Identifying and Analyzing SHL Ice Hockey Match Styles Based on Event Data Aggregation and K-Means Clustering

Yanjie Lyu, Qingxuan Cui, Huaide Liu, Han Xia, Yi Yang

Linköping University, Linköping,Sweden

1 Introduction

As a high-speed and dynamic team invasion sport, ice hockey's game strategy and tactical choices have a crucial impact on the result of the game[1]. While traditional post-match statistics such as goals, assists, +/- values, etc., provide information about the result, they often fail to reveal the overall strategic tendencies adopted by the team during the game, the so-called "game style"[2]. Understanding the game styles of different teams not only helps to assess team characteristics and prepare for games, but also provides a reference for player recruitment and roster construction [3].

With the development of data collection technology, detailed event data makes it possible to quantitatively analyze the style of the game[4]. At present, relevant institutions or organizations have begun to explore the use of these data for analysis. For example, clustering based on the positioning of players during the game to distinguish player roles and styles[5], research that focuses on specific behaviors such as passing networks or space utilization patterns[6], or dedicated to evaluating the value of players or actions through machine learning models[7]. In recent years, a promising direction has been to draw on the experience of other team projects (such as field hockey), and utilize clustering algorithms to process statistical data so that macroscopic game styles can be identified[2].

This paper aims to apply a similar methodology to SHL hockey game data with the following contributions:

1. Identify distinct styles of play present in the SHL by applying a K-Means clustering algorithm to aggregated event statistics from individual team games.
2. Characterize the identified styles using statistical metrics and visualization techniques, such as radar charts and heatmaps.
3. Conduct a preliminary analysis of the winning percentage of matches between the different styles of play.

While the core task of the LINHAC challenge is to identify event sequence patterns that lead to specific outcomes, we argue that first identifying and understanding macro-level styles of play at the team level provides a crucial foundation and contextual backdrop for subsequent, more granular sequence analyses. The styles identified in this study offer essential context for further investigations into which success or failure event sequences are more likely to emerge under particular styles of play.

2 Background

The data basis for this paper is the Swedish Ice Hockey League (SHL) detailed game event data provided and licensed by Sportlogiq Inc. This kind of refined event data records the details (such as time, player, team, coordinates, result, etc.) of every pass, dribbled, shot, zone transition, etc., during the game. However, it is challenging to extract meaningful patterns from high-dimensional, time-series event data, especially macro team tactical styles[8].

Subsequently, K-Means clustering, a widely used unsupervised algorithm, is employed to partition matches into distinct tactical style clusters [9]. To clearly interpret and present these tactical styles, the study utilizes visualization methods: radar charts illustrate the multi-dimensional performance profiles of different styles, and heatmaps visualize the winning rates between competing tactical styles, revealing potential interactions and constraints among them.

3 Methodology

This study adopts a multi-stage analysis process, aiming to identify, quantify and interpret the game styles at the team level from the SHL match event data, and ultimately evaluate the relative effectiveness of different styles. This process integrates methods such as data processing, feature engineering, unsupervised clustering, supervised learning interpretation, and adversarial analysis.

3.1 Data Preparation and Feature Extraction

The research data is derived from the SHL race event log provided by Sportlogiq. The original data was first preprocessed and sorted by ‘gameid‘ and ‘compiledgametime‘. This was followed by meticulous feature engineering designed to translate discrete match events into continuous or count-type variables capable of capturing a team’s tactical intent and execution efficiency. The key steps include: (a) Created numerical or categorical identifiers for core game events such as passing, shooting, ball possession entering, blocking, clearing, etc.; (b) Combined the event outcome field to quantify the successfully executed actions such as the number of successful passes and the number of successful area entries; (c) For the carry event, by calculating the spatio-temporal differences between adjacent events, the carry_duration and carry_distance were extracted to reflect the mode of advancing with the ball. The goal of this part is to build a rich feature set to lay the foundation for the subsequent aggregation of team performances.

3.2 Team Performance Aggregation and Metric Construction

To achieve the transformation of the analytical scale from micro events to macro team performance, aggregating the variables extracted by feature engineering according to team (teamid) and game (gameid) is a proper way. In order to be able to calculate the key efficiency indicators and capture the performance in other dimensions, the aggregation strategy is determined based on the nature of the variables: the number of event occurrences (such as num_passes, num_successful_passes, which are the basis for calculating the efficiency indicators) can be obtained by summation the corresponding indicator variables, The expected goals (xg_allattempts), dribbling time and distance, etc. can be calculated as the average value of a single game. Subsequently, using these aggregated results, efficiency indicators such as pass_success_rate were constructed by calculating the corresponding ratios.

3.3 Game Style Identification using K-Means Clustering

To explore and discover the potential, data-driven game style in the SHL competition, K-Means clustering algorithm was adopted. Given that K-Means is sensitive to the scale of input features, 13 aggregated features covering aspects such as ball control, offense, defense and efficiency were selected , and they were standardized to ensure that each feature has zero mean and unit variance. Then, the KMeans algorithm was applied to assign each team-match sample point to one of the preset three clusters. The choice of K=3 was based on the initial consideration of the interpret ability of the results and the purpose of identifying the main style types.

3.4 Evaluation of Inter-Style Effectiveness

Finally, in order to preliminarily evaluate the effect of the identified game styles in actual confrontations, the competition results under different style combinations were analyzed. By matching the two opposing teams in the same match and their style tags and combining with the game results, the average winning rate (scoring rate, win=1, draw=0.5, loss=0) of each style when facing the specific opponent's style was calculated. This pairwise comparison helps to reveal the underlying restraint relationship among styles.

4 Results and Discussion

Through K-Means clustering method in this study, three game styles with significantly different characteristics were successfully identified. With the evaluation of the winning rate of different styles, it is possible to depict the portraits of these styles, identify the key driving factors and explore the practical effects.

4.1 SHL Game Style Profiles and Key Features

The three main game styles discovered by K-Means clustering have been initially named "Defensive Counterattack", "High-Pressure Offense" and "Puck Control Play". Figure 1 visually presents the average performance differences of these three styles on 13 standardized aggregated features.

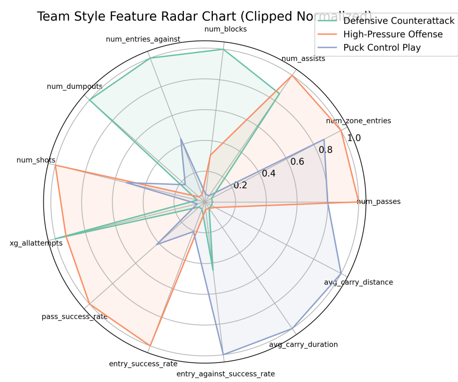

Fig. 1. Team Style Feature Radar Chart (Clipped Normalized)

By synthesizing the information from Figure 1, the profiles of the three playing styles are clearly depicted:

(a) **Defensive Counterattack** This style excels in defensive metrics, while scoring low on most offensive and possession metrics. This highlights its strategy of prioritizing solid defense and relying on quick transitions.

(b) **High-Pressure Offensive** This style leads in offensive outputs, with a notably high number of passes, reflecting its aggressive approach of high-frequency pressing and creating numerous shooting opportunities.

(c) **Puck Control Play** This style is distinguished by exceptionally high average carry distance and average carry duration, as well as the highest entry success rate, with number of passes being the most significant differentiating factor for this style (Figure 2). This indicates its core strategy of controlling the tempo of the game through long periods of possession and high success rate in advancing the ball. However, its relatively low number of shots and expected goals from all attempts (Figure 1) suggest potential limitations in converting possession advantage into concrete offensive threats.

4.2 Style Effectiveness and Matchup Dynamics

The performance of different playing styles was evaluated in this study, and the results show significant differences in their win-loss outcomes.

Figure 2 clearly shows a potential counteracting relationship:(a)The **Defensive Counterattack** style not only effectively suppresses the **Puck Control**

Play style(win rate 82%), but also slightly outperforms the **High-Pressure Offensive** style(win rate 55%); (b)The **High-Pressure Offensive** style, while strongly dominating the **Puck Control Play** style(win rate 86%),is at a disadvantage when facing the **Defensive Counterattack** style (win rate 45%); (c)The **Puck Control Play** style struggles against both other styles, with win rates of 18% and 14%, respectively.

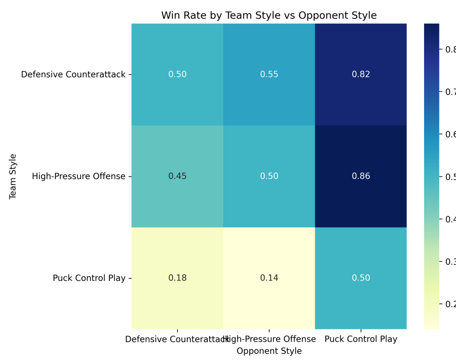

Fig. 2. Win Rate by Team Style vs Opponent Style

The Puck Control Play style has a significantly low win rate and a high loss rate, which contrasts with its theoretical advantage in controlling the game tempo. To further explore the defensive issues it may face in practice, the spatial distribution of average goals conceded by teams of each style during matches was analyzed. Through a heatmap visualization, the differences in goal-conceding areas across playing styles are revealed, providing a spatial perspective on the potential defensive weaknesses of the Puck Control Play style.

Figures 3 to 5 illustrate the spatial distribution of average ball loss locations for teams employing the three playing styles.

In Figure 3, the Average Possession Loss for Puck Control Play style teams reveals that possession posses are primarily concentrated in the defensive zone and near the neutral zone blue line. This suggests instability in puck control during defensive-to-offensive transitions or when organizing plays through the neutral zone. Such spatial patterns may expose Puck Control Play style teams to higher risks of turnovers under aggressive forechecking, thereby creating counterattack opportunities for opponents and negatively impacting overall game outcomes.

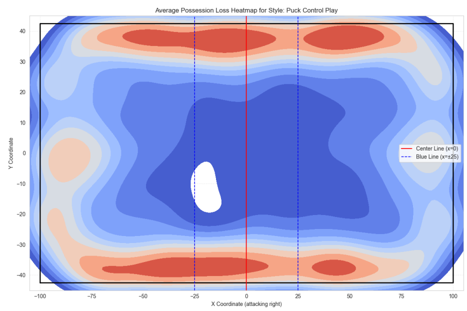

Fig. 3. Average Possession Loss Heatmap for Style: Puck Control Play

Figure 4 illustrates the Average Possession Loss for teams employing a Defensive Counterattack strategy. Compared to other styles, turnovers are more concentrated and occur farther from their own goal, primarily on both sides of the center red line. This pattern reflects the tactical focus on solid defensive positioning and swift counterattacks. Most puck losses take place during contested plays in the neutral zone and do not directly threaten the defensive zone, which may partly explain the higher win rate associated with this playing style.

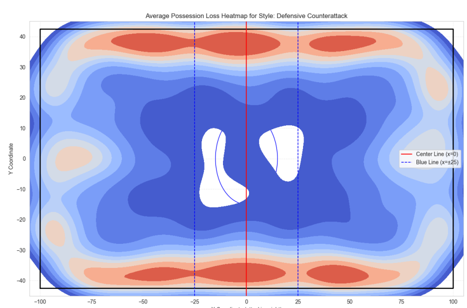

Fig. 4. Average Possession Loss Heatmap for Style: Defensive Counterattack

Figure 5 presents the Average Possession Loss for teams employing a high-pressure offensive strategy. Turnovers are concentrated in the offensive zone and around the offensive blue line. This pattern suggests that during aggressive forechecking and rapid transitions, possession may be lost due to rushed plays or passing errors, resulting in puck losses high up the ice. While this approach entails a higher turnover risk, the fact that these losses occur far from the team's own net reduces the immediate threat of conceding goals. This reflects a tactical trade-off between offensive intensity and tolerance for risk in puck management.

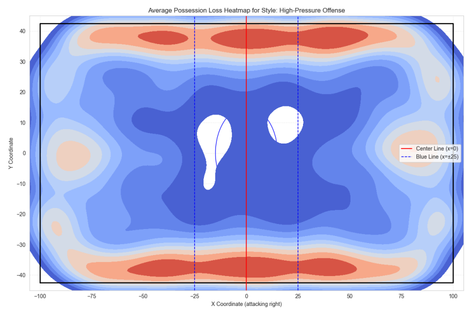

Fig. 5. Average Possession Loss Heatmap for Style: High-Pressure Offense

4.3 Implications and Limitations

This study successfully applied a combined approach of K-Means clustering to identify and quantify three distinct playing styles from aggregated event data in the SHL. The results provide a data-driven perspective on the tactical diversity of SHL teams and offer practical implications for coaching staff in pre-game preparation, tactical planning, and opponent scouting. The identified key features underscore the foundational role of these basic actions in shaping broader tactical styles.

However, several limitations of this study should be acknowledged. First, the aggregated statistics used in this study omit the temporal sequencing and complex contextual dependencies of events, which differs from the core requirement of the LINHAC challenge.The macro-level styles identified in this study should be seen as a foundation or a hierarchical variable for more detailed sequence analysis, rather than the final goal. Secondly, the choice of K=3 clusters is based on preliminary exploration, lacking more rigorous quantitative metrics (such as silhouette coefficients) or validation through comparisons with multiple K values. Finally, the analysis is based on a specific dataset, and the generalizability of the conclusions remains to be tested on broader datasets (such as multiple seasons or different leagues).

Despite these limitations, this study demonstrates the feasibility of applying data mining techniques to ice hockey event data for identifying and analyzing playing styles, providing a foundation for further in-depth exploration of tactical patterns, event sequences, and their relationship with match outcomes.

5 Summary

This paper applied the K-Means clustering algorithm to identify three distinct playing styles from aggregated event data in SHL matches: “Puck Control Play”, “Defensive Counterattack”, and“High-Pressure Offense”.Each style was clearly characterized through radar charts , highlighting their unique attributes in areas such as passing, shooting, dribbling, defense, and effectiveness. Preliminary analysis of style-versus-style win rates suggests that the "Defensive Counterattack" style achieved a higher win rate within this dataset. This paper provides a foundational framework for the quantitative understanding of playing styles in ice hockey.

6 Future Work

This paper serves as an initial exploration and could be extended in the following directions:

- **Methodological Refinement**: Determine the optimal number of clusters K using techniques such as the elbow method; adopt more precise definitions of match outcomes (e.g., final result); and account for stylistic dynamics across different phases of the game (e.g., by periods or overtime).
- **Incorporating Spatial Dimensions**:Integrate spatial information more thoroughly, such as computing metrics for specific pitch zones or applying spatial clustering techniques, to provide a richer description of playing styles[10].
- **Focus on Sequential Patterns**:To better align with the LINHAC competition task, future research should emphasize the **analysis of event sequences** that lead to key outcomes (e.g., goals, successful area entries) and examine how these patterns relate to the competition styles identified in this paper.
- **Model Selection**:Explore alternative clustering algorithms, such as fuzzy clustering methods that allow for partial membership across clusters [3], which may better capture the nuanced nature of playing styles.
- **Data Expansion and Validation**:Extend the analysis to additional seasons or other leagues to assess the generalizability and stability of the identified playing styles.

7 Code Access Link

The code used in this paper can be accessed here:
https://github.com/path2morepro/HockeyAnlysis.git

References

1. Leonardo Lamas, Junior Barrera, Guilherme Otranto, and Carlos Ugrinowitsch. Invasion team sports: strategy and match modeling. *International Journal of Performance Analysis in Sport*, 14(1):307–329, 2014.
2. Felicity Lord, David B Pyne, Marijke Welvaert, and Jocelyn K Mara. Identifying and analysing game styles and factors influencing a team's strategy in field hockey. *Journal of Sports Sciences*, 40(8):908–919, 2022.
3. Rasmus Säfvenberg, Niklas Carlsson, and Patrick Lambrix. Identifying player roles in ice hockey. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*, pages 131–143. Springer, 2023.

4. Guiliang Liu, Wang Zhu, and Oliver Schulte. Interpreting deep sports analytics: Valuing actions and players in the nhl. In *Machine Learning and Data Mining for Sports Analytics: 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings 5*, pages 69–81. Springer, 2019.
5. Axel Rosendahl and Anton Olivestam. Player type classification in ice hockey using soft clustering, 2024.
6. Boyang Zhang and Tommy Löwendahl. Exploring the dynamics of ice hockey strategies using yolov8 and gephi in sports education.
7. Oliver Schulte. Valuing actions and ranking hockey players with machine learning. In *Linköping Hockey Analytics Conference*, pages 2–9, 2022.
8. Felicity Lord, David B Pyne, Marijke Welvaert, and Jocelyn K Mara. Methods of performance analysis in team invasion sports: A systematic review. *Journal of sports sciences*, 38(20):2338–2349, 2020.
9. Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
10. Claude B Vincent and Byron Eastman. Defining the style of play in the nhl: An application of cluster analysis. *Journal of Quantitative Analysis in Sports*, 5(1), 2009.