

Analyzing Volleyball Match Data from the 2014 World Championships Using Machine Learning Techniques

Jan Van Haaren

KU Leuven — Department of Computer Science
Celestijnenlaan 200A, 3001 Leuven, Belgium
jan.vanhaaren@cs.kuleuven.be

Jesse Davis

KU Leuven — Department of Computer Science
Celestijnenlaan 200A, 3001 Leuven, Belgium
jesse.davis@cs.kuleuven.be

Horesh Ben Shitrit

PlayfulVision
Ch. de la Raye 13, 1024 Ecublens, Switzerland
horeshb@playfulvision.com

Pascal Fua

EPFL — Computer Vision Laboratory
Station 14, 1015 Lausanne, Switzerland
pascal.fua@epfl.ch

ABSTRACT

This paper proposes a relational-learning based approach for discovering strategies in volleyball matches based on optical tracking data. In contrast to most existing methods, our approach permits discovering patterns that account for both **spatial** (that is, partial configurations of the players on the court) and **temporal** (that is, the order of events and positions) aspects of the game. We analyze both the men's and women's final match from the 2014 FIVB Volleyball World Championships, and are able to identify several interesting and relevant strategies from the matches.

Keywords

Sports analytics, Spatial data, Strategy detection

1. INTRODUCTION

Optical tracking enables analyses that go beyond descriptive statistics based on event data by providing sufficient context about the game state such that it is possible to gain insights into tactics and strategies. However, automatically detecting and understanding tactics in dynamic sports such as volleyball is a challenging problem as tactics comprise complex interactions among multiple players that evolve across both time and space. Hence, strategy detection must consider and evaluate a huge number of possible spatio-temporal movement patterns. Existing approaches, which largely look at basketball and soccer, tackle this problem by restricting the analysis in some way such as only considering a specific strategy [9], a short time window [4], or a small subset of players [5, 3].

This paper addresses automatically discovering attacking strategies in volleyball matches based on camera-tracking data of the players and the ball as a pattern mining task [1]. Specifically, we consider the following two tasks:

Task 1: Identify a team's attacking patterns in volleyball matches that occur frequently in *won* rallies and infrequently in *lost* rallies.

Task 2: Identify attacking patterns in a volleyball match that are used by one team but not the opposing team.

In contrast to most existing approaches, we attempt to identify patterns that account for both **spatial** and **temporal** aspects of the game. That is, we want to model (partial) configurations of players' positions on the court as well as how play evolves over time. To illustrate this, consider the following simple pattern automatically discovered by our approach:

```
IF player #13 performs the dig AND NEXT
   player #1 performs the set
   in the front center zone AND NEXT
   player #8 performs the spike
   in position 81 of the court
THEN the attack is likely to be successful.
```

This pattern is temporal as the dig occurs first, the set second, and the spike third. The pattern is spatial as it states the location on the court where the set and spike occur (see Figure 1 for a description of the locations).

In order to automatically identify patterns like the one just shown, we use a relational-learning based approach. As much as possible, we attempt to employ a data-driven approach that can automatically determine which players and characteristics of the game state are relevant to the strategy and should be included in the pattern. We analyze data from both the men's and women's final match from the 2014 FIVB Volleyball World Championships. Our top-ranked discovered patterns represent strategies that are both interesting and relevant from a volleyball perspective.

2. BACKGROUND ON VOLLEYBALL

Volleyball [10, 2] is a ball sport that is played by two teams of six players each. A volleyball court is 18 meters (59 feet) long and 9 meters (29.5 feet) wide. Each team occupies one half of the court, which is 9 meters by 9 meters. The halves are separated by a net whose top is 2.43 meters above the floor in men's competitions and 2.24 meters in women's competitions. The overall goal is to score points by grounding the ball on the opponent's court.

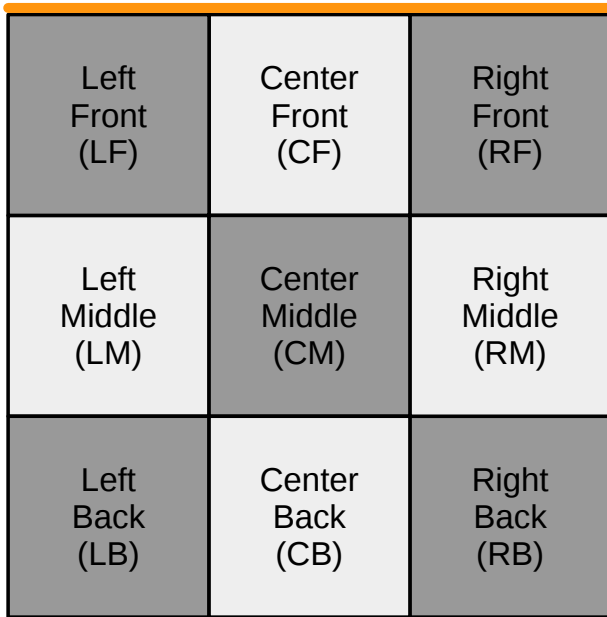
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

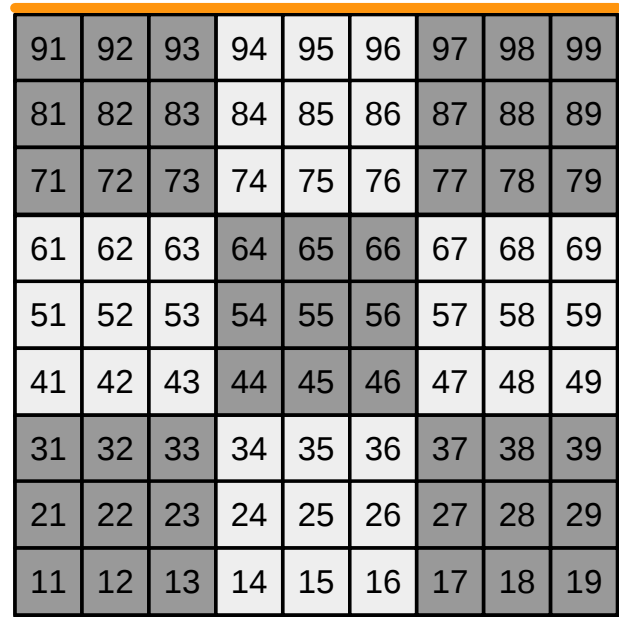
© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2939672.2939725>

Figure 1: The division of the pitch in higher-level zones (left) and lower-level positions (right). The direction of play is from bottom to top, where the thick orange line at the top represents the net.



Locations encoded as 3 meter by 3 meter zones.



Locations encoded as 1 meter by 1 meter positions.

Volleyball matches are won by the first team to win three sets. A set is won by the first team to score 25 points and lead by two points. However, the fifth set is typically played to only 15 points. Each set consists of rallies and one team is awarded a point at the end of each rally. A rally starts by serving the ball from behind the back-line over the net into the opponent's court. The opponent may touch the ball up to three times to prevent it from hitting the court and to get the ball back over the net. A rally ends either when one of the teams makes a kill by grounding the ball on the opponent's court or when a team makes a foul.

The players follow a rotation scheme, where they must rotate one time in clockwise direction after their team wins the serve. Nevertheless, players do have different roles in the team and are free to move within their half of the court after the serve. Typically, the best offensive players will move towards the net, while the best defensive players will move to the back of the court. Although the rotation scheme imposes some restrictions, this tactical freedom allows teams to adopt a wide variety of match strategies. Therefore, volleyball players need to master the following six basic skills:

Serve The serve is the skill of moving the ball from behind the back-line into the opponent's court. While many different types of serves are used, the most popular type is the jump serve, where the server first tosses the ball high in the air and then jumps to hit it.

Dig The dig is the defensive skill of preventing the ball from hitting the court when the ball is nearly touching the floor after a serve or attack from the opponent. The dig is a reflex-based skill which often requires a player to dive towards the ball.

Pass The pass is very similar to the dig. However, in addition to preventing the ball from hitting the court, this

defensive skill also involves moving the ball towards a team mate that is well-placed to set up an attack.

Set The set is the offensive skill of pushing the ball into the air such that a team mate can hit it into the opponent's court. The *setter*, who is the player performing the set, coordinates the offensive play of the team by deciding who will eventually attack the ball.

Spike The spike is the offensive skill of hitting the ball such that the opponent cannot prevent it from touching their court. The *spiker*, who is the player performing the spike, first makes a few steps and then jumps to swing at the ball.

Block The block is the skill of stopping or altering an opponent's attack by players standing at the net. An *offensive* block aims at keeping the ball into the opponent's court, while a *defensive* block aims at getting the ball under control by slowing it down.

Typically, the dig or pass is the first contact that the team makes with the ball, the set the second contact, and the spike the third contact. In this paper, we will focus on these three skills.

3. DATA

The data were collected at the FIVB Volleyball World Championships finals that were held in Poland¹ and Italy² in 2014. PlayfulVision [7] recorded several men's and women's

¹<http://poland2014.fivb.org/en>

²<http://italy2014.fivb.org/en>

matches, including the final matches of both tournaments. They first captured each match using 8 video cameras placed at different angles at 30 frames per second and then used their ball and player tracking code to automatically determine the locations of the players and the ball in each frame. Furthermore, a human manually annotated each frame with the skills performed by the players (i.e., serve, dig, set, spike, and block) for both finals. The data set does not distinguish between digs and passes.

In this paper, we focus on the final matches of the world championships for which both the tracking information and the annotations are available. Table 1 shows the number of sets, rallies, and attacks in both matches as well as relevant statistics for each team such as the number of won rallies and how often they performed each of the skills.

We divide each rally into a series of attacks, where each attack consists of a sequence of consecutive skills performed by the same team. In this work, we only consider attacks that involve each of a dig, a set, and a spike. Based on this preprocessing, we construct positive and negative examples for four settings on each task. In task 1, the positive examples are a team's attacks that result in a point, while the negatives are all the team's other attacks. In task 2, the positives are the attacks from one team (e.g., Brazil), while the negatives are the attacks from the opponent (e.g., Poland). Table 2 lists the number of positive and negative examples for each setup.

To represent an attack, we take a snapshot of the pitch configuration at the time that a skill is performed. Each snapshot is described by the performed skill (dig, set or spike) as well as information about the location of each player and the ball. Each snapshot describes the locations on two levels of granularities: high-level zones (see Figure 1 left) and lower-level positions (see Figure 1 right).

4. RELATIONAL-LEARNING APPROACH TO STRATEGY DISCOVERY

Pattern mining typically focuses on finding patterns that occur frequently (that is, many times) in the data. Our problem has several other important characteristics that we must account for and which distinguish it from standard pattern mining. First, each of our tasks requires differentiating between two classes of examples, either successful and unsuccessful attacks or attacks done by two different teams. Second, we want to be able to simultaneously reason about multiple different levels of granularity in the data. For example, we may want to represent the court with both high-level zones and lower-level positions as illustrated in Figure 1. Furthermore, we would like to be able to discover patterns that involve specific players in a specific position as well as patterns that involve any player in a specific position. Third, this problem is inherently relational and it is crucial to find patterns that account for relationships such as changes over time. For example, we may want to know how the configuration of players changes between a dig and a set. Finally, we have specific knowledge about the game of volleyball and we would like to be able to incorporate it into the pattern mining process.

Based on the above requirements, we pursue an approach based on inductive logic programming (ILP) [6]. ILP is a relational learning approach that permits modeling multiple granularities, capturing relationships, and incorporating

background knowledge. ILP represents data using a subset of first-order logic, which is a commonly used representation language for relational data. Interestingly, this subset is equivalent to relational algebra, which forms the basis for relational databases and the SQL query language. Thus, another way to view ILP's approach to rule construction is that it learns the WHERE clause of an SQL query for a given set of tables in a relational database (the FROM part of the query), and a target attribute (the SELECT part of the query).

The goal of ILP is to learn a model that distinguishes positive examples from negative examples. The model is a set of IF-THEN rules. The IF portion of a rule is a set of conditions and the THEN portion has an outcome. If all the conditions in the IF portion are met, then the outcome can be expected (usually with a certain probability). An advantage of rules of this form is that they are easy for domain experts to interpret. However, we do have to make a number of modifications to the standard ILP setup to adapt it to our needs. ILP is traditionally used for learning classifiers, and hence it has a preference for smaller models. That is, models with as few rules as possible with each rule being as short as possible. In contrast, for knowledge discovery we want to find all interesting rules. Furthermore, all other things being equal, we would prefer a detailed pattern that gave as much information about the strategy (e.g., assigns positions and actions to as many players as possible). Thus, unlike model learning, we tend to prefer more specific or detailed patterns to more general ones. To address these problems, we use a two-step process involving pattern generation and pattern post processing.

Pattern Generation

We use the well-known and publicly available Aleph ILP system [8] to generate rules. Depending on the task, we define a set of positive and negative examples as described in the Data section. Aleph constructs one rule at a time, with the goal of discovering a set of conditions such that the condition applies to as many positive and as few negative examples as possible. In other words, it tries to maximize the precision of the rule or the conditional probability that the outcome is true given that the condition is true.

Instead of learning one rule at a time, we define a set of criteria and attempt to identify rules that satisfy the criteria. We consider patterns of up to size 15. Each pattern must apply to at least five examples (i.e., a frequency threshold). Each pattern must have a precision of at least 25%. Given the size of the search space (e.g., there are on the order of 10^{22} patterns of size 15 alone), an exhaustive approach is infeasible so we employ a beam search. We collect all rules found by the beam search that meet these conditions.

When constructing a pattern, the learning algorithm can make use of both levels of granularity for the location and it automatically decides, based on the data, which to use. In fact, a single pattern can simultaneously make use of both granularities. The patterns can refer to specific players (e.g., player 1 is in the front center zone) or generic ones (e.g., a player is in the front center zone). Referring to generic players allows us to find patterns that generalize across different lineups for a team as teams may substitute during the match.

		Men's final		Women's final	
		Brazil	Poland	China	USA
Sets	Total sets	4	4	4	4
	Won sets	1	3	1	3
	Lost sets	3	1	3	1
Rallies	Total rallies	185	185	188	188
	Won rallies	92	93	94	94
	Lost rallies	93	92	94	94
Attacks	Total attacks	246	248	300	306
Skills	Total serves	86	89	95	93
	Total digs	124	120	167	173
	Total sets	105	106	145	154
	Total spikes	112	112	148	156
	Total blocks	39	42	51	52

Table 1: Statistics for the men's and women's final at the 2014 World Championships.

		Men's final		Women's final	
		Brazil	Poland	China	USA
Task 1	Positive examples	51	49	55	61
	Negative examples	18	17	25	36
Task 2	Positive examples	96	95	132	144
	Negative examples	95	96	144	132

Table 2: The number of positive and negative examples for each task.

Pattern Post Processing

Even with our constraints, a huge number of patterns were generated. For task 1, there were 53,212 patterns generated for Brazil, 79,907 for Poland, 51,593 for China, and 77,509 for the USA. For task 2, there were 190,956 patterns generated for Brazil, 197,467 for Poland, 269,753 for China, and 327,731 for the USA. We post processed the patterns and only retained those that included all three skills (dig, set, and spike) we were interested in. This resulted in 1,527 patterns for Brazil, 1,301 for Poland, 242 for China, and 2,037 for the USA in task 1. For task 2, there were 65,484 patterns generated for Brazil, 41,763 for Poland, 73,011 for China, and 84,639 for the USA.

Then, we ranked the patterns by the number of specific players and locations the patterns contained. This enforces a preference for longer, more specific patterns. We broke ties by considering pattern coverage (number of positives the pattern applies to), pattern precision, and pattern length.

5. RESULTS AND DISCUSSION

We present the top-ranked pattern for each of the eight setups we consider. Most of these patterns capture the same offensive strategy, which involves attacking over one side of the court. The patterns do have some small variations, such as about whether the attack is on the left or right side. This attacking pattern is a well-known volleyball strategy. Next, we discuss the top-ranked pattern in each setup in more detail.

Task 1: Which offensive strategy is most successful for each team?

Figures 2 through 5 visualize the offensive patterns that occur in successful attacks but not unsuccessful ones. In all figures, numbers represent specific players, the capital letter X represents the location of the ball, and other capital letters (A, B, and C) represent generic players. That is, they denote the presence of a player in that location, but do not specify which player is there, and hence account for the fact that teams can substitute players. The net is shown in orange at the top of the image.

Figure 2 shows the top-ranked pattern for the Poland men's team. The setter, who is the player with jersey number 5, is already close to the net at the time of the dig. He sets the ball to the front left zone, where another player B spikes it. The pattern covers five successful attacks and no unsuccessful ones. Looking at the location of the spike, the Polish team had 36 successful (including the five covered by this pattern) and 16 unsuccessful spikes in the front left zone.

Figure 3 shows the top-ranked pattern for the Brazil men's team. This pattern is very similar to the one for the Polish team, but with two important exceptions. First, the spike occurs on the right side of the court. Second, the spike occurs in a very specific location, at the back right corner of the front right zone. The pattern covers five successful attacks and no unsuccessful ones. Here, the specific location is important to the pattern as, in the entire front right zone, Brazil had 22 successful attacks (including the five covered by this pattern) and 13 unsuccessful attacks. In terms of the specific position where player 4 was located, Brazil attempted 11 spikes of which nine were successful. Thus the

team was much more successful in this location than in the zone in general. The Polish team was less successful from this specific location, attempting 15 spikes of which 10 were successful. Also, in contrast to the women's final, the USA and China only attempted four spikes in aggregate in this specific location and none of them were successful.

Figure 4 shows the top-ranked pattern for the USA women's team. This again illustrates an attack from the side of the court at the front. Given that player B spikes the ball and the ball is located in the front left zone, that is the location where the spike was performed. Notice that a player denoted A moves from the left middle to the left front between the dig and the set. Thus, the setter could have set to either side for the spike. The pattern covers five successful attacks and no unsuccessful ones. The USA attempted 56 spikes in the front right zone of which 37 were successful (including the five covered by this pattern) and 19 were unsuccessful.

Figure 5 shows the top-ranked pattern for the China women's team. This pattern is less specific than the previous three, but it still shows the same general scheme of attacking from the side of the court on the front. Given that the ball is in the front rightmost corner next to the net at the time of the spike, we can infer that this is where player 8 is located. The pattern covers five successful attacks and no unsuccessful ones. Over all spikes in this specific position, China had eight successful spikes (including the five covered by this pattern) and two unsuccessful ones.

Task 2: Which common offensive strategy distinguishes a team from its opponent?

Figures 6 through 9 show the offensive patterns that distinguish between two teams playing against each other. The patterns are illustrated in the same manner as for Task 1.

Figure 6 shows the top-ranked pattern employed by the Poland men's team that is not used by the Brazilian team in their match. Figure 7 shows the top-ranked pattern for the Brazil men's team that is not used by the Polish team. These patterns are quite similar, with a set in the middle and a spike by player 10 on the outside in the front. The difference is that the Polish number 10 attacks on the right and the Brazilian number 10 on the left. One possible explanation is that these players have different dominant hands. Typically in volleyball the spiker wants his dominant hand closest to where the ball is coming from (that is, right-handed players want to spike from the front left, and left-handed players want to spike from the front right). We could not verify this hypothesis.

Figure 8 shows the top-ranked pattern for the USA women's team that is not used by the Chinese team. Figure 9 shows the top-ranked pattern for the China women's team that is not used by the USA team. These patterns are slightly less informative as they do not indicate the location of the spike. All the patterns tell is that the ball is set in the front center zone.

How do men's and women's volleyball compare to each other?

Another natural question to ask is how do the men's and women's game compare to each other. The statistics in Table 1 show that there are some commonalities between them. Namely, the number of rallies is roughly the same in both finals. However, the women's match features many more

attacks than the men's final. Furthermore, the number of digs, set, spikes, and blocks is much higher in the women's final than in the men's final. This is most likely due to the faster pace of men's volleyball, which makes it harder to gain control of the ball after an attack from the opponent.

6. CONCLUSION

We proposed a relational-learning based approach to identifying different strategies based on optical tracking data from volleyball matches. In contrast to most existing work, our approach is able to discover spatio-temporal patterns on multiple different levels of granularity that characterize successful attacking play in volleyball. Specifically, it discovered several strategies that make sense from a volleyball perspective. Ultimately, knowledge about what types of patterns work well can be used to alter match tactics and training sessions.

Acknowledgments

We thank the International Volleyball Federation (FIVB) for supporting the Volleyball Tracking System-project (VTS) and allowing PlayfulVision to use their video footage and data for academic research. Jan Van Haaren is supported by the Agency for Innovation by Science and Technology (IWT). Jesse Davis is partially supported by the KU Leuven Research Fund (C22/15/015), and FWO-Vlaanderen (G.0356.12, SBO-150033).

7. REFERENCES

- [1] C. Aggarwal and J. Han, editors. *Frequent Pattern Mining*. Springer, 2014.
- [2] FIVB. Official Volleyball Rules, 2015. [Online; accessed 12-February-2016].
- [3] K. Knauf and U. Brefeld. Spatio-Temporal Convolution Kernels for Clustering Trajectories. In *Proceedings of the KDD Workshop on Large-Scale Sports Analytics*, 2014.
- [4] P. Lucey, A. Bialkowski, P. Carr, Y. Yue, and I. Matthews. "How to Get an Open Shot:" Analyzing Team Movement in Basketball Using Tracking Data. In *Proceedings of the MIT Sloan Sports Analytics Conference 2014*, 2014.
- [5] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews. Assessing Strategy Using Spatio-Temporal Data. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2013.
- [6] S. Muggleton and L. De Raedt. Inductive Logic Programming: Theory and Methods. *The Journal of Logic Programming*, 19:629–679, 1994.
- [7] PlayfulVision. Volleyball Tracking System (VTS), 2016. [Online; accessed 10-February-2016].
- [8] A. Srinivasan. The Aleph Manual. *Machine Learning at the Computing Laboratory, Oxford University*, 2001.
- [9] J. Wiens, G. Balakrishnan, J. Brooks, and J. Guttag. To Crash or Not to Crash: A Quantitative Look at the Relationship Between Offensive Rebounding and Transition Defense in the NBA. In *Proceedings of the MIT Sloan Sports Analytics Conference 2013*, 2013.
- [10] Wikipedia. Volleyball — Wikipedia, The Free Encyclopedia, 2015. [Online; accessed 7-December-2015].

Figure 2: Frequent successful offensive pattern by Poland in the men's final. The setter, who is the player with jersey number 5, is already close to the net at the time of the dig. He sets the ball to the front left zone, where another player B spikes it. Here, player B could denote any other Polish player.

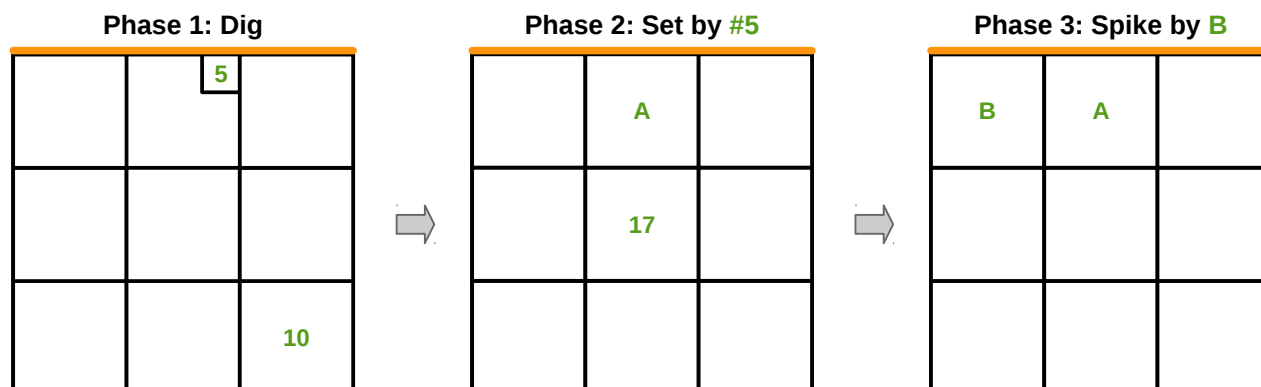


Figure 3: Frequent successful offensive pattern by Brazil in the men's final. The setter, who is the player with jersey number 1, is already close to the net at the time of the dig. He sets the ball from the front center zone to the back right corner of the front right zone, where the player with jersey number 4 spikes it. X denotes the location of the ball.

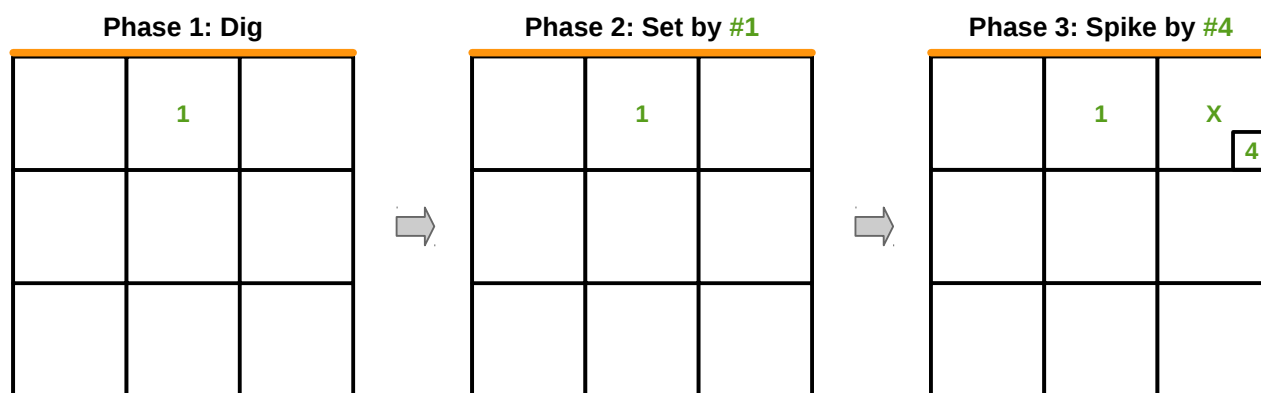


Figure 4: Frequent successful offensive pattern by the USA in the women's final. Player 1, who is the setter, is in the front center at the time of the dig. Player 1 sets the ball to the front left where player B spikes it. A player denoted A moves from the left center to the left front between the dig and the set. It appears that the setter could set to either side for a spike. X denotes the location of the ball.

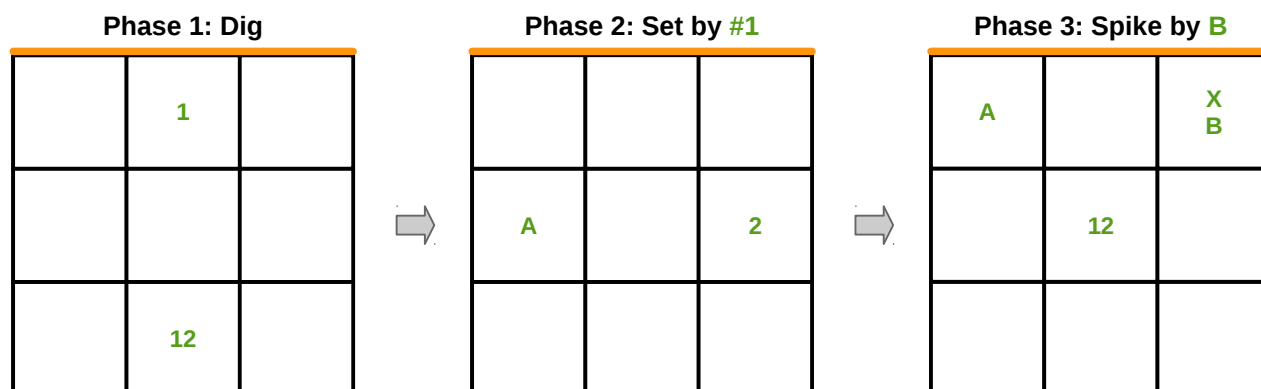


Figure 5: Frequent successful offensive pattern by China in the women's final. In this pattern, player 5 sets the ball and it is spiked by player 8 in the front right corner of the court. X denotes the location of the ball.

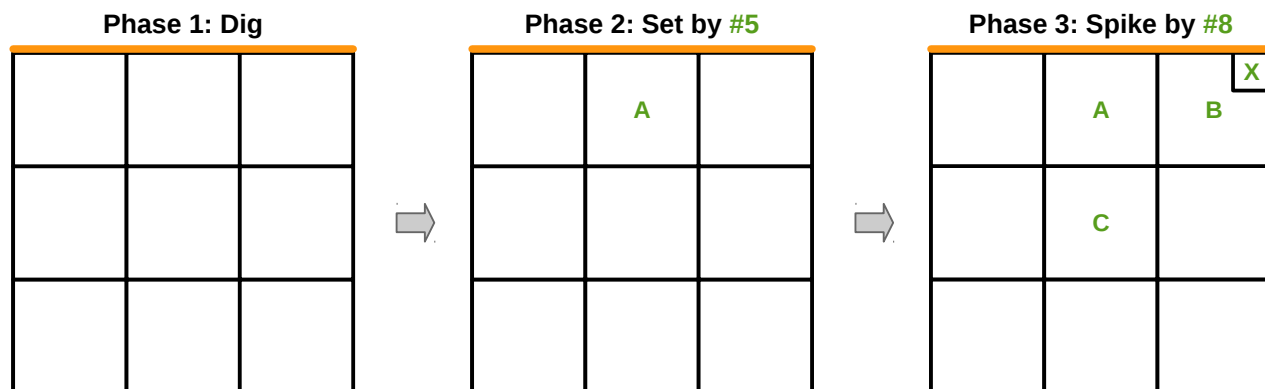


Figure 6: Frequent offensive pattern by Poland that Brazil does not frequently employ in the men's final. Player with jersey number 5 sets the ball from the front center to the front right where it is spiked by player number 10. X denotes the location of the ball.

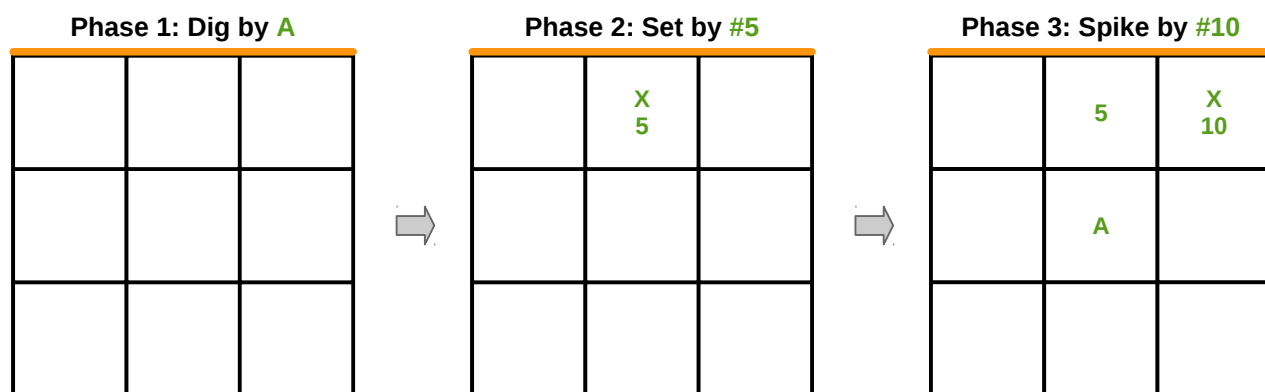


Figure 7: Frequent offensive pattern by Brazil that Poland does not frequently use in the men's final. The player with jersey number 1 is located in the front center at the time of the dig. He sets the ball from the front center to the front left. Player 10 moved from the middle left zone at the time of the dig to the front left zone, where he spiked the ball.

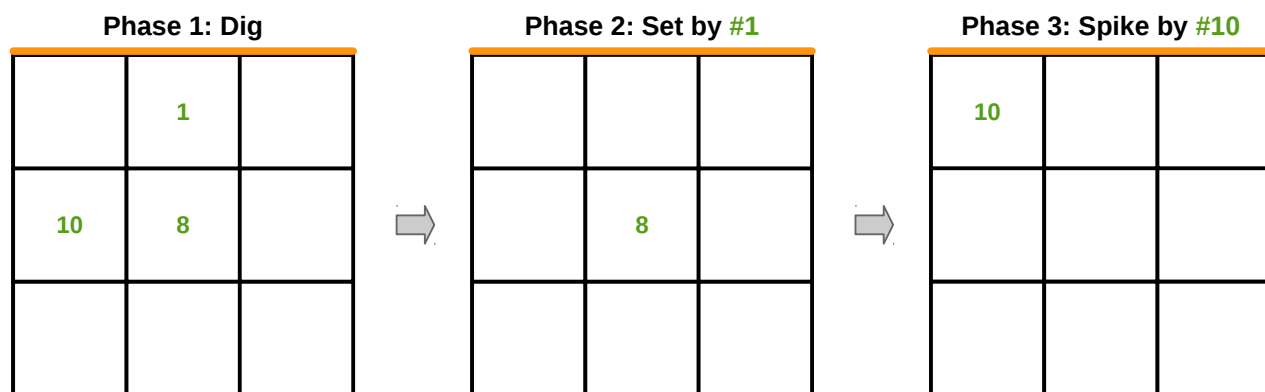


Figure 8: Frequent offensive pattern by the USA not used by China in the women's final. The ball is set by the player with jersey number 1 in the front center zone. Afterwards, another player spikes the ball. X denotes the location of the ball.

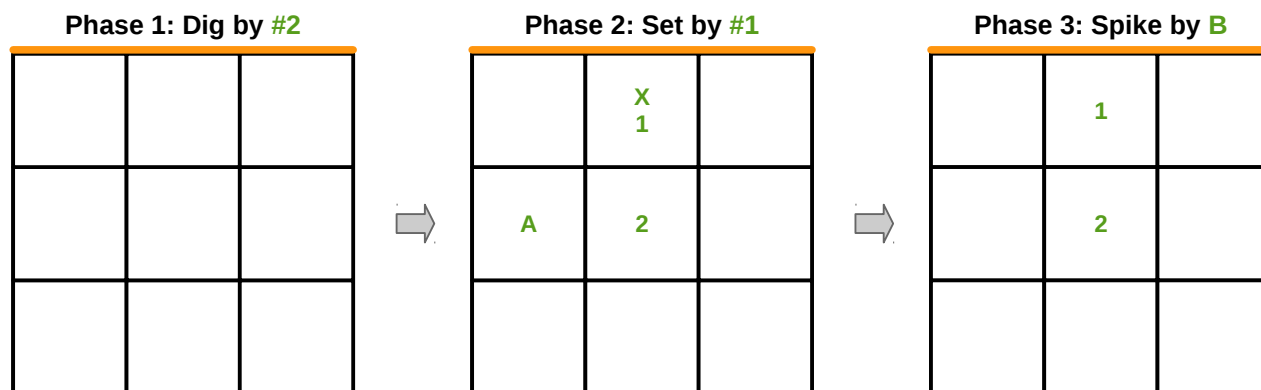


Figure 9: Frequent offensive pattern by China not used by the USA in the women's final. The ball is set by player 5 in the front center zone. Afterwards, player 1 spikes the ball. X denotes the location of the ball.

