Preliminary title:

# Robotic Navigation in Urban Environments using Reinforcement Learning and Remote Sensor Locations, and the effect of Uncertainty on Learning.

Word count: 2295

Table of Contents

[1 Introduction 3](#_Toc10636718)

[1.1 Justification for the study 3](#_Toc10636719)

[1.2 Literature Review 4](#_Toc10636720)

[1.3 Motivation and concept 9](#_Toc10636721)

[References 11](#_Toc10636722)

# 1 Introduction

## 1.1 Justification for the study

We live in an exciting time, where technology is an inevitable part of our daily life. Technology aid citizens of modern *smart cities* in varies ways, whether it is *Internet of Things* (IoT) devices monitoring traffic, weather conditions or your air condition at home, or autonomous instances (robots) delivering your parcel or driving you home (perhaps not quite yet but it does not seem far away). The influence of IoT devices is expected to increase exponentially over the next decade, with the number of IoT devices projected to exhibit an annual growth rate of around 35% through 2025 by various sources[[1]](#footnote-1).  
The field of robotics is estimated to grow by a factor of 3 to 7 over the coming 20 years, as a combination of technological advances and increased demand for autonomous assistance, across industries as well as private households (Ghaffarzadeh, 2018).   
Some of the technological advancement mentioned by (Ghaffarzadeh, 2018) is contributed to, recent as well as still to come, advances in the field of *deep learning*. Recent advancement in deep learning, as of the possibility to extract high-level features from raw sensor data (Mnih, 2013), has led to breakthroughs in *computer vision* and therethrough *reinforcement learning* (RL) [(Mnih, 2013), (Mnih et al, 2015), (Silver et al, 2016)].  
RL is one of many ways to *control* robots [(Zuo et al., 2014), (Zhang et al., 2017), (Faust, A. and Francis A., 2019), (Zeno et al., 2016)], and the use of RL can make robots truly autonomous, enabling exploration of unknown environments and learning of unknown tasks.  
This dissertation seeks to explore the combination of RL and remote sensor data for robotic navigation in dynamic environments, under uncertainty about the quality of the remote sensor data.  
The objective of the study is to be elaborated further in section 1.3 below.  
The project proceeds as follows; section 1.2 outlines the literature review of relevant fields, section 1.3 outlines the motivation for conducting the presented study, including where it fits in the literature review. section 2.. **[To be written at a later point].**

## 1.2 Literature Review

The following literature review outlines the major contributions, to the development of the field *RL* and *mobile* *robotic navigation* (MRN)*,* in a sparse format, in the light of the limited extent of this dissertation. The implication is that the history of RL is left out[[2]](#footnote-2), and the presented literature review starts from the breakthrough of *deep reinforcement learning*. Furthermore, the literature review on the field of robotics is limited to MRN. The interested reader should consult (Pandey, 2017), (Sutton and Barto, 2018) and (Juliani, 2018) among others, on which the below is largely inspired, along with footnotes for more detail.

The field of robotics is a vast field spanning many domains, with the subfield of interest herein being mobile robotics, and more specifically MRN. The subfield of mobile robotics covers indoor and outdoor domains, as well as private to industrial applications. Navigation is essential[[3]](#footnote-3) in many of the areas of mobile robots, and the three core functionalities of MRN are *self-location*, *path planning* and *map building/interpretation*. Many methods are utilised to handle the above-mentioned functionalities[[4]](#footnote-4), such as Fuzzy logic (Pandey, 2017), Simulated annealing (Pandey, 2017), unsupervised on-line classification (Kim et al, 2006), Kalman filter based on vision data (Grießbach et al., 2014) and reinforcement learning [(Zuo et al., 2014), (Zhang et al., 2017), (Faust, A. and Francis A., 2019)]. The methods present different opportunities and challenges, and some are more suitable than others for different domains. The focus of this dissertation is the challenges that dynamic urban environments presents, for which autonomous learning is suited (Faust and Francis, 2019), and in the light of the recent advances in deep learning, is deep RL the method of choice. RL can deal with all three core functionalities of MRN, yet self-location is indirect through the exploration being carried out.

The following presents a review of some major contributions to deep RL, before reviewing contributions to MRN using deep RL.

*Keep going in the same way if things are getting better, and otherwise move around.* (Selfridge, 1978).

The essence of RL is well captured in the quote from Selfridge above; behaviour associated with positive feedback should dominate behaviour associated with negative feedback, and ultimately result in an *optimal* behaviour.   
The application area of RL span a wide range of areas, yet RL is well suited for a wide range of games, because the structure of many games resembles with the dynamics of reinforcement learning. Some of the first cases were (Shannon, 1950) and (Samuel, 1959), studying the games of chess and checkers respectively.  
More recent, and deeply influential, studies are (Bellemare et al., 2013) and in continuation hereof (Mnih et al., 2013).  
(Bellemare et al., 2013) introduced a platform for researchers, for the purpose of exploring existing algorithms and encourage progress in *domain-independent AI technology* (Bellemare et al., 2013). The platform, Arcade Learning Environment (ALE), provides researchers with easy access to hundreds of Atari 2600 games, which are suitable for benchmarking algorithms. One of the first to explore the opportunities embedded in ALE was (Mnih et al., 2013), with the introduction of a *deep* structure for the neural network to learn the Q-function, which at the same time became the start of a new era. Their motivation came from recent advances in the field of deep learning, and the most influential implications was the ability to train the neural network on raw image input, combined with the use of *experience replay*. The power of using raw data is the term *raw*, i.e. no handcrafted features, which allows the neural network to truly uncover potentially non-linear structures of the data. The use of experience replay allows the agent to exploit the data to a greater extent, implying faster training, without the risk of overfitting to the data, by continuously updating its experience (Lin, 1993).

The time following these two papers, for the field of reinforcement learning, has been high-paced and revolutionary. Many important papers emerged in relatively short time, with some of the more interesting contributions being (Mnih et al., 2015) and (Silver et al., 2016).  
(Mnih et al., 2015) was especially interesting because of the generalisable ability of their implementation, which obtained above-professional-human level on 49 Atari games, and smashed the performance of all previous work evaluated on the ALE. Using their own words;

*This work bridges the divide between high-dimensional sensory inputs and actions, resulting in the first artificial agent that is capable of learning to excel at a diverse array of challenging tasks.* (Mnih et al., 2015)*.*

(Silver et al., 2016) obtained master[[5]](#footnote-5) level in the boardgame GO, which was regarded as one of the grand challenges for artificial intelligence, because of its enormous state space consisting of possible moves. The architecture employed consisted of two different neural networks, one to learn the policy function and one to learn the value function, combined with a tree-search implementation to locate the most visited move doing simulation.  
The before mentioned papers is perhaps some of the most influential papers in recent time, yet many great contributions has been seen over the last six years, developing novel and sophisticated methods. For now, these novel and sophisticated methods are left uncomment, as at least one will be explained later.

The recent advances in both deep learning and reinforcement learning has given rise to some interesting applications within the subfield of autonomous navigation, through the papers (Zuo et al., 2014), (Zhang et al., 2017) and not the least (Faust, A. and Francis A., 2019).  
(Zuo et al., 2014) embed Q-learning in a small robot, to safely and smoothly navigate it out of a maze, using input from internal sensors giving information about the robots relative position.   
(Zhang et al., 2017) goes a step further, with their objective of outlining a framework which is generalisable across many tasks. Their implementation builds on the assumption that the reward structure of a state, and thereby the future states following, can be expressed as a linear combination of *successor* features. More specifically, the idea is that different, yet somewhat similar, tasks can be expressed as linear combination of the same features. The implication is that transferring between tasks ends up being a question about obtaining new weights, compared to initialising weights, for the base model of the new task, which substantially reduces training time.

Perhaps the most interesting reference at this point, within the subfield of autonomous navigation, is a collection of three novel papers. (Faust, A. and Francis A., 2019) presents the work done in all three recent papers combining automated reinforcement learning and sampling-based planning.  
Automated reinforcement learning implies searching the reward space and the parameter space of a neural network, to obtain an optimal combination leading to the desired behaviour. In relation to sampling-based planning, (Faust, A. and Francis A., 2019) utilises *probabilistic roadmaps[[6]](#footnote-6)* before advancing to *simultaneous localisation and mapping[[7]](#footnote-7)*, to support the planning procedure by locating feasible roadmaps for the robot to follow.

The final contribution needed to be highlighted, considering the previously mentioned recent advancements, is based on the paper by (Juliani et al., 2018). (Juliani et al., 2018) introduce a toolkit for Unity – *Unity ML-agents Toolkit*, with the aim of taking the recent advances within artificial intelligence and RL even further. They note themselves;

*As the state of the field becomes more developed, existing environments and the benchmarks based on them become less informative, the need for novel environments presents itself.* (Juliani et al., 2018).

Ground-breaking discoveries are based on the formulation of new grand challenges, and the introduction of Unity ML-agents Toolkit brings endless possibilities, in terms of formulating tasks and environments with ever increasing complexity. By the authors;

*When examining the areas of human intelligence typically studied, we find four major ones which have been the focus in recent years, and which we expect to continue to be the focus into the future: sensory, physical, cognitive, and social. … To challenge algorithms within these four domains, environments should be able to test all four axes of intelligence simultaneously.* (Juliani et al., 2018).

The increasing desire to humanise robots, is well supported by the functionalities of Unity, based on the above citation. Unity seems as an obvious choice of training ground, for autonomous instances, irrespectively of the application area, to sustain the highest degree of assistance from day one. Unity is the choice of software for this dissertation.

## 1.3 Motivation and concept

The interest of the author lies in learning and how learning in general can be supported. Urban environments present interesting challenges, in terms of learning, because of the dynamic and adverse nature of cities. One common way to support learning, is to provide external information, and let the agent figure out how to use that information for optimal learning – that is what we see in the lecture room and that is what this dissertation seeks to explore the effect of for learning in autonomous instances.   
This dissertation seeks to explore the effect of external information (location of remote sensors) for learning in a simplified setting, since interacting with remote sensors (IoT devices) within a real-life city raises serious ethical questions and imply computational challenges beyond the scope of this dissertation.

The idea behind using the location of remote sensors as external information to guide learning, comes from the assumption that a remote sensor represents a potential conflicting object (a pedestrian, and the sensor reflecting their cell phone). Areas with high density of remote sensors is potentially beneficial to avoid when searching for an optimal route to the goal destination. Therefore, is the navigation taking place referred to as global navigation, within the literature. The aim of the conducted research is not to avoid individual objects, which require implementations out of the scope for this dissertation, but to avoid choosing a global optimal path which conflicts with local dynamics. The use of the external information can be seen as equipping the agent with near-local vision, allowing the agent to take action on a higher level, before a potential conflict occurs and low-level action is needed. This is, to the knowledge of the author, not something that is covered in the literature at this point.

Another interesting angle is that sensors aren’t perfectly reliable (Zeno et al., 2016), because they can suffer from *sensor noise*, *malfunction* etc. That implies some degree of uncertainty around the appearances of remote sensors, which potentially could affect learning. This is also something that this dissertation explores in the simulation models.

The simulation models are built in Unity and the base environment is a 3D grid world, meant to represent an urban environment. The models have increasing complexity, in terms of the objects present, to better capture the challenges of real-life cities. Autonomous agents are created using Unity ML-agents Toolkit, enabling the use of state-of-the-art deep learning methods to facilitates learning in the agents.  
Areas with varying degree of densities, of remote sensors, are represented as areas associated with varying degree of negative reward, should the area be passed through. Different degree of density could be modelled in different ways, which will be discussed at a later point.

## References

1. **Sutton, R. S. and Barto A. G. (2018). Reinforcement Learning – An introduction. Second Edition, MIT Press, Cambridge, MA.**
2. **Sutton, R. S. and Barto A. G. (1998). Reinforcement Learning – An introduction. First Edition, MIT Press, Cambridge, MA.**
3. **Samuel, A. (1959). Some studies in machine learning using the game of checkers. IBM J. of Research and Development 3, 210-229.**
4. **Shannon, C. E. (1950). Programming a computer for playing chess. Philosophical Mag. 41, 265-275.**
5. **Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M. (2013). Playing atari with deep reinforcement learning. ArXiv:1312.5602.**
6. **Bellemare, M. G., Naddaf, Y., Veness, J., Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. Journal of Artificial Intelligence Research, 47:253–279.**
7. **Long-Ji Lin. (1993). Reinforcement learning for robots using neural networks. Technical report, DTIC Document.**
8. **Mnih V, Kavukcuoglu K, Silver D, et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540): 529-533.**
9. **Silver, David, et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature 529.7587: 484-489.**
10. **Zuo, B. et al. (2014). A reinforcement learning based robotic navigation system.** [**2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6960119)**.**
11. **Zhang, J. et al. (2017). Deep Reinforcement Learning with Successor Features for Navigation across Similar Environments.**
12. **Faust, A. and Francis A. (2019). Long-Range Robotic Navigation via Automated Reinforcement Learning. Google AI Blog. Online:** <https://ai.googleblog.com/2019/02/long-range-robotic-navigation-via.html> **Accessed: 01/06/2019.**
13. **Faust, A. et al. (2018). PRM-RL: Long-range Robotic Navigation Tasks by Combining Reinforcement Learning and Sampling-based Planning. *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia (2018), pp. 5113-5120.**
14. **Francis, A. et al. (2019). Long-Range Indoor Navigation with PRM-RL.**
15. **Juliani, A. et al. (2018). Unity: A General Platform for Intelligent Agents.**
16. **Columbus, L. (2018). 2018 Roundup of Internet Of Things Forecasts and Market Estimates. Online:** <https://www.forbes.com/sites/louiscolumbus/2018/12/13/2018-roundup-of-internet-of-things-forecasts-and-market-estimates/#449f1bef7d83> **accessed 04/06/2019.**
17. **Ghaffarzadeh, Dr. K. (2018). New Robotics and Drones 2018-2038: Technologies, Forecasts, Players. Online:** <https://www.idtechex.com/en/research-report/new-robotics-and-drones-2018-2038-technologies-forecasts-players/584> **Accessed 04/06/2019.**
18. **Zeno, P. J., Patel, S., and Sobh, T. M (2016). Review of Neurobiologically Based Mobile Robot Navigation System Research Performed Since 2000. Journal of Robotics, vol. 2016, Article ID 8637251, 17 pages.**
19. **Pandey, Dr. A. (2017). Mobile Robot Navigation and Obstacle Avoidance Techniques: A Review. International Journal of Robotics and Automation, May.**
20. **Grießbach, D., Baumbach, D., and Zeuv, S. (2014). Stereo-Vision-Aided Inertial Navigation for Unknown Indoor and Outdoor Environments. International Conference on Indoor Positioning and Indoor Navigation.**

1. Ericsson and IoT Analytics among others (Columbus, 2018). [↑](#footnote-ref-1)
2. See (Sutton and Barto, 2018) for a comprehensive coverage. [↑](#footnote-ref-2)
3. (Pandey, 2017) [↑](#footnote-ref-3)
4. Consult the references for many more methods than the listed. [↑](#footnote-ref-4)
5. Page 5, (Silver et al., 2016). [↑](#footnote-ref-5)
6. See (Faust et al. 2018) for details. [↑](#footnote-ref-6)
7. See (Francis et al. 2019) for details. [↑](#footnote-ref-7)