sensitivity analysis. In terms of trend patterns, the results may not be expressive enough as it does not consider a combination of univariate temporal patterns such as peaks and troughs. For temporal relations, our framework does not provide a holistic understanding of pairwise relations, as it only calculates the distribution. The importance of relations is thus expressed only through frequency.

A major limitation is that the framework does not scale well to the numbers of attributes, pattern types, and types of relations. A possible approach to alleviate this is to develop a guiding system that suggests potentially interesting selections to explore. During the analysis, the analyst can interactively construct a knowledge graph or several graphs for different conditions or classes of situations, such as fast attack or gradual approach in football. A knowledge graph contains pattern types of different abstraction levels and relations between them, including temporal and hierarchical.

We also foresee several potential improvements and directions for future work to enhance our framework. For trend patterns, one possible direction is to construct combined univariate patterns from basic patterns, thus increasing expressiveness of the results. For temporal relations, future work could incorporate approaches capable of dealing with multiple pairwise relations, such as a network-based approach where nodes represent temporal patterns and edges represent relations between them. This could provide a more holistic understanding of the relations and allow for the identification of important nodes using graph centrality measures.

An important consideration is the accessibility and user-friendliness of any tools developed to support this framework. Our primary objective was the development and validation of the framework itself, while the prototype tools we implemented served mainly as proof of concept. These tools were not optimized for end-user adoption and would require further user-centered design for broad accessibility. Conceptually, our framework is simple enough to understand to understand and adopt without specialized technical skills.

However, the practical implementation of the framework for end-users would necessitate domain-specific enhancements to facilitate its use. Our proof-of-concept implementation has demonstrated the types of analyses possible with the framework, but further development is needed to make these operations user-friendly in specific domains of application.

6. Conclusion

This paper presented a framework that unifies various methods for the abstraction of multivariate time series (MVTS) data. The unification is achieved by integrating these different methods into a *cohesive workflow*, which allows to understand dynamic phenomena through the lens of temporal relations, the identification of basic behavior patterns and the examination of temporal relations among these patterns. Our framework is designed to identify basic behavior patterns and examine the temporal relations among these patterns, taking into account temporal lags and varying duration of the patterns. This feature enhances the understanding of complex interactions among multiple attributes, making the framework valuable for analysts.

The effectiveness and versatility of our framework were demonstrated through its application to mobility data during the COVID-19 pandemic and football (soccer) data. Despite its strengths, the framework has some limitations, such as the need for further enrichment to handle intricate variable interactions and the integration of more complex patterns. These limitations provide avenues for future work.

In conclusion, our framework offers an approach to abstracting MVTS, with a focus on understanding temporal relations. By integrating various methods into a single workflow, it enables analysts to effectively explore and comprehend complex temporal relations in MVTS data.

7. Acknowledgements

This work was partly supported by Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the *Lamarr Institute for Machine Learning and Artificial Intelligence* (Lamarr22B), and by EU in projects *SoBigData++* and *CrexData* (grant agreement 101092749).

References

[1] Allen, J.F., 1983. Maintaining knowledge about temporal intervals. Communications of the ACM 26, 832–843. doi:10.1145/182.358434.

[2] Andrienko, N., Andrienko, G., 2023. It's about time: Analytical time periodization. Computer Graphics Forum doi:10.1111/cgf.14845.

[3] Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., Wrobel, S., 2020. Visual analytics for data scientists. Springer.

[4] Andrienko, N., Andrienko, G., Miksch, S., Schumann, H., Wrobel, S., 2021. A theoretical model for pattern discovery in visual analytics. Visual Informatics 5, 23–42. doi:10.1016/j.visinf.2020.12.002.

[5] Bertin, J., 1983. Semiology of graphics. University of Wisconsin Press.

[6] Combi, C., Oliboni, B., 2012. Visually defining and querying consistent multi-granular clinical temporal abstractions. Artificial intelligence in medicine 54, 75–101. doi:10.1016/j.artmed.2011.10.004.

[7] Deng, Z., Weng, D., Xie, X., Bao, J., Zheng, Y., Xu, M., Chen, W., Wu, Y., 2022. Compass: Towards better causal analysis of urban time series. IEEE transactions on visualization and computer graphics 28, 1051–1061. doi:10.1109/TVCG.2021.3114875.

[8] Fonseca, S., Milho, J., Travassos, B., Araújo, D., 2012. Spatial dynamics of team sports exposed by voronoi diagrams. Human movement science 31, 1652–1659. doi:10.1016/j.humov.2012.04.006.

[9] Google, 2020. COVID-19 community mobility reports. https://www.google.com/covid19/mobility/. Accessed: 2022-4-24.

[10] Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society 37, 424–438. doi:10.2307/1912791.

[11] Hochheiser, H., Shneiderman, B., 2001. Visual queries for finding patterns in time series data. University of Maryland, Computer Science Dept. Tech Report, CS-TR-4365 .

[12] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[13] Islind, A., Óskarsdóttir, M., Steingrímsdóttir, H., 2020. Changes in mobility patterns in europe during the COVID-19 pandemic: Novel insights using open source data. arXiv.org doi:10.48550/arXiv.2008.10505.

[14] Keogh, E., Chu, S., Hart, D., Pazzani, M., 2004. Segmenting time series: a survey and novel approach, in: Data Mining in Time Series Databases. WORLD SCIENTIFIC. volume 57 of *Series in Machine Perception and Artificial Intelligence*, pp. 1–21. doi:10.1142/9789812565402_0001.

[15] kicker, 2018. Furioser BVB! nach Robert-Lewandowski-Doppelpack drehen reus & co. auf - borussia dortmund baut vorsprung auf bayern münchen auf insgesamt sieben punkten aus. https://www.kicker.de/dortmund-gegen-bayern-2018-bundesliga-4243356/analyse. Accessed: 2023-6-20.

[16] Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Review 51, 455–500. doi:10.1137/07070111X.

[17] Köthur, P., Witt, C., Sips, M., Marwan, N., Schinkel, S., Dransch, D., 2015. Visual analytics for correlation-based comparison of time series ensembles. Computer graphics forum: journal of the European Association for Computer Graphics 34, 411–420. doi:10.1111/cgf.12653.

[18] Lee, T.Y., Shen, H.W., 2009. Visualization and exploration of temporal trend relationships in multivariate time-varying data. IEEE transactions on visualization and computer graphics 15, 1359–1366. doi:10.1109/TVCG.2009.200.

[19] Li, J., Chen, S., Zhang, K., Andrienko, G., Andrienko, N., 2019. Cope: Interactive exploration of co-occurrence patterns in spatial time series. IEEE Transactions on Visualization and Computer Graphics 25, 2554–2567. doi:10.1109/TVCG.2018.2851227.

[20] Lin, J., Keogh, E., Wei, L., Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. Data mining and knowledge discovery 15, 107–144. doi:10.1007/s10618-007-0064-z.

[21] Liu, S., Weng, D., Tian, Y., Deng, Z., Xu, H., Zhu, X., Yin, H., Zhan, X., Wu, Y., 2023. ECoalVis: Visual analysis of control strategies in coal-fired power plants. IEEE transactions on visualization and computer graphics 29, 1091–1101. doi:10.1109/TVCG.2022.3209430.

[22] Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer Berlin Heidelberg. doi:10.1007/978-3-662-02691-5.

[23] van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of Machine Learning Research 9, 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[24] Mannila, H., Toivonen, H., Inkeri Verkamo, A., 1997. Discovery of frequent episodes in event sequences. Data mining and knowledge discovery 1, 259–289. doi:10.1023/A:1009748302351.

[25] Monroe, M., Lan, R., Lee, H., Plaisant, C., Shneiderman, B., 2013. Temporal event sequence simplification. IEEE transactions on visualization and computer graphics 19, 2227–2236. doi:10.1109/TVCG.2013.200.

[26] Murphy, K.P., 2002. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. thesis. University of California.

[27] Patel, P., Keogh, E., Lin, J., Lonardi, S., 2002. Mining motifs in massive time series databases, in: 2002 IEEE International Conference on Data Mining, 2002. Proceedings., pp. 370–377. doi:10.1109/ICDM.2002.1183925.

[28] Sacchi, L., Larizza, C., Combi, C., Bellazzi, R., 2007. Data mining with temporal abstractions: learning rules from time series. Data mining and knowledge discovery 15, 217–247. doi:10.1007/s10618-007-0077-7.

[29] Shahar, Y., 1997. A framework for knowledge-based temporal abstraction. Artificial intelligence 90, 79–133. doi:10.1016/S0004-3702(96)00025-2.

[30] Shirato, G., Andrienko, N., Andrienko, G., 2023. Identifying, exploring, and interpreting time series shapes in multivariate time intervals. Visual informatics 7, 77–91. doi:10.1016/j.visinf.2023.01.001.

[31] Shneiderman, B., 2002. The eyes have it: a task by data type taxonomy for information visualizations, in: Proceedings 1996 IEEE Symposium on Visual Languages, IEEE Computer Society Press. pp. 336–343. doi:10.1109/vl.1996.545307.

[32] Tufte, E.R., 1990. Envisioning Information. Graphics Press, Cheshire, CT.

[33] Walker, J., Borgo, R., Jones, M.W., 2016. TimeNotes: A study on effective chart visualization and interaction techniques for Time-Series data. IEEE transactions on visualization and computer graphics 22, 549–558. doi:10.1109/TVCG.2015.2467751.

[34] Wongsuphasawat, K., Gotz, D., 2011. Outflow : Visualizing patient flow by symptoms and outcome. IEEE VisWeek Workshop on Visual Analytics in Healthcare , 25–28.

[35] Yi, J.S., Elmqvist, N., Lee, S., 2010. TimeMatrix: Analyzing temporal social networks using interactive Matrix-Based visualizations. International journal of human-computer interaction 26, 1031–1051. doi:10.1080/10447318.2010.516722.

[36] Zhao, J., Chevalier, F., Pietriga, E., Balakrishnan, R., 2011. Exploratory analysis of time-series with ChronoLenses. IEEE transactions on visualization and computer graphics 17, 2422–2431. doi:10.1109/TVCG.2011.195.

Author Information

- Title
  - Exploring and visualizing temporal relations in multivariate time series

- Author names and affiliations
  - Gota Shirato
    - Fraunhofer IAIS
      - Schloss Birlinghoven, 1, 53757 Sankt Augustin, Germany
    - University of Bonn
      - Regina-Pacis-Weg 3, 53113 Bonn, Germany
    - gshirato@uni-bonn.de
  - Natalia Andrienko
    - Fraunhofer IAIS
      - Schloss Birlinghoven, 1, 53757 Sankt Augustin, Germany
    - City, University of London
      - Northampton Square, London EC1V 0HB, UK
    - natalia.andrienko@iais.fraunhofer.de
  - Gennady Andrienko
    - Fraunhofer IAIS
      - Schloss Birlinghoven, 1, 53757 Sankt Augustin, Germany
    - City, University of London
      - Northampton Square, London EC1V 0HB, UK
    - gennady.andrienko@iais.fraunhofer.de
- Corresponding author: Gota Shirato

Ethical Approval

This study does not contain any studies with human or animal subjects performed by any of the authors.

Author Contributions

Gota Shirato: Conceptualization, Methodology, Software, Validation, Investigation, Writing - Original Draft, Visualization
Natalia Andrienko: Conceptualization, Methodology, Writing - Review & Editing, Supervision
Gennady Andrienko: Conceptualization, Methodology, Writing - Review & Editing, Supervision

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: