Harini Ramaswamy

Mr. Kosek

Mobile Web Senior Research Lab

May 9, 2023

LaxGo: Lacrosse Playing Analysis Through Motion Capture and K-Means Clustering Artificial Intelligence Algorithm

Table of Contents

1. Abstract
2. Introduction
    a. Inspiration
    b. Definition and History of Motion Capture
    c. Idea of Project
3. Background Technologies
    a. HTML (Hypertext Markup Language)
    b. CSS (Cascading Style Sheets)
    c. JavaScript
    d. Python
    e. Vicon System
4. Main Objectives
    a. Road Map of Tasks
    b. Parameters of Success
5. Procedures
    a. Motion Capture Data Collection
        i. Reference to Research
        ii. Technologies, Softwares, and Formulas
    b. Frontend Coding (HTML & CSS)
    c. Backend Coding (Python)

6. Application
    a. Statistical Data Analysis
    b. Combination of Prototype Designs (Frontend and Backend)
7. Conclusion and Discussion
    a. Discussion of Problems in Project
    b. Discussion of Improvements on Project
    c. Concluding Statement
8. References

Abstract

This research project aimed to develop a web application for lacrosse coaches to optimize their team's performance by placing players in positions that maximize their skill sets. The application includes different tabs for accessing raw datasets, viewing basic player position details, and finalizing positions after running the program. The K-means clustering algorithm was selected for this project as it was found to be the most effective at analyzing trends between the data, compared to hierarchical clustering and supervised decision tree algorithms. The results of this study will enable lacrosse coaches to better organize their team, ultimately leading to improved performance on the field. The results of this project are still in development as the final product is being pieced together. The work shows good promise in reaching the final results and the research paper and presentation will be updated accordingly afterwards.

Introduction

Inspiration

I've been playing lacrosse since the fourth grade and have always had a passion for the sport. This year, after getting back to the varsity team at TJ since I last played during my freshman year, I've been inspired to create a product that I can see being used in my real life. My coaches often ask the players for their input on what positions they'd prefer to play, and I thought that if there was a quantitative way to predict the best positions to increase our chances of success, I wanted to make it! With our great success this season and currently advancing to the district semi finals round, I wanted to find out how our team could improve even more. With the interactive data collection method I aimed to make with the Vicon motion capture, it was a creative way to involve my friends and teammates in this project that I am passionate about.

Definition and History of Motion Capture

Motion capture, also known as mocap, is a technology that captures the movements of humans and animals and translates them into digital data for use in video games, movies, and other forms of media. The roots of motion capture can be traced back to the 18th century, but it wasn't until the 20th century that motion capture technology began to take shape with the development of the first electronic sensors that could track movement (Gray). In the 1970s and 80s, motion capture began to be used in film and television, but it was expensive and limited in its capabilities. With the advent of digital technology in the 1990s, motion capture became more accessible and advanced, allowing for more realistic and complex animations (Lum). Today, motion capture is a widely used technology in the entertainment industry and continues to develop and evolve with advancements in computer technology and machine learning.

The Vicon system has been in the motion capture industry for over thirty years after first being introduced in the UK in 1984. Their company offers all of the necessary items for those looking to research or utilize motion capture software, including mocap rigs, cameras, sensors, and even databases and software that enable users to do further research. This was the system that was utilized throughout the course of this research project.

Idea of Project

The objective of this project is to create a web application where lacrosse coaches can organize the players on a team into spots in a play that would optimize their skill sets and use them where they would be most effective. Firstly, the web application is presented with various tabs where the user can access different aspects of the design. There is a tab for accessing the raw datasets and inputting one's own data, another for seeing the basic details of each player's positions on the field, and a final one for the finalized positions after the program has been run. As a result of this study, lacrosse coaches will be able to enhance their team's organization and achieve better on-field performance.

Background Technologies

In this section, background information on each software and technology used through the course of this research project will be provided. In order to gain insight and understanding of the tools, specifics about the history, development, technology, and uses of each aspect has been given.

HTML (Hypertext Markup Language)

HTML, or Hypertext Markup Language, is a critical technology for web development, allowing developers to create and structure content in a way that is easily readable and accessible by both humans and machines. To use HTML effectively, a basic understanding of web development concepts and programming languages is necessary, including HTML syntax and markup, as well as an understanding of web design principles and best practices (S).

HTML documents are created using a text editor or an integrated development environment (IDE), and the resulting code is saved with an .html file extension. In this project, HTML is used to create the user interface and website.

CSS (Cascading Style Sheets)

Cascading Style Sheets (CSS) is a language used for describing the presentation and layout of HTML documents. CSS allows developers to separate the content of a web page from its presentation, making it easier to maintain and update the design of a website. CSS works through selectors, which are used to target specific elements on a web page.

To use CSS effectively, developers must have a strong understanding of the syntax and structure of CSS rules, as well as knowledge of CSS properties and values. They must also be familiar with CSS layout techniques, such as grid and flexbox, and understand how to use media queries to create responsive designs that adapt to different screen sizes.

JavaScript

JavaScript is a high-level programming language that works by allowing developers to add interactivity and functionality to HTML documents, making it possible to create web

applications and user interfaces. JavaScript is used to control the behavior of web pages, including handling user input, modifying the content of a web page, and communicating with web servers. It is also used for animations, form validation, and other interactive features. JavaScript can be used with a variety of web development frameworks and libraries, such as React, Angular, and Vue. In this project, JavaScript is used to branch together the frontend and backend components.

Python

Python is a general-purpose language that can be used for a variety of tasks, including web development, scientific computing, data analysis, artificial intelligence, machine learning, and more. Its simplicity, ease of use, and large number of libraries and tools make it a popular choice for many developers. It also has a vast standard library, which provides a wide range of modules for tasks such as file I/O, networking, and regular expressions.

To use Python effectively, developers need to have a good understanding of programming concepts such as variables, data types, functions, loops, and control structures. In this project, Python is used for implementing an artificial intelligence program using the K-means clustering method which is explained later in the paper.

Vicon

The Vicon system uses infrared cameras to track the movement of reflective markers, which are placed on the subject's body or objects of interest. The system can track the position and orientation of each marker in 3D space, allowing for detailed analysis of movement and biomechanics. To use the Vicon system, a good understanding of motion capture technology and

the principles of biomechanics is necessary. Additionally, there must be ample space to be able to set up and calibrate the cameras and markers and process the captured data using specialized software (Vicon).

Each marker used in the Vicon system is typically a small, reflective sphere or cube, which is designed to reflect the infrared light emitted by the cameras. The markers are placed on specific points on the subject's body or on objects of interest, and their positions are tracked by the cameras as the subject moves through the capture volume. The markers that were placed on the test subjects in this project are detailed in the procedures section of this paper.

Main Objective

Road Map of Tasks

The first approach for this project was solidifying the pathway of tasks that is necessary to follow in order to attain the final product.

Steps:

1. Set up new site and make the simple HTML forms and the various tabs
2. Set up a database on the backend to collect all of the player data that’s being collected
3. Finalize the setup and calibration of the Vicon Motion Capture camera system
4. Finalize the setup and calibration of the Vicon Motion Capture camera system
5. Find out the technical parts of creating human renders and “statues” with the motion capture
6. Create the tab on the website with the plays and the “statues” for demonstration
7. Create a method to sort through the list of statistics to find fastest, strongest, etc.

8. Put the best person for each spot into their respective places on the website tab for ready plays
9. Finalize all three tabs of the app, the choosing method, the “statues” and complete project

Although this was the most preliminary version of the road map, there were various changes that were made along the way. Certain steps were altered as they were not feasible with the materials and time frame provided, and others were altered to match the new aims of the project. These changes are reflected throughout the paper and elaborated on in the conclusion and discussion section of the paper.

Additionally, there was initial research towards the beginning of the project about existing solutions to this topic or other resources and softwares that provided users with tailored plays matching their players abilities. Although there are simulators for basketball and football, there does not seem to be one for girls lacrosse that is easily accessible. To continue, any websites that are currently available are mainly simulators that show the playing field and not programs that input players into them.

Applications

Statistical Data Analysis

Unsupervised learning algorithms are frequently utilized for recognizing patterns and trends within unprocessed datasets or to group similar data into a predetermined number of categories (Seldon). Additionally, it is commonly employed as a technique in the preliminary exploration stage to enhance the comprehension of the datasets. The model is capable of efficiently handling massive arrays of data, without requiring human supervision. Hence, it is

well-suited for providing insights into undiscovered trends and relationships within the data itself.

In reference to the Lax-Go project, a K-means clustering algorithm is the best choice (compared to hierarchical clustering or a supervised decision tree algorithm) because it can analyze the trends between the data. For example, it can find the association between taller and heavier players and their success performing as a goalie. Similarly, it can understand the importance of faster sprinting and shooting speed for attackers who are meant to quickly run in and shoot. However, these placements are not simply averaged and inputted into their spots. With the K-means clustering, the accuracy of placements increases and the algorithm learns how to analyze the statistics given.

Procedures

Motion Capture Data Collection

a) Reference to Research

In this project, motion capture will be used to analyze general player movement patterns and, more specifically, measure the shooting speed of each player. Recently, the Vicon motion capture system has been very helpful for gait analysis and kinematic gait patterns in particular circumstances (Yeung et al., 2021). An open-source study by Goldfarb et al., (2021) used the Vicon motion capture camera system Toolkit to study gait analysis and overall demonstrate the abilities of the Vicon package. Alongside providing technical information about the Vicon motion capture system, they provided snippets of code and explained the function of each part. They use an open-source framework for ingesting, parsing, and analyzing Vicon mocap data and performing gait analysis. Their Vicon Toolkit provides tools to work directly with the marker

data and other Vicon Nexus system outputs. This was particularly relevant to this research as it presented starting points for how to manage the software and code the cameras to collect my desired data values.

One such study performed by Moore et al., (2022) compared the performance of rugby players between different playing positions on the field. The researchers collected data from pre-season rugby players that were all subject to the same training conditions and were categorized according to the field and gym sessions to make their observations more specific. This provided research for the beginning of this project by helping create the plan of how to identify the ideal player for each position as well as the mathematical side of the project.

b) Technologies, Softwares, and Formulas

The reflective markers that are necessary in order to track movement were placed on each subject as shown in Figure 1 below. This figure, taken from ResearchGate, shows the specific names and positions of each marker placement. This was done by velcro straps that were secured around each subject. Additionally, they held a T-shaped wand with 7 additional markers that served as the lacrosse stick in this project. Holding this bat, each subject could measure their shooting speed and could follow through with the proper motion capture measurements.

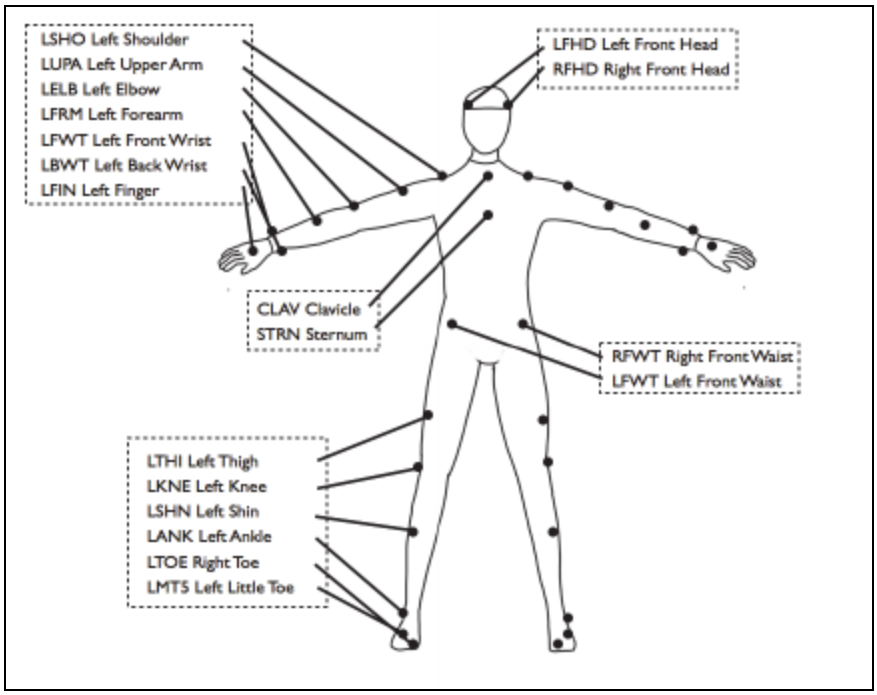

Figure 1

Frontend Coding (HTML & CSS)

The frontend experimentation was done in various ways. First, the assisted website builder Webflow was used to gain an idea of the layout and elements necessary for the frontend. As seen in Figure 2 below, the orientation of the website and individual elements were planned

out according to desired size and position.

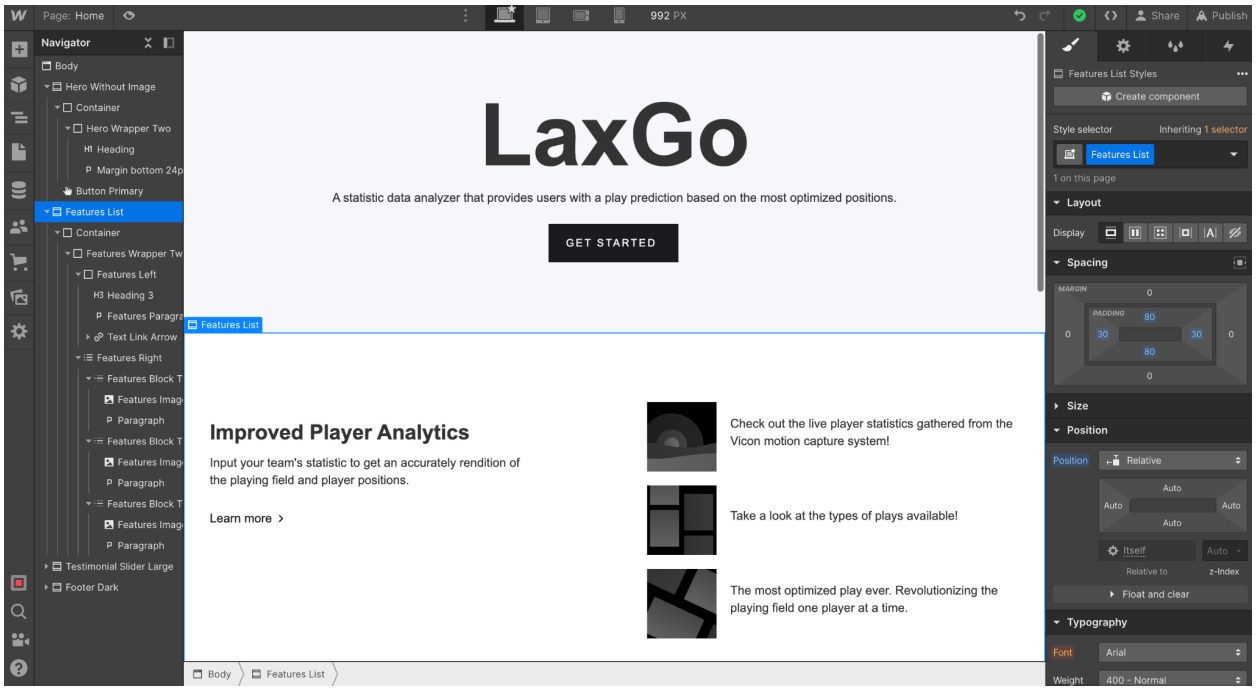

Figure 1

For example, planning the layouts of the tabs, the sizing of the header, and all the other small aspects was done with the help of Webflow. Additionally, it was an easy way to have an inspiration for the actual HTML coding. The first aim was to utilize the Webflow design as is and code it to attach to the frontend, but after encountering several problems, the aim switched to copying the layout through HTML. Next, HTML and CSS templates were used to build the website from scratch. The preliminary designs sufficed for testing purposes and ended up being developed to appear as the design above.

Backend Coding (Python)

This project is aimed to use a K-means unsupervised machine learning algorithm for recommending positions to each player.

To begin, the Pandas package was imported, an open-source Python package that is predominantly utilized for data analysis, data science, and machine learning operations. It is constructed on a foundation package called Numpy, which offers assistance for arrays that have multiple dimensions. Secondly, the player data was loaded into a CSV file in order to select the features to use for the clustering. The parameters selected for this file, as shown in the code snippet below, were height, weight, sprinting speed, and shooting speed. Height, weight, and sprinting time were measured outside of the project, and shooting speed was acquired through motion capture. An important feature of K-means clustering is ensuring that the data is standardized/scaled to make it compatible with one another (University of Cincinnati). Standardization means changing the mean to equal 0 and standard deviation equal 1. These first few steps are shown below.

```
import pandas as pd
import numpy as np
from sklearn.cluster import KMeans

# Load player data from a CSV file
player_data = pd.read_csv("player_data.csv")

# Define the features to use for clustering
features = ["height", "weight", "sprinting_speed", "shooting_speed"]

# Normalize the data to have mean=0 and variance=1
normalized_data = (player_data[features] - player_data[features].mean()) / player_data[features].std()
```

The k-means clustering was applied with k = 3 with cluster labels being assigned to each player. This means that all the players that were measured were divided into the three clusters that the program is working with. We include the cluster assignments to the player information and determine the positions of the players as well as the ideal number of players for each position. For example, there is only one goalie, three midfielders, four attackers, and four defenders.

```
# Apply k-means clustering with k=3
kmeans = KMeans(n_clusters=3, random_state=0).fit(normalized_data)

# Get the cluster labels for each player
cluster_labels = kmeans.labels_

# Add the cluster labels to the player data
player_data["cluster"] = cluster_labels

# Define the player positions and the optimal number of players per position
positions = {"attacker": 4, "midfielder": 3, "defender": 4, "goalkeeper": 1}
```

Afterwards, we group the players according to their respective clusters and allocate them to their optimal positions. We sort the players in each cluster for a specific position based on the most significant feature (e.g., goalkeeper's height) and select the top few players according to those demographics. The maximum number of players cannot exceed the allocated amount, which is why the top few are analyzed and selected. Lastly, the data is displayed in the console.

```
# Group the players by cluster
cluster_groups = player_data.groupby("cluster")

# Place the players in their optimal positions
for position, num_players in positions.items():
    position_players = pd.DataFrame()
    for _, group in cluster_groups:
        # Sort the players in the cluster by the feature that is most important for the position
        sorted_group = group.sort_values(by=features.index(position.split("_")[0]), ascending=False)
        # Take the top num_players players from the sorted group
        position_players = pd.concat([position_players, sorted_group.head(num_players)])
    # Print the players in the position
    print(f"{position.capitalize()}s:")
    print(position_players[["player_name", "cluster"]])
```

This version of the code inputs the result into the console with assignments specified to each player in each cluster, however, the final product places each name in the play simulator to show their role on the field in action.

Conclusions

Discussion of Problems Faced

There were various problems that arose during the course of this project. Firstly, deciding what algorithm approach to take was the first important step of the project process. The choices varied between supervised and unsupervised machine learning algorithms and the subcategories within each type. Supervised algorithms require human insight to accurately label data ready for supervised learning. This creates a tedious and intensive work process to standardize the data beforehand. Supervised algorithms are also generally used to classify unseen data into established categories and forecast trends and future change as a predictive model, like weather or music chart predictions. After much trial and error with a supervised decision tree algorithm, it was clear that it was not the most efficient choice for the project. After this, the approach was rethought and switched to the unsupervised algorithm.

Additionally, there was the constraint of materials that was present throughout the project. Although there was ample time since the beginning of the year, there was an initial slow down during the brainstorming and learning portions of the year. After that, the necessary materials and knowledge needed to be gained before proper progress could be made.

Discussion of Improvements on Project

Future improvements can also be made to this project, including doing more in-depth calculations and research with the Vicon motion capture system, as well as looking into more sophisticated and possibly complicated algorithms for the optimization process. More specifically, there could have been more solid experimentation and implementation of the motion

capture system. Due to certain restraints, the project is still in progress but shows consistent promise of reaching the end goals.

Concluding Statement

This research will allow us to creatively explore a new field of sports and motion capture by combining algorithmic predictions with statistical analysis. I want somebody to gain insightful information by using my web app and be able to use a cohesive shortcut to making plays for their sports team. More specifically, through my years of playing lacrosse, I hope to see my web app be in use for real people in my life!

Works Cited

Goldfarb, N., Lewis, A., Tacescu, A., & Fischer, G. S. (2021). Open source Vicon Toolkit for motion capture and Gait Analysis. Computer Methods and Programs in Biomedicine, 212, 1-13. https://doi.org/10.1016/j.cmpb.2021.106414

Gray, Ali. "A Brief History of Motion-Capture in the Movies." *IGN*, 11 July 2014, https://www.ign.com/articles/2014/07/11/a-brief-history-of-motion-capture-in-the-movies. Accessed 9 May 2023.

Lum, Rachel. "A Brief History of Motion Tracking Technology and How it is Used Today." *Medium*, 9 January 2019, https://medium.com/@lumrachele/a-brief-history-of-motion-tracking-technology-and-how-it-is-used-today-44923087ef4c. Accessed 9 May 2023.

S, Astari. "What Is HTML? Hypertext Markup Language Basics Explained." *Hostinger*, 2 May 2023, https://www.hostinger.com/tutorials/what-is-html. Accessed 9 May 2023.

Seldon. "Supervised vs Unsupervised Learning Explained." *Seldon*, 16 September 2022, https://www.seldon.io/supervised-vs-unsupervised-learning-explained. Accessed 8 May 2023.

University of Cincinnati. "K-means Cluster Analysis · UC Business Analytics R Programming Guide." *UC Business Analytics R Programming Guide ·*, https://uc-r.github.io/kmeans_clustering. Accessed 8 May 2023.

Vicon. "Confluence Mobile - Documentation | VICON." *Confluence Mobile - Documentation | VICON*, https://docs.vicon.com/pages/viewpage.action?pageId=107483982. Accessed 9 May 2023.

Yeung, L.-F., Yang, Z., Cheng, K. C.-C., Du, D., & Tong, R. K.-Y. (2021). Effects of camera

viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2. Gait & Posture, 87, 19-26. https://doi.org/10.1016/j.gaitpost.2021.04.005