Visual Informatics 9 (2025) 100234

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Research article

# Contextualized visual analytics for multivariate events

Lei Peng [a], Ziyue Lin [a], Natalia Andrienko [b,c], Gennady Andrienko [b,c], Siming Chen [a,*]

[a] School of Data Science, Fudan University, Shanghai, China
[b] Fraunhofer IAIS, Sankt Augustin, Germany
[c] City, University of London, London, UK

ARTICLE INFO

*Article history:*
Received 30 December 2024
Received in revised form 16 March 2025
Accepted 18 March 2025
Available online 21 March 2025

*Keywords:*
Visual analytics
Event analysis
Contextualized analysis
Interactive exploration
Visualization design

ABSTRACT

For event analysis, the information from both before and after the event can be crucial in certain scenarios. By incorporating a contextualized perspective in event analysis, analysts can gain deeper insights from the events. We propose a contextualized visual analysis framework which enables the identification and interpretation of temporal patterns within and across multivariate events. The framework consists of a design of visual representation for multivariate event contexts, a data processing workflow to support the visualization, and a context-centered visual analysis system to facilitate the interactive exploration of temporal patterns. To demonstrate the applicability and effectiveness of our framework, we present case studies using real-world datasets from two different domains and an expert study conducted with experienced data analysts.



## 1. Introduction

In some event analysis scenarios, analysts are focused on a specific type of event. For example, a football analyst may want to analyze passing events during a football game. A typical approach is to extract and analyze the occurrences of such events. However, event analysis goes beyond the moments of occurrence. Analysts may also be interested in the causes, consequences, and other aspects related to the events. In many cases, this information is likely to be found from before and after the events. From this perspective, focusing solely on the moment of the occurrence may not provide sufficient information for event analysis. Therefore, it is meaningful to reframe event analysis tasks from a contextualized perspective.

For a multivariate time series with events recorded, we can define an event context as a multivariate time slice surrounding the point of time when an event occurs. Shifting the analysis object from the event occurrence to the event context introduces several challenges:

The first challenge is that event context, compared to event occurrence, is a more complex object to visualize. Its visual representation should be easy to understand, and should allow analysts to observe the changes over time within the context. Secondly, analyzing multivariate event contexts requires handling increased data volumes and multiple dimensions. Finally, the visual analysis of event contexts could be a flexible and open-ended process, requiring the integration of human insight.

To deal with these challenges, we propose a framework for visually representing and exploring multivariate event contexts. The framework includes three main components: a design of visual representation for multivariate event contexts, a data processing workflow, and a context-centered visual analysis system.

We apply this framework to two real-world datasets and present case studies. In the first case, we analyze passing event contexts in a football match, identifying several distinct behavior patterns across different passing event contexts and offering insights into tactical strategies. In the second case, we analyze harsh braking event contexts in a truck's driving records, identifying several driving patterns and various responses to road conditions. These case studies demonstrate the general applicability of our framework to various domains. We also conduct an expert study to collect feedback on our analysis framework and system, which helped validate the effectiveness of our approach.

Overall, our study makes the following contributions:

- We propose a multivariate event context analysis framework that helps analysts identify and explore temporal patterns within event contexts through interactive visual analysis.
- We introduce a method for visualizing multivariate event contexts, which includes a visual representation design based on color coding and high-dimensional context information ordering.
- We provide a context-centered visual analysis system to support the processing and exploration of event contexts.

* Corresponding author.
*E-mail address:* simingchen@fudan.edu.cn (S. Chen).

https://doi.org/10.1016/j.visinf.2025.100234
2468-502X/

## 2. Related works

Our work is closely related to existing studies on multivariate data visualization and event visual analysis. This section provides an overview of the related works in these fields.

### 2.1. Multivariate data visualization

Multivariate data visualization is a widely studied area. Cui (2019) provides a literature review on visual analytics, in which he categorizes and discusses the applications of multivariate data visualization. A primary approach is to represent multivariate information in a single chart or view, such as using parallel coordinate plots. 3D visualization and animation can also increase the number of variables represented in a single view. Another approach to visualizing multivariate data is to construct multiple related views. The information that can be displayed in a single chart is limited. Therefore, in recent studies, multiple views are often adopted for multivariate data visualization. Multiple views are often realized as interactive visual analysis systems, such as Smartadp (Liu et al., 2016), which analyzes large-scale GPS trajectory data for billboard placement, and Srvis (Weng et al., 2018), which solves ranking decision problems by combining spatial environmental information. Wang et al. (2022) also provide a multi-view visual analysis system for spatio-temporal data to help developers observe when, where and how the model failure happens in autonomous driving systems.

Multivariate data poses challenges to visualization due to its complexity. Excessive volume, multiple dimensions, and information redundancy may all cause difficulties in visualization. Therefore, data mining techniques are widely used to simplify multivariate data before visualization. The corresponding techniques include dimensionality reduction, projection, clustering, sampling, and others. Nonato and Aupetit (2018) provide a literature review on multivariate projections (MDPs) for visual analytics, categorizing MDP methods, transformations, analysis tasks, and layout enhancements. Kisilevich et al. (2010) provide an overview of spatio-temporal clustering, especially trajectory clustering. Andrienko and Andrienko (2013) support using clustering and interactive analysis for large-scale spatio-temporal data, and apply color assignments to projected data. TPFlow (Liu et al., 2018) models multivariate spatio-temporal data as tensors, proposing an algorithm that automatically segments the data into uniform partitions and extracts potential patterns from each of them. Deng et al. (2023) develop a data mining framework to extract evolution patterns from large-scale spatio-temporal series and propose a technique called GeoChron which leverages the evolution pattern to organize and visualize large-scale spatio-temporal series in a pattern-aware and narrative-preserving way.

In our work, we use various multivariate visualization methods, utilizing both single-chart representations of multiple dimensions (e.g., parallel coordinate plots) and multi-view interactions. We also employ data mining techniques such as dimensionality reduction and clustering to support data visualization.

Our work is innovative in the visualization of event contexts. To the best of our knowledge, there has been limited research dedicated to the visualization of event contexts. Andrienko and Andrienko (2024) visualizes event contexts in matrices. And in our work, we propose a visual representation specifically designed for multivariate event contexts.

### 2.2. Event visual analysis

Substantial work has been done on the detection, comparison, and visual analysis of event sequences. In these studies, flow charts, Sankey diagrams, and their variations are often used to visualize the temporal order, synchronization, and interaction of events. Due to the complexity of event analysis tasks, each work focuses on specific aspects of event analysis or approaches them from different perspectives.

For example, some studies approach event analysis from a sequence analysis perspective. Eventpad (Cappers and van Wijk, 2017) offers a graph-oriented approach to analyzing event sequences, enabling both temporal analysis of the sequences and structural analysis of associated multivariate data. DecisionFlow (Gotz and Stavropoulos, 2014) processes high-dimensional temporal event sequences of thousands of different types, supporting interactive real-time analysis of high-dimensional event sequence data. Sequence Braiding (Bartolomeo et al., 2020) focuses on event alignment and comparison. IVESA (Bernard et al., 2024) focuses on the analysis of time-stamped event sequences (TSEQs), which are time-oriented event occurrence sequences without value information. Some other studies focus on causality analysis in event sequences. Jin et al. (2020) applies Granger causality testing to explore causality in spatio-temporal events. VAC2 (Zhu et al., 2024) provides a combined causality visual analysis system to help users effectively explore combined causes in temporal event sequence data.

There are also application-focused event studies. PassVizor (Xie et al., 2020) deals with football passing event sequences, helping to reveal dynamic patterns and tactics in passes. Pmu tracker (Arunkumar et al., 2022) provides a visualization platform for epicentric event propagation analysis in the power grid.

Meanwhile, there are studies that focus on the contextual nature of events. Although treating events as sequences is valuable, examining events within contexts can be equally important. For instance, Frequence (Perer and Wang, 2014) is an interactive frequent event sequence mining and visualization interface that considers the user's temporal context by defining pattern duration in sequences. CoNTA (Cappers and van Wijk, 2016) supports contextual analysis of network traffic alerts. Andrienko et al. (2011) propose a conceptual model supporting the analysis of movement data with spatio-temporal context. Chen et al. (2019) introduce a workflow for analyzing movement events, including event contextualization, context pattern detection, and exploration. In their work, they visualize and compare event contexts based on a single variable.

In our work, we define event contexts as multivariate time slices surrounding the event occurrences, and focus on the exploration of patterns within them. Our definition of "context" is distinguished from the concept "episode" as used in other works on temporal pattern analysis, such as Shirato et al. (2023b) and Andrienko et al. (2023). The key difference lies in the event-centered nature of "context", whereas "episode" does not emphasize the presence of events. In the realm of temporal pattern analysis, our approach is explicitly event-oriented.

To our knowledge, there have not been studies that approach multivariate event visual analysis from a contextualized perspective. Thus, in this paper, we provide a framework for the contextualized visual analysis of multivariate events.

## 3. Overview

In this chapter, we will first introduce our target data type and provide a specific definition of event contexts. Next, we will present the analysis tasks derived from discussions with experts. Finally, we will give an overview of the multivariate event context analysis framework we propose and the methods employed.

### 3.1. Data and definitions

Our framework is designed for multivariate time series data, which should include time and several other numerical variables. Specifically, when the data contains variables representing spatial information, it is referred to as trajectory data. Though spatial variables are not essential in our target data, our framework includes functions for trajectory analysis.

Occurrences of events and their types should be recorded in the data. Furthermore, it is important to note that the framework is designed for analyzing recurring events instead of sporadic or anomalous events. Therefore, the data should include multiple records of the same event type. For instance, the tracking data in a football match might record multiple occurrences of passing events.

To use our framework, time series with uniform time intervals is the most suitable. For time series with non-uniform time intervals, methods such as interpolation are needed to unify the intervals before extracting event contexts.

For a time series that meets the above conditions, with $T$ time points and $n$ variables, we denote it as $X = \{x_t\}^{T\times n}$, where $x_t$ is the multivariate data entry at time point $t$. The time intervals in this series are equal and denoted as $\Delta t$. $Event(t) \in \{0, 1\}$ is a binary variable that indicates whether a target event occurs at time point $t$. If a target event occurs at $t$, we define its context $C_t$ as a time slice around $t$. To be specific,

$$C_t = \{\mathbf{X}[t - a\Delta t,\ t + b\Delta t] \mid \mathrm{Event}(t) = 1\}, \quad a, b \in \mathbb{N}$$

The event context $C_t$ includes $(a + b + 1)$ data entries, where $a$ entries are before the event, and $b$ entries are after the event. The values of $a$ and $b$ are selected by the analyst based on the specific analysis scenario.

### 3.2. Analysis tasks

Our framework is designed for experienced data analysts. We aim to propose an visual analysis framework that help analysts in gaining insights into events from a contextualized perspective. We discussed with several experts who have more than five years of experience in data analysis, and gathered their requirements for analyzing multivariate event contexts. During the discussions, the experts raised several needs related to the mining of contexts, the recognition of temporal patterns, and the interpretability of these patterns. From the discussions with experts, we derive the following analysis tasks:

- T1: Visualize multivariate event contexts.
- T2: Identify event contexts with similar temporal patterns.
- T3: Interpret the meanings of temporal patterns within similar event contexts.

### 3.3. Analysis framework

To support the analysis tasks, we propose the following analysis framework as shown in Fig. 2. In the framework, the target data goes through a data processing workflow to generate visual representations of event contexts. Then, we provide a context-centered visual analysis system that allows users to identify and interpret patterns within contexts interactively. In the system, users can also perform real-time data processing tasks such as clustering.

The visual representation of event contexts is designed based on a "building block" metaphor. It enables comparisons based on similarity and the identification of temporal patterns within the contexts. A detailed introduction to the visual representation of event contexts is in Section 4.

The data processing workflow consists of 4 steps. In Step 1 Event context extraction, the event contexts $C_t$ are extracted from original data $X$. The collection of all the data entries included in extracted event contexts is denoted as a data subset $C = \{C_t\}$. In Step 2 Projection and coloring, $C$ is reduced to two dimensions to enable the use of a 2-D colormap, which assigns a representative color to each data entry. In Step 3 Clustering and Step 4 Reordering, event contexts are clustered and ordered based on similarity. Details about the data processing workflow will be discussed in Section 5. For each step, various statistical algorithms will be considered and compared.

## 4. Visual representation of multivariate event contexts

In order to visualize multivariate event contexts (T1), we first work on the design of the visual representation for event contexts.

Multivariate event context is a complex object. It contains multiple variables as well as temporal information. First, we consider how to represent multivariate information in the visual representation. Visualizing event contexts requires a simplification of information, since it is impractical to visually display all the variables within a context. One feasible approach is to retain similarity only. More specifically, we can set a design goal: when comparing two visual representations of contexts, the viewer should be able to assess their similarity. We can obtain the similarities between data entries through dimensionality reduction. However, results of dimensionality reduction algorithms involve unavoidable distortions. Therefore, the representation that we design allows only approximate judgments of similarity.

Secondly, we consider how to represent temporal information in the visual representation. In the definition provided in Section 3.1, a multivariate event context is a combination of data entries at multiple time points. We thus consider designing the visual representation of a event context as a combination of representations of various time points. The advantage of this design lies in that it aligns the data level definition and the visual representation of an event context, making it easy to understand.

Therefore, we have developed the visual representation for multivariate event contexts as described below. We first project all the data entries included in the event contexts (denoted as $C$ in Section 3.3) onto a 2-D plane. As shown in Fig. 3, using a 2-D colormap, each data point can be assigned a corresponding color. In visualizations, color coding is regularly used to represent changes in data values. Color coding can represent both changes in a single dimension such as changes over time (Bernard et al., 2012), and changes in two dimensions such as changes over a 2-D plane (Andrienko and Andrienko, 2023). We use color coding to represent the proximity of the projected data. In terms of comparison, similar colors indicate closer proximity. For example, with reference to the colormap, we can see that red points are closer to orange points, while blue points are farther away.

Assume that we have an example event context with 5 data entries (where $a = b = 2$). Now, each data entry has a color representing it, as shown in Fig. 4-a. Next, as illustrated in Fig. 4-b, we represent each data entry as a colored block and stack them from top to bottom in time order. To be specific, in this case, block $t_3$ in the middle represents the data entry in which the event occurs. Block $t_1$ and $t_2$ represent the two data entries before the event. Block $t_4$ and $t_5$ represent the two data entries after the event. This design can be seen as a "building block" metaphor, where each colored block represents a data entry, and assembling them together forms the event context. It allows analysts to observe temporal changes within the context. For example, a transition from blue to yellow can be observed in this example context, indicating the potential temporal changes at the variable level.

Moreover, we consider how to display a large number of event contexts and make it easier for analysts to identify common patterns among them. When individual event contexts are placed together as shown in Fig. 5-a, their colors make them comparable for analysts. To visualize multiple contexts, we arrange them horizontally together as shown in Fig. 5-b. However, this is not sufficient for the task of pattern identification (T2). When there is a large number of contexts, clustering and ordering are needed to improve the arrangement of contexts based on similarity, as illustrated in Fig. 5-c. After this rearrangement, contexts that look most similar will be placed together. This approach makes it easier for analysts to identify common patterns among event contexts.

In summary, using this "building block" metaphor, we represent each data entry as a colored block and assemble them into the visual representation of an event context. The representations are further arranged and displayed based on similarity to ensure that similar contexts are placed together, making pattern identification easier for analysts. In the next section, we will discuss in detail all the methods used to generate and arrange these visual representations.

## 5. Data processing workflow

In this section, we present the data processing workflow. The workflow is used to extract event contexts from raw data and represent them visually as described in the previous section. The workflow includes 4 steps: event context extraction, projection and coloring, clustering, and reordering. We will detail each step of the data processing workflow and discuss the selection of statistical methods.

### 5.1. Step 1: Event context extraction

As is mentioned in Section 3.1, the target data for extracting event contexts is multivariate time series, with events of interest recurring over time. Analysts need to first determine the time interval and the length of event contexts. By definition, they need to choose the ideal time interval $\Delta t$ and the number of data entries they would like to include before and after the event (i.e, $a$ and $b$).

We can offer some guiding principles for making choices of time interval and context length. Firstly, the choice of context length should depend on the nature of the specific event. For example, constructing event contexts in minutes is acceptable for driving records, but may not be suitable for fast-changing events in football matches. As for the time interval, the rate of change in the phenomenon under analysis should be taken into consideration. Besides, the choice of the time interval is constrained by the quality of the raw data. The smaller the granularity of the raw data, the more choices the analyst has in choosing the time interval. However, when raw data granularity meets the requirements, we still advise analysts not to include too many data entries in the context, as this may lead to overlaps in the visualization.

Fig. 6 illustrates the context extraction process. We denote the data subset $C = \{C_t\}$ as the collection of all the data entries extracted from original data $X$. In some cases, due to the potential temporal overlap of event contexts, certain data entries might be included in multiple event contexts. But they will not be duplicated in data subset $C$.

Before moving on to the next steps, some data preprocessing measures are suggested for the data subset $C$. We suggest removing outliers and normalizing the data because the algorithms we use, such as k-means, are sensitive to outliers. To reduce multicollinearity in dimensionality reduction, we also recommend eliminating highly correlated variables.

### 5.2. Step 2: Projection and coloring

After extracting data subset $C$ from the original data, it is reduced to two dimensions so that a representative color can be assigned to each data entry using a 2-D colormap. These colors are used in constructing the visual representation of event contexts. Closer colors indicate greater similarity.

The choice of dimensionality reduction method is worth discussing. In our practice, we experimented with MDS, t-SNE, and auto-encoders, which are commonly used methods for dimensionality reduction. MDS and t-SNE are manifold learning algorithms, while auto-encoders are self-learning neural networks. Our experiments show that MDS has a high time cost for large data sets, making it unsuitable for the task. Auto-encoders are faster but produce unstable results, with projection distributions varying significantly across runs. In contrast, t-SNE offers better stability and has acceptable time costs. Another advantage of t-SNE is that it preserves local neighborhoods within data. In our analysis, we focus on groups of similar event contexts as our goal is to identify patterns. Therefore, it is essential to accurately recognize contexts close to each other in the original space. However, it is important to note that clusters farther apart in t-SNE results may not be comparable in terms of distance. The t-SNE algorithm gives close positions only to close neighbors, while slightly less similar data points can be placed quite far apart, which will result in getting a very different color.

In the rest of this paper, we use t-SNE as the default projection method. However, other methods may have advantages in different scenarios.

### 5.3. Step 3: Clustering

In most cases, analysts need to handle a large number of event contexts. Therefore, To assist analysts in identifying contexts with shared patterns (T2), the visual representations of event contexts should be displayed in an organized way. First, similar event contexts need to be grouped together so that analysts can study them in clusters or smaller groups.

K-means, a partition-based clustering algorithm that can efficiently handle large numbers of samples, is used here to cluster event contexts. Event contexts need to be transformed in order to make them processable for k-means. To be specific, for each event context, the matrix $C_t$ is flattened into a vector $vec(C_t)$ with $(a + b + 1) \times n$ dimensions, where $vec(C_t) = [x_{(t-a\Delta t)}, x_{(t-(a-1)\Delta t)}, \ldots, x_{(t+b\Delta t)}]$. K-means is then applied on these vectors.

K-means requires a predefined number of clusters. In practice, there is no standard answer for the optimal number of clusters, and analysts cannot determine beforehand how many groups the contexts should be divided into. Therefore, in our work, a human-in-the-loop approach is adopted, allowing analysts to decide the number of clusters. Reference information will be provided for analysts in the visual analysis system to support real-time clustering. Analysts can continually adjust the number of clusters until they achieve the ideal results.

### 5.4. Step 4: Reordering

K-means only provides an initial grouping of event contexts. Before proceeding with pattern identification (T2), the arrangement of event contexts needs to be further optimized within each group. Fig. 7-a shows a group of contexts arranged in their original order (typically the time order of their occurrences). In this order, the contexts are visually disorganized. Ideally, similar looking context representations should be close to each other, as shown in Fig. 7-b. This will improve visual coherence, and


**Fig. 1.** The visual analysis system for multivariate event contexts. The system has 5 views: the Timeline (A) which shows the distribution of event occurrences over time; the Context View (B) which shows the visual representations of contexts; the Reference View (C) which provides reference information for analysts; the Parallel Coordinates View (D) which shows the original values of variates and supports filtering; and the Space View (E) which shows the trajectories of contexts.

also ensure that neighboring event contexts are the most similar. Therefore, the reordering of event contexts within clusters is another important step in the data processing workflow.

Event contexts can either be reordered based on their original form $C_t \in \mathbb{R}^{(a+b+1)\times n}$ or on their projected forms $C'_t \in \mathbb{R}^{(a+b+1)\times 2}$ produced in Step 2. It is essential to recognize that statistical methods define similarity in different ways. Similarity encoding can vary significantly due to differences in definitions of distance and the inherent random variations present in some algorithms. Therefore, although the colors of contexts are assigned based on similarity, reordering using other statistical methods does not guarantee consistent results. To avoid inconsistency, we suggest reordering based on $C'_t$ instead of $C_t$. This is because the similarity of the projected 2-D data aligns with the similarity of their assigned colors. As is shown in Fig. 8, reordering on projected data yields better visual coherence.

Essentially, reordering event contexts is a similarity-based vector sorting task on $vec(C'_t)$. One potential approach is dimensionality reduction again, such as using MDS or t-SNE to reduce event context vectors to one dimension and then sorting them based on the projected values. Another approach we consider is principal component analysis (PCA), since ordering vectors by the first principal component value is also reasonable, as it explains most of the variance in the original data. We also experiment with the OPTICS algorithm, a density-based clustering algorithm that generates a similarity sequence as a byproduct. In short, OPTICS works by sequentially selecting the nearest neighbors of a given point, and then arranging similar objects together according to the sequence produced.

Fig. 9 shows the reordering results produced by MDS, t-SNE, PCA, and OPTICS. Compared to the original order, all of the results show improvements. Among these, OPTICS has several advantages. It offers slightly better visual coherence than the other methods, and is also the most time-efficient. Another advantage of OPTICS is that, due to its operating principle, it tends to place hard-to-sort examples at the end of each cluster, which is visually beneficial.

Some analysts may prefer t-SNE for consistency, as it is previously used in data projection. However, t-SNE is highly sensitive to parameter settings. In practice, we observe that without proper parameter tuning, t-SNE may fail to converge. Perplexity, a crucial parameter for t-SNE, must be adjusted based on data size. In this case, perplexity requires manual adjustment for each cluster since the number of event contexts per cluster may vary significantly. In contrast, OPTICS does not have this issue. Its parameter selection does not significantly affect the output. Therefore, due to its advantages in parameter selection flexibility, time efficiency, and ordering characteristics, we choose OPTICS as the default reordering method.

After completing these 4 steps of the data processing workflow, analysts can obtain organized visual representations of event contexts. In the next section, we will introduce the context-centered visual analysis system that allows analysts to engage in data processing and exploration.

## 6. Context-centered visual analysis system

This section provides a detailed introduction to the interactive visual analysis system (Fig. 1) designed to support exploratory analysis. This system serves as the platform for demonstrating the visualization of multivariate events (T1). Moreover, T2 and T3 will primarily be accomplished through the interactive analysis of analysts using the system. The system has 5 views: the Context View, the Reference View, the Parallel Coordinates View, the Timeline, and the Space View. The Context View displays the visual representations of contexts. The Reference View provides reference information related to clustering and projection. The Parallel Coordinates View allows researchers to examine and filter original data values. The Timeline and the Space View display the temporal and spatial information of contexts. Interactivity and cross-view linking are implemented across all views.

### 6.1. Context View (B)

The Context View (Fig. 1-B) is the core of the system and the center of interactive exploration. Here, the visual representations of event contexts are displayed according to clustering results. At the top of the view, analysts can select the number of clusters and perform real-time clustering. Additionally, contexts within each cluster will be reordered based on similarity automatically.

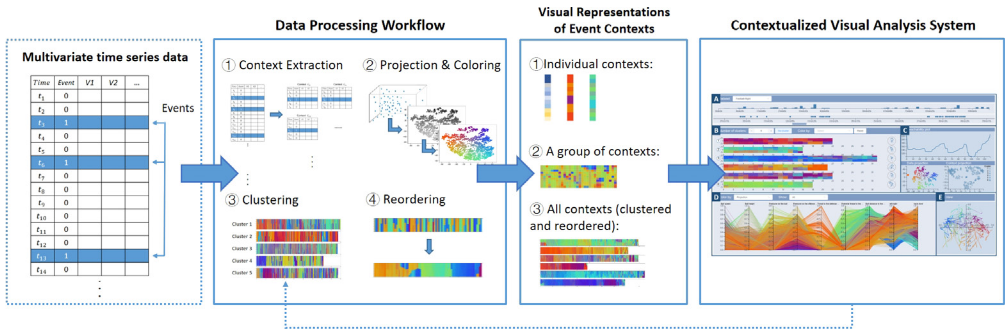


**Fig. 2.** The multivariate event context analysis framework. The framework consists of a design of visual representation for multivariate event contexts, a data processing workflow, and a context-centered visual analysis system.

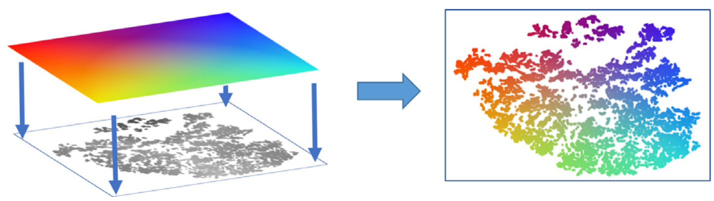

**Fig. 3.** With data projection and a 2-D colormap, each data entry can be assigned a representative color based on similarity.

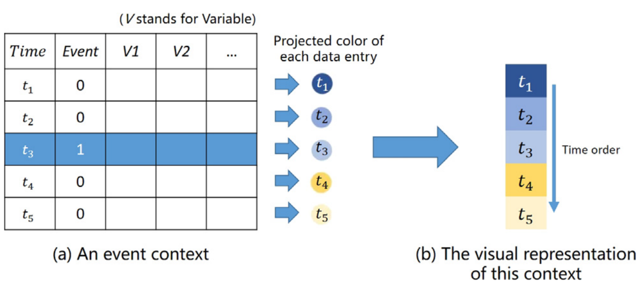


**Fig. 4.** An event context and its visual representation. We use colored blocks to represent each data entry, and stack them together to form the visual representation of an event context.

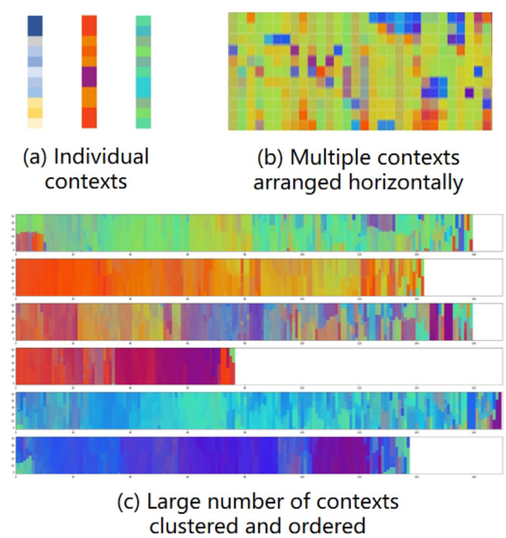


**Fig. 5.** The display of the visual representations of event contexts.

By default, contexts are colored according to the projected color. Analysts can choose to color the contexts by any variable using the dropdown menu at the top. This single-variable coloring function helps analysts trace specific visual patterns back to the original variables.

Additional interactions with the Context View can be carried out through brushing. Since clustering and reordering tend to place similar contexts together, analysts can easily brush and select groups of similar contexts for further exploration. Analysts can select contexts either partially or fully, and the selected parts will be filtered and highlighted across all views.

### 6.2. Reference View (C)

The Reference View (Fig. 10) provides analysts with reference information related to projection and clustering. It consists of 3 sub-views:

- Reachability Plot: Positioned at the top, the Reachability Plot is generated by OPTICS. This sequence represents the density-based clustering structure of all contexts. Analysts can generally estimate the number of potential clusters by observing the number of valleys in the reachability plot.
- Data Projection View: Positioned in the lower left, this view displays the 2-D projections and colors of data entries in $C = \{C_t\}$, representing the results of Step 2 in the data processing workflow. Here, analysts can observe the distribution of the projected data entries and the proximity of their corresponding colors.
- Context Projection View: Positioned in the lower right, this view shows the 2-D projection of entire contexts (i.e, the 2-D projection of $vec(C_t)$ using t-SNE), allowing analysts to see how contexts are distributed on a 2-D plane. This view serves as a reference for clustering.

The Reference View responds to interactions in the Context View. After re-clustering, different marks will represent different clusters in the Context Projection View, allowing analysts to examine clustering results. When contexts are brushed and selected in the Context View, they will also be highlighted in the Data Projection View and the Context Projection View.

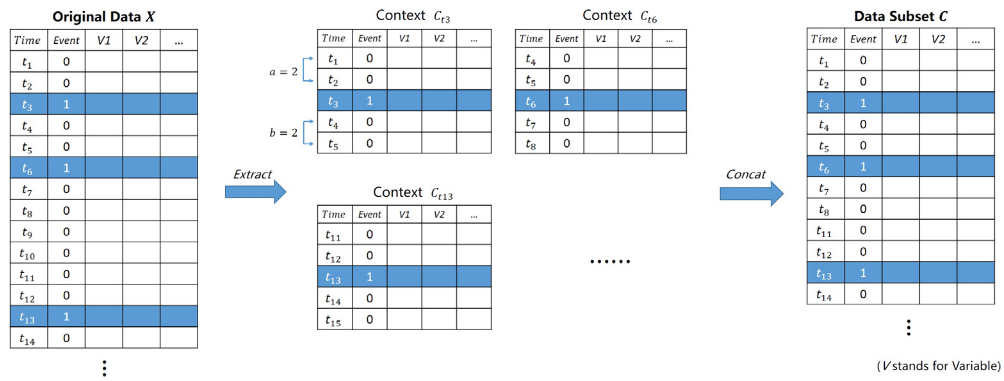


**Fig. 6.** For each event occurrence recorded in the original data $X$, an event context can be extracted. The extracted contexts are time slices with the same length, columns, and time intervals. Then, all the data entries included in event contexts form the data subset $C$.

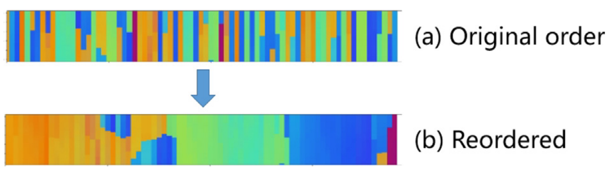


**Fig. 7.** A cluster of contexts before and after reordering.

### 6.3. Parallel Coordinates View (D)

The Parallel Coordinates View (Fig. 1-D) enables analysts to examine the original variables in detail, which is important for pattern interpretation (T3). In this view, the polylines are displayed in projected colors of corresponding data entries. Analysts can also choose to color the lines by single variable or by time from event using the dropdown menu at the top (Fig. 11).

The Parallel Coordinates View and the Context View will respond to each other (Fig. 12). In the Parallel Coordinates View, analysts can brush on the axes to restrict variable values. Brushing multiple axes allows for multiple restrictions. Therefore, the Parallel Coordinates View assists analysts in filtering event contexts under specific situations during exploration.

### 6.4. Timeline (A) and Space View (E)

The Timeline (Fig. 1-A) and the Space View (Fig. 1-E) are used to display temporal and spatial information respectively. The Timeline has 2 parts. The upper part is a histogram showing the distribution of events over time. The lower part is draggable, with all the events and their information marked at their time of occurrence.

The Space View is used to draw trajectories when spatial variables are included in the event contexts. As shown in Fig. 13, we plot trajectories using scatter points, and the colors of the trajectories correspond to those in the visual representations of contexts. Within a trajectory, the points increase in size over time.

When brushing happens in the Context View, the selected event contexts will be highlighted on the Timeline, and their trajectories will be plotted in the Space View.

## 7. Case study

In this work, our goal is to propose a general framework which is applicable in different event context analysis scenarios. Therefore, in this section we will apply the framework to datasets from two different domains. The first dataset consists of data collected from a football match, where we identify and explore various passing context patterns. The second dataset contains driving records, where we explore driving patterns in harsh braking contexts.

### 7.1. Case 1: Passing event contexts in a football match

In Case 1, we demonstrate how analysts can use our framework to identify passing event contexts with similar patterns (T2) and analyze the meanings of these patterns in terms of player behavior and tactical considerations (T3).

The data are collected from a match in the 2018–19 Bundesliga season. This data records the positions of all players and the ball on the field every 40 ms, resulting in around 3.5 million records throughout the game. Additionally, the dataset logs ball possession status and specific events such as passing and shooting. Based on the original records, some calculated metrics are also included in the dataset, such as positions within the team space. The team space is a spatial reference system proposed by Andrienko et al. (2019) for football analysis. The team space represents the relative placements of the players within a team. In the team space, the coordinates of an object represent its relative position within the current team formation. In this case, the Space View represents the team space (as shown in Fig. 14). Other quantitative indicators in the dataset include measures of threat and pressure, defined by Andrienko et al. (2017). Pressure assesses the defensive team's force exerted on the ball or attacking players.

In the football match, we focus on the abundant passing events. We take 25 data entries both before and after a pass at 40-millisecond interval, forming a 2-second context around the pass. We extract 130 passing event contexts of one team, and import them into the system for exploration. We include variables such as ball speed, ball height, pressure on the ball, pressure on the attacking team, threat to the defensive goal, potential threat to the attacking team's goal, distance of the ball from the defensive goal, and the ball's location within the team space.

#### 7.1.1. Determining the number of clusters

We first need to choose an appropriate number of clusters. This decision can be guided by the Reference View and the current clustering result displayed in the Context View. In this case, the reachability plot shows the existence of at least 3 valleys (Fig. 15-a). The Context Projection View also suggests that the contexts may be divided into roughly 3 groups. Therefore, we begin by selecting 3 clusters and perform clustering within the system, yielding the results shown in Fig. 15-b. With 3 clusters, k-means produces results that align well with context projection.

However, when observing the Context View, analysts may find that this result is too general, with too many contexts in the same

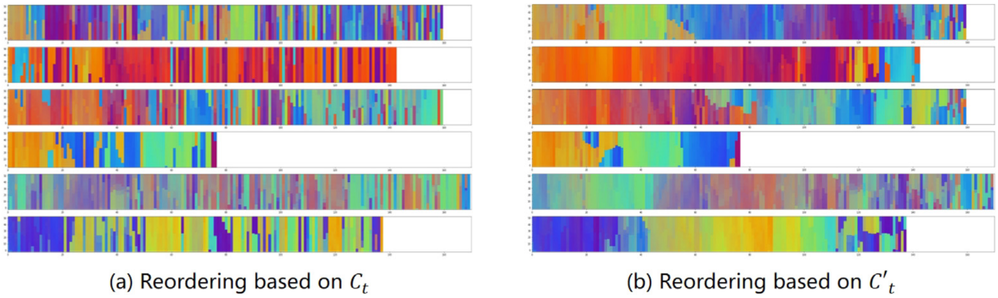


**Fig. 8.** Comparison of reordering based on original contexts $C_t$ and their projected forms $C'_t$.

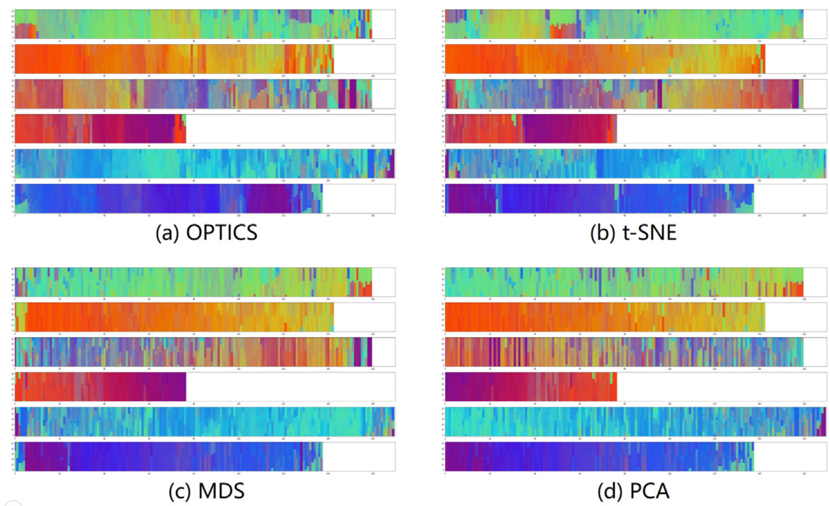


**Fig. 9.** Comparison of reordering results produced by OPTICS, t-SNE, MDS, and PCA.

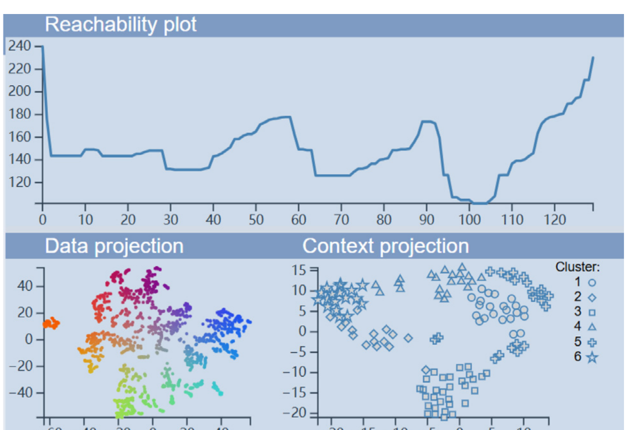


**Fig. 10.** The Reference View. This view consists of 3 sub-views: the Reachability Plot (top), the Data Projection View (lower left), and the Context Projection View (lower right).

cluster. We can increase the number of clusters as needed and compare the results. Re-clustering can be done directly within the system. Fig. 15-c shows the results when the number of clusters is set to 5. Compared to 3 clusters, increasing the cluster count leads to more detailed groupings. Still, the choice of cluster number has no universally correct answer. It is a decision made by the analyst based on a comprehensive consideration over the Reference View and the Context View. In this case, we proceed with 5 clusters.

### 7.1.2. General features of clusters

Once the number of clusters is determined, analysts can explore the general characteristics of each cluster through interactions. For example:

Cluster 1 (Fig. 16-a): Passing contexts in this cluster share a common trait. They occur at the farthest distance from the opponent's goal, closest to the team's own goal. Although the team is in possession, they are not yet in a position to attack and are, in fact, facing considerable pressure from their opponents.

Cluster 2 (Fig. 16-b): This cluster has a positional characteristic, with the ball trajectories generally located in the rear-right section of the team space. These passes are closer to the opponent's goal than those in Cluster 1, but they are still relatively distant. The coloring reveals that contexts in Cluster 2 are not entirely homogeneous. There are more distinct patterns in Cluster 2, which will be explored in the following.

Cluster 3 (Fig. 16-c): Compared to Cluster 2, passes in Cluster 3 are closer to the opponent's goal. Contexts in Cluster 3 can also be further divided to see more specific sub-patterns.

Cluster 4 (Fig. 16-d): Most of the passes in this cluster are concentrated in the front of the team space, indicating that these are forward passes close to the opponent's goal. Some passes within this cluster are very close to the goal, suggesting that the team is in an active offensive phase, seeking scoring opportunities.

Cluster 5 (Fig. 16-e): These passes are concentrated in a specific location within the team space, positioned at a moderate distance from the opponent's goal. Cluster 5 also contains distinct sub-patterns.

In summary, we can broadly describe the characteristics of passing event contexts within each cluster. However, there are more specific patterns within clusters. The next step is to brush

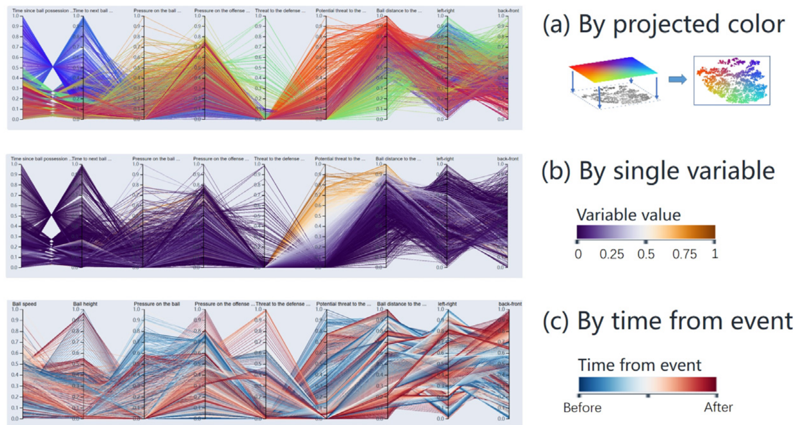


**Fig. 11.** Analysts can choose to color the lines in the parallel coordinates view by projected color, a single variable, or time from event.

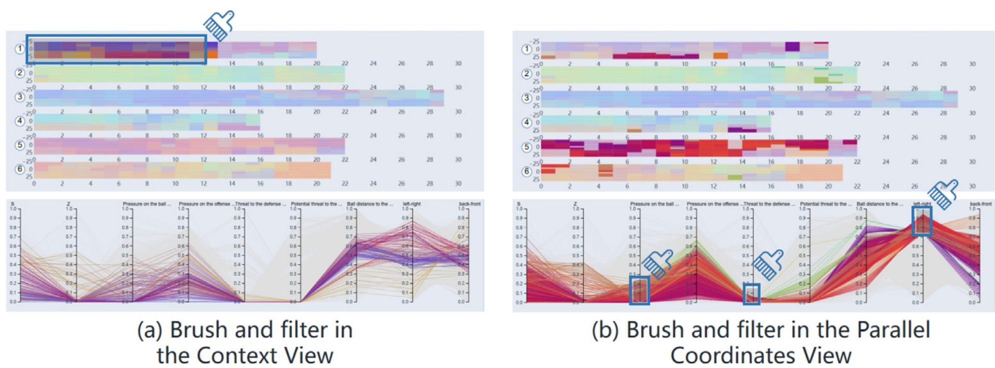


**Fig. 12.** The Context View and the Parallel Coordinates View will respond to each other. Brushing and filtering can be done in both views.

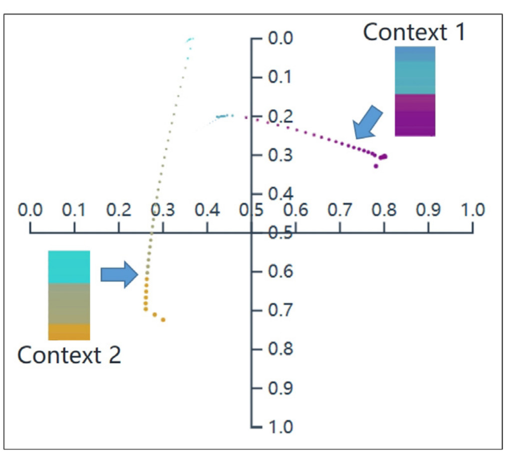


**Fig. 13.** Two trajectories plotted in the Space View and their corresponding event context visual representations.

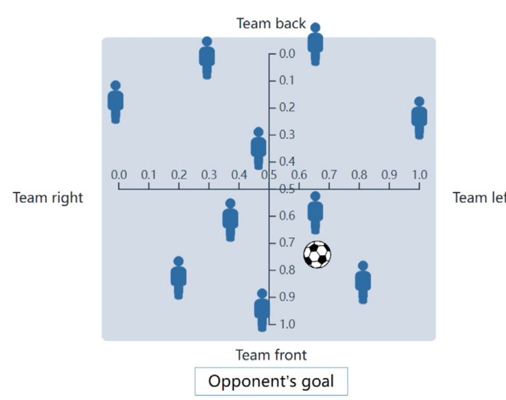


**Fig. 14.** The Space View represents the team space in Case 1. The orientation of the team is as indicated in the figure.

and explore these similar contexts individually and investigate the behavioral patterns they represent. As mentioned, our system applies OPTICS to reorder the contexts within each cluster, arranging them by similarity. This allows analysts to brush sub-patterns directly within each cluster.

*7.1.3. Exploring specific patterns*

We can take a set of contexts from Cluster 3 as an example (Fig. 17). These six contexts share the same pattern. Their projected colors shift from purple-gray to purple-red. This pattern may interest analysts as it indicates variable-level changes around the pass. After brushing these contexts, we observe similar passing trajectories in the Space View. In the team space, these passes are generally horizontal, directed towards the left side of the team. By changing the coloring based on different variables, we find that two other variables show significant changes around the pass: the ball speed and the pressure on the ball. As shown in Fig. 17, ball speed increases significantly after the pass, while defensive pressure on the ball decreases substantially. This suggests that these passes are intended to relieve defensive pressure by transferring the ball horizontally.

In Cluster 2, we can also identify a set of unique passing event contexts (Fig. 18). Their visual representations show a color transition from cyan to light brown. After brushing these contexts, we

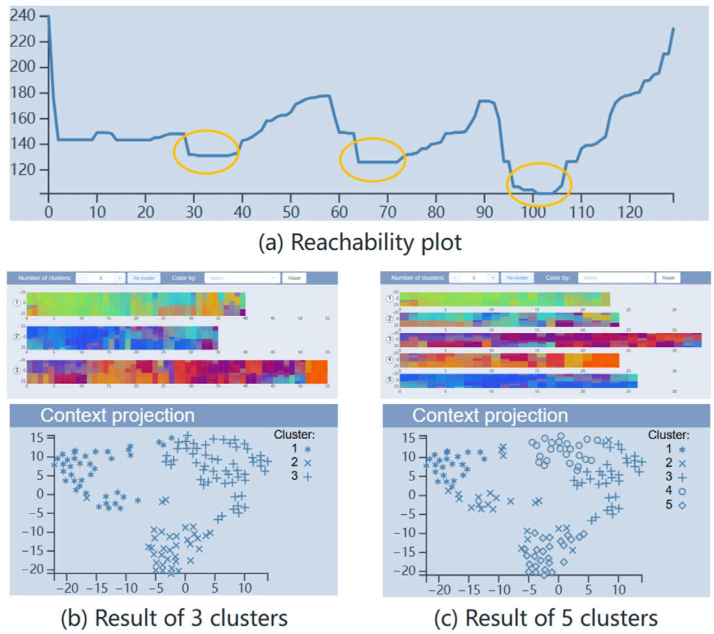


**Fig. 15.** Clustering results of passing event contexts with different cluster number.

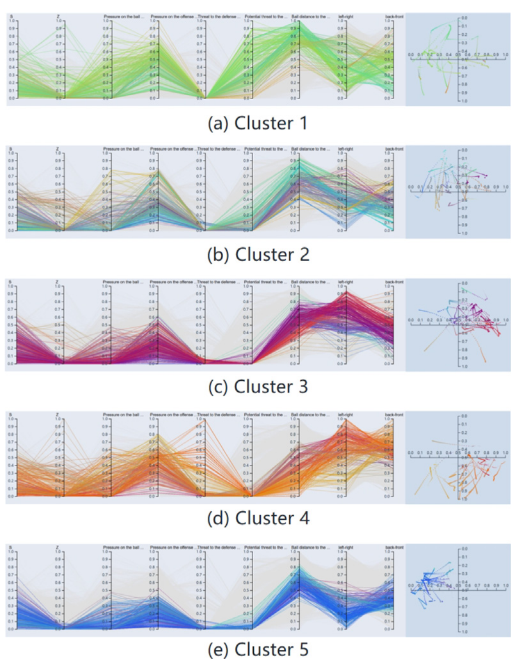


**Fig. 16.** The Parallel Coordinates Views and the Space Views of cluster 1 to 5.

observe similar trajectories in the Space View. It becomes apparent that these are long-range passes from the back of the team's formation towards the front. When we adjust the coloring based on different variables, we notice several variables change as the pass occurs. As illustrated, the ball speed increases significantly after the pass. Notably, the ball's height also rises considerably, indicating that these are powerful long passes. Additionally, by examining the distance from the ball to the opponent's goal and the potential threat to the team's own goal, we see that these passes start near the team's own goal, and are directed away from it. In most instances, the threat to the team's own goal decreases after the passes. Therefore, we can deduce the behavior pattern under these passing contexts. These are long passes happened

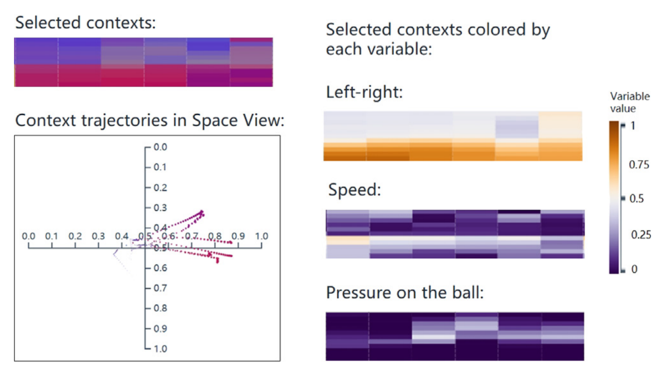


**Fig. 17.** A set of passing event contexts selected from Cluster 3.

near the team's own goal, directed towards the opponent's goal. The likely objective of these passes is to get the ball away from the team's own goal and shift towards an offensive stance.

#### 7.1.4. *Pattern exploration starting from specific situations*

In the previous section, we demonstrate an exploration process that begins from the Context View. Analysts first identify visual patterns of interest and then investigate the underlying behavior via interactive analysis. However, this is not the only exploration path. If analysts are interested in a specific situation, they can start from the Parallel Coordinates View by restricting variable values to filter particular situations.

For instance, an analyst may be interested in contexts where the team has a high chance of scoring (Fig. 19). The analyst could first set the distance to the opponent's goal to a range of 0–0.3, representing close proximity to the goal. To narrow down further, the analyst could set the pressure on the ball to a low range, indicating minimal defensive pressure on the ball. This setting filters a situation where the team is close to the opponent's goal with low defensive pressure, maximizing the scoring opportunity. Most of these filtered passing event contexts appear in Cluster 4. Returning to the Context View, the analyst can then compare these contexts to examine their similarities and differences (Fig. 20):

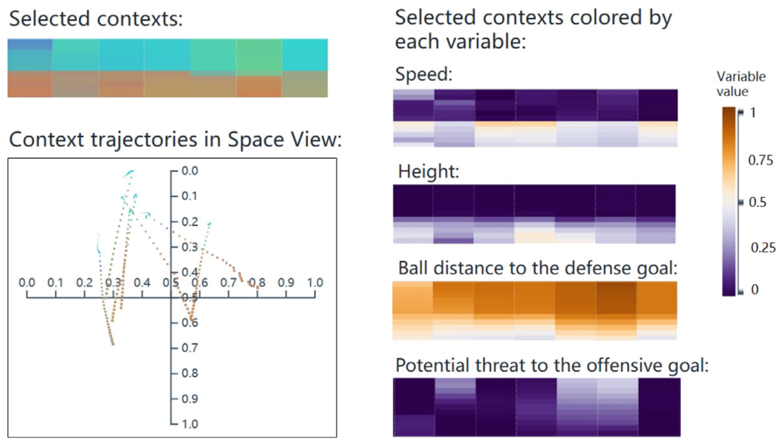


**Fig. 18.** A set of passing event contexts selected from Cluster 2.

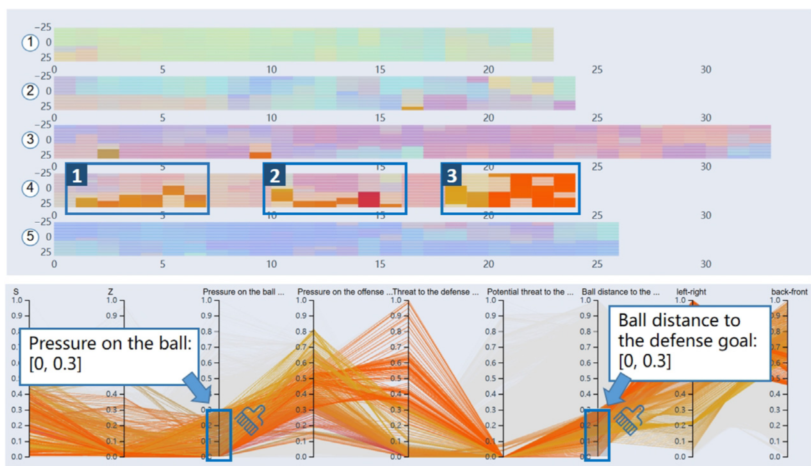


**Fig. 19.** The filtered contexts where the team has a high chance of scoring.

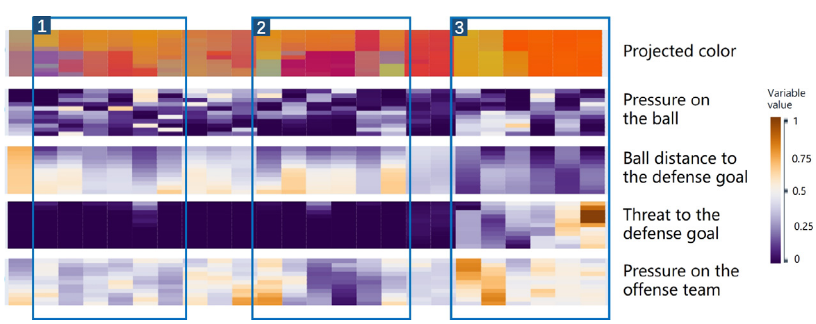


**Fig. 20.** Exploration of the filtered contexts by variate in Cluster 4.

The first group of contexts shows high pressure on the ball before the pass, indicating that the player successfully bypassed defenders with the pass, reducing the pressure level. In the second group, initial defensive pressure is lower, and the passes bring the ball significantly closer to the opponent's goal while avoiding defenders. The third group of contexts occurs very close to the opponent's goal, with the ball in the closest possible position throughout the match and with the highest threat level towards the goal. Clearly, this group of passing event contexts represents the most offensive scenarios in the entire match.

If the analyst is interested in contexts where the team's own goal is under threat (Fig. 21), they can set the distance to the opponent's goal to a range of 0.7–1.0, indicating proximity to the team's own goal. The analyst can narrow down further by setting a high threat level to the team's goal, filtering a situation where their goal faces serious danger and has a risk of conceding. These filtered contexts appear in Cluster 1. After returning to the Context View and coloring by the threat to their own goal, two main patterns emerge (Fig. 22):

In the first two groups, the passes help clear the ball away from the goal, reducing the threat of conceding. In the third group, the players choose to pass towards their own goal, which increases the potential threat but decreases the immediate pressure on the ball. This set of contexts illustrates a defensive tactic where the team responds to high pressure near their goal by passing back to evade opponents.


**Fig. 21.** The filtered contexts where the team's own goal is under threat.

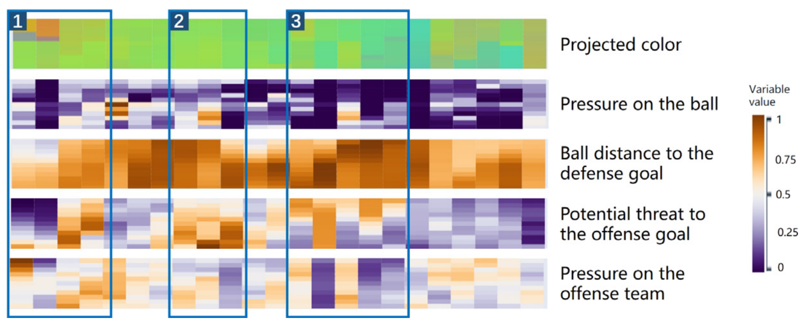


**Fig. 22.** Exploration of the filtered contexts by variate in Cluster 1.

In this case study using football data, we demonstrate how to use the framework we proposed to explore patterns within passing event contexts. We show how analysts can interactively perform clustering and exploration within the system. Given that football players typically cooperate in formations during a match, it is likely that the contexts with shared patterns can reveal the team's strategy in similar situations, or more specifically, the coordination between specific players within the team.

### 7.2. Case 2: Harsh braking event contexts in driving records

In Case 2, we focus on the harsh braking event contexts in driving records. Compared to Case 1, this case requires constructing event contexts from datasets with uneven time intervals. Additionally, we analyze contexts with longer time spans and more complex patterns. We identify harsh braking event contexts with similar patterns (T2) and analyze the meanings of these patterns in terms of driver behavior and responses to road conditions (T3).

We use data from the records of a truck driving in Greece over a period of 3 months. The data include vehicle information such as model and fuel tank size, as well as movement data like location, direction, speed, engine status, mileage, and fuel level. Driving events such as harsh braking, sharp turns, and rapid accelerations were recorded at corresponding timestamps.

This data is previously used in the work of Chen et al. (2019), in which the authors discussed the length of event contexts with domain experts, and decided to choose ±5 minutes. We adopt this choice and construct 10-minute long event contexts with 60-second time intervals by averaging the original data entries. In this case, we focus on the contexts of harsh braking events. In total, we extract 300 harsh braking event contexts and import them into the system for analysis. The included variables are latitude, longitude, altitude, speed, and steering angle.

In the Reference View (Fig. 23), the reachability plot shows around 6 distinct valleys. Here we skip the step of determining the number of clusters which has been discussed in Case 1, and choose to have 6 clusters. Given the larger number and longer time span of the harsh braking event contexts (10 min), the visual representations show more complex patterns than those of passing event contexts.

There are some interesting contexts in Cluster 5 (Fig. 24-a). In these contexts, the visual representations transition from purple to cyan. After coloring by different variables, the following pattern is revealed: the vehicle remains almost stationary during the first half of the context, then gradually begins to move. After a harsh braking event, the vehicle maintains a high speed and goes straight while the altitude gradually decreases. This pattern suggests that the driver starts, brakes harshly upon approaching a slope, and then maintains a high speed while going downhill. This driving pattern recurs frequently, and the paths are quite consistent. We can infer that this might be a slope that the truck often drives through on its routine route.

In some event contexts in Cluster 1 (Fig. 24-b), there are more intricate patterns which may interest the analyst. After coloring by different variables, it becomes evident that this pattern is most closely related to changes in the steering angle. In the early part of the contexts, the driver begins moving forward while turning the steering wheel to the right. After a harsh brake, the driver quickly steers left, then switches back to the right, and finally returns the wheel to the left at the end of the context. This pattern represents the driver's behavior on a winding road. The reason for the harsh brake is also revealed. To go through the road, the driver needs to adjust the steering angle frequently. To safely complete these turns, the driver brakes to decelerate.

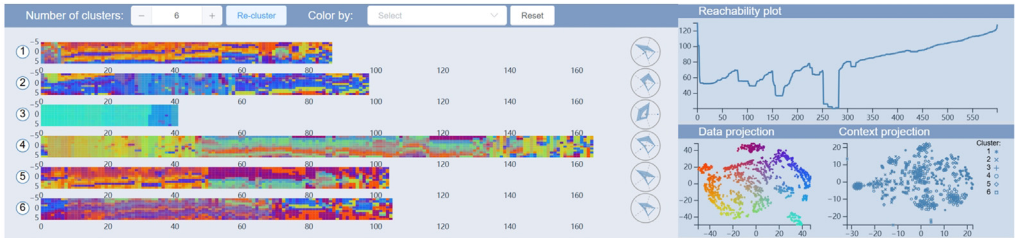

**Fig. 23.** Harsh braking event contexts in driving records.

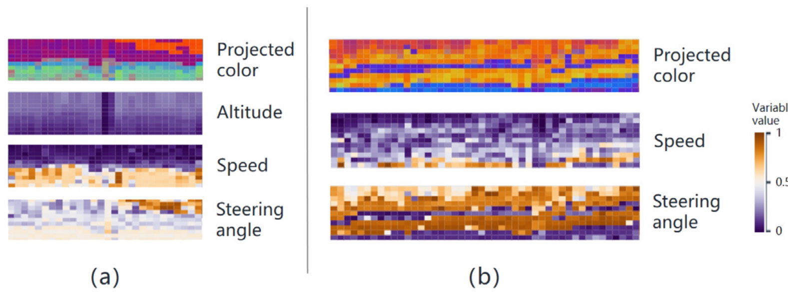


**Fig. 24.** (a) A set of harsh braking event contexts from Cluster 5. (b) A set of harsh braking event contexts from Cluster 1.

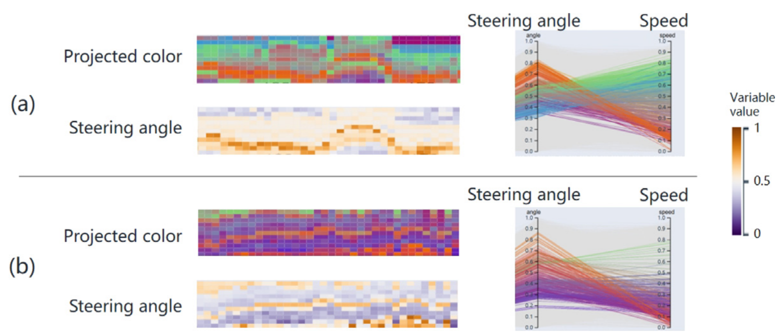


**Fig. 25.** Harsh braking event contexts with patterns related to steering angle in Cluster 4 and Cluster 6.

In Clusters 4 and 6, similar patterns related to steering angle can be observed. The contexts in Cluster 4 (Fig. 25-a) represent a driving sequence involving a left turn, a harsh brake, and a right turn. The contexts in Cluster 6 (Fig. 25-b) show an even more complex pattern where the driver switches between left and right turns frequently. Examining the relation between steering angle and speed in the Parallel Coordinates View reveals a consistent behavior: when the steering angle is large (with values approaching 0 or 1), the driver tends to decelerate. Conversely, when going straight (with angle values near 0.5), the speed is often higher. This pattern aligns with common driving practices, since reducing speed during turns and increasing it on straight paths helps maintain safe control over the vehicle.

In this case study of driving data, we focus on harsh braking event contexts with a longer time span. These event contexts show more complex and diverse patterns. Using the analysis framework we proposed, we are able to identify and analyze the underlying patterns in these more intricate contexts. The framework assists analysts in identifying groups of event contexts that occur under similar road conditions and enables them to observe the driving behavior patterns associated with these contexts.

## 8. Expert study

In order to examine the effectiveness of our analysis framework, we conducted an expert study with five data analysis experts. In this section, we will present the feedback from experts on our framework and visual analysis system.

### *8.1. Participant background*

We invited five experts with over five years of experience in data analysis to participate in our study. Three of these experts (denoted as E1, E2, and E3) had at least three years of experience in visual analytics, and had also developed visual analysis systems. E1 had additional experience in event analysis, including anomalous event and user interaction event analysis. The other two experts (denoted as E4 and E5) had no experience in visual analytics.

### *8.2. Study procedure*

The study began with a 20-minute tutorial introducing the concept and definition of event context and our method for visualizing it. We also introduced our visual analysis system and demonstrated how to interpret patterns within a set of event contexts using the system.

Next, we invited participants to use our visual analysis system and complete four operational tasks:

**Task 1:** Perform real-time clustering with the help of the Reference View.

**Task 2:** Select a single event context of interest, explore and identify its feature.

**Task 3:** Select a group of event contexts with common temporal patterns, explore and interpret the patterns.

**Task 4:** Start from a situation of interest, filter and explore the event contexts under this situation.

Finally, we conducted interviews with each expert. We asked two sets of questions, one about our visual analysis system and the other about our overall analysis framework.

### 8.3. Operational tasks

For the operational tasks, we used the same dataset as in Case Study 1, which included 130 passing event contexts from a football match. We will detail the experts' performance and feedback in each task.

In **Task 1**, experts were asked to perform real-time clustering with the help of the Reference View. This task aimed to test whether our design for real-time clustering is feasible, and whether sufficient information was provided for analysts to determine the number of clusters. All the experts were able to understand the information provided by the Reference View, and completed clustering within three minutes. We recorded the number of clusters chosen by the experts and asked for the reasons of their choices. Among the experts, E5 referenced the Reachability plot and the Context Projection View, and chose to directly select three clusters. The other experts preferred to compare the clustering results of different cluster numbers before making a decision. The number of clusters chosen varied from three to seven, reflecting their individual preferences. E1 considered *"cluster numbers within a certain range are all acceptable."*

In **Task 2**, we asked experts to select an event context of interest from the context view and analyze its feature. This task aimed to assess if analysts understood the definition and visual representation of event contexts. We found that experts tended to choose contexts with significant color variation. E3 specifically chose a context with the greatest color difference by checking the distribution of colors in the Data Projection View. We observed that experts preferred to use the Parallel Coordinate View in this task, and were able to interpret the reason of color changes from the perspective of at least one variable.

In **Task 3**, we further invited experts to select a group of event contexts with common temporal patterns and interpret these patterns. This task aimed to test whether our system could effectively support pattern identification and interpretation. We found that experts increasingly used the single-variable coloring function in the Context View, as it could help observing temporal patterns across multiple contexts based on a specific variable. In this task, experts showed varying interests and analysis paths. For example, E2, a football enthusiast, showed great interest in this task and enjoyed integrating domain knowledge into the exploration. During the analysis, E2 identified three different passing patterns: backward passes, long passes, and offensive plays. E3 selected a set of orange contexts and discovered that they were distributed distinctly in the Context Projection View. E3 sought to explore what made these contexts unique, and found out that these were passes occurred closest to the opponent's goal, posing the greatest threat. E3 also referred to the Timeline to identify the sequential relationships between these offensive passes. Overall, every expert was able to identify event contexts with similar patterns, and was able to interpret the patterns using various functions in the system.

In **Task 4**, we asked the experts to think of a specific situation and use the Parallel Coordinates View to filter event contexts under that situation. This task aimed to assess if our system could support event context analysis starting from a specific situation. The experts thought of various situations and successfully filtered and analyzed the contexts. For example, E2 first filtered a situation involving offensive plays by restricting the pressure on the ball and the threat to the defensive goal. E2 then restricted the left–right variable in the team space to observe attacks from both the left and the right side of the team. E2 concluded that the team was more proficient in attacking from the right side. E3 was curious about a situation where the football was close to the opponent's goal but did not pose any threat. E3 found that there were indeed passes that occurred under such situation. E1, E2 and E5 all noticed that the event contexts they filtered under a specific situation were mostly clustered together and located close to each other. They concluded that our method effectively groups similar event contexts together.

### 8.4. Expert feedback

At the end of the expert study, we conducted interviews with each expert. We organize the expert feedback collected from these interviews as follows.

#### 8.4.1. On the visual analysis system

We first asked experts a set of questions related to the usage of the visual analysis system. The experts agreed that the system enabled them to perform clustering, identify event contexts with similar patterns, and further explain these patterns.

We also asked the experts which views or functions in the system they found most helpful. Almost all experts chose the Context View and the Space View, describing them as *"intuitive"* or *"rich of information"*. Regarding the interpretation of temporal patterns within contexts, we observed two preferences among the experts. E1, E4, and E5 preferred using the Parallel Coordinates View, while E2 and E3 favored single-variable coloring function in the Context View.

Regarding the system's usability, the experts gave positive feedback. However, we noticed that E4 and E5, who had no prior experience in visual analysis, required more time to learn the system and raised more questions during the study. E4 mentioned that *"for those unfamiliar with visual analytics, understanding and using the numerous interactive features in the system requires learning."* E5 mentioned that *"there are many views in the system thus I sometimes get lost"*. We agree with E4's opinion that *"tutorials and demonstrations of the system are crucial."* Regarding our system, we consider providing detailed tutorials and demonstrations essential, especially for users unfamiliar with visual analytics.

#### 8.4.2. On the analysis framework

We then asked the experts a set of questions about our analysis framework, including their opinions on our definition of event context, visual representations, and analysis methods.

The experts approved our definition of event context and the design of its visual representation. E4 commented that *"the visualization design is rational and effective; though it is a little challenging, it can be understood with training and explanation."*

The experts also agreed that our method for analyzing event contexts effectively supports the tasks of pattern identification and interpretation. E3 commented that *"the method successfully grouped similar patterns together, enabling analysts to clearly see the temporal patterns."* E1 commented that *"it provides overviews as well as details."*

#### 8.4.3. Advantages of contextualized analysis of events

We discussed with experts about the advantages of analyzing events from a contextualized perspective. We derive from the discussions that the main advantage of a contextualized perspective is it introduces the temporal dimension into individual events. This enables causal and situational analysis of events, and thus enhances the practical applicability.

E1, E4, and E5 mentioned that without a contextualized perspective, it is impossible to analyze the cause, course, and consequence of an event. E2 approved of the necessity of analyzing the context of events, as every event occurs under a situation. From this perspective, the preceding of the context represents the situation faced, the event itself represents the response to this situation, and the succeeding of the context represents the result of that response. Therefore, understanding the whole context is essential in many analysis scenarios.

Furthermore, the experts comments that the strengths in causal and situational analysis make the contextualized perspective beneficial in practical applications. For instance, E1 mentioned that *"in many analytical scenarios, the context allows analysts to see more tactics or strategies."* E2 mentioned that *"situational analysis within contexts is particularly beneficial for tasks such as capability assessment for sports players and responsibility determination for traffic accidents."*

#### 8.4.4. Suggestions for improvement

At last, we collected suggestions for improvement from the experts. For the system, E3 and E5 suggested for more interactions starting from the Timeline which would help analysts better understand the connections between sequential events. E4 suggested adding more operational hints to the system interface to guide users. Regarding the analysis method, E2 and E5 believed that incorporating automated result generation capabilities could make our framework suitable for users with less analytical experience. We agree with these suggestions and regard them as valuable directions for future work.

## 9. Discussions

In this section, we will discuss some of the design choices, limitations, and future directions of our work.

### 9.1. Color usage

Our visualizations presented have a high reliance on colors. Color serves as the primary visual channel in our design. We choose to use colors due to the effectiveness and scalability of 2-D colormaps in representing 2-D information (Bernard et al., 2015). However, this design choice imposes certain requirements on the color resolution capabilities of the analysts and their equipment. Our approach relies on users' ability to identify and compare the differences in colors. Therefore, it has certain applicability limitations.

### 9.2. Dimensionality reduction methods

Our approach has a high reliance on dimensionality reduction methods, which are applied multiple times at different levels of data processing. During the analysis process, it is essential for analysts to have a understanding of the dimensionality reduction method they use. For example, the default method used in this work, t-SNE, is a neighborhood-preserving dimensionality reduction algorithm. When performing similarity analysis based on projected colors, analysts must be aware that only colors relatively close are comparable. Otherwise, incorrect similarity judgments may occur.

### 9.3. Scalability

Although the increase in data volume reduces the efficiency of the statistical methods we use, the primary scalability challenges in our work stem from visualizations. The increase in data volume can slow down the response speed of the visual analysis system and cause overlaps in views such as the Context View and the Parallel Coordinate View.

For individual event contexts, the Context View can clearly display the color changes in contexts with fewer than about 30 data entries. For contexts with more data entries, it is recommended to increase the time interval and reduce the number of data entries. Another feasible solution, as we adopted in Case 1, is to include all data entries in data processing, but to sample within the contexts when drawing them in the system.

As for the overall data volume, we tested the system with datasets of varying sizes. We observed that the primary limiting factor in the system's performance is the number of contexts. Significant congestion and overlap occur in the Context View when the total number of contexts exceeds about 500. As a solution to handling large amounts of contexts, we suggest clustering the contexts and importing each cluster into the system separately for analysis.

### 9.4. User involvement

In this work, we leave substantial room for the analyst's involvement within our framework. As for data processing, the decisions on the time interval, the length of contexts, as well as the number of clusters, are made by analysts. This is because the selection of these parameters can vary greatly depending on the specific analytical scenario.

As for visual analysis, we also do not prescribe a fixed analytical process for analysts. This is because we observed that the process of exploring event contexts is typically open-ended and non-linear. From our expert study, we also observed that the analysts have various interests, and their interests tend to emerge gradually during their exploration. Therefore, we only suggest that analysts first complete the clustering. After that, as demonstrated in Case 1, analysts can either start from interesting patterns in the Context View or from filtering specific situations using the Parallel Coordinates View.

Overall, in this work we do not aim at providing an automated or structured solution to users. Instead, we choose to leave more decision-making space for users to ensure flexibility and adaptability. However, this choice also results in limitations in our target user group. Based on expert suggestions, in future work, we will work on incorporating more automation and intelligence to reduce our reliance on user knowledge.

## 10. Conclusion

In this study, we present a framework for the visual analysis of multivariate event contexts, which consists of a design of visual representation, a data processing workflow, and a context-centered visual analysis system. By incorporating a contextualized perspective in event analysis, the framework facilitates the identification and interpretation of temporal patterns within and across events. We present case studies using real-world datasets from two different domains and an expert study conducted with experienced data analysts to demonstrate the applicability and effectiveness of our framework.

## CRediT authorship contribution statement

**Lei Peng:** Writing – original draft, Visualization, Software. **Ziyue Lin:** Visualization. **Natalia Andrienko:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Gennady Andrienko:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Siming Chen:** Writing – review & editing, Supervision, Methodology, Conceptualization.

## Ethical approval

Informed consent was obtained from all participants prior to their involvement in the study. Signed consent forms are on file and available upon request.

## Declaration of competing interest

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.

## Acknowledgments

The authors want to thank the reviewers for their suggestions. This work is supported by Natural Science Foundation of China (NSFC No. 62472099 and No. 62202105), 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 project *CrexData* (grant agreement No. 101092749).

## References

Andrienko, N., Andrienko, G., 2013. A visual analytics framework for spatio-temporal analysis and modelling. Data Min. Knowl. Discov. 27, 55–83.
Andrienko, N., Andrienko, G., 2023. It's about time: Analytical time periodization. In: Computer Graphics Forum. Wiley Online Library, e14845.
Andrienko, N., Andrienko, G., 2024. Exploring relationships between events in context.
Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., Hecker, D., Weber, H., Wrobel, S., 2019. Constructing spaces and times for tactical analysis in football. IEEE Trans. Vis. Comput. Graphics 27, 2280–2297.
Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., Von Landesberger, T., Weber, H., 2017. Visual analysis of pressure in football. Data Min. Knowl. Discov. 31, 1793–1839.
Andrienko, G., Andrienko, N., Heurich, M., 2011. An event-based conceptual model for context-aware movement analysis. Int. J. Geogr. Inf. Sci. 25, 1347–1370.
Andrienko, N., Andrienko, G., Shirato, G., 2023. Episodes and topics in multivariate temporal data. In: Computer Graphics Forum. Wiley Online Library, e14926.
Arunkumar, A., Pinceti, A., Sankar, L., Bryan, C., 2022. Pmu tracker: A visualization platform for epicentric event propagation analysis in the power grid. IEEE Trans. Vis. Comput. Graphics 29, 1081–1090.
Bartolomeo, S.Di., Zhang, Y., Sheng, F., Dunne, C., 2020. S equence b raiding: Visual overviews of temporal event sequences and attributes. IEEE Trans. Vis. Comput. Graphics 27, 1353–1363.
Bernard, J., Barth, C.M., Cuba, E., Meier, A., Peiris, Y., Shneiderman, B., 2024. Ivesa-visual analysis of time-stamped event sequences. IEEE Trans. Vis. Comput. Graphics.
Bernard, J., Steiger, M., Mittelstädt, S., Thum, S., Keim, D., Kohlhammer, J., 2015. A survey and task-based quality assessment of static 2D colormaps. In: Visualization and Data Analysis 2015. SPIE, pp. 247–262.
Bernard, J., Wilhelm, N., Scherer, M., May, T., Schreck, T., 2012. Timeseriespaths: Projection-based explorative analysis of multivariate time series data.
Cappers, B.C., van Wijk, J.J., 2016. Understanding the context of network traffic alerts. In: 2016 IEEE Symposium on Visualization for Cyber Security. VizSec, IEEE, pp. 1–8.
Cappers, B.C., van Wijk, J.J., 2017. Exploring multivariate event sequences using rules, aggregations, and selections. IEEE Trans. Vis. Comput. Graphics 24, 532–541.
Chen, S., Andrienko, G.L., Andrienko, N.V., Doulkeridis, C., Koumparos, A., 2019. Contextualized analysis of movement events. In: EuroVA@ EuroVis. pp. 49–53.
Cui, W., 2019. Visual analytics: A comprehensive overview. IEEE Access 7, 81555–81573.
Deng, Z., Chen, S., Schreck, T., Deng, D., Tang, T., Xu, M., Weng, D., Wu, Y., 2023. Visualizing large-scale spatial time series with geochron. IEEE Trans. Vis. Comput. Graphics.
Gotz, D., Stavropoulos, H., 2014. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. IEEE Trans. Vis. Comput. Graphics 20, 1783–1792.
Jin, Z., Guo, S., Chen, N., Weiskopf, D., Gotz, D., Cao, N., 2020. Visual causality analysis of event sequence data. IEEE Trans. Vis. Comput. Graphics 27, 1343–1352.
Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S., 2010. Spatio-Temporal Clustering. Springer.
Liu, D., Weng, D., Li, Y., Bao, J., Zheng, Y., Qu, H., Wu, Y., 2016. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. IEEE Trans. Vis. Comput. Graphics 23, 1–10.
Liu, D., Xu, P., Ren, L., 2018. Tpflow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. IEEE Trans. Vis. Comput. Graphics 25, 1–11.
Nonato, L.G., Aupetit, M., 2018. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. IEEE Trans. Vis. Comput. Graph. 25, 2650–2673.
Perer, A., Wang, F., 2014. Frequence: Interactive mining and visualization of temporal frequent event sequences. In: Proceedings of the 19th International Conference on Intelligent User Interfaces. pp. 153–162.
Shirato, G., Andrienko, N., Andrienko, G., 2023b. Identifying, exploring, and interpreting time series shapes in multivariate time intervals. Vis. Informatics 7, 77–91.
Wang, J., Li, Y., Zhou, Z., Wang, C., Hou, Y., Zhang, L., Xue, X., Kamp, M., Zhang, X.L., Chen, S., 2022. When, where and how does it fail? a spatial–temporal visual analytics approach for interpretable object detection in autonomous driving. IEEE Trans. Vis. Comput. Graphics 29, 5033–5049.
Weng, D., Chen, R., Deng, Z., Wu, F., Chen, J., Wu, Y., 2018. Srvis: Towards better spatial integration in ranking visualization. IEEE Trans. Vis. Comput. Graphics 25, 459–469.
Xie, X., Wang, J., Liang, H., Deng, D., Cheng, S., Zhang, H., Chen, W., Wu, Y., 2020. Passvizor: Toward better understanding of the dynamics of soccer passes. IEEE Trans. Vis. Comput. Graphics.
Zhu, S., Sun, G., Shen, Y., Zhu, Z., Xia, W., Chang, B., Tang, J., Liang, R., 2024. Vac 2: visual analysis of combined causality in event sequences. IEEE Trans. Vis. Comput. Graphics.