DEPARTMENT OF COMPUTER SCIENCE

THE UNIVERSITY OF HONG KONG

Robot Path Planning in Wireless Communication

Using Reinforcement Learning

FYP Interim Report

Author: Anushka Vashishtha (3035299404)
Supervisor: Dr. C. Wu
Advisor: Dr. Y. C. Wu
Submission Date: January 20, 2018

I confirm that this fyp interim report is my own work and I have documented all sources and material used.

, January 20, 2018 Anushka Vashishtha (3035299404)

Acknowledgments

I would like to thank my final year project's supervisor Dr. C. Wu for her advice. Next, I want to thank Dr. Y. C. Wu my advisor who always steered me in the right direction whenever I needed it. I would also like to thank Dr. Shuai Wong, a post-doctoral candidate at the Department of Electrical and Electronic Engineering. The door to his office was always open whenever I ran into trouble or had questions. He specially provided me assistance related to parts where knowledge in electrical engineering was crucial.

Abstract

In the past few decades, there has been an influx in the number of internet of thing devices being used worldwide, and the amount of data which they are producing is estimated to be 100s of trillion gigabytes per year [1]. This tremendous reliance on IoT devices, generates a situation where we have to find efficient ways to collect data from them as well as charge them, specifically in the case of tiny IoT devices like an RFID or Bluetooth. Using a traditional method like battery is not a viable option for miniscule size IoT devices. On the other hand, charging cables are not suitable, as it is not only expensive to purchase them in abundance considering each device, but also not practical in inaccessible areas. Henceforth, this project proposes the deployment of an unmanned ground vehicle in designated areas to wirelessly charge [2] and collect data from clusters of tiny IoT devices.

The objective of this report is to explore different methods like MINLP, Reinforcement Learning, Deep Reinforcement Learning and if time permits, Multi-armed Bandit Optimisation, in order to plan the path of the UGV so that it can charge the devices, meanwhile optimising the energy consumed by it and the total path taken. Results from MINLP and Reinforcement Learning have been included and compared. As a next major step, findings from methods like Deep Reinforcement Learning and Multi-armed Bandit Optimisation will be added in this report and then they will be subsequently compared. All of the above methods are compared extensively on the basis of their efficiency and speed, and ultimately the one which gives the best result in a real world environment is chosen.

This report demonstrates that if the performance of the chosen method is promising, then such a vehicle can actually be deployed and can help in charging and gathering data in real life, for example from packages kept in a warehouse and marked by an RFID. Moreover, it leads to reduction in charging cable usage which can help the environment.

Abbreviations

Here are some abbreviations used in this report:

Table 1.

IoT	Internet of Things
UGV	Unmanned ground vehicle or robot
MINLP	Mixed Integer Non-Linear Programming
RFID	Radio Frequency Identification
MTC	Machine-Type Communications
UWB	Ultra-Wide Band

Contents

List of Figures

List of Tables

1. Introduction

This section embarks with a detailed description on the general project background, followed by a problem statement and objective. In addition, this section also provides with an outline for the remaining sections of the report.

1.1. Project Background

In the last years, IoT has taken the centre stage in the technology world by creating one of the fastest growing markets and it has been predicted by Forbes that more than 30% of the companies in manufacture, Internet of Vehicle as well as retail, have already adopted IoT devices in 2015. Notably, IoT devices will soon be worth 1700 billion U.S. dollars, as they are expected to outnumber 20 billions in 2020 [3]. In addition to that, IoT is already estimated to be generating 100s of trillion gigabytes of data per year and this figure is only increasing [1]. In the next decade, almost every device will be connected to the internet, ranging from sensors, vehicles, wearable electronics to other embedded systems like refrigerators [1].

To prepare for the future, design engineers are working on finding efficient solutions, specifically in order to power as well as to communicate with billions of tiny IoT devices, since providing sufficient energy to them particularly, is quite a difficult task. Relying on traditional resources like batteries cannot meet the requirement due to the miniscule size of such a device. Additionally, using charging cables can suffice the requirements but the downfall here is that they must be bought in abundance and then, repaired and disposed sustainably. Moreover, maintenance of such a cable becomes a challenge, when IoT devices work in inaccessible areas.

To solve this problem, a UGV is used to wirelessly charge [4] (and communicate with) a cluster of small scaled IoT devices like RFIDs, Bluetooth and UWB as they not only require short distance MTC (distance $\leq$ 10m) but also do not need enough power for charging (typically 1 µW ~1 mW)[3]. As a result, cables will become an obsolete solution for powering the IoT devices and this techniques will also resolve the difficulties involved in purchasing and maintenance. Therefore, rather than buying say 100 different cables for 100 different IoT devices, there will now be just one UGV to power all of the devices and also collect data from them, if required.

1.2. Objective

From a Computer Science prospective, this problem can be abstracted to be a path planning problem in a graph. The UGV will be capable of interacting with the devices and charging the devices present at different locations in a graph, one by one. To achieve this, the given report will present different approaches such as MINLP, Reinforcement Learning, Deep Reinforcement Learning and if time permits, Multi-armed Bandit Optimisation. After individually receiving the results for the optimum paths from the various approaches mentioned above, the report will depict a systematic comparison of their performances to conclude which has a promising solution theoretically and if it is practically viable to apply any of the approaches in a real world environment.

1.3. Outline

The remainder of this report will proceed as follows. First, there is a detailed description about the prior work which is done in the field of robot wireless charging. Next, the report explores different methodologies in detail on how to plan the path of the UGV. Here, the relevance of each method, technical side and step by step procedure involved in execution of each algorithm is explained in depth. Next, the report discusses the result and covers the main difficulties encountered till now. Adding on to this, conclusion as well as the future planning of the project is discussed wherein the project schedule is presented.

2. Previous Work

Through this project, Reinforcement Learning is being implemented for the first time in order to plan the path of a UGV (also referred to as robot) so that it can charge IoT devices. Nevertheless, there has been extensive research done on various robot wireless charging approaches which are explained below.

A traditional scheme for the robot wireless charging is to deploy multiple static transmitters. However, this is cost expensive and not adaptive to network changes. Additionally, there are two other references which can help in robot wireless charging.

1. Wirelessly powered two way communication with non-linear energy harvesting model: Rate regions under fixed and mobile relay [5].

2. Near-Optimal Velocity Control for Mobile Charging in Wireless Rechargeable Sensor Networks [6].

Recently, the above two references use mobile robot for charging, but they adopted fixed paths and assumed complete knowledge of user channels [5][6]. This can lead to excessive energy usage. To this end, path planning with or without channel knowledge is needed, as in real life situations it's not practical to know the complete picture of where each IoT device is located.

Therefore, this project advocates the use of Reinforcement Learning so that the robot can learn from the environment and can explore where each IoT device is located. At the same time the robot is also charging the devices and aiming to minimise energy consumption. This is quite cost effective as only the robot is required for learning and charging without any additional hardware. Moreover, while planning the path, the robot can also have the ability to communicate with IoT devices.

3. Methodology

This section includes description of model of the environment as well as the different phases involved in the development of the project.

Charging Region Model

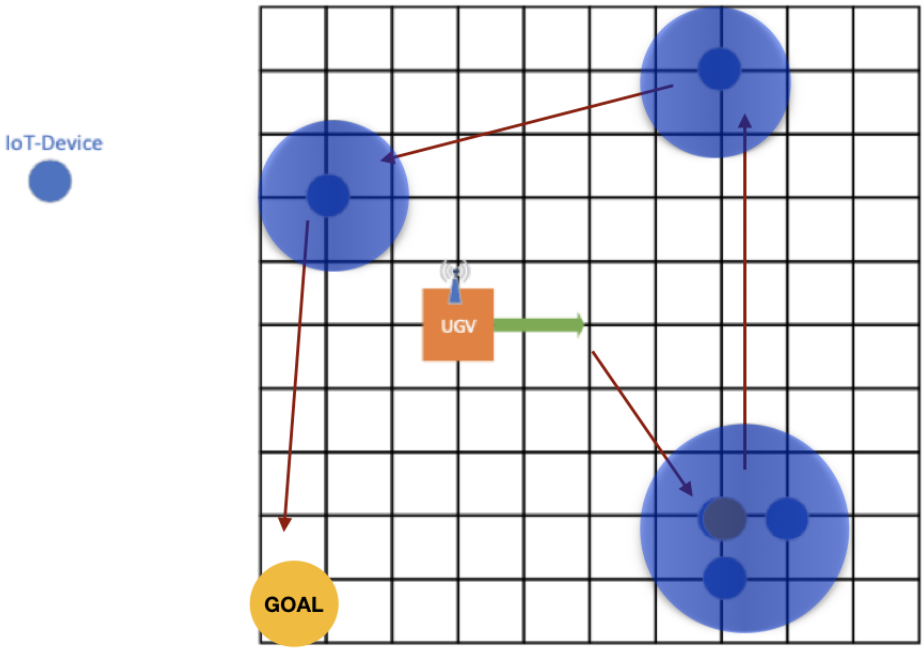

Figure 3.1.: Charging Region Model

The abstraction of the robot path planning problem gives the following model:

There is a set of IoT devices $D = \{d_1, \ldots, d_k\}$ which are positioned on the vertices $V = \{v_1, \ldots, v_M\}$ of an equal-distance-grid-shaped-graph representing the operation area. The UGV is located on one of the vertex v_U and the goal is present at another

vertex v_G. In each time-step, the UGV can move one step along the grid. The UGV has to interact (i.e. provide power or communicate) with the devices by entering the charging region [appendix A.1] of each device one by one and ultimately reach the goal. A certain amount of energy is required by UGV in moving and that is the energy which is reduced through path planning.

3.1. Phase I

The first phase to obtain the solution for the path planning problem, involved the use of MINLP which uses branch and bound approach for solving problems.[7] It is used to solve convex or non-convex optimisation problems with discrete variables and non-linear functions which can be placed as either an objective function or as a constraint.[7]

These properties of MINLP made it particularly useful in deciphering the robot path planning problem because of two reasons. First, the constraint on path planning, which is sub-tour elimination, is of non-linear nature and second, the location of each IoT device on the grid satisfies the requirement for the presence of a discrete variable.[7]

Subsequently, a CVX optimisation solver Mosek was used to handle combinatorial difficulty of optimising over discrete variable sets together with the issue of handling a non-linear function in order to solved MINLP.

This table presents the parameters used in the MINLP problem.

Table 3.1.

$\mathbf{v}$	Visit a point in the grid or not (Boolean)
$\mathbf{X}$	Link between two points in the grid or not (Boolean)
α_1, α_2	Toning parameter: pioneer's 3DX robot experiment result at MIT(constant)
a	Velocity of UGV (constant)
$\mathbf{D}$	Distance between two point in the grid or not (Boolean)
$\mathbf{W}$	Summation of X values
M	length & width of the grid
K	Total number of IoT devices

The mathematical description of objective function (represents the total moving energy of UGV) and constraints involved in MINLP is as follows:

$$\min_{\mathbf{v},\mathbf{X},\{\lambda_m\}} \quad \left(\frac{\alpha_1}{a} + \alpha_2\right) \mathrm{Tr}(\mathbf{D}^T\mathbf{W})$$

$$\text{s.t.} \quad v_1 = v_M = 1, \ (\textit{select starting and end points}) \tag{3.1}$$

$$v_m \in \{0,1\}, \ \ \forall 2 \leq m \leq M-1, \ (\textit{selection is binary}) \tag{3.2}$$

$$\sum_{m \in \mathcal{C}_i} v_m \geq 1, \ \ \forall i = 1, \cdots, K, \ (\textit{charge all IoT users}) \tag{3.3}$$

$$W_{m,j} \in \{0,1\}, \ \ \forall m, j, \ \ W_{m,m} = 0, \ \ \forall m, \ (\textit{flow selection is binary}) \tag{3.4}$$

$$\sum_{j=1}^{M} W_{1,j} = 1, \ \sum_{j=1}^{M} W_{j,1} = 0, \ (\textit{flow from starting point}) \tag{3.5}$$

$$\sum_{j=1}^{M} W_{M,j} = 0, \ \sum_{j=1}^{M} W_{j,M} = 1, \ (\textit{flow to end point}) \tag{3.6}$$

$$\sum_{j=1}^{M} W_{m,j} = v_m, \ \sum_{j=1}^{M} W_{j,m} = v_m, \ \ \forall m = 2, \cdots, M, \tag{3.7}$$

$$(\textit{flow passing selected points; no flow passing abandoned points})$$

$$\lambda_m - \lambda_j + \left(\sum_{l=1}^{M-1} v_l - 1 \right) W_{m,j} + \left(\sum_{l=1}^{M-1} v_l - 3 \right) W_{j,m}$$

$$\leq \sum_{l=1}^{M-1} v_l - 2 + J\left(2 - v_m - v_j\right), \ \ \forall 2 \leq m, j \leq M-1, \ m \neq j, \tag{3.8}$$

$$v_m \leq \lambda_m \leq \left(\sum_{l=1}^{M-1} v_l - 1 \right) v_m, \ \ \forall m \geq 2. \tag{3.9}$$

$$(\textit{guarantee flow connected})$$

In Phase I when there is complete knowledge about the environment, then MINLP gives the most optimal solution and when there is incomplete knowledge then it helps by giving the lower bound.

3.2. Phase II

Thereafter, the model is put in a Reinforcement Learning setting and Q learning is applied on it. The goal of Q-learning is to learn from the environment, wand tell an agent what action to take under what state. It does not require a model of the environment and can handle problems with stochastic transitions and rewards.

Here the state space, the actions and the rewards are proposed to be defined in the following way:

State $S = \{(x,y)|x,y \in [M]\}$ $(\textit{location of each point on the grid})$

Action $A = \{"move\ left\ UGV", "move\ up\ UGV", "move\ down\ UGV", "move\ right\ UGV"\}$

$$f(x_i, y_i) = \begin{cases} 10 + (5 * x) & \text{if } (x_i, y_i) = v_G,\ x = \text{ no. of IoT devices charged} \\ 10 & \text{if } (x_i, y_i) \text{ in charging region of a particular IoT for the first time} \\ -1 & \text{otherwise.} \end{cases}$$

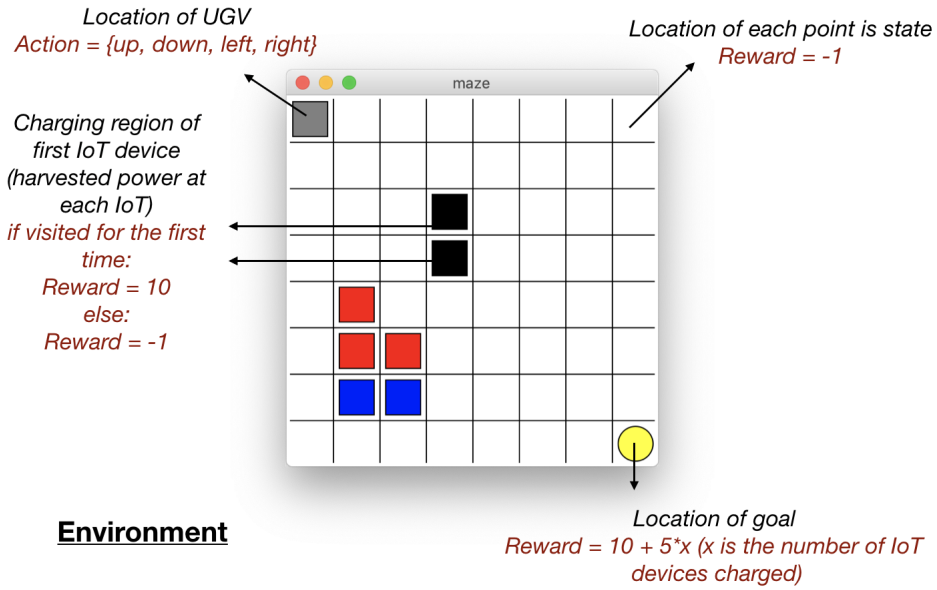

Figure 3.2.: Q Learning Environment

3.3. Phase III

Next, in this project Deep Reinforcement Learning will be used along with policy gradient method to approach the problem. Deep Reinforcement learning involves applying the standard Reinforcement Learning state-action model as describe in Phase II, along with neural network. Neural networks are a computing system and they consist of a collection of connected nodes. They are used to work together and process complex data inputs and therefore will be able to take in a model with a very large state space.

4. Results

This section presents the results from different methods, and comparisons between the methods.

4.1. Phase I

In this section, all the results and inferences obtained after the application of MINLP are discussed. Implementation is carried out in MATLAB using CVX solver Mosek.

In fig. 4.1 we do not consider the charging case (section 3.1) which means that UGV will not consider charging IoT devices as a constraint while planning its path.

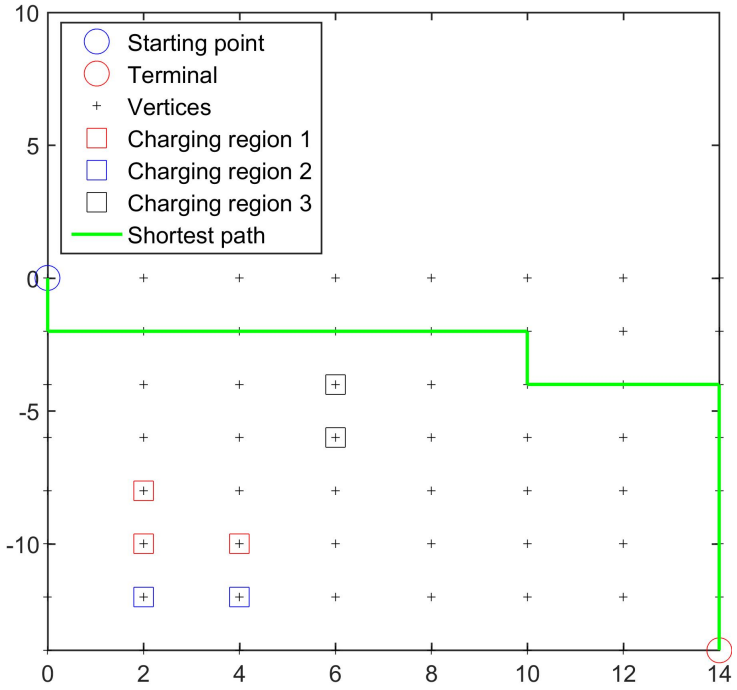

Figure 4.1.: Shortest Path by MINLP

Next fig. 4.2 shows the path taken by the UGV where sub-tour eliminations (sec-

tion 3.1) is not involved as one of the constraints[8]. Therefore, there will be various small tours in the grid apart from the path between UGV's initial position and goal(terminal), which are highly undesirable.

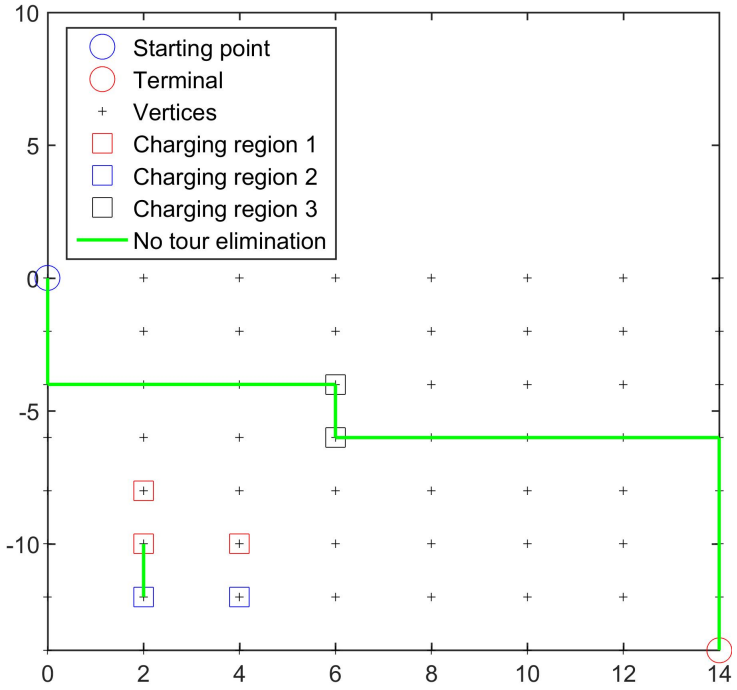

Figure 4.2.: No Sub-tour Elimination for MINLP

Finally, we have the optimal solution after considering all the constraints from section 3.1. Here, the energy value obtained is 241 Joule by taking the path shown in fig. 4.3.

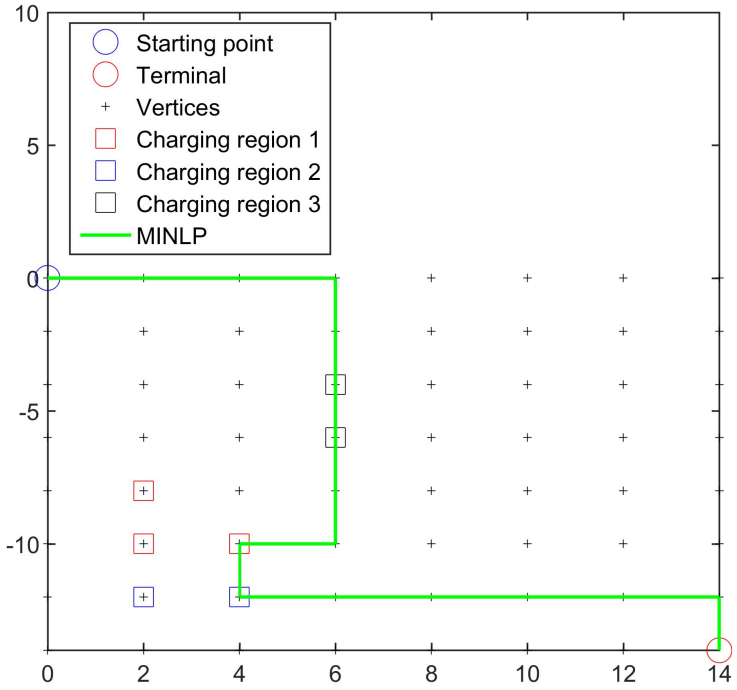

Figure 4.3.: Optimal Solution by MINLP

4.2. Phase II

In this section, all the results and inferences obtained after the application of Q-Learning are discussed. Implementation is carried out in python.

The fig. 4.4 shows that when a high reward is given to the UGV (say 100 + 5*each device charged) when it reaches the goal, then the UGV has a very low tendency to charge all IoT devices and therefore we will not get the desired path.

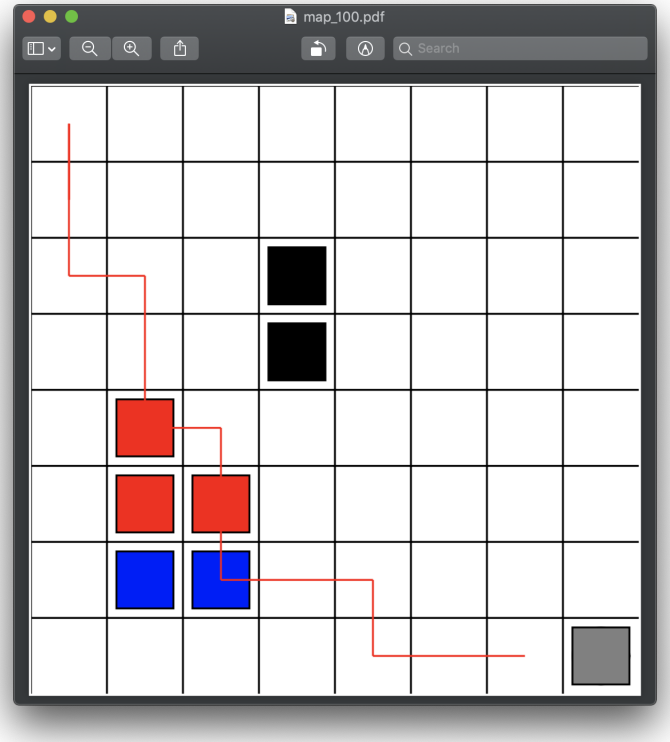

Figure 4.4.: High Reward for Goal

Finally after implementing all the reward values from section 3.2, we get the result by Q Learning in fig. 4.5. The lowest energy value consumed is 252 Joule after training for 100 epochs.

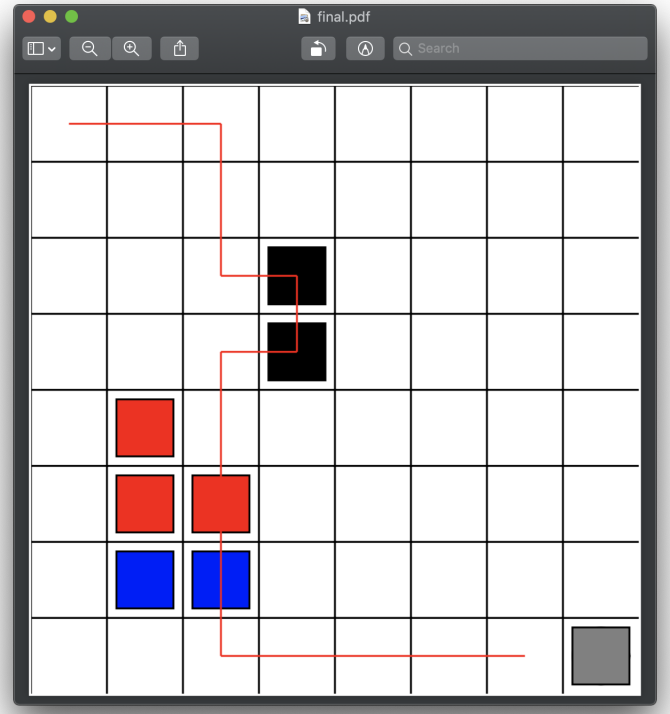

Figure 4.5.: Solution by Q Learning

4.3. Comparison of Results

This section compares results obtained from MINLP and Q-Learning.

In fig. 4.6 a graph is plotted between energy consumption by UGV in moving and number of epochs. Here as the UGV gets trained for more epochs, the energy consumed by it in moving decreases drastically. The Blue line shows the energy consumption for Q-Learning and the red line serves as a lower bound and shows the energy consumption for MINLP(energy is calculated in section 3.1). This comparison is not fair as it is not certain that every path obtained by Q-Learning after a certain number of epochs, involves all the IoT devices to be charged.

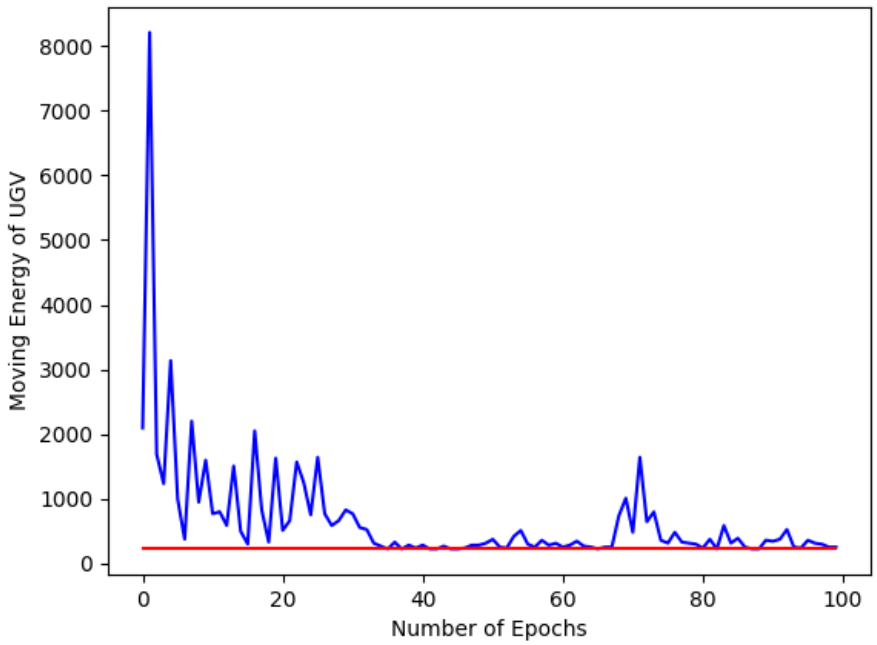

Figure 4.6.: Moving Energy Comparison: MINLP and Q-Learning

Therefore, we only take those energy values from Q-Learning where path involves charging all the IoT devices. The final result obtained is shown in fig. 4.7

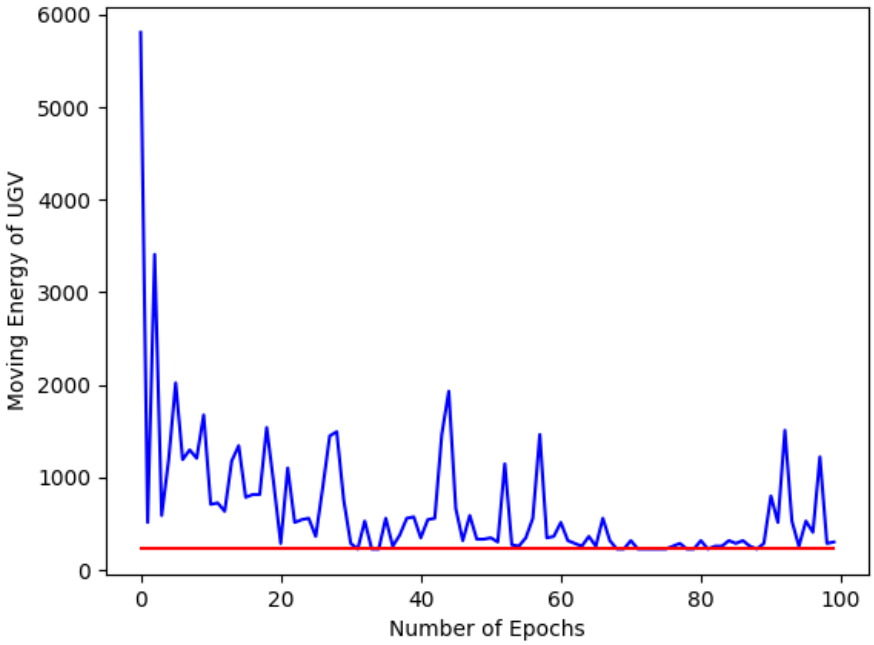

Figure 4.7.: Moving Energy Comparison: MINLP and Q-Learning(all devices charged)

5. Difficulties Encountered

The main difficulties encountered in the development of this project are explained as follows:

1. Formulating the model for the project initially appeared as a difficult task, since there was no prior work available in this area. Therefore, different action sets and reward functions were considered before choosing the most suitable one for Phase I modelling.

2. Understanding the electrical engineering concepts involved in robot wireless charging was a challenging work as it required quite high level knowledge in that field.

3. Limited experience in TensorFlow and MATLAB was one of the biggest hurdle in project development as all the implementation is carried out with these software.

4. Most importantly all the main methods like Reinforcement Learning, Deep Reinforcement Learning and Multi-armed Bandit Optimisation are quite new for me, therefore considerable amount of time was spent to get a basic foundation in order to understand the mechanism of each of them.

5. Problems related to the installation of CVX software due to licensing issues, delayed the progress of the project significantly

6. Conclusion

The main aim of this project is comparison of different approaches so as to choose the most efficient path planning method for the deployment of a UGV. In order to achieve this, this report presented results obtained from the first and second methods which are MINLP and Q-Learning respectively. Subsequent results from methods ranging from Deep Reinforcement Learning and Multi-armed Bandit Optimisation will be added in the future. Comparison of outcome obtained from reinforcement learning completed the interim stage of the project and showed that after training the UGV on Q-Learning algorithm for sufficient number of epochs, the UGV was able to take similar path as that of MINLP method and the energy consumption reduced significantly and became close to that of MINLP method which serves as a lower bound in this report.

Finally, after the comparison of different methods, if the results are optimum for real world application, then the deployment of such a UGV for wirelessly charging (and communication with) IoT devices will be feasible. This will not only make cables obsolete but also will play a big role in data collection and charging in various sectors ranging from manufacturing to retail.

7. Future Planning

This section describes the remaining milestones as well as some of the future research work which can be done in this project if time permits.

1. Tackling the situation when variable power is given by UGV to an IoT device according to how far away it is from an IoT device
2. Limited energy present in UGV
3. MINLP application in Continuous charging model (fig. 7.1)

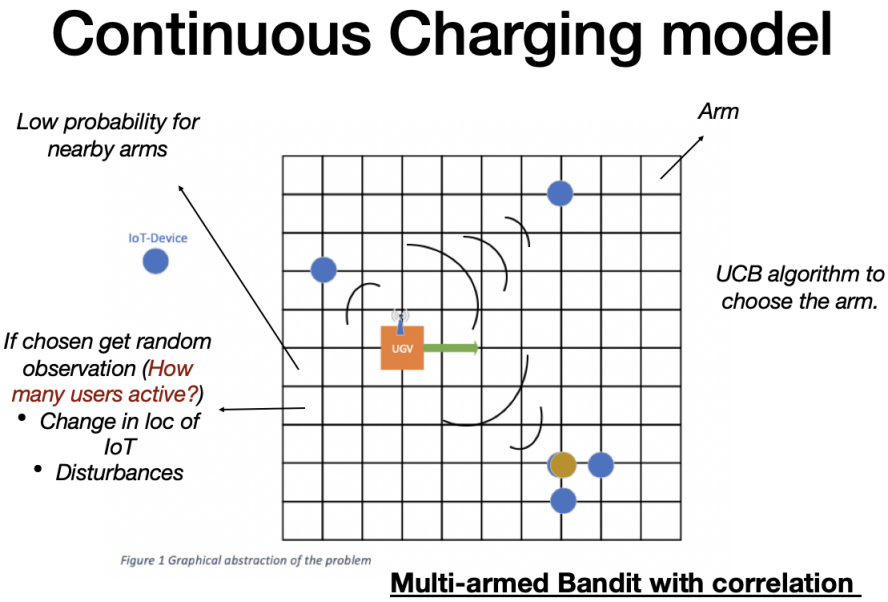

Figure 7.1.: Continuous Charging Model

The table 7.1 illustrates the already achieved milestones and the tentative plan on project development in the upcoming months. Until now deliverables 1 and 2 have

been finished, thorough background research has been done, creation of simulated environment has been achieved and result from MINLP and Q-learning have been obtained. In the future, the main focus is on acquiring results from Deep Reinforcement Learning method and Multi-armed Bandit Optimisation.

Table 7.1.

Dates	Milestones	Status
September 30	Deliverable 1 1. Project Plan 2. Project Website	Completed
October	Working into MATLAB, Python and TensorFlow. Reading up on MINLP, Reinforcement Learning and Deep Reinforcement Learning.	Completed
Nov - Dec	Development of demo application. Creating simulated environment for MINLP application and applying it.Applying Reinforcement Learning and comparing them	Completed
Jan	Deliverable 2 1. Demo Application 2. Interim Report	Completed
Dec - Feb	Applying Deep Reinforcement Learning with Policy gradient method.	In progress
Mar - Apr	Considering variable power given by UGV, limited energy present in UGV as further extensions to the problem. Comparing the results obtained from the different approaches. If time permits then working into Continuous charging model (Multi armed Bandit Optimization)	Scheduled
April	Deliverable 3 1. Finalized Implementation 2. Finalized Report	Scheduled

Bibliography

[1] A. Somov and R. Giaffreda, "Powering iot devices: Technologies and opportunities," *IEEE Internet of Things Newsletter*, Nov. 2015.

[2] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," *CoRR*, vol. abs/1803.07123, 2018. arXiv: 1803.07123. [Online]. Available: http://arxiv.org/abs/1803.07123.

[3] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, and S. He, "Narrowband internet of things: Implementations and applications," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2309–2314, Dec. 2017, ISSN: 2327-4662. DOI: 10.1109/JIOT.2017.2764475.

[4] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 39–50, Aug. 2013, ISSN: 0146-4833. DOI: 10.1145/2534169.2486015. [Online]. Available: http://doi.acm.org/10.1145/2534169.2486015.

[5] S. Wang, M. Xia, K. Huang, and Y. Wu, "Wirelessly powered two-way communication with nonlinear energy harvesting model: Rate regions under fixed and mobile relay," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8190–8204, Dec. 2017, ISSN: 1536-1276. DOI: 10.1109/TWC.2017.2758767.

[6] Y. Shu, H. Yousefi, P. Cheng, J. Chen, Y. J. Gu, T. He, and K. G. Shin, "Near-optimal velocity control for mobile charging in wireless rechargeable sensor networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 7, pp. 1699–1713, Jul. 2016, ISSN: 1536-1233. DOI: 10.1109/TMC.2015.2473163.

[7] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1–131, 2013. DOI: 10.1017/S0962492913000032.

[8] G. Laporte, "The traveling salesman problem: An overview of exact and approximate algorithms," *European Journal of Operational Research*, vol. 59, no. 2, pp. 231–247, Jun. 1992. [Online]. Available: https://ideas.repec.org/a/eee/ejores/v59y1992i2p231-247.html.

A. Appendix

A.1. Charging Model

If the transmit power at UGV is P, then the harvested power at IoT user k is $\mathrm{Y}\Big(|g_k|^2 \cdot P\Big)$, where g_k is the wireless channel from UGV to user k, and Y is the function representing the energy conversion process and is given by

$$\mathrm{Y}(P_{\text{in}}) = \left[\frac{P_{\max}}{\exp(-\tau P_0 + \nu)} \left(\frac{1 + \exp(-\tau P_0 + \nu)}{1 + \exp(-\tau P_{\text{in}} + \nu)} - 1 \right) \right]^+, \tag{A.1}$$

where the parameter P_0 denotes the harvester's sensitivity threshold and $P_{\max}$ refers to the maximum harvested power when the energy harvesting circuit is saturated. The parameters τ and ν are used to capture the nonlinear dynamics of energy harvesting circuits. For the Powercast energy harvester P2110, we have $\tau = 274$, $\nu = 0.29$, $P_{\max} = 0.004927$ W and $P_0 = 0.000064$ W.