

Discovering and Visualizing Tactics in a Table Tennis Game Based on Subgroup Discovery

Pierre Duluard^{1,3}, Xinqing Li^{2,3}, Marc Plantevit⁴, Céline Robardet^{2,3}, and Romain Vuillemot^{1,3}

¹ École Centrale de Lyon

² INSA Lyon

³ LIRIS CNRS UMR 5205

⁴ Epita

Abstract. We report preliminary results to automatically identify effective tactics of elite table tennis players. We define these tactics as subgroups of winning strokes that table tennis experts seek to identify in order to train players and adapt their strategy during play. We first report how we identify and classify these subgroups using the weighted relative accuracy measure (WRAcc). We then present the subgroups using visualizations to communicate these results to our expert. These exchanges allow rapid feedback on our results and makes it possible further improvements to our discoveries.

Keywords: Data Mining · Sports Data Visualization · Table Tennis.

1 Introduction

Table tennis is a racket sport ranked amongst the most popular physical activities played at both amateur and elite levels. It is practiced and followed by millions of sports enthusiasts, especially as an Olympic discipline since the Olympic Games of Seoul 1988. Thus, many international federations and clubs train players all around the world at various levels. Academics have also focused on this sport in many areas from video tracking to data mining and visualization. In this work we contribute to this area of research by reporting on a close collaboration with an organization in charge of training elite players, the **Table Tennis National partner (TTN)**⁵. This organization recently annotated videos of elite player games to evaluate descriptive game statistics (e.g., number of wins per type of serve). They sought to improve analysis of such datasets to train elite players and improve their tactical preparation before games. They also sought to get such analysis during games to provide insights to players on which strategy to focus on.

The challenge in this work is to reveal hidden patterns from table tennis datasets which contain short, yet rich stroke sequences, grouped by rally. Table tennis games usually contain up to a hundred rallies composed of a series of (on

⁵ Anonymized

average) half a dozen strokes each. Each rally ends with a score increased by one for one of the two players. Rallies are composed of multi-variate strokes with laterality (side of the racket being used), type of stroke and impact zone on the table. Many other parameters enter into account to analyze rallies such as who serves, players scores and previous sets won by each player.

Table tennis sport recently gained interest from data mining and visualization [9, 7]. Two main contributions have been proposed to visualize tactics in racket sports. The iTTVis [9] offers an interactive visualization system that works for analyzing and exploring table tennis data. The system is divided into several parts, and each part presents the evolution of the match from a different aspect. It also summarizes the statistical correlation of inter and intra stroke attributes. In other words, it shows the relationships of attributes of each stroke and between strokes. It does, however, rely on expert visual analysis of each rally to identify tactical patterns. Tac-Miner [7] is more advanced on this point. It focuses more on tactics than on the match. By merging several matches into a single presentation, it allows the analysis of several matches against the same opponent. It presents a tactic either globally, using frequency and win rate, or precisely, using detailed attributes. Players and study attributes can be selected. Compared to iTTVis, it lacks integrity in the evolution of the whole match, but shows more comparative and correlation analysis of different tactics.

In a first approach, we performed an exploratory data analysis using simple statistics such as data distribution and frequency calculation. This gave us a good overview of the games. We also explored several synchronized views deployed for our partner as a dashboard to quickly explore data without technical expertise (Figure 1). This approach allowed our experts to develop a deeper understanding of the tactical possibilities from the data. However, they did not identify complex tactics involving action sequences. After several weeks of discussions, we have identified the following issues to address regarding tactical analysis using sequences:

- What are the most effective serve and hit zone combinations?
- Are there recurring behaviors to win a rally?
- How to characterize player profiles?

These questions are so far under-explored by analysts or are resolved by subjective analysis. Recent research has focused on intra-stroke analysis whereas we aim to focus on the tactical level by exploring discriminant sub-sequences to identify tactics. In particular, we focus on winning stroke combinations to characterize the players' tactics. In the following, we also provide visual representations so that experts can quickly grasp the result and get context about the sequences.

2 Methodology

The goal of our research is to discover useful tactics that lead to success. According to the **TTN**, a tactic consists of two consecutive strokes for a player, which

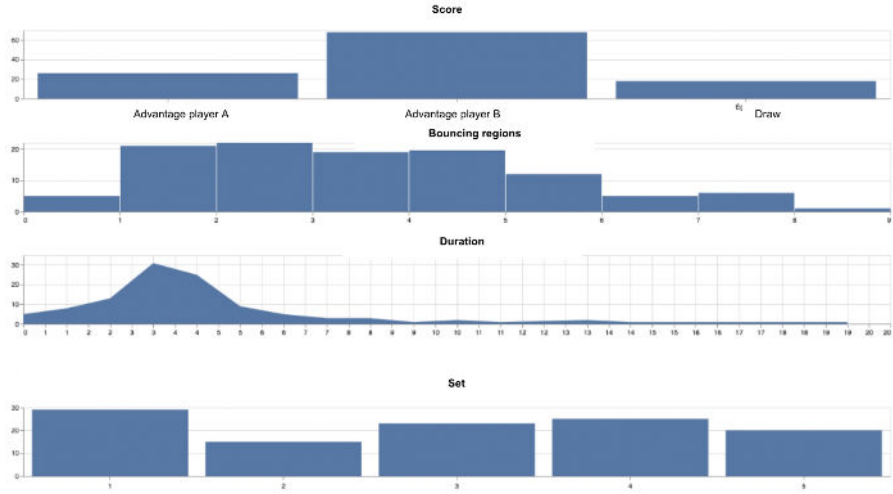


Fig. 1. Game statistics for a table tennis game between two players.

means that a tactic consists of three consecutive strokes. Indeed, the player, who serves, perfectly controls the first two strokes of an exchange, which can thus constitute a tactic, while thereafter he acts according to the actions of his opponent. Finding useful tactics is not exclusive to table tennis competitions, but pertains to many other sports. We find a similar problem in an article on football [2], which addresses the problems of low repetition between items of a sequence and the inequality of sequence length. It uses Dynamic Time Wrapping (DTW) [6] to compute the similarity between two sequences of different length, then uses CM-SPADE [3] to find frequent sub-sequences. However, unlike soccer, table tennis data has close relationships between players and stroke order. In football games, we define a sequence as a list of consecutive moves within a limited time interval, which can include several continuous moves by the same player or moves by different players at the same time. In contrast, the sequences set in a table tennis match are actually a rally. Thus, in a list of consecutive moves, the two players appear alternately in the sequence, which implies a strong correspondence between the sequence of strokes and the player. The use of the DTW, which can associate items with different positions in the two sequences, causes in this situation problems of correspondence because we cannot say that a stroke of player A is similar to a stroke of player B in our situation. This is why we turn to the extraction of frequent and/or discriminating sub-strings. These concepts are formally presented below.

2.1 Dataset

The annotated dataset we work with comes from a single table tennis game. It includes match information that corresponds to the following hierarchical order:

$Match \rightarrow Set \rightarrow Rally \rightarrow Stroke$. In table tennis, players take turns hitting the ball with their racket and bouncing it off the opponent’s side of the table (with the exception of the serve, which must bounce off both sides). Thus, a rally is lost by a player if he fails to return the ball as described above. A set is won by a player when he reaches 11 points or more, with a difference of 2 points between him and his opponent. The information included at each level is shown in Figure 2. In this work, the level of analysis of the game relates to the rallies,

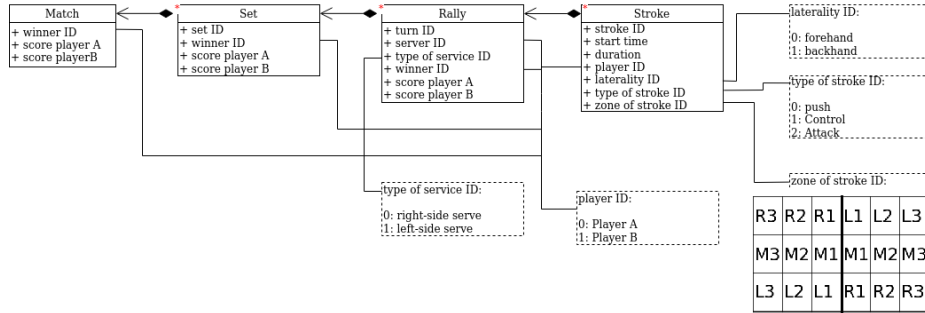


Fig. 2. Table tennis game data model. In this work we are mainly interested in stroke sequences whose attributes are described in table Stroke. Notice the locations of the 9 possible impact zones on the table.

that is to say the sequences of strokes until winning by one of the two players. Additional information provides context to the rallies, and we will use the score result as the criteria for success in such a rally (and possibly the tactics used in it). Notice that a sequence of strokes (i.e., a rally) has the following structure:

- It begins with a serve (Rss for a right side serve and Lss for a left side serve of a player) which hits an impact zone (9 possible areas Z1 to Z9, see Figure 2) on the opponent’s table part.
- There follows a sequence of strokes described by the type of stroke (C for a Control, A for an Attack and P for a Push), the laterality (B for a Backhand and F for a Forehand) and the hit zone (Z1 to Z9).

Those choices are justified in Section 2.4. It is possible to use a graph to represent a set of rallies, as shown in Figure 3. The nodes of the graph represent the strokes and the edges represent the transitions between the strokes (the edges are then labeled with a number equal to the number of rallies concerned by the transition). We also order the nodes so that we read the rallies from left to right: the leftmost node is the serve and the rightmost node indicates the winner of the rally. The square nodes represent the strokes of Player A, while the oval ones represent the strokes of Player B.

For example, this graph represents three rallies. The first one is **Player A’s Forehand Right side serve in R3** → **Player B’s Forehand Attack in M3** → **Player A’s Forehand Control in M3** → ... → **Player B wins**. The

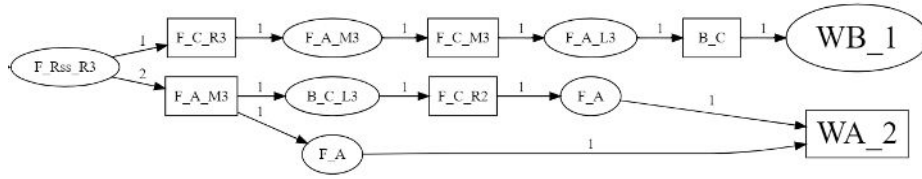


Fig. 3. Example of rally represented by a graph: each node is a stroke (player A is represented as squares, player B as circles).

two other rallies begin similarly with **Player A's Forehand Right side serve in R3** → **Player B's Forehand Attack in M3** and are then different until **Player A** wins.

2.2 Tactics in Table Tennis

We define tactics as a sequence of consecutive strokes whose goal is to win the point for the player who uses it. According to **TTN**, a tactic can be seen as a sequence of three consecutive strokes even if the three strokes are not related to the same player. However, we assume the server perfectly controls his first two strokes of an exchange. Thus, with his service, he reduces the possible strokes of his opponent. This echoes previous work in tactical analysis [7, 8]. A tactic occurs several times during the game and generalizes a sequence of strokes by identifying the key elements that characterize it (laterality, type or area).

2.3 Mining Frequent and Discriminant Sequential Pattern

Each such sequence is associated to a label that indicates the player who won the rally (WA if it is player A, WB otherwise). The set of such sequences is denoted \mathcal{D} in the following.

From the sequences that represent rallies we consider subsequences that occur frequently in the data [3]. The occurrence of one sequence in another is specified in Definition 1.

Definition 1 (Occurrence of a sequence in another one). *A sequence $S_A = X_1, X_2, \dots, X_k$, where X_1, X_2, \dots, X_k are itemsets, is said to occur in another sequence $S_B = Y_1, Y_2, \dots, Y_m$, where Y_1, Y_2, \dots, Y_m are also itemsets, if and only if there exists integers $1 \leq i_1 < i_2 < \dots < i_k \leq m$ such that $X_1 \subseteq Y_{i_1}$, $X_2 \subseteq Y_{i_2}$, ... $X_k \subseteq Y_{i_k}$. It is denoted by $S_A \sqsubseteq S_B$. The support of S_A in the database \mathcal{D} is the number of sequences $S \in \mathcal{D}$ where $S_A \sqsubseteq S$ divided by the total number of sequences in \mathcal{D} .*

A sequence is considered as a tactic if it satisfies two constraints:

1. it is an alternating sequence of consecutive strokes played by each of the players,

2. its frequency is higher than a threshold **MinSupp**.

Definition 2 (Alternate sequence). *An sequence $S_A = X_1, \dots, X_n$ is alternate if all the itemsets with an even index are played by a player, and the odd ones by the other player. Furthermore, the itemsets are consecutive in the sequences $S \in \mathcal{D}$ where S_A is considered to occur: considering $S = Y_1, \dots, Y_m$, we have $S_A \subseteq S$ if there exists an integer $1 < i < m - n + 1$ such that $X_1 \subseteq Y_i$, $X_2 \subseteq Y_{i+1}, \dots, X_n \subseteq Y_{i+n-1}$.*

To be able to answer the question of interest of the **TTN**, we perform supervised descriptive rule discovery [4]. In our context, it consists in discovering alternate sequential patterns whose supporting rallies are mainly won by a given player j . This is what is called subgroup discovery [1]. The quality of an alternate sequence to describe the tactic of player j is measured by the Weighted Relative Accuracy measure (WRAcc) [5]. It requires defining a measure of support on player j 's winning rallies, $\mathbf{Supp}(S, \mathcal{D}_j)$, as the number of sequences with label W_j where the sequence S occurs, divided by the total number of sequences with label W_j .

Definition 3 (Weighted relative accuracy). *Weighted relative accuracy of an alternate sequence S to characterize the winning rallies of player j is defined by*

$$\begin{aligned} \text{WRAcc}(S, W_j) &= P(S) \cdot (P(W_j|S) - P(W_j)) \\ &= \mathbf{Supp}(S, \mathcal{D}) \cdot \left(\frac{\mathbf{Supp}(S, \mathcal{D}_j)}{\mathbf{Supp}(S, \mathcal{D})} - \mathbf{Supp}(\langle \rangle, \mathcal{D}_j) \right) \end{aligned}$$

with $\langle \rangle$ the empty sequence that generalizes all the sequences of the dataset.

We use SPADE [3] to compute frequent alternate sequences. We adapt it by modifying the sequence containment used in the algorithm.

2.4 Summary of assumptions

A first category of assumptions is related to the choices of possible values for each attribute when annotating a stroke. These choices were made by the **TTN**.

According to the **TTN**, for the types of strokes, the choice of the three values (attack, control and push) allows to describe the player's intention, which is for the **TTN** more important than knowing the exact technique used for the stroke.

Regarding the zones on the tennis table, the separation into 3 side zones (L, M and R) is enough to describe whether a ball is sent to the player, or to his forehand or his backhand. According to the **TTN** the separation into 3 depth zones (1, 2 and 3) is sufficient to describe the player's intention.

A second category of assumptions concerns how to define a tactic. Indeed, it was agreed with the **TTN** that a tactic would be an alternation by each player of three consecutive strokes. Indeed, the rallies are only composed of 3 to 4 strokes on average and it is therefore useless to be interested in longer tactics.

In addition, it was decided that all the tactics of a player in a winning rally were winning. This choice makes it easy to classify the tactics.

3 Results

We applied the methodology introduced in the previous section to a match that opposes elite players (**Player A** and **Player B**) during an international game. In order to effectively communicate our results with the **Table Tennis National partner**, we also designed several visualizations of table tennis sequences.

3.1 Presentation of the obtained alternate sequences

In this part, we use the SPADE algorithm to determine the most frequent tactics (with **MinSupp**=5%) then we calculate the WRAcc measure for each of them in order to only keep the most relevant ones. The most interesting subgroups for each player are represented in Tables 1 and 2

WRAcc	Frequency	Winrate	Player A's stroke	Player B's stroke	Player A's stroke
0.02984	12.9%	78.6%	M1	Forehand	Forehand
0.02984	12.9%	78.6%	M1	Forehand	Attack
0.02984	12.9%	78.6%	M1	Forehand,Push	Forehand
0.02984	12.9%	78.6%	M1	Forehand,Push	Attack
0.02881	14.8%	75.0%	M1	Push	Attack
0.02469	5.6%	100.0%	Right-side Serve	Forehand	R3

Table 1. Player A's tactics extracting by SPADE with high WRAcc.

WRAcc	Frequency	Winrate	Player B's stroke	Player A's stroke	Player B's stroke
0.03086	5.6%	100.0%	Forehand	Push	R3
0.02881	11.1%	66.7%	Forehand	Control	L3
0.02778	8.3%	77.8%	R3	Control	Forehand
0.02778	9.3%	70.0%	Forehand	Control	Attack,L3
0.02778	8.3%	77.8%	R3	Forehand,Control	Forehand
0.02675	19.4%	66.7%	Forehand	Control	Attack

Table 2. Player B's tactics extracting by SPADE with high WRAcc.

The previous tactics are calculated on the complete dataset, thus including rallies where **Player A** is a server as well as those where it is **Player B**. However, according to the **TTN**, the tactics used by a player in a rally strongly depend on whether this player is serving or receiving for this rally. This is why we have reapplied our method on a reduced dataset only composed of rallies where one of the two players is serving. Best **Player A**'s tactics when **Player A** is serving are represented in Table 3. The other configurations, such as best **Player A**'s tactics when **Player B** serves, and similar results for **Player B**'s tactics are given in Appendix A.

WRAcc	Frequency	Winrate	Player A's Stroke	Player B's Stroke	Player A's Stroke
0.05024	16.4%	88.9%	Forehand R_ss, M1	Forehand Push	Forehand
0.03967	18.2%	80.0%	Forehand R_ss	Forehand Push	Forehand Attack
0.03041	7.3%	100.0%	R3	Backhand	Forehand Attack, R3
0.02744	10.9%	83.3%	Forehand	Control	Backhand Control
0.03802	5.5%	100.0%	Backhand, R3	Backhand	Forehand Attack, R3

Table 3. Best **Player A**'s tactics when **Player A** is server.

These tactics seem to be more relevant because of their higher WRAcc. For example, Table 3 shows that **Player A**'s best tactic when serving is to start with a right side serve in M1. Thus, this serve tends to force **Player B** to do a push, which allows **Player A** to take the lead.

In order to compare them, we can also look at the losing tactics by selecting tactics with the worst WRAcc in each case. Worst **Player A**'s tactics when he serve are given in Table 4 and the other configuration in appendix A.

WRAcc	Frequency	Winrate	Player A's Stroke	Player B's Stroke	Player A's Stroke
-0.09355	25.0%	21.4%	R_ss	Forehand	L3
-0.06347	10.9%	0.0%	Forehand R_ss, M2	Forehand	Control, M3
-0.05289	9.1%	0.0%	Forehand R_ss, M2	Attack	Control

Table 4. Worst **Player A**'s tactics when **Player A** is server.

In contrast to the right side serve in M1, Table 4 shows that tactics starting with a right side serve in M2 are losing for **Player A**. This is because the M2 service is risky. Indeed, the objective of this serve is to be short enough so that the opponent cannot make an offensive stroke but long enough so that it is difficult to make a push. According to Table 4 line 3, **Player B** made offensive returns on this serve, which means that **Player A** made too long serves in M2. To get a better idea of how **Player A** uses these tactics, we can look at the evolution of the use of these different serves in Figure 4.

Figure 4 presents the types of serves used during the match: on the ordinate we have the types of serves and on the abscissa the rallies ordered in time. The color of each point identifies the winner of each rally. On Figure 4, we can highlight the fact that **Player A** started the game with a risky strategy by using right side serves in M2. However, as seen in Figure 4 and Table 4, this strategy was not successful. **Player A** then used right side serves in M1 from the 7th rally. It was a less risky serve that allowed him to regain the advantage over **Player B**.

3.2 Visualization of the tactics

To explore data and visualize subgroups discovered through data mining, we can use the graph representation of rallies described in Section 2.1. Examples of this

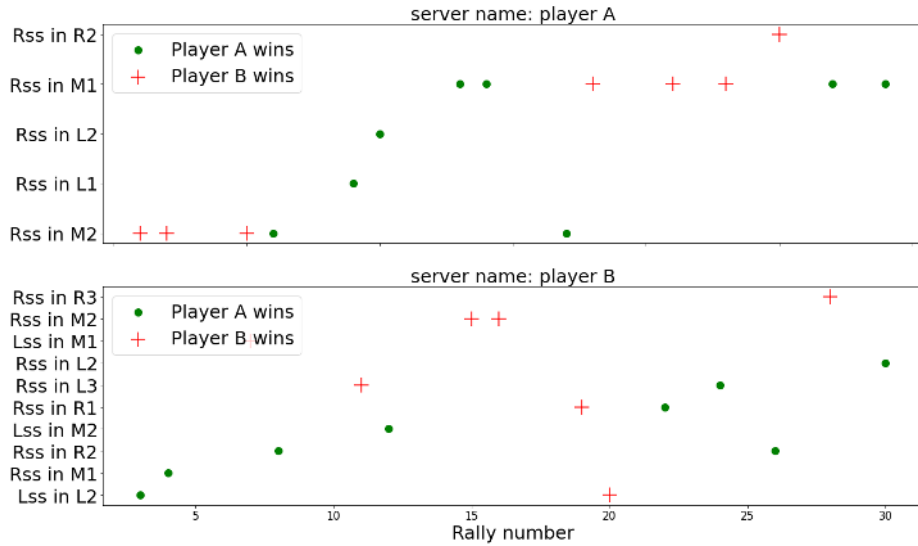


Fig. 4. Evolution of the serves used by **Player A** and **Player B**.

representation are given for the tactics analyzed in section 3.1.

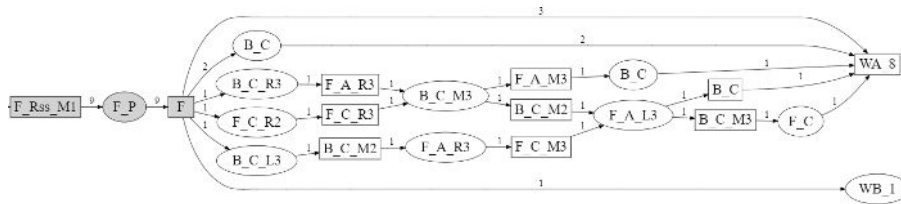


Fig. 5. Graph representation of best **Player A**'s tactic when he serves.

Figure 5 represents the following **Player A**'s tactic : **Player A**'s **forehand Right-side serve in M1** → **Player B**'s **Forehand Push** → **Player A**'s **forehand**. This corresponds to the first tactic in Table 3.

Figure 6 represents the following **Player A**'s tactic: **Player A**'s **forehand Right-side serve in M2** → **Player A**'s **attack** → **Player A**'s **control**. It corresponds to the second tactic in Table 4.

This representation allows to easily read the sequence of strokes present in rallies containing the selected tactic. This visualization reflects the risky nature of the right side serve in M2 because the rallies are short, while the tactic using the right side serve in M1 gives longer rallies, although it allows **Player A** to take the lead.

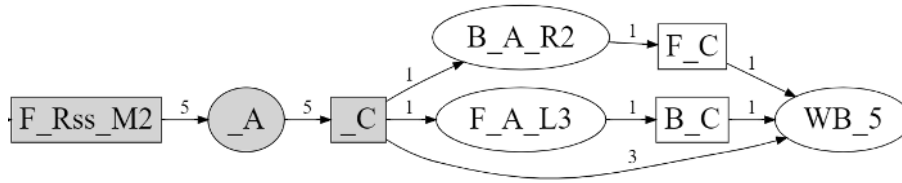


Fig. 6. Graph representation of one of worst **Player A**'s tactic when he serves.

4 Conclusion and Perspectives

We have introduced some preliminary results for discovering and visualizing tactics in a single Table Tennis game, based on subgroup discovery. Given Tennis Table data, we show that it is possible to discover some tactics that have a positive impact on the score of the player (i.e., the fraction of points that are won increases). We believe that such method can support knowledge discovery from Tennis Table games and provide insights for both the players and their coaches. However, a number of potential limitations need to be considered for future research to make this method effective in practice. First, a tactic – no matter how effective it is – must be used wisely. If a tactic is always used by a player, the opponent will adapt and its effectiveness will decrease through the game. It is therefore important to provide more context to a tactic (e.g., momentum of the match, score, set). A promising direction is to monitor the effectiveness of a tactic and to detect the adaptation of the opponent. Some links with Bayesian Nash Equilibrium should be investigated. Eventually, it is important to study sets of tactics instead of tactics individually. Finally the visual representation of all the tactics and their use within or between games is a challenge that remains to be addressed.

5 Acknowledgment

This project was partially funded by Action Transversale at LIRIS Lab. We thank **Table Tennis National partner** for the annotated game they provided us and their time to provide us with feedback on the results. We also thank the reviewers for their thoughtful comments.

References

1. M. Atzmueller. Subgroup discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.
2. Decroos, Tom, Van Haaren, Jan, and Davis, Jesse. Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 223–232, 2018. Association for Computing Machinery.
3. Fournier-Viger, Philippe and Gomariz, Antonio and Campos, Manuel and Thomas, Rincy. Fast Vertical Mining of Sequential Patterns Using Co-occurrence Information. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 40–52, 2014.
4. P. Kralj Novak, N. Lavrač, and G. I. Webb. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *J. Mach. Learn. Res.*, 10:377–403, jun 2009.
5. R. Mathonat, D. Nurbakova, J.-F. Boulicaut, and M. Kaytoue. SeqScout: Using a Bandit Model to Discover Interesting Subgroups in Labeled Sequences. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 81–90, Oct. 2019.
6. M. Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.
7. J. Wang, J. Wu, A. Cao, Z. Zhou, H. Zhang, and Y. Wu. Tac-Miner: Visual Tactic Mining for Multiple Table Tennis Matches. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):2770–2782, June 2021. Conference Name: IEEE Transactions on Visualization and Computer Graphics.
8. J. Wang, K. Zhao, D. Deng, A. Cao, X. Xie, Z. Zhou, H. Zhang, and Y. Wu. Tac-Simur: Tactic-based Simulative Visual Analytics of Table Tennis. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):407–417, Jan. 2020. Conference Name: IEEE Transactions on Visualization and Computer Graphics.
9. Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, and H. Zhang. iTTVis: Interactive Visualization of Table Tennis Data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):709–718, Jan. 2018. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

A Appendix

WRAcc	Frequency	Winrate	Player A’s Stroke	Player B’s Stroke	Player A’s Stroke
0.05340	11.3%	100.0%	Push	Forehand Push	Attack
0.04450	9.4%	100.0%	Backhand Push	Forehand Push	Attack
0.04450	9.4%	100.0%	Push, M1	Forehand	Backhand
0.03560	13.2%	85.7%	Backhand	Forehand	Backhand Attack
0.03560	7.5%	100.0%	Backhand Push	Forehand Push	Backhand Attack

Table 5. Best **Player A**’s tactics when **Player B** is server

WRAcc	Frequency	Winrate	Player B's Stroke	Player A's Stroke	Player B's Stroke
0.05981	11.3%	100.0%	R3	Forehand Control	Forehand
0.05304	20.8%	72.7%	Forehand	Control	Attack
0.04984	9.4%	100.0%	Forehand, R3	Forehand Control	Forehand
0.04984	9.4%	100.0%	Forehand	Forehand Control	Attack
0.04984	9.4%	100.0%	Attack	Forehand Control	Forehand

Table 6. Best **Player B's** tactics when **Player B** is server

WRAcc	Frequency	Winrate	Player B's Stroke	Player A's Stroke	Player B's Stroke
0.03471	9.1%	80.0%	Control	Attack, L3	Forehand
0.03174	5.4%	100.0%	Control	Backhand Attack,M3	Backhand Control
0.03174	5.4%	100.0%	Attack	Backhand Control	Attack
0.02711	10.9%	66.7%	Backhand	L3	Forehand
0.02413	7.3%	75.0%	Backhand Control	Backhand Attack	Control

Table 7. Best **Player B's** tactics when **Player A** is server

WRAcc	Frequency	Winrate	Player A's Stroke	Player B's Stroke	Player A's Stroke
-0.09078	20.8%	9.1%	Control	Attack	Control,R3
-0.08971	17.0%	0.0%	M3	Forehand	R3
-0.08188	22.6%	16.7%	Control	Attack	Control

Table 8. Worst **Player A's** tactics when **Player B** is server

WRAcc	Frequency	Winrate	Player B's Stroke	Player A's Stroke	Player B's Stroke
-0.09790	20.8%	0.0%	Forehand	Forehand	M1
-0.09683	23.6%	7.7%	Forehand	Attack	L3
-0.09576	27.3%	13.3%	Forehand	Forehand	Backhand, M3

Table 9. Worst **Player B's** tactics when **Player B** is server

WRAcc	Frequency	Winrate	Player B's Stroke	Player A's Stroke	Player B's Stroke
-0.05322	12.7%	0.0%	Backhand	Attack, R3	M3
-0.04562	10.9%	0.0%	Forehand Push	Forehand	Control, R3
-0.04562	10.9%	0.0%	Forehand Push	Control	Control, R3

Table 10. Worst **Player B**'s tactics when **Player A** is server