Drafting Detection in Triathlons

CS701 Final Project Report

Zeke Mostov
Middlebury College
Middlebury, VT
emostov@middlebury.edu

Jenny Melcher
Middlebury College
Middlebury, VT
jmelcher@middlebury.edu

ABSTRACT

Drafting during the cycling portion of triathlons is illegal according the USA Triathlon rules as drafting benefits the drafter and gives them an unfair advantage over the rest of the competitors. Drafting rules are outlined in the USAT guidelines and should be enforced by referees and officials, but officials are usually focused on the leaders of the race and many competitors ride unchecked. This program was developed as a proof of concept that drafting can be detected by examining GPS data that racers upload to Strava after an event. This program analyzes these GPX files to determine whether drafting occurred, by which rider, and for how long. This method of policing can check any athlete who records and uploads their race. Additionally, this program creates a map visualization so that the user can visualize the distance between the bikers throughout the ride. In the future, this technology could be used to monitor the entire field of athletes if race directors provided accurate GPS recording devices to racers.

Keywords

Drafting; Cycling; Triathlon; GPX; GPS; Track matching

1. INTRODUCTION

Drafting during the biking segments of triathlons is illegal and creates a huge disparity in ease of competition and thus fairness between those who draft and those who do not. While not athletes are competing at a professional level, fairness still matters for awards within each individual age group. To avoid drafting penalties, athletes are required to stay outside of the drafting zone of another biking during triathlons except for a 15 second window during which they can enter and pass the leading biker[10]. This biking zone is an area 25 feet long behind and 6 feet wide around the leading rider.[10] See Figure 1.

Athletes cover over 100 miles in the cycling leg of an Iron-Man and unfortunately, referees are few and far between at triathlons and are usually focused on the few elite competitors. Thus, most drafting violations go unnoticed and unpenalized. A more efficient technological solution should be implemented to monitor cheating behavior via drafting during the cycling segment of triathlons. This project proves that drafting can be detected using GPS data that users upload to Strava, a social-media application that logs workouts and routes. This project shows that comparing GPS data can serve as a more efficient and fair method of rule enforcement during the cycling portion of Triathlon races.

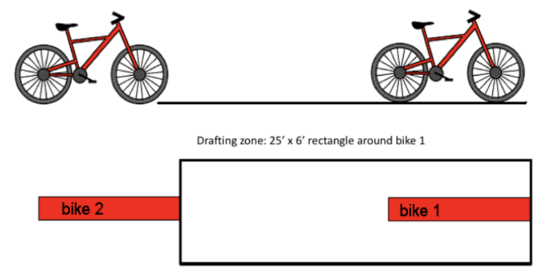
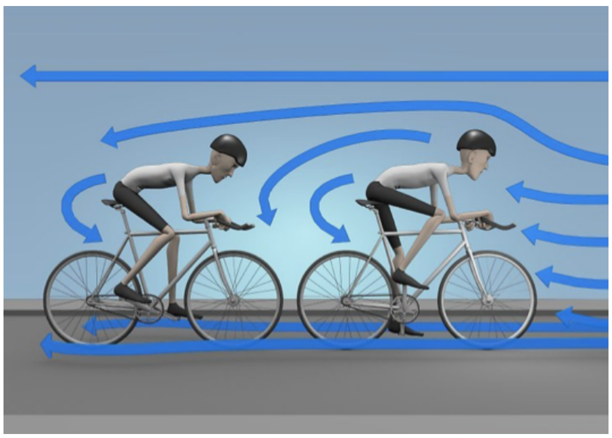


Figure 1: Drafting Zone

2. PROBLEM STATEMENT

A triathlon is a race consisting of a swimming portion, a cycling portion, and a running portion in that order.

Triathlons are competitive by age groups so that while there is an overall winner, there are winners for each age group in each gender category as well. Triathlons cover a variety of distances from sprint races covering just a few miles, to "Ironman" races covering 112 miles total. With racers spread out along this length of course, over three different disciplines, rule enforcement can be a challenge in triathlons. Generally, penalties for rule violations are given as amounts of time added to a racer's finish time, slowing their total race time, or disqualification. Drafting during the cycling portion of the race is a rule violation that often results in a time penalty when caught[10]. Drafting is illegal in triathlons because these races are intended to be individual competitions and each racer should be competing individually. Drafting gives one rider an advantage and is considered a form of external aid, so the rider is not truly competing individually. Wind resistance accounts for the majority of the drag that a rider experiences, typically about 80 percent of drag[8]. Drafting another rider reduces the wind resistance that the drafter is experiencing, allowing them to cycle with less effort than other riders who are experiencing wind resistance. Drafting also allows the drafter to ride in a low pressure area behind the leading rider so that they experience some pull from the rider ahead of them. These two benefits of drafting allow the drafter to expend much less energy than other riders while maintaining the speed of the leading rider. While drafting clearly provides an unfair advantage, it often goes undetected by race officials. Typically, race officials monitor the course on motorized vehicles and they report the bib numbers of racers they find in violation of rules. Unfortunately, race officials are fairly sparse, particularly for longer races. It is impossible to monitor every racer over the whole course with this system. This means that drafting is likely going undetected. Triathlon rule enforcement needs an update. Increased use of technology for violation detection could increase fairness among the field of competitors within triathlon races. With riders using GPS tracking devices to record their swims, rides, and runs, the data necessary to analyze these efforts already exists. In fact, many of these competitors are also uploading their own data to Strava. Strava takes GPS data from whatever recording device an athlete is using and converts it to a GPX file so that each athletic endeavor that an athlete records becomes a "post" just like a social media post on their Strava news feed with an attached GPX file. Analyzing files uploaded to Strava is a good starting point for rule enforcement via athlete's own recordings during triathlons.



3. RELATED WORK

One of the biggest inspirations for this project was a Strava Labs creation called FlyBy[7]. After a Strava user uploads a Strava FlyBy detects whether other users completed the same or a similar course around the same time, accumulates these files, and creates an animated map where a user can watch their progress compared to those nearby users. The FlyBy feature shows how riders may be incredibly close together at some parts of the course but spread out during other parts. This visualization prompted us to detect drafting by comparing rides in a similar manner. Additionally, there was also some prior work from individuals taking GPX files from strava and manipulating them on their own to produce visualizations of their ride, or to analyze and manipulate the data in other ways[6][2]. Many of these projects included work calculating distance between coordinate points given in the GPX files using the Haversine formula. We followed a similar procedure to calculate distance between different riders rather than distances between points of the same ride. For the map visualization part of this project, prior work using GeoJSON to show changes over time was beneficial in implementing an animated map[4]. While similar methods and ideas were implemented in previous work, not many addressed drafting detection. Only one article suggested detecting drafting, but by using mobile smartphones rather than GPS data[9].

4. METHODS

- Data

  To start, this project required GPX files taken from Strava. Strava allows users to download GPX files of their own activities, but not those of other users. We collected our own data by riding bikes at varying distances apart and we also gathered data from old rides and reached out to coaches and friends to collect GPX files from them as well. During this process, we were able to differentiate between the accuracy of different recording devices and quickly realized that not all GPX files are accurate enough to detect drafting. Specialized bike computers such as the Wahoo Edge or the Garmin Edge 520 provided extremely detailed data and many more coordinate points than some common triathlon watches such as the Forerunner 920XT. Thus, we narrowed down which devices could be used to detect drafting and eliminated data from these less accurate devices.

- Parsing and comparing GPX files

  The whole project was coded in Python. The GPX files were parsed into separate data frames using Gpxpy so that each data frame included columns for seconds, longitude, latitude, altitude, and time. These two data frames were merged based on the seconds column so that each rider had coordinate points for each time point in the data frame[5]. The distance between the point for rider 1 and the point for rider 2 at every time point was calculated using the Haversine formula. This formula calculates the great circle distance between two coordinate points. Although the distances might have been small, it was still important to account for the curvature of the earth[1]. The distance calculated between two points at the same time stamp was used to determine whether drafting behavior occurred. If the distance between the two points was less than 25 feet for more than 15 seconds, it was likely that there was drafting.

- Analysis of file quality

  Due to inconsistency in data collection intervals, some GPX files are not conducive to producing legitimate results. All GPS recording devices analyzed by us did not record GPS coordinates if connection was poor. Poor GPS connection can result in data gaps of one second or more. In order to indicate to the user file quality,we came up with a simple algorithm that indicates a files poor quality if more than 1 percent of moving time is not accounted for with GPS coordinates.

- 3 Dimensional Haversine Formula

  In order to calculate the distance between points we used the Haversine formula[1]. Due to the round shape of the earth, simply calculating a straight-line distance has the potential to lead to significant inaccuracies. We used a two dimensional haversine formula of latitude and longitude for all distance calculations. We did not use a three dimensional haversine calculation that would have included altitude because all the devices tested by us had inconsistent altitude outputs. We observed that this is likely due to the fragility of the barometric mechanisms used by the devices to calculate altitude.

- Setting Drafting Tolerance Based On Real Word Data

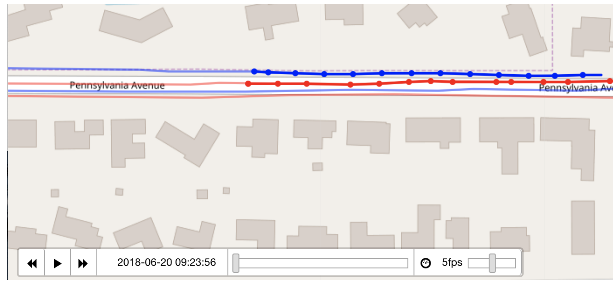
  In order to conduct track matching the program merges dataframe of each ride based on time. Distance between both riders is then calculated at each time interval. If the distance between two riders is 25 feet or less, the program then does further analysis to determine which rider is in front. The tolerance of 25 feet was set by extensive testing on real world data as well as USAT regulations[10]. We created data sets with speeds and route curvature similar to popular triathlons. Using different devices, test runs were connected at various distance intervals to then compare against how the data reflected the distance. We then obtained data from coaches of triathletes doing various exercises at different drafting distances in order to expand on our own data set. We found through this data collection and analysis that a tolerance of 25 feet was appropriate and erred on the side of caution and provided ample room to evaluate and red flag potential drafting participants. We wanted to maximize the range to bring attention to potential cheating for further analysis and evaluation, but we do not claim to be a final word. In a real world deployment it might make more sense to allow more flexibility to minimize those falsely accused of cheating.

- Mapping

  To create an animated map using Python, folium was used for base maps. These are really leaflet maps which can be coded in Python using folium[3]. The individual tracks of the riders were plotted on to the maps through folium. Then, GeoJSON was used to create the animation[4]. Each ride was broken up into smaller segments of three points so that the animation flowed more smoothly. Users can zoom in and out of the map and control the speed at which the rides play.

5. RESULTS

The final product of this project is a command line program that asks a user to input two GPX files to analyze. Once input, the program will determine whether the quality of the data is good enough to determine drafting. If so, the program will offer to output the total time and distance of each ride. Next, the program will offer the animated map file as an html file that can be opened by the user. Finally, the program will print whether each rider broke the drafting



rule during the ride. Finally, the program will offer to print each instance of drafting which shows which rider was leading and which was following and for how many consecutive seconds drafting occured. When opened, the animated map shows a track that each rider followed, and it starts automatically playing the animation of the ride so that the user can watch the riders over the entire course.

6. DISCUSSION

This project shows that drafting can indeed be detected by comparing GPX files taken from Strava. Not only can the program determine who was following who, but it can also determine how long drafting occured for so that the user knows how certain it is that USAT drafting rules were violated. This project serves as a proof of concept that GPS data can be used to detect drafting and that a more efficient and fair form of rule enforcement can be implemented using technology. In the future, race directors might provide each racer with a GPS recording device to collect and analyze this data in order to assign penalties at the end of the race. Right now, timing chips are given to each racer and returned at the end of each race. GPS recording devices might work the same way but collect more detailed data. One limitation of this study that may help inform the implementation of this plan is that the GPS recording device must be of high quality. Certain watches and bike computers such as the Garmin 920XT does not record points frequently enough to reliably detect drafting. Further, one limitation of our project that should be improved upon is the run time that the analysis of the data frames necessitates. To compare multiple files or an entire competitive field, a different data structure may be necessary. One possibility for this could be KDTrees or hierarchical clustering. A revised version of this project may also include NumPy arrays to speed up operations. While the data frames allowed easy visualisation and understanding of the data, NumPy arrays could be implemented instead once the clear visualisation of the data is not necessary.

7. ACKNOWLEDGMENTS

Thank you to Professor Grant for his instruction in CS701 and the opportunity to take on this project. We would also like to thank Nate Wilson for providing us with some additional data sets to test our project on.

8. REFERENCES

[1] Calculate distance and bearing between two latitude/longitude points using haversine formula in JavaScript. URL: https://www.movable-type.co.uk/scripts/latlong.html.

[2] Exploring GPX files. URL: https://ocefpaf.github.io/python4oceanographers/blog/2014/08/18/gpx/.

[3] Folium âĂŤ folium 0.8.3 documentation. URL: https://python-visualization.github.io/folium/.

[4] Jupyter notebook viewer. URL: https://nbviewer.jupyter.org/github/python-visualization/folium/blob/master/examples/Plugins.ipynb.

[5] Numpy and scipy documentationÂű. URL: https://docs.scipy.org/doc/.

[6] qgis - matching GPS tracks. URL: https://gis.stackexchange.com/questions/81551/matching-gps-tracks.

[7] Flyby. 2018.

[8] Andy G. Edwards and William C. Byrnes. Aerodynamic characteristics as determinants of the drafting effect in cycling. *Medicine and Science in Sports and Exercise*, 39(1):170–176, 2007.

[9] Iztok Fister, Dusan Fister, Simon Fong, and Iztok Fister, Jr. Widespread mobile devices in applications for real-time drafting detection in triathlons. 5(3):310–. 310. URL: http%3A%2F%2Flink.galegroup.com%2Fapps%2Fdoc%2FA350977900%2FAONE%3Fu%3Dvol_m58c%26sid%3DAONE%26xid%3D56e80e6f.

[10] USA Team. Article v: Cycling conduct, 2018. [Online; accessed 11-May-2019]. URL: https://www.teamusa.org/USA-Triathlon/About/Multisport/Competitive-Rules#Article_V.