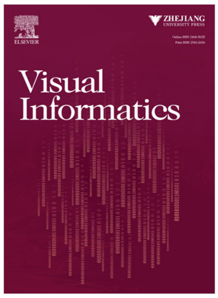
Journal Pre-proof



Identifying, exploring, and interpreting time series shapes in multivariate time intervals

Gota Shirato, Natalia Andrienko, Gennady Andrienko

PII: S2468-502X(23)00001-3
DOI: https://doi.org/10.1016/j.visinf.2023.01.001
Reference: VISINF 160

To appear in: *Visual Informatics*

Received date : 15 November 2022
Revised date : 25 December 2022
Accepted date : 4 January 2023

Please cite this article as: G. Shirato, N. Andrienko and G. Andrienko, Identifying, exploring, and interpreting time series shapes in multivariate time intervals. *Visual Informatics* (2023), doi: https://doi.org/10.1016/j.visinf.2023.01.001.

Identifying, exploring, and interpreting time series shapes in multivariate time intervals

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

We introduce a concept of episode referring to a time interval in the development of a dynamic phenomenon that is characterized by multiple time-variant attributes. A data structure representing a single episode is a multivariate time series. To analyse collections of episodes, we propose an approach that is based on recognition of particular patterns in the temporal variation of the variables within episodes. Each episode is thus represented by a combination of patterns. Using this representation, we apply visual analytics techniques to fulfil a set of analysis tasks, such as investigation of the temporal distribution of the patterns, frequencies of transitions between the patterns in episode sequences, and co-occurrences of patterns of different variables within same episodes. We demonstrate our approach on two examples using real-world data, namely, dynamics of human mobility indicators during the COVID-19 pandemic and characteristics of football team movements during episodes of ball turnover.

Keywords: temporal patterns, multivariate time series, time intervals

1. Introduction

Everything that happens in the world can be conceptualized as a sequence of episodes representing various events or developments of dynamic phenom-

*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)

ena. The term ‘episode’ means (in the context of our research) a time interval during which something happens or develops. The happening or development can be characterized by multiple time-variant attributes, or features. A data structure containing values of multiple features attained at consecutive time units is called multivariate time series. Our research presented in this paper deals with collections of episodes described by multivariate time series where all features are numeric, i.e., represent measurements rather than categories.

A chronologically ordered sequence of values of a single numeric attribute forms a certain pattern [6]. When such a sequence is represented by a polygonal line along a time axis, the pattern is visually perceived as a certain geometric shape. There are shapes, i.e., patterns, that are not only readily detectable by a human eye but also readily interpretable; moreover, their meanings are denoted by specific terms, such as ‘increase’, ‘decrease’, ‘peak’, etc. Temporal variation of a single feature within an interval can thus be described as one of these simple patterns or a sequence of several simple patterns. Obviously, this can be done for each individual attribute of a multivariate time series. However, the resulting description does not provide immediate holistic understanding of the joint behaviour of the attributes.

The research problem we want to solve is how to proceed from recognition of temporal development patterns of individual attributes to identifying and understanding patterns of their joint development in a set of episodes. To investigate this problem, formulate specific analysis tasks, find approaches to fulfil these tasks, and test the efficacy of these approaches, we use two real-world example datasets: mobility data upon the COVID-19 pandemic and collective movement in football games.

Our research presented in this paper aims to support the following analysis tasks.

- T1: Identify major **temporal patterns** in the variation of **individual features** within the episodes.
- T2: Study the **temporal distribution** of the identified univariate temporal patterns.
- T3: Investigate the **transitions** between univariate temporal patterns in consecutive episodes.
- T4: Investigate the **co-occurrence** of univariate temporal patterns of different features within episodes.

The tasks were defined based on the theoretical model for pattern discovery [6], which is further referred to as "the pattern theory". We do not strive to cover all possible tasks in analysing time series but consider the tasks relevant to the analysis process in which higher-level patterns are constructed from lower-level patterns. In this process, task T1 extracts lower-level patterns and tasks T2-T4 aim to discover different types of higher-level patterns formed by the lower-level patterns.

For T1, we introduce an algorithm to extract temporal patterns from univariate time series. T2 is supported by a timeline display and, when appropriate, by circular charts with the circumference representing a temporal cycle. The latter can facilitate detection of periodic re-occurrence patterns in the temporal distribution. For T3, we propose bipartite graphs showing frequencies of pattern transitions. T4 can be fulfilled by interacting with a co-occurrence network.

The rest of this paper is structured as follows. Section 2 discusses the related work. Section 3 introduces the proposed techniques and approaches using the example of the mobility data during the COVID-19 pandemic. Section 4 demonstrates the generality of our approach 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

We introduce previous approaches to pattern detection, interpretation and visualization applicable to multivariate time series in episodes.

2.1. Conceptual foundations

Collins et al. [12] define a pattern as a holistic representation of multiple (data) items abstracted from the individual items. The concept of a data pattern and the existing definitions in different research disciplines have been extensively discussed by Andrienko et al. [6], who argued that patterns are formed by relationships between data items. A data pattern involves elements of at least two sets, for example, time units and values of a numeric attribute. A pattern is made by intrinsic relationships between the elements within these sets and the correspondences between the elements from the different sets. The former depend on the nature of the sets and the latter are defined in the data. The intrinsic relationships between elements of one of the sets create a particular arrangement of the corresponding elements of the other set. A pattern is the manner in which the elements of the second set relate one to another throughout this arrangement.

For example, the intrinsic relationships between time units are temporal ordering and temporal distance, i.e., the amount of time that passed between two units. The intrinsic relationships of ordering and distance (i.e., difference) exist also between values of a numeric attribute. A data set specifies what attribute values corresponds to which time units. The intrinsic temporal relationships between the time units create a temporal arrangement, i.e., a sequence, of the corresponding attribute values. A pattern is the manner in which the values differ one from another along this sequence: whether values that are further in the sequence are greater or smaller than the preceding values or nearly equal to them. Depending on these relationships, we identify the pattern as increase, decrease, or constancy.

The definition of a data pattern as a system of relationships implies that visual discovery of data patterns can be enabled by visualizations satisfying two requirements: (1) appropriately represent the pattern-forming relationships according to the types of data components and (2) facilitate holistic perception of multiple data items. Thus, in a case of a numeric time series, a line chart (a.k.a. time plot) is a suitable visualization: two axes appropriately represent the ordering and distance relationships between time units and between numeric attribute values, the positions of points in this coordinate system accurately represent the correspondences between the time units and the attribute values, and holistic perception is facilitated by connecting the points by lines. A temporal pattern is thus perceived as a particular shape of the resulting polygonal chain. Hence, discovery of temporal patterns can be done by identifying shapes.

According to the pattern theory [6], data patterns that have been discovered can be treated as new elements of data to which the subsequent analysis steps are applied. The analysis involves determining relationships between the patterns throughout arrangements created by elements of other data components, e.g., how the patterns vary along time or how they are distributed over space.

2.2. Temporal pattern extraction and classification

A comprehensive overview of visual analytics approaches for temporal data can be found in monographs by Aigner et al [1], Andrienko et al [5] and

Tominski and Schumann [41].

An important pre-requisite for pattern extraction is segmentation of multivariate time series into semantically meaningful episodes. Papers by Bernard et al [9] and Gharhabi et al [16] propose visual interactive and computational approaches to segmentation. Further works propose semantic segmentation based on TimeMask [4] and its extensions [3].

There exist two major building blocks for temporal pattern extraction and classification. First, there exist methods that search for patterns specified by their shapes. Second, similarity measures (also called distance functions) are used for quantifying similarity and detecting patterns in time series.

Several papers proposed libraries of temporal patterns for univariate [31] or multivariate [44] time series. Das et al. apply a data-driven approach for identifying patterns with interpretable and recognizable shapes [14]. Algorithms for measuring similarity to pre-defined patterns were proposed for detecting time series that contain the given patterns [30] and, in contrast, for detecting dissimilar subsequences in time series [26]. Other approaches to pattern detection and analysis include representations of time series as aggregates [25] or as sequences of symbols [28].

In our work, patterns are identified by means of a new algorithm that calculates the largest triangle within a time series for determining the pattern shape. The idea of the algorithm originates from Steinarsson, who aimed at downsampling time series for visual representation [39]. Unlike the most common approaches, which are based on computing similarities to earlier defined shapes, either taken from a library or sketched by a user, our approach takes into account geometric characteristics of a time series fragment and provides a useful opportunity to represent the patterns in a highly schematic and compact manner using two or three points.

Apart from the research on extraction of predefined patterns and on recognition of pattern types, there are also works where time series patterns are identified implicitly by means of clustering assuming that each cluster defined a certain pattern. The main idea has been exemplified by val Wijk and van Selow [42], who clustered daily univariate time series and investigated the distribution of the clusters over a year. Schreck et al. [37] proposed to treat time series of two variables as trajectories in 2D space. Long time series were divided into episodes, the trajectories from the episodes were clustered, and the original time series were represented as sequences of the averaged trajectory shapes generated for the clusters. From the perspective of our work, these approaches are interesting for their focus on exploring the distribution of temporal patterns rather than solely on pattern detection and extraction.

2.3. Visualization of multivariate time series and episodes

An obvious approach to visualization of multivariate time series is to create multiple visualizations representing the time series of the individual variables. For example, Janetzko et al. [24] create multiple horizon graphs [22] to visualize multiple time series characterizing episodes of a football match. Hao et al.[20] focus on showing the occurrences of earlier detected frequent patterns (motifs) in long time series represented by line graphs. Pham et al. [33] complement multiple area charts showing variation of singular variables with a temporally ordered sequence of radar charts showing combinations of values of the variables. Other authors strive to create a compact view, such as Kaleidomaps [8], where each time series is represented by a heat map embedded in a sector of a circle. A popular technique utilized in visual exploration of multivariate time series data is applying dimensionality reduction to the combinations of attribute values corresponding to the time steps [10, 40]. In these works, the authors are dealing with continuous time series rather than episodes.

In visualizing episodes characterized by multivariate time series, it is necessary to address:

1. **When** the episodes happened: representing their temporal references in linear [13] or cyclic time [32, 11] or structural (calendar) models [42];
2. **What** happens within each episode: temporal dynamics of attributes, usually represented either by displaying time lines [13] or animating representations such as scatter plots [35, 36, 21]. Zhao et al. proposed an interactive visualisation for episodes which facilitated comparison of timelines with different attributes [45];
3. **How** multiple episodes relate to each other: what are transitions between the episodes. This can be represented, for example, by a node-link diagram with nodes representing patterns and links - transitions between them [29].

In analysing the times of the episode occurrences (when), not only the temporal distribution of the episodes is of interest but also the temporal relationships between episodes. Allen and Ferguson systematically introduce all possible pairwise relations between time intervals [2]. These relationships can be represented graphically using triangular logic introduced by Van de

Weghe [43]. Qiang et al. [34] used this approach for representing temporal relationships between episodes. Lee and Shen [27] propose techniques for visual exploration of temporal relationships between occurrences of user-defined patterns (called "trends" by the authors) in multivariate time series. They transform the time series into a sequence of states characterized by different combinations of trends and propose a visual representation in the form of a matrix with columns corresponding to the states and rows to the trends of the different variables.

Our paper uses several visual representations that combine ideas from the mentioned earlier works. Specifically, the idea of our timeline view (Fig. 1) is similar to the visualization of state sequences by Lee and Shen [27], the circular charts (Fig. 6) utilize the idea of Ringmaps [46], and the representation of temporal patterns by colours in various displays follows the ideas of van Wijk and van Selow [42].

3. Visual analytics approach

In this section we introduce our visual analytics (VA) approach that helps analysts to explore and understand large sets of episodes characterised by multivariate numeric time series.

3.1. Essence of the approach

The key idea of our approach is to abstract each individual time series within each episode to a temporal pattern. All patterns are assigned to a finite (preferably small) set of classes, or pattern types, which can be denoted by semantically meaningful labels or somehow encoded in a symbolic form. Hence, each individual time series is represented by a reference to the corresponding pattern class, and each episode is represented by a combination of pattern classes of the multiple attributes. The following analysis is done using this representation of the episodes. For the sake of brevity, we shall henceforth use the term 'pattern' to refer to a pattern class.

According to the pattern theory [6], we treat the temporal patterns that have been obtained as new elements of data. We strive to find higher level patterns in the distributions these new elements with respect to the other components of the data, which are the set of the episodes considered as discrete objects and the time with its intrinsic relationships of temporal ordering and distances; see Section 2.1.

Hence, based on the pattern theory summarised in Section 2.1, our approach includes two stages:

1. Detect and abstract temporal patterns of singular attributes appearing in the episodes.
2. Treating the univariate temporal patterns as data elements, study the distribution of these "elements" within the set of episodes and along time.

In this approach, we deal with patterns of two levels of complexity and sophistication. The first stage discovers lower-level patterns formed by temporally ordered numeric values. The second stage aims to discover higher-level patterns formed by these lower-level patterns due to their relationships and thereby imposed arrangements within and across the episodes. Within the episodes, the univariate patterns of multiple attributes are linked by the relationships of *co-occurrence*. Across the episodes, the univariate patterns are linked by relationships of *temporal ordering and temporal distance*.

The task **T1** formulated in Section 1 refers to the first stage and the remaining tasks to the second stage. The task **T4** focuses on the relationships of co-occurrence within episodes. The expected type of higher-level patterns is which univariate patterns tend to frequently co-occur and which do not occur together. The task **T3** focuses on the temporal ordering and strives to find patterns of frequent or infrequent occurrence of one lower-level pattern immediately after another. The task **T2** focuses on more distant temporal relationships regarding the arrangement of the lower-level patterns along the time axis and, when appropriate, within temporal cycles. The expected types of higher-level patterns include tendencies to occur earlier or later in time or at certain positions in a cycle, to re-occur more or less frequently in different time periods, to occur in a particular sequence, etc.

As stated by the pattern theory [6], pattern discovery is supported by faithful visual representation of relevant relationships. Taking into account the aforementioned relationships that are relevant for tasks **T2-T4**, we propose the following visualizations to support these tasks:

- **T2**: A timeline display of the temporal patterns (Fig. 1), where the horizontal axis represents the linear ordering relationships between time intervals, plus circular diagrams (Fig. 6), where positions in circles represent the cyclic arrangement relationships.

- **T3**: A bipartite graph of immediate transitions between patterns of the same attributes (Fig. 7).
- **T4**: A co-occurrence network (Fig. 8).

The task **T1** can be fulfilled in different ways, for example, by dividing the time series into intervals and encoding the interval averages by symbols according to the value ranges in which the averages fall. The resulting codes are called SAX patterns [28]. In our paper, we propose another method, which is based on the recognition of the geometric shape that would be formed when the time series is represented graphically by a line chart. It should be noted that the visual analytics techniques we propose for the tasks **T2-T4** do not depend on the method of extracting and encoding temporal patterns and on the choice of labels to denote the patterns.

We demonstrate our approach on example of Google Mobility data [18]. Continuous time series of daily mobility indicators were divided into disjoint episodes.

3.2. Approach introduced by example

The COVID-19 pandemic has impelled local authorities and/or governments to regulate people's mobility. Such policies generate changes in mobility, which are typically sporadic across a certain period. We should distinguish those sporadic patterns from seasonal repetitions in mobility data. For example, we can expect the increasing number of people staying at home and the decreasing number of those going out during the Christmas season. Moreover, different categories of places have different patterns of mobility even during the same time interval. Here, our interest is to visualize temporal patterns across episodes and to investigate how the mobility changes over time across different categories of places.

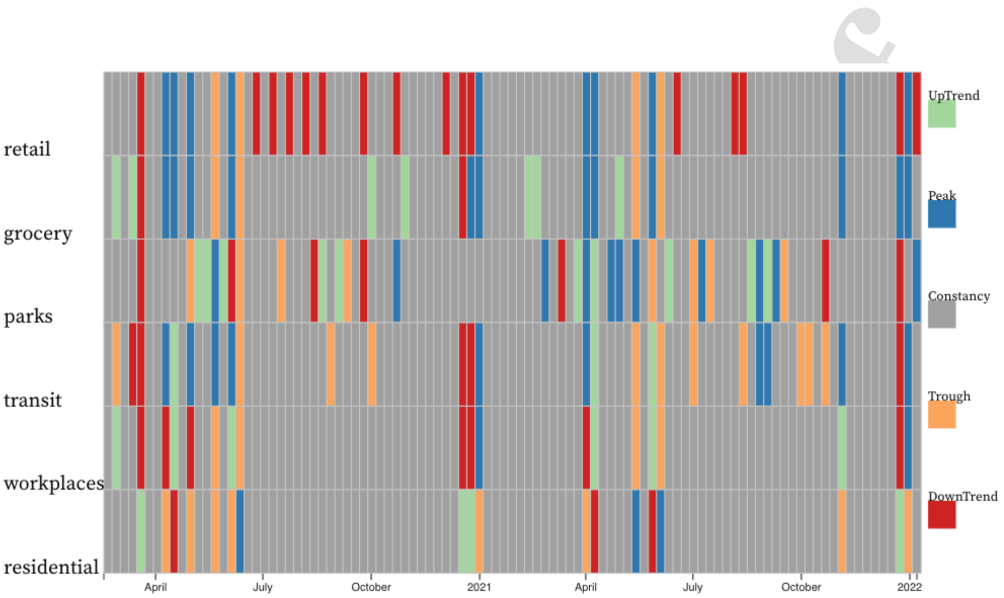


Figure 1: Timeline of temporal patterns in Google Mobility Data.

Data Description

We preprocess the mobility data provided by Google [18] to obtain multivariate time series. Since the COVID-19 outbreak around February 2020, Google has been daily publishing anonymized mobility data for 6 different categories of places (namely, retail and recreation, supermarkets and pharmacies, parks, public transport, workplaces, and residential) from different regions. The data consist of daily visitor numbers to these categories of places relative to baseline days before the pandemic outbreak. Baseline days represent a normal value for each day of the week and are given as the median value over the five-week period from January 3rd to February 6th 2020. The values in the published data are expressed as percentages of the changes from the baseline values.

From the continuous time series, we extract the time intervals of weekdays (i.e., 5 time steps for each week) with the corresponding segments of the time series. Mobility data for weekends are excluded from the analysis because changes from weekdays to weekends are very prominent and therefore obscure the longer-term changes of the mobility behaviours. We process the mobility data for Germany collected between the 17th of February, 2020 and the 7th of January, 2022 (i.e., almost for two years), which results in 99 episodes of the length of five time steps.

For validation purposes, we acquired values of eight policy indicators (namely, closing of schools, workplace, and public transport, cancelling public events, and restrictions on internal and international movement) from the Oxford COVID-19 Government Response Tracker [19].

T1. What are the major patterns of individual attributes?

In the introduction, we mentioned the existence of simple, easily perceivable and interpretable patterns of temporal variation of numeric attributes. These patterns can be schematically represented by lines of particular geometric shapes. Let us use the term "elementary pattern" for a pattern that can be represented (in abstraction from minor fluctuations) by a single straight line. There are three elementary temporal patterns: up-trend, constancy, and down-trend. More complex patterns can be considered as sequences of these. Fig. 2 illustrates how elementary temporal patterns can make more complex temporal patterns. Any temporal pattern starts with one of the elementary patterns, and a sequence of two or more temporal patterns can make a composite pattern such as a peak or a trough.

When a sequence consists of the same kind of elementary pattern (i.e., up-trend, constancy, or down-trend), we can simply consider it as a single temporal pattern. For example, a sequence of two up-trend patterns makes a single up-trend pattern and this temporal pattern makes a peak pattern with a subsequent down-trend (i.e., up→up→down makes peak). Note that when the sequence gets longer, it can create a more complicated shape. A long time series often looks like an oscillation. It can be simplified by means of temporal smoothing. We assume that the episodes under analysis are short, so that the time series include a small number of time steps and thus can be represented by sufficiently simple patterns. Longer episodes can be subdivided into shorter ones to enable such representation. Another possibility is to downsample the time series, i.e., reduce the number of time steps by dividing a long sequence of time steps into a small number of intervals and taking a single representative value (e.g., the mean or median) from each interval.

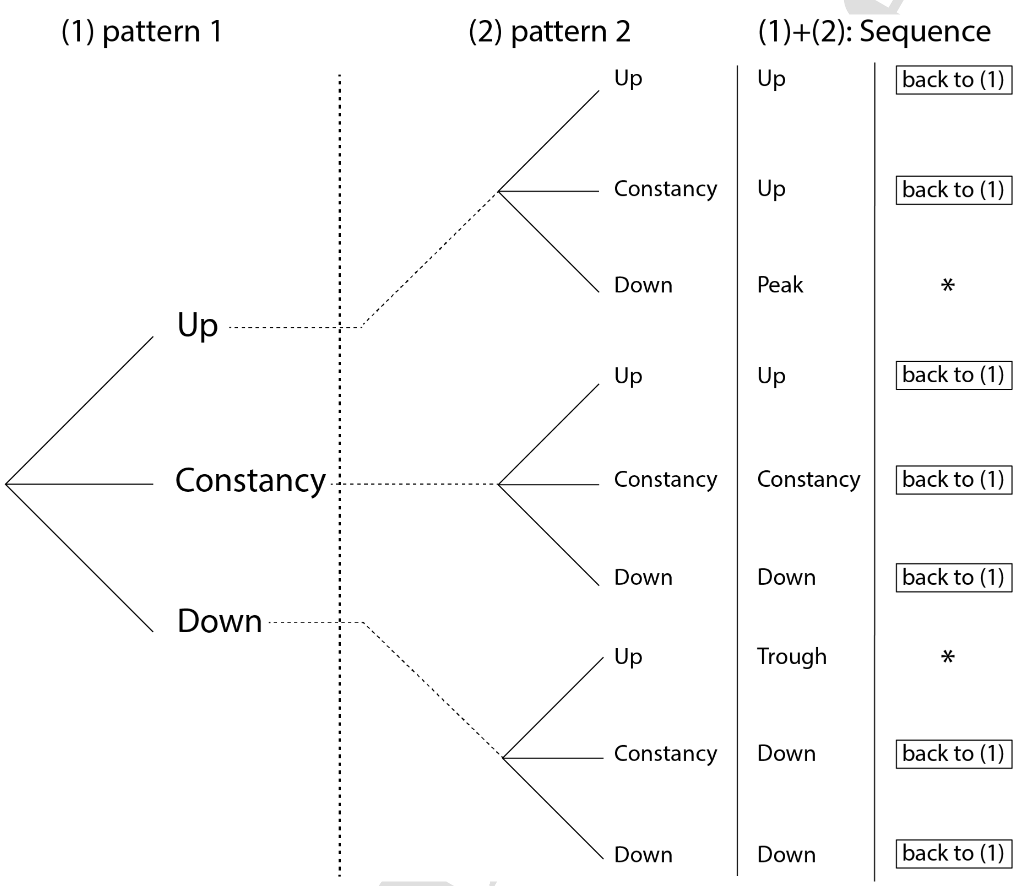


Figure 2: Possible sequences of elementary temporal patterns. For example, a temporal pattern consisting of up-up-constancy-down-down will be classified as a peak. A long time series including a peak or a trough (marked *) may require a subtle adjustment to distinguish different temporal patterns.

We assign an episode to one of the five temporal patterns to represent the most prominent shape of the time series: up-trend, peak, constancy, trough, down-trend. To determine a temporal pattern, we adapt the main idea from the algorithm of Steinarsson [39], which was devised for downsampling of time series , i.e., reducing the number of points used to represent the time series. This matches very well our goal to transform time series into simple shapes that can be represented by very few points. The method is based on finding the data point that makes the largest triangle when connected to the first and last data points in a time interval. Fig. 3 shows an example of the largest triangle in a time series. Fig. 4 illustrates the work of the pattern determination algorithm, which is explained below.

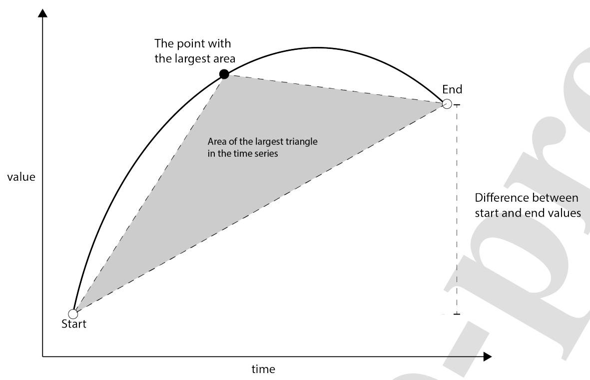


Figure 3: The largest triangle in time series.

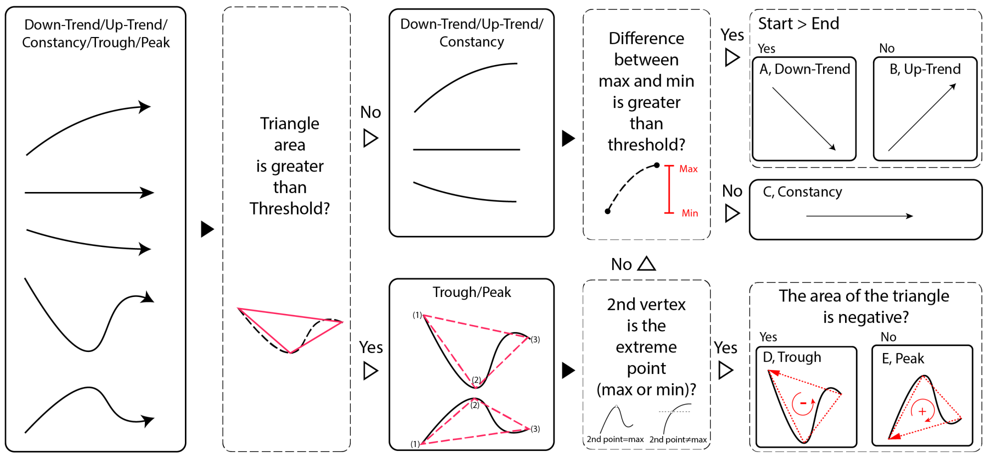


Figure 4: The process of pattern determination. Time series will be classified into either A. Down-Trend, B. Up-Trend, C. Constancy, D. Trough, or E. Peak.

A time series can be classified as peak or though when the area of the largest triangle is greater than a chosen threshold. To find the largest triangle, we take the first and the last points of the time series as the first two vertices of a triangle and test all intermediate data points one after another as potential third vertices of the triangle. From these points, we take the one that makes a triangle with the largest area among all.

If the absolute value of the area of the largest triangle is above the threshold, the time series has either a peak or a trough; otherwise, it can be classified as a trend (up or down) or constancy. The value of the area is treated as negative when the order from the start, via the extreme, to the end points is counter-clockwise. Otherwise, the area has a positive value. The time series has a peak with a positive area and a trough with a negative area.

When the time series is neither peak nor trough, meaning that the values do not significantly deviate from the straight line connecting the first and last points, the time series has either of the following patterns: an up-trend, a down-trend, or a constancy. Imagine a time-distance graph for uniform velocity, where distance increases at the same pace. In this case, no triplet of the points makes a triangle, and we define the area to be zero. This pattern determination is relatively straightforward; when the difference between the start and end values is larger than a chosen threshold, the time series has either an up-trend or a down-trend, otherwise it has a constancy. Then the time series has a down-trend when the start value is greater than the end, and an up-trend happens when the end value is greater than the start.

Results of pattern detection depend on two thresholds that we use for determining peaks vs. troughs and identifying constancy patterns. The specific values of the thresholds are not essential for demonstrating our approach. Generally speaking, these thresholds are application-specifics, and domain knowledge may be needed for setting them properly. In our example, we've performed self-assessment to choose appropriate values based on several trials. For the assessment, we used a visualization with time series translated to a common starting point, as in Fig. 5.

Table 1 presents the distribution of temporal patterns in the mobility data. We observe constancy as the most frequent among the patterns. This observation can be confirmed by time series visualizations in Fig. 5.

The types of patterns our algorithm aims to extract can be categorised as *patterns of value change*, while, for example, SAX patterns [28] can be seen as *patterns of value magnitude*. Our algorithm ignores the magnitudes of values and considers only the differences with respect to the first value of a time series. This needs to be taken into account when assessing the suitability of our algorithm for specific analysis goals. Another important note is that the algorithm allows extraction of a more refined set of pattern types than

Table 1: The overview on temporal patterns in their frequency (Peak/Trough threshold = 0.1, Constancy threshold = 0.2). We observe constancy as the majority.

	Peak	UpTrend	Constancy	DownTrend	Trough
retail_and_recreation	10	0	76	9	4
grocery_and_pharmacy	12	4	77	2	4
parks	11	9	64	7	8
transit_stations	10	3	70	5	11
workplaces	2	6	80	7	4
residential	3	4	81	3	8

we consider in our examples. Thus, for the peak and trough patterns, it is possible to introduce subtypes based on whether the final value of the time series increased, decreased, or remained nearly the same as the first value. For the up- and down-trends, it is possible to distinguish steep and gradual increase or decrease. An appropriate level of pattern abstraction can be chosen in accord with the goals of analysis. In our examples, we extract and use highly abstracted patterns; however, the exploratory techniques we demonstrate can also be applied to an extended set of patterns.

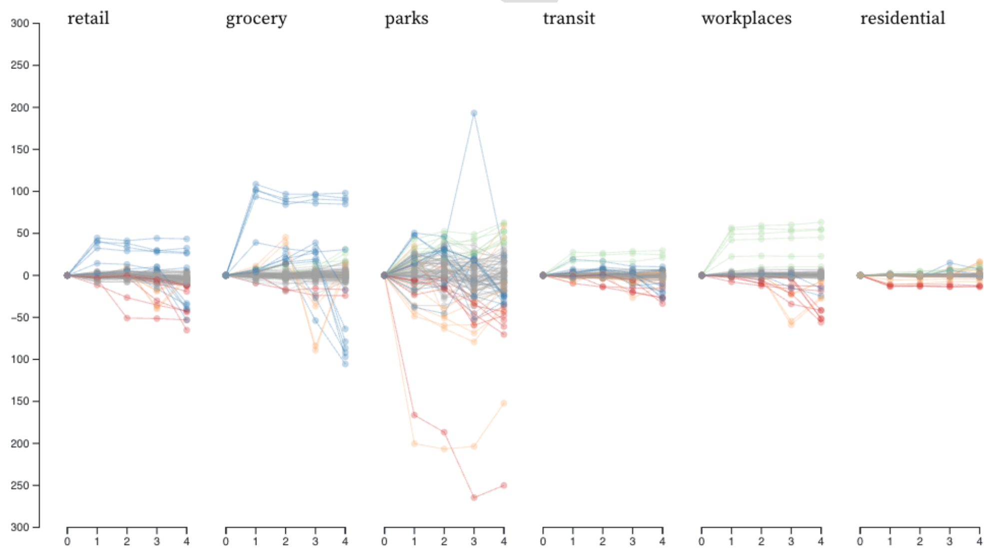


Figure 5: Actual time series with values shifted to align the start points.

T2. What is the temporal distribution of the patterns?

For this task, we propose two visualizations focusing on different types of relationships between time intervals. The timeline view (Fig. 1) focuses on the relationships of linear ordering, which are represented by positions on a straight horizontal time axis. The circular view (Fig. 6) focuses on the relationships of cyclic temporal arrangement between the episodes. In a circular chart, the years are represented by rings, and episodes (weeks of data) are blocks of the rings arranged clockwise. In both views, the temporal patterns of the individual episodes are represented by colour coding.

The timeline view (Fig. 1) reveals periods of stable mobility behaviour (i.e., prevalence of the constancy patterns) and periods of changes, in which all mobility indicators or some of them are non-constant. It shows when different patterns of the individual indicators occurred, what pattern combinations existed, and when they took place. The prevailing combination throughout the entire time span was the combination of six constancy patterns. Other combinations are rare and require more attention to be identified. For example, the combination of simultaneous down-trends of the visits of all places except homes and an up-trend of the staying at home occurred in the third week of March, when the first lockdown was issued. Similar combinations (differing by just one constituent pattern) in the Christmas periods of 2021 and 2022. These were followed by combinations of the trough in staying home and peaks in visiting all place categories except for parks.

This re-occurrence of similar patterns at the ends of two years can also be noticed by looking at the circular charts (Fig. 6). Each chart facilitates identification of seasonal and sporadic temporal patterns of a single attribute. In Fig. 6 (a), we clearly see that some temporal patterns re-occur annually. These recurrent patterns can be attributed to seasonal variations represented in the data. For example, down-trends are seen in the 'retail and recreation', 'public transport', and 'workplaces' features at the end of each year while we observe an up-trend in the 'residential' feature. We can conjecture that people travel less and prefer to stay home in the Christmas season. While the circular charts are good in revealing periodic repetitions of single-feature patterns, detection of re-occurring combinations requires integrating information from six charts; hence, holistic perception of pattern combinations is not supported by this representation. The timeline view, on the opposite, supports holistic perception of combinations but does not show periodicity as clearly as the circular charts.

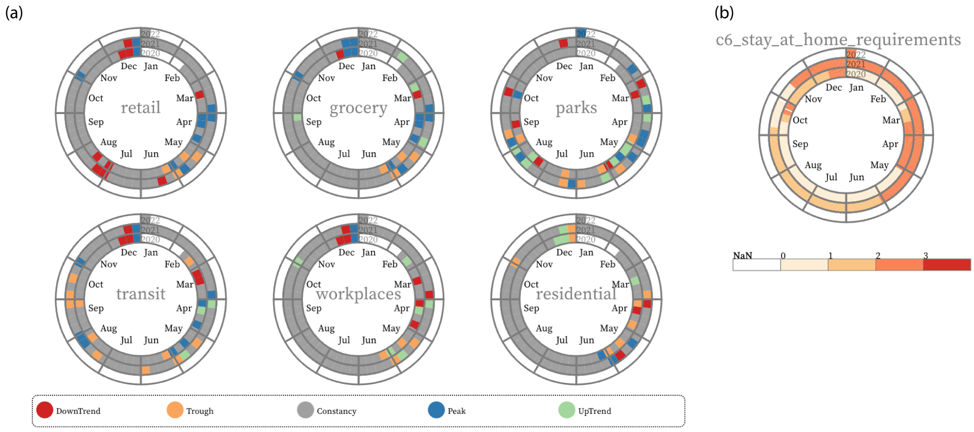


Figure 6: Circular displays of temporal patterns for features in Google Mobility Data (a) and stay-at-home requirements level in Germany (b). In all plots inner ring represents year 2020, middle - 2021, outer ring - 2022. Values in (b) mean 0: no measures announced, 1: recommended not leaving house, 2: required not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips, 3: required not leaving house with minimal exceptions (e.g. allowed to leave once a week, or only one person can leave at a time, etc), NaN: no data [19]

The circular charts also help to detect sporadic occurrences of temporal patterns, which may be caused by factors or events that do not occur regularly. For instance, the German government required closing (or working from home) for some sectors or categories of workers. Fig. 6 (b) shows that the stay-at-home requirement level goes from 0 (no measures) to 2 (require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips) in the middle of March, 2020. In Fig. 6, as well as in Fig. 1, we see the effect of this measure: the residential category shows an up-trend at this time while the others have a down-trend. Moreover, we also see that the 'grocery and pharmacy' category has an up-trend in the week before the down-trend, which suggests that people went to groceries to stockpile products of everyday use (e.g., food and toilet paper) in preparation for the forthcoming restrictions or possible good shortages.

T3. Are there frequent transitions between univariate temporal patterns over sequential times?

We create bipartite graphs to represent transitions of univariate temporal patterns between consecutive time intervals. It helps to find patterns of temporal succession and adjacency between the same and different temporal patterns of feature variation. In Fig. 7, there are six bipartite graphs, one per feature, consisting of three components: two proportionally segmented bars and curved lines linking the bar segments. The segmented bars show the overall proportions of the occurrences of the different patterns in the episodes. The segments are painted in the colours corresponding to the patterns using the same encoding as in the timeline view and the circular charts. The opacity and the stroke width of the linking lines represent the frequency of the transitions between the classes of the temporal patterns represented by the bar segments.

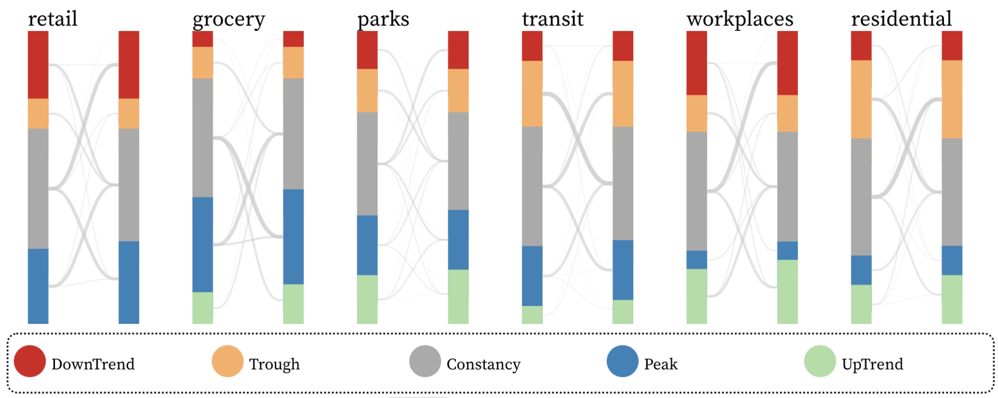


Figure 7: Bipartite graph of transitions between different patterns.

This representation can be interactively modified for focusing on selected patterns only. For example, most frequent transitions between constancy patterns are subject to be omitted for the sake of better visibility of the other transitions.

T4. Which patterns frequently co-occur?

To answer this question, we build a co-occurrence network, where nodes represent the temporal patterns of the features and edges connect patterns of different features that co-occur in the same episodes (Fig. 8, left). The size of a node represents the frequency of the temporal patterns appearing in the dataset and the opacity and the stroke width of an edge represent the frequency of co-occurrence. For example, we see that the decreasing pattern of visiting residential places and the increasing pattern of visiting workplaces frequently co-occur with the increase of the use of transit station. Note that, same as in the transition graph, the co-occurrence between two constancy patterns is obvious and therefore omitted from the chart.

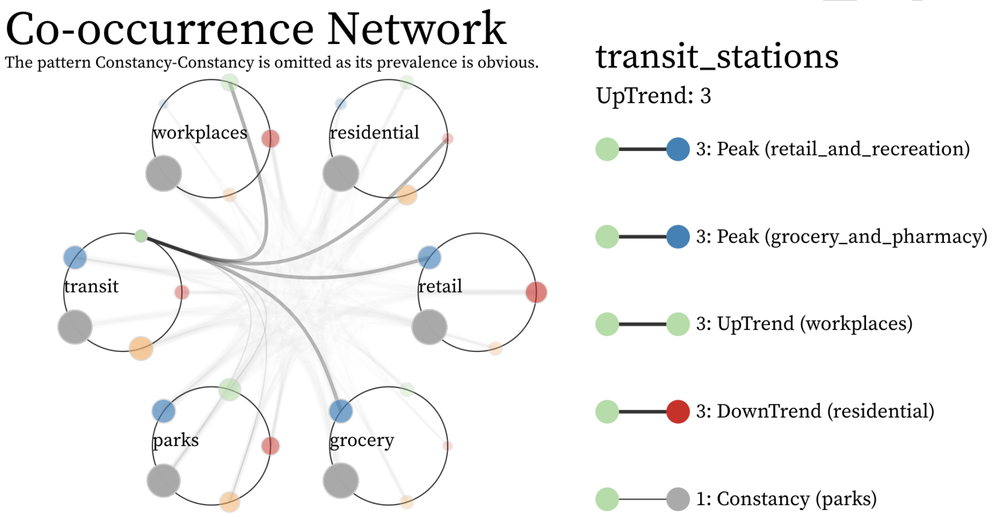


Figure 8: Co-occurrence network with the up-trend of 'transit stations' highlighted.

An analyst can interactively select a node in the network for displaying the most frequent co-occurrences of the respective temporal pattern with the temporal patterns of the other features. This interactive exploration reduces clutter in the chart and facilitates finding important relationships. Thus, the right part of Fig. 8 demonstrates the effect of selecting the node representing the up-trend pattern of 'transit stations'. It shows that this pattern occurred only three times in our data set, and in all cases it occurred together with the peak pattern of 'retail and recreation' and 'grocery and pharmacy', the up-trend pattern of 'workplaces', and the down-trend of 'residential'. This reveals a re-occurring multivariate temporal pattern (i.e., a combination of univariate patterns) in the data set.

4. Case Study: Teams' behaviours in football

We demonstrate the generality of our approach by applying it to episodes around ball possession change from a professional football (or soccer) match. Different types of changes of possession exist in football, each of which forces both teams to switch their tasks from attacking to defending or vice versa.

The team can apply different tactics. For example, after regaining the possession possible options are either to approach the opponent's goal (i.e., execute a counter-attack) or to remain at own side to protect the possession.

While different types of transitions are typically visible to a human eye, experts such as video analysts often have to watch the game to sub-categorize the transitions (e.g., label them as counter-attacks or securing the possession), which is a time-consuming and daunting task. Our intention in this study is to investigate which multivariate temporal patterns appear in transition episodes in football. We characterize these episodes by spatial features of collective movement.

Data Description

We extract episodes from positional data of players in one professional football match. We choose time intervals based on the occurrence of a specific event, i.e., change of possession. Each time interval consists of players' positions for ten seconds around transitions and the change of possession occurs exactly in the middle of the episode. As a consequence, we acquire 115 episodes, each lasting 10 seconds (i.e., 250 timesteps, given that the raw data has a sampling rate of 25 Hz), with 63 episodes seeing the home team gaining the possession and 52 episodes featuring the away team. Next, we characterize time intervals by spatial features that can be computed from positional data: compactness of the team, distance from their own goal, and velocity. For each team, we compute **team width** (i.e., distance perpendicular to the side line, between the most left-positioned field player and the most right-positioned one), **team depth** (i.e., distance parallel to the side line, between the farthest player from the goal and the nearest one, except the goalkeeper), and **distance from the center of the team to the own goal**. Since we observe a strong correlation of average velocities between players of the two teams, we calculate a common **average velocity** for the 20 infield players of both teams.

4.1. T1. What are the major patterns of individual attritbutes?

As Fig. 9 shows in grey, episodes consisting of many time steps (250 in our case) may have complex temporal patterns consisting of multiple fundamental patterns. As discussed in T1 in Section 3, temporal patterns need to be sufficiently simple to allow easy interpretation. Complex patterns can be simplified by omitting excessive details, which can be achieved through downsampling of the time series. We use the same algorithm [39] introduced in T1 to downsample the episodes. The red dot lines in Fig 9 show how time series with 250 timesteps are downsampled into 5 timesteps. We begin by applying the downsampling technique to each half of the episode in order to get a representative value that would form the greatest triangle with two ends in the divided half. Then, using our algorithm on the downsampled episode, we classify temporal patterns. Fig. 10 illustrates all downsampled time series with colors.

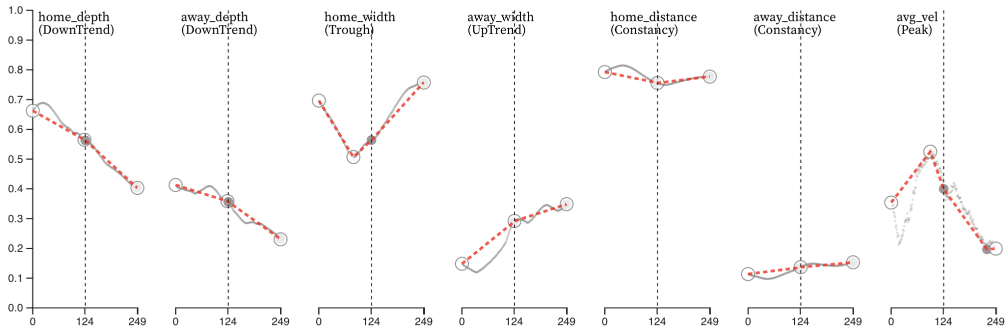


Figure 9: An overlay of the downsampled time series and the original time series for the home depth, away depth, home width, away width, home distance to their goal, away distance to their goal, and average velocity. The larger three points including the start and end points indicate the points used to classify the temporal pattern, and together with the other smaller two points they form the downsampled time series. The downsampled time series is colored to indicate the classified temporal pattern while the original time series is grey. The dotted line in the middle means the middle point of the episode. Time is shown on the horizontal axis in frames (1/25th of a second) while the normalized attribute values ranging from 0 to 1 are shown on the vertical axis.

Table 2: Frequency of temporal patterns for each feature in the football data set. Two numbers in each cell represent two types of episodes where the home team begins by defending (left) and when the away team begins by defending (right). (Peak/Trough threshold = 0.05, Constancy threshold = 0.1).

	Peak	UpTrend	Constancy	DownTrend	Trough
home_depth	14, 16	14, 13	19, 6	10, 6	6, 11
away_depth	21, 12	10, 15	12, 9	14, 2	6, 14
home_width	6, 11	10, 4	10, 6	10, 21	27, 10
away_width	11, 2	7, 13	12, 15	23, 5	10, 17
home_distance	0, 6	8, 7	33, 30	21, 9	1, 0
away_distance	1, 0	23, 10	30, 25	9, 10	0, 7
avg_vel	31, 26	5, 5	1, 0	7, 0	19, 21

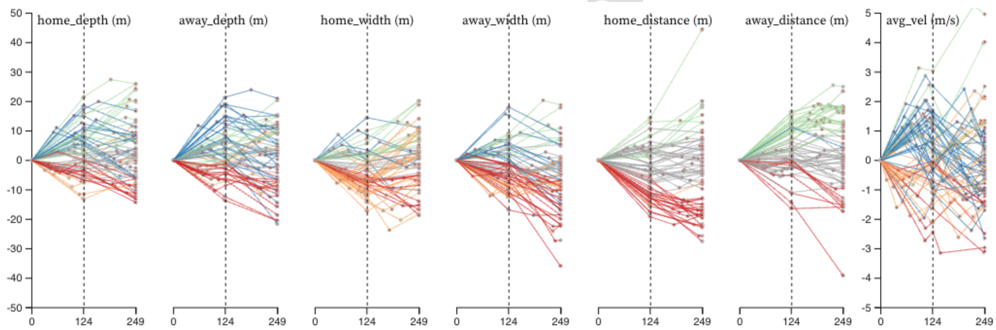


Figure 10: All downsampled time series. Colors indicate classified temporal patterns. Time is shown on the horizontal axis in frames (1/25th of a second) while the changes of the attribute values with regard to the initial point are represented by the vertical positions. The axes are labelled according to the measurement units of the original (not normalized) attributes.

Table 2 summarizes the detected patterns. For the attribute `home_width`, we observe a prevalence of patterns with increase towards the episode end (i.e., up-trends and troughs) over decreasing patterns (37 vs. 16) in the episodes when the home team begins the episode by defending (left side of the cell). This means that the home team tends to expand after they gain the possession, which is a known behaviour in football [15]. We find the opposite patterns (i.e., down-trends and peaks) to be the majority in `away_width` (34 out of 63). Second, we observe a similar number of up-trend patterns in `home_distance` as down-trend patterns in `away_distance`, as well as the similar number of down-trend patterns in `home_distance` as up-trend patterns in `away_distance`. Fig. 11 confirms this finding with the centroids of both teams following similar trajectories. Third, we see most of patterns appear as peaks or trough (50 out of 63) in `avg_vel`. We can assume that the change of possession can accelerate or decelerate players abruptly rather than monotonically. Finally, the significant difference between the both teams may be the trough pattern of the distance from the goal. We observe only one trough pattern in `home_distance` while seven in `away_distance`. Different tactics, such as having the away side attempt more counter-attacks than the opponent, can account for this variation.

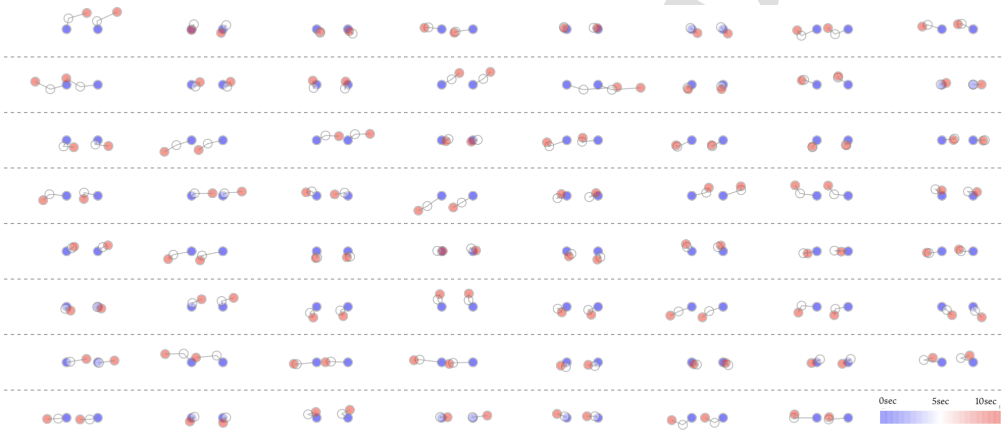


Figure 11: The team centroid shifts during the episode. Each row depicts the shift in the centroid for both teams over the course of eight episodes (left: home, right: away). The colors reflect the progression of time, from blue to white to red.

4.2. T2. What is the temporal distribution of the patterns?

We use a linear ordering to represent the temporal distribution of the temporal patterns. In Fig. 12, rectangles that represent episodes are aligned more sparsely than in Fig. 1 since the time intervals are selected according to specific events.

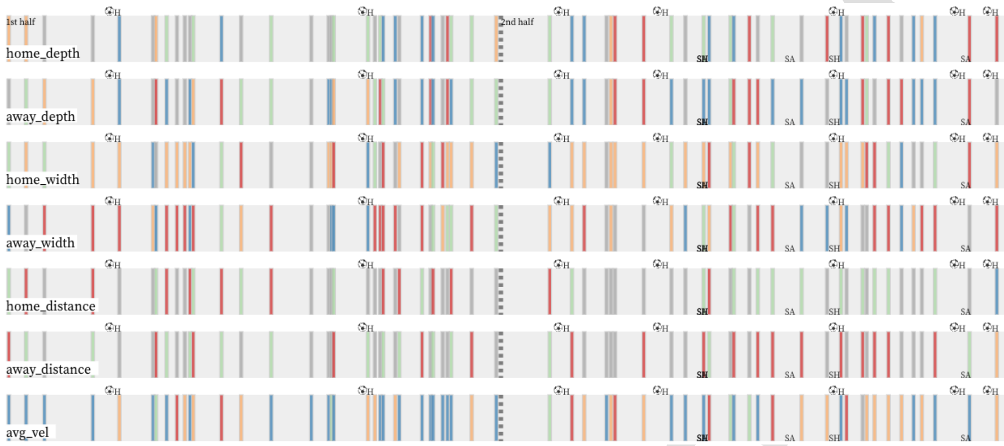


Figure 12: Timeline of temporal patterns in football data. Markers at the top and bottom of each row indicate goals and substitutions (Ball: goal, S: substitution. H: home, A: away).

Fig. 13 shows a circular view of the temporal distribution of the patterns. Two arcs in each chart represent the temporal axes, where inner arcs represent the first half of the match and outer arcs represent the second half. Although no periodic repetitions can be expected, this view can facilitate understanding the data as a circle refers to a clock face, which allows to compare the first and second halves.

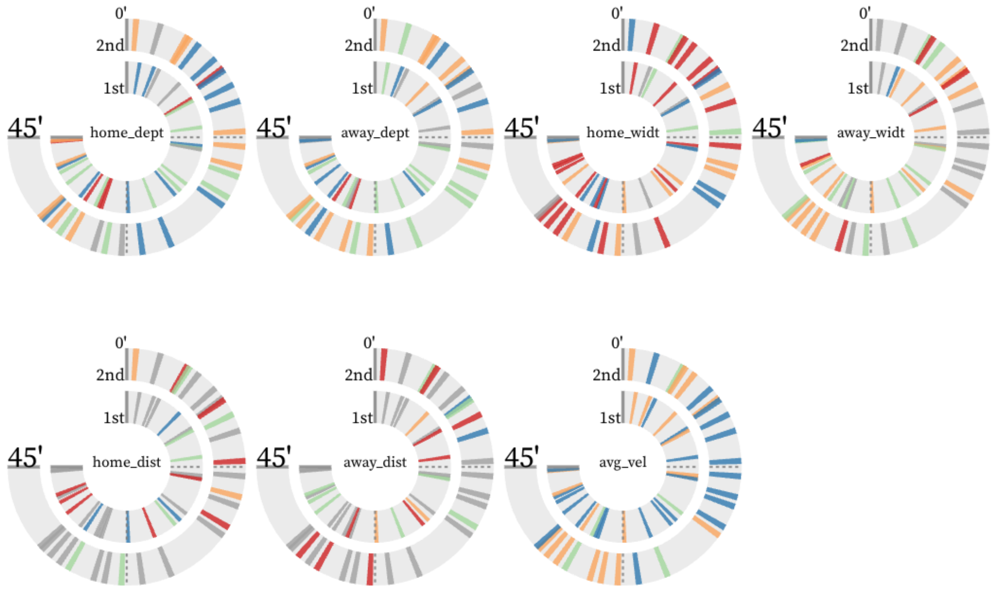


Figure 13: Circular time view of temporal patterns in football data.

4.3. T3. Are there frequent transitions between univariate temporal patterns over sequential times?

This task is not applicable to this dataset since episodes appear sporadically.

4.4. T4. Which patterns frequently co-occur?

Fig. 15 shows a co-occurrence network applied to the episodes (left) and the five multivariate temporal patterns that most frequently co-occur with the up-trend pattern of `home_distance` (right), where the home team gains the ball possession in the middle (at 5 seconds).

One third of the patterns with increasing `home_width` toward the episode end (i.e., up-trends and troughs) co-occur with the combination of `avg_vel`'s peak, `home_distance`'s down-trend, and `away_distance`'s up-trend. Fig 14 illustrates the movement of the team centroids in these episodes. We further identify from the footage that the defending team slowly rebuilds the attack after collecting long balls deep in their own side. Other combinations such as with `avg_vel`'s trough or `home_distance`'s up-trend mainly consist of counter-attacks, collecting balls relatively near to the opponent's goal, or immediate regains of possession by the defending team. Similar tendency is found in the co-occurrence of increasing patterns of `away_width` when the away team is defending. However, we observe more counter-attacks with `avg_vel`'s trough (21% vs 14%), which implies that the away team tends to attack fast after they gain the possession.

The fact that the home side finished the season in the top three and the other team in the relegation zone explains these distinct tactics. While the away side may have preferred long balls to possession, the home team may have felt secure in controlling the ball against the opponent.

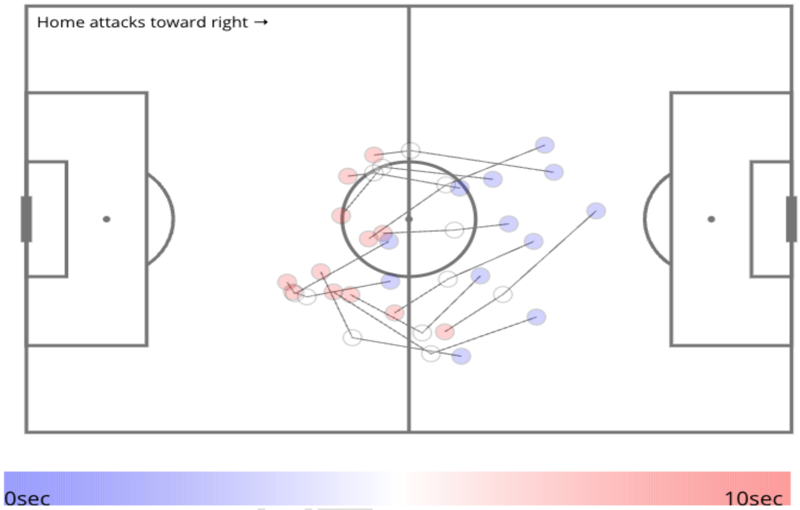


Figure 14: Movement of centroids during episodes with `avg_vel`=peak, `home_distance`=down-trend, and `away_distance`=up-trend when the home team is defending

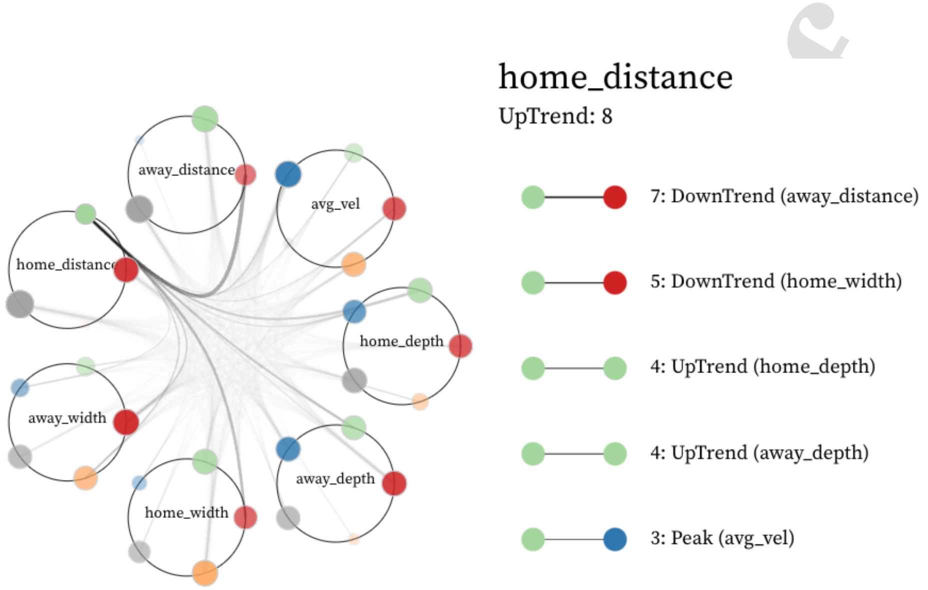


Figure 15: Features that have co-occurrence with the up-trend pattern of 'home_distance' (left) and the top 5 co-occurrent features (right) in episodes from Match 1, where the home team defends.

4.5. Summary of findings

Our approach enabled us to identify similar and distinctive behaviours for the two teams. Temporal patterns show players often play wide when they are attacking and narrow when they are defending. Additionally, quick acceleration and deceleration in response to a change of possession is observed. The co-occurrence chart reveals two typical tactics used by both sides when they gaining possession of the ball: either executing counter-attacks or gradually rebuilding the attack. After obtaining possession of the ball, the home team often carefully connects passes while the away team typically attempts quick counter-attacks.

5. Discussion

With this paper, we are proposing a view of time-varying phenomena as a sequence of episodes, i.e., time intervals encapsulating fragments of the temporal behaviours of the phenomena. The term "behaviour" here refers to any kinds of changes. Episodes can be described by values of multiple attributes specified for different time slices within the intervals and thus forming multivariate time series. The rationale for introducing episodes as units of behaviour is that they can be short enough to allow abstractive perception and representation of each time series as a single easily interpretable temporal pattern. Hence, the behaviour encapsulated in an episode can be represented by a combination of patterns made by the multiple attributes.

Based on the premise that simplification and abstraction are essential for understanding a phenomenon, i.e., building a mental model of it [7], we explored in our research the analytical potential of computer-supported abstraction of time series to temporal patterns and explicit representation of these patterns for involvement in subsequent analysis. The idea is that the patterns substitute the original elementary data [6] and are themselves treated as data to be analysed. We considered several analysis tasks that can be posed when dealing with such data and defined visual analytics techniques that can support these tasks.

In our exploratory study, we neither tried to create a complete task taxonomy for analysis of temporal patterns of episodes nor strove to design novel visualisations. The goal was to investigate the principal possibility of analysing data transformed into temporal patterns. Our study showed that this approach can be quite useful. By using abstractions of elementary data, it allows considering the behaviour of a phenomenon at a yet higher level of abstraction, namely, at the level of relationships between the patterns. This contributes to obtaining an overall understanding of the behaviour or revealing its essential features. It can be noted that the very idea of the approach is generic, i.e., potentially applicable to any type of data.

Given that transformation of data to patterns can be beneficial, a valid question is what kinds of patterns should be considered and how to obtain them from data. This question requires a specific answer for each distinct type of data, because patterns are formed by type-specific intrinsic relationships between data elements [6]. We have proposed an answer to this question for data consisting of time series of values of numeric attributes. We wanted to represent such data by patterns that are well understood by humans and, preferably, denoted by commonly understandable terms. We considered a set of basic patterns that can be represented graphically as particular geometric shapes and are commonly labelled as up-trend, peak, constancy, trough, and down-trend. We propose an algorithm for automatic recognition of these patterns and representation of episodes by combinations of patterns. We acknowledge the possibility to consider other sets of patterns requiring other algorithms for extraction, but we would like to note that the same visualisation and exploration techniques may be applied to transformed data regardless of the specific pattern "vocabularies" used for encoding the data.

The visual analytics techniques that we described in this paper are intended to support exploration of (A) the temporal distribution of the different types of patterns and (B) relationships between the temporal patterns, namely, temporal ordering of patterns in a sequence of episodes and co-occurrence of patterns within episodes. A and B are the two major classes of analytical tasks relevant to time-referenced data in general. The most common representation of such data is by some kind of visual marks along a time axis, and we apply it in our timeline view. A circular representation of time is also frequently used, particularly, to reveal and explore cyclic changes. We also propose two time-abstracted and aggregated representations of the data in the form of graphs showing sequential ordering relationships between patterns of the same attribute and co-occurrence relationships between patterns of different attributes. Using graphs to visualise relationships is one of the most common design choices along with the use of a time axis-based display to visualise a temporal distribution. The visualisations we describe in the paper should be considered as mere examples of numerous possible implementations of these fundamental designs.

Thus, there are many methods for laying out nodes of a graph [17]. Most of the existing algorithms are not suitable for visualising relationships between patterns, which requires the nodes representing the patterns of the same attribute to be grouped together and separated from nodes referring to other attributes. We address this requirement in our design of the co-occurrence network (Figs. 8 and 15) by arranging groups of nodes in circles. A more usual design that could satisfy this requirement is the chord diagram [23] using a circular layout, where groups of nodes are arranged in arcs and separated from other groups by gaps. Figure 16 demonstrates how the same data as in Figs. 8 and 15 can be visualised in the form of chord diagrams. In our design, the grouping of nodes is much better noticeable than in a chord diagram. A disadvantage of our design is intersections between some of the graph edges and the circles that visually link nodes belonging to groups. The circular layout, as in a chord diagram, is potentially suitable for visualising hierarchical networks by increasing the number of outer circles; however, this is not needed in our case. The circular layout may also be more scalable to a greater number of nodes given its simple structure; however, showing a large number of node groups with sufficient separation between them may be problematic. Since there is no universally effective layout, the choice should depend on properties of data and user preferences.

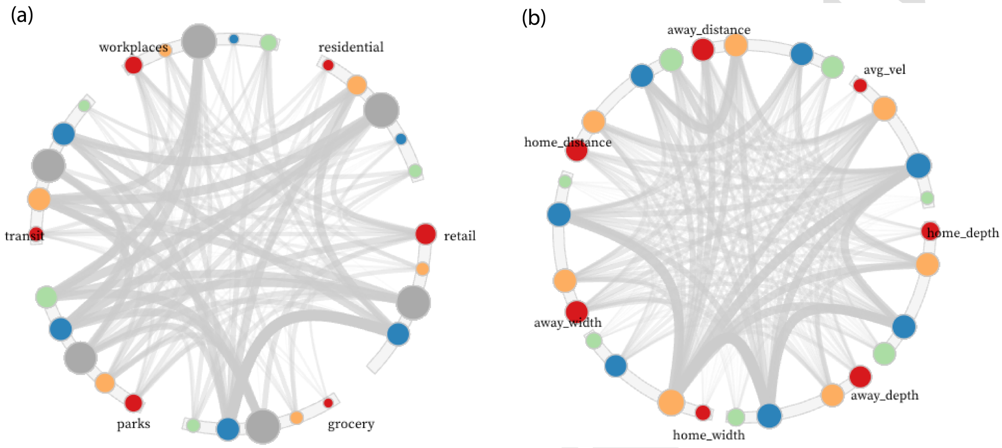


Figure 16: An alternative design of co-occurrence chart with two data sets: (a) Google Mobility Data (2) Football Data

We consider our work as just a first step in the research on analysis of episodes as a way of representing complex dynamic phenomena. We envisage continuous systematic research in this direction. Our exploratory study shows how this representation can be utilised leveraging the possibility of condensing and abstracting elementary data. While we see that this approach has good potential, we admit that the set of techniques we have developed is not yet sufficiently powerful. In particular, it provides quite limited opportunities for exploration of multi-attribute temporal patterns, i.e., combinations of single-attribute patterns. The co-occurrence network shows only pairwise co-occurrence relationships but does not support joint perception and analysis of multiple patterns occurring together in episodes. We see the problem of representing and analysing multivariate temporal patterns as a challenge for future research that requires significant attention and concentration of effort.

Hence, one of the next steps in the future research should be towards finding methods for the integration of multiple single-attribute temporal patterns into composite multi-attribute patterns that can be perceived and treated as units. We see a possibility to achieve this goal with the help of topic modelling. Our experiments [38] showed that this idea deserves further investigation. Another step should be towards methods for comprehensive analysis of temporal relationships between patterns not limited to co-occurrence and sequential ordering. Our initial idea is to consider temporal neighbourhoods of patterns and try to find re-occurring combinations of patterns whose neighbourhoods overlap.

6. Conclusion

We have introduced a concept of episode as a relatively short fragment in temporal development or behaviour of a dynamic phenomenon. We have suggested that data describing episodes may have the form of time series of values of multiple attributes. Limiting our focus to numeric attributes, we have presented an approach to analysis of such data by means of automated abstraction of the time series to temporal patterns represented as categorical labels. We have demonstrated possible ways of visualising abstracted data for analysing the temporal distribution of the patterns and relationships between patterns within and across episodes. Our study has shown that decomposition of complex behaviours into episodes and characterising episodes by temporal patterns of multiple attributes is a promising approach to analysis of dynamic phenomena. We call for further research in this direction, particularly, to find ways to consider and analyse combinations of single-attribute patterns holistically as integrated patterns incorporating multiple aspects of the behaviour.

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), by EU in projects *SoBigData++* and *CrexData*, and by DFG within priority research program *SPP VGI* (project *EVA-VGI*).

References

[1] Aigner, W., Miksch, S., Schumann, H., Tominski, C., 2011. Visualization of time-oriented data. Springer. doi:10.1007/978-0-85729-079-3.

[2] Allen, J.F., Ferguson, G., 1994. Actions and events in interval temporal logic. J. Logic Comput. 4, 531–579.

[3] Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., Hecker, D., Weber, H., Wrobel, S., 2021a. Constructing spaces and times for tactical analysis in football. IEEE Transactions on Visualization and Computer Graphics 27, 2280–2297. doi:10.1109/TVCG.2019.2952129.

[4] Andrienko, N., Andrienko, G., Camossi, E., Claramunt, C., Cordero-Garcia, J.M., Fuchs, G., Hadzagic, M., Jousselme, A.L., Ray, C., Scarlatti, D., Vouros., G., 2017. Visual exploration of movement and event data with interactive time masks. Visual Informatics 1, 25 – 39. doi:https://doi.org/10.1016/j.visinf.2017.01.004.

[5] Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., Wrobel, S., 2020. Visual Analytics for Data Scientists. Springer.

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

[7] Andrienko, N., Lammarsch, T., Andrienko, G., Fuchs, G., Keim, D., Miksch, S., Rind, A., 2018. Viewing visual analytics as model building. Computer Graphics Forum 37, 275–299. doi:10.1111/cgf.13324.

[8] Bale, K., Chapman, P., Barraclough, N., Purdy, J., Aydin, N., Dark, P., 2007. Kaleidomaps: A new technique for the visualization of multivariate time-series data. Information Visualization 6, 155–167. doi:10.1057/palgrave.ivs.9500154.

[9] Bernard, J., Dobermann, E., Bögl, M., Röhlig, M., Vögele, A., Kohlhammer, J., 2016. Visual-interactive segmentation of multivariate time series, in: Proceedings of the EuroVis Workshop on Visual Analytics, Eurographics Association, Goslar, DEU. p. 31–35.

[10] Bernard, J., Wilhelm, N., Scherer, M., May, T., Schreck, T., 2012. Timeseriespaths : Projection-based explorative analysis of multivariate time series data, in: Journal of WSCG, pp. 97–106.

[11] Buono, P., 2016. A circular visualization technique for collaboration and quantifying self, in: Proceedings of the International Working Con-

ference on Advanced Visual Interfaces, Association for Computing Machinery, New York, NY, USA. pp. 348–349.

[12] Collins, C., Andrienko, N., Schreck, T., Yang, J., Choo, J., Engelke, U., Jena, A., Dwyer, T., 2018. Guidance in the human—machine analytics process. Visual Informatics 2, 166 – 180. doi:10.1016/j.visinf.2018.09.003.

[13] Crisan, A., Fisher, S.E., Gardy, J.L., Munzner, T., 2021. GEViTRec: Data reconnaissance through recommendation using a Domain-Specific visualization prevalence design space. IEEE Trans. Vis. Comput. Graph. PP.

[14] Das, G., Lin, K.I., Mannila, H., Renganathan, G., Smyth, P., 1998. Rule discovery from time series., in: KDD, pp. 16–22.

[15] Fonseca, S., Milho, J., Travassos, B., Araujo, D., Lopes, A., 2013. Measuring spatial interaction behavior in team sports using superimposed voronoi diagrams. Int. J. Perform. Anal. Sport 13, 179–189.

[16] Gharghabi, S., Yeh, C.C.M., Ding, Y., Ding, W., Hibbing, P., LaMunion, S., Kaplan, A., Crouter, S.E., Keogh, E., 2019. Domain agnostic online semantic segmentation for multi-dimensional time series. Data mining and knowledge discovery 33, 96–130.

[17] Gibson, H., Faith, J., Vickers, P., 2013. A survey of two-dimensional graph layout techniques for information visualisation. Information Visualization 12, 324–357. doi:10.1177/1473871612455749.

[18] Google, 2022. COVID-19 community mobility reports.

[19] Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H., 2021. A global panel database of pandemic policies (oxford COVID-19 government response tracker). Nat Hum Behav 5, 529–538.

[20] Hao, M., Marwah, M., Janetzko, H., Sharma, R., Keim, D.A., Dayal, U., Patnaik, D., Ramakrishnan, N., 2011. Visualizing frequent patterns in large multivariate time series, in: Visualization and Data Analysis 2011, SPIE. pp. 194–203. doi:10.1117/12.872169.

[21] Haroz, S., Kosara, R., Franconeri, S.L., 2016. The connected scatterplot for presenting paired time series. IEEE Trans. Vis. Comput. Graph. 22, 2174–2186.

[22] Heer, J., Kong, N., Agrawala, M., 2009. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. pp. 1303–1312. URL: http://idl.cs.washington.edu/papers/horizon.

[23] Holten, D., 2006. Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans. Vis. Comput. Graph. 12, 741–748.

[24] Janetzko, H., Sacha, D., Stein, M., Schreck, T., Keim, D.A., Deussen, O., 2014. Feature-driven visual analytics of soccer data, in: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 13–22. doi:10.1109/VAST.2014.7042477.

[25] Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2001. Dimensionality reduction for fast similarity search in large time series databases. Knowl. Inf. Syst. 3, 263–286.

[26] Keogh, E., Lin, J., Lee, S.H., Van Herle, H., 2007. Finding the most unusual time series subsequence: algorithms and applications. Knowl. Inf. Syst. 11, 1–27.

[27] 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, 1359–66. doi:10.1109/TVCG.2009.200.

[28] Lin, J., Keogh, E., Wei, L., Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. Data Min. Knowl. Discov. 15, 107–144.

[29] Liu, L., Mei, S., 2016. Visualizing the GVC research: a co-occurrence network based bibliometric analysis. Scientometrics 109, 953–977.

[30] Lu, Y., Wu, R., Mueen, A., Zuluaga, M.A., Keogh, E., 2022. Matrix profile XXIV: Scaling time series anomaly detection to trillions of datapoints and ultra-fast arriving data streams, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. pp. 1173–1182.

[31] Mauceri, S., Sweeney, J., Nicolau, M., McDermott, J., 2021. Feature extraction by grammatical evolution for one-class time series classification. Genet. Program. Evolvable Mach. 22, 267–295.

[32] Peuquet, D.J., 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the Association of american Geographers 84, 441–461.

[33] Pham, V.V., Nguyen, N.V.T., Li, J., Hass, J., Chen, Y., Dang, T., 2019. Mtsad: Multivariate time series abnormality detection and visualization. 2019 IEEE International Conference on Big Data (Big Data) , 3267–3276.

[34] Qiang, Y., Delafontaine, M., Versichele, M., De Maeyer, P., Van de Weghe, N., 2012. Interactive analysis of time intervals in a two-dimensional space. Inf. Vis. 11, 255–272.

[35] Robertson, G., Fernandez, R., Fisher, D., Lee, B., Stasko, J., 2008. Effectiveness of animation in trend visualization. IEEE Trans. Vis. Comput. Graph. 14, 1325–1332.

[36] Rosling, H., 2006. The best stats you've ever seen. https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen. Accessed: 2021-11-5.

[37] Schreck, T., Tekušová, T., Kohlhammer, J., Fellner, D., 2007. Trajectory-based visual analysis of large financial time series data. SIGKDD Explor. Newsl. 9, 30–37. URL: https://doi.org/10.1145/1345448.1345454, doi:10.1145/1345448.1345454.

[38] Shirato, G., Andrienko, N., Andrienko, G., . What are the topics in football? Extracting time-series topics from game episodes. IEEE VIS 2021 poster URL: http://geoanalytics.net/and/papers/vis21poster.pdf.

[39] Steinarsson, S., 2013. Downsampling Time Series for Visual Representation. Ph.D. thesis. University of Iceland.

[40] Tanisaro, P., Heidemann, G., 2019. Dimensionality reduction for visualization of time series and trajectories, in: Felsberg, M., Forssen, P., Sintorn, I., Unger, J. (Eds.), Image Analysis, Springer International Publishing. pp. 246–257. doi:10.1007/978-3-030-20205-7_21.

[41] Tominski, C., Schumann, H., 2020. Interactive Visual Data Analysis. CRC Press.

[42] Van Wijk, J.J., Van Selow, E.R., 1999. Cluster and calendar based visualization of time series data, in: Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99), pp. 4–9.

[43] Van de Weghe, N., Docter, R., De Maeyer, P., Bechtold, B., Ryckbosch, K., 2007. The triangular model as an instrument for visualising and analysing residuality. J. Archaeol. Sci. 34, 649–655.

[44] Yeh, C.C.M., Kavantzas, N., Keogh, E., 2017. Matrix profile VI: Meaningful multidimensional motif discovery, in: 2017 IEEE International Conference on Data Mining (ICDM), IEEE. pp. 565–574.

[45] Zhao, J., Drucker, S.M., Fisher, D., Brinkman, D., 2012. TimeSlice: interactive faceted browsing of timeline data, in: Proceedings of the International Working Conference on Advanced Visual Interfaces, Association for Computing Machinery, New York, NY, USA. pp. 433–436.

[46] Zhao, J., Forer, P., Harvey, A.S., 2008. Activities, ringmaps and geovisualization of large human movement fields. Information Visualization 7, 198–209. doi:10.1057/PALGRAVE.IVS.9500184.

Author Information

- Title
  - Identifying, exploring, and interpreting time series shapes in multivariate time intervals
- 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

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: