Tracking behaviour on videos and predicting running performance.

ERIC VAN DAM
STUDENT NUMBER: 2020165

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE SCHOOL OF
HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Supervisor: Prof. dr. E.O. Postma

Second Reader: Prof. M. Postma

Tilburg University
School of Humanities and Digital Sciences Department of Cognitive Science &
Artificial Intelligence Tilburg, The Netherlands
June 2019

Acknowledgements

I would first like to thank my thesis advisor dr. E.O. Postma of the Department of Cognitive Science & Artificial Intelligence School of Humanities and Digital Sciences at Tilburg university. Prof. Postma's experience and expertise steered me in the right the direction whenever he thought I needed it. I also would like to thank prof. J.S. Olier Jauregui, for his contribution during the meetings and his expertise with complex models. Finally, I must express my very profound gratitude to my parents and to my girlfriend for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this master thesis. This accomplishment would not have been possible without them. Thank you.

Abstract

Predicting human behaviour is a complicated task because humans do not always make rational decisions and therefore could be unpredictable. Capturing behaviour based on video tracking to make predictions was employed in several studies. The current research attempted to predict running behaviour based on the body compositions of the runners which were extracted from video clips. This research used a data set featuring extracted coordinates of 18 anatomical points from the starting positions of young football players running for a total of 66 meters in order to predict the time needed to finish. Ridge, lasso, and elastic net regressions were performed to determine the extent to which running performance can be predicted based on the starting positions of the runners. The results demonstrated that the elastic net model performed the best over the test set. This model achieved a mean squared error of 3.16 and an average absolute error of 1.46 seconds. The elastic net model was better than the baseline model at predicting running time. However, the number of observations within the data set was small, and consequently the results are based on a small amount of data. Predicting behaviour based on extracted coordinates from human pose methods, on the other hand, was a relatively new task. This research revealed that predicting running behaviour based on the human pose is a technique that could be applied in the future for predicting other human behaviour based on video tracking. Future research should examine whether a larger data set would achieve better results and whether using more sophisticated algorithms to extract body coordinates could influence the results.

Keywords: *predicting behaviour, video tracking, elastic net, ridge, lasso, running performance.*

Table of contents

1. Introduction

Running is a key aspect of many sports. In the Netherlands alone, 1.5 million people run on a weekly basis, and this number is increasing every year (NOC*NSF, 2017). In sports such as basketball, baseball, football, and field hockey, improving one's running technique is an important method to improve performance (Di Prampero, Botter, & Osgnach, 2014). One element of improving performance is analysing videos of athletes to gain insight into specific performance and to monitor players' development (Cust, Sweeting, Ball, & Robertson, 2018). These videos could be used also to track behaviour or make predictions about the future behaviour of the athletes. With a growing collection of literature that recognizes the importance of data science and human pose estimation methods, the body compositions of the athletes could be used for predicting behaviour based on video tracking.

Human pose methods are based on extracting key points from videos and transforming them into coordinates. Most studies that used human pose methods focussed on 3D analysis (Zhao, Li, & Pietikainen, 2006; Pfister, West, Bronner, & Noah, 2014), in which researchers place multiple cameras at a fixed distance from one another to capture every movement in the 3D space. An advantage of such an approach is automatically capturing running performance because the cameras extract nearly all the key points of the body. However, such analyses are expensive and difficult to apply in practice because the experiments are often conducted in laboratories.

This leaves a gap for 2D human pose estimation methods. Previous research used human pose to detect difficult objects (Yao & Fei-Fei, 2010) or distinguish different clothes in an image (Yamaguchi, Luis, Ortiz, & Berg, 2012). Research by Singh (2016) used human pose for multiple purposes, such as distinguishing different clothes and body parts. These approaches are based on images and do not include extracting human poses from videos. In a recent development, Cao et al. (2016) devised the OpenPose algorithm. This method automatically extracts 18 anatomical key points from an image or video and therefore could be useful in analysing running technique.

The previous studies employing human pose methods were focussed only on detecting or distinguishing different body parts or certain objects within a frame. Another gap within this field is if human behaviour can be predicted based on video tracking. A study by Felsen, Agrawal, and Malik (2017) examined whether the subsequent move of players in water polo and basketball could be predicted based on visual inputs. Their model achieved accuracies of 70% for the basketball data and 36% for the water polo data. Other research by Truscott and Belden (2019) used data from the 2016 Tour de France professional bicycle stage race, which was filmed from a helicopter to provide an overhead view of the pelotons. Video clips were computed into individual images per frame, after which a set of image processing algorithms were employed to extract rider locations and network structure to examine peloton behaviour.

This research attempts a task similar to that of the research by Felsen et al. (2017) and Truscott and Belden (2019) by examining whether behaviour could be predicted based on video tracking. The current research used an already-prepared data set of young football players running back and forth between cones separated 16.5 metres. Key points of their bodies were extracted and transformed into coordinates using the OpenPose algorithm. Previous research using this data set focussed on correlation between these coordinates and running performance (Van der Meijden, 2019; Van Leeuwen, 2019). However, both studies were limited to correlation and were unable to make predictions about the performance or behaviour of the runners.

The specific objective of this study is to predict behaviour in the form of running performance based on frame-based key point coordinates extracted from 2D video sequences of the runners' starting positions. This research aims to predict the time required to run one lap of a track based on the extracted coordinates of the runner at the start. Therefore, the following research question is formulated:

To what extent can running performance be predicted from body coordinates extracted from a frame of the starting position?

To answer this research question and examine whether behaviour in the form of running performance can be predicted, four regressions were performed. Multiple linear regression was used as baseline, while three regularisation models were used to predict the outcome of runners on the unseen test set. These regularisation models were lasso, ridge, and elastic net. Based on the literature, elastic net is expected to outperform the lasso and ridge regularisations because it combines the two shrinkage penalties of the other models and therefore could make better predictions (Zhang, Tian, Bai, Xiahou, & Hancock, 2017; Zou & Hastie, 2005; Marafino, John Boscardin, & Adams Dudley, 2015). These algorithms were selected because the number of observations within the data set was limited to 50, and these algorithms perform adequately with a limited number of observations.

Outline

This research comprises six chapters, each divided in several subchapters. Chapter 1 contains the introduction of this research, the research question, and the practical and academic relevance. Then, Chapter 2 discusses the related work, and Chapter 3 describes the method and analysis used for this research. This chapter also establishes the steps taken during pre-processing. Chapter 4 presents the results of the of this research. Chapter 5 contains the discussion, which is divided into the goal of the research, the limitations, and the recommendations for future research. The final chapter presents the conclusions.

2. Related work

This chapter is divided into six sections. The first section describes studies about predicting behaviour from video tracking. The second section sets out the theoretical dimensions of the research into running performance. Section three examines the importance of a sprint's start. Afterward, the fourth section briefly explains the OpenPose algorithm. Section five investigates how previous studies have used this data set, and the final section discusses different techniques used for feature selection.

2.1 Research on predicting behaviour from video tracking

Running performance in the form of speed or time could be considered how someone is behaving. Analyzing the runners' starting positions and observing how the runners behave according to these positions could be used for many other tasks. This section details several papers in which video tracking prediction tasks were performed.

Predicting certain behaviour based on video tracking is becoming popular within the field of science. Research by Felsen et al. (2017) examined whether the subsequent move of basketball and water polo players could be predicted based on visual input. They transformed images into an overhead view to predict who would next possess the ball. Their model achieved accuracies of 70% and 36% for the basketball and water polo data sets, respectively. Both models performed better than the predetermined baseline and could extract generic game strategies. This implies that behaviour can be predicted based on video tracking and therefore contributes to the current research in which a similar task is performed. Another study that attempted to predict the following move of basketball players based on video input was conducted by Su, Hong, Shi, and Park (2017). Their study used first-person videos to predict future behaviour. This approach is relatively new, as most research is focussed on data with a view on the entire field rather than through the eyes of the players. This research not only examined the direction of the ball but also whether an algorithm could predict the direction in which the players would run.

Researchers investigated also the effect of grouping behaviour based on video tracking within the field of sports (Truscott & Belden, 2019). This research used data from the 2016 Tour de France professional bicycle stage race, which was filmed from a helicopter to provide an overhead view of the pelotons. After the video clips were computed into individual images per frame, the rider locations and network structure were extracted to examine peloton behaviour. The object of this research was to map peloton behaviour and therefore lacked making future predictions.

These previous studies illustrate that predicting behaviour based on video tracking is a feasible task. Because not every study was able to make predictions about behaviour, this research attempts to make predictions based on video input. The approach for this thesis was similar as Truscott and Belden (2019) by dividing the video clips in single frames in order to make analyses.

2.2 Research on analysing running performance

Running performance can be measured and analysed in various ways, and this section explains some of these methods. A previous study found a positive correlation between longer stride length and faster speed (Williams, Netto, Kennedy, Turner-Bryndzej, & Campbell, 2018). In the study, children ran a 30-metre track and were recorded with cameras from different angles to analyse their running technique and performance. The purpose of the study, however, was to analyse specific biochemical factors and running performance; prediction was not one of the aims.

Research by Herrmann et al. (2018) examined whether running speed can be predicted from body composition. They discovered that body composition is a superior predictor for running speed than is body mass index. Nagano, Fujimoto, Kudo, and Akaguma (2016) conducted a study that used image processing to predict running speed. They used a new method with a mechanical centroid to predict walking and running speeds. The results were similar to those of the studies in which multiple digital cameras were used to obtain 3D kinematics (Nagano et al., 2016).

2.3 Research on a sprint's start

Previous studies indicate that the start of a sprint is crucial for a good performance (Brown & Vescovi, 2012; Eriksen, Kristiansen, Langangen, & Wehus, 2009). The study by Eriksen et al. (2009) analysed Usain Bolt's world record sprint during the 2008 Olympics in order to obtain the quickest time that he could achieve. They learned that while his start was excellent, the lack of acceleration near the end cost him a faster finish.

Since the start is crucial for the performance of a sprint, it is sensible that researchers have attempted to optimise the start position of sprinters. A study by Eikenberry et al. (2008) investigated differences in the positions of the feet. They found that the starting time can be minimised if runners start with their dominant foot. Further research examined the positions of the legs and hands and the angle of the forward lean (Slawinski et al., 2010). They also confirmed that the start of a sprint is crucial in obtaining the best performance.

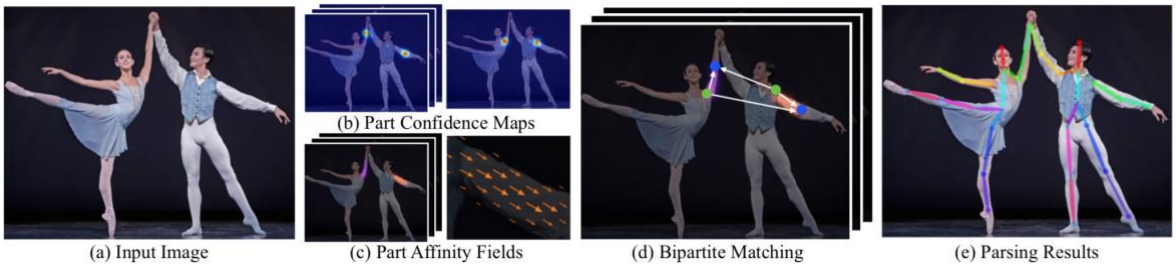
2.4 OpenPose

This research employs a data set of extracted body coordinates analysed with the OpenPose technique developed by Cao et al. (2016). OpenPose is the first real-time multi-person system to detect key points of human arms, shoulders, and feet on single images. Figure 1 below depicts OpenPose's pipeline method. The advantages of this method are its efficiency, practical applicability, and ability to detect multiple people in 2D space. However, OpenPose is not flawless; it struggles to analyse images of people taken from the side. Furthermore, the videos were recorded with a smartphone rather than a professional camera. The lower quality of the videos makes it more difficult for OpenPose to correctly

detect the key points. As a result, in some cases, the algorithm incorrectly identified the position of the arms and legs: what the algorithm identified as the right arm was in some cases the left arm.

DensePose (Guler, Neverova, & Kokkinos, 2018) was developed as a successor to OpenPose. The DensePose algorithm can extract additional key points of the body from images to create a 3D model of the human body. It produces highly accurate results in real time (Guler, Neverova, & Kokkinos, 2018). Compared to OpenPose's, DensePose's algorithm is more complex and computationally expensive. Therefore, despite minor flaws, the OpenPose algorithm was used instead.

Figure 1. *A pipeline of the OpenPose algorithm: (a) is the single frame of the input image; (b) displays a convolutional network detecting confidence maps of the different body parts; (c) displays part association; (d) displays the body parts being associated with the subjects; and (e) displays the detection of all body parts (Cao et al., 2016).*



2.5 Previous research with current data set

This section briefly explains what has already been examined in the current data set. Three previous studies are referred to in order to establish what is already known in addition to the gaps that the current research is intended to fill.

Previous research on this data set investigated whether a correlation exists between running performance and extracted features. Research by Van Leeuwen (2019) revealed that body pose while running on a track correlated with speed. Furthermore, differences in the first 16.5 metres of the track were caused by different starting positions of the runners. Van der Meijden (2019) discovered that running technique correlated with sprinting performance; in particular, bringing the upper body upright after acceleration correlated with sprinting performance. Despite both studies finding correlations, neither could make any predictions about running performance.

One study that did make predictions about running performance from this data set was conducted by Luttik et al. (2018). They performed a ridge regression to predict speed. Their model, however, offered limited predictive power, as it always predicted a value near the mean.

2.6 Regularisation methods

Feature extraction can be used to locate the most informative features in the data set. Cui, Bai, Zhang, Wang, and Hancock (2019) discovered that the models provided different results with respect to selecting the most informative feature. They compared lasso regression with elastic net to understand

the differences in finding the most informative features. Another shrinkage model is ridge regression, which has been found to outperform lasso regression (Hastie, Tibshirani, & Friedman, 2008).

Previous research revealed that lasso and elastic net feature selection perform well on high-dimensional data (Zhang, Tian, Bai, Xiahou, & Hancock, 2017; Zou & Hastie, 2005). In addition, Zou and Hastie (2005) found that elastic net is particularly useful when the number of features is larger than the number of observations. Furthermore, they state that lasso is less useful when the data have more features than there are observations. Research by Marafino, John Boscardin, and Adams Dudley (2015) established that elastic net can reduce the number of features from thousands to hundreds with only a small impact on the overall performance.

This research examines the extent to which running performance can be predicted based on the runners' starting positions. Based on existing literature, elastic net regression is expected to perform best on the current data set because the number of features is larger than the number of observations (Zou & Hastie, 2005). Furthermore, lasso and ridge regression are expected to perform well on the data set and are therefore included in the analyses.

3. Method

This chapter describes the methodology used in this study. It presents the algorithms and software used and describes the data set, task of the runners, pre-processing, experimental procedure, and evaluation method used to compare the results.

3.1 Algorithms and software

The software used to analyse the running technique of young football players is the programming language Python. This programming language was used for the pre-processing and conducting experiments to obtain results. The coordinates were extracted using the OpenPose algorithm. To predict the running time, four regressions were performed. Multiple linear regression was used as a baseline to understand the results of lasso, ridge, and elastic net regressions. To test the performance of the models, the mean squared error and mean absolute error were calculated. This was performed for all four regressions. For more information, see Section 3.5.5, evaluation method.

3.2 Task of the runners

This section explains how the data were gathered and establishes the specific task of the runners. Filming, gathering, and extracting coordinates from the videos was already performed and is not part of this research. This research used only the data set derived from these videos.

The videos consist of 50 football players aged 10 to 12 years running four times along a track of 16.5 metres—back and forth twice—for a total distance of 66 metres. The OpenPose algorithm detects 18 body parts and translates these points into coordinates. For more information, see Sections 2.3 and 3.3 or the paper by Cao et al. (2016). A smartphone camera was used to capture the videos in full-HD (1080 × 1920 pixels) resolution with 60 frames per second. The camera was set on a tripod and at a fixed distance. Not all videos were on water level, which resulted in slightly different coordinates in some cases. The lengths of the videos depend on the speed of the runners and vary from 16 to 26 seconds.

The specific task of the runners was to run as back and forth between the cones as quickly as possible. The timer began when they passed the first cone and ended when they passed the final cone. It is not a normal sprint because the runners did not start from a block and abruptly stopped at the turning points. Figure 2 illustrates the track and the position from which the video was recorded.

Figure 2. *A still image from one of the videos. The athlete (on the right of the image) runs back and forth between the cones.*

3.3 Data set

The original data set consists of 50 videos of young football players running back and forth between two cones. The cones were spaced 16.5 metres apart, and the children ran back and forth twice. The features from the 50 videos were already extracted and transformed into coordinates using the OpenPose algorithm, as described in Section 2.3. This resulted in 50 files in the comma-separated values (CSV) format, one file for each of the runners with their body coordinates. The data set thus consisted of 50 rows and 56 features after pre-processing. This is explained in greater detail in Section 3.4.

The unprocessed data set consists of six features: FileIndex, FrameIndex, KeypointIndex, KeypointX, KeypointY, and KeypointScore. The most important features for this research were KeypointIndex, which has a value between 1 and 18 and refers to the different body parts; KeypointX and KeypointY, which are specific coordinates of the body parts; and KeypointScore, a combination of the X and Y coordinates. For a sample overview of one of the CSV files, refer to Table 1 below.

Table 1. *Sample of a CSV file. FileIndex identifies the runner; FrameIndex refers to the frame number, zero indicating the first frame; KeypointIndex refers to the body part (e.g. arm, ankle, elbow); and KeypointX, KeypointY, and KeypointScore refer to the coordinates of the body parts in the frame.*

FileIndex	FrameIndex	KeypointIndex	KeypointX	KeypointY	KeypointScore
22.0	0.0	1.0	1817.0	565.0	0.653976
22.0	0.0	2.0	1809.0	565.0	0.605477
22.0	0.0	3.0	1808.0	589.0	0.260387
22.0	0.0	4.0	1806.0	605.0	0.203959
22.0	0.0	5.0	1824.0	565.0	0.669655

3.4 Pre-processing

This section explains the pre-processing performed to prepare the data set for analysis. As described in Section 3.2, there were 50 CSV files with the extracted coordinates of the runners. These files first needed to be merged into a single data set with only the starting frames of each runner. Each body part becomes a feature, and because each body part has an X, Y, and Score index, this resulted in a data set with 54 features of extracted body coordinates (18 body parts multiplied by three indexes). Including the FileIndex of the runner and the time taken to complete the course, the data set contained a total of 56 features. See Table 2 below for a sample of the data set.

3.4.1 Extracting starting positions and creating new features

This section explains in detail how the data set was prepared for analysis and how the 50 CSV files were added one at a time to the new data set. First, all data with a frame index of 0.0 were extracted. This frame index refers to the first frame and therefore corresponds to the frame displaying the starting

positions of the runners. Second, the KeypointIndex had a value between 1 and 18, referring to one of the 18 body parts. These indexes were assigned to the correct body parts, which resulted in three coordinates—an x-, y-, and s-score—per body part. Third, the time the runners took to complete the course was calculated by dividing the number of frames by the number of frames per second, which was 60. The time taken to complete the course is expressed in seconds. Finally, this resulted in a data set with 56 features and a row for each of the 50 runners, as displayed in Table 2.

***Table 2**. A sample of a preprocessed data set with 56 features and 50 rows. fileIndex refers to the different runners. noseX, noseY, noseS, and the other similarly labelled columns refer to the specific body parts with the extracted coordinates; the X and Y coordinates refer to specific positions within the frame, while the S-score is a combination of the X and Y coordinates. All these coordinates had a frameIndex of zero in the original data set. The final feature in this data set is the time taken to complete the course.*

	fileIndex	noseX	noseY	noseS	neckX	neckY	neckS	RshoX	RshoY	RshoS	...
0	0.0	1760.0	643.0	0.716575	1765.0	660.0	0.701018	1754.0	660.0	0.693718	...
1	1.0	1796.0	650.0	0.651613	1808.0	664.0	0.683265	1799.0	664.0	0.645674	...
2	2.0	1821.0	541.0	0.696741	1825.0	555.0	0.726528	1816.0	555.0	0.690350	...
3	3.0	1782.0	635.0	0.651314	1793.0	647.0	0.674857	1786.0	648.0	0.627948	...
4	4.0	1764.0	616.0	0.734196	1769.0	636.0	0.736007	1755.0	638.0	0.731777	...
5	5.0	1732.0	611.0	0.724840	1744.0	623.0	0.685525	1734.0	626.0	0.661228	...
6	6.0	1753.0	649.0	0.714526	1761.0	664.0	0.683010	1751.0	664.0	0.642278	...
7	7.0	1758.0	646.0	0.695697	1762.0	661.0	0.708279	1752.0	663.0	0.695161	...

3.4.2 Missing values and outliers

After pre-processing, the data set can be analysed. The first step is exploratory data analysis, which involves checking for missing values and possible outliers within the data set. Missing values are easily detected by searching for undefined values. No missing values were within this data set. However, some -1 values existed, which were flagged as missing, as all other values were non-negative. These values were analysed, and it was discovered that one feature had a value of -1 in nearly every observation. This feature was the X coordinate from the right ear. The explanation for this is that the videos recorded the subjects from the side, and as the right ear was not visible during the start of the videos, the algorithm had difficulty detecting these coordinates. These missing values could have been replaced by the coordinates of the subsequent frame, in which the ear was detected. However, in the following 100 frames, these values were still missing; once the ears were detected, the new coordinates no longer refer to the starting positions of the runners. Therefore, the feature RearX was not included in the model.

To determine outliers within the data set, a boxplot was used. The boxplot revealed three potential outliers in the end time of the runners, two of which were nearly twice the mean. These two

outliers could have significantly influenced the model, as there were only 50 observations. Therefore, to avoid these two outliers potentially influencing the model's performance, they were excluded from the analysis. The third potential outlier was within an allowable range of three times the Z-score, and since the number of observations was already limited, this observation was not excluded from further analysis. After handling missing values and outliers, the data set consisted of 48 observations with 55 features.

3.5 Experimental procedure

This section explains the experimental procedure used to obtain the results. Four regression models were employed: multiple linear, lasso, ridge, and elastic net. The results of the multiple linear regression were used as a baseline to understand the performance of the other models on unseen test data. The four regressions were performed using the sklearn package in Python. The error stands for the difference between the predicted and actual values; see Section 3.5.5 for further explanation regarding calculating the mean squared error (MSE) and mean absolute error (MAE).

3.5.1 Regression 1: multiple linear

A multiple linear regression is a basic model. Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed data (Hastie et al., 2008). The results from this model were used as a baseline to understand the results from the other models and to put the obtained results into perspective. Leave-one-out cross validation (LOOCV) was used to obtain the mean error on the test set. This type of cross validation was used instead of the normal cross validation due to the limited data within the data set. Each time, a different single observation was used as the test set, while the other 47 observations were employed as training sets. After fitting the model and testing the performance on the unseen test observation, the mean squared error (MSE) and mean absolute error (MAE) were computed; see Section 3.5.5 for an explanation. This process was repeated until all observations were used as a test set. The mean error was then calculated by dividing the sum of the errors by 48. These mean errors over all the test sets were then used as baselines to examine the performance of the other models.

3.5.2 Regression 2: ridge

Ridge regression is a simple technique to reduce model complexity and prevent overfitting which may result from multiple linear regression (Hastie et al., 2008). Ridge regression is a type of regression similar to least squares. The only difference is that the coefficients were estimated by minimising a slightly different quantity. Ridge uses a shrinkage penalty in which the coefficients shrink towards zero depending on the alpha parameter, in which alpha must always be greater than zero (Cui et al., 2019). As with least squares, ridge regression seeks coefficient estimates that fit the data well by

making the residual sum of squares (RSS) small. A disadvantage of ridge is that it shrinks the coefficient estimates towards zero but not to exactly zero, which means that the model always includes every feature (Zheng & Liu, 2011).

Different alphas were tested to obtain the alpha parameter that resulted in the lowest error using LOOCV. Each fold featured a different alpha value, and the alpha value that resulted in the lowest error was used in the model. Regarding the bias-variance trade off, as alpha becomes larger, the variance decreases, and the bias increases. After determining the best alpha parameter, LOOCV was used to obtain the mean error over all $n - 1$ folds that were used as test sets, where n represents the number of observations in the data set. These errors, expressed in MSE and MAE, were compared with those of the baseline models and the results obtained from the other models to provide an indication of the ridge regression model's performance.

3.5.3 Regression 3: lasso

Lasso regression is similar to ridge regression in that it is a type of regression in which the coefficient estimates shrink towards zero. In contrast to ridge, however, lasso sometimes forces estimates to shrink exactly to zero should the tuning parameter alpha be sufficiently large (Zheng & Liu, 2011). This is the primary advantage of lasso over ridge because feature coefficients could be set to zero and therefore be excluded from the model should these features result in having little predictive power. The performance of the lasso depends on selecting the best alpha parameter, as the size of alpha determines whether features shrink to zero. When alpha = 0, lasso gives the least squares fit, and when alpha becomes sufficiently large, lasso gives the null model in which all coefficient estimates equal zero. The shrinkage process is important in identifying the most informative features (Cui et al., 2019). However, it remains a challenge to find these informative features without a possible loss of information.

As described in the previous section, to obtain the best results, lasso regression depends on the alpha parameter. Multiple alphas in a range of 0.001 to 20 were used to obtain the alpha parameter that provided the lowest error. This was performed using LOOCV with different alphas. After finding the best alpha parameter, the model was fitted and tested against the unseen observations. This was executed using LOOCV, which splits the data into a number of folds equal to the number of observations. Each observation was used as a test set to obtain the average error over all 48 test sets. These errors, expressed in MSE and MAE, were compared with the baseline model and the other models to provide an indication of the performance of the lasso regression model.

3.5.4 Regression 4: elastic net

Elastic net regression is a combination of ridge and lasso regarding their penalties. Elastic net has an advantage over the other models because it considers grouping the effect of the predictors (Zou & Hastie, 2005). For this reason, highly correlated features are grouped and therefore are in or out of the

model together. The alpha in this model is the mixing parameter between the ridge (a = 0) and lasso (a = 1). The other parameter is the L1 ratio, which is a combination of the L1 and L2 penalties of ridge and lasso. Previous research found that elastic net feature selection performs well on high-dimensional data (Zhang et al., 2017; Zou & Hastie, 2005). They discovered also that elastic net is particularly useful when the number of features is larger than the number of observations. Since the current research used a data set that has more features than observations, elastic net was selected to examine whether it could outperform the other models and predict running performance.

To find the optimal parameters for this model, LOOCV was used to determine the parameters that resulted in the lowest error. This was executed using the 'gridsearch' package in Python. These parameters were then used in the model to obtain the error for each test set. The average error was calculated using LOOCV, which splits the data into a number of folds equal to the number of observations. Each observation was then used as a test set to obtain the average error over all 48 test sets. These errors, expressed in MSE and MAE, were compared with the baseline analysis and the other analyses to provide an indication of the performance of the elastic net regression analysis.

3.5.5 Evaluation

To evaluate the performance of the models, the MSE and MAE for the unseen test data were computed. This was performed to assess the models' accuracy and to obtain their predictive power. Mean squared error is a single value that provides information regarding the regression line's goodness-of-fit. The smaller the MSE value, the better the fit, as smaller values imply smaller magnitudes of error. Mean absolute error measures the average magnitude of the errors in a set of predictions without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation, in which all individual differences possess equal weight. For each model, this resulted in 48 MSEs and MAEs. The average error over these folds was used to examine the performance of the models. Mean squared error and mean absolute error are calculated using the following equations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right|$$

4. Results

This chapter discusses the results of the regression analyses. To test the extent to which running performance can be predicted from extracted coordinates of body parts at a sprint's start, four regressions were performed. Table 3 below provides an overview of how well the models performed on the test data, in which multiple linear regression was used as baseline to understand the performance of the other models. The running task lasted a mean of 18.9 seconds (standard deviation = 1.7 seconds) and a median of 18.8 seconds. The baseline model achieved an MSE score of 15.55 and a MAE score of 3.55. The MAE reveals the size of the error on the test set in terms of seconds.

The results of the lasso regression indicated that an alpha parameter of 0.04 provided the best results for this model. The model achieved an MSE of 5.48 and a MAE of 2.79 seconds over the 48 test sets. The lasso model achieved lower errors on the test set than on the baseline set by the multiple linear regression.

The alpha parameter for the ridge regression was set to 0.8, as this produced the best results. The average errors on the test sets revealed an MSE of 5.32 and an MAE of 2.5 seconds. As with the lasso regression, the ridge model performed better than the baseline model on the test set. However, the ridge model performed slightly better than the lasso regression regarding errors.

The results from the elastic net regression demonstrated the lowest errors compared with the other three models. The best results were obtained by setting the alpha parameter to 0.8 and the L1 ratio to 0.2. After setting these parameters, the model achieved an average MSE of 3.16 and an average MAE of 1.46 seconds over the 48 test sets using LOOCV. This model outperformed lasso and ridge and was far better than the baseline model. Figure 3 illustrates the predicted and actual values over all the test sets for the elastic net regression. As displayed in the figure, the model always predicts a value close to the mean. The figure reveals a discrepancy for the final sprinting number: the actual value is well above the predicted value.

Figure 3. *Differences between the actual values (orange bars) and predicted values (blue bars) for the elastic net model.*

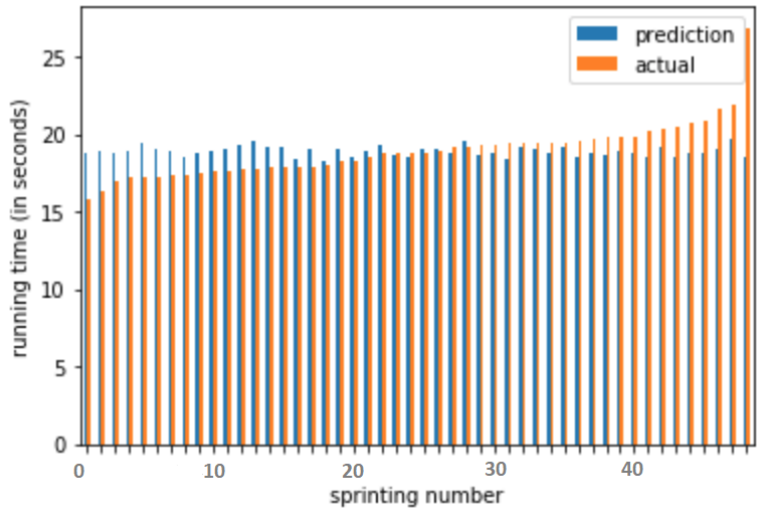
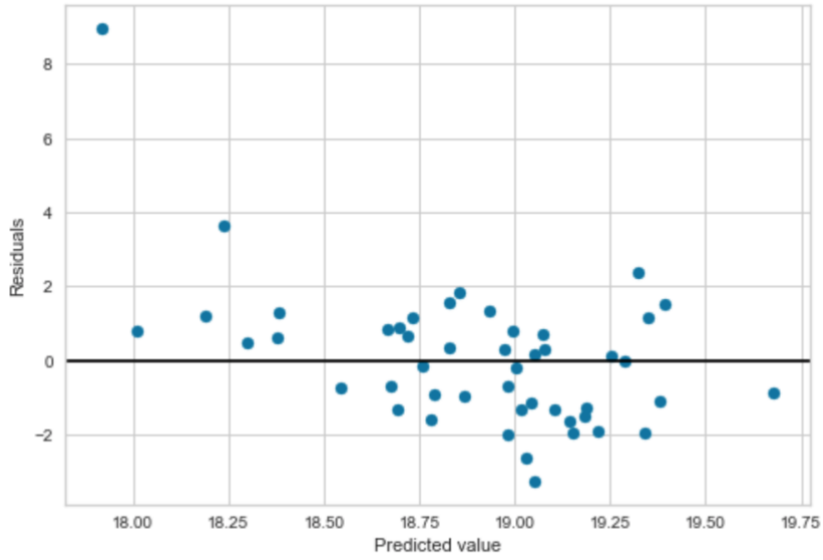


Figure 4 displays how the residuals are distributed compared with the actual value. The distance from the line at 0 is how poor the prediction was for that value. Since Residual = Observed – Predicted, positive values for the residual (on the y-axis) indicate that the prediction was too low, and negative values indicate that the prediction was too high; 0 indicates that the guess was exactly correct. As with Figure 3, the residuals reveal one major error in the upper-left corner. This residual plot does not illustrate a pattern, this indicates that the model is good fit and is not overfitting.

Figure 4. *Predicted value (x-axis) versus the residuals (y-axis) to examine how far the predictions are from the actual values in the elastic net model.*



To view the results with respect to the research question '*To what extent can running performance be predicted from body coordinates extracted from an image of the starting position?*', elastic net performed best on the test set in terms of average errors. The model predicted running time with an average MAE of 1.46 seconds on the test set. Compared to the simple baseline model, elastic net regression performed much better regarding errors.

Table 3
How the models performed on the unseen test set. The mean squared error and mean absolute error scores represent the errors on the unseen test data. The *A* values are the optimal alphas for each model. 'L1' refers to the L1 ratio, which was used in only the elastic net regression.

		Score	
	Models	MSE	MAE
	Lasso Regression (A = 0.04)	5.48	2.79
	Ridge Regression (A = 0.8)	5.32	2.50
	Elastic Net (A = 0.8, L1 = 0.2)	**3.16**	**1.46**
Baseline	Multiple Linear Regression	15.55	3.55

5. Discussion

In Section 5.1, the results of the models are evaluated regarding the research question stated in the introduction. Section 5.2 discusses the limitations of this research, and Section 5.3 proposes ideas for future work on the topic of human pose estimation.

5.1 Goal of this research

The object of this research was to investigate whether behaviour can be predicted based on video tracking. This was examined by attempting to predict the running performance of children based on the positions of their body parts at the start of the sprint.

Additionally, previous research made attempts to predict behaviour based on video tracking (Felsen et al., 2017; Su et al., 2017). Both studies could track and predict behaviour based on video input. The current research performed a similar task by predicting the time it would require to run the track based on the positions of the runners' body parts at the start. This human pose method in combination with predicting behaviour is a different approach than previous research. This research could therefore be the beginning of more behaviour prediction using video tracking and human pose methods. The goal would be to automatically extract body compositions from video footage and directly make predictions about people's future behaviour based on this composition. An example could be to use security cameras to detect potential terrorists at an airport based on their body compositions.

Based on the literature, the expectations were that the elastic net regression would outperform the lasso and ridge regressions and would perform better than the baseline (Cui et al., 2019; Marafino et al., 2015). This was expected not only because elastic net combines the penalties used in lasso and ridge but also because the model considers grouping the effect of the predictors rather than examining predictors separately. Furthermore, as this was the case, it was expected also that elastic net would perform better when the number of features was larger than the number of observations (Zou and Hastie, 2005).

The results indicate that all the models performed better than the baseline. The best performing model was, as expected, the elastic net model, with an MSE of 3.16 and an MAE of 1.46 seconds. Further examination of the results reveals that the elastic net model made predictions close to the mean. The model consequently has difficulty predicting values further from the mean and therefore loses predictive power. The model might predict values close to the mean because the alpha parameter which produced the lowest error was selected. A consequence of this approach could be that the lowest average error is obtained by making predictions not far from the mean. However, an error of 1.46 seconds on average is still much better than the baseline and indicates that the model can make predictions. Hastie et al. (2008) learned that the ridge model performs better than the lasso model regarding errors. Despite the limited data, this research could confirm these findings because ridge

performed slightly better, with an MAE of 2.5 seconds, than did lasso, with an MAE of 2.79 seconds. In contrast to previous work using this data set by Van der Meijden (2019) and Van Leeuwen (2019), this research succeeded in building a model that can predict running time based on the extracted coordinates.

5.2 Limitations of the research

This study encountered several limitations that must be highlighted to place the results in the right perspective. First, the number of observations within the data set was small, and consequently, the results are based on a small amount of data. The performance of the models and the results of this research are based on only 48 observations. This reduces the significance of the research and limits the use of this model for other purposes in the field. Second, as described in Section 2.3, the OpenPose algorithm made errors in identifying arms and legs because the videos were recorded from the side rather than from the front. Finally, not all data from the data set were used in this research; as the primary objective was to investigate whether the starting position could predict running time, only the frames from the starting positions were used. Therefore, this research did not consider that the turning points during the run could greatly influence performance.

5.3 Future work

Considering the limitations of the research, some recommendations can be made for future work. Tracking and predicting behaviour based on videos is a complicated task because human behaviour could be difficult to predict. This research made use of only one frame extracted from the videos to predict behaviour; future research should investigate whether predicting behaviour based on multiple frames would be more accurate. When the model includes more frames, it could be possible to better predict behaviour. Using multiple frames could make the model more complicated; therefore, it would be better to use more sophisticated methods, such as a neural network.

As discussed in Section 5.2, the number of observations was small. These experiments could be conducted on a much larger data set to determine whether similar results could be obtained. A recommendation for recording the videos is to film the athletes from the front rather than from the side. To better analyse running performance, the running could take place on an athletics track where the runners can begin from a starting block and where no turning points are needed.

Another recommendation concerns the OpenPose algorithm used to extract the coordinates from the runners. As mentioned in Sections 2.3 and 5.2, OpenPose made mistakes in extracting body part coordinates from the videos. Guler et al. (2018) developed DensePose as a successor to OpenPose. This newer algorithm can extract additional body key points from images to make a 3D model of the human body and produce highly accurate results in real time. Therefore, future research should examine whether more accurate predictions can be made if the features are extracted using

DensePose. A comparison between the two algorithms could be made to determine the extent to which the limitations of OpenPose influence the results.

6. Conclusion

This study examined the extent to which running performance of the athletes could be predicted based on coordinates at the sprint's start. This task could be considered predicting behaviour using video tracking. This approach is not completely new, as previous research also made attempts to predict behaviour based on video clips (Felsen et al., 2017; Su et al., 2017). Predicting behaviour based on extracted coordinates from human pose methods, on the other hand, was a relatively new task. This research demonstrated that predicting running behaviour based on human pose is a technique that could be applied in the future to predict other human behaviour based on video tracking.

In this thesis, three regularisation models were evaluated to assess the extent to which running performance can be predicted from the coordinates of the starting position. The results from this research were compared with a baseline model and expectations derived from the literature. Although this research used limited data, the elastic net model could predict running time with a mean absolute error of 1.46 seconds. The models made fewer errors than did the baseline model and performed in line with the literature. Elastic net was expected to perform best on this data set, and it did. All models, however, made predictions close to the mean, and they consequently lost predictive power. Elastic net, on the other hand, can predict to a certain extent the end time of runners based on starting positions and could therefore be used in other prediction tasks. Future research should determine whether applying more sophisticated models, such as neural networks, could be used to improve the performance of the models.

References

Brown, T. D., & Vescovi, J. D. (2012). Maximum Speed. *Strength and Conditioning Journal*, *34*(2), 37–41. https://doi.org/10.1519/ssc.0b013e31824ea156

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2016). Realtime multi-person 2d pose estimation using part affinity fields. *Computer Vision Foundation*, 7291–7299.

Cui, L., Bai, L., Zhang, Z., Wang, Y., & Hancock, E. R. (2019). Identifying the most informative features using a structurally interacting elastic net. *Neurocomputing*, *336*, 13–26. https://doi.org/10.1016/j.neucom.2018.06.081

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2018). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *Journal of Sports Sciences*, *37*(5), 568–600. https://doi.org/10.1080/02640414.2018.1521769

Di Prampero, P. E., Botter, A., & Osgnach, C. (2014). The energy cost of sprint running and the role of metabolic power in setting top performances. *European Journal of Applied Physiology*, *115*(3), 451–469. https://doi.org/10.1007/s00421-014-3086-4

Eikenberry, A., McAuliffe, J., Welsh, T. N., Zerpa, C., McPherson, M., & Newhouse, I. (2008). Starting with the "right" foot minimizes sprint start time. *Acta Psychologica*, *127*(2), 495–500. https://doi.org/10.1016/j.actpsy.2007.09.002

Eriksen, H. K., Kristiansen, J. R., Langangen, Ø., & Wehus, I. K. (2009). How fast could Usain Bolt have run? A dynamical study. *American Journal of Physics*, *77*(3), 224–228. https://doi.org/10.1119/1.3033168

Felsen, P., Agrawal, P., & Malik, J. (2017). What will Happen Next? Forecasting Player Moves in Sports Videos. *The IEEE International Conference on Computer Vision*, 3342–3351.

Goss, D. L., & Gross, M. T. (2012). A review of mechanics and injury trends among various running styles. *NORTH CAR- OLINA UNIV AT CHAPEL HILL*.

Guler, R. A., Neverova, N., & Kokkinos, I. (2018). DensePose: Dense Human Pose Estimation In The Wild. *arXiv*.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning* (2nd ed.). -: Springer.

Herrmann, F. R., Graf, C., Karsegard, V. L., Mareschal, J., Achamrah, N., Delsoglio, M., . . . Genton, L. (2018). Running performance in a timed city run and body composition: A cross-sectional study in more than 3000 runners. *Nutrition*, *61*, 1–7.

Luttik, D. T., Van Leeuwen, H., Van Lieshout, C., Oosterwaal, M., Paalman, J., & Van de Water, M. (2018). Moving Towards Better Motion Analysis: Extending Multi-person Pose Estimation to Analyse and Prescribe Running Behavior. -, .

Marafino, B. J., John Boscardin, W., & Adams Dudley, R. (2015). Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk

stratification from nursing notes. *Journal of Biomedical Informatics*, *54*, 114–120. https://doi.org/10.1016/j.jbi.2015.02.003

Nagano, A., Fujimoto, M., Kudo, S., & Akaguma, R. (2016). An image-processing based technique to obtain instantaneous horizontal walking and running speed. *Faculty of Sport and Health Science*, *51*, 7–9.

NOC*NSF. (2017). *Lidmaatschappen en sportdeelname NOC*NSF over 2017*. Retrieved from https://www.nocnsf.nl/nieuws/sportdeelname-in-nederland-stijgt-aantal-lidmaatschappen-bij-sportclubs-blijft-stabiel

Nuijten, W. (2018). Running behavior of children [Dataset]. Retrieved from Not published

Pfister, A., West, A. M., Bronner, S., & Noah, J. A. (2014). Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology*, *38*(5), 274–280.

Singh, D. (2016). Human Pose Estimation: Extension and Application. *International Institute of Information Technology Hyderabad*.

Slawinski, J., Bonnefoy, A., Levêque, J., Ontanon, G., Riquet, A., Dumas, R., & Chèze, L. (2010). Kinematic and Kinetic Comparisons of Elite and Well-Trained Sprinters During Sprint Start. *Journal of Strength and Conditioning Research*, *24*(4), 896–905. https://doi.org/10.1519/jsc.0b013e3181ad3448

Su, S., Hong, J. P., Shi, J., & Park, H. S. (2017). Predicting Behaviors of Basketball Players from First Person Videos. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1501–1510.

Truscott, T., & Belden, J. (2019). Peloton tracking and analysis from the 2016 Tour de France. *Utah State University*.

Van der Meijden, K. (2019). Analyzing sprint features with 2D Human Pose Estimation. -, .

Van Leeuwen, M. (2019). Prediction of running performance from video sequences. -, .

Williams, S., Netto, K., Kennedy, R., Turner-Bryndzej, J., & Campbell, R. (2018). Biomechanical correlates of running performance in active children. *Journal of Science and Medicine in Sport*, *22*, 65–69.

Yamaguchi, K., Luis, M. H. K., Ortiz, E., & Berg, T. L. (2012). Parsing clothing in fashion photographs. *CVPR*.

Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object interaction activities. *2010 IEEE Conference*, 17–24.

Zhang, Z., Tian, Y., Bai, L., Xiahou, J., & Hancock, E. (2017). High-order covariate interacted Lasso for feature selection. *Pattern Recognition Letters*, *87*, 139–146. https://doi.org/10.1016/j.patrec.2016.08.005

Zhao, G., Li, G., & Pietikainen, M. (2006). 3d gait recognition using multiple cameras. In Automatic face and gesture recognition. *7th international conference*, 529–534.

Zheng, S., & Liu, W. (2011). An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Computers in Biology and Medicine*, *41*(11), 1033–1040. https://doi.org/10.1016/j.compbiomed.2011.08.011

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x