REU Final Reports

Research Experiences for Undergraduates (REU) on Computational Modeling Serving the City

8-20-2021

Agent-based Activity Generation of Runners for City Infrastructure Planning

Quang Le
Portland State University

Citation Details

Le, Quang, "Agent-based Activity Generation of Runners for City Infrastructure Planning" (2021). *REU Final Reports*. 24.
https://pdxscholar.library.pdx.edu/reu_reports/24

Agent-based Activity Generation of Runners for City Infrastructure Planning

Quang Le

Abstract: Since the pandemic started, many gyms and indoor classes have been shut down to mitigate the spread of Coronavirus. Many people have been forced to get onto pavement streets to get some fresh air while running around and coping with the new reality. There are over 60 million runners in the U.S., and that number is growing rapidly during this time without any sign of stopping once life gets back to normal. In this project, an agent-based model has been developed to generate a set of routes that runners would take in their daily run in a neighborhood of Portland city. The street network and data from OpenStreetMap are downloaded by the OSMnx package in python before being transformed to a grid world and having runners roam around with several preferences of behaviors. A heat map is generated at the end to compare to the Strava heat map which uses aggregated, public data. This approach is the first step to get to know more about runners' behavior and transferable to other neighborhoods or cities as well.

Keywords: runners, agent-based modeling, running simulation.

1. Introduction

Runners make up a crucial part of every community along with a large number of pedestrians and cyclists. Throughout the pandemic, many people, some might be for the first time, have been sent onto the street for a run and getting some air since many gyms and indoor classes have been shut down [1]. According to the data from a fitness-tracking app, Runkeeper, the number of runners is increasing rapidly since the virus took over our world without any sign that running will be stopped once we get back to normal [2], [3]. In this project, an agent-based model is developed to learn more about runner's behaviors and to generate a set of routes that runners are using in their daily run.

Running produces a tremendous benefit to all people both physically and mentally. An increase in physical activity like running has also helped many smokers, who have tried to get rid of this habit before but fail, gradually identify themselves as a runner and decrease their smoking identity or even quit smoking [4]. There are over 60 million runners in the U.S., but to the best of our knowledge, runners are never considered as part of any city planning. There is not a single model about runners compared to an increasing number of models dedicated to pedestrians or cyclists. By using the MESA framework in python, this model will be the first one ever to have a number of agents follow some specific behaviors of runners in real life and generate a set of routes along with the level of being used shown in the heat map. This automatic generated heatmap is then compared to the Strava heat map, which uses public aggregated data and is updated monthly, to validate the efficiency of our model at the end. The complex behavior of runners is encoded in this model by using the state machine technique.

2. Literature Review

There is only one study about recreational runners and their environmental preferences, which was published this year. In this study, researchers found that people run an average of 3.9 days/week with over 50% of people in the age group from 19 to 54 running less than 30km per week compared to only 41% of people aged over 54. This study also suggests that recreational runners often prefer to run on asphalt/paved surfaces, and they also prefer routes that are well-lit, quiet, and have few intersections that could potentially interrupt the continuous run. Men seem to value the continuity of routes and having pedestrian safety features in place, whereas women value more access to green and blue spaces. The study also points out that women worry about their safety more than men which is why a higher percentage of women want adequate lighting and high visibility. Amenities on the running path (public toilets, water fountains,..) are also important, especially for those aged 65+ with the highest rate in this section [5].

In another study, researchers point out that pedestrians and cyclists have a significant risk of being injured or killed in a collision with a vehicle at night. Therefore, a number of cyclists and runners were recruited to learn about their strategies and route choices in order to increase their visibility and safety under low light conditions, i.e. at dawn, dusk, or nighttime. Routes that are brightly lit were revealed as prefer choices of participants to increase their conspicuity. Moreover, runners state that they'd rather run alongside main roads to encounter more traffic than running on quiet roads, which might put their safety at risk. Runners also have more concerns about visibility if they run on the road of the routes without a sidewalk [6]**.**

A number of agent-based models has been developed to investigate the route choice and preferences of cyclists with the purpose of either commute or leisure. Among those, one model was created to incorporate all related attributes of infrastructure to see how they would affect cyclist's decisions in choosing routes for commute in Berlin. Several factors, which were found as an agreement from many studies before, were taken into consideration in this model such as travel time, route length, slopes, pavement surface conditions, and riding smoothness on the road. After testing this model in a small illustrative scenario with 9 separate links of a simple graph, this model was then simulated on a real-world scenario in Berlin through nodes and links from MATSim networks of this city in Netlogo. Through the simulation, it shows that attributes of the infrastructure including travel time, slopes, cycling infrastructure, and pavement surface have an impact on the decision-making behaviors of cyclist [7].

Another earlier analysis was conducted to study the bicycle route choice preference in Texas, U.S. with the data from a web-based stated preference survey of Texas bicyclists. The participants in this analysis including both commuter bicyclists and non-commuter bicyclists with the purpose of exercising, recreation, or visiting friends or family which are similar to the group of our study. All of the attributes were chosen for examination in this analysis including bicyclist characteristics such as demographic, on-street parking, type of bicycle lane, width, and continuity of it, roadway grade, the number of stop signs, red lights, and cross streets, the volume of traffic, speed limit, and travel time. After doing analysis, the results show that bicyclists prefer minimal parking to no parking on the route if possible. They also have a preference for cycling on continuous bicycle facilities, lower traffic volume, and lower roadway speed limit with fewer stop signs, red lights, and cross streets. Moderate hills are another preference over flat terrain, and commuting bicyclists have a great sensitivity to the travel time [8].

3. Methodology

3.1 Simple Virtual World

3.1.1 Virtual environment

A virtual environment is first created by using the MESA framework which is a modular framework for building, analyzing, and visualizing agent-based models in Python 3 [9]. In this virtual environment, a grid world of 50x50 cells is initialized to represent a simple map with yellow cells as roads, orange as trails, bright green as grass, dark green as forests, and red cells as houses. Each agent (runner) in this model is presented as a black dot on top of the background cells with their own type of behavior to follow while running (Fig. 1). Every type of behavior in this virtual environment has its own state machine which could be assigned to agents if some specific conditions are met.

3.1.2 Types and preferences

There are two types of behaviors in this virtual environment. The first type would run only on the road as straight as possible and avoid running over repeated routes. When it comes to an intersection, it always prefers to run on roads that have not been used in the order of straight, right, and left. If all the available roads around an intersection have been used, the runner would have to choose the first road entering this intersection in order to get out of this possible loop.

Whereas, the second type of runner starts on the normal road but then tries to get to the closest trail as soon as possible and spends most of its time running on the trail. From the initial point on the normal road, this runner relies on a state from the corresponding state machine to get to the trail as soon as it can. This second type has similar preferences with type one when runners are on an intersection of the trial as they have to consider choosing a road that hasn’t gone yet in the order of straight, right, and left. If all the roads have been used, this second type will choose randomly one road instead of choosing the first road entering this intersection like type one. All the agents with either type one or type two in this model have a distance goal as well. As soon as their distance has gone is over the distance goal and stands at an intersection, runners would use the same state to follow the closest path to get back home. The state machines for type 1 and type 2 are presented in Fig. 2 and Fig. 3.

3.1.3 Heatmap

A heatmap is automatically generated at the end of each simulation by matplotlib package in python to show all the routes that are used and their levels of being used by runners. The busiest routes are represented with the brightest color, whereas, routes with lower activity are illustrated with a darker color or black color if there is no activity during the simulation.

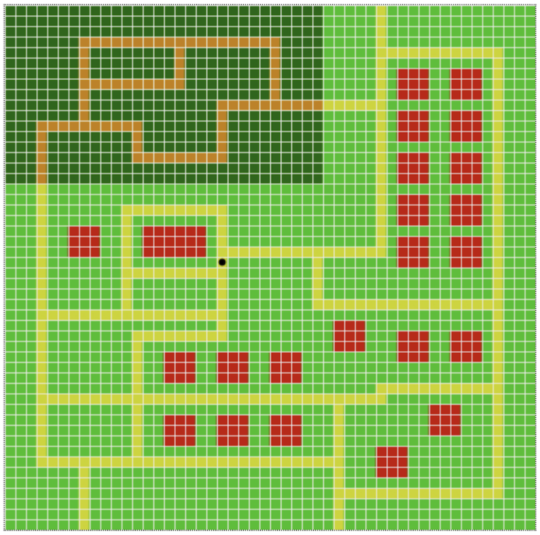


Fig. 1: Simple Virtual World created by MESA framework

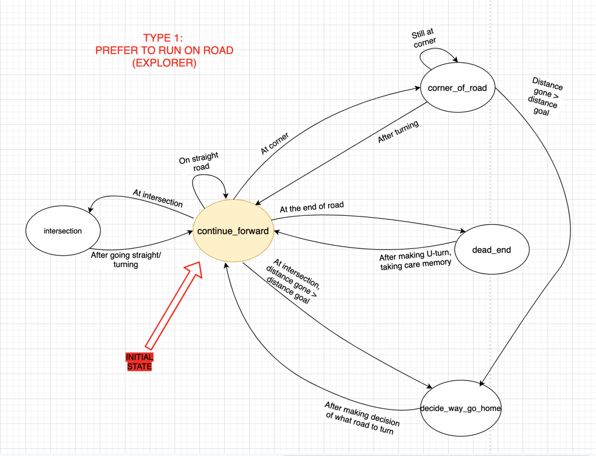


Fig. 2: State machine for type 1.

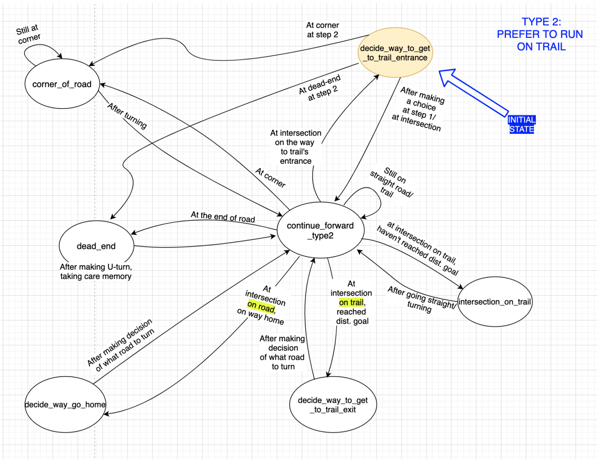


Fig. 3: State machine for type 2.

3.2 Virtual city

3.2.1 Networks and data from OpenStreetMap

To create a grid world for a city or neighborhood, specifically Brentwood - Darlington neighborhood in Portland, that is similar to the grid world from the virtual environment, the OSMnx package is used to download, analyze, and visualize the street networks from the OpenStreetMap (OSM) data [10]. From every network downloaded from this OSMnx package, its data consists of all the nodes (intersections or dead ends) and edges (sections of roads) information that makes up the entire street network. Importantly, nodes contain the real coordinates of latitude and longitude while edges contain the starting and ending nodes. Several other tags are useful when modeling this area including *street_count* - the number of edges around, and *highway* - a type of crossing if exists at that specific node which could be *'traffic_signals,' 'stop,' turning_circle,'* or *'crossing.'* Besides the starting and ending nodes in edges data, other useful tags are *name* - the name of the street that edge lies, *highway* - the type of that edge or road in real life, *oneway* - either that road is one way or not, and *length* - the total length of that edge.

3.2.2 Highway of edge

It is important to understand several highway tags of the edges or types of roads from this area that would be transferred from the network to a grid world:

highway=residential: roads that are around residential areas.

highway=tertiary: roads commonly connect minor streets (residential streets in this case) to the major roads.

highway=secondary: larger roads than residential and tertiary roads with moderate traffic.

highway=primary: larger roads than the previous 3 types of roads. They could be considered as major routes linking to different towns/neighborhoods with more than 2 lanes and a lot of cars.

highway=footway: roads could be accessed by pedestrians, and in this area, they are roads around or inside of parks.

3.3.3 Transfering network to a grid world

After downloading nodes and edges data by using the OSMnx package, all the nodes must be represented by a cell, and edges must become those cells connecting between two nodes or two cells in this grid world. The coordinate (x, y) of each cell in the grid world has to be a pair of integer numbers greater than 0. Therefore, from the nodes' data, the minimum latitude and longitude have to be found to set them as the smallest y and x in the grid world as (0,0) or (1,1) to have some space outside. The maximum latitude and longitude also have to be found in order to get the gap between min. and max. in latitude and longitude. With the network streets of Brentwood - Darlington neighborhood, those numbers are found to create a grid world of 202x84 with the distance gap of x and y between cells is 0.000187. Each cell is responsible for a range of latitude and longitude, and once that latitude and longitude of nodes fall within that range of a specific cell, the cell is assigned to be a node or intersection/dead-end in the grid world. After getting all the nodes appeared in the grid world, an algorithm is applied to create those cell connections for every edge. A completed grid world for Brentwood - Darlington neighborhood is created and shown in Fig. 5.

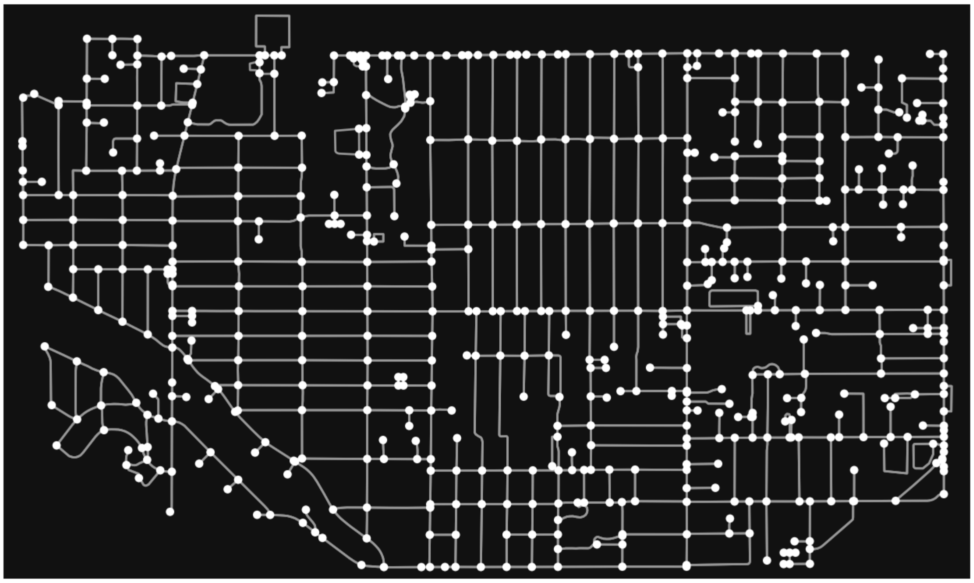


Fig. 4: Street network from Brentwood - Darlington neighborhood downloaded by OSMnx.

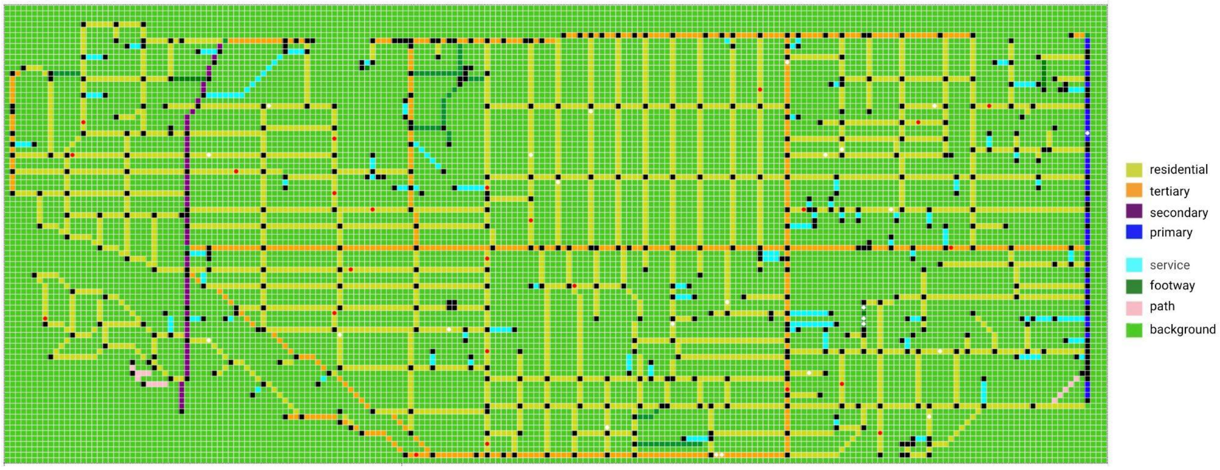


Fig. 5: Grid world from Brentwood - Darlington neighborhood.

3.3.4 Runners

According to the data from Statista.com, the population of Brentwood - Darlington neighborhood is over 17.000 people with the ratio of men vs women is 1:1. Therefore, this model initializes 250 male agents and 250 female agents that add up to 500 agents in total. Besides gender attribute, each agent also has a fitness level attribute with either *Low, Moderate, High,* or *Very High.* An average person could run 0.3mi which is about 5km in 30 minutes, so the people with moderate fitness level are set to run from 5 to 7.5km for each time. From the study about recreational runners, people run on average 3.9 days/week [5], so this model is simulated with 7 days, and runners will have to pick 4 days in a week to go for a run. Therefore, the weekly distance running range for the group of people with moderate fitness level is set from 20 to 30km.

The percentage of men and women running with a specific range of distance per week (Table 1) from the study [5] is also used as an assumption in this model. However, with the weekly distance range from 51 - 60km, the percentage is low for both men and women, and the range over 60km is very high and could be overrepresented for a general population. Therefore, these 2 weekly distance ranges are dropped, and the remaining weekly distance ranges are categorized as Low with 5-20km, Moderate with 21-30km, High with 31-40km, and Very High with 41-50km (Fig. 6)

Km Distance/Week	Men (%)	Women (%)
<10	4.6	11.6
10-20	15.7	26.6
21-30	17.5	24.5
31-40	17.8	15.2
41-50	19.3	13.7
51-60	8.1	3.7
60+	17	4.8

*Table 1: Percentage of men and women with corresponding weekly distance range [5]***.**

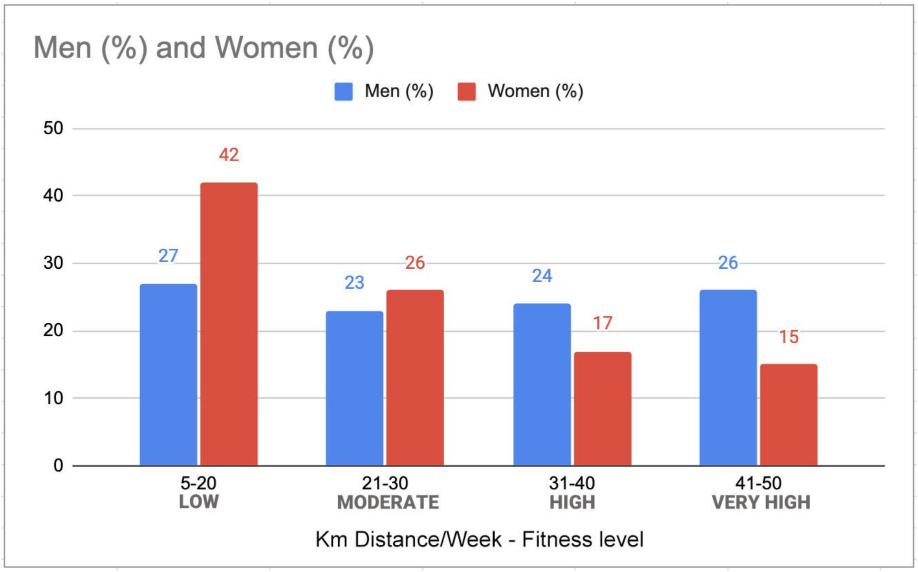


Fig. 6: Percentage of men and women with Low - Moderate - High - Very High fitness level.

3.3.5 Preferences

There are 4 preferences of behaviors that runners follow in this model:

Preference 1: Running as straight as possible and avoid repeating routes. When the runners come to the end of the road, they will choose randomly to either turn right or left to continue their runs.

Preference 2: Running on a straight road back and forth. However, to make this behavior more complex and unpredictable, there are possibility rates that if a random number falls into, runners could make a turn while running on this straight road. These possibility rates depend on the gender attribute of runners and what type of roads they first start.

Preference 3: Getting to the park as soon as the runners can by following the shortest route. There is also another possibility rate of running into another alternative road and continuing until getting to the park then running around it for a while.

Preference 4: Going home by following the shortest route. However, there is a possibility rate to run into another alternative road as well.

3.3.6 State machines

At first, all agents must be checked if they live within the range of 480m from the closest park and a random number falls within the rate that they want to go to the park. Preference 3 is assigned to those agents who want to run around the park when they start if satisfying those 2 conditions, otherwise, preference 1 or 2 is assigned to the agents to start running around and following their behaviors. After agents follow the behavior of preference 3 and run around the park for a while, their preference could be changed to 1 or 2 to continue running outside of the park area. Once the distance has gone and the estimated distance back home adds up to the distance goal, agents are then assigned with preference 4 to start going home until it gets to the initial position to stop completely and its state changed to 'rest' (Fig. 7).

Each state machine of each preference is simple in this model. There are 2 states in each preference including *"_continue_forward"* to keep running on the roads and *"_intersection"* when encountering an intersection or dead-end that need to decide what road to take in order to follow its behavior (Fig. 8). Depending on what preference that runners follow at that time, the state *"_intersection"* will execute the corresponding intersection function to decide the road to continue that follows its behavior.

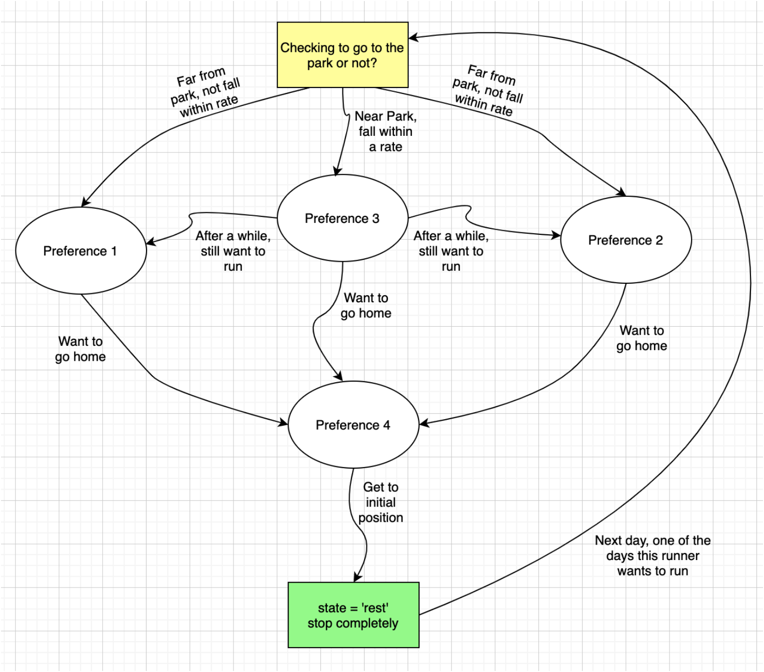


Fig. 7: Machine to help runners choose preferences.

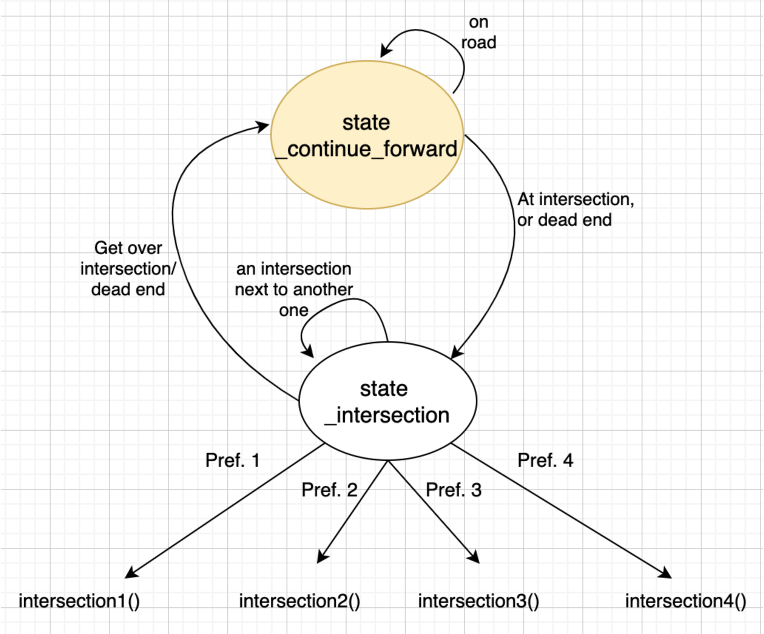
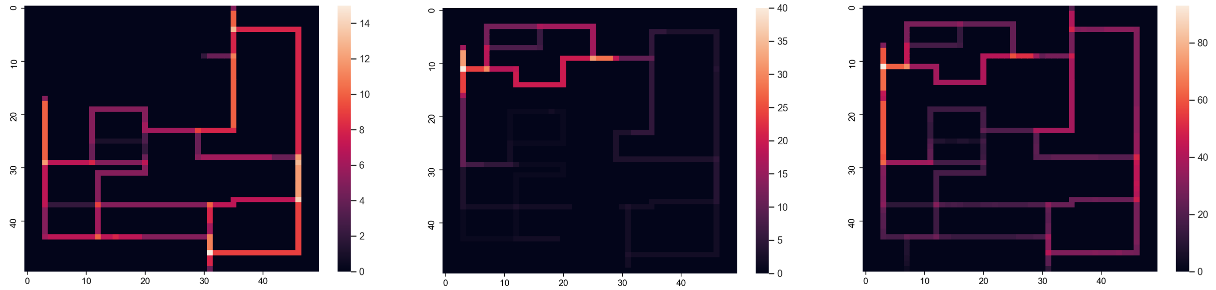


Fig. 8: State machine of each preference.

4. Results & Discussions

4.1 Simple Virtual World

The preferences of type 1 and type 2 of the simple virtual world are simulated separately and then together to see how runners follow their own behavior. The first type of behavior is tested with 10 agents and the distance goal of 160 grid cells. Similarly, the second type of behavior is also simulated with 10 agents and 160 grid cells for the distance goal as well. From the results (Fig. 9a and 9b), the heat maps show clearly all agents only follow their own preferences as all the normal roads are used in the first heat map. Whereas, the second heat map reveals the trail is used a lot by runners with type 2 along with a low level of use of the normal roads leading to the trail. When putting 2 types of behaviors together with 50 agents running over 160 grid cells, the heat map shows all of the roads are used with the level from low to high. Some roads are illustrated with a bright color as they are used to run a lot, especially, the normal roads that could be used to run by runners with type 1 and to access the trail by runners with type 2 (Fig. 9c).



(a) Runners with type 1. *(b) Runners with type 2.* *(c) Runners with type 1 and type 2.*

Fig. 9: Results from the simple virtual world.

4.2 Virtual City

The model is tested on the grid world of Brentwood - Darlington neighborhood in Portland. 500 agents were initialized and scattered randomly around the map to follow their own behaviors. After simulating this model with 7 days, and runners pick 4 days of a week to run, the heat map (Fig. 10) shows the roads around the park are used a lot by runners which are illustrated with the brightest color in this heat map. In addition, the main roads which are tertiary and secondary roads are used at a moderate to a high level since these roads are straight and have moderate traffic, so they are preferred to run a lot by runners. Most of the residential roads in the west-central area are used at a moderate to a high level as well due to the fact that they are long straight roads and connecting to the main routes. In contrast, most of the roads in the southeast area aren't used much since there are many dead ends and not many long straight roads that runners could stick around. The primary road on the right border is used with a low level as well because runners in this model don't like the heavy traffic on this type of road, and most of them prefer to turn or stay on tertiary or residential roads rather than stay on this road (values for variables are included in Appendix A).

In the second trial run of this model, a few variables were adjusted to aim for a better result and get closer to the Strava heat map. The possibility rate that male agents could be assigned to preference 1 when starting on the residential roads was decreased a little bit. An increase in the possibility rate for agents to turn to tertiary or secondary roads if they start on the primary road was also made. Another possibility rate that was increased in this trial run was the rate for agents to turn to secondary roads if starting on a residential road (values for variables are included in Appendix B). After running this trial, the heat map produced from this trial (Fig. 11) was getting more accurate than the first trial.

Although the heat maps of this model don't match completely with the Strava heatmap (Fig. 12), it captures those preferences of behaviors on what roads that runners use to run a lot in 7 days. However, there are a few limitations that could have affected this model including the lack of completed routes in a few corners and a few existed traffic signals that are missed in the data.

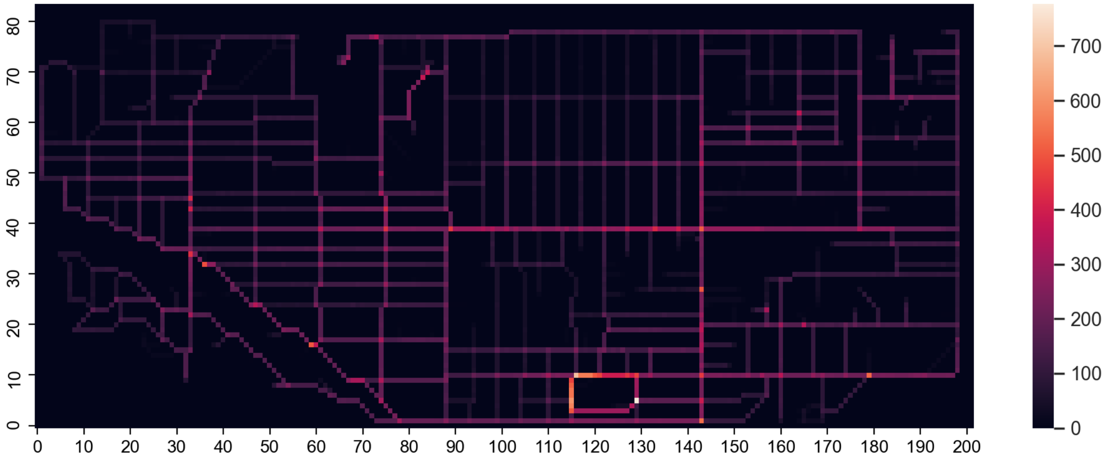


Fig. 10: Heat map from the model (1st trial).

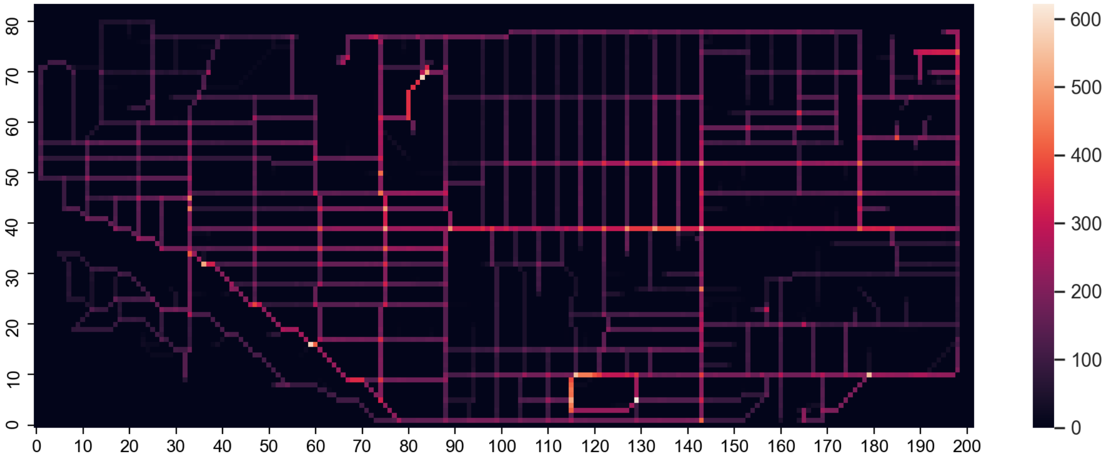


Fig. 11: Heat map from the model (2nd trial).

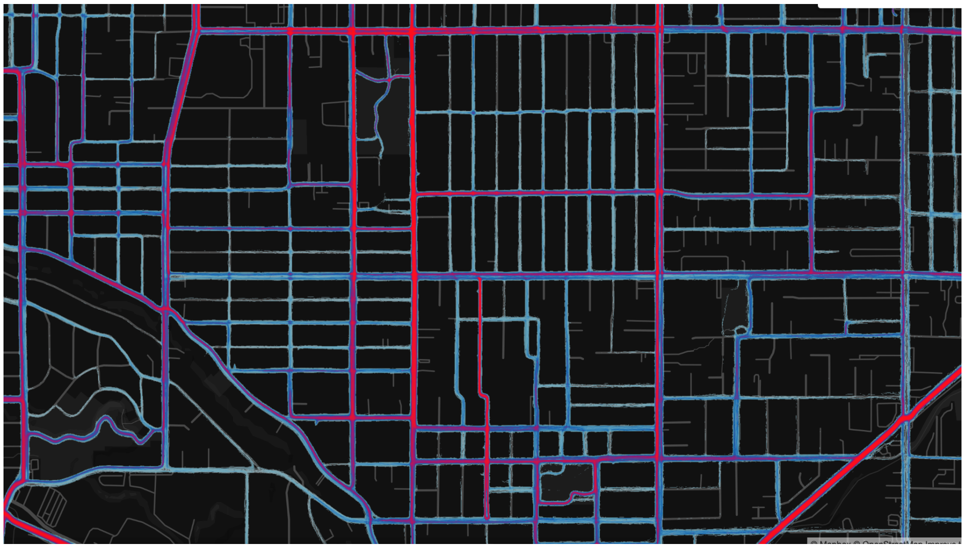


Fig. 12: Strava heat map.

5. Discussion/Conclusion

An agent-based model has been developed to simulate several behaviors of runners from the Brentwood - Darlington area in Portland, and a heat map is generated automatically at the end to show the roads and level of use by runners. The number of agents in this model could be increased to 1000 or 10,000 with some areas having a higher density than others to make the results more accurate. Not only simulated 7 days a week, but this model also gives us the capability to run the simulation with 1 month or 1 year if necessary. More preferences could be added to simulate more behaviors of runners in real life. Importantly, the interaction between runners could be another thing that future work might want to focus on.

Since this is the first-ever model about recreational runners, this is the first step to learn more about runners' behaviors and get to know the specific routes that they are going to take in their daily runs. This is contributing significantly to the city infrastructure planning to design paths or running zones appropriately to create more balance between runners and cyclists or runners and pedestrians in any community. Furthermore, this approach could be used to model other neighborhoods or cities as the street networks are available from OpenStreetMap data and could be downloaded using the OSMnx package in Python before transferring to the grid world and simulate runners similarly.

6. Acknowledgement

This REU Site is supported by the National Science Foundation under grant no 1758006. The author of this paper would also like to thank Dr. Christof Teuscher and Philippe Proctor for their valuable advice, guidance, and support throughout this research internship.

References

[1] S. Annear, “If you feel like there are more people out there running, you’re not alone,” *BostonGlobe.com*, Apr. 06, 2020. https://www.bostonglobe.com/2020/04/07/nation/if-it-feels-like-there-are-more-people-out-there-running-youre-not-alone/ (accessed Jul. 23, 2021).

[2] B. Stanic, “New Study Explores The World’s New-Found Love of Running,” *Lofficielarabia.com*, Jun. 10, 2020. https://www.lofficielarabia.com/fitness%20&%20health/new-study-explores-the-world-s-new-found-love-of-running (accessed Jul. 28, 2021).

[3] B. Galic, “126 Running Statistics You Should Know,” *LIVESTRONG.COM*, Feb. 18, 2021. https://www.livestrong.com/article/13730338-running-statistics/ (accessed Jul. 27, 2021).

[4] C. S. Priebe, M. Beauchamp, K. Wunderlich, and G. Faulkner, “‘I’m a runner not a smoker’: Changes in identity as predictors of smoking cessation and physical activity,” *Psychology of Sport and Exercise*, vol. 49, p. 101702, Jul. 2020, doi: 10.1016/j.psychsport.2020.101702.

[5] N. Schuurman, L. Rosenkrantz, and S. A. Lear, “Environmental Preferences and Concerns of Recreational Road Runners,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, Art. no. 12, Jan. 2021, doi: 10.3390/ijerph18126268.

[6] F. Fylan *et al.*, “Increasing conspicuity on night-time roads: Perspectives from cyclists and runners,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 68, pp. 161–170, Jan. 2020, doi: 10.1016/j.trf.2019.11.016.

[7] D. Ziemke, S. Metzler, and K. Nagel, “Modeling bicycle traffic in an agent-based transport simulation,” *Procedia Computer Science*, vol. 109, pp. 923–928, Jan. 2017, doi: 10.1016/j.procs.2017.05.424.

[8] I. N. Sener, N. Eluru, and C. R. Bhat, “An analysis of bicycle route choice preferences in Texas, US,” *Transportation*, vol. 36, no. 5, pp. 511–539, Sep. 2009, doi: 10.1007/s11116-009-9201-4.

[9] D. Masad and J. Kazil, “Mesa: An Agent-Based Modeling Framework,” Austin, Texas, 2015, pp. 51–58. doi: 10.25080/Majora-7b98e3ed-009.

[10] G. Boeing, “OSMNX: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks,” *Computers Environment and Urban Systems*, vol. 65, pp. 126–139, Jul. 2017, doi: 10.1016/j.compenvurbsys.2017.05.004.

Appendix A: Values of the parameters in the 1st trial.

Table A1: Possibility rate of going to the park and running around it when first start.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Male	0.4	0.35	0.2	0.15
Female	0.6	0.45	0.4	0.3

Table A2: Possibility rate going straight or turning at the traffic signals position.

Gender	Go straight/get over traffic signals	Make a turn
Either male or female	0.2	0.8

Table A3: Possibility rate assigning runners to follow preference 1 when starting on residential streets.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Male	0.3	0.4	0.6	0.7
Female	0.1	0.2	0.3	0.4

Table A4: Possibility rate assigning runners to follow preference 2 when starting on tertiary or secondary streets.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Either male or female	0.9	0.85	0.8	0.8

Table A5: Possibility rate making a turn while following Preference 2.

Type of road when start	Turn to residential roads	Turn to tertiary roads	Turn to secondary roads	Turn to primary roads
Residential	0.3	0.4	0.4	0.2
Tertiary or secondary	0.1	0.4	0.4	0.4
Primary	0.2	0.5	0.5	None

Appendix B: Values of the parameters in the 2nd trial.

Table B1: Possibility rate of going to the park and running around it when first start.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Male	0.4	0.35	0.2	0.15
Female	0.6	0.45	0.4	0.3

Table B2: Possibility rate going straight or turning at the traffic signals position.

Gender	Go straight/get over traffic signals	Make a turn
Either male or female	0.2	0.8

Table B3: Possibility rate assigning runners to follow preference 1 when starting on residential streets.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Male	0.3	0.35	0.4	0.5
Female	0.1	0.2	0.3	0.4

Table B4: Possibility rate assigning runners to follow preference 2 when starting on tertiary or secondary streets.

Gender	Low fitness level	Moderate fitness level	High fitness level	Very High fitness level
Either male or female	0.9	0.85	0.8	0.8

Table B5: Possibility rate making a turn while following Preference 2.

Type of road when start	Turn to residential roads	Turn to tertiary roads	Turn to secondary roads	Turn to primary roads
Residential	0.3	0.4	0.5	0.2
Tertiary or secondary	0.1	0.4	0.4	0.4
Primary	0.2	0.75	0.75	None