J Vis (2019) 22:1241–1255
https://doi.org/10.1007/s12650-019-00598-x

REGULAR PAPER

Dongming Han · Jiacheng Pan · Fangzhou Guo · Xiaonan Luo · Yingcai Wu · Wenting Zheng · Wei Chen

RankBrushers: interactive analysis of temporal ranking ensembles

Received: 2 July 2019 / Accepted: 19 July 2019 / Published online: 23 September 2019

Abstract Temporal ranking ensembles indicate time-evolving multivariate rankings. Such data can be commonly found in our daily life, for example, different rankings of universities (QS, ARWU, THE, and USNews) over year and those of NBA players over season. Effective analysis and tracking of rankings allow users to gain insights into the overall ranking change over time and seek the explanation for the change. This paper introduces a novel visual analytics approach for characterizing and visualizing the uncertainty, dynamics, and differences of ranking ensemble data. A novel visual design is proposed to characterize the evolution pattern, distribution, and uncertainty of a large number of temporal ranking ensembles. The evolutionary ranking ensembles are progressively explored, tracked, and compared by means of an intuitive visualization system. Two case studies and a task-driven user study conducted on real datasets demonstrate the effectiveness and feasibility of the implemented system.

Keywords Visualization · Temporal ranking ensembles · Uncertainty

1 Introduction

Visualizing and analyzing ranking data has attracted much attention (Gousie et al. 2013; Gratzl et al. 2013; Shi et al. 2012). Although presenting ranking data with a ranking list is simple and effective in many situations, the complexity of ranking data in the real world greatly reduces the capability of ranking lists. In a more general scenario, a set of items may be ranked by different users, rating agencies, and models at multiple time steps, e.g., the searching results of the same keyword on different search engines like Google, Yahoo!, and Bing at different times.

It is difficult to analyze temporal multivariate datasets that are not originally ranked. Such datasets are generated from multiple sources. Different sources usually have different scales, and data size may change along time. Ranking the data (or the quantile normalization) is the simplest way to make the data comparable. Transforming data into rankings can benefit some data analysis tasks, such as analyzing a set of stocks based on multiple stock valuation models and finding similar NBA players with the same average score growth.

It is of great importance to inspect the ranking changes of one or more individuals in a group of entities, e.g., the trends of an NBA player in NBA player technical statistics rankings. We call dynamic multivariate rankings as *temporal ranking ensembles*. This kind of data are uncertain, temporal, and may contain

D. Han · J. Pan · F. Guo · Y. Wu · W. Zheng · W. Chen (✉)
State Key Lab of CAD and CG, Zhejiang University, Hangzhou 310058, Zhejiang, China
E-mail: chenwei@cad.zju.edu.cn

X. Luo
Guilin University of Electronic Technology, Guilin, China

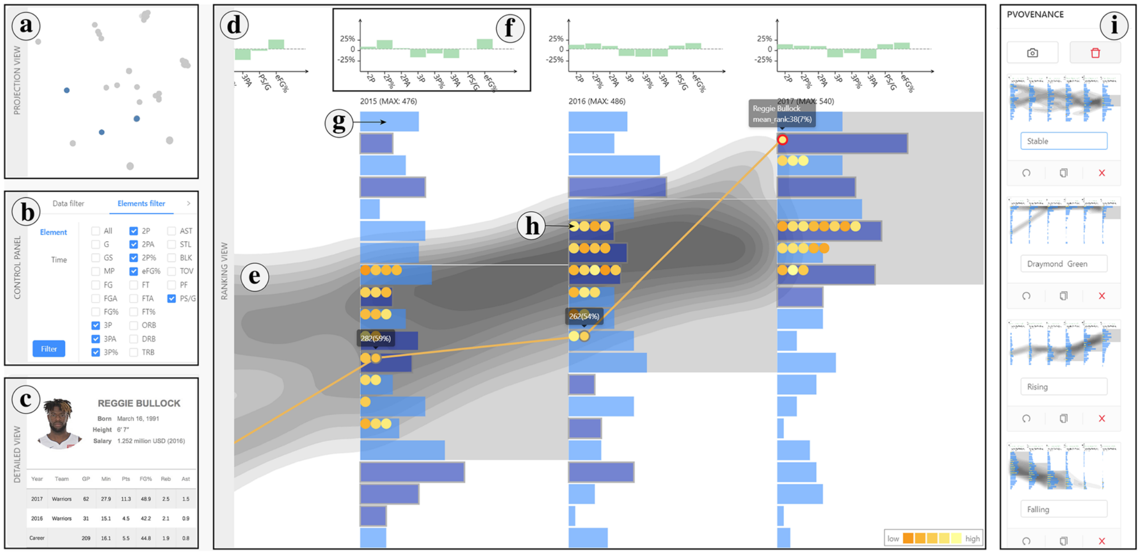


Fig. 1 The overview of RankBrushers. **a** The projection view. **b** The control panel. **c** The detailed view. **d** The ranking view. **e** The heatmap that depicts the evolution of ranking ensembles. **f** The bar chart that visualizes the deviation of elements of ranking ensembles. **g** The histogram that depicts the distribution of elements of ranking ensembles. **h** The circle glyph that represents a ranking ensemble. **i** The provenance view that lists the snapshots of the system

potential outliers and data misalignment due to the differences in the quality of data sources, the length of time steps, and the integrity of entity sets. The analytical tasks of temporal ranking ensembles can be classified into two categories. The first category is pattern analysis. Tracking and analyzing the changes in ranking ensembles over time help users identify the patterns, find ensembles with the desired trend, and find interesting ensembles different from the others in the desired trend. Because of the features of temporal ranking ensembles, similar trends do not indicate the consistency well. The second category is uncertainty and distribution analysis. Observing the uncertainty of ensembles and the distribution of elements in ensembles enables users to identify the consistency of different rankings, and the change of ensembles at an exact time step or along the timeline. Observing the evolutionary pattern of the distribution and the uncertainty can assist users in exploring ranking data.

There are few studies dedicated to the visualization of ranking data. LineUp (Gratzl et al. 2013) elegantly proposes an intuitive solution for analyzing multi-attribute rankings based on the flexible combination of attributes and refinement of parameters. RankExplorer (Shi et al. 2012) employs a flow-based visualization to analyze the ranking changes in temporal ranking data. However, it remains challenging to address the dynamics and uncertainty of temporal ranking ensembles, especially when the number of ensembles is large. Challenges for that are twofold. First, the evolutionary pattern of the ensembles is hidden in the temporal ensembles. Multiple values at each time step of an ensemble increase the complexity of both data processing and visualization. Second, there is no solution for visualizing the uncertainty and distribution of multiple ensembles. The box plot is effective for displaying a single ensemble, but cannot handle the visualization and comparison of a large number of ensembles. Meanwhile, analyzing time-varying ranking ensembles is inseparable from the comparison of ensembles. Existing ranking visualization techniques are not designed for temporal ranking ensembles (Gousie et al. 2013; Gratzl et al. 2013; Shi et al. 2012), thus cannot properly depict the uncertainty and the dynamics of the data.

In this paper, we contribute RankBrushers (Fig. 1), an interactive visualization approach that addresses the challenges of analyzing temporal ranking ensembles. We emphasize our attempts on two aspects. First, we propose a novel visual encoding scheme based on heatmap and histogram to depict the distribution and uncertainty of temporal ranking ensembles. Second, we design and implement an intuitive visual interface that supports comprehensive visualization and analysis of temporal ranking ensembles. The continuous evolution of temporal ranking ensembles over time can be explored and identified iteratively. In summary, the contributions of this paper are as follows:

- A novel visual design that characterizes the evolutionary pattern, distribution, and uncertainty of a large number of temporal ranking ensembles; and

- An visual analytics system that supports users to drill down in large-scale temporal ranking ensembles by hovering, brushing, and filtering.

2 Related work

In this section, we review related work in ranking data visualization, uncertainty visualization, and ensemble data visualization.

2.1 Ranking data visualization

The ranking is a common and effective way to display observations in a priority order based on users' interests. Although making ranked lists is a simple way to visualize ranking data, it is time-consuming and unpractical when analytical tasks become complex or when the data are large and high dimensional. Seo and Shneiderman (2005) used a ranked list with color encoding for exploring high-dimensional ranking data. Visualization techniques such as heatmap (Kidwell et al. 2008), scatter plot (Sun et al. 2010), clickstream sequence (Wei et al. 2012), multidimensional scaling (Chen et al. 2015; Zhao et al. 2019; Zhou et al. 2019a, b, 2017), and small multiples (Xia et al. 2017) are widely used to compare rankings. LineUp (Gratzl et al. 2013) introduces a series of linked stacked-bar charts to explore and analyze the changes in multi-attribute rankings by combining attributes and refining parameters interactively. However, it is difficult to find items with similar evolutions in a series of linked bar charts. Our work can characterize the evolution pattern and explore elements of the ranking ensembles.

Recently, the visualization of temporal ranking data attracts much attention (Weng et al. 2018). Parallel coordinates, which clearly show the ranking evolutions of multiple items, are used to display the rankings in Rank Clocks (Batty 2006). However, when the number of ranking items increases, severe visual clutter of lines is introduced. Flow-based techniques are used in RankExplorer (Shi et al. 2012) and TrajRank (Lu et al. 2015) to show the changes of items among different clusters with their evolutionary information. Xia et al. (2017) used several visual designs for displaying the temporal evolution of ranking data. However, the technique was not originally designed for temporal rankings ensembles, in which the ranking data are not only temporal but also multivariate.

2.2 Uncertainty visualization

Uncertainty spreads throughout the entire data analysis pipeline, including acquisition, transformation, and visualization (Riveiro 2007; Huang et al. 2019; Wang et al. 2016). In the past decades, a large number of techniques have been developed (Xia et al. 2017; Seipp et al. 2019). Glyph-based methods encode uncertainties into well-designed glyphs (Chen et al. 2019b). For example, vector glyphs can be adapted to represent uncertain information, such as variability in direction and magnitude (Wittenbrink et al. 1996). Potter et al. (2010) presented a new summary plot that incorporates a collection of descriptive statistics to highlight the salient features of the data. Hlawatsch et al. (2011) invented flow radar glyphs to visualize unsteady flow with static images. Various visual variables (Pothkow and Hege 2011) and visual representations (Dinesha et al. 2012) can be employed to show uncertainty. Chen et al. (2018) design a visual analysis system for understanding the human behaviors, detecting outliers, and exploring interesting patterns with heterogeneous spatiotemporal data.

Visual variables, such as color (Grigoryan and Rheingans 2004), brightness (Dinesha et al. 2012), and blurriness (Lee and Varshney 2002), can also be employed to encode uncertainty.

Similar to glyph-based approaches, geometry-based techniques adapt the basic geometry to represent uncertainty, including point (Grigoryan and Rheingans 2004), line (Zehner et al. 2010), cube (Schmidt et al. 2004), and surrounding volume (Pothkow and Hege 2011). Texture-based visualization techniques have proven to be useful in improving the perception of uncertainty-affected regions (Botchen et al. 2006). Being more specific efficient techniques in visualizing uncertainty are adopted in our work.

2.3 Ensemble data visualization

Ensemble data have been widely used in scientific fields like meteorology (Demir et al. 2014) and hydrology (Zappa et al. 2008). An ensemble dataset is generated by simulating multiple numerical models through user-defined parameters (Wilson et al. 2009; Guo et al. 2018). They can be regarded as one type of uncertainty. Visualizing, exploring, and analyzing ensemble data are challenging because of the intrinsic properties (e.g., multivariate, high-dimensionality, and uncertainty) of the data (Ma et al. 2018, 2017a, b, 2016). EnsembleLens (Xu et al. 2019) constructed ensembles and applies anomaly detection algorithms with multidimensional data based on ensemble analysis. LDA ensembles (Chen et al. 2019a) use topic modeling methods to generate ensembles and analyze the topic representing topical behaviors.

Means and variances of scalar quantities are the straightforward approaches in measuring the aggregated distribution of ensemble data (Potter et al. 2009; Wilson et al. 2009). In this paper, we show the evolution of ranking ensembles based on the means of elements and visualize the variances to remind users of the uncertainty of ranking ensembles at each time step.

3 Design goals

Suppose we have a series of ensemble rankings, $\mathbb{R} = \{R^t : t = 1, 2, \ldots, T\}$. At any time step t, $R^t = \{E^t_m : m = 1, 2, \ldots, M\}$, where M is the number of ranking ensembles or the length of the ranking list. For any ranking ensemble $E^t_m = \{e^t_{mn} : n = 1, 2, \ldots, N\}$, N denotes the number of elements (data models or data sources). The ranking ensembles at each time step also form a time series, $\mathbb{E}_m = \{E^t_m : t = 1, 2, \ldots, T\}$, which we call the mth temporal ranking ensembles, and $\mathbb{R} = \{\mathbb{E}_m : m = 1, 2, \ldots, M\}$.

In the design process, we organize several rounds of brainstormings to elicit the design requirements. The design goals are derived based on the collected design requirements as well as the literature review.

*G*1 *Multi-perspective Visual Representation* Due to the size and complexity of temporal ranking ensembles, it is impossible to directly visualize all of their content at once. Users need to explore and analyze the data from different perspectives to find patterns they are interested in, such as temporal trends, inter-rank uncertainty, and differences between items. Therefore, it is of great importance to provide a multi-perspective visual representation that present data subsets, depict the relations between data items, and reveal data patterns in different forms.

*G*2 *Multilayer Visual Design* In visualizing complex temporal ranking ensembles, different forms of data are displayed as visual objects and stacked in layered views. To make users judge the corresponding data subsets intuitively and involve attributes of each visual object, a multilayer visual design is required to reasonably distinguish the overall pattern, statistical data, and individual objects. It helps users gain the multilayer uncertainty of ranking ensembles like the temporal layer, the single-time layer, the evolution layer, and different constructions layer.

*G*3 *Multi-step Drill-Down Workflow* In analyzing ensemble datasets, the discovery of hidden patterns relies on the proper choice of presenting subsets according to the multilayer uncertainty. However, users require guidance during exploration, directed by inspecting similar items, the reasoning of rankings, and finding abnormals. As such, we propose a multi-step drill-down workflow that allows users to iteratively filter ranking ensembles to locate subsets of interests for further exploration, switch between different perspectives to check the findings, and come up with potential patterns through presented relations.

4 Our approach

In this section, we introduce the analysis pipeline and visual designs of RankBrushers.

4.1 Pipeline

The pipeline of RankBrushers is shown in Fig. 2. The system first loads the ranking data and processes them into temporal ranking ensembles. Then, the distance matrix is calculated by the DTW algorithm. Ensembles are visualized according to the projection result of t-SNE based on the similarity matrix. Besides, trends and deviations of ranking ensembles are calculated and visualized. Users can analyze patterns of temporal ranking ensembles progressively by interactions.

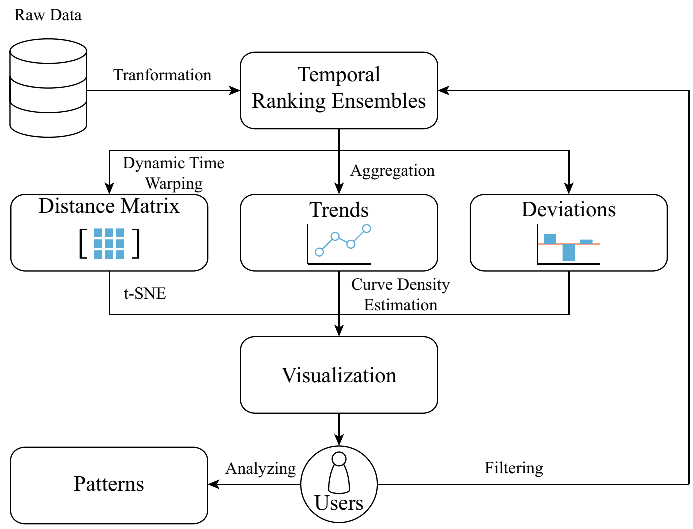


Fig. 2 The analysis procedure of RankBrushers

4.2 Interface

The interface consists of five components: a *projection view* that provides an overview of the temporal patterns of ranking ensembles; a *ranking view* that visualizes the details of ranking ensembles in an aggregated way; a *deviation view* that visualizes the deviation of elements of ranking ensembles; a *control panel* that filters ranking ensembles and their elements iteratively; a *provenance view* that saves and backtracks the snapshots.

4.2.1 The projection view

The projection view (Fig. 1a) provides an overview of all temporal ranking ensembles (G1) by projecting them onto a 2D plane with t-SNE. The similarity between temporal ranking ensembles is calculated by four steps: (1) calculate the average of the elements in ranking ensembles at each time step; (2) normalize all ranking ensembles at each time step; (3) calculate the similarity between two averaged temporal ranking ensembles by dynamic time warping (DTW) (Müller 2007), note that two ensembles are translated by $\mathbb{E}'_m = \mathbb{E}_m - \min(\mathbb{E}_m)$.

According to the similarity, ranking ensembles with similar evolution patterns are in the same clusters from the projection view. Thus, users can select clusters of ranking ensembles by lasso interaction. Moreover, the details of the selected ranking ensembles will be shown in the ranking view (G3).

4.2.2 The ranking view

The ranking view (Fig. 1d) visualizes the details of ranking ensembles. The ranking view consists of three major components, including a series of vertical histograms, circle glyphs, and a heatmap.

The histograms We overlay histograms (Fig. 1g) at each time step to show the distribution of elements in ranking ensembles at each time step (G1, G2) on the heatmap. Ranking lists at each time step are divided into a series of intervals and elements are binned and aggregated. The bars represent the ranking interval. From top to bottom, the ranking is from high to low. The height of each bar in the histograms shows the number of elements in the corresponding ranking interval.

The circle glyphs Circle glyphs (Fig. 1h) are embedded in the histograms to represent ranking ensembles at each time step (G2). At a specific time step, the position of a ranking ensemble's circle glyph is determined by the average of elements in the ensemble. The corresponding ranking of a circle glyph is the same as the ranking interval of each bar in the histograms. The color of the circle glyphs encodes the deviation of the elements in ranking ensembles, which indicates the uncertainty (G1). A line links two circle glyphs, which represent the same ranking ensemble, and represents the trend of a ranking ensemble. The combination of circles and rectangles appears in many existing works and has good performances (Yue et al. 2019).

The heatmap We use a heatmap (Fig. 1e) as the basic representation to enable users to identify the evolution of ranking ensembles (G1). The heatmap is generated by three steps: (1) determine the position of ensembles at each time step; (2) connect the same ensembles at different time steps; (3) use curve density estimation (CDE) to generate the heatmap of the lines. In our first design, a sankey diagram (Riehmann et al. 2005) is used. However, there will be a large number of lines among time steps when the number of ranking ensembles is large, resulting in severe visual clutter. Therefore, we use the heatmap to avoid visual clutter, which is also well-suited for displaying the overall patterns (G2).

We use a contoured design for the heatmap rather than a continuous color mapping to avoid too many colors in the view and make it easier to distinguish when multiple evolution patterns are superimposed. Moreover, there is a switch to turn off the heatmap for avoiding to observe circles representing ranking ensembles. In the evaluation section, we will introduce our experiments about the effects of the heatmap and compare it with other visual designs.

4.2.3 The deviation view

The bar charts on top of the axes (Fig. 1f) encode the deviation of the elements in the entire ranking ensembles at the time steps (G1). It is effective for analyzing the impact of the elements for ranking ensembles. The baseline in the bar charts represents a zero deviation. The height of a bar encodes the deviation. The bar above the baseline represents that the deviation of the element is higher than the average ranking, and the bar below the baseline represents a lower deviation (G2).

4.2.4 The control panel

The control panel (Fig. 1b) enables users to filter ranking ensembles and elements in ensembles (G3). It helps users to iteratively adjust the parameter combination and observe updated visualization results in other views.

4.2.5 The provenance view

The provenance view (Fig. 1i) enables users to record snapshots of the ranking view and the deviation view when they find interesting patterns. Snapshots are juxtaposed vertically to support visual comparison among different evolution patterns. Users can restore or delete snapshots in this view (G3).

4.3 Interactions

Hovering When users hover on a bar in the histogram of the ranking view, certain circle glyphs will be highlighted if the elements of the corresponding ensembles belong to the ranking interval. When users hover on a circle glyph, circle glyphs representing the same ranking ensemble and the links connecting them will be highlighted. Besides, certain bars in the histogram would be highlighted if the elements of the corresponding ensembles belong to the ranking interval, which shows the distribution of elements in the ensemble. Average rankings and names of ranking ensembles are shown by tooltips around the circle glyphs at each time step. Users can observe the patterns and the uncertainty of a specific ranking ensemble.

Brushing Hierarchical brushing in the ranking view assists users in selecting ranking ensembles. It enables users to filter specific patterns of ranking ensembles by brushing on different ranking intervals at different time steps. For example, when users are interested in ranking ensembles with an increasing pattern, they can brush increasing ranking intervals step by step along the time axis. Visualizations will be updated according to the filtering result.

Lasso Lasso is supported in the projection view to enable users to select clusters of ranking ensembles. We support two modes of lasso, including a standard lasso and a semiautomatic lasso. The second mode is an enhancement of the first one by automatically selecting a cluster of ranking ensembles after users select a specific ranking ensemble as a seed.

Details on demand RankBrushers provides the different level of detail of ranking ensembles by linking the projection view, the ranking view, and the detailed view together. Starting from the projection view, users can select clusters of the ranking ensembles and explore the overall temporal pattern and the element distribution of clusters. When users find a ranking ensemble with an abnormal or interesting pattern, they can further analyze its details in the detailed view. The element distribution helps users decide which elements are unnecessary so that they can be filtered in the control panel. In this way, users can analyze temporal ranking ensembles iteratively by following the loop of selecting, filtering, and exploring.

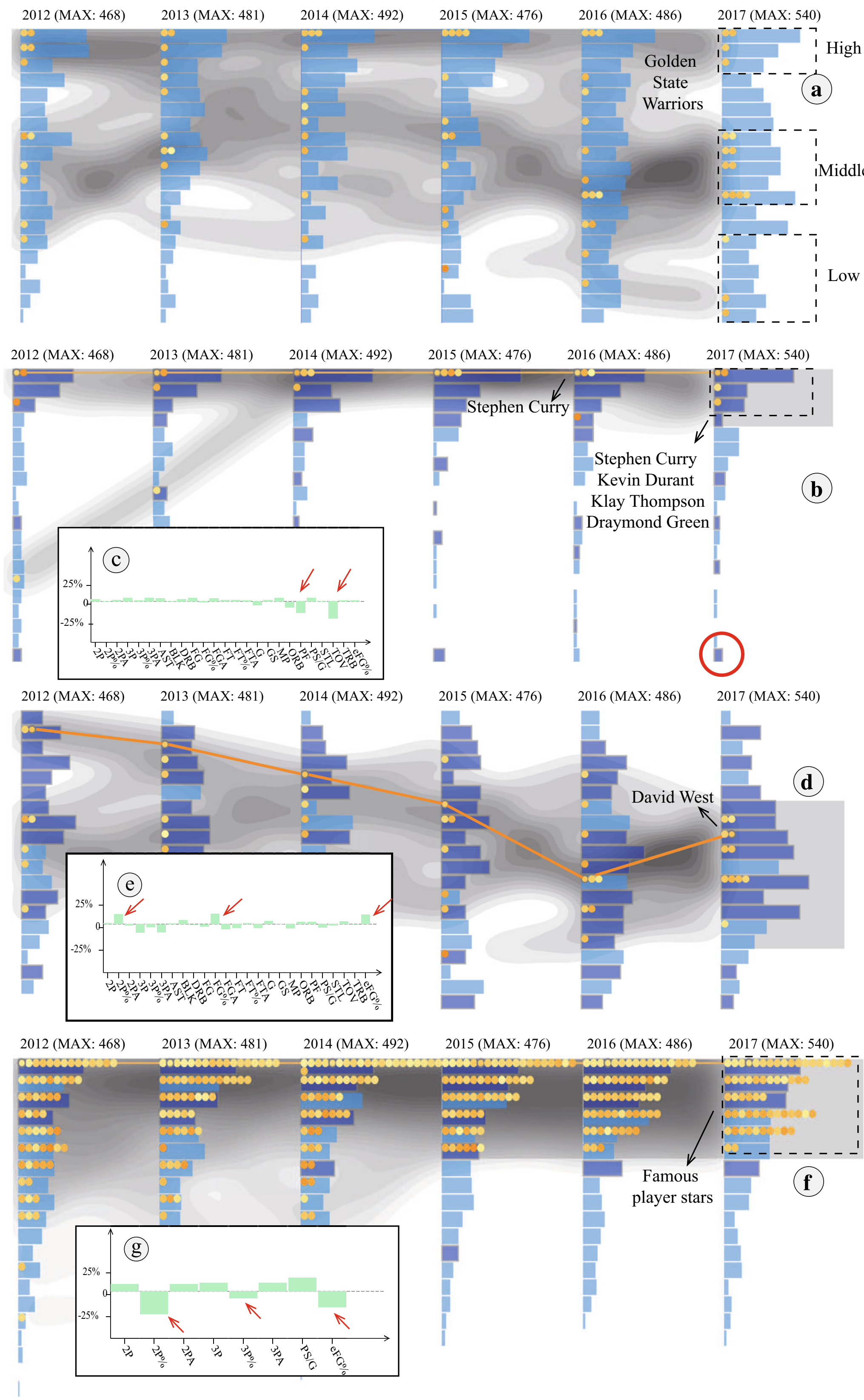
4.4 Usage scenario

We demonstrate the usage and analysis pipeline of RankBrushers by introducing how an NBA fan, Ted, explores and analyzes the NBA ranking dataset. The dataset records stats of 934 NBA players in each season from 2012–2013 to 2017–2018. Stats of an NBA plater contain 25 technical statistics. By ordering all NBA players under different statistics in each season, original dataset is transformed into a temporal ranking ensembles dataset. Additional information includes positions players play and teams to which players belong. The more specific technical statistics name: FG%: field goal percentage, 3P: 3-point field goals per game, 3PA: 3-point field goal attempts per game, 3P%: FG% on 3-Pt FGAs per game 2P: 2-point field goals per game, 2PA: 2-point field goal attempts per game, 2P%: FG% on 2-Pt FGAs per game, eFG%: effective field goal percentage per game, PS/G: points per game, TOV: turnovers per game, PF: personal fouls per game.

Ted wants to explore the temporal ranking ensembles of NBA players in Golden State Warriors (GSW) which was the champion in season 2017–2018. He selects players who played for GSW in season 2017–2018 in the control panel. Ranking ensembles of these players are visualized in the ranking view (see in Fig. 3a). He observes that players cluster into three categories in the high, middle, and low ranking intervals (G1, G2). The high, middle and low ranking intervals are a visual effect. Here, we use this word just to explain the visualization. After brushing on corresponding rank intervals, he finds out that the three categories of ensembles have different evolution patterns. He hovers on circle glyphs to explore the trend of different players (G1). He finds that Stephen Curry, Kevin Durant, Klay Thompson, and Draymond Green in the higher-ranking interval (Fig. 3b) are very stable over the years; they are all-star players and in the peak period (Fig. 3b). The evolutions of players in the middle of the ranking list are from high to middle or always in the middle, like David West, a famous player, has passed the peak period (the highlighted lines in Fig. 3d). The players in lower-ranking intervals are new players with a short trend. He brushes the different ranking intervals to focus them (G1). He notices that players in different ranking intervals have different deviations of elements from the bar chart. “TOV” and “PF” of players in the high-ranking interval are low (Fig. 3c). “FG%,” “eFG%,” and “2P%” of players in the middle-ranking interval are high (Fig. 3e). “PF” of players in the lower-ranking interval is high.

He finds the highest ranked player Stephen Curry has elements which in the lowest-ranking interval (Fig. 3b). Moreover, the lowest ranked player Chris Boucher has elements which in the highest-ranking interval. He clicks the circle glyph for seeing the evolution and distribution of elements in the detailed view (G1). He finds each element of Chris Boucher is lower ranked except the elements “PF,”, and the evolutions of elements of Stephen Curry is more stead over the years, and the color of lines in the plot violin shows the lower-ranked elements: “TOV” and “PF”(G1). He feels that these phenomena may be caused by minutes played. So he brushes players in higher-ranking intervals and sets “TOV” and “PF” into the element filter for refreshing the ranking view, and then sets “MP” as the parameters, finally sets eight elements about score ability (“2P,” “2PA,” “2P%,” “3P,” “3PA,” “3P%,” “PS/G”). He svae snapshots of each step in the provenance view (G3). He finds they almost have the same stable evolution, higher ranking with “MP” and elements of score ability, but lower ranking with “TOV” and “PF.” He finds Draymond Green is different from other players in the evolution with elements of score ability; Draymond Green has a lower-ranking score ability in 2012 and then grew rapidly to 2015, with fluctuations in recent years.

He finds these players are very close in the projection view which is recalculated by new elements of score ability. It means they have a similar evolution, so he selects surrounding player ranking ensembles from them in the projection view for looking up the other players who have a similar evolution (G3). Then, he finds lots of famous player stars (Kyrie Irving, Lebron James, and so on) in the refreshed ranking view with similar evolution pattern (Fig. 3f). He brushes the players from the highest-ranking interval. He finds they have a very high and stable temporal pattern. Then, he finds the deviation of elements of shooting (“2P%,” “3P%,” “eFG%”) is the lowest from the bar chart (Fig. 3g). He selects the three elements for refreshing; some players with higher-ranked elements of score ability have lower-ranked elements of shooting, and the entire evolution is more downward. The highest players are Stephen Curry and Kevin Durant, who are known as a shooter. The lowest players are Blake Griffin and Matt Barnes, who are not


2012 (MAX: 468)
2013 (MAX: 481)
2014 (MAX: 492)
2015 (MAX: 476)
2016 (MAX: 486)
2017 (MAX: 540)
Golden State Warriors
High
a
Middle
Low
Stephen Curry
Stephen Curry
Kevin Durant
Klay Thompson
Draymond Green
b
c
25%
0
-25%
2P 2P% 2PA 3P 3P% 3PA AST BLK DRB FG FG% FGA FT FT% FTA G GS MP ORB PF PS/G STL TOV TRB eFG%
David West
d
e
Famous player stars
f
g
2P 2P% 2PA 3P 3P% 3PA PS/G eFG%

◀ **Fig. 3** Usage: The two kinds of the temporal pattern of player ranking ensembles in different ranking intervals. **a** The evolution of NBA players played GSW in the 2017–2018 season. **b** The evolution of players from **a** in the higher-ranking interval of the 2017–2018 season. **c** The deviation of elements from **b** in the 2017–2018 season. **d** The evolution of players from **a** in the 2017–2018 season. **e** The evolution of players from **a** in 2017–2018 season. **f** The evolution of NBA players played GSW in the 2017–2018 season with score ability ranking. **g** The evolution of players from **f** in 2017–2018 season

known as a shooter. When he brushes the lowest players, he finds evolutions of them are significantly downward. Meantime, he finds that the ranking of the element "3P%" is higher than other elements from the detailed view and the bar chart. The three point skills of the two players are a little better than the other skills.

5 Evaluation

5.1 User study I: effectiveness of heatmap-based encoding for ranking ensembles

We conducted a user study to evaluate the effectiveness of our approach for representing the evolution of the ranking ensembles. Specifically, we focused on testing how visual encoding approaches affect the user's recognition of different patterns through a variety of synthetic datasets. We tested four different visual encoding approaches (Heinrich and Weiskopf 2013), including lines, curves, bundling, and heatmap, as illustrated in Fig. 4. Many existing works focused on visualizing evolution with different heatmap techniques (Roberts et al. 2019; Fua e al. 1999; Heinrich et al. 2011; Zhou et al. 2008). We only tested a based heatmap method in this study.

The lines visual encoding approach (E_L) connects successive points representing rank values of each ranking ensemble with straight lines, forming a series of polygon lines. The curves visual encoding approach (E_C) is similar to E_L except that curves are used. (In this study, we used B-spline curves.) The bundling approach (E_B) bundles the neighboring curves using the bundling algorithm in Gansner et al. (2011). The rendering parameters such as opacity, the width of lines, bandwidth of heatmap, and delta of edge bundling function were determined in the pilot studies to emphasize potential patterns as much as possible. We use an estimation method-inspired kernel density estimation (KDE). The method is implemented with a 2D normal kernel. Three parameters can be customized. The thresholds parameter of the KDE is set to 15. The cell size and the bandwidth of the method are set to 30 and 40, respectively. All of these parameters determine the estimation granularity.

Datasets Two synthetic datasets were tested: One has 1800 commonplace ranking ensembles (which would randomly evolve within a small range) along with 200 special ranking ensembles (which would evolve in a certain pattern); the other one has 2000 commonplace ranking ensembles (Fig. 4d). All those ranking ensembles would evolve in ten time steps.

We introduced three kinds of patterns (Fig. 4) : continuously fluctuating pattern (Fig. 4a), rising and then falling pattern (Fig. 4b), and continuously rising pattern (Fig. 4c).

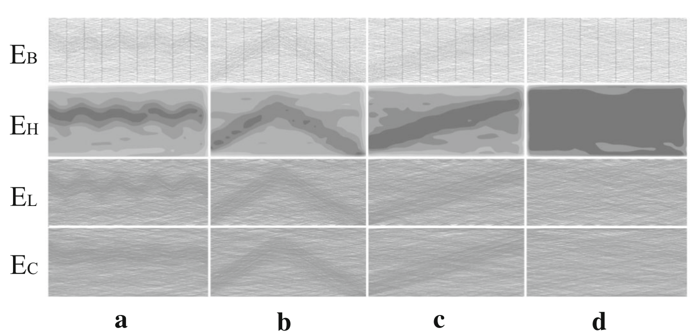


Fig. 4 The examples of the patterns produced by the different approaches. **a** The continuously fluctuating pattern. **b** The rising and then falling pattern. **c** The continuously rising pattern. **d** No pattern. Four visual encoding approaches: bundling (E_B) heatmap (E_H) Lines(E_L) Curves (E_C)

Hence, we got four kinds of datasets: three of them contained special patterns, and the other was generated without any pattern. To avoid bias, we randomly generated these synthetic datasets for three times so that we got 12 different datasets.

Variable The primary variable to test in this user study was the visual encoding of the ranking ensembles, which contains the four approaches: heatmap (E_{H}), line (E_{L}), curve (E_{C}), and edge bundling (E_{B}).

Hypotheses There are different levels of information loss rate with various visual encodings. One with high information loss rate such as E_{H}, however, is more intuitive in visualizing a general trend pattern because it integrates the details. It is important to balance the information loss and detail integration. E_{H} tends to be suitable for large-scale data because the usage of CDE delivers a high-level abstraction of the underlying data. Thus, the participants can finish the study trials more efficiently by using E_{H}. However, using E_{H} may lead to heavy information loss about the evolution of ranking ensembles, and consequently wrong understandings to the ranking ensembles. In contrast, E_{L} and E_{C} keep all information about the evolution of ranking ensembles so that the accuracy would be higher. Surely, the visualization processes are longer than E_{H}. In terms of the efficiency, E_{C} behaves worse than E_{L} because its curved lines distort the trends. Compared with E_{L}, it is also more time-consuming for discovering patterns and may lead to a higher error rate. Among all these designs, E_{B} has a moderate performance as well as a middle error rate. Based on these observations, we hypothesize:

H1: Participants will spend less time with E_{H} than the others.
H2: There will be a higher error rate in E_{H} than the others.
H3: Participants will spend more time with E_{C} than E_{L}, and there will be a higher error rate in E_{C}.

Participants and Apparatus We recruited 14 participants (seven males and seven females, aging from 20 to 30, average 23.15, median 23). None of them was involved in developing our approach and system. All the participants are college students with background knowledge in visualization and computer graphics, normal or corrected visual acuity, and no color blindness. Tasks are completed on a regular personal computer with Chrome Browser installed. Fourteen participants participated and completed our user study.

The heatmap was filled by a set of different gray colors just the same as our system. The curve, line, and edge bundling were all filled by a translucent gray color [rgba(128, 128, 128, 0.3)]. All the parameters were conducted by our pilot study, as discussed above.

Procedure In total, we tested four kinds of approaches with four kinds of patterns as mentioned above. Each participant should perform 48 randomized study trials (4 approaches × 12 datasets) as a session, yielding a total of 672 (18 trials × 14) trials.

Before the formal study, we conducted an informal study for participants to practice until they got familiar with all the tasks, approaches, and patterns.

We randomized the order of the study trials in each formal study session. Moreover, in each study trial, participants were asked to select one of the four answers about the pattern (as mentioned above) in this trial. After that, a 5-point Likert scale was used for participants to rate their confidence in the answer, the aesthetics, and the comprehensibility of the visual encoding approach in this trial from 1 to 5.

We did not limit the time of each trial, and each formal study session took about 15 min.

Results We recorded the answering time and accuracy (1 for true, 0 for false) for each trial. The answering time is the time participants spent on selecting the kind of pattern, but not the total time of each trial (which includes computation time, rendering time, answering time, and rating time). In Fig. 5, we summarized our results with 95% confidence intervals, including accuracy, answering time, confidence, aesthetic, and comprehensibility.

For answering time, we did ANOVA test for multiple approaches comparisons and accuracy; we did the Friedman test.

Accuracy There was no significant difference in accuracy among different approaches at $p < .05$ ($\chi^2(3) = 3.5357, p = .31616$), which rejected *H2* and partially rejected *H3*.

Answering time In Fig. 5, it is obvious that the mean answering time of E_{H} is less than others. After the ANOVA test, we found a significant difference among those approaches at $p < .05$ ($f = 4.59249, p = .003447$). So we did pairwise t tests between those approaches. Significant statistics difference was found in multiple comparisons ($p < .05$), Except for two comparisons: one was E_{C} and E_{L}, and the other was E_{B} and E_{L}. The above findings supported *H1* but partially rejected *H3*.

Confidence, Aesthetic, and Comprehensibility Overall, the mean values of all the results about participants' subjective feelings ranked E_{H} as the best and E_{C} as the worst. The performance of E_{B} and E_{L} was similar.

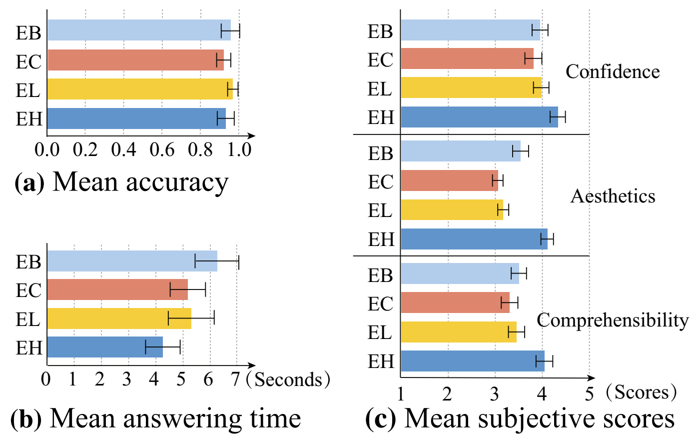


Fig. 5 The result of different approaches in the user study 1. Error bars represent 95% CIs. **a** The mean accuracy, **b** the mean answering time, **c** the mean confidence, aesthetics, and comprehensibility

In conclusion, we rejected *H2* and *H3*. *H1* was supported by our findings. Nevertheless, participants' subjective feelings showed that E_{H} was their favorite, and E_{C} was less preferred.

Feedbacks and Discussion We found that E_{H} behaves much better in answering time than the other three approaches. We believed that E_{H} conducts a smooth view of the ranking ensembles' evolution, which reduces the visual disturbance and emphasize patterns. Quite a big part of our participants mentioned that "Heatmap is the most intuitive approach for pattern exploration." This was consistent with our hypotheses. However, the heatmap also has a bias in the visualization and mislead users to understand the visualization. There are some existing works that focus on the problems of the heatmap. In this study, the users did not have this confusion about this problem.

Moreover, we found that there is no significant difference in accuracy among different approaches. It rejected *H2*. In terms of computation and rendering cost, E_{H} would work better than the other three approaches (especially with large-scale datasets) which were verified by our pilot study. Since there is no significant difference in accuracy, and E_{H} boosts pattern exploration with low cost, we chose E_{H} as our visual encoding approach.

As for *H3*, the findings rejected it. Nevertheless, we found that E_{C} distorts the evolution trends, so E_{C} might not work well with some subtle patterns such as the continuously fluctuating pattern shown in Fig. 4a.

5.2 User study II: effectiveness of RankBrushers

We conducted a task-based user study to evaluate the effectiveness of RankBrushers in temporal ranking ensembles.

Participants and Data We recruited ten participants (six males and four females). These participants are identified by S1–S10, respectively. We used the NBA dataset above in this user study.

Tasks and Procedure We demonstrated RankBrushers for each participant. We introduced each view and interaction of the system for each participant. Next, we gave them practice with answers. After that, the official dataset was loaded. The participant was asked to complete a series of tasks in Table 1 using RankBrushers. We set one or two tasks for each design goal so that each visualization component could be properly covered by the four tasks. After finishing tasks, participants used freely in our system to explore the dataset. Finally, the participants were asked to rate different aspects of RankBrushers on a 5-point Likert scale. The questions in the questionnaire are listed in Table 2.

Passing rates Each session of the task-based user study lasted 15–20 min. The passing rates (Fig. 6a) of four evaluation tasks are satisfactory. On average, the participants scored 3.94 out of the 4 on average. There were two participants misunderstood the color encoding of the circle glyph. Moreover, most of the participants completed the task very smoothly.

Post-study rating As shown in Fig. 6b, the results show that our system has a very high-level rating. Most of the participants agree that our system was intuitive and easy to use (Q1, Q2). They also agree that the heatmap is very useful for helping them analyze the evolution (Q5), especially when the number of

Table 1 The evaluation tasks of user study II

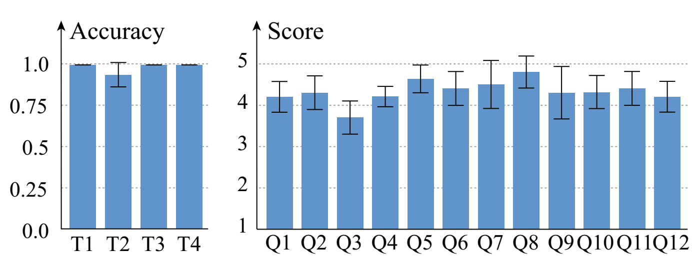
No.	Question	Goal
E1	Among the 2017–2018 HOU active players, find the top five players with the highest average ranking in the 2017–2018 season, and the player with an increasingly high average ranking in his career	G1, G2
E2	Among the 2017 BOS active players, find the top five players whose ranking ensemble is the most uncertainty in the 2016–2017 season, and which element has the most deviation with the element value	G1, G2
E3	Among the 2017–2018 BOS active players find how much players who have the element ranking in the lowest-ranking interval in the 2017–2018 season	G1
E4	Among the 2017–2018 CLE active players, find the players whose average ranking by scoring (3p%, 3PA, 2P, 2PA, 2P%, PS/G, eFG%) are top 10 in the 2017–2018 season, and then find the players have the lowest ranking from the average ranking of 2 points (2p%, 2PA, 2P%) from the ten players in 2017–2018 season	G1, G3

Table 2 The post-study questionnaire of the user study II

No.	Question
Q1	Is the interface of RankBrushers intuitive?
Q2	Is the interface of RankBrushers easy to use?
Q3	Is the interface of RankBrushers easy to learn?
Q4	Is the interface of the projection view intuitive and easy to use?
Q5	Does the heatmap help you analyze the evolution of the ranking ensemble?
Q6	Does the brushing help you analyze the evolution of the ranking ensemble?
Q7	Does the bar chat help you analyze the deviation of the elements?
Q8	Does the histogram help you analyze the distribution of the elements?
Q9	Does the color of the circle glyph help you analyze the uncertainty of the ranking ensembles?
Q10	Is the interface of the ranking view intuitive and easy to use?
Q11	Is the interface of the detailed view intuitive and easy to use?
Q12	Is the interface of the provenance view intuitive and easy to use?

ranking ensembles increases. They show high interest in the histogram (Q8), which is useful for analyzing the distribution of the elements. The bar chat that represents the deviation of elements also gets a high rating (Q7). However, the average rating for the question: "easy to learn" (Q3) was the lowest. But Visual guidance techniques (Ceneda et al. 2017) can help users to get familiar with our system quickly. Also, we find that the rating for the projection view was lower than other views. The reasons are: (1) there was less visual encoding in the projection view, so participants were not interested in it; and (2) they pay more attention to the single player with more detail and storytelling which is difficult to get from the projection view.

Feedback Participants were interviewed after they used our system. Most of them agreed that our system is very intuitive and useful. "The system is very smooth, I can easily get what I want from the visual components." Other two participants said that they found some interesting patterns from our system. One of them said "I have never known that Stephen Curry, with a high average ranking, has a particularly low technical ranking." The other said "I was surprised to find that Clint Capela's career average ranking grown so fast."



(a) Evaluation tasks **(b)** Post-study questionnaire

Fig. 6 Analysis of passing rates (**a**) of evaluation tasks (Table 1) and user rating (**b**) of post-study questionnaire, where (Table 2)

6 Discussion

The evaluation shows the effectiveness of our approach. In this section, we discuss the limitations and implications of our approach.

Limitations Our method is effective for analyzing temporal ranking ensembles. However, it still has much space to improve. First, the complexity of the backend algorithm is high. The complexity of DTW is $O(n^2m^2)$, and the complexity of t-SNE is $O(n^2)$, where n is the number of ranking ensembles and m is the length of time steps. Therefore, the current backend does not scale well on large datasets in real time. We plan to use parallel computing to accelerate these algorithms. Second, the heatmap in the ranking view does not show the uncertainty of trends. We plan to encode uncertainty of ranking ensembles into the heatmap in the future. Third, interactions in our method are limited, although current interactions can support users to analyze the data. In the future, we plan to support more interactions, including the comparison of different ranking ensembles and query ranking ensembles by sketching. Moreover, we plan to apply anomaly detection algorithms to calculate the anomaly scores for each ranking ensembles and visualize them as a ranking list. Finally, we use "SVG" to render each element on the web page; if the number of elements is more than 1000, the interactions of the system will be stuck. Moreover, when the number of ranking ensembles increases, the circle glyphs cause interface confusion. We plan to design new glyphs with aggregated and expanded interactions.

Implication In this paper, we present a visual analytics approach to analyze an important form of data, temporal ranking ensembles. Such data form can be found in various fields and scenarios, for example, university rankings, NBA player rankings, country rankings on the economy, and nutrition rankings of food. Our method is general and comprehensive and can support users to explore the data from different aspects, including temporal trends, inter-rank uncertainty, and differences between items. The visual design and the system are not only design for the temporal ranking ensembles but also apply the general ranking data and the multi-source heterogeneous ranking data. It is general and applicable to other scenes, like exploring the connections between the scholar rankings and the university rankings. Our system is flexible for setting the parameters to adjust the visualization effect, the algorithms, and the relation of mapping between the different components. In particular, the heatmap visual design is easy to expand to other scenes, like the graph layout and the high-dimensional parameter adjustment. We also can flexibly assemble the visual components for the visualization of different tasks of ranking data on demand.

7 Conclusion

This study introduces a novel visual design for analysis temporal ranking ensembles, including the temporal pattern and the uncertainty of ranking ensembles, the deviation, and the distribution of the elements. Moreover, then we present a novel visual analytics system, RankBrushers, that supports users to drill down in the large-scale temporal ranking ensembles by hovering, brushing, and filtering. Finally, we conducted two user studies to evaluate the effectiveness of heatmap-based encoding and the effectiveness of our system, respectively.

Acknowledgements This research supported by National Key Research and Development Program (2018YFB0904503) and National Natural Science Foundation of China (U1866602, 61772456, 61761136020).

References

Batty M (2006) Rank clocks. Nature 444(7119):592–596

Botchen RP, Weiskopf D, Ertl T (2006) Interactive visualization of uncertainty in flow fields using texture-based techniques. In: 12th international symposium on flow visualisation, pp 4051–4056

Ceneda D, Gschwandtner T, May T, Miksch S, Schulz H-J, Streit M, Tominski C (2017) Characterizing guidance in visual analytics. IEEE Trans Vis Comput Graph 1:111–120

Chen H, Zhang S, Chen W, Mei H, Zhang J, Mercer A, Liang R, Qu H (2015) Uncertainty-aware multidimensional ensemble data visualization and exploration. IEEE Trans Vis Comput. Graph 21(9):1072–1086

Chen S, Wang Z, Liang J, Yuan X (2018) Uncertainty-aware visual analytics for exploring human behaviors from heterogeneous spatial temporal data. J Vis Lang Comput 48:187–198

Chen W, Guo F, Han D, Pan J, Nie X, Xia J, Zhang X (2019a) Structure-based suggestive exploration: a new approach for effective exploration of large networks. IEEE Trans Vis Comput Graph 25(1):555–565

Chen S, Andrienko N, Andrienko G, Adilova L, Barlet J, Kindermann J, Nguyen PH, Thonnard O, Turkay C (2019b) LDA ensembles for interactive exploration and categorization of behaviors. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2019.2904069

Demir I, Dick C, Westermann R (2014) Multi-charts for comparative 3D ensemble visualization. IEEE Trans Vis Comput Graph 20(12):2694–2703

Dinesha V, Adabala N, Natarajan V (2012) Uncertainty visualization using HDR volume rendering. Vis Comput 28(3):265–278

Fua Y, Ward MO, Rundensteiner EA (1999) Hierarchical parallel coordinates for exploration of large datasets. In: IEEE visualization 1999, Proceedings, 24–29 October 1999. San Francisco, CA, USA, pp 43–50

Gansner ER, Hu Y, North S, Scheidegger C (2011) Multilevel agglomerative edge bundling for visualizing large graphs. In: 2011 IEEE pacific visualization symposium (PacificVis). IEEE, pp 187–194

Gousie MB, Grady J, Branagan M (2013) Visualizing trends and clusters in ranked time-series data. In: IS&T/SPIE electronic imaging. International Society for Optics and Photonics, pp 90170F–90170F

Gratzl S, Lex A, Gehlenborg N, Pfister H, Streit M (2013) Lineup: visual analysis of multi-attribute rankings. IEEE Trans Vis Comput Graph 19(12):2277–2286

Grigoryan G, Rheingans P (2004) Point-based probabilistic surfaces to show surface uncertainty. IEEE Trans Vis Comput Graph 10(5):564–573

Guo H, Liu H, Li R, Wu C, Guo Y, Xu M (2018) Margin & diversity based ordering ensemble pruning. Neurocomputing 275:237–246. https://doi.org/10.1016/j.neucom.2017.06.052

Heinrich J, Weiskopf D (2013) State of the art of parallel coordinates. In: Eurographics (STARs), pp 95–116

Heinrich J, Bachthaler S, Weiskopf D (2011) Progressive splatting of continuous scatterplots and parallel coordinates. Comput Graph Forum 30(3):653–662

Hlawatsch M, Leube P, Nowak W, Weiskopf D (2011) Flow radar glyphs—static visualization of unsteady flow with uncertainty. IEEE Trans Vis Comput Graph 17(12):1949–1958

Huang Z, Lu Y, Mack E, Chen W, Maciejewski R (2019) Exploring the sensitivity of choropleths under attribute uncertainty. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2019.2892483

Kidwell P, Lebanon G, Cleveland W (2008) Visualizing incomplete and partially ranked data. IEEE Trans Vis Comput Graph 14(6):1356–1363. https://doi.org/10.1109/TVCG.2008.181

Lee CH, Varshney A (2002) Representing thermal vibrations and uncertainty in molecular surfaces. In: Electronic Imaging 2002. International Society for Optics and Photonics, pp 80–90

Lu M, Wang Z, Yuan X (2015) Trajrank: exploring travel behaviour on a route by trajectory ranking. In: 2015 IEEE pacific visualization symposium (PacificVis). IEEE, pp 311–318

Ma Y, Lin T, Cao Z, Li C, Wang F, Chen W (2016) Mobility viewer: an eulerian approach for studying urban crowd flow. IEEE Trans Intell Transp Syst 17(9):2627–2636

Ma Y, Chen W, Ma X, Xu J, Huang X, Maciejewski R, Tung AKH (2017a) EasySVM: a visual analysis approach for open-box support vector machines. Comput Vis Media 3(2):161–175

Ma Y, Xu J, Wu X, Wang F, Chen W (2017b) A visual analytical approach for transfer learning in classification. Inf Sci 390:54–69

Ma Y, Tung AKH, Wang W, Gao X, Pan Z, Chen W (2018) Scatternet: a deep subjective similarity model for visual analysis of scatterplots. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2018.2875702

Müller M (2007) Dynamic time warping. In: Information retrieval for music and motion, pp 69-84

Pothkow K, Hege H-C (2011) Positional uncertainty of isocontours: condition analysis and probabilistic measures. IEEE Trans Vis Comput Graph 17(10):1393–1406

Potter K, Wilson A, Bremer P-T, Williams D, Doutriaux C, Pascucci V, Johnson CR (2009) Ensemble-vis: a framework for the statistical visualization of ensemble data. In: IEEE international conference on data mining workshops, 2009. ICDMW'09. IEEE, pp 233–240

Potter K, Kniss J, Riesenfeld R, Johnson CR (2010) Visualizing summary statistics and uncertainty. In: Computer graphics forum, vol 29. Wiley, Hoboken, pp 823–832

Riehmann P, Hanfler M, Froehlich B (2005) Interactive sankey diagrams. In: Proceedings of the IEEE symposium on information visualization, p 31

Riveiro M (2007) Evaluation of uncertainty visualization techniques for information fusion. In: 2007 10th international conference on information fusion. IEEE, pp 1–8

Roberts RC, Laramee RS, Smith GA, Brookes P, D'Cruze T (2019) Smart brushing for parallel coordinates. IEEE Trans Vis Comput Graph 25(3):1575–1590

Schmidt GS, Chen S-L, Bryden AN, Livingston MA, Rosenblum LJ, Osborn BR (2004) Multidimensional visual representations for underwater environmental uncertainty. IEEE Comput Graph Appl 24(5):56–65

Seipp K, Gutiérrez F, Ochoa X, Verbert K (2019) Towards a visual guide for communicating uncertainty in visual analytics. J Comput Lang 50:1–18

Seo J, Shneiderman B (2005) A rank-by-feature framework for interactive exploration of multidimensional data. Inf Vis 4(2):96–113

Shi C, Cui W, Liu S, Xu P, Chen W, Qu H (2012) Rankexplorer: visualization of ranking changes in large time series data. IEEE Trans Vis Comput Graph 18(12):2669–2678

Sun M, Lebanon G, Collins-Thompson K (2010) Visualizing differences in web search algorithms using the expected weighted hoeffding distance. In: Proceedings of the 19th international conference on World wide web. ACM, pp 931–940

Wang X-M, Zhang T-Y, Ma Y-X, Xia J, Chen W (2016) A survey of visual analytic pipelines. J Comput Sci Technol 31(4):787–804

Wei J, Shen Z, Sundaresan N, Ma K-L (2012) Visual cluster exploration of web clickstream data. In: 2012 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 3–12

Weng D, Chen R, Deng Z, Wu F, Chen J, Wu Y (2018) SRVis: towards better spatial integration in ranking visualization. IEEE Trans Vis Comput Graph 25(1):459–469. https://doi.org/10.1109/TVCG.2018.2865126
Wilson AT, Potter KC (2009) Toward visual analysis of ensemble data sets. In: Proceedings of the workshop on ultrascale visualization. ACM, pp 48–53
Wittenbrink CM, Pang AT, Lodha SK (1996) Glyphs for visualizing uncertainty in vector fields. IEEE Trans Vis Comput Graph 2(3):266–279
Xia J, Hou Y, Chen YV, Qian ZC, Ebert DS, Chen W (2017) Visualizing rank time series of wikipedia top-viewed pages. IEEE Comput Graph Appl 37(2):42–53
Xu K, Xia M, Mu X, Wang Y, Cao N (2019) Ensemblelens: ensemble-based visual exploration of anomaly detection algorithms with multidimensional data. IEEE Trans Vis Comput Graph 25(1):109–119
Yue X, Shu X, Zhu X, Du X, Yu Z, Papadopoulos D, Liu S (2019) Bitextract: interactive visualization for extracting bitcoin exchange intelligence. IEEE Trans Vis Comput Graph 25(1):162–171
Zappa M, Rotach MW, Arpagaus M, Dorninger M, Hegg C, Montani A, Ranzi R, Ament F, Germann U, Grossi G et al (2008) MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmos Sci Lett 9(2):80–87
Zehner B, Watanabe N, Kolditz O (2010) Visualization of gridded scalar data with uncertainty in geosciences. Comput Geosci 36(10):1268–1275
Zhao Y, Luo F, Chen M, Wang Y, Xia J, Zhou F, Wang Y, Chen Y, Chen W (2019) Evaluating multi-dimensional visualizations for understanding fuzzy clusters. IEEE Trans Vis Comput Graph 25(1):12–21
Zhou H, Yuan X, Qu H, Cui W, Chen B (2008) Visual clustering in parallel coordinates. Comput Graph Forum 27(3):1047–1054
Zhou Z, Ye Z, Liu Y, Liu F, Tao Y, Su W (2017) Visual analytics for spatial clusters of air-quality data. IEEE Comput Graph Appl 37(5):98–105
Zhou F, Lin X, Liu C, Zhao Y, Xu P, Ren L, Xue T, Ren L (2019a) A survey of visualization for smart manufacturing. J Vis 22(2):419–435
Zhou Z, Meng L, Tang C, Zhao Y, Guo Z, Hu M, Chen W (2019b) Visual abstraction of large scale geospatial origin-destination movement data. IEEE Trans Vis Comput Graph 25(1):43–53