Knowledge-Assisted Ranking: A Visual Analytic Application for Sports Event Data

David H.S. Chung, Matthew L. Parry, Iwan W. Griffiths, and Robert S. Laramee ■ *Swansea University*

Rhodri Bown ■ *Welsh Rugby Union*

Philip A. Legg ■ *University of the West of England*

Min Chen ■ *University of Oxford*

Organizing sports video data for performance analysis can be challenging in cases with multiple attributes, and when sorting frequently changes depending on the user's task. The proposed visual analytic system allows interactive data sorting by learning a user's sort requirement dynamically through a knowledge-assisted process.

Event ranking (such as to determine relevance or prioritize actions) is an important task in visual analytics.[1,2] However, this task becomes challenging when sorting involves several data dimensions and the way in which each dimension influences the sorting is not well defined. Such a ranking task is commonplace in practical visual analytics, where we often encounter a request for organizing data into some kind of order without precise specification of the relevant sort keys and a sorting function. Although analytical methods such as multidimensional scaling (MDS) or principle component analysis (PCA) may help in some applications,[3] they focus on the discovery of the most influential attributes in the data, rather than the discovery of a sorting function for an ad hoc sorting task.

In this work, we propose a novel knowledge-assisted approach to such a visual analytics task for sorting sports event data. Our concept is inspired by the card-sorting method,[4] a user-centered design that lets a user categorize a set of items into familiar groups or structures. Card sorting has been previously used to effectively classify symbols in cartography,[5] organize online course sites,[6] and cluster multivariate glyphs.[7] We apply a similar approach to rankings instead. In a knowledge framework,[8] we can summarize the situations as follows:

- Users have tacit knowledge about ranking events but do not have formal knowledge of a sorting function. They may have partial knowledge about sort keys because they typically speculate on a set of influential attributes.
- Although users can rank a given set of events using their tacit knowledge (because they define the ordering), this does not scale up to a large number of events. It is generally easy for users to order a few of the most representative events (such as successful, neutral, or failed). The task becomes inefficient when the number of events increases significantly and ineffective for events with a similar principle criterion (such as the degree of success) but a diverse set of conditions (left or right, earlier or later, different players involved, and so on).
- The system does not have any a priori knowledge about the expected ranking outcome because the ranking requirement is not predefined. It also does not have the formal knowledge about

 Published by the IEEE Computer Society

a sorting function. If the system has a sorting function, it can perform event sorting in a scalable and consistent manner.

We thereby developed a visual analytics system that enables users to provide the system with some of their tacit knowledge by selecting a small set of events (typically three to seven) and ranking them in order as an example for the system. The users may also provide their partial knowledge about possible attributes (for example, data dimensions) that should be considered. This partial knowledge is not essential, but it can significantly reduce the amount of computation. We use regression analysis to convert the input into a sorting function and a measure of influence of different sort keys. The system then provides users with a visualization of the sorted results. The former is shown in a glyph-based sorting canvas, and the latter in a parallel coordinates plot. Users can interactively refine the sorting results by adjusting model parameters or reactivate the knowledge discovery process by refining their initial specification of the example set or the speculated data dimensions.

Satisfactory results can normally be obtained within a few iterations, and users can produce a sorted set of events for supporting further analytical tasks, such as to compile various statistical indicators or to analyze video clips in a sports coaching session. Our results show our visual analytic approach can significantly enhance the usability of multivariate sorting. Here, we demonstrated its usefulness and provide feedback from a range of domain experts using a rugby case study.

Problem Specification

In modern sports, coaches and analysts are now faced with a deluge of data due to the introduction of various digital technologies that support match analysis and training, especially in high-level teams. Our visual analytics system was developed over a two-year period as we closely collaborated with the Wales National Rugby team, who use videos extensively to analyze team and player performance. Such videos often have to meet specific criteria, such as how successful a strategy is in some conditions. The criteria can also change frequently, depending on the user's task—for example, analyzing offensive or defensive plays. The rugby analysts are tasked with the crucial role of finding these key instances. However, current limitations with existing software means this is performed manually, so it can take a considerable amount of time to search the available footage.

Our approach aims to alleviate this problem by enabling analysts to sort events in a flexible manner based on their sorting requirements. To fully consider the challenges involved in sorting rugby events, we first briefly describe the game and provide some background.

Rugby Union

Rugby Union is a popular team sport that consists of two teams of 15 players each who advance an oval ball across a rectangular field (up to 144 m long by 70 m wide) with two H-shaped goal posts at either end. The game is played primarily by carrying the oval ball from one end of the pitch to the other. Points can be scored in several ways: grounding the ball in the opposition goal area (which is called a *try*) or by kicking the ball between the H-shaped post from a conversion, penalty kick, or drop goal. The ball can move from one team to another from tackles and set pieces. Each match is played in two 40-minute halves, where the objective is to score more points than the opponent.

Rugby Event Analysis

The analysis of rugby performance heavily relies on using notational data. This involves tagging video footage with semantic notations from which statistics on individual teams or players can be derived. A rugby game is coded into a series of facets known as a *phase ball event*, which describes the period of play a team has possession of the ball. Each facet then encodes additional data attributes (descriptors) that describe the event in detail:

- *start event*, the type of event in which play is started (such as a scrum, kick reception, or lineout);
- *gain*, the distance gained toward the goal area;
- *territory start position*, the spatial position on the pitch where the team received possession relative to the goals;
- *time*, the event's starting time;
- *tortuosity*, the tortuosity of the ball path; and
- *number of phases*, the number of times a team has the ball between successive breakdown in play.

Although quantitative analyses are helpful for getting an overview of a match, the rugby analysts consider that this alone is not enough to paint a full picture of a game. They therefore examine the key instances by watching the video associated with a phase ball event. Currently, this search process is performed using systems such as SportsCode to browse and select such events based

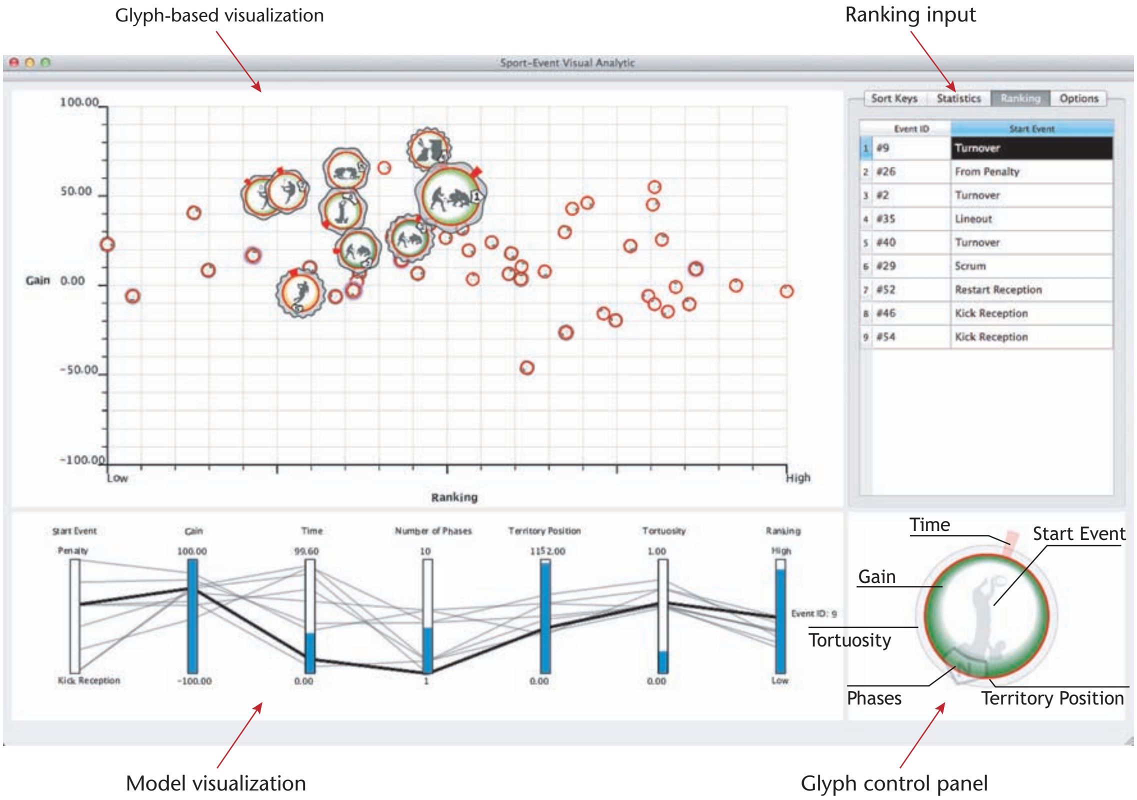


Figure 1. User interface consisting of four main views. The glyph-based visualization shows the sorted events of a match and lets a user select and import events to the system. Once events are imported, the ranking input view is used to specify a sorting requirement. The model visualization view allows the user to analyze how the current model parameters and accuracy correspond to their ranking input. The ranking model can then be exported to one of the primary axes in the glyph-based visualization for viewing the sorted results. The axes can also be modified by clicking on a component in the glyph control panel.

on a predefined attribute. Given the range of requirements for different types of tasks, searching clips by some fixed criteria is time consuming, and often the analysts have to filter through numerous video clips that are not relevant because the sort is not well-defined. Furthermore, this approach does not scale well to multiple matches.

To address these challenges, our knowledge-assisted ranking framework allows users to specify their sorting requirements without depending on specific knowledge about individual sort keys to support this task.

Visual Analytics for Multivariate Sorting

Our visual analytic system closely integrates a knowledge-assisted process to enhance the exploration and sorting of sports event data. By training an analytical model with a user's knowledge on ranking, the system constructs a multivariate sort query that can be used on various matches for retrieving the desired events or associated video clips in flexible manner. (See the Web extra video at https://youtu.be/Cs6SLtPVDQQ for a demonstration of the proposed visual analytic system.)

System Overview

Figure 1 shows the four main system views:

- *Glyph-based visualization* shows the sorted events of a match using glyphs and is the main interface that allows users to select and import events into the ranking input view.
- *Ranking input* allows users to specify and configure their sorting requirement to the system.
- *Model visualization* uses parallel coordinates to convey how the events correspond to the individual attribute contributions and the accuracy of the resulting model, as defined by the ranking input.
- *Glyph control panel* lets users interactively control the primary axes within the glyph-based canvas by clicking on the corresponding glyph attribute.

Each of the system views are linked to support interactive data exploration. For example, brushing glyphs in the glyph-based view will update the corresponding polylines in the model visualization. We use a glyph-based canvas as our main interface

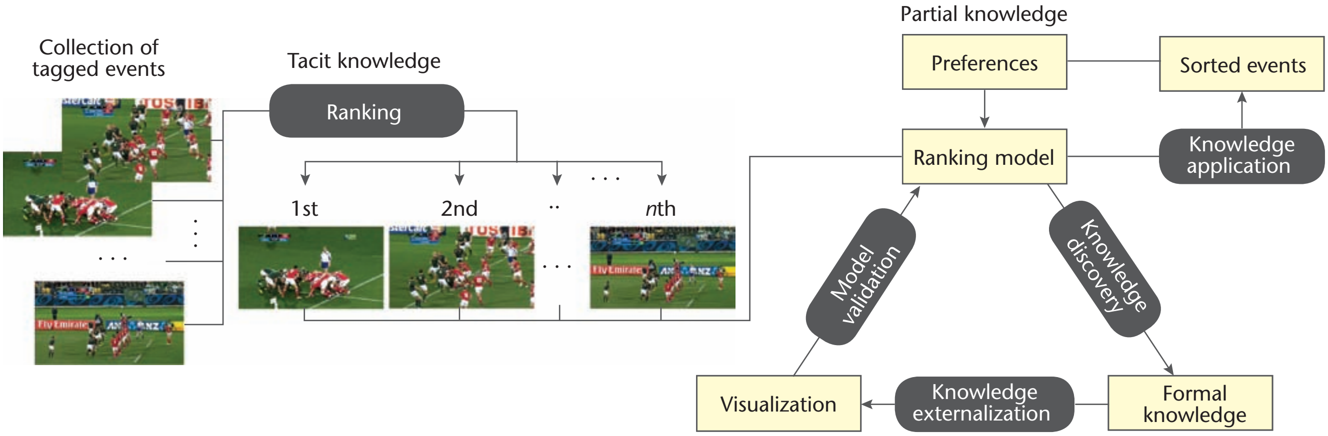


Figure 2. A knowledge-assisted ranking framework. The framework consists of five steps: the user's ranking input as an example to the system; a knowledge discovery process to predict a set of sortable attributes combined into a function; knowledge externalization to convey the analytical model through visualization; model validation based on ranking analysis; and using the model to interactively analyze, rank, and replay match videos (knowledge application).

for selecting and importing events to the system. Metaphorically, the glyphs represent the "cards," as in a card-sorting methodology. In the ranking input view, users can specify an event's rank by dragging the event to the appropriate position in the table. The selected event is then highlighted in black, both in the table and model visualization. The corresponding glyph is also highlighted by magnifying its size to help users visually navigate between each of the different views. We also provide tool tips and a statistics dashboard that displays additional information about an event. To sort events using the learned ranking function, users can export the analytical model to one of the axes in the glyph-based view using a drop-down menu. The user can then playback the video associated with each event data for detailed analyses.

Knowledge-Assisted Ranking Framework

The core framework of our visual analytics system involves converting a user's ranking (sort query) into a function that can be explicitly applied to sort the data. This process involves defining a relationship between the ranking input and the set of sort keys (data dimensions) as illustrated in the first two steps in Figure 2.

Let $e_1, e_2, \ldots, e_n$ be a subset of events, and $e_{i,j}$ be its jth attribute value for m attributes. By placing them into some order $e_{s_1} < e_{s_2} < \ldots < e_{s_n}$, we can model the ranking as $\mathbf{y} = \mathbf{E}\beta$, where $\mathbf{E}$ is an $n \times m$ matrix and $\beta_j \in \mathbb{R}$, $j < m$ are the weights or contribution of each sort key. The goal then is to estimate the weights β such that an event's rank y_i is preserved. Typically, a user may guess these weights during the ranking process. However, this is impractical because the criteria for sorting can frequently change depending on the user's task.

One effective approach for predicting such weights and a potential ranking function is to use regression modeling, which is a common method within statistical forecasting.[1] In this work, we employ three different analytical models: multiple linear regression, polynomial regression, and logistic regression. We then solve the ranking system by approximating the sort key weights using a least-squares fitting[1] that generalizes to

$$\hat{\beta} = \left(\mathbf{E}^{\mathbf{T}}\mathbf{E}\right)^{-1}\mathbf{E}^{\mathbf{T}}\mathbf{y}\ .$$

To ensure that a solution to β exists (that is, that the matrix $\mathbf{E}^{\mathbf{T}}\mathbf{E}$ is invertible), we remove any constant column vector from the model. This problem may occur because our data contains both ordinal and categorical values, for example, if the ranking input contains only a set of scrum events. The least-squares solution is applicable when the system of equations $\mathbf{E}$ is over-determined (that is, for $n > m$). Conversely, $\mathbf{E}$ is under-determined if there may be a lack of suitable training data. Generally, such a system may have infinitely many or no solutions. We can pick one of these solutions such that $\hat{\beta}$ is minimized subject to the constraint $\mathbf{y} = \mathbf{E}\beta$. This is solved using the method of Lagrange multipliers:

$$\hat{\beta} = \mathbf{E}^{\mathbf{T}}\left(\mathbf{E}\mathbf{E}^{\mathbf{T}}\right)^{-1}\mathbf{y}\,.$$

Once the ranking model has been trained, we validate and visualize the model parameters to the user as part of an analytical loop. Rather than simply presenting the model as a black box, this lets users assess whether the sort query is reliable and empowers them to infer some of their own knowledge into the knowledge-discovery process.

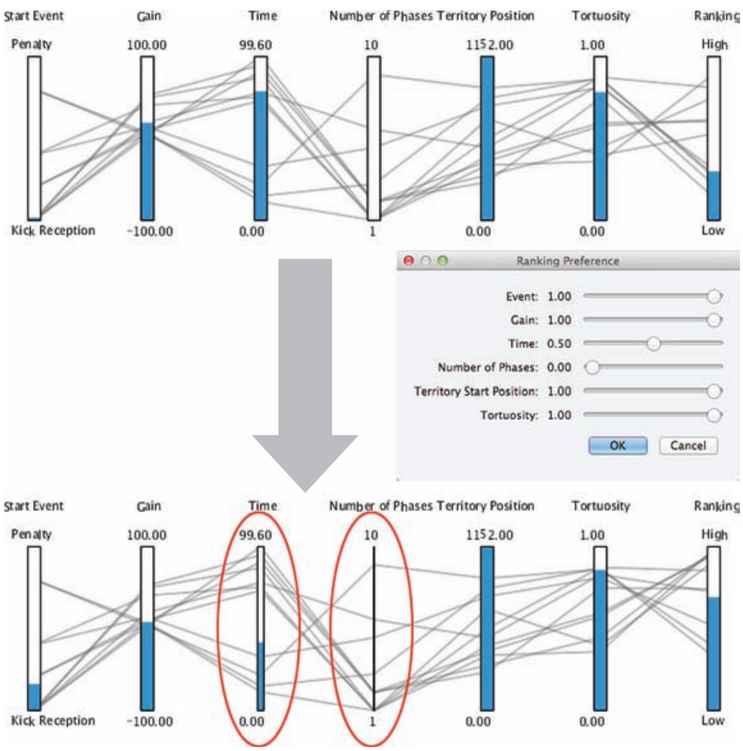


Figure 3. Refining the model parameters. Users can adjust the contribution of each attribute w_j in the ranking model by scaling the axis widths using sliders in the ranking preferences. This example shows how modifying the weights (see the red circles) can result in an improved model, as shown by the larger ranking gauge (see the far right slider).

Regression Evaluation

Given a ranking input, the system needs to compare this against the ranks predicted by the regression model. A common approach used in statistical modeling would be to compute its mean squared error (MSE):[9]

$$\mathrm{MSE} = \frac{1}{n - dof - 1}\sum_{i=1}^{n}\left(\hat{y}_1 - y_i\right)^2 ,$$

where n is the number of events, dof is the degrees of freedom, $\hat{y}$ is the predicted value, and y is the actual value. At this level, the computed values $\hat{y}$ and y represent ranking scores, which are used later to determine the events rank. By choosing a set of scores (such as, $y_i \in [0,1]$), it is easy to observe that the predicted ranks will be preserved when MSE = 0. However, this does not hold for MSE > 0. We address this by incorporating two comparison metrics to help validate different models: a ranking confidence τ and a mean ranking error (MRE).

The ranking confidence τ measures the model's accuracy based on a percentage of events in which its predicted rank matches the order defined by the ranking input. Let $f : E \mapsto \mathbb{R}$ be the trained ranking model and $\phi : \mathbb{R}^2 \mapsto \{0,1\}$ be a binary mapping that returns 1 if the order between two subsequent events $f\left(e_{s_i}\right) < f\left(e_{s_{i+1}}\right)$ for all i = 1, ..., n − 1 is correct. We derive the ranking confidence as

$$\tau = \frac{1}{n-1}\sum_{i=1}^{n-1}\phi\left(f\left(e_i\right), f\left(e_{i+1}\right)\right).$$

When ranking events, such as a set of key moments within a match, we often find that more significant events (the winning goal, for example) can be ranked more easily and accurately than events that are less significant (such as a player making a foul). This concept has been well-established within event-based detection such as video storyboarding.[10] Because such events are more accurate in terms of their ranking, the model's accuracy should therefore take this weighting into account when being compared. We incorporate this by modulating the ranking confidence using a Gaussian function $G(x)$, where $x = (n - 1) - i$. The parameter σ in $G(x)$ is predefined, and we set $\sigma = 2$ as the default.

Our third comparison metric we compute for each model is the MRE, which measures the average difference between an event's actual rank s_i and its predicted rank t_i as given by

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left\|s_i - t_i\right\| .$$

Each of the three comparison metrics allow us to examine how the predicted ranking from different analytical models compares with the actual ranking of the training data. We next describe how we use these metrics to choose the optimal regression for different types of sorting requirement.

Model Selection

The discovery of performance indicators (sort keys) that influence the user's ranking is particularly sensitive to the regression technique used, as Figure 3 shows. The weight of each attribute (for example, the blue gauges) in the model can change and may significantly impact the overall accuracy.

We incorporate each of the three comparison metrics into our system to validate the model using a weighted contribution. For each model, we compute its performance, $P = \lambda_1\mathrm{MSE} + \lambda_2(1 - \tau) + \lambda_3\mathrm{MRE}$, and choose the model with the smallest value. By default, we set each weighting term to be equal ($\lambda_i = 1/3$, for example). However, this can be customized according to the user's preference. The resulting model will give a predicted ranking that is most similar to the sort requirement as defined by the ranking input. We also provide the user with the ability to manually choose between different regressions. This lets the user analyze the different sets of performance indicators that may correlate better to their sort query (even though the predicted ranking may be less similar).

Model Interaction

When a user ranks a set of events based on some ad hoc requirement, they can often make intuitive or educated guesses on specific sort keys that may affect their ranking criteria. We liken this to partial knowledge. Thus, we allow the user to refine the model parameters by applying additional weightings $w_j \in [0,1]$ to the sort key weights β such that the model is defined as $y_i = f(w, \beta, e_i)$. We incorporate this into our system as a series of interactive sliders. Moving the sliders will scale the axis widths in the model visualization (see Figure 4). This enables a user to explore new sorting strategies and understand its impact on the predicted ranking. Optionally, users can choose to remove a sort key parameter from the model completely ($w_j = 0$). Removing a sort key can significantly reduce the computational cost and parameter space as well as potentially resulting to an improved model.

Model Visualization

Figure 1 shows the model visualization. To visualize the analytical model, we adopt the use of parallel coordinates, which is a well-established technique in multivariate analysis.[11] This provides a visual representation of the attribute weights and the overall accuracy of the trained model. Each attribute dimension is plotted as vertical gauges that are then filled according to the amount of contribution within the model. Natalia Andrienko and Gennady Andrienko use a similar approach to visualize and weight multiple criteria in a decision-making application context.[12] They visualize the combined result on a separate axis.

We follow this method to convey the model's accuracy by plotting it as an additional axis gauge and filling the gauge according to its ranking confidence τ. This enables the user to inspect the model's quality and identify which attributes contribute well for a given ranking. The user can also choose to adjust the weights manually using the sliders that adjust the axis widths for that particular attribute. For each event in the match, we render a polyline to help provide context to the model. Allowing the user to brush the polylines in the parallel coordinates or within a linked view (such as a glyph-based canvas) provides a facility to verify that the model is performing as expected by observing the ranking outcome.

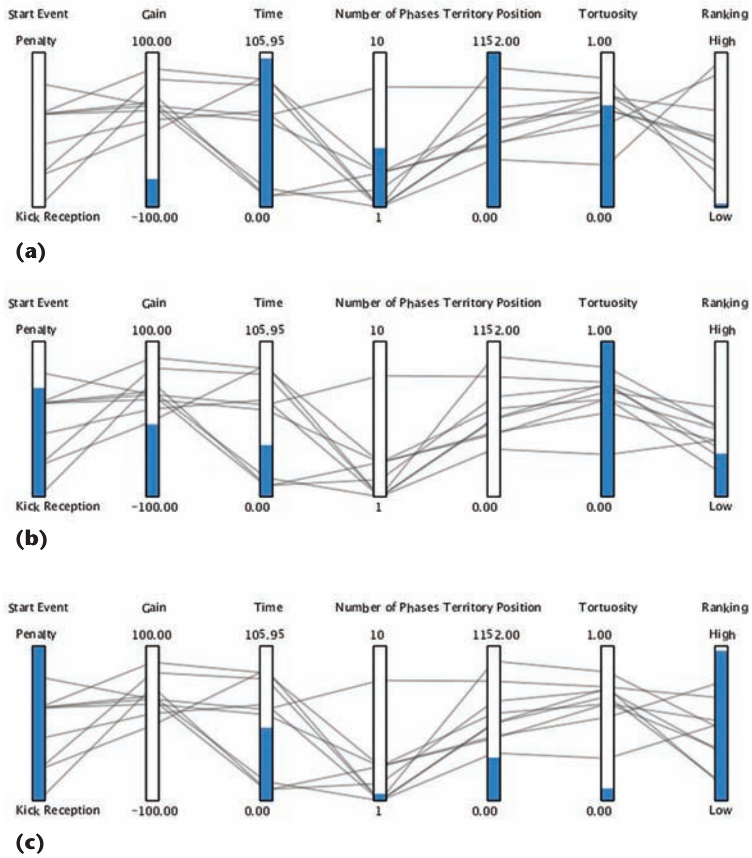


Figure 4. Visual comparison of the ranking models using (a) linear, (b) polynomial, and (c) logistic regression in parallel coordinates. Each attribute's contribution is depicted using gauges that correspond to each axis. To convey the model's overall accuracy, the ranking model is plotted as an additional axis gauge that encodes the ranking confidence *t*. Each regression model may discover a different set of key performance indicators.

Glyph-Based Visualization

Glyph-based visualization is an effective tool for representing multivariate data.[13,14] Our approach takes advantage of the work by Philip A. Legg and his colleagues that demonstrates the use of glyphs for rugby.[15] Each glyph encodes an event data, which we then position along two primary axes. Although interactive multivariate sorting is the focus of this work, we are careful not to confuse the end user with an unfamiliar visual design. We therefore adopt the glyphs used in earlier works[1,15] to encode the event properties as indicated on the glyph control panel in Figure 1. The glyphs that have a purple halo indicate events that resulted to a point scored (see Figure 5, for example). Other visual design choices can be used depending on the application context.[14] The glyph-based canvas is the primary interface for importing and selecting events (the glyphs) into the ranking input view. We found glyphs to be an intuitive mechanism for selecting and ranking events. This is due to similarity with our card-sorting metaphor,

Empirical Study on Formal Rankings

To supplement the motivation of this work, we performed an empirical study using five participants (three computer scientists and two sports scientists) to investigate the difficultly of formalizing a ranking for an ad hoc task in the context of rugby. Each participant had knowledge in both rugby and visualization.

We tasked the participants with identifying and ranking a set of events that highlight the most important positive outcomes of a match. We consider positive outcomes in rugby to be when a team gains an advantage either by scoring or winning a set piece, such as a penalty or free kick. We designed the study such that the participants' recognition of an event's importance is the tacit knowledge we are trying to formalize. During each session, we presented the same match containing 12 events using a basic system with two views, as in Figures 1b and 1c in the main article. This system represents a similar environment, albeit more advanced, to current notational software for selecting events and playing back video clips.

To help us analyze the confidence of a participant's result, for each task outlined in Figure A, the users choose an additional meta-answer: (a) I am reasonably confident about my answer, (b) I am unsure about my answer, or (c) I do not know how to do this.

For tasks 1 and 2, we compared the difficulty of ranking a small set of events (five) to a relatively larger sample (10). Figure A illustrates our results, where the events are ranked from worst to best (1–5 for task 1 and 1–10 for task 2). The majority of the participants were fairly confident with their results in task 1. In contrast, they became unsure of their ranking for task 2. We observed that users were able to establish the rank of important events more easily based on some clear objective feature (for example, the most gain) than for events of less importance. This would suggest and support the use of a moderated ranking confidence t, which we incorporated into our system.

For task 3, we asked the users to identify a set of influential attributes that affected their ranking. Because they defined the sorting outcome, the participants could confidently speculate about a set of performance indicators that determined their ranking. However, it is clear from task 4 that combining each attribute and formally specifying their ranking proved to be a challenge. While a typical participant could perhaps describe such a formalization in an abstract manner, they acknowledged that this would be too difficult to define into an analytical form that can then be used for event organization.

Lastly, we demonstrated our visual analytic system to the users by importing their rankings into the model. We found the discovered sort keys to be consistent with the participant's ranking. All the participants were impressed with the system and believed that such a tool would be useful for sorting event data in a more effective and efficient manner.

Task	Optional meta-answer	Result											
		e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}
1. Identify and rank five events from best to worst.	(a)	3		1					5	4	2		
	(a)		4	5		2		3	1				
	(b)		2					4	3	1		5	
	(a)		4					1	5	2			3
	(b)			1					5	3		4	2
2. Identify and rank 10 events from best to worst.	(a)	4	8	2		1		7	10	9	3	6	5
	(b)	10	7	8		5	4	6	1	2	9	3	
	(b)	2	7		10	1	4	9	3	5		6	8
	(b)	7	8	2	3			4	10	5	1	9	6
	(b)	2		3	6	4	8		10	5	1	9	7
3. Identify a set of attributes that may affect the ranking.	(a)	Gain (high), tortuosity (low), number of phases (low)											
	(b)	(Tortuosity + number of phases), (gain + territory position)											
	(b)	Tortuosity, number of phases, start event											
	(a)	Gain, start event, number of phases											
	(a)	Gain, number of phases											
4. Formulate a ranking based on the set of attributes.	(c)	N/A											
	(c)	N/A											
	(c)	N/A											
	(a)	Combination of high gain, low tortuosity, and a weighted start event (such as turnover is more important than scrum)											
	(b)	Sequences containing high gain or high number of phases from various start events											

Figure A. Empirical study results for sorting rugby events. Each subrow within a task corresponds to five participants along with their optional meta-answer. For tasks 1 and 2, their rankings are shown from the 12 possible events ei and are ranked from worst to best (1–5 for task 1 and 1–10 for task 2). The color map emphasizes the worst (darker) and best (lighter) events.

which is proven to be an effective approach for a sorting task.[7]

Because of the inherent occlusion when using glyphs,[14] we support interaction that enables the user to adjust the length of the sorting axes to help declutter the visualization. During the event-selection process, we also find that nonselected glyphs (such as transparent glyphs) can sometimes interfere with this view because of their large size. To address this problem, users can interactively reduce the size of such glyphs so they appear as small red markers (see Figure 6).

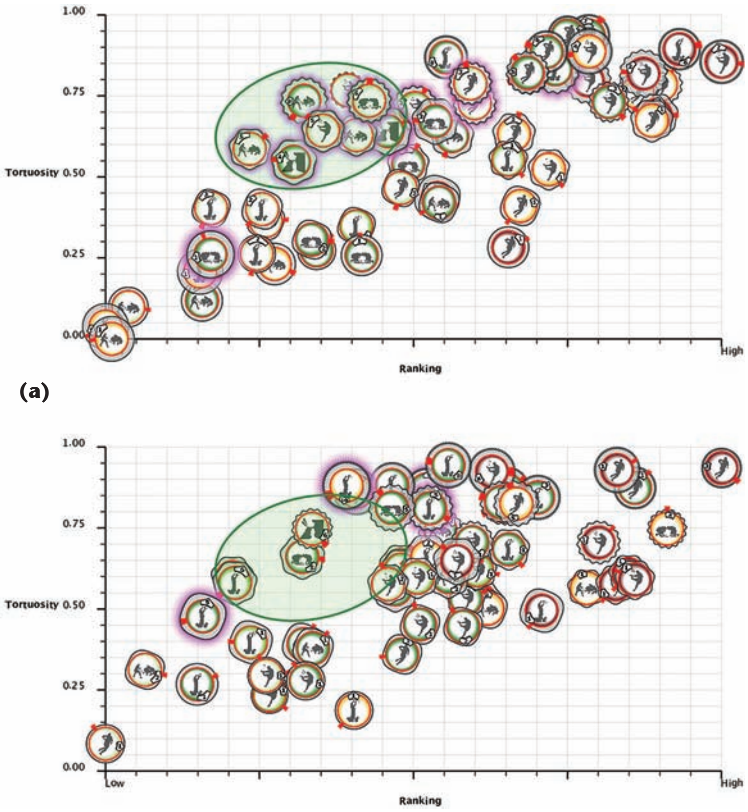
Sorted Event Replay

Sports analysts often rely on making semantic observations that can only be gained by watching videos to determine the relevance of an event. Thus, videos are especially important for specifying a ranking in the system. Because the data is associated with single or multiple video clips, we incorporate a video playback user option for viewing the sorted events (see Figure 7). Brushing events in the glyph-based view or parallel coordinates lets users replay smaller subsets, which enables them to more effectively choose, view, and rank the events than using the results of a typical search query.

Case Study

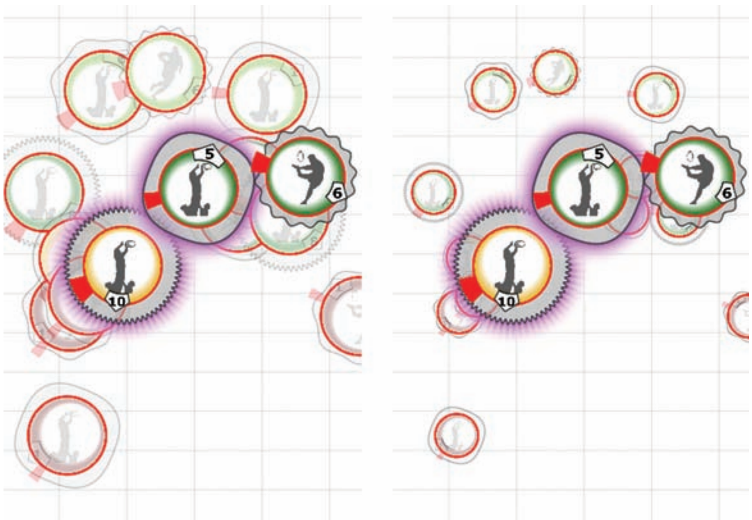
Finding (and often formulating) a successful strategy against opposition teams is a critical task in professional sports. We worked in close collaboration with the Welsh Rugby Union, where coaches and analysts perform such a task primarily by browsing video clips obtained from notational data. The limitations in current software means this is done manually. Such a process is time consuming and does not scale well to multiple matches. Our system presents a novel approach for organizing match videos and event analysis. Here, we present a case study comparing two matches taken from the recent World Cup as part of our evaluation. These two matches present an interesting case to the user because of their huge point differential (81–7 and 16–17, respectively). The analysts we worked with wanted to investigate which strategy led to such a high score and what was so different about the other game.

To begin with, the analysts chose a set of representative events as a training example for the system by selecting the glyphs in the glyph-based canvas. Their initial action was to layout the glyphs according to gain (a typical performance indicator) by changing the primary axes using the glyph control panel to help with this search process. Although the amount of gain is important,



(a) (b)

Figure 5. Visual comparison of two matches. The events are sorted according to successful traits that resulted in points scored as defined by the model shown in Figure 4 and tortuosity. (a) In the first match, the analyst observed a group of events where a high percentage of points are scored (see the green circle). (b) There were significantly fewer points scored in the second match.



(a) (b)

Figure 6. Interaction and occlusion. (a) Brushing glyphs in the glyph-based canvas renders the glyph in focus, while nonselected glyphs are drawn as red markers to indicate their position. (b) Nonselected glyphs can also be interactively scaled by the user to reduce the amount of visual occlusion.

Figure 7. Video playback of sorted events. Four different broadcasting feeds that recorded the event can be viewed.

according to our analysts, a combination of other factors is influential, such as where the players received the ball (territory start position) and how much they worked the opposition (tortuosity).

Potential events were identified quickly based on the glyphs features. After importing these events into the ranking input, the analysts then watched each video to help determine their rank based on how successful the outcome was. Our domain experts were used to performing such a routine task in their normal workflows.

Once the system was trained, the analysts could visually assess the quality of the resulting model in the parallel coordinate view (see Figure 4). During this process, they observed that phases were not a significant attribute for their ranking and refined the attribute weights further to discover an improved model indicated by the amount of blue in the ranking axis, as shown in Figure 4.

Figure 5 illustrates the sorted results according to the ranking function derived by the analysts for both matches. From ranking the events, they were able to discover a cluster of glyphs in one match where a high percentage of points were scored; Figure 5a depicts these with the highlighted purple glyphs. The analysts found it useful to compare and visualize the difference between entire matches in a single overview. The system revealed that the second match had fewer occurrences of events with similar features, which visually suggests why there were not as many scoring opportunities in that game. Two events can be observed within this region, but they did not result in points scored. Investigating this further through video revealed that two poor kicks during play caused the possession to be turned over. Our visual analytic system helped the analysts identify such events quickly and effectively. More importantly, it allowed the analysts to use this to convey to players and coaches what needed to be improved and could lead to new strategies.

Domain Expert Feedback

As part of our case study, we gathered qualitative feedback from three domain expert users: a rugby analyst, the head coach of a university rugby team, and an international rugby player. Each user commented on our software following a hands-on demonstration:

> **Analyst:** Using the software has enabled us to discover new key performance indicators that we wouldn't have recognized before, which ultimately helps save time as we do not need to watch as many irrelevant videos. It's a totally different way of looking at our data. Previously, we would only look at match heuristics such as the territory that we're in, or the gain in isolation, but being able to combine the two attributes (or more) now makes this a lot more meaningful. This is great for comparing matches. The visualization clearly identifies any differences in events, and we can then investigate those clips further and see why they're different.

> **Head coach:** Analysts have reams and reams of stats, which all have to be computed and interpreted manually. The system here is a good way at grouping clips. For instance, if we're defensively bad for a couple of games, you could press a few buttons and it's all there

for you, rather than going through manually, create a database from the first game, then add to it from the second game. Every coach will be looking at different things. For example, I might be looking at "Do we move forward when we catch the ball?" Where this is useful is that it can show the best case and worst case and also be able to look at examples in the middle. The flexibility of the whole model is its strength.

Rugby player: The software is useful as it allows you to break up the game by what you want to see. For instance, it would be irrelevant to show the Heineken cup team (which is the elite competition) all the clips with the squad involved in the LV league competition as they would be with a completely different team. Its main feature is the scalability to sort events from an archive of matches.

This feedback shows the importance of organizing relevant events in sports and that our visual analysis system is a useful approach to support such a task.

Among several possible modeling techniques, we used three different regression analyses to train our ranking model. Because the predicted contributions of each attribute in the model are sensitive to the type of regression (see Figure 3), the role of visual analytics becomes more important because it allows the user to verify whether the discovered performance indicators correspond with their knowledge interpretation. The system may also benefit from a wider range of different models for identifying parameters with a better fit to the ranking input.

For this application, we typically train the model using a relatively small sample size (five to 10 events) to generate a good ranking accuracy. However, this is not the case for all training data. A larger ranking input could result in a more robust model, but without further testing, we do not know how much is enough to improve the accuracy. Ranking more events can also restrict the practicality of the system as they become more difficult to rank, which we found in our initial pilot study. (See the "Empirical Study on Formal Rankings" sidebar for more details.)

In terms of the algorithm for training the model, the scalability of our approach is highly generalizable, and it can be easily applied to other domains and larger datasets. Extending our system to other team sports such as football, basketball, and hockey would simply require adapting the event mapping in the glyph-based design, as in earlier work.[15] A potential issue is with the scalability of our visualization such as the glyph-based canvas. The system currently supports loading a single match, although this could be extended to multiple matches. Because of the use of large glyphs, visualizing several matches at once in this view will create more visual clutter. Likewise, higher dimensionality could also affect the visibility of the model parameters in the parallel coordinates view as a result of over-plotting gauges.

For future work, we would like to evaluate how the system performs over existing software and validate the accuracy of our ranking model across different matches by examining the relevance of the sorted results. In addition, we would like to investigate its scalability using larger data sets.

References

1. D.H.S. Chung et al., "Glyph Sorting: Interactive Visualization for Multi-dimensional Data," *Information Visualization*, vol. 14, no. 1, 2013.
2. S. Gratzl et al., "Lineup: Visual Analysis of Multi-attribute Rankings," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2277–2286.
3. D.H. Jeong et al., "IPCA: An Interactive System for PCA-Based Visual Analytics," *Proc. IEEE VGTC Conf. Visualization*, 2009, pp. 767–774.
4. G. Rugg and P. McGeorge, "The Sorting Techniques: A Tutorial Paper on Card Sorts, Picture Sorts and Item Sorts," *Expert Systems*, vol. 14, no. 2, 1997, pp. 80–93.
5. R.E. Roth et al., "Card Sorting for Cartographic Research and Practice," *Cartography and Geographic Information Science*, vol. 38, no. 2, 2011, pp. 89–99.
6. A. Doubleday, "Use of Card Sorting for Online Course Site Organization within an Integrated Science Curriculum," *J. Usability Studies*, vol. 8, no. 2, 2013, pp. 41–54.
7. I. Borg and T. Staufenbiel, "Performance of Snow Flakes, Suns, and Factorial Suns in the Graphical Representation of Multivariate Data," *Multivariate Behavioral Research*, vol. 27, no. 1, 1992, pp. 43–55.
8. R. Chang et al., "Learning-Based Evaluation of Visual Analytic Systems," *Proc. 3rd Beyond Time and Errors: Novel Evaluation Methods for Information Visualization Workshop* (BELIV), 2010, pp. 29–34.
9. F.S. Acton, *Analysis of Straight-Line Data*, Wiley, 1966.
10. M.L. Parry et al., "Hierarchical Event Selection for Video Storyboards with a Case Study on Snooker Video Visualization," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 12, 2011, pp. 1747–1756.

11. A. Inselberg, "The Plane with Parallel Coordinates," *The Visual Computer*, vol. 1, no. 2, 1985, pp. 69–91.
12. N. Andrienko and G. Andrienko, "Coordinated Views for Informed Spatial Decision Making," *Proc. Int'l Conf. Coordinated and Multiple Views in Exploratory Visualization*, 2003, pp. 44–54.
13. R. Borgo et al., "Glyph-Based Visualization: Foundations, Design Guidelines, Techniques and Applications," *Eurographics State-of-the-Art Reports*, 2013, pp. 39–63.
14. M.O. Ward, "A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization," *Information Visualization*, vol. 1, nos. 3-4, 2002, pp. 194–210.
15. P.A. Legg et al., "Matchpad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis," *Computer Graphics Forum*, vol. 31, no. 3, part 4, 2012, pp. 1255–1264.

David H.S. Chung *is a software developer at We Predict. This research was performed while he was a research assistant in the Sports Video Visualization Research Group at Swansea University. His research interests include data visualization, image processing, and machine learning. Chung has a PhD in computer science from Swansea University. Contact him at D.H.S.Chung@swansea.ac.uk.*

Matthew L. Parry *is a software engineer at SportsViz. This research was performed while he was a research assistant on the Sports Video Visualization Project at Swansea University. His research interests include data visualization and visual analytics. Parry has an MPhil in computer science from Swansea University. Contact him at matthewparry@outlook.com.*

Iwan W. Griffiths *is an associate professor in the School of Engineering at Swansea University. He is also a member of the Innovation and Engagement Group in the A-STEM Research Centre and is technical director of SportsViz. His research interests include biomechanics, mathematical modeling and visualization of biomechanical and performance data, and computer science applied to sports. Griffiths has a PhD in physical chemistry from Swansea University. Contact him at I.W.Griffiths@swansea.ac.uk.*

Robert S. Laramee *is an associate professor in the Department of Computer Science at Swansea University. His research interests include data visualization, computer graphics, and human–computer interaction. Laramee has a PhD from the Institute of Computer Graphics and Algorithms at the Vienna University of Technology. Contact him at R.S.Laramee@swansea.ac.uk.*

Rhodri Bown *is a performance analyst with the Wales National Rugby Team at the Welsh Rugby Union. His research interests include sports performance analysis and sport science. Bown has a BS in sport and exercise science from Cardiff University. Contact him at RBown@wru.co.uk.*

Philip A. Legg *is a senior lecturer in computer science at the University of the West of England. This research was performed while he was a research associate at Swansea University. His research interests include machine learning, visual analytics, and image processing. Legg has a PhD in computer science from Cardiff University. Contact him at phil.legg@uwe.ac.uk.*

Min Chen *is a professor of scientific visualization at Oxford University and a fellow of Pembroke College. He also leads the visualization activities at the Oxford e-Research Centre. His research interests include a broad spectrum of interdisciplinary research topics, ranging from the sciences to sports and from digital humanities to cybersecurity. He is an associate editor in chief of IEEE Transactions on Visualization and Computer Graphics and codirector of the Wales Research Institute of Visual Computing. He is a fellow of British Computer Society, European Computer Graphics Association, and Learned Society of Wales. Contact him at min.chen@oerc.ox.ac.uk.*