Visualizing Rugby Game Styles Using Self-Organizing Maps

Peter Lamb, Hayden Croft

Abstract—Rugby coaches and analysts often use notational data describing match events to assess their team’s performance and to devise strategic plans for upcoming matches. However, given the volume and complexity of the data available, it is difficult for them to recognize high-dimensional relationships among the available performance variables. A nonlinear approach using self-organizing maps (SOM) can help visualize the performance of a team and its opponents as well as the subsequent suitability of certain game styles, given the style of the opponent.

Index Terms—computer graphics, sports analytics, visual analysis, self-organizing maps.

1 INTRODUCTION

Rugby is a dynamic, evasive, and highly possession-oriented game. During matches, teams try to score points in an opponent’s in-goal area while in possession of the ball in one of two ways: tries or goal kicks. Possession and territory is gained and defended by performing one of several actions, such as passes, carries, and kicks, which can result in numerous events such as points scored, resetting the attack (phase), as well as set plays (known as a set piece), such as scrums, which are described and illustrated in Figure 1. The effectiveness of a series of actions in the match not only depends on their execution, but also on the performance of the opposing team and the state of the match (time left in a match, score, momentum, and so on). Therefore, a match represents a back-and-forth coupling between planned offensive strategies, their execution (known as tactics), and the defensive team’s strategies and tactics.

Coaches and analysts often use notational data describing match events (such as the type of action, the players involved, the location on the field, and the outcome) to assess their team’s performance and to devise strategic plans for upcoming matches. Typically, analysts will look at game summary statistics derived from notational data, usually focusing on a few key performance indicators (KPIs). In practice, KPIs are determined using various statistical approaches, but they are often simply chosen based on the experience and intuition of an analyst or coach.

Since the introduction of professionalism in rugby in 1995, commercial entities (such as OPTA and Prozone) have collected notational data and provided it to teams on a much larger scale than was previously feasible. Thus, the volume and complexity of the notational data available highlights shortcomings of the current status quo. In particular, coaches and analysts are not likely to be capable of recognizing high-dimensional relationships among the available performance variables.

Here we present a nonlinear approach using self-organizing maps (SOM) for visualizing the performance of a team and its opponents as well as the subsequent suitability of certain game styles, given the style of the opponent.

2 SELF-ORGANIZING MAPS

The SOM is an unsupervised neural network useful for clustering high-dimensional data and visualizing those clusters on a low-dimensional output map. These characteristics make it an attractive tool to help convey the nonlinear relationships between many recorded variables to a largely nontechnical end user (such as coaches and performance analysts). SOMs have the following characteristics:

- The output map consists of a lattice of nodes, each of which has an associated prototype vector with values that are achieved through an iterative process.
- The dimensionality of the prototype vectors matches that of the input that is, the number of KPIs used.
- The competitive learning algorithm and the neighborhood function dictate that similar nodes and their
prototype vectors are located in similar map regions, thus preserving the topology of the input data.

- The number of map nodes is far fewer than the number of input vectors, which further promotes the organization of similar nodes into similar map regions. In our case, there were about 3.9 input vectors per output node.

Further details on SOMs and their clustering are available in previous work [2].

The map nodes therefore represent prototypical team performances, and the map regions can collectively represent styles of play. We can look at visualizations of the individual components’ values to interpret map regions and subsequently characterize them qualitatively (see https://github.com/ilarinieminen/SOM-Toolbox). For example, styles based on territorial kicking tend to be clearly separated from styles that depend on frequent passing and maintaining possession.

Input data can be represented on the output map by identifying the single best-matching node in the output for a given input (team performance), which is defined by the node with the shortest Euclidean distance to the input. Labels are assigned to each team’s performance, such as the team’s geographical region, its win-loss record and standing in the league, and point differentials. We then use these labels as a filter to identify all the best-matching nodes for team performances matching various criteria. For each instance, if a particular node is declared the best-matching node, we say that node has been activated, or hit.

To implement and evaluate our approach, we use the dataset from the 2013–2015 seasons of New Zealand’s Independent Timber Merchants Cooperative (ITM) Cup, an annual league tournament in which players play for their home regions. There are 14 teams representing distinct regions in New Zealand. The dataset consists of 193 matches, or 386 individual team performances. Each team performance is represented by 31 KPIs that were selected from a large set of variables based on the following criteria:

- All aspects of a match were to be represented: ball carrying, phase length, kicking tactics, possession location, infringements (breaking a rule or law), tackling, and set pieces.
- There should be a balance between the number of KPIs selected to represent each match aspect.
- Referee related variables were excluded.
- Redundant variables were excluded.

Redundant variables are deemed as those covering the same aspects of a match. For example, because tackle attempts and the percentage of tackles made are included, we considered successful tackle counts redundant. This was also true for tackles in different field zones and for a number of passing variables.

3 Map Region Characterization

The SOM provides a clustering of the nodes, and we use the $k$-means clustering algorithm to determine the cluster partitioning (see Figure 2). Using the ITM Cup dataset, we found that five clusters balance the trade-off of having meaningful cluster divisions while maintaining sufficient node hit frequencies to analyze specific team performances.

The clusters in Figure 2 can be interpreted as unique game styles based on the numerous combinations of KPI values recorded. The two clusters at the bottom of the map (clusters 2 and 4 in Figure 2) represent high-possession game styles. Matches that were represented by nodes in cluster 4 are characterized as games in which possession was maintained in the attacking half with a high amount of ball carrying, frequent passing and off-loading, and long phases. Cluster 4 is also associated with a high frequency of errors and a medium-to-high frequency of infringements. Similar to cluster 4, cluster 2 is characterized by a high amount of ball carrying, but very little passing that is, territory is gained by off-loads, hit-ups, and low kicks. Match performances in cluster 2 have a low rate of infringements and are able to maintain possession deep in the attacking zone, known as the attacking 22, for a high proportion of the match; this is considered a tough, aggressive style similar to Rugby Union’s cousin, Rugby League, which is very popular in Australia. The central cluster, cluster 3, is a low possession, defensive style match with high quality tackling. Cluster 1
represents a defensive game style with a low frequency of ball carries, high proportion of time spent in the defending half, a high frequency of kicks in play, and a high number of total phases. Cluster 5 is characterized by a high ruck speed and a high phase total in other words, a game style based on a high frequency of quick forward attempts in an effort to give the defense little time to reorganize after each attempt and the ensuing ruck. (A ruck is formed when the ball is on the ground with players on both teams surrounding it; they try to move the ball with their feet or move the opposing players back so that the ball emerges for teammates to pick up.)

Characterizing and visualizing rugby play in such a way is unique because it preserves and focuses on the (potentially) nonlinear relationships between the many KPIs used to train the network. The importance of certain KPIs in determining success depends on the style being played. For example, ruck speed seems to only be important in the context of a low-passing, low-possession game style. These unique insights are useful in characterizing an upcoming opponent’s style, but also, more generally, in determining the effectiveness of counterstrategies. Because each match performance is represented by two input vectors, one for the performance of each team, the best-matching nodes for each team let us expose the coupling of the game styles.

4 Game Style Analysis

We can isolate a counterstrategy’s effectiveness by selecting all performances represented by a certain game style and showing the hits of the opponent’s performance in response to that game style.

In Figure 3, we selected all the teams with a performance that was represented by the nodes in cluster 4 and show the best-matching nodes of their opponents’ performances. This lets us see the coupling between the tactics of one team and those of its opponent. In Figure 3, the red nodes show opponent losses (or equivalently wins by the selected team) and the blue nodes show opponent wins (or selected team losses). The opponents of teams who played the cluster 4 game style which was a successful game style seemed to have a greater chance of winning if they played the same or a similar style, although this was only accomplished about half of the time. The more dissimilar the opponents’ game styles were to cluster 4, as shown by the red nodes scattered across other regions of the map, the more ineffective they were against the selected teams. Teams who tend to play a cluster 4 game style may find it informative to know whether their opponents are showing signs of successfully dealing with their tactics.

The most successful game style was cluster 1 (see Figure 3b): this style was rarely met with the same style by the opponent. Although it is difficult to determine which team dictates the styles of play, it is clear that cluster 4 (the high possession- and passing frequency style) provides a good match-up against the defensive, kicking style of cluster 1. Furthermore, teams preparing for matches against opponents who tend to play cluster 1 games are advised to make their best effort to avoid executing game styles characterized by clusters 2 and 5.

Figure 4 shows the predominant game type for a specific team (team A) and the performance of its opponents. Roughly half of team A’s matches were best represented by an aggressive style of attacking from all field zones with high frequencies of off-loads, hit-ups, and low kicks. In general, this is not a successful style of play (with an overall 40–64 win-loss record), but team A was one of the best at this style.

Figure 4 shows the performance of team A’s opponents for the cluster 2 matches it played. The red nodes signify the performance style of the opponents that team A beat, and the blue nodes show the opponents who beat team A. Team A was more successful when its opponent tried to play a similar game style. Or from the opponent’s perspective, cluster 1 seemed to represent a good counterstrategy to team A when it executed aggressive tactics.

5 Analyst-Coach Interface

Using the following case study, we can outline in practice how SOM information can be fed back from the analyst to the coaches and players. After characterizing the game styles identified by the SOM, the analyst can reduce the number of variables to a short list of KPIs. Domain-specific expertise would guide the variable selection. The analyst will also consider whether high or low values of a variable are unique to a specific game style, or if they distinguish two potentially competing game styles, as well as the practicality of implementation.

Preview meetings can be held before each match to plan strategies. Based on thresholds, usually of range quartiles, the players are then given qualitative information related to the short-listed KPIs, with statements such as “try to kick more” or “keep up the kicking”, instead of trying to achieve a specific number of kicks.

Fig. 2. The self-organizing map (SOM) output. The hexagonal cells represent nodes. The k-means algorithm was used for cluster partitioning, which resulted in five clusters. The colors indicate node cluster membership, and the numbers are indices to respective clusters.
The KPIs can then be tracked live during the match to estimate whether game styles for each team are developing as expected. If the analyst decides a strategic adjustment is required, that information can be passed on to the coach, who communicates it to the players at various breaks in the match.

Lastly, in a review meeting after each match, the analyst and coaches can provide feedback to the players. Feedback from the analyst is usually related to the previously identified KPIs, whereas coach feedback tends to be more specific, for example, with video footage of key plays.

Future directions in rugby performance analysis using game statistics may involve focusing on game styles in certain zones of the pitch and whether or not the team is in possession of the ball. For example, when in the defending 22 without the ball, a team’s tactics differ significantly compared with being in the attacking 22 in possession of the ball. These details are lost in the current system when summing up a team’s performance across the match. The difficulty in extracting these details, if SOMs are used, will be in introducing new maps for each region and each ball possession combination, which all require different interpretations.

Player tracking systems may present promising opportunities for characterizing game styles. Describing each player using \((x, y)\) locations on the pitch would allow us to record offensive and defensive configurations. Actions could also be attributed to players and their associated locations. To date, the player interactions on a pitch must be theorized, where they behave as dynamical systems influenced by constraints such as the pitch dimensions, laws of the sport, and the opponent’s strengths and weaknesses. To our

Fig. 3. Game style analysis: (a) opponent performance for cluster 4 (36–22 win-loss record) and (b) opponent performance for cluster 1 (45–27 win-loss record). The shaded areas indicate the selected teams’ game style, the red nodes indicate losses by their opponents (or wins by the selected teams), and blue nodes indicate wins by their opponents (or selected team losses). The hexagon sizes increase with increasing hit frequency.

Fig. 4. Game style analysis showing opponent performance for team A (cluster 2, 7–6 win-loss record). The shaded areas indicate team A’s game style, the red nodes indicate the performance style of the opponents that team A beat, and blue nodes show the opponents who beat team A. Team A was more successful when its opponent tried to play a similar game style.
knowledge, efforts at studying these interactions have not yet led to findings of interest to coaches and practitioners. Until more can be made of player interactions, derived from player trajectory data, the current analysis may be re ned by further consideration of the variables chosen for input.

ACKNOWLEDGMENTS

We thank Opta for supplying the data used in this article.

REFERENCES


Peter Lamb is a lecturer in the School of Physical Education, Sport, and Exercise Sciences at the University of Otago. His research interests include visualizing and interpreting multisegment and muscle movement patterns in sporting and health contexts and the development of software tools to aid practitioners, coaches, and clinicians in interpreting biomechanical and sport performance data. Lamb has a PhD in biomechanics from the University of Otago. Contact him at peter.lamb@otago.ac.nz.

Hayden Croft is a senior lecturer at Otago Polytechnic and a performance analyst with Otago Rugby and Southern Steel Netball. His research interests include sports analysis, biomechanics and video technology. Croft has a master’s of physical education (biomechanics) from the University of Otago and is currently working toward his PhD at AUT in New Zealand. Contact him at haydenc@op.ac.nz.