Universiteit Leiden
The Netherlands

Opleiding Informatica

Determining Good Tactics for a Football Game using Raw Positional Data

Davey Verhoef

Supervisors:
Arno Knobbe
Rens Meerhoff

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

August 11, 2018

Abstract

This thesis is about trying to find a model which helps a football club to play more effective and this way improve their way of playing. There is a lot of money involved in the world of football. Many clubs spend a lot of money on new players. These new players do not always make these football clubs better. There are many examples of expensive players who did not succeed at their new club.

This thesis tries to find a way to improve the way a football club plays a football match without having to pay a lot of money on new players. The aim is to research the current team of a specific football club and find the best way of playing for this team with the players they have at the moment, instead of trying to find the best players and pay a lot of money to attract them.

This thesis is made up of two parts. The first part is about researching situations which happen a lot in a football match. There are a few choices defined per situation to look at. The goal is to find the best choice in these situations, so the football team has prove that a certain choice is better in a certain situation than other choices.

The second part is about calculating what the best positions are on the field for this team. If a team knows which positions on the field are apparently effective for them to be in, they can try to get the ball in those positions more often. Combining this with the first part, they can find the best choices in common situations to get into the better positions on the field for them. This way, we try to improve the way of playing for a certain football team, by analysing big data and proving more effective ways to play the game.

Contents

1 Introduction **1**

1.1 Thesis Overview . . . 3

2 Methodology **4**

2.1 Data . . . 4

2.2 Common game situations . . . 5

2.2.1 Goal kicks . . . 5

2.2.2 Throw ins . . . 6

2.2.3 Passes made by defenders . . . 7

2.2.4 T-test . . . 8

2.3 Assigning values to all positions on the field . . . 9

2.3.1 Gaussian function . . . 10

2.3.2 Calculating values, approach 1 . . . 11

2.3.3 Calculating values, approach 2 . . . 12

2.3.4 Heatmaps . . . 13

3 Experiments **15**

3.1 Common game situations . . . 15

3.1.1 Goal kicks . . . 15

3.1.2 Throw ins . . . 19

3.1.3 Passes defenders . . . 22

3.2 Assigning values to all positions on the field . . . 26

3.2.1 Approach 1 . . . 26

3.2.2 Approach 2 . . . 27

3.3 Difference approach 1 and 2 . . . 28

4 Conclusions **29**

4.1 Common game situations . . . 29

4.2 Assigning values to all positions on the field . . . 32

4.3 Overall conclusion . . . 33

5 Future Work 34

Bibliography 37

Chapter 1

Introduction

Football is the most popular sport in the world. It is not only the most watched sport in the world, but millions of Euros are involved in football each year. Last year, Paris Saint Germain bought Neymar from FC Barcelona for more than 222 million Euros. This kind of money is paid for just one player.

Why does a football club pay this kind of money for just one player? They have to win this money back somehow. Football clubs get a lot of money from sponsors. The better a football club performs, and the better (and more popular) players they have in their team, the more money they get from sponsors. This is one (important) way to get money. Another way is to win games, and to win important tournaments. If they win the UEFA Champions League for example, they can earn up to 100 million Euros with this.

This explains the amount of money clubs pay for a player. But how do clubs decide which players to buy? All clubs employ scouts, who travel all around the world to watch a lot of matches to identify top-level players. One strategy to get new players for less money, is to have scouts in your organization who are able to identify young talented players who can be bought for relatively little money.

This is one way to get a good team, to buy good players. In this thesis, we look at another way to try and win matches. We do not look for good players who can be bought, but we try to find the best choices a specific team should make in certain situations, and we tried to find the positions on the field which are best for this team.

There are certain statistics which are available for every match played, such as the percentage of ball possession, the number of shots per team, and the number of passes made. This tells us to a certain extent which team played better, but this is not always true. In many situations, the team which had the most shots, the most passes, and the most possession, was the better team. But there are also teams that focus less on having a lot of shots, passes or possession.

One great example of this is a game played in the UEFA Champions League on 23 October 2012, between Celtic and FC Barcelona. FC Barcelona has a higher budget than Celtic has, so they can pay more money to attract new players, or pay more salary to keep players at their club. Overall, the expectation is that a team

with a higher budget has better players in their team than a team with a lower budget. As a result of this, they are expected to win the game when they play against each other, but this is not always case, see Figure 1.1.

Figure 1.1: Match statistics Celtic - Barcelona, 23 October 2012

What we see here is that Celtic almost did not have the ball, they had very few passes, and they only shot 5 times. FC Barcelona however had a lot of ball possession, a lot of passes, and a lot of shots. But in the end, Celtic won the game. This is a great example of a football match where the better team with the better players and far better statistics in the match still lost. Is this because Celtic did something really well, or because they were just very lucky?

In this thesis, we analyse raw positional data to find conclusions to help football teams play more effectively. What we have done is distinguish certain choices in common game situations, and determine which choice is the best choice for the researched team to make in these situations.

After this, we calculated a value for every position in the field. With this values, we can make a graphical representation of the field with all values. This helps us to quickly see which positions on the field are more favourable to be in for this team. When this team is in possession of the ball in one of these positions, they have a better chance to score a goal.

By combining these two parts, we have positions on the field which are more favourable for this team to be in, and we can determine certain choices to make in certain common game situations to get in to these positions. This way we can improve the way of playing a football match for a certain team without buying new players for a lot of money, but just with the players which are available for this team.

This is an approach which is gives us specific results for the researched team, but it is not developed for just one team. It can be used for every football team, as it uses the data which is measured for that team to

calculate all the results. What we deliver is a standard approach that finds ways to improve the way of playing of a football team.

1.1 Thesis Overview

This thesis, which is written as a bachelor thesis with Arno Knobbe as supervisor and Rens Meerhoff as second supervisor for the study Computer Science followed at LIACS, is made up of five chapters. This chapter contains the introduction; Chapter 2 contains the methodology of our experiments; Chapter 3 includes the experiments which we have done with their results; Chapter 4 contains a summary of the conclusions which we discussed more extended in Chapter 3; Chapter 5 discusses what can be done to extend this research.

Chapter 2

Methodology

As mentioned before, this thesis consists of two parts. The first part is about finding the best choice players could make in a certain situation. The second part is about assigning a value to every position on the field. In this chapter, we will describe the methodology used to find these choices and these values.

2.1 Data

The data that was available for this research is raw positional data of 9 matches played by PSV in the Dutch Eredivisie. The x- and y-coordinate of the players and the ball were measured 10 times per second. These coordinates are expressed as values between -1 and 1. The system that was used to measure these coordinates was not able to measure height. This means that the data we have is 2-dimensional data, so when the ball was in the air, this system mapped the ball to a 2-dimensional coordinate which sometimes results in measurement errors in the data we have. Sometimes the data tells us the ball went from one side of the field all the way to the other side in less than a second. Also, the system was not always able to track the ball, or (one or more of) the players, which results in a coordinate of (-10, -10). If we were looking at one of the players or the ball in a frame where they could not be measured, we ignored this frame.

We also have the event data of these matches. This is a database which mentions all important actions in the game, such as *goal kicks, passes, goals.* These actions are tracked by humans, and not by a computer system like the measured raw positional data. The consequence of this is that it is not very precise. There is always a chance that the moment the action happened is not exactly the moment the database tells us, because it is tracked by the human eye. This is the reason that we use the raw positional data in this research and filter the moments out with the help of this data. The event data is only sometimes used as guideline for when a certain action approximately occurred. We confirm this timing by calculating the exact moment with the help of the raw positional data.

2.2 Common game situations

We looked into three different classes of common situations in a football game, and determined possible choices the player could make in that situation. For each of these situations, we looked at different things to determine if the choice made was a good or a bad choice. The situations we looked at are goal kicks, throw ins, and passes made by defenders.

The reason to look at these situations is to see if there is a certain choice in one of these situations which is better than the other choices. We research these choices to find a choice in a particular situation in the game which leads to a better situation in the game than when one of the other available choices was made. This way, we find a way to make football players more efficient, by showing them that a particular choice in a particular situation is the best choice to make. By providing players and trainers with this information, we hope to give them insight in how to improve their game with possibly a small simple change in their play. We are not trying to come up with a whole new complicated tactic, but we improve the way a football team plays by giving them simple instructions to do this. An example of such an improvement would be to always try to pass the ball in a certain direction. This is because the data tells us that when playing the ball in that direction, it leads to an improvement of the game situation, or it leads to a better situation than when the ball was passed in another direction.

We chose goal kicks, throw ins, and passes made by defenders as the situations to look at, because these are common situations in a football game. There are a lot of these situations every game, so if we can improve one or more of these situations, we automatically improve a great part of a match. Some people are convinced that a player should always try to find a way to get the ball forward, as this is the way to the opponent goal. But maybe this is not a faster way to the goal when a player is in one of the situations we look at. It is possible that it is more efficient to first pass the ball back before passing the ball forward, or maybe it is more efficient to take a short goal kick instead of kicking the ball far away to get to the goal sooner. These are the things we are trying to find out.

2.2.1 Goal kicks

When looking at goal kicks, we distinguished between two choices when taking the goal kick: taking it short and taking it long. When a goal kick is taken long, the distance of the ball to the opponent goal gets smaller quickly in little time. This opposed to taking the goal kick short, when the ball stays closer to their own goal. To put it very simple, one could say it is better to kick the ball long, because it gets closer to the opponent goal sooner. In this research, we look if this is true. We also looked at the number of players behind the ball, both of their own team as the players of the opponent team, after a certain period of time after kicking the ball.

First of all, we need to know when all the goal kicks took place. We used the event data for this. The event data contains a definition of the action which took place at a certain time. For goal kicks, the definition of the action which took place is simply set as *goal kick*.

By looking at the line up of each of the games we have the measured data from, we know that *Jeroen Zoet* was the goal keeper of PSV in all of the games played. So we specified our search for all the goal kicks by adding the condition of the player who took the goal kick, which needed to be *Jeroen Zoet*, because we will be researching PSV in this thesis.

Now we had determined all the goal kicks taken by PSV. We found that the amount of goal kicks we found was not enough, so we decided to broaden the concept *goal kick*. Instead of only looking at goal kicks taken by the keeper of PSV, we looked at every goal kick, free kick, or pass made by the keeper. This gave us a lot more data to work with, which will result in more reliable results.

The next thing we need to do is to specify when a goal kick is taken *long* and when it is taken *short*. Also, we need to specify when we call a situation in the game better than another situation. We call a goal kick short, if the ball is within 25 meters of the goal after 2.5 seconds. We measured this by following the ball 2.5 seconds after the keeper took the goal kick, and looked at if the ball was within 25 meters of the goal the whole time. If this was the case, we say the goal kick was taken short. Every goal kick where the ball was further than 25 meters away within 2.5 seconds, we specified as a long taken goal kick.

The goal of football is to score goals, and to do this, the ball needs to get close to the opponent goal and eventually go in. This is why we looked at the distance of the ball to the opponent goal, because we are looking to get the ball as close as possible to the opponent goal. Intuitively, one would probably say that it is better to take a long goal kick if you want the ball to be close to the opponent goal quick. This is true, but there is also a chance that the ball gets back real quick because you effectively pass the ball to the opponent. What we are looking for is the distance of the ball to the opponent goal, in a time period of a hundred seconds after taking the goal kicks. This way we can see if the ball stays close to the opponent goal, or gets back quickly.

The other thing we look at is the number of players of the opponent which are passed. It is better for the attacking team if there are fewer opponents in front of them, because this means there are less obstacles in their way to the goal. For this measurement, it is also true that you can pass many opponent players quickly by kicking the ball long, but again we want to see how this holds over time. So for this measurement we also look at the first hundred seconds after the goal keeper kicked the ball.

2.2.2 Throw ins

For the throw ins, we looked at two different possibilities, throwing the ball forward and throwing the ball backward. Again, intuitively one might say it is more efficient to try and throw the ball forward, as it brings the ball closer to the opponent goal. We looked at throw ins at least 20 meters away from the back line, and researched the difference in the absolute distance of the ball to the opponent goal, and the relative distance. We look at both the absolute and the relative distance, because when a throw in is rewarded on the opponent half of the field, the absolute ball distance to the opponent goal is already smaller than when the throw in is on the teams own half of the field.

Just like with the goal kicks, we needed to know when the throw ins occurred. It is easier to get all the throw ins than to get all the goal kicks. We needed to check fewer situations for the throw ins than for the goal kicks. What we needed to check is that the ball is out of play at the side line, and is in play again a frame later. This gave us all the throw ins that occurred in the game, but we did not want to use all of them. We wanted to use only the throw in made by PSV, and only the throw ins in a certain range of the field.

The throw ins were filtered on the team which was in possession. These throw ins were filtered again, and we took only the throw ins which occurred further than 20 meters away from the back line. We did this because we wanted throw ins where there is a good possibility to either throw it forward or backward. If the ball is too close to the backline, the player will never make the choice to throw the ball backward, as he would then throw it over the back line.

We also needed to make sure that it really was a throw in, and not a measurement error. Because the data is not really accurate as stated before, we did not manage to catch every measurement error, but we can catch quite a lot of them by checking the ball speed. What we have seen a lot in this data is that the ball seems to jump from one place to the other, which indicates a measurement error. To filter these errors out of the found throw ins, we checked if, in the period of a second before until a second after the throw in, the ball speed exceeded 100 kilometers per hour. If this is the case, we assume that this is a measurement error, and we do not look at this moment.

2.2.3 Passes made by defenders

The last situations we looked at, are the situations where a defender passes the ball. We determined three possible choices for defenders to make when passing the ball. They can pass the ball forward, wide, or backward. Instead of one comparison, we did three. For all three choices, we measured the difference between one of these choices and the other two choices together. This gives us three sets of results.

To determine all the passes made by defenders, we first identified all defenders by looking at the table where all players with corresponding player id are stored. With this information, we can determine all moments where one of the defenders is in possession of the ball. For all these moments where a defender is in possession of the ball, we looked at when the distance of the ball to the defender in possession increases for 8 frames in a row. If this is the case, we specify this as a pass made by the defender.

Now that we determined all the passes made by defenders, we need to distinguish between the three choices they can make for making the pass: forward, wide, or backward. This is not so hard to do, as we just have to check the distance of the ball to the goal of the opponent, at the moment the pass was made, and one second after passing the ball. If this distance decreases by more than one meter, we classified it as a pass forward. If this distance increases by more than one meter, we classified it as a pass backward. All other situations are classified as a wide pass.

2.2.4 T-test

To statistically determine if any of the choices made is better than the other choices, we used a t-test. We used a two-sample location test of the null hypothesis, which is also known as the Student's t-test [Ste18].

The Student's t-test does not only tell us the difference between the means of the two datasets, it also tells us how significant this difference is. So the t-test not only tells us which of the two choices is better, it also tells us how big the chance is that the difference happened by chance.

This is expressed by a t-value and a p-value. The t-value expresses the ratio between the difference between the two groups and the difference within the groups. A larger t-value means that there is a greater difference between the groups, that the difference between the groups is more likely to be significant, and that the results are repeatable.

Every t-value comes with a p-value. The p-value tells us if the results occurred by chance. The greater the p-value, the more likely it is that the results occurred by chance. P-values are usually written as decimal, and can be interpreted as a percentage. So, a p-value of 0.05 can be seen as a 5% chance that the results happened by chance. A maximum p-value of 0.05 is usually used to determine if the results are valid.

In the three researched situations, we look at two or three possible choices players can make per situation. A t-test is done for all these situations. The data sets used for the t-test are described per situation. In each of these situations, we search for a high t-value with a corresponding low p-value. If we can find these values, we can determine a choice in this situation which is a significantly better choice than the other possible choices.

2.3 Assigning values to all positions on the field

In the previous section, we looked at common situations in a game (goal kicks, throw ins, passes made by defenders) and tried to come up with the best possible choice in such a situation. It is nice to know what the best possible choice is in these situations, but this is not all there is to a football game of course. What we did after researching the common situations, is to try and assign a value for every place in the field, which indicates how good it is for this football team to be in possession of the ball at this place in the field at any given time.

We did this by laying a grid over the field, so that the field is built up of small boxes which we can give a value. The size of the boxes we chose is one by one meter. The size of the fields played on is 105 by 68 meters, so this gives us 105 * 68 = 7 140 boxes. The boxes can have a value between 0 and 1, where 0 is a really bad situation, and 1 is the best possible situation.

Initially, all the boxes have the value of 0, except for certain specified moments where we give certain boxes the value of 1. At first, we wanted to give all the moments where a goal was scored the value 1. We only have the data of 9 matches, so this does not give us much goals. There are only 26 goals made by PSV in this matches. This is too little information to start the calculation with. Because of this, we needed to think of something else to have more activation points for our calculations, and still have reliable results.

For a team to score a goal, they first have to get close to the goal. The closer they get, the better chance they have to score a goal. This is why we chose to give every situation where the ball is within one meter of the goal the value 1. This increases the amount of measured frames that will activate the calculations from 26 to 23538. This gives us enough activation points to start the calculation with.

How do we calculate the value of the other boxes? We tried two approaches. Both approaches start with taking a frame out of one of the games randomly, and check in which box the ball is at that moment. Then, we take the next frame and do the same thing. The difference between the two approaches is the way of calculating the values for the positions.

In the first approach, we take the current value of the two boxes we just got. We determine the difference of these values, and calculate the distance between the two boxes. With this information, we update the values of the boxes using a gaussian function. An explanation of a gaussian function follows below. The second approach starts the same, but the amount for updating the values is calculated differently. A more extensive explanation for both approaches follows below.

In the beginning, all the frames where the ball is not within one meter of the goal of the opponent in this frame and the frame hereafter have the value 0. This means that the values of the boxes will not be updated, as there is no difference in value in these frames. This is why we need a certain amount of frames where the value of the box where the ball is in is 1, to start the change in values for the boxes. If we would have chosen only the goals as the frames where the value is 1, there are only 26 possible moments initially where the values of other boxes are changed. To have all the boxes updated enough to have a reliable value in all boxes, we

would have to loop through all the data for a very long time, because it takes a long time to have the values distributed throughout the field. By initially setting the value to 1 for every situation where the ball is within one meter of the goal, the activation of the distribution of the value throughout the field starts quicker.

2.3.1 Gaussian function

We use the Gaussian function to distribute the change of the values over the field. We do this so that not only the box where the ball is in at the frame we are looking at is updated, but so that the whole field is updated. Each box has a distance to the box where the ball is in at the frame we are looking at, and this distance corresponds to a Gaussian value. We calculated the difference in value between the frame we are looking at and the frame hereafter, and for each box we multiply this difference by the value of the Gaussian for this box. This gives us the value we need to use to update the value the box had up to now. The Gaussian value depends on the distance between the two boxes, the Gaussian value gets smaller when the distance between the boxes is bigger. See Figure 2.1 for a graph of a Gaussian.

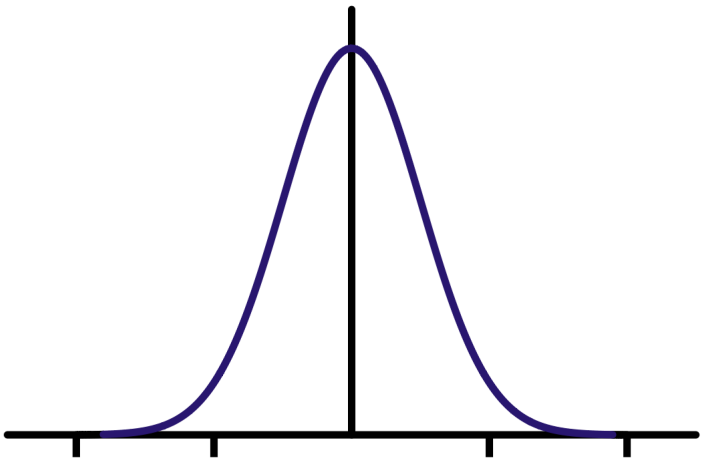

Figure 2.1: The graph of a Gaussian is bell shaped

The formula of a Gaussian is:

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$$

where a is the height of the curves peak, b is the position of the centre of the peak, c controls the width of the bell (the standard deviation), and x is the distance between the boxes.

The surface under the graph has to be 1. To achieve this, the height of the graph has to be calculated by the following formula:

$$a = \frac{1}{c\sqrt{2\pi}}$$

The position of the centre of the graph is always at 0 in our experiment, because the box we are looking at should have the highest change value. So the value of b is always 0.

A good value for c, the standard deviation, was harder to find. We saw that when taking a too small value for c, we needed to loop over the same data a lot to get the values distributed over the whole field. When the value for c was too big, the distribution went too far over the field too quick, resulting in very high values all over the field. We found that a value of 5 was a good fit for the dataset we have. If we fill in the formula for the height of the graph, we get the value of a:

$$a = \frac{1}{5\sqrt{2\pi}} \approx 0.0797884561$$

2.3.2 Calculating values, approach 1

The values are updated by randomly selecting a frame out of one of the games. We will call this *frame 1* in the rest of this text. When we have such a frame, we first check if the ball is in possession of the team we are analysing. If it is not, we take another random frame. If the ball is in possession of the team we are analysing, we take the value of the position where the ball is located at in this frame, and the value of the position where the ball is located at in the next frame. This next frame will be called *frame 2* in the rest of the text. The difference between these two values (value *frame 2* – value *frame 1*) is the value we are going to work with.

Now that we have this value, we need to do something with the Gaussian. The last variable that needs to be filled in the Gaussian function is the distance. This distance is the distance between the first point we looked at (the position of the ball in *frame 1*), and all the other positions on the field. So for every box in the field (we divided the field in boxes, as explained above), the distance between this box and the box where the ball is located at in *frame 1* is calculated.

With this distance, a Gaussian value can be calculated for every position in the field by filling in the distance in our Gaussian function. We also have the difference between the values of the positions in *frame 1* and *frame 2*. By multiplying these two values (Gaussian value and difference of the values), we get the amount to update the current value of this box with. This updating is done by simply adding this value to the value which this box currently has. This whole updating ritual is done until all the frames out of all the games have been processed.

Note that if the ball comes from a position that has a higher value and goes to a position that has a lower value, the calculated difference of these positions is negative. This means that the entire field is updated with a negative value, so the values of the boxes will decrease.

It might seem that it is not right to update the whole field, because there are positions on the field which are so far away from the position where the ball was at that moment, that they have nothing to do with this. This is true, and this is where the Gaussian value is for. The Gaussian value decreases as the distance increases, and eventually goes to 0. This means that when the distance is big enough, the Gaussian value will be 0. The result of this is that the value with which the value of this box is updated is also 0, because there is a multiplication with 0. This makes sure that there is not too much updated in the field, and the results remain reliable.

2.3.3 Calculating values, approach 2

The calculation in the first approach assumes that it is always better to be closer to the goal, as the values are updated positively when the ball moves in the direction of the goal. This is not totally true, because there are a lot more variables to take into account when playing a football game. This is why we did other calculations after we did this one, taking more variables into account.

Instead of only looking at the position of the ball and which team is in possession, we looked at more things, and in another way. Instead of only calculate a new value if the team is in possession, we also look at the direction the ball travels. We will calculate a new value if the team we are looking at is in possession, or if the ball travels in the direction of the opponent goal. We do this because if the team we are looking at is not in possession, but the ball did travel in the right direction, this is probably because the opponent team were not able to play the ball forward. If this is the case, the team we are looking at is doing a good job, because without actually having the ball, they get the ball closer to the goal of the opponent by their play without the ball.

It is not always necessary to be in possession of the ball to get it into a better position in front of the opponent goal. If the team puts on good pressure on the ball when the opponent is in possession of the ball, they might be able to force the opponent to play the ball back in the direction of their own goal. If this is the case and they manage to win the ball, they are immediately in a good position close to the opponent goal.

There are two more variables we looked at. The first one is the number of players of the own team behind the ball, where behind the ball means that the distance of these players to the goal of the opponent is greater than the distance of the ball to the goal of the opponent. We decided that for our calculations it is better for a team to have few players behind the ball. When there are few players of the own team behind the ball, this automatically means that they are between the ball and the goal of the opponent. This gives the player in possession of the ball more options to pass the ball to another player who is closer to the goal than the player in possession is, and we said that closer to the goal is better as the goal of playing football is to score goals.

The other variable which is used in this new calculation is the number of players of the opponent team behind the ball. This is also seen from the perspective of the attacking team. What they want is to have as much players of the opponent team as possible behind the ball, as this gives them more space to attack. The more opponent players behind the ball, the more opponent players they passed on the field, the easier it gets to score a goal.

With these new variables, we distinguish four situations. First we look at the team in possession and the direction the ball is travelling. It is possible that the team we are researching is not in possession of the ball, and the ball is travelling in the wrong direction. If this is the case, we do nothing.

It is also possible that only one of the two is true, so either the right team is in possession and the ball is travelling in the wrong direction, or the wrong team is in possession and the ball is travelling in the right direction. If this is the case, we calculate a temporal value by the following formula:

$$tempValue = min(0.7, \frac{1}{22} * valuableTeamMates + \frac{1}{22} * passedOpponents)$$

where

$$valuableTeamMates = 11 - numberOfOwnPlayersBehindTheBall$$

As explained above, it is better for the attacking team to have less players behind the ball, so for this calculation we start with the value 11 (the number of players in a football team) and subtract the number of players behind the ball. This way we get a bigger value as there are less players behind the ball. It works the other way around for the number of opponent players behind the ball, so we just take the number of opponent players behind the ball as the value for this.

The last possible situation is where the team is in possession of the ball and the ball is travelling in the right direction. If this is the case, we have almost the same formula as before, but with one addition:

$$tempValue = min(0.9, \frac{1}{22} * valuableTeamMates + \frac{1}{22} * passedOpponents + 0.25)$$

Because there are two positive things going on here, the ball in possession of the right team, and the ball travelling in the right direction, this needs to have a higher value. This is why it has the same formula, but with 0.25 added to the value, and a new higher maximum of 0.9.

Now that we have calculated a tempValue, we need to use it somehow. In the first implementation of calculating the values for the game, we took the difference of the two values (the value of the position where the ball is in *frame 1* and in *frame2*) as the value we were updating the rest of the field with. With the new calculation where we use multiple variables, we use the tempValue to determine the value with which we are going to update the field. The new formula to calculate this value:

$$value = (1.0 - tempValue) * valueOfFrame1 + tempValue * valueOfFrame2$$

The result of this is that tempValue handles how much the value of *frame 1* and how much the value of *frame 2* is used to update all values. This results in a higher usage of the value of *frame 2* when this is a good position, so in a higher update value. This way the field is updated based on multiple variables, instead of just looking at where the ball is and update the values based on this position.

2.3.4 Heatmaps

The methods described above results in a value between 0 and 1 for all positions on the field. It is hard to conclude something when only looking at the results this way. This is why we wanted a graphical

representation of the information. We chose to display this information with a heatmap.

A heatmap is a graphical representation of data. Basically, a heatmap translate every value to a colour. When we make a graph with the x-axis from 0 to 105, and the y-axis from 0 to 68, we display the whole football field. This way we have a graphical representation of our calculated values, where it is easy to identify good and bad positions.

Chapter 3

Experiments

3.1 Common game situations

Our research question for this part is the following:

Based on measured data of multiple football matches, can we determine if a certain choice in a common situation is better than other possible choices in the same situation?

We will answer this research question with the help of three subquestions per situation, starting with the goal kicks.

3.1.1 Goal kicks

Based on measured data of multiple football matches, can we determine if taking a goal kick long results in a better situation in the game than taking a goal kick short?

Below are graphs for both the mean of the measured data and the results of the performed t-test. In the graph with the mean of the data, it is easy to see what the difference is between the distance of the ball to the opponent goal for long goal kicks and short goal kicks. As explained above, not only the mean of the data is important for us to get a conclusion out of the results, but also the value of the measured data. The value of the measured data can be seen in the graph with the t-value and p-value of this data.

In Figure 3.1, we see the means of the measured ball distance for long goal kicks and short goal kicks. The y-axis shows the distance of the ball to the opponent goal in meters. The x-axis shows the number of frames. We know that there are 10 frames measured per second, so the figure shows 105 seconds. We got the goal kicks by looking at the event data. Because the event data is not always totally accurate, we started by looking 5 seconds before the goal keeper took the goal kick according to this event data. By doing this, we can be sure we always capture the moment when the ball was kicked.

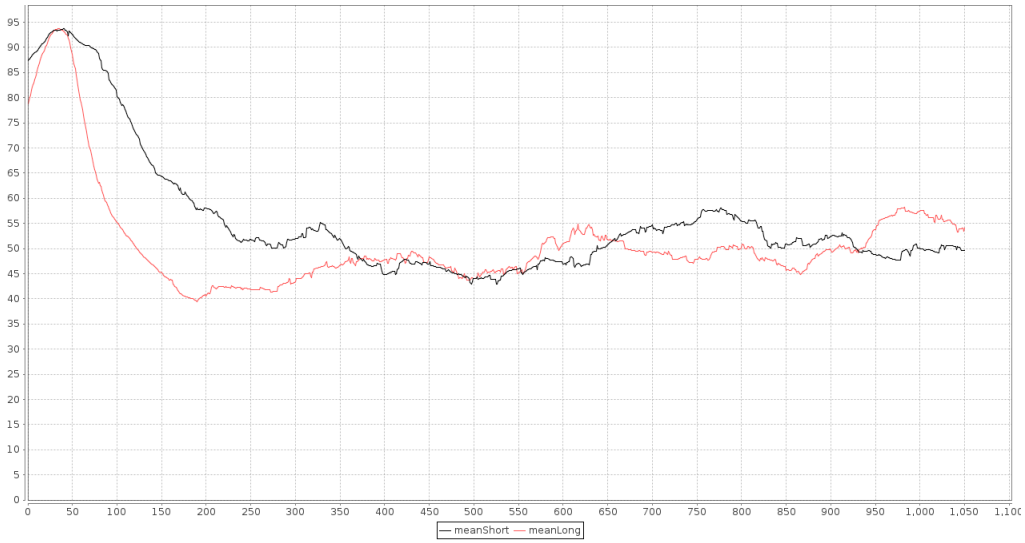

Figure 3.1: Mean ball distance to opponent goal after goal kicks

We will not look at the first 2.5 seconds after taking the goal kicks, because we used these 2.5 seconds to distinguish between a long and a short goal kick. The following seconds is what we are really looking at. We see that after about 37 seconds in the graph the distance of the ball to the opponent goal is almost the same for both long and short goal kicks. This lasts until 56 seconds in the graph, from where on short goal kicks have a small advantage for 10 seconds over long goal kicks. Then after 66 seconds in the graph, long goal kicks take over the advantage. This period lasts a little longer, as the long goal kicks have an advantage over short goal kicks for 32 seconds. After this 32 seconds, when we are at 94 seconds in the graph, short goal kicks take over again.

What we see in this graph is that there is a small period of time that the short goal kicks have a small advantage over long goal kicks, but that it switches after this period. Then for a longer period, the long goal kicks have a little bit bigger advantage over short goal kicks. The last switch occurs after 94 seconds into the graph, which is about one and a half minute after taking the goal kick. We will not take this into account, because we think that after one and a half minute, the goal kick does not have any influence any more on where the ball is.

As stated before, we do not only want to see the means of the data, but we also want to know how valuable the data is. This is where the t-value and p-value come in, which we can see in Figure 3.2.

What we see here is that the t-value with the corresponding p-value in the first 30 seconds in the graph show a significant difference between the long and short goal kicks. This is exactly what we expect, as we use this first period of time to distuinguish between long and short goal kicks. In the small period between about 59 seconds and 64 seconds in the graph, there is an advantage for short goal kicks. However, if we look at the p-value in this period, we see that this value is quite high which tells us that this difference is not significant.

Figure 3.2: t-value and p-value ball distance to opponent goal after goal kicks

After 64 seconds in the graph, the advantage for long goal kicks take over. This lasts for half a minute. In the period 64 seconds after taking the goal kicks and before 73 seconds after taking the goal kick, the p-value is above 0.05 which means the difference is not significant. Between 73 and 78 seconds in the graph, the p-value is below 0.05. The advantage for long goal kicks last until 95 seconds after taking the goal kicks. There is a period of about 30 seconds where there is an advantage for long goal kicks. In this period, there is a smaller period of 5 seconds where the p-value is below 0.05. We find that this period is too small to say that there is a significant advantage for long goal kicks.

This means that after about 70 seconds after taking a goal kick, there is a significant difference between taking this goal kick long or short, with an advantage for taking the goal kick long.

Overall, there is no long period of time with a significant difference between long and short goal kicks. There is this small period of about 5 seconds where there is a significant difference between the different goal kicks. As this period is part of a bigger period of time where there is an advantage for long goal kicks over short goal kicks, we might say that it is better to take a long goal kick instead of a short goal kick.

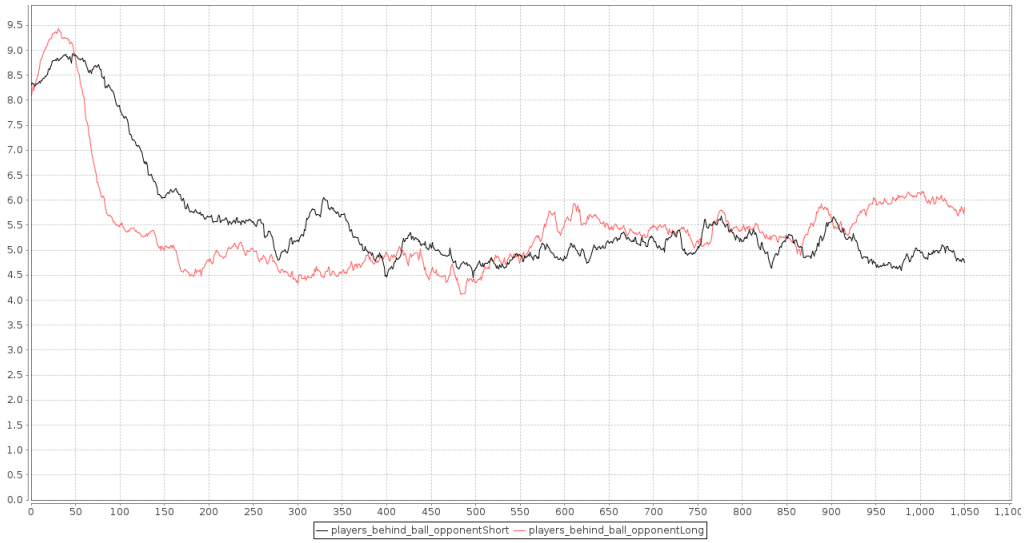

Figure 3.3: Number of opponent players passed after goal kicks

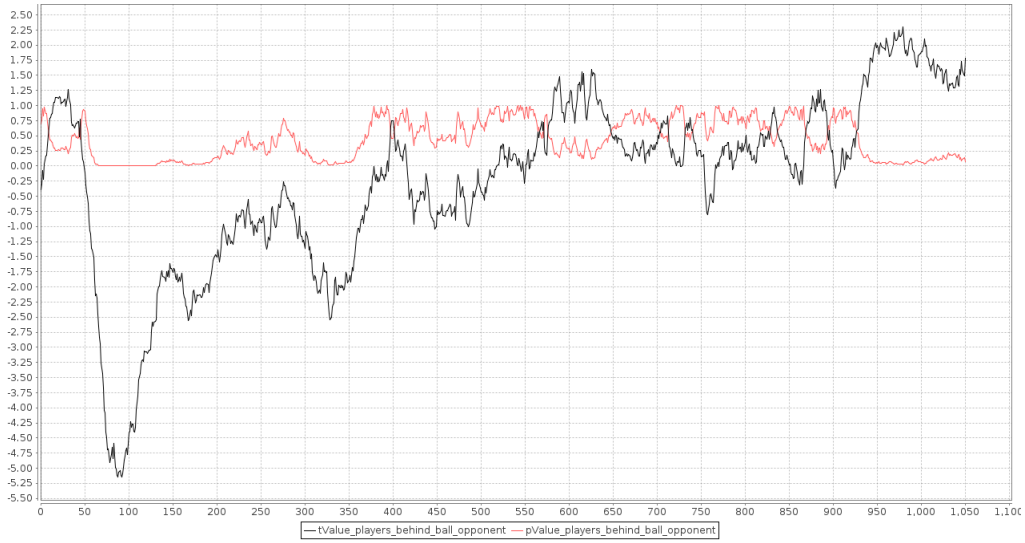

Figure 3.4: T-value and p-value of opponent players passed after goal kicks

We also looked at the number of opponent players behind the ball in the same period of time after the goal kicks. As the distance of the ball to the goal determines how good or how bad of a situation the attacking team is in, also the number of opponent players passed is important. The mean of the number of players who are between the ball and the goal is shown in Figure 3.3. The attacking team wants this number to be as small as possible, because less players between the ball and the goal means that they have passed more players, so they have less obstacles between them and the opponent goal.

What we see in this graph is that the number of opponent players between the ball and the goal is quite similar after taking a long or a short goal kick. After one and a half minute, there is an advantage for short goal kicks. Figure 3.4 confirms this. This graph shows high p-values up until 95 seconds in the graph. After this time, the p-value is close to 0 and there is a clear advantage for the short goal kicks. This is the same as we saw in Figure 3.2, where the t-test also showed an advantage for the short goal kick after one and a half minute.

If we combine the results of these two graphs, we can say that there is a slight advantage for long goal kicks. We can say this because there is a slight advantage for long goal kicks in the ball distance to the opponent goal one minute after taking the goal kicks, for a period of about half a minute. In this period of half a minute, there are 5 seconds where the p-value tells us that the difference is significant. The difference in the rest of this period is not really significant and could be a result of chance. In the same period of time, there is no advantage for one or the other if we look at the number of opponent players passed. For both the distance of the ball to the goal as for the number of opponent players passed, there is a slight advantage for short goal kicks one and a half minute after taking the goal kick.

Because there is a slight advantage for long goal kicks in terms of ball distance to the goal one minute after taking the goal kicks, we could say that if we have to choose between these two options, we would choose taking goal kicks long. There is no clear advantage for one of the other, but just a slight advantage, so we would only recommend this if we really have to choose between one or the other.

3.1.2 Throw ins

Based on measured data of multiple football matches, can we determine if throwing in the ball forward results in a better situation in the game than throwing in the ball backward?

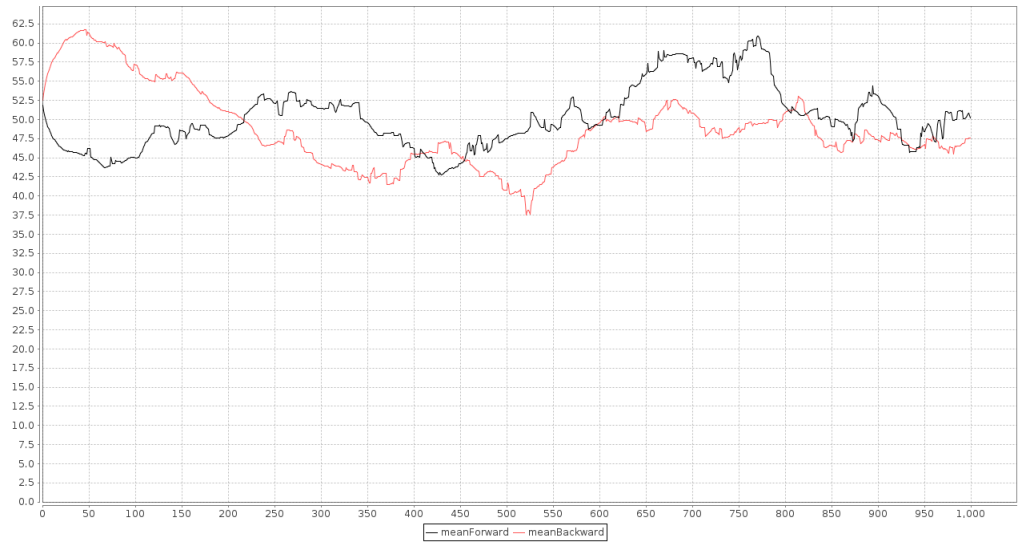

Figure 3.5: Mean ball distance to opponent goal after throw ins

We can see in Figure 3.5 that the distance to the opponent goal quickly changes in favour of throwing the ball backward, but also changes back again quickly. What we see in this graph with the means of the ball distance to the opponent goal, is that the distance in favor of one or the other goes back and forth. Overall it seems like there is a slight advantage for throwin the ball backward, but we can only say this after looking at the results of the t-test.

Figure 3.6: t-value and p-value ball distance to opponent goal after throw ins

Looking at Figure 3.6, we can see that the advantage for one or the other goes back and forth, but we have seen this already in Figure 3.5. The p-value in this graph is close to 0 at two points. Only around frame 525 and around frame 770. But it is only a very small time that the p-value is this low, it is not a period of time. This means that throwing the ball backwards has a significant advantage over throwing the ball forward at two exact moments in the game, and not periods in the game. Because of this, we can not say that it is significantly better to throw the ball backward.

This graph shows the distance of the ball to goal of the opponent, and it shows us that there is a slight advantage for throwing the ball backward when looking at this distance. The reason for this can be that it is just better for this team to throw the ball backward. It can also mean that this team chooses to throw the ball backward more often when the throw in is closer to the opponent goal. This is why we also measured something else, the absolute distance the ball has traveled after the throw in. We did this because the place where the throw ins are done differ a lot in terms of distance of the ball to the opponent goal, so we think this is not a fair distance to measure.

Absolute distance

The absolute distance the ball traveled does not tell us the distance of the ball to the goal of the opponent, but it tells us how much closer (or further away) the ball got after the throw in. This still is not a very clear way to measure the progress the play made because there are many other elements in play to take into account. We think it is a better way to measure the progress than taking the distance of the ball to the goal of the opponent.

What we see in Figure 3.7 is that the advantage of throwing the ball forward quickly switches to throwing the ball backward. This happens after about 21 seconds already. Then there is a small period of time where the forward throw takes back advantage again, and after this small period of time the throw backward wins back advantage and keeps advantage for the rest of the measured period. We have seen before that the averages of

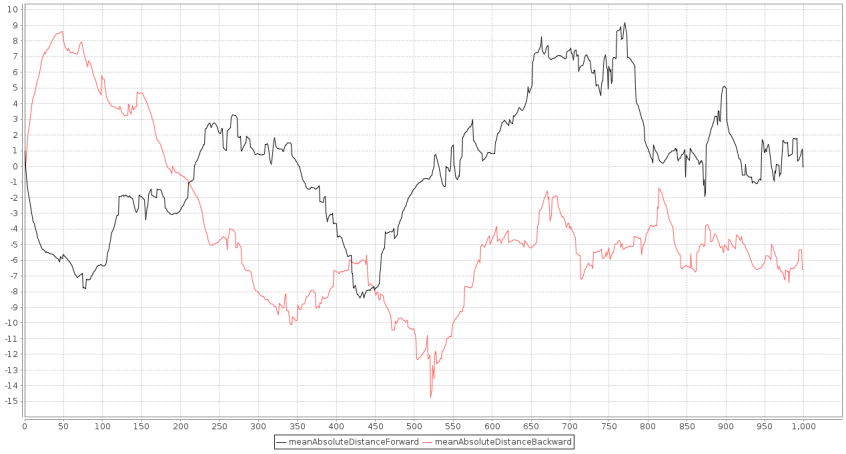

Figure 3.7: Mean of the absolute distance the ball traveled after throw ins

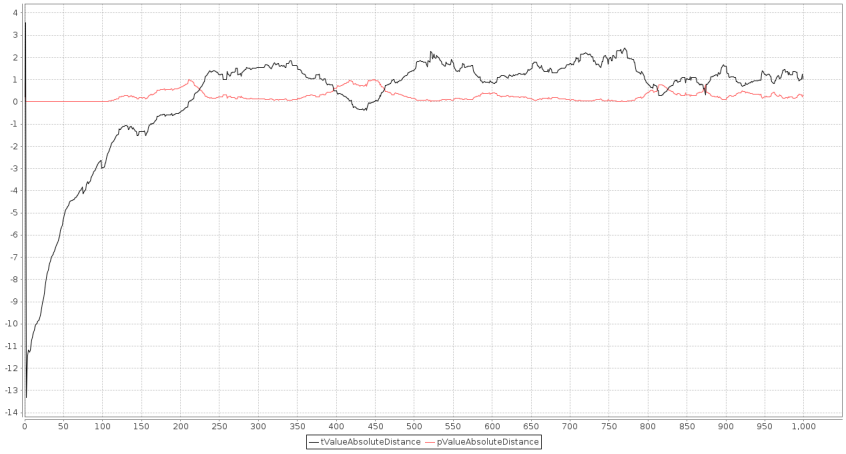

Figure 3.8: t-value and p-value of the absolute distance the ball traveled after throw ins

the measured data do not always show a significant difference. So we did a t-test again, and we can see the results in Figure 3.8.

For the first time, the t-test gives us significant results. This graph has a small p-value for a very long time, which indicates a significant difference between the two measured datasets. In the period between 23 seconds and 35 seconds after throwing the ball, there is a big difference with the advantage of throwing the ball backward, and the p-value in this period tells us that this difference is significant and not based on chance. The same is true for the period between 50 seconds and 80 seconds after throwing the ball. These are two, quite large, periods of time, with both a significant advantage for throwing the ball backward.

Because of this graph, we are able to conclude that there it is better, when this team has a throw in, to throw the ball backward than it is to throw the ball forward. By throwing the ball backward, this team has a much bigger change to cover more distance to the goal of the opponent, than it has when throwing the ball forward. There are of course more elements in the game to take into account, but if we do not look at these other elements (and we did not), it is better for this team to throw the ball backward than to throw the ball forward.

3.1.3 Passes defenders

Based on measured data of multiple football matches, can we determine if one of the three choices for a defender with the ball, passing the ball backward, passing the ball wide, or passing the ball forward, is better than the other two choices?

Passes backward vs passes wide and passes forward

The first t-test we will look at is the t-test for passing the ball backward versus passing the ball wide or forward. We start with Figure 3.9 where the mean distance to the opponent goal is shown. What we see is that in the beginning passing the ball backward has a disadvantage against passing the ball forward or wide. This disadvantage decreases quickly and 42 seconds after passing the ball, the disadvantage is gone. Then there is a period of about 15 seconds where the distance is amost equal for the two datasets. And the rest of the graph shows that the advantage shifts back and forth between the two datasets.

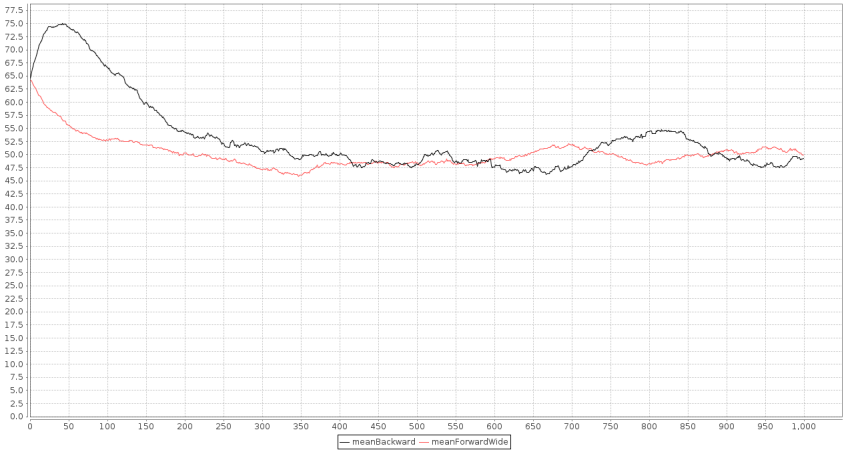

Figure 3.9: Mean distance of passes backward versus passes forward and wide

Figure 3.10: t-value and p-value of passes backward versus passes forward and wide

All we can say after looking at this graph is that there is an advantage in the first period of time after passing, for passes made wide or forward, over passes backward. This advantage decreases quickly, and even disappears after 42 seconds. To be able to say something more conclusive, we need to look at the results of the t-test in Figure 3.10. This graphs backs up what we saw in Figure 3.9. It shows us that in the first period after the passes are made, there is an advantage for passing forward or wide. Looking at the p-value, we see that this advantage is significant for a smaller period. The p-value is close to 0, only for the first 20 seconds. After this 20 seconds, there still is an advantage for passing forward or wide, but the p-value indicates that this advantage is not significant.

This situation does not tell us much. To be able to conclude more about the three possible choices, we need to take a look at the two other executed t-tests.

Passes wide vs passes backward and passes forward

The next t-test we look at, is the t-test where we compare passes wide with passes backward and passes forward. What we look at here is one dataset where the distance in the first period of time after passing the ball does not change much. The other dataset contains the passes where the distance does change a lot in the first period after passing the ball, but it changes in both ways. What we expect is that these changes in distance cancel each other out. This will likely result in two datasets which are very alike.

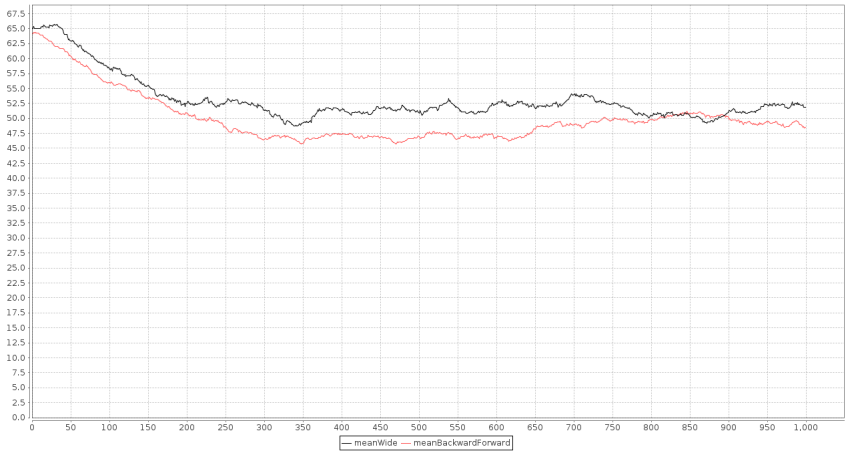

Figure 3.11: Mean distance of passes wide versus passes backward and forward

To see if our expectation is right, we begin by looking at Figure 3.11. This graph shows us that from the beginning to almost the end of the graph, there is an advantage for passing the ball forward or backward over passing the ball wide. The difference in distance of the two datasets is fairly small, so it looks like the difference is not really significant. To be sure of this, we look at Figure 3.12.

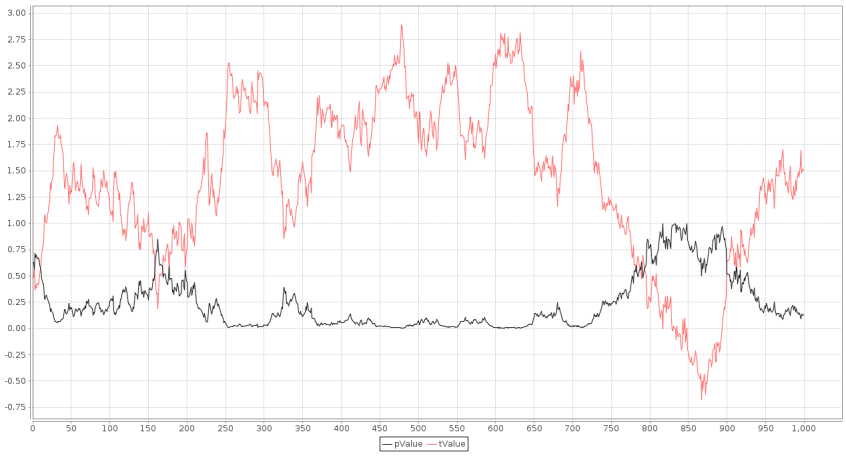

Figure 3.12: t-value and p-value of passes wide versus passes backward and forward

In this graph, we see that there is a relatively small difference between the two datasets. The most important thing we see in this graph is that the p-value is almost never close to 0. It happens a few times, but only for a very small period of time. This is too little time to say that there is a significant difference. All we can say after looking at the results of the t-test for these datasets, is that our expectation was correct, and that there is no significant difference for one of the two datasets.

After two t-test, we are not yet able to point out a choice that is significant better than the other two. All we know now is that there is a slight advantage in the first 20 seconds for passing forward or wide over passing backward. This is not much, so we need the last t-test to either show us that passing the ball forward has an advantage over passing the ball backward or wide, or to show us that there is no significant advantage for one of the three over the other two.

Passes forward vs passes backward and passes wide

The last t-test we will look at is where all passes forward are one dataset, and all passes backward and wide are the other dataset. The mean distance is shown in Figure 3.13, and the corresponding t-value and p-value are shown in Figure 3.14. We can see in Figure 3.13 that the distance to the opponent goal is smaller in almost the complete graph after passing the ball forward. We know now that only looking at the mean distances do not tell us everything, so we need to look at the t-value and p-value to determine if the difference is significant.

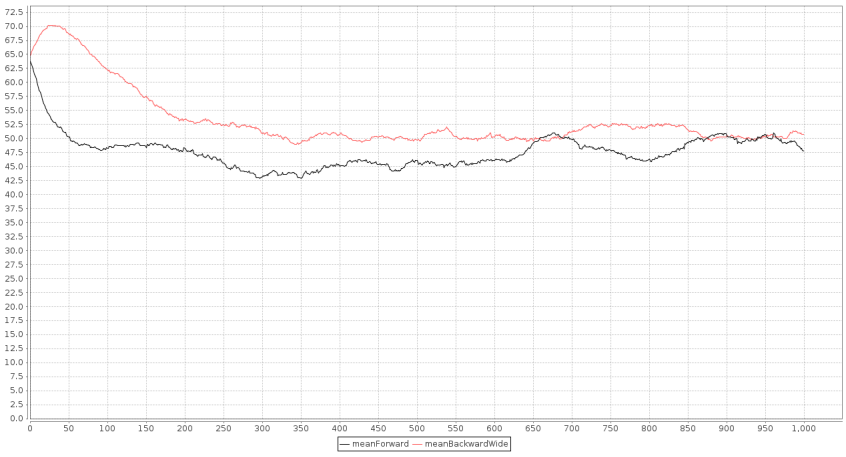

Figure 3.13: Mean distance of passes forward versus passes backward and wide

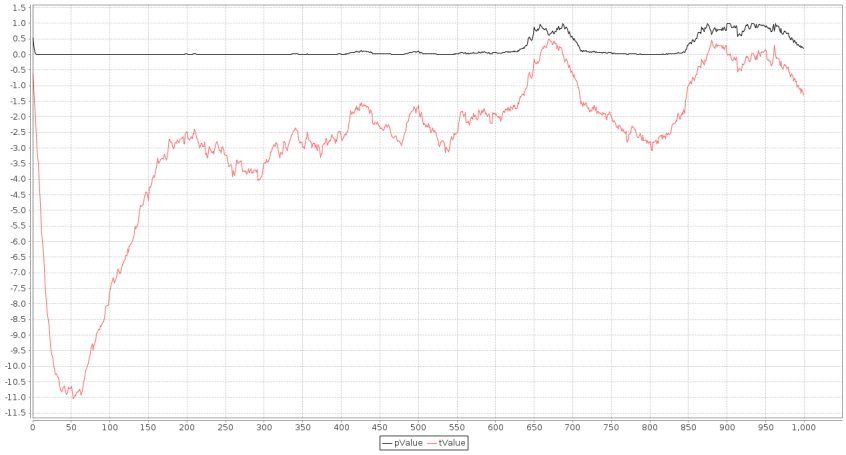

Figure 3.14: t-value and p-value of passes forward versus passes backward and wide

Looking at Figure 3.14, we can see that the p-value is very close to 0 almost all the time where the mean distance of forward passes has an advantage over the mean distance of backward and wide passes. This is a clear win for passing the ball forward. Combining this with the results of the previous two t-tests, where there was no significant advantage for one of the other two choices over the rest, we can safely say that it is better for a defender to pass the ball forward, than to pass the ball backward or wide.

3.2 Assigning values to all positions on the field

Based on measured data of multiple football matches, can we assign a value to every place in the field, to indicate how good or how bad it is for a football team when they are in possession of the ball at that place in the field at any given time in the game?

3.2.1 Approach 1

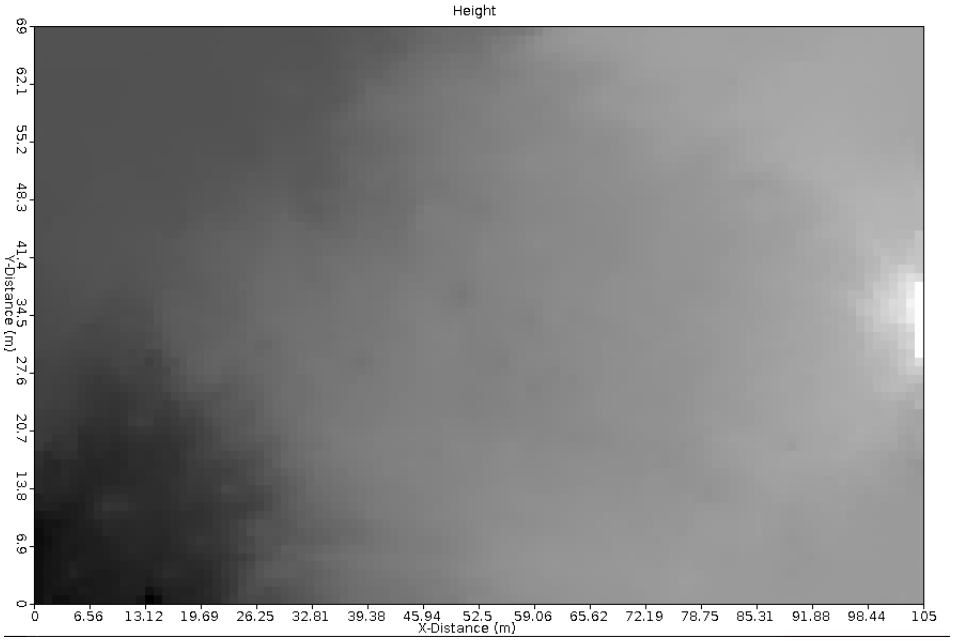

Figure 3.15: First heatmap

After calculating values for the whole field as described in Chapter 2, we got the heatmap shown in Figure 3.15. The range of colours in this heatmap is from black to white, where black is a bad position and white is a good position. This a heatmap with calculations over 9 games. In this heatmap, the team we looked at is playing from left to right. Of course they did not play from left to right all the time in all those games (actually they played 9 halves from left to right, and 9 halves from right to left), but when they played from right to left, we mirrored the positions.

What we see is that all positions within one meter of the opponents goal are coloured white. This is because, as described before, we gave all positions within one meter of the opponent goal the value 1 to be able to do this calculations. What we also see is that both left and right corners at the defence side are darker than the middle. The right corner is even darker than the left corner. This means that the ball was played to a better position more often through the left back side than through the right back side.

Both corners at the offence side of the field are lighter than the middle of this side of the field. There is not much difference between the left and right corner. This means that based on this calculations, it is better for this team to attack from the sides, than to attack through the middle. If we look close to the goal, we see a little more light spots at the left side of the goal than at the right side of the goal. So if they need to choose one of the two sides to attack from, it would be better to attack from the left side.

3.2.2 Approach 2

The result of the second approach is shown in Figure 3.16. There is a big difference between this heatmap and the heatmap in Figure 3.15. Where the first heatmap looked like some sort of cloud over the field, in this second heatmap a path can be seen.

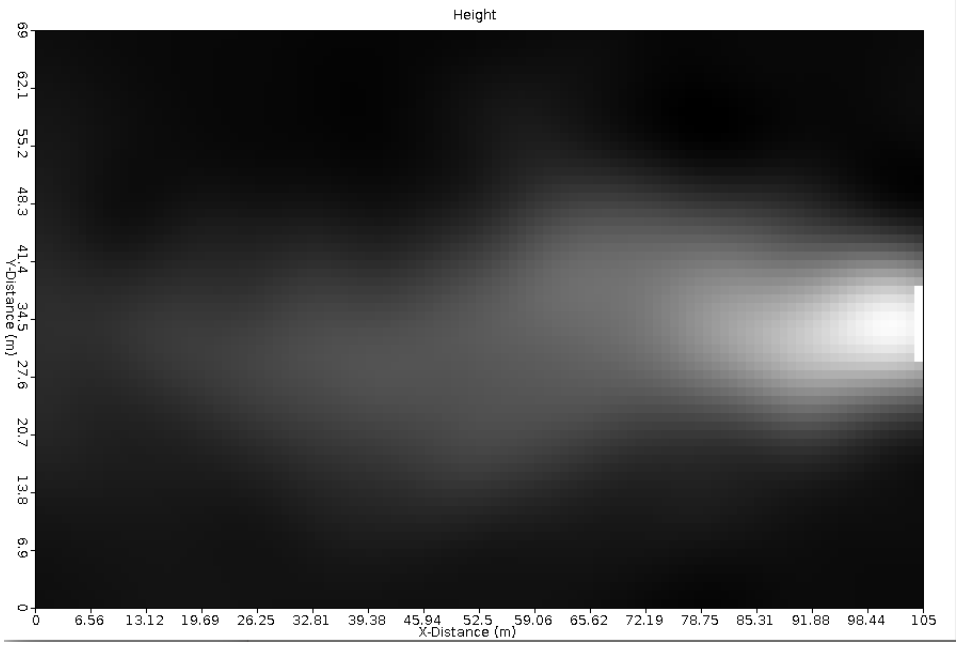

Figure 3.16: Heatmap after calculations described in approach 2

There is a path roughly through the middle of the field. Where in the first heatmap the sides of the field were relatively light, in this new heatmap the sides are quite darker than the middle of the field. Further away from the goal, there are some lighter parts at the sides of the field, but when close to the goal the lighter area is clearly in the middle.

This does make sense, as it is easier to score a goal when right in front of the goal, instead of at the side of the goal. When a player is in possession at the side of the field, he will probably dribble more to the middle of the field, or cross the ball to the middle to a better position. This corresponds to the higher values in the middle of the field than at the sides of the field.

3.3 Difference approach 1 and 2

There is a big difference between the heatmaps obtained after the different approaches. This difference is caused by the different formulas for calculating the values. Why is this difference so big?

In the first approach, we only looked at the direction of the ball, and only when the right team was in possession of the ball. This resulted in a change of value whenever the ball went from one point to another. Basically, this calculation results in a heatmap which shows the places on the field where the ball has been most before ending close to the goal. It is explainable that the sides of the field at the half of the opponent are lighter than the rest of the field, as most attacks originate at the sides. This also makes sense when looking at the players who were playing for PSV in these games. They played with wingers on the sides, so it makes sense that most attacks originated there.

In the second approach, not the sides but the middle is light. A reason for this is the variables we chose to do the calculations with. An important part of these calculations is the direction the ball is travelling. This is calculated by comparing the distance of the ball to the goal in *frame 1* with the distance of the ball to the goal in *frame 2*. The distance of the ball to the goal is smaller when the ball is in the middle of the field than when the ball is at the side of the field. This means that when the ball went from the side of the field to the middle of the field, this is seen as the right direction. When the ball traveled the other way around, this is seen as the wrong direction, as the distance of the ball to the goal increases.

This section started with the question:

Based on measured data of multiple football matches, can we assign a value to every place in the field, to indicate how good or how bad it is for a football team when they are in possession of the ball at that place in the field at any given time in the game?

This experiment shows that it is possible to assign a value to every place in the field to indicate how good or bad this position is. This experiment also shows that it is important to take multiple variables into account when calculating this values, and that a graphical representation is a good way to show the results.

In the end of this experiment we used 4 variables to calculate the values of the positions in the field. This gave us a totally different result than the first approach. It is important to choose the right variables and give the right value to the variables in the calculations. This experiment is not final and can be extended by adding more and more variables to the calculations. The more variables used, the more reliable the results will be.

Chapter 4

Conclusions

4.1 Common game situations

We examined three different common game situations: goal kicks, throw ins, and passes made by defenders. Every situation has its own results with corresponding conclusion, so these conclusions will be split per situation.

Goal kicks

We looked at the difference between taking a goal kick long and taking a goal kick short. Obviously after taking a goal kick long, the ball will be closer to the goal of the opponent at the first period after taking the goal kick than when taking the goal kick short. So we did not look at the period of time directly after taking the goal kick, but we looked at the period of time after 2.5 seconds after taking the goal kick.

What we saw is that there are about 5 seconds where there is a significant difference between the long and short goal kicks. This is between 73 and 78 seconds after taking the goal kick. These 5 seconds are too short to say that there is a big advantage for taking a goal kick long over taking a goal kick short. But if we have to choose between one or the other, we would choose taking goal kicks long, because our calculations tell us that there is a slight advantage for taking a goal kick long over taking a goal kick short.

Throw ins

We did not look at the distance of the ball to the goal of the opponent when examining the throw ins, but we looked at the distance the ball traveled in the direction of the goal of the opponent. The reason for this is that the distance of the ball to the opponent goal can differ much for different throw ins.

The results for the throw ins are very clear, there is a convincing winner. The winner between throwing the ball forward or throwing the ball backward, is backward. Almost the whole measured period of time after the throw ins were taken is in favour of the throw ins which were thrown backward. Not only the mean distance traveled in the direction of the opponent goal, but also the results of the T-test are clear. The p-value is close to 0 at most times, with an advantage for throwing the ball backward.

It might not be the result one would expect, but it is possible that there is a reason for this result. It might have to do with the position where the throw ins were taken. We did not distinguish between different positions on the field where the throw ins were taken, but we looked at all the throw ins which are more than 20 meters away from one of the two back lines.

It is possible that for example all throw ins on their own half of the field were thrown backward and all throw ins on the opponent half of the field were thrown forward. When the ball is on their own half of the field, it is easier to cover more distance, as there is more distance between the ball and the opponent goal than when the ball is on the opponent half of the field.

Passes made by defenders

When researching passes made by defenders, we looked at three possible choices, instead of two as with the previous researched situations. We distinguished between passes backward, passes wide and passes forward. We did this by comparing all three choices against the other two choices.

When comparing passes backward to passes forward and passes wide combined, there was no significant advantage for any. The same was true when comparing passes wide to passes forward and passes backward combined.

When we compared passes forward to passes backward and passes wide combined, there was a significant advantage for passes forward. Almost the whole period after the defender passed the ball, there is an advantage when passing the ball forward. Since there was no winner when comparing the other two situations, we can say that passing the ball forward is significantly better than passing the ball backward or passing the ball wide.

It is important to take into account here that in this research, we did not look at the situation the defenders are in. It is easy to say that the defenders have to pass the ball forward because this leads to a better game situation. It is not always possible for defenders to pass the ball forward. This depends on the situation they are in, if there is much pressure from the opponent, it is not always possible for a defender to pass the ball forward.

4.2 Assigning values to all positions on the field

We managed to assign a value to every position on the field by recalculating every value of all positions for all frames in all the games. The first calculation we did was just looking at which team was in possession, and taking the mean of the values of the position where the ball is in, and the position the ball was in the next frame. Then with a Gaussian, we distributed the change in value over the field.

This approach resulted in a heatmap which looked like a sort of a cloud over the field. The values increased when closer to the opponent goal, and the values are slightly higher at the sides of the field than at the middle of the field. The difference in values between the sides and the middle of the field is not really big, and we only looked at the team in possession and the movement of the ball. To get better results, we tried another approach after this.

The other approach used more variables than only the team in possession and the movement of the ball. We added the direction the ball is travelling, the number of players of their own team behind the ball, and the number of opponent players behind the ball. All variables have effect on the amount the values will be changed every frame. Obviously it is better to be in possession of the ball than to not be in possession of the ball. It is positive if the direction the ball is travelling in is in the direction of the goal of the opponent. It is better to have more players of your own team in front of the ball (so less behind the ball), and it is better to have less players of the opponent in front of the ball (so more players of the opponent behind the ball).

This resulted in a completely different heatmap. Instead of something that looks like a cloud over the field, a sort of a path has been created. Further away from the goal of the opponent, this path is wide with a few outliers to the sides. Closer to the opponent goal, the best values are in the middle. This makes sense, because when trying to score a goal, it is easier to be in front of the goal than to be at the side of the goal.

Of course these are not all variables to take into account, when calculating a value for every position on the field as we did. The more variables in the equation, the better. This is a good way to improve this research to be able to get better results. What we have shown with this research is that it is possible to assign a value to every position of the field, and that it is better to use more variables in the equation when calculating this values to get better results.

4.3 Overall conclusion

What looks like two different researches can be combined by looking at which positions on the field are best to be in, and by looking at how the team can get the ball there with which choice in a common situation in the game. When we evaluated the results of the choices in the common game situations, we assumed that it is better to be closer to the goal of the opponent. This assumption was confirmed by the results of the calculations of the values of the positions in the field, but the calculations for this can be adjusted by adding more or other variables which might give other results.

This is no problem for combining these researches, because we measured the distance of the ball to the goal of the opponent for the common game situations. If the calculations of the values of the positions in the field give us other results, we can adjust our view of what is good for the distance of the ball in the first research.

The most important thing we have shown with this research is that it is possible to define certain choices for common game situations and to measure the distance of the ball to the goal of the opponent after these choices. This can be combined with assigning a value to every position in the field which shows us what positions in the field are favourable over other positions. This way we can define the best places in the field to aim for, when looking at the common game situation.

By doing this, we improve the way of playing football for the researched team. This way they can get the ball in the best positions on the field quicker, in order to score goals, and eventually to win games. By getting the ball in the best positions on the field as quick and effective as possible, we increase the change of scoring goals. And like Johan Cruijff said: *"To win you have to score one more goal than your opponent."*

Chapter 5

Future Work

With this research, we demonstrated that it is possible to find the best choices for certain moments in the game, and to assign values to positions in the game. This can be very useful to improve the effectiveness of the way the game is played by a football team. This is a good way to improve the way of playing of a football team without spending a lot of money on expensive new players.

Although it is a good thing that we have shown that this is possible, it can and should be improved before it is used. We showed that if we do not look at the situation a defender is in, the best choice for the defender is to pass the ball forward. The next thing to do is to look at which situation the defender is in, and if passing forward is still the best choice in every situation.

In the second part of this research, we had two different results for the values of all positions on the field, with two different calculations. We saw a different result when using more variables. Depending on the situation, the next step is to determine which variables are important for which situations. When this is determined, new calculations can be made using these new variables for these situations.

In this research, we extracted our own features out of the raw data. These features were used in the different calculations. This was a time-consuming process, because for every feature one has to think how to calculate this out of the raw data. A solution for this is to use tried methods to calculate these features. A way to do this is to use the software tool SOCCER [PM11].

By using a tool like this to calculate features out of the raw data, a lot of time can be saved. This time can be used to improve the formulas used for the calculations. The most important thing to do to improve these formulas is to add more variables. The more variables added, the better, because a game of football is really complex and consists of many variables. To analyse the game, it is important to take into account as much of these variables as possible.

An example of a variable to add to the calculations is the space controlled by the players [FS05, TH00], which can be defined by voronoi diagrams. Other variables to include in the calculations are the variables which lead to possession switches [dW16]. These are just some example variables which can be used to improve

the calculations introduced in this thesis. It has to be researched which variables are best to use for these calculations.

When the calculations for the values assigned to all positions on the field are improved and new values are calculated, the first part of this research can also be adjusted. In this thesis, we assumed that it is better to get closer to the goal of the opponent, so this is what we used to define the good and the bad choices in the common game situations. When it is known which positions on the field are best to be in for the researched team, the definition of good and bad choices in common situations can be adjusted.

Bibliography

[dW16] R. de Winter. Analysing possession switches in football using subgroup discovery. Bachelor Thesis LIACS, 6 2016.

[FS05] A. Fujimura and K. Sugihara. Geometric analysis and quantitative evaluation of sport teamwork. *Inc. Syst Comp Jpn*, 36, 2005.

[PM11] J. Perl and D. Memmert. Net-based game analysis by means of the software tool soccer. *International Journal of Computer Science in Sport*, 10, 2011.

[Ste18] Stephanie. T test (students t-test): Definition and examples, 2018. [Online; accessed 10-August-2018].

[TH00] T. Taki and J. Hasegawa. Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings Computer Graphics International*, 1000.