Knowledge-Assisted Ranking: A Visual Analytic Application for Sport 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 and Min Chen
University of Oxford

Abstract

Organizing sport video data for performance analysis can be challenging, especially when this involves multiple attributes, and the criteria for sorting frequently changes depending on the user's task. In this work, we propose a visual analytic system to convert a user's knowledge on rankings to support such a process. The system enables users to specify a sort requirement in a flexible manner without depending on specific knowledge about individual sort keys. We use regression techniques to train different analytical models for different types of sorting requirements. We use visualization to facilitate the discovery of knowledge at different stages of the visual analytic process. This includes visualizing the parameters of the ranking model, visualizing the results of a sort query for interactive exploration, and the playback of sorted video clips. We demonstrate the system with a case study in rugby to find key instances for analyzing team and player performance.

Event ranking (e.g., for determining relevance or prioritizing actions) is an important task in visual analytics [6, 8]. 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 one often encounters a request for organizing data into some kind of order without precise specification of the relevant *sort keys* and a *sorting function*. Although some analytical methods such as multidimensional scaling (MDS) or principle component analysis (PCA) may help in some applications (e.g., [10]), 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 sport event data. Our concept is inspired by the method of *card sorting* [14], a user-centered design that allows a user to decide how to categorize a set of items into groups or structures they are familiar with. Card sorting has been previously used to effectively classify symbols in cartography [13], organize online course sites [7], and cluster multivariate glyphs [3]. We apply a similar approach to rankings instead. In a knowledge framework [5], we can summarize the situations as follows:

- Users have *tacit* knowledge about ranking events, but do not have the *formal* knowledge as to a sorting function. They may have *partial* knowledge about sort keys as they typically speculate a set of influential attributes.

- Although users can rank a given set of events

1

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 place a few most representative events (e.g., success, neutral, failure) into order. The task becomes inefficient when the number of events increases significantly, and ineffective (i.e., less 'accurate') for events with a similar principle criterion (e.g., how successful), but a diverse set of conditions (left or right, earlier or later, different players involved, etc.).

- On the other hand, the system does not have any a priori knowledge about the expected ranking outcome, since the ranking requirement is not predefined. Of course, it does not have the formal knowledge about a sorting function either. 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 3-7), and ranking them in order as an example for the system. The users may also provide their *partial knowledge* about possible attributes (e.g., data dimensions) that should be considered. This partial knowledge is not essential, but can reduce the amount of computation significantly. We use regression analysis to convert the input to a sorting function and a measure of influence of different sort keys. The system then provide 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 through model parameters, or re-activate 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 compiling various statistical indicators, and analyzing video clips in a sport coaching session.

Problem Specification

In modern sports, especially in high-level teams, coaches and analysts experience a deluge of data due to the introduction of various digital technologies for supporting match analysis and training. This work was developed from collaborating closely with the Wales National Rugby team over a 2 year period, who use videos extensively for analyzing performance indicators. Such videos often have to meet a specific criteria, for example, how successful a strategy is in some conditions, and can also frequently change depending on the user's task, for example, analyzing offensive or defensive play. 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, and can take a considerable amount of time to search. Our approach aims to alleviate this problem by enabling analysts to sort events in a flexible manner based on their sorting requirement. To fully consider the challenges involved in sorting rugby events, we provide a background to the game.

Rugby Union

Rugby Union is a popular team sport which consists of two teams (of 15 players) who advance an oval ball across a rectangular field (up to 144m long by 70m 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: A *try*, which involves grounding the ball in the opposition goal area, or through 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.

2

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 *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, which include:

- **start event** — the type of event in which play is started (e.g., scrum, kick reception, lineout).

- **gain** — the distance gained towards the goal area.

- **territory start position** — the spatial position on the pitch where the team received possession relative to the goals.

- **time** — the starting time of the event.

- **tortuosity** — the tortuosity of the ball path.

- **number of phases** — a count of the phases.

Although quantitative analyses is helpful for getting an overview of a match, the rugby analysts consider that this alone is not enough to paint a right picture of a game. They therefore examine the key instances through watching the video associated with the phase ball event. Currently, this search process is performed using systems such as SportsCode to browse and select such events based on a pre-defined attribute. Given the range of requirements for different types of tasks, searching clips by some fixed criteria is time consuming, and more than often the analysts have to filter through numerous video clips that are not relevant as the sort is not well defined. Furthermore, this approach does not scale well to multiple matches. We introduce a knowledge-assisted ranking framework that allows a user to specify their sorting requirements without depending on specific knowledge about individual sort keys to support this task.

Visual Analytics for Multivariate Sorting

We developed a visual analytic system that closely integrates a knowledge-assisted process to enhance the exploration and sorting of sport 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.

System Overview

The system (Figure 1) contains four main views:

- Glyph-based Visualization — this view shows the sorted events of a match using glyphs, and is the main interface which users can select and import events into the ranking input view.

- Ranking Input — this view allows the user to specify and configure their sorting requirement to the system.

- Model Visualization — this view uses parallel co-ordinates to convey how the events correspond to the individual attribute contributions and accuracy of the resulting model as defined by the ranking input.

- Glyph Control Panel — this graphical interface allows the user to interactively control the primary axes within the glyph-based canvas by clicking on the corresponding glyph attribute.

Each of the views in the system are linked to support interactive exploration of the data. 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 for selecting and importing events to the system. Metaphorically, the glyphs represent the 'cards' as in our 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
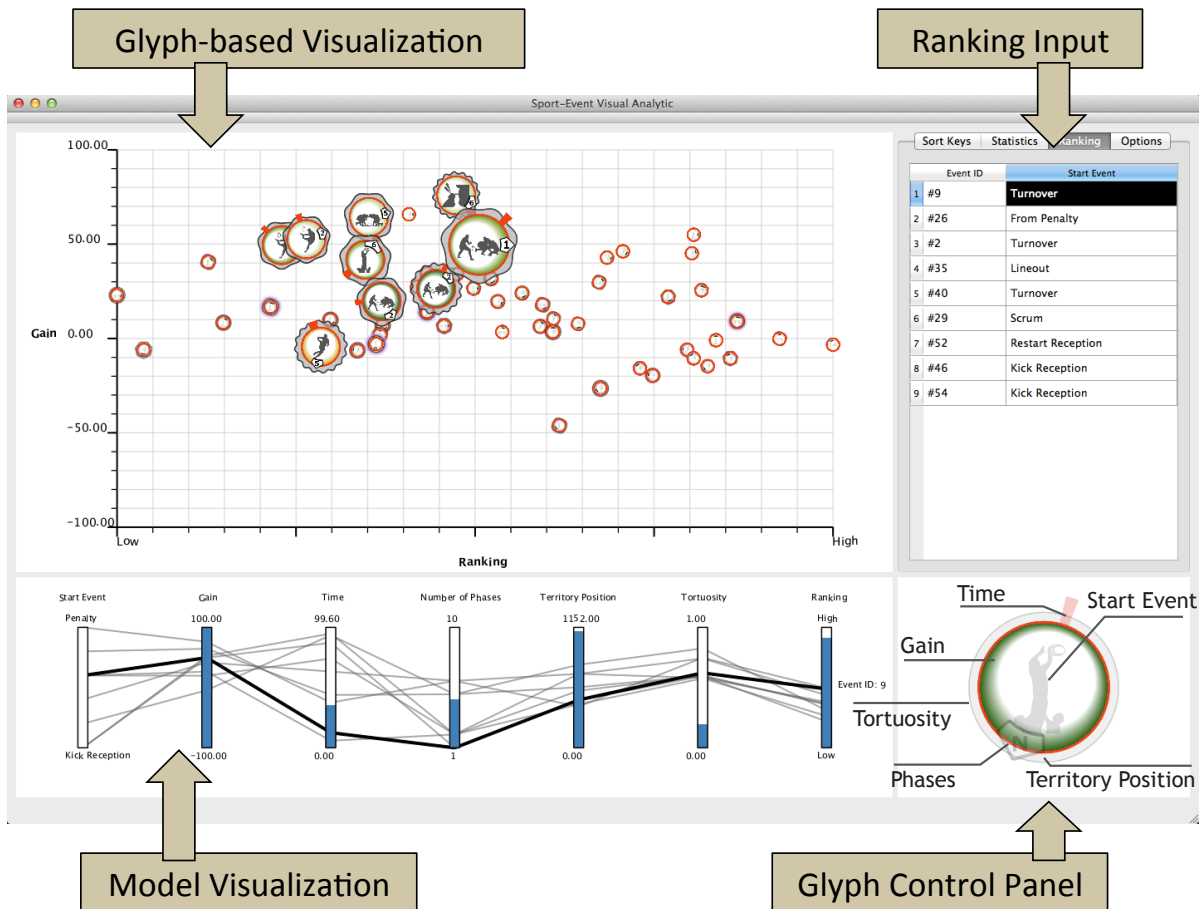
Figure 1: The user interface contains four main views. The glyph-based visualization shows the sorted events of a match, and allows a user to select and import events to the system. Once events are imported, the ranking input view is used to specify a sorting requirement. The model visualization view allows the user to analyse 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.

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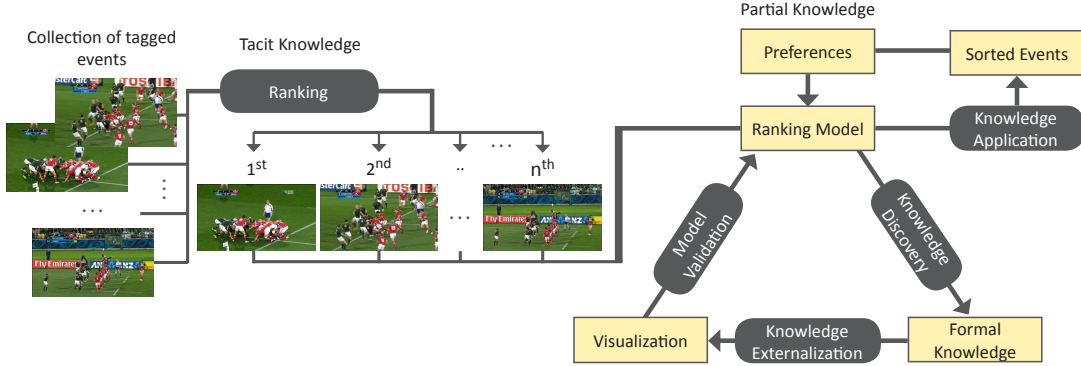
Figure 2: A knowledge-assisted ranking framework. It consists of five steps: the user's **ranking** input as an example to the system, a **knowledge discovery** process to predict a set of sortable attributes combined into a function, **knowledge externalization** to convey the analytical model through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

ship between the ranking input and the set of sort keys (i.e., data dimensions) as illustrated in the first two steps in Figure 2. Let $e_1, e_2, \ldots, e_n$ be a subset events, and $e_{i,j}$ be its j-th 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 since the criteria for sorting may frequently change depending on the user's task.

One effective approach for predicting such weights and a potential ranking function is through regression modelling, 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] which generalizes to:

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

To ensure a solution to β exists, (i.e., the matrix

$\mathbf{E^T E}$ is invertible), we remove any constant column vector from the model. This problem may occur since our data contains both ordinal and categorical values, for example, if the ranking input contains only a set of *scrum* events. The least square solution is applicable when the system of equations $\mathbf{E}$ is *over-determined* (i.e., 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:

$$\hat{\beta} = \mathbf{E^T} (\mathbf{EE^T})^{-1} \mathbf{y} \tag{2}$$

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 allows the user to assess whether the sort query is reliable, and empowers the user to infer some of their own knowledge into the knowledge discovery process.

Regression Evaluation
Given a ranking input, the system needs to compare this against the ranks predicted by the regres-

5

sion model. A common approach used in statistical modelling would be to compute its Mean Squared Error (MSE) [1]:

$$MSE = \frac{1}{n - dof - 1} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad (3)$$

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 is used later to determine the events rank. By choosing a set of scores (e.g., $y_i \in [0,1]$) 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 accuracy of the model 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 $f(e_{s_i}) < f(e_{s_{i+1}})$ for all $i = 1, \dots, n-1$ is correct. We derive the ranking confidence as:

$$\tau = \frac{1}{n-1} \sum_{i=1}^{n-1} \phi(f(e_i), f(e_{i+1})) \qquad (4)$$

When ranking events, for example, a set of key moments within a match, we often find that more significant events (e.g., the winning goal) can be ranked more easily and 'accurately' in comparison to events that are less significant (e.g., a player making a foul). This concept has been well-established within event-based detection such as video storyboarding [12]. Since such events are more accurate in terms of their ranking, the accuracy of model 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 pre-defined, and we set $\sigma = 2$ as default.

(a) Linear Regression

(b) Polynomial Regression
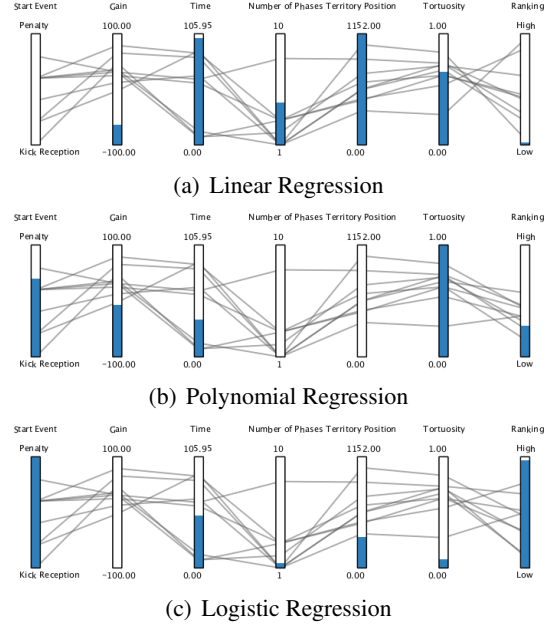
(c) Logistic Regression

Figure 3: Visual comparison of the ranking models using (a) Linear, (b) Polynomial and (c) Logistic regression in parallel co-ordinates. The contribution of each attribute is depicted using gauges that correspond along each axis. In order to convey the model's overall accuracy, the ranking model is plotted as an additional axis gauge which encodes the ranking confidence τ. Note that each regression model may discover a different set of key performance indicators.

Our third comparison metric we compute for each model is the Mean Ranking Error (MRE). The MRE measures the average difference between an event's actual rank s_i and its' predicted rank t_i as given by:

$$MRE = \frac{1}{n} \sum_{i=1}^{n} ||s_i - t_i|| \qquad (5)$$

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

6

Model Selection

The discovery of performance indicators (sort keys) that influences the user's ranking is particularly sensitive to the regression technique used as shown in Figure 3. Notice how the weight of each attribute (e.g., the blue gauges) in the model can change, and may have a significant impact to 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 $P = \lambda_1 MSE + \lambda_2(1 - \tau) + \lambda_3 MRE$, and choose the model with the smallest value. By default, we set each weighting term to be equal (e.g., $\lambda_i = \frac{1}{3}$). 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 ability for a user to manually choose between different regressions. This allows the user to analyse 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 or may not 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 $y_i = f(\mathbf{w}, \beta, e_i)$. 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 to 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.
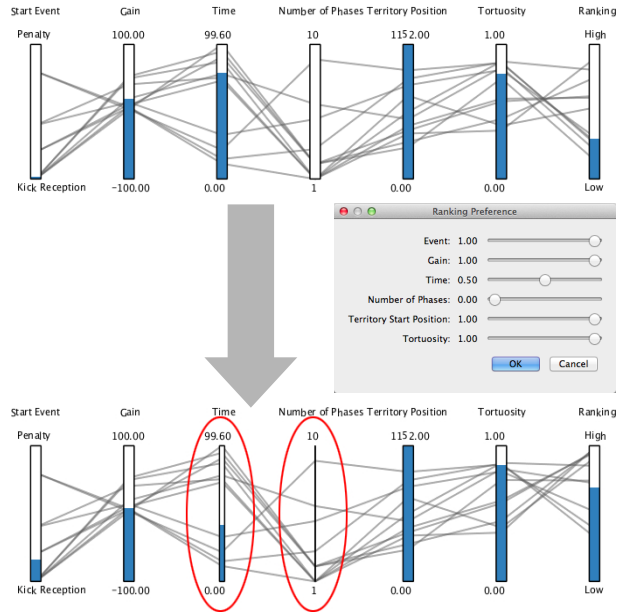

Figure 4: Refining the model parameters. Users can adjust the contribution of each attribute w_j in the ranking model by scaling the axis widths using sliders in the ranking preferences. This example shows how modifying the weights (highlighted in the red circle) can result to an improved model as shown by the larger ranking gauge (see right most axis).

Model Visualization

Figure 1 shows the model visualization. In order to visualize the analytical model, we adopt the use of parallel co-ordinates which is a well-established technique in multivariate analysis [9]. 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 which are then filled according to the amount of contribution within the model. A similar approach is used by Andrienko and Andrienko [2] to visualize and weight multiple criteria in a decision making application context. They visualize the combined result on a separate axis. We also follow this method to convey the accuracy of the model, by plotting this as an additional axis gauge and filling the gauge ac-
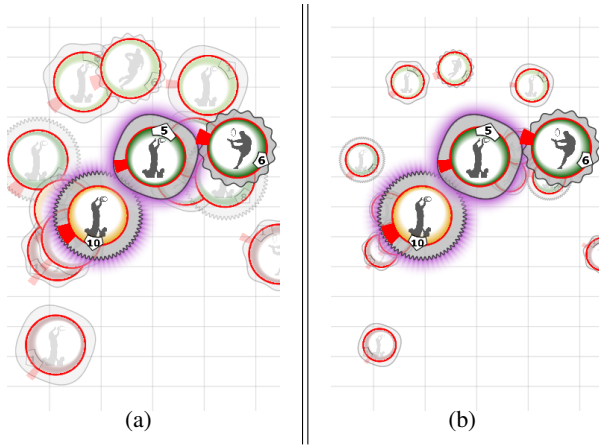
(a)   (b)

Figure 5: Brushing glyphs in the glyph-based canvas
(a) renders the glyph in focus, while non-selected
glyphs are drawn as red markers to indicate their po-
sition. Non-selected glyphs can also be interactively
scaled by the user (b) in order to reduce the amount
of visual occlusion.

Figure 6: Video playback of sorted events. Four dif-
ferent broadcasting feeds that recorded the event can
be viewed simultaneously for detailed analyses.

cording to its ranking confidence τ. This enables the
user to inspect the quality of the model and to iden-
tify which attributes contribute well for a given rank-
ing. The user can also choose to adjust the weights
manually through interaction with the visualization
using sliders that adjusts the axis widths for that par-
ticular attribute. For each event in the match, we ren-
der a polyline to help provide context to the model.
By allowing the user to brush the polylines in the
parallel co-ordinates or within a linked view (e.g.,
glyph-based canvas), it provides a facility to verify
the model is performing as expected by observing
the ranking outcome.

Glyph-based Visualization
Glyph-based visualization is an effective tool for
representing multivariate data [4, 15]. We take ad-
vantage of the recent work by Legg *et al.* [11] who
demonstrate the usability of glyphs in rugby. Each
glyph encodes an event data, which we then position
along two primary axes. Although interactive multi-
variate sorting is the focus of this work, we are care-
ful not to confuse the end-user with an unfamiliar

visual design. We therefore adopt the glyphs used
in [6, 11] 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 7 for example). Other
visual design choices such as the number of design
options presented in [15] may be used depending
on the application context. The glyph-based canvas
is the primary interface for importing and selecting
events (i.e., the glyphs) into the ranking input view.
We found glyphs to be an intuitive mechanism for
selecting and ranking events. This is due to similar-
ity with our card sorting metaphor, which is proven
to be an effective approach for a sorting task [3].

Interaction and Occlusion.   Due to the inherent
occlusion of using glyphs [15], we support inter-
action that enables the user to adjust the length of
the sorting axes to help de-clutter the visualization.
We also find that during the event selection process,
non-selected glyphs (e.g., transparent glyphs) can
sometimes interfere with this view due to their large
size. In order to address this problem, users can in-
teractively reduce the size of such glyphs so they ap-
pear as small red markers (see Figure 5).

Sorted Event Replay
Sporting analysts often rely on making semantic ob-
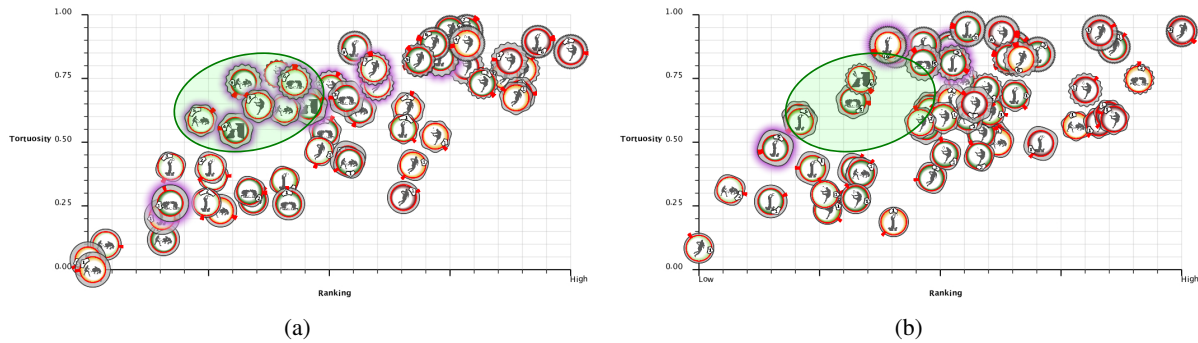servations that can only be gained through the con-

8

Figure 7: 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. The analyst observed a group of events highlighted in the green circle (a) where a high percentage of points are scored, which is significantly less in the second match (b).

text of watching videos in order to determine the relevance of an event. This tool is especially important for specifying a ranking to the system. Since the data is associated with a single or multiple video clips, we incorporate a video playback user-option for viewing the sorted events (see Figure 6). Brushing events in the glyph-based view or parallel coordinates allows for smaller subsets to be replayed, which enable users to choose, view, and rank the events in a much more effective manner than the results of a typical search query.

Case Study: "What is a successful strategy to score?"

Finding (and often formulating) a successful strategy against opposition teams is a critical task in professional sport. We have 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 performed manually. Such a process is time consuming and does not scale well to multiple matches. This system presents a novel approach for organizing match videos and event analysis. After spending time with the analysts using the system, we present a case study comparing two matches taken from the

recent World Cup as part of our evaluation. Both matches present an interesting case to the user due to the huge point differential (81-7 and 16-17 respectively). The analysts would like to investigate what strategy led to such high points scored, and why this was so different to the other game. We detail the process below.

To begin with, the analysts chose a set of representative events as a training example to the system by selecting the glyphs in the glyph-based canvas. Their initial action is 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. While the amount of gain is important according to our analysts, a combination of other factors such as where they received the ball (territory start position) and how much they worked the opposition (tortuosity) is influential. Potential events is identified quickly based on the glyphs features. After importing these events into the ranking input, the analysts then watch each video to help determine their rank based on *how successful* the outcome is. Our domain experts are used to performing such a routine task in their usual workflow. Once the system is trained, the analysts can visually assess the quality of the resulting model in the parallel coordinate view (see Figure 4).

9

During this process, they observed that phases was not a significant attribute to 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 7 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 (see Figure 7(a)) where a high percentage of points is scored depicted by the highlighted purple glyphs. What the analysts found interesting is the ability to compare and visualize the difference between entire matches in a single overview. The system reveals the second match to have fewer occurrences of events with similar features, which visually suggests why there were not as many scoring opportunities in this game. Two events can be observed within this region, however they did not result in points scored. Investigating this further through video reveals two poor kicks during play caused the possession to be turned over. Our visual analytic system helps analysts identify such events quickly and effectively. More importantly, it allows analysts to use this to convey to players and coaches what needs to be improved, and may lead to new strategies.

Domain Expert Feedback. We report qualitative feedback from three domain expert users: a rugby analyst, the head coach of a university rugby team, and an international rugby player. After testing and a hands-on demonstration of our software, we held a consultation with each user.

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 com-bine 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."*

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

Discussion

Among several possibilities of modelling techniques, we used three different regression analyses to train our ranking model. Since the predicted contributions of each attribute in the model is sensitive to the type of regression (see Figure 3), the role of visual analytics becomes more important as it allows the user to verify whether the discovered performance indicators correspond with their knowledge

interpretation. The system may also benefit with 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 (5-10 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 we do not know *'how much'* is enough to improve the accuracy without further testing. Ranking more events can also restrict the practicality of the system as they become more difficult to rank which is reported from our initial pilot study (see Appendix. A).

The scalability of our approach in terms of the algorithm for training the model is highly generalizable, and 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 [11]. A potential issue is with the scalability of our visualization such as the glyph-based canvas. The system currently supports loading a single match, though this could be extended to multiple matches. Due to 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 co-ordinates view as a result of over plotting gauges.

Conclusion and Future Work
In this work, we have proposed a knowledge-assisted visual analytic process for interactive sorting of sport event data. Users provide their knowledge by ranking a set of events as input to the system. We use regression analysis to discover a set of influential sort keys and a function to sort the events according to the user's sort requirement. This allows a user to perform ad hoc sort queries in flexible manner without depending on specific knowledge about individual sort keys. We find our visual analytic approach can significantly enhance the usability of multivariate sorting, and demonstrate its usefulness

Task	(optional) meta-answer	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 5 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 (e.g., turnover is more important than scrum)											
	(b)	Sequences containing high gain or high number of phases from various start events											

Figure 8: Table showing the empirical study results for sorting rugby events. Each sub-row within a task corresponds to five participants along with their optional meta-answer (see Appendix for details). For task 1 and 2, their ranking is shown from the 12 possible events e_i, and are ranked from worst-to-best with 1-5 and 1-10 respectively. A color-map is applied to emphasize the worst and best events.

in rugby along with feedback from a range of domain experts.

For future work, we would like to evaluate how the system performs over existing software, and to 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 to larger data sets. Since the methods in our framework are generalizable, it would be interesting to apply our technique to other sports and application areas.

Appendix A: Empirical Study on Formal Rankings
To supplement the motivation of this work, we perform an empirical study using 5 participants (3 computer scientists and 2 sport 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.

Experiment design. 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 when a team gains an advantage either through scoring, or winning a set piece such as penalties and free kicks. The study is designed such that importance is the tacit knowledge we are trying to formalize. During each session, we presented the same match containing 12 of such events using a basic system with two views as in Figure 1(b) and (c). 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, the users provide an additional meta-answer: (a) I am reasonably confident about my answer, (b) I am unsure about my answer and (c) I do not know how to do this, with each task outlined in Figure 8.

Results. For task 1 and 2, we compare the difficulty of ranking a small set of events (e.g., five), to a relatively larger sample (e.g., ten). Figure 8 illustrates our results, where the events are ranked from worst-to-best with 1-5, and 1-10 respectively. We notice that the majority of participants were fairly confident with their results in task 1. In contrast, they became unsure of their ranking for task 2. We observed during the process that users were able to establish the rank of important events more easily based on some clear objective feature (e.g., the most gain), than events of less importance. This would suggest, and support the use of a moderated ranking confidence τ which we incorporate in our system.

In task 3, we asked the users to identify a set of influential attributes that affected their ranking. Since they define the sorting outcome, the participants could speculate a set of performance indicators confidently which determined their ranking. However, it was clear from task 4 that combining each attribute and formally specifying their ranking proved to be challenging. Whilst 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 which can then be used for event organization. Finally, 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.

References

1. F. S. Acton. *Analysis of Straight-line data*. Wiley, 1966.

2. N. Andrienko and G. Andrienko. Coordinated views for informed spatial decision making. In *International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pages 44–54, July 2003.

3. I. Borg and T. Staufenbiel. Performance of snow flakes, suns, and factorial suns in the graphical representation of multivariate data. *Multivariate Behavioral Research*, 27(1):43–55, 1992.

4. R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics State of the Art Reports*, pages 39–63, 2013.

5. R. Chang, C. Ziemkiewicz, R. Pyzh, J. Kielman, and W. Ribarsky. Learning-based evaluation of visual analytic systems. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaluation methods for Information Visualization*, BELIV '10, pages 29–34, 2010.

6. D. H. S. Chung, P. A. Legg, M. L. Parry, R. Bown, I. W. Griffiths, R. S. Laramee, and

M. Chen. Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization*, 2013.

7. A. Doubleday. Use of card sorting for online course site organization within an integrated science curriculum. *Journal of Usability Studies*, 8(2):41–54, 2013.

8. S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013.

9. A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.

10. D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. ipca: an interactive system for pca-based visual analytics. In *IEEE VGTC conference on Visualization*, pages 767–774, 2009.

11. P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. *Computer Graphics Forum*, 31(3pt4):1255–1264, 2012.

12. M. L. Parry, P. A. Legg, D. H. S. Chung, I. W. Griffiths, and M. Chen. Hierarchical event selection for video storyboards with a case study on snooker video visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1747–1756, 2011.

13. R. E. Roth, B. G. Finch, J. I. Blanford, A. Klippel, A. C. Robinson, and A. M. MacEachren. Card sorting for cartographic research and practice. *Cartography and Geographic Information ScienceCartography and Geographic Information Science*, 38(2):89–99, 2011.

14. G. Rugg and P. McGeorge. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93, 1997.

15. M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4), 2002.