

הפקולטה להנדסה

המעבדה לעיבוד אותות

Multi-agent reinforcement learning of traffic routing in networks of autonomous vehicles

אלון רייכר

ליעד גם

פרויקט שנה ד' לקראת תואר ראשון בהנדסה

מנחה: ליאון רביב

מנחה אקדמי: פרופ' אמיר לשם

אוקטובר 2023

תוכן עניינים

[Abstract 1](#_Toc145444017)

[Thanks 3](#_Toc145444018)

[Presenting the Problem 5](#_Toc145444019)

[Theoretical Background 6](#_Toc145444020)

[Alternative Solutions 10](#_Toc145444021)

[Chosen Solution 11](#_Toc145444022)

[Results 14](#_Toc145444023)

[UML and chosen code 20](#_Toc145444024)

[Comparison between the different solutions 23](#_Toc145444025)

[Conclusions 25](#_Toc145444026)

[Ideas To Expand Upon 27](#_Toc145444027)

[Bibliography 29](#_Toc145444028)

[Appendix 31](#_Toc145444029)

[Appendix A 31](#_Toc145444030)

[Appendix B 31](#_Toc145444031)

# Abstract

In the modern era of the 21st century, with the always-increasing number of vehicles on the roads, getting from one point to another takes more time. That, in turn, leads to higher energy consumption, pollution traffic jams, etc [[1]](#_Bibliography). One possible solution to the problem is using AI-driven cars (or- agents) to navigate more efficiently.

With this project, we hope to help create, implement, test, analyze, and compare different algorithms to solve the problem. We also aim to provide an easy-to-use, realistic environment for testing different routing algorithms, extracting agents’ data, and being able to replay previous runs of the simulation.

The simulator we created is one part of a larger project, whose purpose is to help navigate autonomous cars in urban areas. This is the simulator, which the agents run on, and is used to collect data about times, routes, the learning process of the agents, and other statistics. Our runs and testing were mainly done using the map of Tel Aviv, provided by Open Street Map (OSM) with the osmnx python library.

The traffic is modeled by referring to the speed of each road, and not to individual cars in the road.

The project includes embedding different routing methods and collecting data about them, with a focus on using the shortest path routing algorithm and comparing it to a Q-learning method.

We conclude that using the shortest path routing algorithm, with the current implementation of the simulation and the q-learning algorithm, gives the best results.

# 

# Thanks

We would like to thank Li-on Raviv and Amir Leshem for guiding us throughout the year and helping us construct and organize the simulator, the agents’ learning process, and consulting us on the implementations of various ideas.

We would also like to thank The Engineering faculty in Bar Ilan for providing the relevant courses for the project.

# Presenting the Problem

In major urban areas, managing traffic efficiently is a major challenge. Our project introduces a car traffic simulator that explores three distinct routing strategies: random, shortest path, and Q-learning. This simulator serves as a tool to allow us to analyze and compare how these strategies influence traffic flow, congestion patterns, and responses to road obstacles in a controlled environment.

We've designed three different approaches for the cars to follow. First, there's the "random" approach, where cars make decisions without a specific plan, much like taking turns by chance. Then there's the "shortest path" approach, where cars always follow the shortest route, like using a GPS. Lastly, we have the "Q-learning" approach, where cars learn from past experiences to figure out the most efficient routes, even when facing unexpected situations.

However, it's not just about smooth driving. Our virtual city throws in challenges like traffic jams and road closures, just like what happens in real life. This is where our simulator becomes valuable. We can use it to see which approach handles these challenges the best.

Our role in this project is similar to that of researchers studying traffic behavior. By observing and comparing how each approach performs, we can determine which strategy is most effective under different circumstances.

# Theoretical Background

Traffic simulation plays a pivotal role in understanding and managing the complex dynamics of vehicular traffic within urban and transportation systems. By employing computational models, traffic simulation allows researchers, urban planners, and policymakers to analyze traffic patterns, assess the impact of infrastructure changes, and formulate effective traffic management strategies.

Dijkstra's Algorithm for Shortest Path:

Dijkstra's algorithm is a well-known method used to find the shortest path between two points in a graph. In the context of our car traffic simulator, Dijkstra's algorithm plays a significant role in the "shortest path" routing strategy. This strategy focuses on guiding vehicles along the most direct route from their starting point to their destination, aiming to minimize the distance traveled.

How Dijkstra's Algorithm Works:

**Initialization**: The algorithm starts by assigning a tentative distance value to each node in the graph, with the initial node's distance set to zero and all other nodes' distances set to infinity.

**Exploration**: The algorithm examines neighboring nodes, updating their tentative distances if a shorter path is discovered.

**Selection**: The node with the smallest tentative distance is chosen as the current node, and its neighbors are explored.

**Termination**: The algorithm selects nodes and explores their neighbors until the destination node is reached or all nodes have been visited.

**Backtracking**: Once the destination node is reached, the algorithm backtracks from the destination node to the starting node, tracing the shortest path.

In the context of our simulator, Dijkstra's algorithm provides a systematic and deterministic approach to route vehicles through the road network. This strategy ensures that vehicles follow the path that minimizes the distance traveled, which can be particularly beneficial for optimizing travel times and reducing congestion under normal traffic conditions.

Q-learning for Adaptive Routing:

Q-learning is a reinforcement learning technique used to make decisions in dynamic and uncertain environments. In the "Q-learning" routing strategy of our car traffic simulator, vehicles learn and adapt their routes based on past experiences and traffic conditions. This approach enables vehicles to make informed decisions, even in the face of changing circumstances and obstacles.

How Q-learning Work:

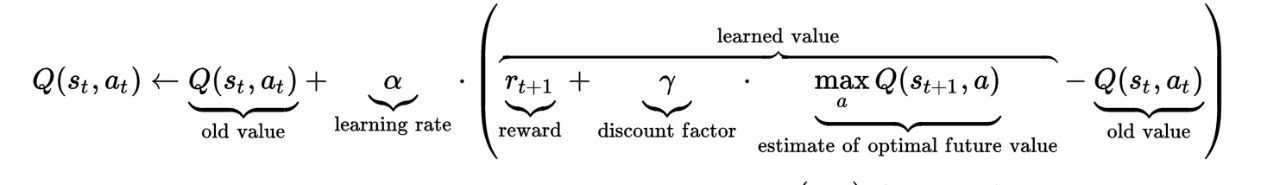
State and Action Space: Q-learning involves defining a state space (representing the environment) and an action space (potential decisions or moves).

Q-Values: Each state-action pair is associated with a Q-value, which represents the expected cumulative reward for taking that action from that state.

Exploration and Exploitation: Vehicles use a balance of exploration (trying new actions) and exploitation (choosing known high-reward actions) to learn optimal strategies.

Updating Q-values: Q-values are updated using a learning rate and discount factor, incorporating rewards obtained from actions and future Q-values.

This is the Q learning update rule:



Learning and Adaptation: Over time, vehicles learn which actions lead to higher rewards in different states, allowing them to adapt their decisions to maximize their cumulative rewards.

In our simulator, the Q-learning strategy empowers vehicles to learn from past experiences, such as traffic congestion and road closures. By adjusting their routes based on the learned Q-values, vehicles can dynamically navigate through the city, optimizing their paths to mitigate obstacles and minimize travel times. This adaptive approach to routing is particularly useful for addressing challenges like congested streets and unexpected road closures, contributing to a more efficient and responsive traffic management strategy.

# Alternative Solutions

For our navigation problem in a congested city, we had 2 alternative solutions:

1. Random navigation:

The first algorithm we implemented is a simple one, when a car needs to move to the next road the car simply chooses a random road and moves to the next road. This is a model that imitates a person who doesn’t know the city and just guesses where his destination is.

This is not a good algorithm because a lot of times the car doesn’t even make it to the destination.

1. Shortest path navigation:

The second approach we tried was shortest-path navigation. This approach uses the Dijkstra algorithm for shortest path finding in a weighted graph.

Our graph's weights are the road's distances from one another, and they can be only positive, Dijkstra algorithm works on a non-negative weighted graph in contrast to Belman Ford, but Dijkstra is faster, that’s why we chose Dijkstra's algorithm.

This approach is faster and more reliable than random routing. In normal conditions, it is guaranteed to find a path from the source to the destination if one exists, but in the other case, this approach does not consider the road's speeds and road blockages.

# Chosen Solution

Based on the weakness of the alternative solutions we chose the solution of Reinforcement Learning to be the navigation algorithm of the cars in the simulator.

Reinforcement learning is a machine learning paradigm that involves training agents to make sequential decisions in an environment to maximize a cumulative reward. It draws inspiration from behavioral psychology, where individuals learn through trial and error by interacting with their surroundings.

In reinforcement learning, an "agent" interacts with an "environment" to learn how to take actions that lead to the most favorable outcomes over time. The agent doesn't require explicit instructions; instead, it learns by receiving feedback from the environment in the form of rewards or penalties based on its actions. The goal of reinforcement learning is to discover the optimal strategy, known as a policy, that guides the agent's decisions to maximize its long-term rewards.

The Learning Process in Reinforcement Learning:

1. The agent starts with little to no knowledge of the environment and how its actions affect outcomes.
2. It interacts with the environment, taking actions and receiving rewards based on those actions.
3. The agent uses the received rewards to update its knowledge and refine its strategy (policy) to make better decisions in the future.
4. Over time, through repeated interactions and learning, the agent improves its policy to achieve higher cumulative rewards.

More specifically, we chose Q learning which is a specific reinforcement learning algorithm.

Q-learning can be effective for several reasons:

1. Adaptability to Dynamic Environments: Navigation often takes place in complex and dynamic environments, where conditions can change rapidly. Q-learning allows an agent to adapt its navigation strategy based on the feedback it receives from the environment. This adaptability is crucial for handling real-world scenarios such as traffic congestion, road closures, and unexpected obstacles.
2. Trial and Error Learning: Q-learning employs a trial-and-error approach, where the agent explores different actions and learns from the outcomes. This process allows the agent to discover which actions lead to better outcomes and adjust its behavior accordingly. In navigation, the agent can learn to avoid routes that consistently result in traffic jams or delays, leading to more efficient routing decisions over time.
3. Exploration and Exploitation: Q-learning balances the exploration of new paths and the exploitation of learned knowledge. This balance is vital for navigation, as it enables the agent to discover better routes while also leveraging its existing knowledge of effective paths. For example, the agent might explore new routes during periods of low traffic and exploit well-known routes during peak hours.
4. Long-Term Planning: Q-learning focuses on maximizing long-term cumulative rewards. In navigation, this means that the agent considers not only the immediate consequences of its actions but also the potential benefits or drawbacks of its choices over the entire route. This long-term perspective allows Q-learning to guide the agent towards routes that lead to the best overall outcomes.

# Results

To see how well the Q-learning routing algorithm works, we ran some tests in our simulations. First, we tried navigating from a random starting point to a destination. This helps us compare how the Q-learning method performs against just picking random paths.

For the Q Learning training, we set the following parameters:

Number of training episodes to 2000, the maximum number of steps in each episode to 100, learning rate to 0.05, discount factor to 0.9, and epsilon to 0.2. The parameter's values were decided by trial and error, we tested and adjusted them until got them to a sweet spot where the agent's training wouldn't take too long but the agent will learn the optimal path.

For the simulation we use the map of Tel Aviv, the starting point is from "Rokach Boulevard" to "Dizengof Center".

After the agent's training finished, we got two diagrams that describe the course of the training:

1. תמונה שמכילה צילום מסך, עלילה, טקסט, קו

   התיאור נוצר באופן אוטומטיBar chart of travel time in each episode and an indicator if the agent made it to the destination:

We can see that at first, the agent mostly hadn’t reached the destination but as training continued the agent got to the destination and on a more optimized path. We can see that by the number of seconds being lower as the episode numbers grow.

1. תמונה שמכילה טקסט, קו, עלילה, צילום מסך

   התיאור נוצר באופן אוטומטיMean rewards over several episodes:

the x-axis is the mean rewards over the last 50 episodes, and we can see that the reward of the agent is growing as we train it more and more.

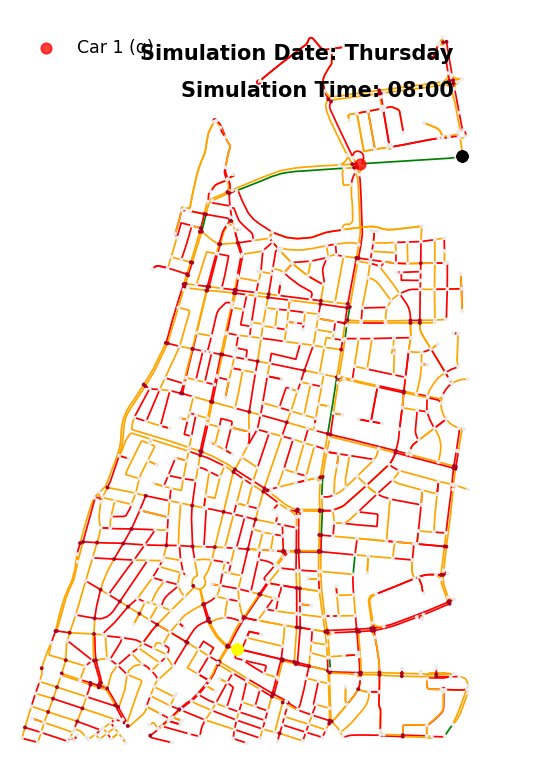
Now let’s look at the result, the agent's route:

תמונה שמכילה תרשים, תוכנית, שרטוט, מפה

התיאור נוצר באופן אוטומטי

The black mark indicates the starting point of the agent, and the yellow mark indicates the ending point.

Additionally, we have an animation of the agent's drive in real-time with the current time road's speeds whereas: red – heavy traffic, yellow - medium traffic, and green – low traffic:

תמונה שמכילה טקסט, תרשים, מפה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, תרשים, מפה

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, מפה

התיאור נוצר באופן אוטומטי

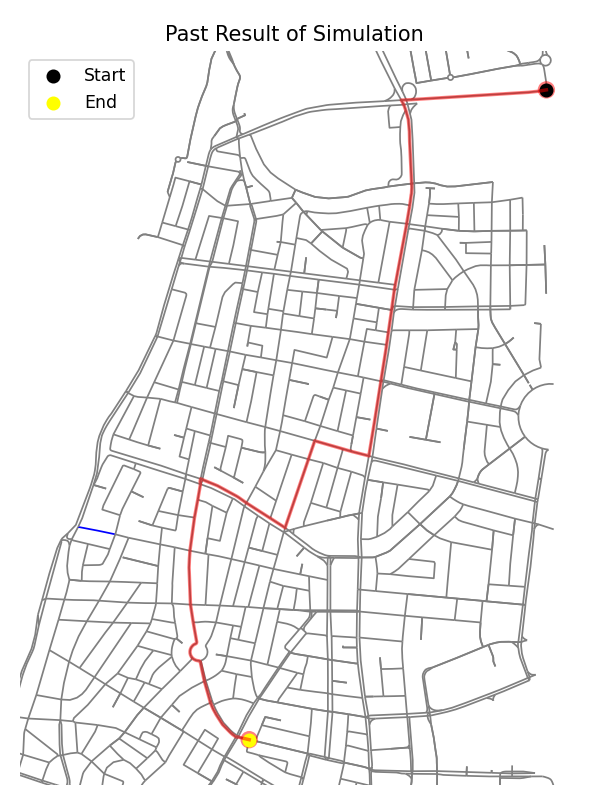
We can see the red dot, representing the q-learning agent, navigates from the starting point to the destination. Additionally, we have the day and the time of the simulation. The agent completed this drive in 19 minutes.

תמונה שמכילה צילום מסך, עלילה, צבעוני, קו

התיאור נוצר באופן אוטומטיNow, to test the q-learning algorithm let’s make a scenario, let’s say that on the following Sunday morning, the agent wanted to make the same route but this time a road in the original route that the agent took was blocked this morning, let’s see how well the agent adapt:

תמונה שמכילה קו, עלילה, צילום מסך, תרשים

התיאור נוצר באופן אוטומטי



We can see that there is a road colored blue in the map, this indicates that the road was blocked. Now we can see that the agent knew the road was blocked and chose a different route this time to avoid the closed road. This route took the agent 22 minutes to complete and that’s why this route wasn’t chosen in the first simulation.

# UML and chosen code

The project was created in the MVC architecture (Model, View, Controller). It is also made of a few thousand code lines, thus we will present only the UML of the model (where the main logic happens) and chosen codes from it. Documentation is available in the Simulator HLD appendix.

A diagram of a computer

Description automatically generated**The UML (also in the HLD appendix)**

Shortest path algorithm usage:

Outer layer in Route.py

def decide\_first\_road(self):  
 if self.src\_node == self.dst\_node:  
 return None  
 first\_road = self.road\_network.get\_next\_road\_shortest\_path(self.src\_node, self.dst\_node)  
 self.current\_node = first\_road.destination\_node.id  
 self.path.append(self.current\_node)  
 return first\_road  
  
  
def get\_next\_road(self):  
 if self.current\_node == self.dst\_node:  
 return None  
 next\_road = self.road\_network.get\_next\_road\_shortest\_path(self.current\_node, self.dst\_node)  
 self.current\_node = next\_road.destination\_node.id  
 self.path.append(self.current\_node)  
 return next\_road  
  
def get\_alt\_road(self):  
 self.path.pop()  
 self.current\_node = self.path[-1]  
 print("get\_alt\_road")  
 adjacency\_list = self.road\_network.node\_connectivity\_dict[self.current\_node] # list of all the adjacent nodes ids  
 for next\_node in adjacency\_list:  
 road = self.road\_network.get\_road\_from\_src\_dst(self.current\_node, next\_node) # this is "road"  
 if not road.is\_blocked:  
 # if the road is not blocked, we check if the any road on the shortest path is blocked  
 # if not, we return the road  
 if self.road\_network.get\_shortest\_path(next\_node, self.dst\_node) is not None:  
 self.current\_node = next\_node  
 return road  
  
 return None

The construction of the distance matrix in the Road\_Network.py class:

def add\_shortest\_path\_to\_matrix(self, src\_id: int, dst\_id: int):  
 *"""  
 Add the shortest path between two nodes to the distances matrix.  
  
 Args:  
 src (int): Source node ID.  
 dest (int): Destination node ID.  
  
 Returns:  
 None  
 """* path = nx.shortest\_path(self.nx\_graph, src\_id, dst\_id, weight='length')  
 path\_length = nx.shortest\_path\_length(self.nx\_graph, src\_id, dst\_id, weight='length')  
  
 #updating the distances matrix  
 previous\_edge\_length = 0  
 for i,node in enumerate(path[:-1]):  
 self.next\_node\_matrix[node][path[-1]] = path[i + 1]#adds the relevant next node to the distances matrix  
 self.distances\_matrix[node][path[-1]] = path\_length - previous\_edge\_length  
 previous\_edge\_length += self.get\_road\_from\_src\_dst(node,path[i + 1]).length  
 return

Explanation:  
The matrix is constructed using the shortest path algorithm of the network library (which by default uses the Dijkstra algorithm). The matrix’s line index is the current node ID, the column index is the destination node ID, and the value in each entry is the node ID of the next node in the shortest path. The agent gets its next node from the matrix, and if the next road is blocked, the agent tries to go through the next best option from its current node.

# Comparison between the different solutions

(insert 2-3 runs from 2-3 cities, long distances and same starting time)

# Conclusions

We observed that with the current parameters of the Q learning algorithm, the best result we get converges to the shortest path algorithm, which is a bit unexpected.

The conclusion we got is that- the function we used for Q learning might not be the optimal one for getting the fastest route, under the current implementation of the simulation.

# Ideas To Expand Upon

We suggest the following ideas for possible changes or to expand upon in the future:

* Adding other algorithms and methods to navigate, such as policy-based learning.
* Tweaking the settings for the Q learning may yield better results- changing the reward function, number of episodes, steps per episode, etc.
* Changing the traffic model (to a microscopic or a mesoscopic model) may yield better results, with the cost of performance.
* Changing the way that synthetic data is made to represent better the real world- for example, considering that cars that drive through a road will get to the adjacent road and affect its speed.
* Adding an algorithm for finding the fastest route for each trip and comparing it to the routes found by other algorithms could help with analyzing the results and give a better way to estimate other solutions.

# Bibliography

[1] Mir Shabbar Ali, M. A. (2014, September 6). *Estimation of Traffic Congestion Cost-A Case Study of a Major Arterial in Karachi.* Retrieved from ScienceDirect: https://www.sciencedirect.com/science/article/pii/S1877705814010078

# Appendix

## Appendix A

"Software Requirements Specification" – more specific software descriptions of the program.

## Appendix B

"Simulator HLD" – provides information about configurable parameters, used data, and user documentation of all the classes in the project.