

Survey on Visual Analysis of Event Sequence Data

Yi Guo, Shunan Guo, Zhuochen Jin, Smiti Kaul, David Gotz, and Nan Cao

Abstract—Event sequence data record series of discrete events in the time order of occurrence. They are commonly observed in a variety of applications ranging from electronic health records to network logs, with the characteristics of large-scale, high-dimensional and heterogeneous. This high complexity of event sequence data makes it difficult for analysts to manually explore and find patterns, resulting in ever-increasing needs for computational and perceptual aids from visual analytics techniques to extract and communicate insights from event sequence datasets. In this paper, we review the state-of-the-art visual analytics approaches, characterize them with our proposed design space, and categorize them based on analytical tasks and applications. From our review of relevant literature, we have also identified several remaining research challenges and future research opportunities.

Index Terms—Visual Analysis, Event Sequences, Visualization

arXiv:2006.14291v1 [cs.HC] 25 Jun 2020

1 INTRODUCTION

EVENT sequence data are found across a vast array of applications and domains. In fields as diverse as computer security, advertising, and healthcare, discrete observations of different types are collected over time and arranged in sequence based on the specific entity for which the event is germane. For example, network logs in computer systems capture timestamped sequences of events (logins, requests, faults, etc.) for specific devices. Similarly, clickstreams used to tailor advertising capture sequences of interaction events for individual users as they navigate websites. Electronic health records, meanwhile, capture events (e.g., diagnoses, procedures) over time for individual patients. The ubiquity of event sequence data reflects both (1) the relative ease with which it can be captured, and (2) the desire to leverage this form of data to gain new insights about real-world systems.

These common goals, however, are challenged by the great heterogeneity that exists within different properties of event sequence data and the types of insights that are sought. For example, event sequences can be high-dimensional (with many event types) or low-dimensional (very few types of events). They can be sparse and irregular over time, or dense and evenly spaced. Events can have zero attributes or many, can be point events or intervals, and can be strictly sequential or occur in parallel. Similarly, the types of analysis tasks can vary widely based on the types of insights one seeks. Are analysts interested in common patterns or rare outliers? Are analysts focused on prediction, or identification of predictive factors for intervention? Are analysts examining a single sequence or comparing across multiple sets of sequences in aggregate? These are just a few examples of the wide variety of data and task challenges which present themselves in event sequence analysis.

These difficult and diverse methodological challenges have motivated a broad range of recent research activities which aim to solve one or more aspects of the event sequence analysis problem. This has led, in turn, to a proliferation of different visual analysis methods and prototypes, each of which has distinct capabilities and advantages in certain contexts. This has resulted in a situation where the state-of-the-art for event sequence data is often difficult to discern. The latest research often offers multiple visual analytics approaches for specific types of challenges. Moreover, the same solution may be effective at addressing difficulties that stem from two or more different challenges. Yet in other cases, open problems remain unaddressed.

The aim of this survey is to provide a comprehensive review and characterization of the state-of-the-art in visual analytics research for event sequence data. Through the collection and analysis of the literature on this topic, we identify key dimensions of the event sequence visual analytics design space. We then use those dimensions, as well as a characterization of different types of event sequence analysis tasks, to organize existing methods and identify common approaches to specific targeted problems. Moreover, we identify areas with little prior work which remain a challenge for future research.

This literature review represents the first (to our knowledge) comprehensive attempt to survey and characterize event sequence data visual analytics methods. In this way, this review promises to help researchers understand key dimensions that unify prior work, how prior research fits together within this complex design space, and which event sequence data analysis challenges remain insufficiently addressed. Moreover, the results can provide value to practitioners as an organized catalog of alternative approaches that are most appropriate for specific types of event sequence data problems. We developed a web-based survey browser¹ to facilitate the exploration of our created taxonomy and reviewed techniques.

- Yi Guo, Shunan Guo, Zhuochen Jin, and Nan Cao are with the Intelligent Big Data Visualization Lab, Tongji University, Shanghai, China. Nan Cao is the corresponding author. E-mail: nan.cao@gmail.com
- Smiti Kaul and David Gotz are with the Visual Analysis and Communication Lab at the University of North Carolina at Chapel Hill, NC, USA. E-mail: smiti, gotz@unc.edu

1. <http://eventvis.idvxlabs.com/>

2 RELATED SURVEYS AND METHODOLOGY

In this section, we first discuss survey papers that are relevant to this work, and then introduce our methodology of selecting papers and creating our taxonomy.

2.1 Related Surveys

This section provides an overview of the surveys that are relevant to visual analysis of event sequence data. Keim et al. [48] proposed a definition and an analytical pipeline for visual analytics, which inspires our formalization of the design space that we discuss later in Section . A prior survey by Sun et al. [99] generalized visual analytics techniques by different data types, among which the review of visual analytics approaches for temporal data is most relevant to our work. Our work, by contrast, focus on a more specific type of temporal data – event sequence data. In addition, some scholars attempted to dive into particular visual forms or visual analytics approaches for a single analytical task that are partially related to our survey. For example, Brehmer et al. [2] formalized the design space for a representative form for visualization of event sequence data – timeline-based visualizations. Jentner and Keim [43] reviewed visualization and visual analytics methods for exploring frequent patterns. Given the broad application of event sequence data, we also notice a larger group of surveys linked to applications where event sequences are commonly collected, such as social media data [116], smart manufacturing [131], and anomalous user behaviors [92]. Different from existing work that summarizes techniques for a particular visualization, visual analytical task, or application related to temporal event sequence, our work aims to provide a more holistic overview of the visual analytics approaches for all types of event sequence data so as to benefit practitioners from a wider range of applications.

2.2 Survey Methodology

This survey aims to obtain an overview of existing visual analytics techniques that are developed for event sequence data. To construct a structured and comprehensive taxonomy, we start by formalizing a design space for developing visual event sequence analysis tools (discussed in Section 3.1). In particular, we leverage the conventional visual analytics pipeline [48] followed by most visual analytics techniques, which revolves around four key components: data, model, visualization, and knowledge. Since deriving knowledge from model and visualizations can be subjective and difficult to standardize, we exclude knowledge inference from the scope of our design space. In addition, user interaction that links the components throughout the pipeline is also indispensable in the visual analytics process. These considerations led to our final proposed design space with with the following four dimensions: *data scales*, *analysis techniques*, *visual representations*, and *interactions*. For each dimension, we further enumerate all alternatives as we conduct our review of the existing studies.

We collect relevant papers from visualization journals and conferences. We followed two main approaches when collecting the papers: reference-driven and search-driven selections. For the reference-driven selection, we utilized a core set of state-of-the-art techniques in this topic known to us in advance as a starting point, and extended the range of work by going through cited and citing publications.

For the search-driven selection, we went through two rounds of paper collection. The first round involves a coarse search of event sequence analysis and visualization techniques from high impact conferences and journals in the field of information visualization and data mining. In particular, we select six visualization conferences: IEEE VAST, IEEE InfoVis, ACM CHI, ACM IUI, EG/IEEE EuroVis and IEEE PacificVis, three visualization journals: IEEE TVCG, IEEE CG&A, Computer Graphics Forum, four data mining conferences: NeruIPS, WWW, ACM SIGKDD, ICML and two journals: IEEE TKDD, ACM TIST. We used two search queries: "event sequence" AND "analysis"; "event sequence" AND "visualization" to collect papers broadly, then reviewed the abstracts and full texts to finalize our selection. We labeled each work with their correspondence in each dimension respectively. Note that event sequence analysis techniques are only labeled in the first two dimensions. In addition, according to Keim et al. [48], the choices of analysis methods, visual representations, and interactions depend on the analytical tasks and application scenarios. Therefore, we also label each collected publication with their motivated tasks and applications. This gives us a full list of nine analytical tasks, which we further organized in to five categories as outlined in Section 4, and seven applications under three major categories as outlined in Section 5.

For each analytical task and application, we went through another round of complementary paper collection for visualization and visual analytics techniques with search queries that combines specific tasks or applications, such as "event sequence summarization" AND "visualization", "medical data" AND "visualization", etc. The entire selection process ended up with 153 most relevant publications of event sequence analysis, and 144 publications of event sequence visualization and visual event sequence analysis. We further refined our selection with 100 most representative and up-to-date event sequence visualization and visual analytics studies to discuss in this paper. Additionally, this survey also includes the review of 8 related surveys, 5 event sequence analysis techniques, 9 visualization techniques in the field of causality analysis yet are not related to event sequence data, and referred to 10 papers regarding the theory of visual analytics, research challenges and opportunities of visual event sequence analysis. This result in a total of 133 papers that are covered in this survey.

The remaining survey is organized as follows. We first introduce the taxonomy of our survey by formalizing our proposed design space and outline visual analytical tasks in Section 2.1. Section 4 elaborates the state-of-the-art solutions developed for each analytical task respectively through an analysis of their corresponding design components of the design space. Then, we provide an overview of applications where event sequence data are commonly observed in Section 5, serving as a more direct guide to practitioners of visual analytics techniques. Finally, we discuss our reflections on research challenges, opportunities in Section 6 and conclude our work in Section 7.

3 TAXONOMY

In this section, we introduce the design space and the collection of visual analysis tasks built from the processes of paper gathering and labeling as mentioned in Section

2.2. The design space and visual analytical tasks form a taxonomy that we further use to structure the survey. Specifically, in Section 4, we partition visual analytics techniques based on their primary analytical task under a consideration that most visual analytics systems are developed around a single analytical task. For each analytical task, we further characterize the relevant papers by exposing the dimensions in the design space that are leveraged to develop each visual analytics method. We also discuss the applications of the visual event sequence analysis techniques in Section 5 to provide domain-specific guidance for practitioners. The applications, however, are not included in our taxonomy, because most of the techniques we collected are developed for event sequence analysis in general cases rather than a specific application.

3.1 Design Space

In the following, we introduce the dimensions of our design space and highlight the key elements (i.e., data scales or techniques) in each dimension that are frequently used for designing and building a visual analysis system for analyzing event sequence data.

3.1.1 Dimension 1: Data Scales

Our proposed design space starts with identifying the granularity of data that the visual analysis is able to cover. For a given event sequence dataset, we summarize the following levels of data granularity.



Event: Individual events represent the finest granularity of event sequences. Each event can be characterized by attributes such as event type, time of occurrence, and duration. Visual analytics techniques often attempt to drill down to individual events to provide users with low-level details of the analysis result. For example, Vistracker [27] identifies anomalous events in trace routes based on event attributes. Carepre [44] predicts upcoming disease based on historical sequence of medical events.



Subsequence: Subsequences are segments of event sequences with the temporal order of events being preserved. Meaningful subsequences can represent the major characteristic of the sequence. EventAction [20] utilizes the number of common subsequences between individuals to measure sequence similarities. MOOCad. [68] leverage anomalous frequent subsequences to facilitate the reasoning of sequence anomalies.



Sequence: An event sequence is the complete record of events that are performed or experienced by a sequence entity (e.g., a patient or a customer). The entire sequence is often analyzed when attempting to get a complete view of the entity's experience. In [35], [72], [129], anomalous entities are detected by analyzing their corresponding progressions of events. Similarly, Guo et al. [33] utilize the embedding of each sequence to estimate the similarity between entities.



Sequence Collection: A collection of sequences are analyzed when summarizing common patterns in the dataset or comparing different groups of sequences. For example, visual summarization techniques [36], [37], [76], [78] aim to provide a summary of patterns and identify entities with common progressions in a collection of sequences. MatrixWave [130] is designed to compare two collections of event sequences and analyze their differences.

3.1.2 Dimension 2: Analysis Techniques

Visual analytics techniques for event sequence data are incorporated with back-end data mining algorithms to support complex analytical tasks. Based on a review of event sequence analysis methods, we identify the following analysis techniques.



Pattern discovery: Pattern discovery aims to find frequently occurring patterns and statistically significant associations of data samples. In the analysis of event sequence data, pattern discovery techniques can be further categorized into frequent pattern mining techniques and similarity analysis techniques based on different analytical goals. Frequent pattern mining techniques are used to uncover common subsequences in the event sequence dataset. For instance, Perer et al. [76] proposed a visual analytics system that employs a SPAM-based algorithm to extract frequent patterns in a collection of event sequences. Similarity analysis techniques utilize event patterns of each sequence to quantify the similarity between sequences. For example, in Eventaction [20] and Similan [114], two different similarity measurements were proposed based on commonness and differences between events across different event sequences.



Sequence inference: Inference is the process of drawing conclusions based on evidence observed in existing data. Conclusions derived from inference techniques are tenable under certain conditions but can be incorrect when applied to unobserved data. Existing inference techniques for event sequence analysis mainly include self-exciting point process and graphical model. Self-exciting point process is a probabilistic model that describes the occurrence probabilities of events over time. The occurrence of upcoming events is influenced by historical events. For example, Hawkes Process is widely employed to model sequential data under the assumption that the impact of the previous event is approximated by a numerical integration over time [64], [119]. Graphical model, on the other hand, presents the conditional dependence between events with a event correlation graph, such as Bayesian Networks [1] and Markov Chain [95].



Sequence modeling: While sequence inference techniques are not capable of making predictions on unobserved data, sequence modeling methods are developed to build a reliable model to characterize observed data while ensuring the model's generalization abilities on unobserved data. Event sequence models are generally specifically designed depending on the analytical tasks, such as classification (e.g., support vector machines, decision trees) and clustering (e.g., k-means). Neural network models, especially recurrent neural networks (RNN) are also commonly applied to model event sequences due to their inherent sequential structure and superior performance comparing to traditional machine learning model. For instance, CarePre [44] employed attention-based RNNs to predict upcoming events based on historical events in sequences, and Guo et al. [35] embedded RNNs into Variational Auto-Encoder to detect anomalous sequences in the dataset.


3.1.3 Dimension 3: Visual Representations


Existing visual analytics techniques leverage a variety of visual representations to display event sequence data and


	Summarization										Prediction & Recommendation					Comparison				Anomaly Detection				Causality	Sum									
	Plaisant et al. 1996	Wongsuphasawat et al. 2001	Cao et al. 2011	Wongsuphasawat et al. 2011	Goiz et al. 2014	Pere et al. 2014	Kwon et al. 2016	Wang et al. 2016	Cappers et al. 2017	Guo et al. 2017	Chen et al. 2018	Law et al. 2018	Kwon et al. 2020	Nguyen et al. 2020	Du et al. 2016	Krause et al. 2016	Du et al. 2017	Jin et al. 2018	Kwon et al. 2018	Wongsuphasawat et al. 2009	Mailik et al. 2015	Zhao et al. 2015	Qi et al. 2019	Guo et al. 2020		Fischer et al. 2012	Zhao et al. 2014	Xu et al. 2016	Nguyen et al. 2018	Guo et al. 2019	Mu et al. 2019			
DS-Single event																																	5	
DS-Subsequence																																		4
DS-Sequence																																		9
DS-Sequence collection																																		16
AT-Pattern Discovery																																		13
AT-Modeling																																		6
AT-Inference																																		5
VR-Chart-based																																		11
VR-Timeline-based																																		18
VR-Sankey-based																																		9
VR-Hierarchy-based																																		5
VR-Matrix-based																																		4
IT-Segment																																		2
IT-Emphasis																																		27
IT-Scaling																																		19
IT-Filter/Query																																		30
IT-Aggregation																																		8
IT-Alignment																																		7
IT-Editing																																		4


Fig. 1. The most cited papers regarding event sequence visualization and visual analytics techniques grouped by different tasks. Each paper is labeled by the relevant design elements in the design space. The rows are grouped and colored by dimensions of our proposed design space: DSS - Data Scales; ATs - Analysis Techniques; VRs - Visualization Representations; ITs - Interaction Techniques.

communicate insightful patterns. The visual representations also determine how events and sequences are organized and aggregated. We identify five categories of visual representation for displaying event sequence data as follows.


 **Chart-based visualizations:** Visualization charts, such as bar charts and scatter plots, are commonly used to display event features and event distributions of event sequences. For instance, Coco [63] uses a table to compare event distributions of two different groups of sequences and a scatter plot to show the number of records containing particular events or subsequences.

 **Timeline-based visualizations:** Timelines are the most intuitive visualizations that organize events of individual sequences successively in a temporal order. Events are generally represented with icons encoded by color, size, or shape to distinguish events with different attributes. For example, VASABI [71] visualizes a sequence as a row of squares colored by event categories.

 **Hierarchy-based visualizations:** In hierarchy-based visualizations, sequences are aggregated into a tree of sequences [113], where each node represents an event placed according to its prefix in the sequence. A variety of visualization designs can be used to display this hierarchical structure of sequences, such as treemaps [100], node-link tree [103], and icicle plots [57], [60].


 **Sankey-based visualizations:** The Sankey-based visualizations organize sequences into the structure of a Sankey diagram [82]. Instead of aggregating sequences into a tree structure as the hierarchy-based visualizations, Sankey-based methods aggregate sequences into a graph, focusing more on providing an overview of transitions between different types of events. Sankey-based

visualizations can be further categorized into two different types of design. The first one is the directed node-link graph in which events are represented by nodes and transitions between events are represented by links [36], [44]. The second one is the traditional Sankey diagram, in which links are further encoded by width, representing the proportions of flow that split and merge among events [37], [111].

 **Matrix-based visualizations:** Matrix-based visualizations are typically used to demonstrate a summary of event frequency or frequent patterns. For example, EventAction [20] incorporates an event matrix to summarize frequencies of events across different time intervals. Mu et al. [68] applied a matrix-based design to present lists of frequent activity patterns in each stage of sequence progression. In addition, a matrix-based design is also utilized to display frequencies of event transitions. For example, Zhao et al. [130] transformed the traditional Sankey diagram into a sequence of matrices to display step-to-step transitions of web clickstream data.

3.1.4 Dimension 4: Interactions

Visual analytics systems usually incorporate rich interactions to empower end users with sufficient flexibility and depth in data analysis. In the following, we summarize seven interaction techniques that are commonly applied in visual analytics systems for event sequence data.

 **Filter/query** allows users to make domain-specific data adjustment or selection based on certain conditions, so as to eliminate noisy and irrelevant data for better analytical performance. The types of filters include *event filters* for filtering specific event types (e.g., [37], [130]), *time filters* (e.g., [27], [57]) for narrowing down to a range of time in the middle of the sequence for exploration, *attribute*

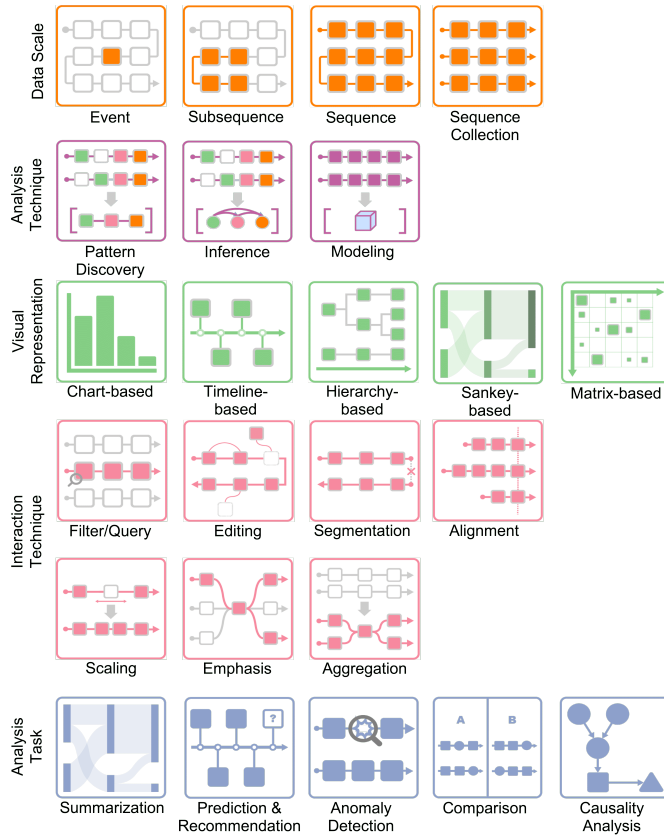


Fig. 2. The taxonomy of this survey is established using a design space for developing visual analysis methods for event sequences with four dimensions: data scale, analysis technique, visual representation, interaction technique, and a organized catalog of visual analytics tasks that includes summarization, prediction & recommendation, anomaly detection, comparison, causal analysis.

filters (e.g., [14]) for retrieving a subset of event sequences based on sequence or event attributes, and *pattern filters* (e.g., [35], [76]) for querying event sequences that contain specific subsequences.

Editing enables users to modify event sequences through *adding new events*, *removing existing events*, *editing event order*, and *editing event duration*, which is commonly employed in what-if simulation of event sequence predictions. The goal is to interactively explore the influence of historical events on the prediction results. For instance, in the CarePre [44] and RetainVis [55], users can edit event sequences to understand how the change of individual events affects the prediction of risks.

Segmentation enables users to split event sequences into sections, which is typically used to narrow the scope of exploration by focusing on sequence segments that are shorter than the entire sequence. Meaningful sequence segments can also indicate event occurring patterns. For example, in MAQUI [57] and DecisionFlow [29], users can segment a set of event sequence by user-specified milestone events to reveal event patterns and correlations.

Alignment refers to arranging multiple sequences to make them aligned on a selected event or time point. This interaction aims to explore and compare patterns before and after the alignment point within a single sequence or across multiple sequences. For instance, Lifelines2 [107] supports interactive alignment of event sequences based on a selected event, so that users can easily

spot precursor, co-occurring, and aftereffect events. Chen et al. [14] allow both sequence alignment and the adjustment of temporal scale to illustrate the temporal distribution of events with respect to a selected event.

Scaling provides analysts an access to zoom in/out the visualizations or inspect data under various granularity. *Zoom in/out* are commonly used in many visualizations, which allows a visualization-level scaling through enlarging or contracting the visual representations to enhance local details or get an overall impression. Additionally, some visual analytics techniques [14], [35], [44], [72] also allows a data-level scaling through *abstract/elaborate* to accommodate the complexity of event sequence. For example, Guo et al. [35], [36] allows a stage-level abstraction and elaboration by aggregating and expanding events within the same progression stage.

Emphasis aims to facilitate the discovery of interesting patterns [92]. This can be achieved through various forms of interactions such as *highlighting*, *sorting*, and *layout adjustment*. *Highlighting* draws users' attention through tweaking basic visual representations (e.g., color, size), which are commonly used in emphasizing sequence groupings, progression pathways, and critical events. *Sorting* emphasizes the ranking of sequences or patterns under specific metrics. For example, Lifeflow [113] allows users to sort progression pathways by the number of records or average time span. *Layout adjustment* enables users to arrange the positions of visual elements in a meaningful way. For example, Guo et al. [37] proposed a layout algorithm that arranges sequence clusters to imply their similarities, which allows users to adjust the similarity threshold to generate different groupings.

Aggregation enables users to interactively merge event sequences, supporting a more scalable exploration of large-scale complex event sequences. For instance, DecisionFlow [29] aggregates sequence with similar occurrence of milestone events so as to enhance visual scalability of large-scale events. CareFlow [74] merges sequences by common event occurrences to reveal frequently observed progression patterns.

3.2 Visual Analysis Tasks

From our review of both data analysis and visual analytics techniques, we summarize the motivating analytical tasks that have gained attention from researchers over the past decade. For simplicity, we further classify these tasks into five high-level categories introduced as follows according to their fundamental objectives.

Summarization: Summarizing event sequences aims to uncover major progression patterns and featured groupings of the sequence entities. The fundamental motivation is to help analysts quickly get an overview of the sequence dataset. A variety of analytical tasks serve the purpose of generating summaries, including sequential pattern mining [60], [76] that discover frequently occurred subsequences from the sequence dataset, progression analysis [29], [36] that reveals time-evolving patterns of latent progression stages, and sequence clustering [31], [71] that segments sequence dataset into groups.



Prediction & recommendation: Prediction & recommendation tasks generally involves analyzing observed event sequences to foresee the upcoming events or sequences, or examining how certain interventions may effect the future trends. The fundamental objective is to make predictive analysis. Typical motivating tasks include making predictions on future events and outcomes [19], [44], and making recommendations on user actions to help achieve certain goals [20]. In addition, due to the importance of interpretability in the applications of sequence predictions, thus, we also include a group of work that visualize the underlying mechanisms of the prediction model [55] to aid result interpretation.



Anomaly detection: Visual anomaly detection for event sequences aims at identifying rare cases that deviates from the majority of the sequence progressions. Anomalies in event sequences can take multiple forms depending on the data scales (i.e., anomalous event, subsequence, sequence). For example, EventThread3 [35] detect anomalous events that derive from normal expected progressions, MOOCad [68] identifies anomalous studying patterns of online students, and FluxFlow [129] captures anomalous spreading process of tweets.



Comparison: Comparison is a common task when investigating similarities and differences between event sequences. Existing visual comparison techniques can be broadly categorized by the scale of comparison targets. For instance, Similan [114] compares individual events of two sequences, while CoCo [62] and MatrixFlow [75] compare two collections of event sequences.



Causality analysis: Causality analysis aims to uncover the causal relationships between event types, promoting a better understanding of which event is very likely to occur after another, or what bring about a certain change to the outcome event. Despite that causality analysis for event sequences have gained much attention in the data mining community, the work in the field of visual analytics under this topic is still very limited, indicating a promising future research direction.

4 VISUAL ANALYSIS TECHNIQUES

In this section, we summarize visual analysis techniques developed for analyzing event sequence data according to the analysis tasks introduced in Section 2.1.

4.1 Visual Summarization

Summarization of event sequences aims to use intuitive representations to reveal major progression patterns and featured groupings of the sequence entities. In many domains such as health informatics [7], [28], [38], [73], [74], [78], [89], social media [57], [76], and career design [36], [37], a variety of analytical tasks serve the purpose of generating summaries, including **explicit summarization**, **implicit summarization**, **progressional analysis**, and **clustering**.

Explicit summarization techniques uncover informative patterns within event sequences using aggregated display overview. All of the sequences are visualized and aggregated into one interface. Existing techniques adopt various visualization approaches to display event sequences, such as timeline-based [7], [49], [58], [78], [107], [126], sankey-based [111], and hierarchy-based [67], [85], [100],

[113] visualizations. Timeline-based visualizations are frequently adopted to reveal temporal information among sequences. For instances, LifeLines [78] and its variant [107] leverage timeline-based visualizations to display the temporal distribution of events in varying time granularities. Sankey-based visualization are adopted to reveal the progression path of event over a period of time. In Outflow [111], alternative clinical pathways within EMRs are visualized using a sankey diagram, the colors of path encode the patient outcomes (Fig. 3(2)). Furthermore, hierarchy-based visualizations, such as tree map and icicle plot, are able to reveal the hierarchical organizations within event sequence data. EventFlow [67] reveals aggregated sequences in a hierarchy-based visualization, and individual sequences are detailed display in a list of timelines (Fig. 3(1)). Similarly, LifeFlow [113] leverage a hierarchy-based visualization to provide an overview of event sequences. Explicit summarization techniques can reveal event sequences with minimum information loss, but interface will become visually messy when the scale of event sequences is large.

Implicit summarization techniques leverage data mining metrics to uncover informative patterns (e.g., frequent patterns) among event sequence data. Existing works that serve the purpose of implicit summarization mainly falls into two categories: query-based techniques and mining-based techniques. *Query-based techniques* [25], [52], [103], [114], [127] enable analysts to create complex queries to extract event sequences of interest. For instances, in COQUITO [52] and CAVA [127], analysts can express complex queries for iterative cohort construction. *Mining-based techniques* leverage advanced sequential pattern mining algorithms to extract insights from complex event sequences [56], [59], [60], [61], [76], [77]. For instances, in Frequence [76] and its variant [77], large scale EMRs data are represented by a set of extracted frequent patterns. The authors used a sankey diagram to reveal the patterns with a color map to encode associated outcomes. Through this view, physicians and clinical researchers can easily understand the important correlations between treatment patterns and associated outcomes. But the frequent patterns do not always correspond to important or meaningful information within the data. Thus, CoreFlow [60] extracts branching patterns in event sequences using the Rank-Divide-Trim three-step procedure, and visualize the patterns as a tree diagram that illustrates an overview of the event flow.

In addition, exploring event sequences by defining queries or using mining algorithms alone may becomes insufficient in some cases. To this end, Law et al. [57] proposed MAQUI, which interweaves *querying* and *mining* to extract informative patterns within a set of event sequences. The authors applied a hierarchy-based visualization and a timeline-based visualization to represent frequent patterns and temporal information, respectively (Fig. 3 (5)). Similarly, Sequence Synopsis combine *querying* and a *mining* method named Minimum Description Length principle to extract informative patterns from event sequences with minimizes information loss. Each extracted patterns are visualized as a series of colored rectangles, where each rectangle represent an type of event.

Progression analysis aims to uncover the evolution of one event during a period of time. Most of the aforemen-

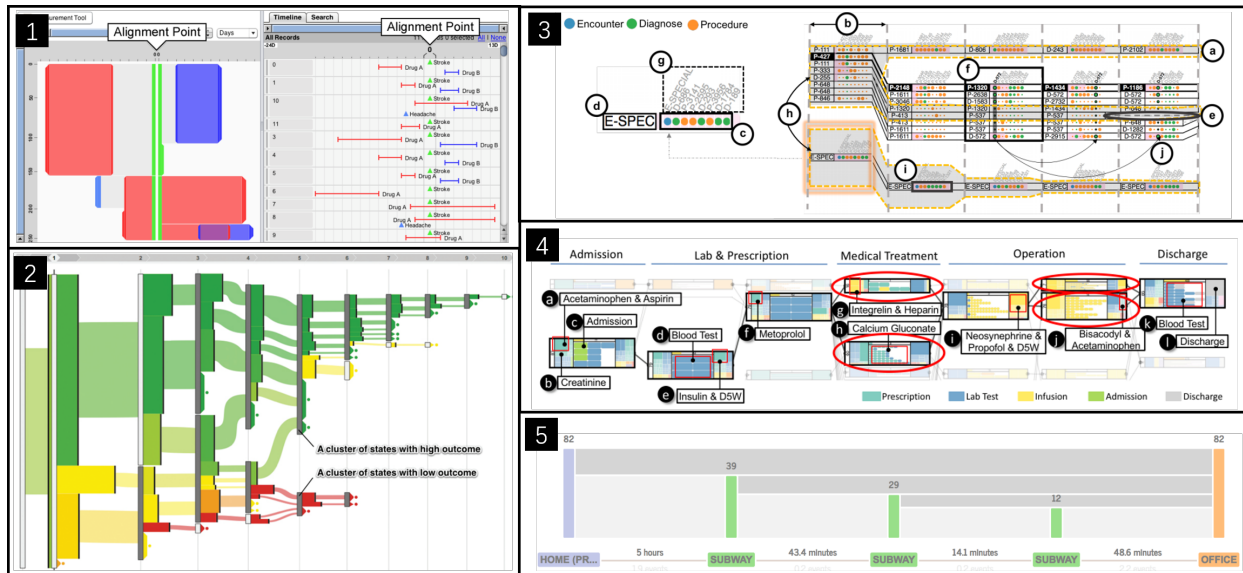


Fig. 3. Selected examples of visual summarization techniques. (1) EventFlow [67] visualizes event sequences in both an aggregated tree-like overview and detailed a timeline display. (2) Outflow [111] visualizes alternative progression paths using color-coded edges that map to patient outcome. (3) EventThread [37] visualizes the threads derived by tensor analysis as segmented linear stripes, following a line map metaphor. (4) EventThread2 [36] uses a node-link visual design to provide a higher-level summary of progression patterns of event sequences. (5) MAQUI [57] applies a hierarchy-based visualization to represent multiple frequent patterns and adopts a timeline to reveal the temporal information.

tioned techniques produce highly summarized results, but fail to show important low-level event details (e.g., single event features) [37]. Visual progression analysis techniques, such as [13], [29], [30], [36], [37], [75], have been introduced to reveal time-evolving patterns of latent progression stages. For instances, in DecisionFlow [29], analysts can use a milestone-based approach to retrieve progression patterns of interest, and visualized them in a hierarchy-based visualization. EventThread [37] has been introduced to summarize latent sequential patterns within a large-scale sequence collection. This technique employs a clustering algorithm to group the summarized patterns into various categories at different stages. In order to clearly reveal the summarized latent patterns, the authors adopted a line map metaphor to display the overall evolution of the latent patterns (Fig. 3(3)). Based on preceding works, Guo et al. further proposed EventThread2 [36], a visual analytics technique that identifies semantically meaningful progressions using a unsupervised algorithm. This technique solves the time scale limitation of EventThread and proposed a new visual design to reveal the progression patterns. It combines a node-link-based cluster view and a timeline-based sequence view to provide a higher-level summary of progression patterns of multiple event sequences by grouping similar segments at each stage (Fig. 3 (4)).

Clustering is the process of finding sequence-wide similarities to achieve sequence groupings. In the clustering analysis of event sequences, a broad range of visual analytics techniques have been developed to empower analysts working with three types of event sequence data, including temporal event sequences, spatiotemporal event sequences and microarray sequences. For temporal event sequences, clustering can be informed by sequence characteristics such as event types and sequence attributes. Cadence system [31] offers a scatter-plus-focus visualization design that supports interactive hierarchical exploration of the space of event type groupings. This system adopts scented navigation cues

to help users navigate complex hierarchies, as well as interactive bar charts and histograms that support additional constraints in categorical and continuous attributes of the target groups. [71], [86], [104], [110] are utilized to cluster individual entities (e.g., works, online users) based on behaviors. VASABI [71] summarise user behaviours by extracting their common tasks, and then identifies the groups of users based on user behaviors. This technique facilitates interactive analysis of user clustering through a hierarchy-based visualization. DICON [3] segments a collection of event sequences into groups based on entity attributes (e.g., age, gender). Each multidimensional cluster is revealed in a hierarchy-based visualization, which allows analysts to understand the event distributions in different groups.

The clustering analysis of spatiotemporal event sequences have been explored in many efforts, such as [42], [53], [83], [102], [122]. Spatiotemporal visual analysis of activity diary data is visualized through VISUAL-TimePacTS [102] on a coordinate plane of time and space. Robinson et al. [83] developed STempo, a geovisualization application to facilitate the exploration of spatiotemporal patterns within event sequence data. Moreover, studies have also introduced visualization tools to cluster microarray sequences [41], [88], [91], [94]. Seo and Shneiderman [91] created the Hierarchical Clustering Explorer that offers a dendrogram and two-dimensional scattergrams, and their dynamic query controls allow users to choose which clusters to display. This model is especially suitable for bioinformatics and microarray data. Moreover, SequenceJuxtaposer [94] facilitates the comparison of biomolecular sequences using a visualization technique called accordion drawing.

In conclusion, visual summarization techniques save user effort by capturing a broad view of event sequence data. To allow for interactive exploration of visual summarization from different perspectives, the aforementioned techniques commonly employ the following interaction techniques within their interfaces: *filter/query* for retrieving



Fig. 4. Selected examples of visual prediction & recommendation techniques. (1) EventAction [20] uses a calendar view to show the temporal information of event sequences. (2) CarePre [44] reveals the medical record of a patient in a timeline-based visualization, and similar patients' medical event sequences are aggregated into a Sankey-based visualization. (3) In RetainVIS [55], predicted risk trajectories are revealed in parallel line charts (middle), and the risk contributors for the patients are displayed in a bar chart (bottom). (4) [34] visualizes the top prediction, alternatives, and their uncertainties in a circular glyph design.

information of user interest, *scaling* for multiple scales visualization, *alignment* for aligning event sequences on selected events or time points, and *sequence editing* for modifying event sequences during analysis.

4.2 Visual Event Prediction & Recommendation

Prediction & recommendation generally involves analyzing observed event sequences to foresee the upcoming events or sequences, or examining how certain interventions may effect the future trends. The fundamental objective is to make predictive analysis. Typical motivating tasks include making predictions on future events and outcomes [19], [33], [34], [44], and making recommendations on user actions to help achieve certain goals [20], [21]. In addition, due to the importance of interpretability in the applications of sequence predictions, a group of work that visualize the underlying mechanisms of the prediction model [51], [55], [96] to aid result interpretation.

Prediction techniques for event sequences have been proposed to predict the next event in a sequence based on historical events. In many domains, event prediction plays an important role in decision-making. For instance, medical researchers and physicians can use this type of techniques to understand potential outcomes of patients under different treatments. The CarePre system [44] leverages deep learning-based RNNs to predict the risk of a patient being diagnosed with certain diseases in the future. In this system, a patient's historical events are displayed in a timeline-based visualization (Fig. 4(2)). Users are allowed to modify these events (e.g., by removing, moving, adding, or adjusting the event's duration) to get different outcomes from the prediction model. Moreover, [34] is a visual analytics system designed for prediction analysis. It employs Recurrent Neural Networks (RNNs) to predict future activities, and review the most probable predictions and possible alternatives in a circular glyph design (Fig. 4(4)). The color

of the first outer ring represents the top prediction for a group of records. Then, depending on the granularity of the analysis, alternative predictions are represented as rings and added to glyph from the inside out.

Recommendation techniques provide reliable suggestions on user actions to help achieve certain goals. Students can adopt this type of techniques to understand their future career development and find an academic plan that suits their desired goals. Du et al. [20], [21] introduced two career path recommendation techniques that provide suggestions and potential outcomes by summarizing the outcome of similar users. In EventAction [20], all of the records that similar to select one are displayed in a list of calendar views (Fig. 4(1)). Recommendation actions are highlighted in the calendar and allow users add into their plans for next round explorations.

In the past few years, deep learning algorithms have demonstrated significant improvements over traditional approaches in the tasks as prediction and classification. For event sequence data, Recurrent Neural Networks (RNNs) are frequently adopted to foresee the upcoming events or sequences, or exam how certain interventions may effect the future trends. However, interpretability is recognized as a primary challenge of deep learning approaches. To address this issue, recent studies have introduced visual prediction techniques to interpret the internal mechanisms of a prediction model [51], [55], [96]. For instances, RetainVIS [55] is a hybrid visual technique for gaining insight into how RNNs model EMR data within the context of diagnosis risk prediction tasks. This technique interprets the relationship between patient records and predicted risk scores. Specifically, patients' medical records and their predicted risk trajectory are visualized in two parallel line charts (Fig. 4(3)), which allow users to understand the progression of predicted diagnosis risks and why such predictions are made. Also, when users hover over the x-axis, they can see

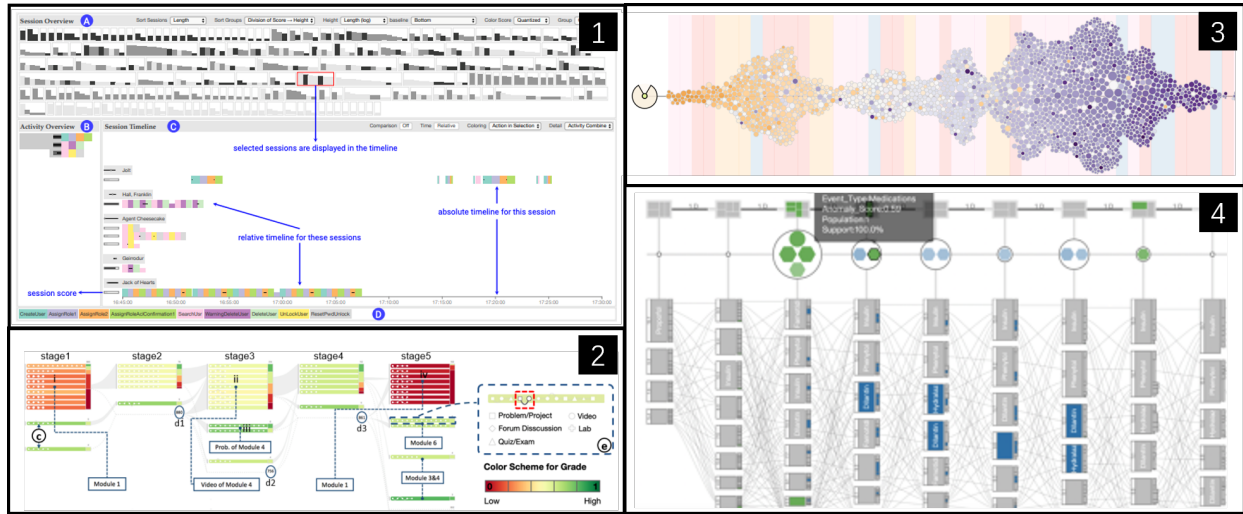


Fig. 5. Selected examples of visual anomaly detection techniques. (1) [72] combines a rectangle glyph design and a timeline-based visualization to reveal anomalies within event sequences. (2) MOOCad [68] employs a Sankey-based visualization to display an overview of the stage segmentation results, and it uses a matrix-based approach to indicate the content patterns of each group within the stage. (3) FluxFlow [129] visualizes anomalous retweeting sequences in a packed circles design. (4) EventThread3 [35] displays anomalous sequences in a line of rectangular nodes ordered by time of occurrence (top). The authors use a circular glyph (middle) to visualize the anomalous events within sequences.

the updated contribution scores of medical events, which represent the importance or contribution of an event to the predicted result. Similarly, LSTMVis [96] focus on the visual analysis of hidden features in RNNs, it allows users to explore hypotheses about RNN hidden state dynamics.

In summary, visual prediction and recommendation techniques contribute to decision-making in many domains. In order to allow users to explore the data from different perspectives, the aforementioned techniques commonly employ the following interaction techniques within their interfaces: *filter/query* for retrieving information of users interest, *emphasis* for adjusting attributes of data to reveal interesting patterns, and *sequence editing* for including a new event or a new feature into the prediction model.

4.3 Visual Anomaly Detection

Visual anomaly detection for event sequences aims at identifying rare cases that deviates from the majority of the sequence progressions. In many application domains, such as social media [4], [6], [129], computer systems [68], [90], [120], clickstream [27], [35], [72], and smart factory [40], [115], [121], various visual techniques have been proposed to serve the task of anomaly detection. As the forms of anomalies vary across different tasks, we broadly divide existing techniques into the following three types: **anomalous events visualization**, **anomalous frequent patterns visualization**, and **anomalous sequences visualization**.

Anomalous events visualization identifies anomalous events within the context of event sequences by uncovering the differences between abnormal and normal events. Many existing techniques incorporate multiple visualization methods in their interfaces to display anomaly events from different perspectives [8], [27], [35], [69], [72], [121]. For instance, EventThread3 [35] detects abnormal events within anomalous sequences based on inferred expected normal progressions (Fig. 5(4)). Anomalous sequences and expected normal progressions are represented in a line of rectangular nodes ordered by time of occurrence. Anomalous events are revealed in circular glyphs to encode critical variables of

anomalous events. In this view, analysts can visually compare the abnormal sequence with normal sequences, and thus potentially understand why anomalies exist. Moreover, in [72], anomalous log sequences are detected by a black-box model, and displayed in a timeline-based visualization. Each event is represented as a colored rectangle, users can verify the anomalous logs and explore the events that contribute to sequence anomaly. Xu et al. [121] extended the Mareys graph to visualize product moving traces in a production line. The visualization of individual products and their processing times improves user understanding of a lines performance, and also helps in better understanding anomalous events, the causes and effects in a production line.

Anomalous frequent patterns visualization is utilized to help users perceive the anomalous frequent patterns that contribute to sequence abnormality. MOOCad [68] is designed to detect anomalous learning patterns within MOOC data (a set of online learning activities sequences) (Fig. 5(2)). To facilitate anomaly detection and reasoning, the large-scale learning sequences are clustered into various groups at different stages. The authors employed a sankey-based visualization to display the overview of the stage segmentation results, and a matrix-based approach to indicate the content patterns of each group within the stage. In this view, users can flexibly explore the anomalous learning patterns via stage comparison, group comparison within stages, and individual path inspection.

Anomalous sequences visualization helps users detect anomalous sequences within sequence collections, uncover the temporal structure of anomalous event sequences, and reveal the deviation of anomalous sequences from normal sequences. For instance, Zhao et al. [129] proposed a flexible timeline visualization technique to discover rumor-spreading processes between Twitter users (Fig. 5(3)). The retweeting sequences are visualized by a packed circles design, where each participating user is represented as a circle. In order to intuitively display the abnormality of sequences, the authors designed a circular glyph for each

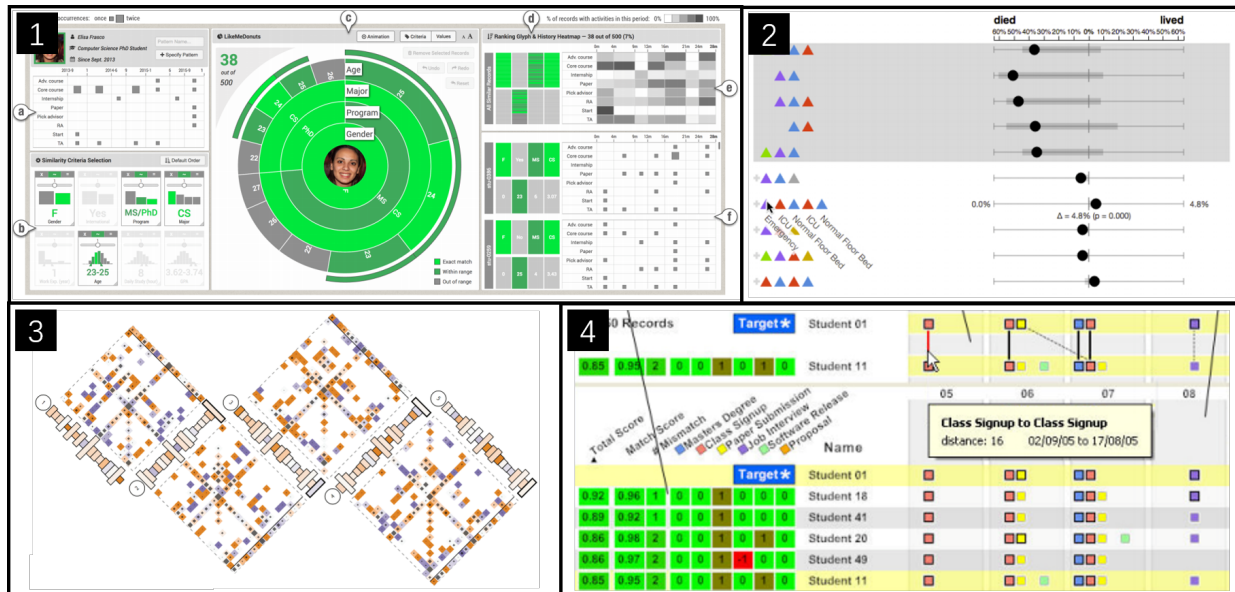


Fig. 6. Selected examples of visual comparison techniques. (1) [22] summarizes the criteria values of similar records as a hierarchical tree. Each of the records and common temporal patterns among the similar records are visualized in the calendar view. (2) In CoCo [63], a combination of medical events is visualized in a timeline, where colored triangles represent medical events. (3) MatrixWave [130] visually compares two web clickstreams in a matrix-based visualization. (4) Similan [114] visualizes each event sequence in a binned timeline. The paired events are connected by lines.

retweeting sequence that summarizes its important aspects such as overall abnormality, contextual polarity, scale, and temporal information. Cao et al. developed TargetVue [6] to detect Twitter users with anomalous behaviors. This technique explores anomalous users via an unsupervised learning model and visualizes the behaviors of suspicious users in three glyphs (Fig. 9(4)). These glyphs are designed to present the users' communication activities, features, and social interactions, respectively. Nguyen et al. [72] proposed a visual analytics approach that aims to detect unusual action sequences of users (Fig. 5(1)). Every sequence is visually summarized in a compact glyph to help analysts spot anomalous sequences, the length and color saturation of glyph represent sequence length and anomaly scores respectively. Also, anomalous sequences are visualized in a timeline visualization, where each event is represented by a colored rectangle whose color maps to event type. Similar designs are proposed by [27] and [35]. In [27], anomalous sequences are displayed in a timeline visualization with color encoding by event type. In [35], Guo et al. employed an MDS projection to visually summarize the abnormality of a dataset and subsequently reveal sequences of interest in a timeline-based visualization.

In summary, visual anomaly detection has been introduced to solve various real-world problems across different application domains. To allow users to interactively explore data from different perspectives, the aforementioned techniques commonly employ the following interaction techniques within their interfaces: *filter/query* for retrieving information of users interest, *emphasis* for adjusting attributes of data to reveal interesting patterns, *scaling* for multiple scales visualization, and *alignment* for align sequences on selected events or time points.

4.4 Visual Comparison

Visual comparison is a common task when investigating the similarities and differences between event sequence data. A variety of visual comparison techniques have been proposed

to solve real-world problems in many domains such as for career path analysis [20], [22], clickstream analysis [72], [130], health information analysis [33], [44], [63], and generic purposes [35], [114]. In our work, we classify the visual comparison techniques for event sequences based on compared targets, including **comparison techniques for event sequences**, **comparison techniques for sequential patterns**, and **comparison techniques for sequence collection**.

Comparison techniques for event sequences are utilized to compare individual events in terms of disorder, missingness or redundancy, and difference in the occurrence of timing and attributes. To facilitate the interpretation of compared results, researchers adopted *juxtaposition design* [114], *superposition design* [44], and *hybrid design* [33], [35] to clearly visualize the similarities and differences between sequences. For instance, Similan [114] shows the similarity of events within two similar sequences via *juxtaposition*. In this technique, each event sequence is visualized in a binned timeline (Fig. 6(4)), similar sequences are placed beneath the target sequence for explicit comparison. In order to reveal the similarity information between two sequences, the pairs of events matched by the Match & Mismatch measure are connected by lines, and events without any links connected to them are missing or they are extra events. Moreover, in CarePre [44], the visual comparison techniques with *superposition design* are utilized to verify the predicted risk of potential diseases. Last but not least, in the most recently published visual comparison technique, Guo et al. [33] searched similar medical records and applied *hybrid design* to convey the differences between target record and its similar records. This technique uses *explicit encoding* to reveal the overall dissimilarity of similar records over time, and it uses *superposition* to represent differences between target record and its top three similar records in detail.

Comparison techniques for sequential patterns are used to investigate the similarity of sequential patterns within two event sequences for diverse applications, such

as for log files [72], [81] and career paths [20], [22]. The career recommendation technique, EventAction [20], uses a calendar view to show event sequences and *juxtaposed* them in a ranked list for visual comparison. In [72], Nguyen et al. adopted visual comparison with *superposition* to indicate the anomaly path within abnormal sequences. Du et al. [22] supports *explicit encoding* and *juxtaposition* of differences for semantically meaningful comparisons (Fig. 6(1)). Specifically, while comparing the target record with the entire dataset, the authors summarized the criteria values of the similar records in a hierarchical tree, where the similarities and differences are explicitly encoded. For a detailed inspection, all records and common temporal patterns are visualized in the calendar views, so that users can juxtapose any two sequences of interest to explore the differences between them.

Comparison techniques for sequence collection aim to find differences between two sets of event sequences in terms of structure, attribute, temporal information [62]. For instance, CoCo [63] leveraged automated statistical analysis to compare the attributes of two distinctly defined cohorts, adopting *explicit encoding* to convey an overview of the differences between the two cohorts (Fig. 6(2)). Moreover, MatrixWave [130] is a matrix-based visualization for comparison analysis of two web clickstream datasets in terms of traffic patterns (Fig. 6(3)). The authors applied *superposition* to represent two related event sequence datasets within one visualization and used *explicit encoding* to reveal the differences between traffic paths at each node. This technique focuses on differences in the occurrence of immediate and pairwise steps among two clickstream datasets.

In conclusion, visual comparison techniques can save analysts' efforts to explore the differences between two event sequences or two groups of event sequences. To facilitate interactive analysis, the aforementioned techniques adopt *filter/query* for retrieving information of users interest, *scaling* for multiple scales visualization, *alignment* for align sequences on selected events or time points, and *emphasis* for adjusting attributes of data to reveal interesting patterns.

4.5 Visual Causality Analysis

Visual causality techniques have been proposed to help users uncover causal relationships among data. Traditional visualizations, such as a directed acyclic graph (DAG) or the Hasse diagram [50], can be employed to illustrate causality to a certain extent. However, they become inefficient as an increased number of variables introduces more edge crossing. Elmqvist et al. successively proposed two visual methods, Growing Squares [23] and Growing Polygons [24], which enhance node representations within a DAG with color-coded squares and polygons that help provide an overview of influences on each event in large systems. They also leveraged animations to dynamically present the temporal ordering of causality. Despite that both methods are effective in uncovering the causal relationships of events, they fail to integrate causal semantics into the graph, which is important for a deeper understanding of causal structures.

To incorporate additional causal semantics, Kabada et al. [46] introduced a set of animations following Michottes rules of causal perception [65] to intuitively illustrate causal strength, amplification, dampening, and multiplicity. Recent

studies have invested more effort in integrating automatic causal analysis algorithms and causality visualizations into a visual analytics system to facilitate interactive causal analysis and reasoning. In [9], Chen et al. proposed a workflow for a visual causal analysis system that aims to support decision-making by providing hypothesis generation and evaluation. This leads to a number of visual analytics systems that are designed to support interactive analysis of data correlation and causation. For example, Zhang et al. [128] introduced a visualization tool that utilized force-directed graphs to display the correlations between numerical and categorical variables in multivariate data. Within their interface, the authors designed a slider bar that allows users to filter the edges corresponding to weak relations. ReactionFlow [18] was developed to facilitate a better understanding of causal relationships between proteins and biochemical reactions in biological pathways. It organizes the causal pathways into a Sankey-based structure to emphasize the downstream and upstream nature of the causal relationships. It uses animation to highlight the flow of activity through a pathway. Wang et al. [105] presented a visual interface to reveal causal relationships in a force-directed graph with a color scheme design that allows analysts to edit and verify causal links according to their domain expertise. They extended this work in [106] with a path diagram visualization to better expose causal relationships between variables.

As prior efforts mainly focus on the causal analysis of multivariate data, few techniques exist to analyze causal relationships among events in event sequence datasets. When dealing with event sequences, three major challenges need to be specifically tackled. First, the temporal nature of event progressions adds additional causal semantics, such as causal delays and causal durations, into the causation of events. Thus, this aspect raises the bar for extracting causal relationships within event sequences. Second, the high dimensionality of events and the latent structure of hierarchies in event types add complexity to the causal graph. This requires a dedicated graph layout mechanism to handle the causal complexity. Lastly, the complexity of temporal event sequences leads to difficulty in investigating event sequence collections.

5 APPLICATIONS

Given the broad applications of event sequence data, in this section, we review visual analysis techniques for event sequences applied in the fields of **Health informatics**, **Internet applications**, and **Industry 4.0**.

5.1 Health Informatics

In health informatics, electronic health records (EHRs) and electronic medical records (EMRs) can be represented as individual event sequences. Each sequence records the medical events of a patient over the course of a clinical process, and each event represents a medical event such as a diagnosis, lab test, medication, or treatment. With ample medical event sequence data and domain knowledge, physicians and medical researchers can extract new knowledge, quantify the effects of changes in care delivery, and potentially guide the formation of best practice guidelines. To extract meaningful information from medical event sequences, a

	Health Informatics						Internet Application								Industry 4.0					Sum							
	Klemm et al. 2014	Perer et al. 2014	Krause et al. 2015	Zhang et al. 2015	Rogers et al. 2019	Guo et al. 2020	Kwon et al. 2020	Viegas et al. 2013	Yuan et al. 2014	Wu et al. 2014	Cao et al. 2015	Cao et al. 2015	Chen et al. 2015	Mueller et al. 2016	Xie et al. 2018	Chen et al. 2019	Goulden et al. 2019	He et al. 2019	Xu et al. 2019		Jo et al. 2014	Xu et al. 2016	Herr et al. 2018	Wu et al. 2018	Sun et al. 2019		
DS-Single Event																										1	
DS-Subsequence																											3
DS-Sequence																											5
DS-Sequence Collection																											14
T-Summarization																											22
T-Progression analysis																											11
T-Prediction&Recommendation																											1
T-Anomaly Detection																											7
T-Comparison																											11
T-Causality Analysis																											0
VR- Timeline-based																											12
VR- Sankey-based																											8
VR- Hierarchy-based																											8
VR- Matrix-based																											2
VR- Chart-based																											13
IT-Segment																											2
IT-Emphasis																											18
IT-Scaling																											15
IT-Filter/Query																											24
IT-Aggregation																											2
IT-Alignment																											1
IT-Sequence editing																											2

Fig. 7. The selected papers regarding visualization and visual analytics of event sequences in different application domains. Each paper is labeled by relevant analysis tasks and design components in the design space. The rows are grouped and colored by dimensions of our proposed design space: DSs - Data Scales; Ts - Analysis Tasks; VRs - Visualization Representations; ITs - Interaction Techniques.

variety of visual techniques have been proposed to serve analysis tasks, including **cohort analysis** [3], [36], [37], [52], [54], [62], [127], **outcome analysis** [29], [31], [74], [84], [108], [111], [113], and **prognosis analysis** [44], [55].

Cohort analysis is a common approach used to uncover correlations between a specific disease risk and the underlying attributes of patients within the cohorts [127]. Medical researches can construct a cohort of patients based on a medical event (e.g., diagnosis, treatment), the attributes of patients (e.g., gender, age), and the patterns of individual sequences (e.g., symptoms progression, treatment progression). Suppose a medical researcher wishes to understand the exposure factors for lung cancer. He can gather the answer by analyzing common attributes within a cohort of lung cancer patients or by measuring the differences between cohorts with or without lung cancer. Following this idea, existing visual techniques for cohort analysis emphasize one of two strategies: cohort summarization [36], [37], [80], [127], or cohort comparison [3], [52], [54], [62], [79].

Cohort summarization techniques, such as CAVA [127] and Chronodes [80], visually summarizing informative patterns within a cohort and uncover the common exposure factors for a disease. CAVA [127] combines chart-based and hierarchy-based visualization to represent the attribute distributions of a cohort (Fig. 8(3)). Then, to further investigate exposure events in the cohort, each patient was assigned a hospitalization risk score based on their medical history. Both the calculated risk scores and event progressions are

visualized by color-coded edges, analysts can intuitively understand how different event progression pathways lead to different hospitalization risk scores and which medical events have higher risks. Moreover, in EventThread2 [36], the clustered medical event sequences and common sequential patterns (e.g., typical care plans) of a cohort are visualized in a sankey-based visualization and a node-link visual design respectively (Fig. 3(4)). User can inspect common sequential patterns of a cohort with the goal to explore those medical events that affect further progression.

Cohort comparison measures differences between two cohorts of patients to determine exposure factors of a condition such as disease or death. COQUITO [52] helps users interactively construct two cohorts and explore exposure events for a disease. It uses a hierarchical tree map and multiple bar charts to provide an overview of statistical information about the cohorts (Fig. 8(1)). Then it leverages PARAMO [70] to compare two cohorts and determine if the constructed cohorts carry exposure events for a disease. CoCo [62] is a visual comparison technique (Fig. 6(2)) that measures the differences between two cohorts under various differentiating metrics. Users can select metrics of interest, such as the most differentiating event subsequences between two cohorts, to explore the medical events or patterns that may influence the incidence of a condition. In CoCo, each row displays the difference between two cohorts, where medical patterns of cohorts are visualized by a timeline-based design. A circle marker is placed horizon-



Fig. 8. Selected examples of visual techniques for health informatics. (1) COQUITO [52] uses a hierarchical tree map and bar charts to provide an overview of statistical information of cohorts defined by users. (2) Composer [84] plots the outcome score trajectories of different procedures in a line chart. (3) CAVA [127] uses a stacked bar chart, a pie chart, and a hierarchical tree map to represent the age, gender, and diagnosis distributions of a cohort, respectively. Both the calculated risk scores and event progressions within the cohort are visualized by color-coded edges.

tally between two cohorts to display the difference between the values and in the direction of whichever cohorts value (e.g., death rate, survive rate) is higher.

Outcome analysis studies the end results of different medical progressions (e.g., symptom progressions, treatment progressions) with the goal of facilitating informed decision-making about diagnosis and treatment options. Existing works, such as Outflow [111] and Frequence [76], reveal the medical progression paths in a sankey-based visualization to uncover the outcomes of different procedures. More specifically, Outflow [111] aggregates medical event sequences from a cohort of patients and visualizes alternative progression paths using color-coded edges that map to patient outcomes (Fig. 3(2)). Similarly, in a series of efforts proposed by Perer et al. [74], [76], [77], the authors extracted frequent progression pathways of a cohort and used a sankey-based visualization to display them, while providing context on which care plans were successful and which were not. These techniques provide an overview of the progression pathways within a cohort, and thus help users understand which factors, medical pathways, or other structures are most associated to the outcome of interest. Nevertheless, as users are not allowed to interactively build the cohorts in some outcome analysis techniques, the analytic capability of these techniques could be hugely impacted when analyzing a sequence collection of different patients. To overcome this issue, DecisionFlow [29] leverages a milestone approach to support users in defining a cohort by highlighting patients with a specific outcome (e.g., a disease). In this technique, the author used a hierarchy-based visualization to interpret how many patients within the cohort have the specific outcome. Users can interactively compare the proportion of patients across different

medical procedures and explore the association between medical events and outcomes. Moreover, Composer [84] enables users to interactively explore the outcomes under different cohorts and treatment plans. This technique employs PROMIS (Patient-Reported Outcomes Measurement Information System) to automatically evaluate the outcome scores of a patient under user-defined treatments, and plots the outcome score trajectories in a line chart (Fig. 8(2)). Medical researches can plot outcome trajectories of different treatments in one chart to determine the optimal treatment for a cohort of patients.

Prognosis analysis predicts the risks of a patient being diagnosed with certain diseases in the future based on the patient's medical history. A series of deep learning-based visual prognosis techniques, such as [15], [44], [51], [55], have been introduced to serve prognosis analysis and interpret the results. For instance, [51], [55] implement RNNs to predict the current and future states of a patient. Retain-Vis [55] enables users to modify individual sequences of medical events (e.g., add or remove medical events, modify visit period) to experiment with how predicted risk changes according to event sequences changes. The authors visualized a patient's predicted risk trajectory and their medical event sequences in two parallel line charts (Fig. 4(3)). In this view, users are able to observe correlations between medical event sequences and prediction risks, and thus understand why such risk predictions are made. CarePre system [44] can predict the risk of a patient being diagnosed with a certain disease and estimates the most influential treatments for a patient based on historical medical records. The patient's historical events are visualized in a timeline-based visualization (Fig. 4(2)), and users are allowed to modify these events (e.g., removing, moving, duration adjustment, adding) to

get different predicted risks. Clinicians can create multiple edited sequences to analyze the predicted results under alternative treatments, and this system thus helps clinicians understand the impact of different treatment options.

5.2 Internet applications

In various internet applications, the activities of users and devices can be recorded as individual event sequences. For instance, social media data contain sequences of timestamped activities (e.g., posting or commenting) for specific users that are recorded over time. Similarly, clickstream data collected from e-commerce websites record how visitors operate and navigate through a web site, and this data can be represented as sequences of timestamped events (e.g., visiting a product page, purchasing a product) generated by visitor actions. Additionally, the system logs collected from a computer system can also be represented as temporal event sequences of device conditions (e.g., usage, temperature, workload). In this section, we provide a review of the visual techniques that have been developed for event sequence data retrieved from **social media platforms**, **e-commerce websites**, and **computer systems**.

5.2.1 Social media

On social media platforms such as Twitter and Facebook, user activities can be recorded as event sequences. Each sequence records the temporal activities of a user over time, where each event represents an online activity such as posting or commenting. Analysis of such event sequence data has exhibited potential for understanding various types of user behavior on social media. Existing efforts have proposed a range of visual analytics techniques to help yield insights about **collective behaviors** [12], [57], [76], [101], [117], [123], [129] and **ego-centric behaviors** [4], [6], [11].

Collective behaviors refer to activities conducted by a temporary and unstructured group of people. On social media, collective behaviors are formulated by groups of social media users through the processes of spreading information and human mobility. To study these collective behaviors and identify behavioral patterns, various visual analytics techniques have been proposed: [5], [12], [98], [101], [129] are designed to study the behavior of *spreading information*, and [57], [76] can be utilized to analyze *human mobility*.

Reposting process refers to how information spreads across space and time on social media platforms. Google + [101] interweaves node-link diagrams and circular map metaphors to visualize message spreading paths. Analysts can easily capture the traces of diffusion between users and identify the importance of a message by its size and diffusion path. Chen et al. [12] used a map metaphor to symbolize the reposting process in a spatial context (Fig. 9(1)). The diffusion structure is visualized using various link metaphors such as rivers, routes, and bridges. This technique highlights the influence of key players, and it enables analysts to explore how these key players promote the evolution of topics and enlarge the influence of the source message. Zhao et al. [129] proposed a flexible timeline visualization technique to reveal the rumor spreading process among Twitter users. Moreover, tracing the spatiotemporal information of diffusion pathways can uncover how information is spread on a global scale, such as [5], [98]. Cao et

al. [5] visually summarized the temporal trends, the social-spatial extent, and community response to a topic using a sunflower metaphor. The original tweets are placed at the center of the circle and linked with geo-groups (users from a same country) once the original tweets are reposted by users in these groups. The retweeting activities are displayed as a sequence of color-coded retweet glyphs moving along pathways that indicate the timing and sentiments of the retweets.

Beside the reposting process, another collective behavior of importance is *human mobility*. The spatiotemporal event sequences retrieved from social media platform, like Foursquare, have recently been used to uncover user mobility patterns and predict mobility decisions. For example, by studying human mobility, advertisement companies can explore the mobility patterns of people such as when and where they go to work and, thus, optimize their advertising strategies. Some visual analytics techniques that leverage pattern mining algorithms have been used to explore common mobility patterns of users such as [57], [76]. MAQUI [57] support interactive exploration of the data collected from Foursquare to uncover the frequent mobility patterns of users.

Egocentric behaviors refer to activities conducted or influenced by a user. An egocentric perspective enables a closer analysis of individual behaviors and thus provides more detailed behavioral patterns [11]. For instance, Cao et al. [4] proposed Episogram, an egocentric representation for visualizing individuals' interaction histories (e.g., posting or reposting content). Episogram visualizes each interaction thread using a vertical line on a timeline and uses a glyph design to represent interaction events among users. Building upon preceding works, Cao et al. developed TargetVue [6] to detect and visualize users with anomalous behaviors on Twitter. TargetVue detects anomalous users via an unsupervised learning model and visualizes the behaviors of suspicious users in three glyphs that represent the users communication activities, features, and social interactions, respectively (Fig. 9(4)). Moreover, [11] proposed a map-based visual technique to summarize the historical diffusion traces initiated by a central user. Users who participated in reposting one central users post are visualized as hex nodes whose color and size encode the user's behaviors and roles. These users are grouped into different regions on the map and linked with the central user, forming the social network of the central user. In this view, if one user leads to a great amount of reposting, analysts can understand how information reaches him and diffuses from him.

5.2.2 E-commerce

Clickstream data collected from e-commerce websites record how visitors operate and navigate through web sites. A visitor online activity can be recorded as an event sequences, in which each event represent a single online activity (e.g., visiting a product page). The increasing availability of such event sequence data permits analysts to extract valuable insights in website design and commercial activities such as advertising. Existing visual techniques have been introduced to explore **frequent visiting traces** [60], [61], [124] and **user behavior patterns** [10], [26], [32], [39], [68].

To facilitate the understanding of frequent visiting traces,

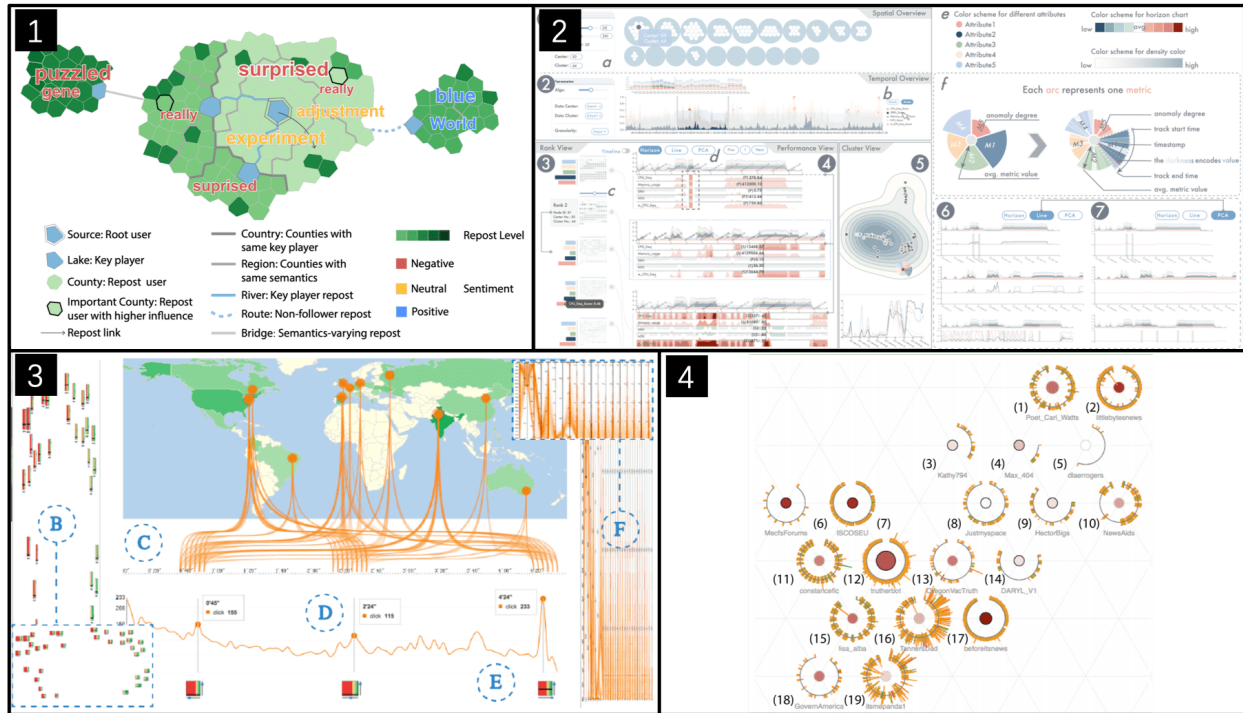


Fig. 9. Selected examples of visual techniques for Internet applications. (1) R-Map [12] uses a map metaphor to symbolize the reposting process in a spatial context. (2) CloudDet [120] combines a glyph design and a stacked line chart to monitor the performance of a computer system. (3) PeakVizor [10] encodes each interaction peak by a glyph in an overview, and the spatial-temporal information of the peaks and the correlation between the peaks are visualized in two additional views. (4) TargetVue [6] visualizes the behaviors of suspicious users in three glyphs that present the users communication activities, features, and social interactions respectively.

Zraggen et al. [124] proposed (s)queries to visualize regular patterns of clickstream data. Moreover, Liu et al. [61] extracted frequent browsing paths from clickstream data and visualized them in a funnel-based visualization. As frequent patterns do not always correspond to important or meaningful information within data, CoreFlow [60] leveraged a tree-based visualization to facilitate branching pattern exploration for browsing paths.

Analyzing clickstream data can help e-commerce companies explore users behavior and optimize their business plans. This idea has been extended to online education platforms with the goal to explore student learning behaviors [10], [32], [39], [68]. For instance, PeakVizor [10] analyzes students' interaction activities to understand how students respond to video material. For example, if an unexpectedly high occurrence of pausing or rewinding is observed at a certain segment, then this segment is probably difficult or confusing and thus requires additional time watching and studying. The authors encoded each peak, representing high pausing or rewinding activity, using a glyph in an event sequence overview. Moreover, spatiotemporal information about the peaks and the correlations between peaks are also visualized in two additional views (Fig. 9(3)). CCVis [32] explores the patterns in online students' clicking behavior, and thus, identifying the course resources that were clicked most and least. It visualizes the critical sequences that lead to different transition probabilities in a node-link diagram, and it use a sankey diagram to display the click behavior patterns.

5.2.3 Computer systems

Computer systems are monitored by regularly sampling profile data that record the timestamped conditions (e.g.,

CPU load, memory usage) of specific devices over time as event sequences. Monitoring and analyzing the profile data are important for identifying devices that are over- or under-allocated, inefficient operations, and nodes that are misbehaving or failing. Muelder et al. [69] proposed a visual technique to portray the behavior of cloud computer systems over time. The authors adopted a stacked graph timeline to summarize the aggregate behavior of cloud computer systems. For detailed inspection, the behavioral lines of each compute node are plotted in a table of line charts. In this view, analysts can efficiently explore the trends and anomalies within a system. Xie et al. [118] leveraged one-class support vector machines to detect anomalous executions in high performance computing clusters. Detected anomalies are visualized in a multi-level visualization system for deeper analysis. Specifically, all of the anomalous compute trees are identified in a scatter plot. Analysts can select the anomalies of interest to inspect their structural patterns in a node-link diagram and their invoked functions in a stacked timeline. [120] provides interactive visualization capabilities that enables analysts to inspect profile data and identify anomalous performances in cloud computing systems. This system combines multiple visualization modes such as glyph design and stacked line charts, to comprehensively monitor the performance of cloud computing systems from different aspects (Fig. 9(2)).

5.3 Industry 4.0

In smart factories, the temporal status of equipment over time can be recorded as an individual event sequences, where each event represent a status (e.g., a equipment condition or a processing event). Monitoring and analyzing these event sequence data can help managers understand

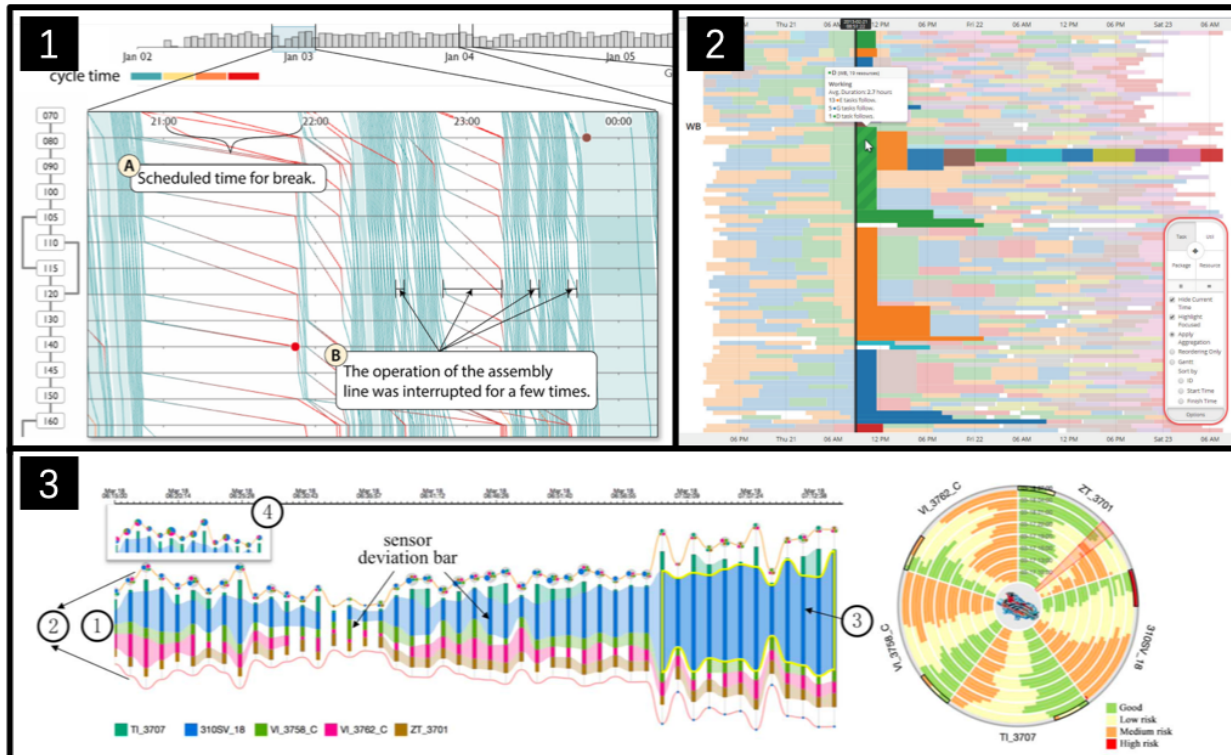


Fig. 10. Selected examples of visual techniques for Industry 4.0. (1) Xu et al. [121] extend a Mareys graph to visualize product moving traces in a production line. (2) In LiveGantt [45], the big picture of the current schedule is visualized in a Gantt chart. (3) Wu et al. [115] employ a stacked timeline to reveal how the real equipment conditions deviate from "normal" conditions in a period of time. The long-term trends of equipment conditions are visualized as a radial visualization to provide users with an overview of equipment conditions during a certain past time period.

factory conditions, quickly respond to various sorts of events, and optimize the productivity of factories. A variety of visual techniques have been introduced to help users exploring anomalous events [8], [40], [115], [121], [132] and optimizing manufacturing plans [45], [97].

In smart factories, an anomalous event (e.g., equipment failure, outlier process) could result in a serious incident or great financial loss. Traditional anomaly detection depends on manually checking every equipment, which is too expensive and inefficient. In contrast, the collected manufacturing data provides more reliable resource for factory managers to analyze anomalies. For instance, Herr et al. [40] analyze event reports of a production line and detected systematic issues in manufacturing processes. Reported events are shown as a time series plot that can help understand the error distribution and recurring error patterns. Xu et al. [121] extended the Mareys graph to visualize product moving traces in a production line (Fig. 10(1)). The visualization of individual products and their processing times improves user understanding of a lines performance, and also helps in better understanding anomalies, the causes and the effects in a production line. The visual technique proposed by Wu et al. [115] provides an interactive interface to monitor the status of equipment in smart factories. The authors estimated normal conditions of equipment based on a training set, and then employed a stacked timeline to reveal how the real equipment data deviate from estimated normal conditions over a shot period of time (Fig. 10(3)). Moreover, in order to visually summarize the long-term trends of equipment conditions, the authors adopted a radial visualization to provide an overview of equipment conditions during a certain past time period.

Analyzing manufacturing data can help managers and factory planners optimize manufacturing schedules. More specifically, in a production line, each machine is responsible only for a specific part of the production process. When the cooperation of machines is not well designed, the production line's overall efficiency will be negatively affected. The event sequence data of production lines record the past and current tasks of machines. By analyzing these data, factory planners can explore and reschedule inefficient plans, such as a manufacturing plan with significant equipment conflict. LiveGantt [45] is an interactive schedule visualization tool that helps managers explore highly concurrent manufacturing schedules from various perspectives. In this technique, the big picture of the current schedule is visualized in a Gantt chart (Fig. 10(2)). Users are allowed to interactively explore the inefficiencies and reschedule manufacturing plans accordingly. PlanningVis [97] is a multi-level visualization system to support interactive exploration and comparison of production plans. This technique juxtaposes heat maps, line charts, and bar charts to visualize the differences between two plans, and thus, optimizing production plans.

6 CHALLENGES AND OPPORTUNITIES

In previous sections, we summarized event sequence visualizations according to our proposed design space, extracted five analytical tasks common in visual analysis techniques for event sequences, and categorized the visual analysis techniques into three typical applications. Through this process, we found several remaining challenges in existing research and promising future research directions that are discussed in this section.

Data quality: The performance of data analysis techniques largely depends on the quality of data [47]. On top of this, the complexity of event sequence data adds difficulty in data recording and leads to more problems for data quality. Typical data quality issues include data incompleteness (e.g., missing events or timestamps), data errors (e.g., errors or inconsistency in event naming), and duplication of data records, each which can mislead statistical analysis results. The issue of data quality implies a need for additional effort to improve data processing to prevent misleading results and inferences gathered from the source data.

Uncertainty: Uncertainty in information is introduced when analyzing event sequence data with quality issues or during user-specified data adjustments such as data transformation and wrangling. This uncertainty can inhibit analysts from making optimal decisions if information about uncertainty is not properly communicated in the visual analytics process [87]. Although some previous studies [19], [34] have incorporated uncertainty information in visual analytics of event sequence data, they focused on only one type of uncertainty information – the probabilistic uncertainty under an event prediction scenario. Therefore, more research is required to study the best ways of incorporating and visualizing other types of uncertainty information, such as bounded uncertainty, during the process of event sequence data analysis.

Scalability: Scalability is a well-recognized challenge in visual analytics [17], [48]. This problem becomes more significant in visual analytics of event sequence data due to the large scale (i.e., large number of sequences) and high dimensionality (i.e., vast number of event types) of most real-world event sequence datasets [29]. Some previous research touches upon this problem mainly through sequence aggregation [112] and event filtering [29], [31] to enhance the visual scalability on the sequence level and event level respectively. However, these summarization techniques hinder the inspection of detailed individual sequences and events, and the problem of how to scale across both sequence summarizations and low-level details still remains. Therefore, there is a demand for a scalable visual analytics pipeline that follows the Visual Information-Seeking Mantra by Shneiderman [93]: “overview first, zoom and filter, then details-on-demand” to allow users to flexibly switch between visual summaries and sequence details.

Heterogeneity: Event sequence data can contain a variety of heterogeneous temporal events. For example, medical health records usually include multiple event types such as diagnostic events, lab test results, vital signs, drug administrations, etc. Events of each event type are observed or recorded with different sampling rates and show different event patterns, which leads to great difficulty for aggregating and organizing data from multiple sources. Most existing techniques choose to assemble all types of events to form a unified process for modeling and display. However, this may hinder the discovery of relationships between event types and distinctive patterns from disparate event types, which is crucial for investigative tasks and sense-making processes [109]. To solve this issue, a visual analytics framework need to be developed, enabling both the integrated analysis of multiple event processes and the investigation of patterns and trends for individual processes.

Multivariate event sequence visualization: Existing visual analytics techniques for event sequences generally characterize events based on their types and timestamps only. Besides these two common event attributes, however, events in a sequence can also be associated with multivariate data. For instance, lab test events in medical data are associated with specific test values, and financial transaction records also contain information about bank accounts and the monetary amount of a transaction. It still remains challenging to visualize multivariate event sequences due to the large number of event attributes a single event may include, coupled with the additional heterogeneity introduced by different data formats of the variables linked to events. Cappers and Wijk [7] provide a starting point of this issue by displaying the distributions of attributes for each individual event using lists of bar charts. However, this method can be limited for the discovery of association between attributes of the same event or between multiple event types. This implies a need for a new visualization design that is able to represent categorical event types and multivariate attributes at the same time.

Interpretability: The chosen analysis model is a critical component in the pipeline of visual analytics [48]. In the pursuit of better analytical performance, recently developed visual analytics tools tend to leverage advanced machine learning or deep learning models with considerably high complexity. These, however, introduce issues of interpretability of the analysis results and a lack of control over the analytical process, both which are essential for high-impact analytical tasks such as precision medicine and financial investments [16]. To address such problems, there has been an increased research investment towards explainable artificial intelligence [66], [96], with the to uncover the inner workings of complex models. Even so, the mechanisms underlying these models can be difficult for non-expert users to understand. Thus, there is a high demand for visual analytics techniques that can organize, transform, and communicate model-level interpretations into comprehensible and actionable guidance. Some recent advancements [15], [35], [44] tackle this issue with a focus on a particular analytical tasks and analysis models, yet more generalizable techniques must be explored and developed in future research.

Causality Analysis: From our review of event sequence analysis techniques, we noticed that causality analysis for event sequence data has gained increased attention in the data mining community over the past years. Many causality analysis techniques have been proposed [119], [125] to uncover the cause-and-effect relationship between events. However, very few visual analytics techniques have been developed for causality analysis of event sequences. Despite that some existing visual analytics methods are developed for analyzing multivariate data [105], [106], the temporal nature and high dimensionality of event sequence data can lead to additional challenges as discussed in Section 4.5, which is worth addressing in future research.

7 CONCLUSION

This paper presents a survey of visual analytics approaches for event sequence data. The survey proposed a taxonomy that includes a design space and a collection of primary

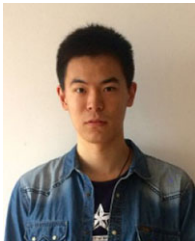
analytical tasks for characterizing the state-of-the-art techniques. In particular, the techniques are partitioned by five categories of analytical tasks, and featured by their corresponding design elements in the design space. It also illustrates the major applications of the techniques through a more domain-specific summary. Finally, the paper discusses the remaining challenges, and points out promising future research directions. With this survey, we connect prior studies in this topic by fitting them together into our taxonomy. We hope our work could provide practitioners with an overview of the alternatives approaches, and help them find the most appropriate design components in developing an effective visual analytics solution that addresses their analytical tasks at hand.



Yi Guo received his M.S. degree in Financial Mathematics from the University of New South Wales, Australia in 2019. He is currently working toward his Ph.D. degree as part of the Intelligent Big Data Visualization (iDV^x) Lab, Tongji University. His research interests include data visualization and deep learning.



Shunan Guo received her Ph.D. degree in Software Engineering from East China Normal University, Shanghai, China. Her research interests include visual analytics and human-computer interaction, especially visual analytics approaches for temporal event sequences. For more information, please visit <http://guoshunan.com/>.



Zhuochen Jin received his B.S. degree in Computational Mathematics from Zhejiang University, China in 2017. He is currently working toward his Ph.D. degree as part of the Intelligent Big Data Visualization (iDV^x) Lab, Tongji University. His research interests include artificial intelligence and data visualization.



Smiti Kaul received her Bachelor's degrees in Computer Science and Mathematical Statistics from Wake Forest University, NC, USA. She is currently working towards an M.S. in Computer Science at the University of North Carolina at Chapel Hill, NC, USA, where she is a part of the Visual Analysis and Communication Lab.



David Gotz received his Ph.D. in Computer Science from the University of North Carolina (UNC) at Chapel Hill, NC, USA in 2005. He is currently an Associate Professor of Information Science with the School of Information and Library Science at UNC Chapel Hill. He directs the Visual Analysis and Communication Lab and conducts research on a range of topics at the intersection of data visualization, HCI, machine learning, and statistical analysis. He is also the Assistant Director for the Carolina Health Informatics Program and an Associate Member of the UNC Lineberger Comprehensive Cancer Center. He spent nearly a decade as a Research Scientist at the IBM T.J. Watson Research Center, New York, NY, USA before returning to join the UNC faculty in 2014.

Before his Ph.D. studies at HKUST, he was a staff researcher at IBM China Research Lab, Beijing, China. He was a research staff member at the IBM T.J. Watson Research Center, New York, NY, USA before joining the Tongji faculty in 2016.



Nan Cao received his Ph.D. degree in Computer Science and Engineering from the Hong Kong University of Science and Technology (HKUST), Hong Kong, China in 2012. He is currently a professor at Tongji University and the Assistant Dean of the Tongji College of Design and Innovation. He also directs the Tongji Intelligent Big Data Visualization Lab (iDV^x Lab) and conducts interdisciplinary research across multiple fields, including data visualization, human computer interaction, machine learning, and data mining.

Before his Ph.D. studies at HKUST, he was a staff researcher at IBM China Research Lab, Beijing, China. He was a research staff member at the IBM T.J. Watson Research Center, New York, NY, USA before joining the Tongji faculty in 2016.

REFERENCES

- [1] D. Bhattacharjya, K. Shanmugam, T. Gao, N. Mattei, K. R. Varshney, and D. Subramanian. Event-driven continuous time bayesian networks. In *AAAI Conference on Artificial Intelligence*, pp. 3259–3266, 2020.
- [2] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Time-lines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2016.
- [3] N. Cao, D. Gotz, J. Sun, and H. Qu. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2581–2590, 2011.
- [4] N. Cao, Y.-R. Lin, F. Du, and D. Wang. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications*, 36(5):72–81, 2015.
- [5] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, 2012.
- [6] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):280–289, 2015.
- [7] B. C. Cappers and J. J. van Wijk. Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):532–541, 2017.
- [8] D. Chankihort, B.-M. Lim, G.-J. Lee, S. Choi, S.-O. Kwon, S.-H. Lee, J.-T. Kang, A. Nasridinov, and K.-H. Yoo. A visualization scheme with a calendar heat map for abnormal pattern analysis in the manufacturing process. *International Journal of Contents*, 13(2):21–28, 2017.
- [9] M. Chen, A. Trefethen, R. Banares-Alcantara, M. Jirotko, B. Coecke, T. Ertl, and A. Schmidt. From data analysis and visualization to causality discovery. *Computer*, (10):84–87, 2011.
- [10] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu. Peakvizor: Visual analytics of peaks in video clickstreams from massive open online courses. *IEEE Transactions on Visualization and Computer Graphics*, 22(10):2315–2330, 2015.
- [11] S. Chen, S. Chen, Z. Wang, J. Liang, Y. Wu, and X. Yuan. D-map+ interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. *ACM Transactions on Intelligent Systems and Technology*, 10(1):1–26, 2018.
- [12] S. Chen, S. Li, S. Chen, and X. Yuan. R-map: A map metaphor for visualizing information reposting process in social media. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1204–1214, 2019.
- [13] Y. Chen, A. Puri, L. Yuan, and H. Qu. Stagemap: Extracting and summarizing progression stages in event sequences. In *IEEE International Conference on Big Data*, pp. 975–981. IEEE, 2018.
- [14] Y. Chen, P. Xu, and L. Ren. Sequence synopsis: Optimize visual summary of temporal event data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):45–55, 2017.
- [15] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*, pp. 3504–3512. Curran Associates, 2016.
- [16] J. Choo and S. Liu. Visual analytics for explainable deep learning. *IEEE Computer Graphics and Applications*, 38(4):84–92, 2018.

- [17] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.
- [18] T. N. Dang, P. Murray, J. Aurisano, and A. G. Forbes. Reaction-flow: an interactive visualization tool for causality analysis in biological pathways. In *BMC proceedings*, vol. 9, p. S6. BioMed Central, 2015.
- [19] F. Du, S. Guo, S. Malik, E. Koh, S. Kim, and Z. Liu. Interactive event sequence prediction for marketing analysts. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2020.
- [20] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Eventaction: Visual analytics for temporal event sequence recommendation. In *Visual Analytics Science and Technology*, pp. 61–70. IEEE, 2016.
- [21] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Finding similar people to guide life choices: Challenge, design, and evaluation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 5498–5544. ACM, 2017.
- [22] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Visual interfaces for recommendation systems: Finding similar and dissimilar peers. *ACM Transactions on Intelligent Systems and Technology*, 10(1):1–23, 2018.
- [23] N. Elmqvist and P. Tsigas. Growing squares: Animated visualization of causal relations. In *Proceedings of the ACM symposium on Software Visualization*, pp. 17–ff, 2003.
- [24] N. Elmqvist and P. Tsigas. Animated visualization of causal relations through growing 2d geometry. *Information Visualization*, 3(3):154–172, 2004.
- [25] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *IEEE Symposium On Visual Analytics Science And Technology*, pp. 167–174. IEEE, 2006.
- [26] X. Fan, Y. Peng, Y. Zhao, Y. Li, D. Meng, Z. Zhong, F. Zhou, and M. Lu. A personal visual analytics on smartphone usage data. *Journal of Visual Languages & Computing*, 41:111–120, 2017.
- [27] F. Fischer, J. Fuchs, P.-A. Vervier, F. Mansmann, and O. Thonnard. Vistracer: a visual analytics tool to investigate routing anomalies in traceroutes. In *Proceedings of the International Symposium on Visualization for Cyber Security*, pp. 80–87, 2012.
- [28] L. Franklin, C. Plaisant, K. Minhazur Rahman, and B. Shneiderman. Treatmentexplorer: An interactive decision aid for medical risk communication and treatment exploration. *Interacting with Computers*, 28(3):238–252, 2016.
- [29] D. Gotz and H. Stavropoulos. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1783–1792, 2014.
- [30] D. Gotz, S. Sun, and N. Cao. Adaptive contextualization: Combating bias during high-dimensional visualization and data selection. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 85–95. ACM, 2016.
- [31] D. Gotz, J. Zhang, W. Wang, and J. Shrestha. Visual analysis of high-dimensional event sequence data via dynamic hierarchical aggregation. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 2020.
- [32] M. C. Goulden, E. Gronda, Y. Yang, Z. Zhang, J. Tao, C. Wang, X. Duan, G. A. Ambrose, K. Abbott, and P. Miller. Ccvis: Visual analytics of student online learning behaviors using course clickstream data. *Electronic Imaging*, 2019(1):681–1, 2019.
- [33] R. Guo, T. Fujiwara, Y. Li, K. M. Lima, S. Sen, N. K. Tran, and K.-L. Ma. Comparative visual analytics for assessing medical records with sequence embedding. *Visual Informatics*, 2020.
- [34] S. Guo, F. Du, S. Malik, E. Koh, S. Kim, Z. Liu, D. Kim, H. Zha, and N. Cao. Visualizing uncertainty and alternatives in event sequence predictions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.
- [35] S. Guo, Z. Jin, Q. Chen, D. Gotz, H. Zha, and N. Cao. Visual anomaly detection in event sequence data. In *IEEE International Conference on Big Data*, pp. 1125–1130. IEEE, 2019.
- [36] S. Guo, Z. Jin, D. Gotz, F. Du, H. Zha, and N. Cao. Visual progression analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018.
- [37] S. Guo, K. Xu, R. Zhao, D. Gotz, H. Zha, and N. Cao. Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):56–65, 2017.
- [38] Y. Han, A. Rozga, N. Dimitrova, G. D. Abowd, and J. Stasko. Visual analysis of proximal temporal relationships of social and communicative behaviors. In *Computer Graphics Forum*, vol. 34, pp. 51–60. Wiley Online Library, 2015.
- [39] H. He, B. Dong, Q. Zheng, and G. Li. Vuc: Visualizing daily video utilization to promote student engagement in online distance education. In *Proceedings of the ACM Conference on Global Computing Education*, pp. 99–105, 2019.
- [40] D. Herr, F. Beck, and T. Ertl. Visual analytics for decomposing temporal event series of production lines. In *22nd International Conference Information Visualisation*, pp. 251–259. IEEE, 2018.
- [41] M. A. Hibbs, N. C. Dirksen, K. Li, and O. G. Troyanskaya. Visualization methods for statistical analysis of microarray clusters. *BMC Bioinformatics*, 6(1):115, 2005.
- [42] O. Huisman and P. Forer. The complexities of everyday life: balancing practical and realistic approaches to modeling probable presence in space-time. In *The Annual Colloquium of the Spatial Information Research Centre*, pp. 155–167. Citeseer, 2005.
- [43] W. Jentner and D. A. Keim. Visualization and visual analytic techniques for patterns. In *High-Utility Pattern Mining*, pp. 303–337. Springer, 2019.
- [44] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, and N. Cao. Carepre: An intelligent clinical decision assistance system. *ACM Transactions on Computing for Healthcare*, 1(1):1–20, 2020.
- [45] J. Jo, J. Huh, J. Park, B. Kim, and J. Seo. Livegantt: Interactively visualizing a large manufacturing schedule. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2329–2338, 2014.
- [46] N. R. Kadaba, P. P. Irani, and J. Leboe. Visualizing causal semantics using animations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1254–1261, 2007.
- [47] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. Van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011.
- [48] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In *Information Visualization*, pp. 154–175. Springer, 2008.
- [49] J. Kiernan and E. Terzi. Constructing comprehensive summaries of large event sequences. *ACM Transactions on Knowledge Discovery from Data*, 3(4):21, 2009.
- [50] B. Koldehofe, M. Papatriantafilou, and P. Tsigas. Distributed algorithms visualisation for educational purposes. In *Proceedings of the Annual SIGCSE/SIGCUE ITiCSE Conference on Innovation and Technology in Computer Science Education*, pp. 103–106, 1999.
- [51] J. Krause, A. Perer, and K. Ng. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 5686–5697. ACM, New York, 2016.
- [52] J. Krause, A. Perer, and H. Stavropoulos. Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):91–100, 2015.
- [53] M.-P. Kwan. Gender and individual access to urban opportunities: a study using space-time measures. *The Professional Geographer*, 51(2):210–227, 1999.
- [54] B. C. Kwon, V. Anand, K. A. Severson, S. Ghosh, Z. Sun, B. I. Frohnert, M. Lundgren, and K. Ng. Dpvis: Visual analytics with hidden markov models for disease progression pathways. *IEEE Transactions on Visualization and Computer Graphics*, 2020.
- [55] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [56] B. C. Kwon, J. Verma, and A. Perer. Peekquence: Visual analytics for event sequence data. In *ACM SIGKDD 2016 Workshop on Interactive Data Exploration and Analytics*, vol. 1, 2016.
- [57] P.-M. Law, Z. Liu, S. Malik, and R. C. Basole. Maqui: Interweaving queries and pattern mining for recursive event sequence exploration. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):396–406, 2018.
- [58] W. Li, M. Funk, Q. Li, and A. Brombacher. Visualizing event sequence game data to understand players skill growth through behavior complexity. *Journal of Visualization*, 22(4):833–850, 2019.
- [59] Z. Liu, H. Dev, M. Dontcheva, and M. Hoffman. Mining, pruning and visualizing frequent patterns for temporal event sequence

- analysis. In *Proceedings of the IEEE VIS Workshop on Temporal & Sequential Event Analysis*, pp. 2–4, 2016.
- [60] Z. Liu, B. Kerr, M. Dontcheva, J. Grover, M. Hoffman, and A. Wilson. Coreflow: Extracting and visualizing branching patterns from event sequences. In *Computer Graphics Forum*, vol. 36, pp. 527–538. Wiley Online Library, 2017.
- [61] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson. Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):321–330, 2016.
- [62] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, and B. Shneiderman. Cohort comparison of event sequences with balanced integration of visual analytics and statistics. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 38–49, 2015.
- [63] S. Malik, B. Shneiderman, F. Du, C. Plaisant, and M. Bjarnadottir. High-volume hypothesis testing: Systematic exploration of event sequence comparisons. *ACM Transactions on Interactive Intelligent Systems*, 6(1):1–23, 2016.
- [64] H. Mei and J. M. Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, pp. 6754–6764, 2017.
- [65] A. Michotte, G. Thines, A. Costall, and G. Butterworth. La causalité perceptive. *Journal de Psychologie Normale Et Pathologique*, 60:9–36, 1963.
- [66] Y. Ming, S. Cao, R. Zhang, Z. Li, Y. Chen, Y. Song, and H. Qu. Understanding hidden memories of recurrent neural networks. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 13–24. IEEE, 2017.
- [67] M. Monroe. *Interactive event sequence query and transformation*. PhD thesis, 2014.
- [68] X. Mu, K. Xu, Q. Chen, F. Du, Y. Wang, and H. Qu. Moocad: Visual analysis of anomalous learning activities in massive open online courses. In *Eurographics/IEEE VGTC Conference on Visualization*, pp. 91–95, 2019.
- [69] C. Muelder, B. Zhu, W. Chen, H. Zhang, and K.-L. Ma. Visual analysis of cloud computing performance using behavioral lines. *IEEE Transactions on Visualization and Computer Graphics*, 22(6):1694–1704, 2016.
- [70] K. Ng, A. Ghoting, S. R. Steinhubl, W. F. Stewart, B. Malin, and J. Sun. Paramo: A parallel predictive modeling platform for healthcare analytic research using electronic health records. *Journal of biomedical informatics*, 48:160–170, 2014.
- [71] P. H. Nguyen, R. Henkin, S. Chen, N. Andrienko, G. Andrienko, O. Thonnard, and C. Turkay. Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 2020.
- [72] P. H. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, O. Thonnard, and J. Zouaoui. Understanding user behaviour through action sequences: from the usual to the unusual. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2838–2852, 2018.
- [73] C. B. Nielsen, S. D. Jackman, I. Birol, and S. J. Jones. Abyss-explorer: visualizing genome sequence assemblies. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):881–888, 2009.
- [74] A. Perer and D. Gotz. Data-driven exploration of care plans for patients. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 439–444. ACM, Paris, 2013.
- [75] A. Perer and J. Sun. Matrixflow: temporal network visual analytics to track symptom evolution during disease progression. In *AMIA Annual Symposium Proceedings*, vol. 2012, p. 716. American Medical Informatics Association, 2012.
- [76] A. Perer and F. Wang. Frequency: Interactive mining and visualization of temporal frequent event sequences. In *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 153–162. ACM, 2014.
- [77] A. Perer, F. Wang, and J. Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics*, 56:369–378, 2015.
- [78] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 221–227. ACM, 1996.
- [79] P. J. Polack, S.-T. Chen, M. Kahng, M. Sharmin, and D. H. Chau. Timestitch: Interactive multi-focus cohort discovery and comparison. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 209–210, 2015.
- [80] P. J. Polack Jr, S.-T. Chen, M. Kahng, K. D. Barbaro, R. Basole, M. Sharmin, and D. H. Chau. Chronodes: Interactive multifocus exploration of event sequences. *ACM Transactions on Interactive Intelligent Systems*, 8(1):1–21, 2018.
- [81] J. Qi, V. Bloemen, S. Wang, J. van Wijk, and H. van de Wetering. Stbins: visual tracking and comparison of multiple data sequences using temporal binning. *IEEE Transactions on visualization and computer graphics*, 26(1):1054–1063, 2019.
- [82] P. Riehmman, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization*, pp. 233–240. IEEE, 2005.
- [83] A. C. Robinson, D. J. Peuquet, S. Pezanowski, F. A. Hardisty, and B. Swedberg. Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data. *Cartography and Geographic Information Science*, 44(3):216–228, 2017.
- [84] J. Rogers, N. Spina, A. Neese, R. Hess, D. Brodke, and A. Lex. Composervisual cohort analysis of patient outcomes. *Applied Clinical Informatics*, 10(02):278–285, 2019.
- [85] P. Rosenthal, L. Pfeiffer, N. H. Miller, and P. Ohler. Visruption: Intuitive and efficient visualization of temporal airline disruption data. *Computer Graphics Forum*, 2013.
- [86] J. Rzeszotarski and A. Kittur. Crowdscape: Interactively visualizing user behavior and output. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*, p. 5562, 2012.
- [87] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2015.
- [88] P. Saraiya, C. North, and K. Duca. An evaluation of microarray visualization tools for biological insight. In *IEEE Symposium on Information Visualization*, pp. 1–8, 2004.
- [89] A. Sarikaya, M. Correll, J. M. Dinis, D. H. O’Connor, and M. Gleicher. Visualizing Co-occurrence of Events in Populations of Viral Genome Sequences. *Computer Graphics Forum*, 2016.
- [90] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein. Time series anomaly discovery with grammar-based compression. In *International Conference on Extending Database Technology*, pp. 481–492, 2015.
- [91] J. Seo and B. Shneiderman. Interactively exploring hierarchical clustering results [gene identification]. *Computer*, 35(7):80–86, 2002.
- [92] Y. Shi, Y. Liu, H. Tong, J. He, G. Yan, and N. Cao. Visual analytics of anomalous user behaviors: A survey. *IEEE Transactions on Big Data*, pp. 1–1, 2020.
- [93] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings IEEE Symposium on Visual Languages*, pp. 336–343, 1996.
- [94] J. Slack, K. Hildebrand, T. Munzner, and K. S. John. Sequencejuxtaposer: Fluid navigation for large-scale sequence comparison in context. In *German Conference on Bioinformatics*, pp. 37–42, 2004.
- [95] L. Stopar, P. Skraba, M. Grobelnik, and D. Mladenic. Streamstoc: exploring multivariate time series on multiple scales. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1788–1802, 2018.
- [96] H. Strobelt, S. Gehrmann, H. Pfister, and A. M. Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):667–676, 2017.
- [97] D. Sun, R. Huang, Y. Chen, Y. Wang, J. Zeng, M. Yuan, T.-C. Pong, and H. Qu. Planningvis: A visual analytics approach to production planning in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 2019.
- [98] G. Sun, T. Tang, T.-Q. Peng, R. Liang, and Y. Wu. Socialwave: visual analysis of spatio-temporal diffusion of information on social media. *ACM Transactions on Intelligent Systems and Technology*, 9(2):1–23, 2017.
- [99] G.-D. Sun, Y.-C. Wu, R.-H. Liang, and S.-X. Liu. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867, 2013.
- [100] J. Trümper, A. Telea, and J. Döllner. Viewfusion: Correlating structure and activity views for execution traces. In *Theory and Practice of Computer Graphics*, pp. 45–52. The Eurographics Association, 2012.
- [101] F. Viégas, M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant, and C. Wren. Google+ ripples: A native

- visualization of information flow. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 1389–1398, 2013.
- [102] K. Vrotsou. *Everyday mining: Exploring sequences in event-based data*. Doctoral thesis, Linköping University, The Institute of Technology, 2010.
- [103] K. Vrotsou, J. Johansson, and M. Cooper. Activitree: Interactive visual exploration of sequences in event-based data using graph similarity. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):945–952, 2009.
- [104] G. Wang, X. Zhang, S. Tang, H. Zheng, and B. Y. Zhao. Unsupervised clickstream clustering for user behavior analytics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 225–236, 2016.
- [105] J. Wang and K. Mueller. The visual causality analyst: An interactive interface for causal reasoning. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):230–239, 2015.
- [106] J. Wang and K. Mueller. Visual causality analysis made practical. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 151–161, 2017.
- [107] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 457–466, 2008.
- [108] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1049–1056, 2009.
- [109] X. Wang, W. Dou, W. Ribarsky, and R. Chang. Visualization as integration of heterogeneous processes. In *Visual Analytics for Homeland Defense and Security*, vol. 7346, p. 73460B. International Society for Optics and Photonics, 2009.
- [110] J. Wei, Z. Shen, N. Sundaresan, and K.-L. Ma. Visual cluster exploration of web clickstream data. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 3–12, 2012.
- [111] K. Wongsuphasawat and D. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop on Visual Analytics in Healthcare*, pp. 25–28. American Medical Informatics Association, 2011.
- [112] K. Wongsuphasawat and D. Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2659–2668, 2012.
- [113] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. Lifeflow: visualizing an overview of event sequences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1747–1756, 2011.
- [114] K. Wongsuphasawat and B. Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 27–34, 2009.
- [115] W. Wu, Y. Zheng, K. Chen, X. Wang, and N. Cao. A visual analytics approach for equipment condition monitoring in smart factories of process industry. In *IEEE Pacific Visualization Proceedings*, pp. 140–149, 2018.
- [116] Y. Wu, N. Cao, D. Gotz, Y.-P. Tan, and D. A. Keim. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia*, 18(11):2135–2148, 2016.
- [117] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1763–1772, 2014.
- [118] C. Xie, W. Xu, and K. Mueller. A visual analytics framework for the detection of anomalous call stack trees in high performance computing applications. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):215–224, 2018.
- [119] H. Xu, M. Farajtabar, and H. Zha. Learning granger causality for hawkes processes. In *International Conference on Machine Learning*, pp. 1717–1726, 2016.
- [120] K. Xu, Y. Wang, L. Yang, Y. Wang, B. Qiao, S. Qin, Y. Xu, H. Zhang, and H. Qu. Clouddet: Interactive visual analysis of anomalous performances in cloud computing systems. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1107–1117, 2019.
- [121] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):291–300, 2016.
- [122] H. Yu. Spatio-temporal gis design for exploring interactions of human activities. *Cartography and Geographic Information Science*, 33(1):3–19, 2006.
- [123] X. Yuan, Z. Wang, Z. Liu, C. Guo, H. Ai, and D. Ren. Visualization of social media flows with interactively identified key players. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 291–292, 2014.
- [124] E. Zraggen, S. M. Drucker, D. Fisher, and R. DeLine. (s—qu)eries: Visual regular expressions for querying and exploring event sequences. 2015.
- [125] W. Zhang, T. K. Panum, S. Jha, P. Chalasani, and D. Page. Cause: Learning granger causality from event sequences using attribution methods. *arXiv preprint arXiv:2002.07906*, 2020.
- [126] Y. Zhang, K. Chanana, and C. Dunne. Idmvis: Temporal event sequence visualization for type 1 diabetes treatment decision support. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):512–522, 2018.
- [127] Z. Zhang, D. Gotz, and A. Perer. Iterative cohort analysis and exploration. *Information Visualization*, 14(4):289–307, 2015.
- [128] Z. Zhang, K. T. McDonnell, E. Zadok, and K. Mueller. Visual correlation analysis of numerical and categorical data on the correlation map. *IEEE Transactions on Visualization and Computer Graphics*, 21(2):289–303, 2014.
- [129] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins. # fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, 2014.
- [130] J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann, and A. Wilson. Matrixwave: Visual comparison of event sequence data. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 259–268, 2015.
- [131] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22(2):419–435, 2019.
- [132] F. Zhou, X. Lin, X. Luo, Y. Zhao, Y. Chen, N. Chen, and W. Gui. Visually enhanced situation awareness for complex manufacturing facility monitoring in smart factories. *Journal of Visual Languages & Computing*, 44:58–69, 2018.