# REINFORCEMENT LEARNING TECHNIQUES FOR IDENTIFYING SOCIAL SPACE MODEL OF HUMAN-ROBOT INTERACTION

**BY**

**MR. PAKPOOM PATOMPAK**

**A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF**
**THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF**
**PHILOSOPHY (ENGINEERING AND TECHNOLOGY)**
**SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY**
**THAMMASAT UNIVERSITY**
**ACADEMIC YEAR 2019**

# REINFORCEMENT LEARNING TECHNIQUES FOR IDENTIFYING SOCIAL SPACE MODEL OF HUMAN-ROBOT INTERACTION

BY

MR. PAKPOOM PATOMPAK

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY (ENGINEERING AND TECHNOLOGY)
SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2019

# REINFORCEMENT LEARNING TECHNIQUES FOR IDENTIFYING SOCIAL SPACE MODEL OF HUMAN-ROBOT INTERACTION

A Dissertation Presented

By

PAKPOOM PATOMPAK

Submitted to
Sirindhorn International Institute of Technology
Thammasat University
In partial fulfillment of the requirement for the degree of
DOCTOR OF PHILOSOPHY (ENGINEERING AND TECHNOLOGY)

Approved as to style and content by

| | |
|---|---|
| Advisor and<br>Chairperson of Thesis Committee | (Asst. Prof. Dr. Itthisek Nilkhamhang) |
| Committee Member and<br>Chairperson of Examination Committee | (Dr. Kanokvate Tungpimolrut) |
| Co-Advisor | NakYoung Chong<br>(Prof. Dr. Chong Nak-Young) |
| Committee Member | (Assoc. Prof. Dr. Waree Kongprawechnon) |
| Committee Member | Toshiaki Kondo<br>(Assoc. Prof. Dr. Toshiaki Kondo) |
| External Examiner: | Prof. Dr. Takayuki Kanda |

DECEMBER 2018

| | |
|---|---|
| Thesis Title | Reinforcement Learning Techniques for Identifying Social Space Model of Human-Robot Interaction |
| Author | Mr. Pakpoom Patompak |
| Degree | Doctor of Philosophy (Engineering and Technology) |
| Faculty/University | Sirindhorn International Institute of Technology<br>Thammasat University |
| Thesis Advisor | Assistant Professor Itthisek Nilkhamhang, Ph.D. |
| Thesis Co-Advisor | Professor Chong Nak-Young, Ph.D. |
| Academic Years | 2019 |

# ABSTRACT

Human-robot interaction in a shared environment is a critical component of service robotics. The ability to perceive, understand and act in a manner that conforms to social convention is the fundamental key to human-robot symbiosis. Notably for navigation tasks, the robot should take into consideration human's social space, defined as the area that we feel comfortable to interact with other humans or robots. The primary task of the robot is to understand and identify this social space by learning and adjusting to human response. Reinforcement learning, which is a machine learning technique that attempts to maximize the accumulated reward through trial-and-error, can be used to update the parameters of the social space model by learning from previous human-robot interactions. However, different reinforcement learning algorithms exist that may or may not be appropriate for human-robot interaction in a real-world scenario.

This thesis focuses on studying the efficacy of reinforcement learning algorithms for parameter adaptation of human's social space model. We study and analyze the performance of popular reinforcement learning algorithms in terms of respective advantages and disadvantages. Methods for accelerating reinforcement learning to meet real-world requirements are explored. Simulation results are presented and compared that examines the efficiency of each reinforcement learning algorithm, as well as its suitability for adapting the parameters of the social space model.

**Keywords**: Social Robotics, Human-Robot Interaction, Social Conventions, Human-Robot Symbiosis, Social Space, Machine Learning, Reinforcement Learning.

# ACKNOWLEDGMENTS

Sirindhorn International Institute of Technology (SIIT) offers three master of engineering programs, namely, Master of Engineering Program in Engineering Technology, Master of Engineering Program in Information and Communication Technology for Embedded Systems, and Master of Engineering Program in Logistics and Supply Chain Systems Engineering. It also offers a Master of Science Program in Engineering and Technology and Doctor of Philosophy Program in Engineering and Technology.

Sirindhorn International Institute of Technology (SIIT) offers three master of engineering programs, namely, Master of Engineering Program in Engineering Technology, Master of Engineering Program in Information and Communication Technology for Embedded Systems, and Master of Engineering Program in Logistics and Supply Chain Systems Engineering. It also offers a Master of Science Program in Engineering and Technology and Doctor of Philosophy Program in Engineering and Technology.

Sirindhorn International Institute of Technology (SIIT) offers three master of engineering programs, namely, Master of Engineering Program in Engineering Technology, Master of Engineering Program in Information and Communication Technology for Embedded Systems, and Master of Engineering Program in Logistics and Supply Chain Systems Engineering. It also offers a Master of Science Program in Engineering and Technology and Doctor of Philosophy Program in Engineering and Technology.

Mr. Pakpoom Patompak

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

Figures Page

# CHAPTER 1
# INTRODUCTION

## 1.1 Importance and Challenge

Mobile robots are trended to provide more and more service in a shared environment with humans, e.g., house, office or co-worker space. The application of mobile robots has ranged from a co-worker robot in the industries to domestic services robot that assists humans in their home, for example, the robot that takes care of the elderly person or handicapped person in the house. For a task that requires robots to move and provide services in a different location, the safe navigation is one of the essential functions that the robot needs to concern. The robot needs to generate a path that does not harm or damage the surrounding environment which includes humans. However, even when the robot is moving safely, sometimes human feel that the motion is not safe because of lacking trust in the technology Kruse et al. (2013). These lacking of trust feeling may occur from the lack of experience to the technology or unfamiliar of the robot appearance. Therefore, the navigation of the robot should not consider only on safe behavior but should increase attention to generate socially competent behavior like moving naturally or considering to the human comfort space. This concept makes the robot to behave more naturally and acceptable for humans to feel safe and comfortable to operate with the robot in the human-robot shared environment.

Human-aware navigation is one of the crucial challenges for human-robot symbiosis. There are two majors constraints that roboticist should consider to design the human-aware navigation. The first is an instance of a safety constraint which is useful to prevent situation potentially harmful to the robot and the surrounding environment Fox et al. (1997). The second is an instance of social constraint such as which is helpful to avoid situations potentially annoying or making discomfort to humans. Early robotics research had mainly focused only on the first constraint. For example, Fox et al. (1997); Stachniss and Burgard (2002) use dynamic window approach which consists of velocity reachable within a short time interval, to define the small neighboring area. This method determines the stop area before the robot colliding with obstacles. However, recent developments have integrated the early research with social science and psychology studies to meet both constraints. For instance, The design of approaching behavior in Althaus et al. (2004). They suggested maintaining a certain distance to the closest person, and the robot body should face the middle of the group. Saulnier et al. (2011) shows the investigation of how different kind of nonverbal cues of the robot can catch the attention of the human. Their study shows that the navigation behavior can serve the messages as the nonverbal communication.

The famous theory from social science that used to develop social navigation is the *Proxemics theory* which is the social and psychological concept Edward (1969). This theory describes how human use surrounding space and effect that population density has on behavior, communication and social interaction. In other words, proxemics theory describes different interpersonal space that the human keep from others depends on social information like cultures, personal traits, age or relationship. These interpersonal spaces can represent the area of privacy that the human does not want to share with others.

The same concept can be applied and integrated into the case of human-robot interaction and human-aware navigation, especially when the robot is required to exhibit a certain level of social competence. However, it is still a challenging problem to formalize this social science and psychological concept into a mathematical model which represent both safety and socially constraints. This work uses the *Social Force Model* (SFM) which introduced by Helbing as the based mathematical model of the human. His work based on the potential field concept which gives the different degrees at different locations in the environment Helbing and Molnár (1995). These degrees can be assigned to useful information like the acceptable or comfort feeling, or the quality of interaction of each human in a shared space. However, the size or boundary of the human area of privacy that human build while interacting with other humans, depends on social factors like acquaintanceship, genders or personal traits

Takayama and Pantofaru (2009). Therefore, to model human social space for human-aware navigation or human-robot interaction, these social factors should consider estimating the human area of privacy more precisely Patompak et al. (2016). Here, the mapping methods like fuzzy inference system played a role as the tool to map this social information and estimated the parameters to calculate the privacy area of each person by SFM Patompak et al. (2016, 2017). This SFM then can be the map with the degree of human's feeling or social constraint that assist the robot in generating the path that respects both safety and socially. Finally, the robot will behave the social navigation according to the generated path.

Even though the understanding of the relationship between an area of privacy of the human and the social factors allows the robot to estimate human's privacy area, but it still has uncertainties that originate from humans and the surrounding environment that the roboticist may overlook when design method to estimate human's privacy area. These uncertainties can be like the difference of cultures or lifestyle of each person, that makes the robot to estimate inaccurate human's privacy area. Therefore, learning ability is another essential ability that roboticist should consider equal to the understanding ability.

In humans learning process, humans try to solve the uncertainties problem based on their experience. Humans improve their performance by obtaining the experience and knowledge from the surrounding environment's feedback signal. This experience must be obtained by interact and observe the results. Then, humans use these results as the base to make better decisions to improve performance when facing the same situation. A framework that using the experience to improve the performance of the agent is similar to one of the machine learning, *reinforcement learning* (RL). The concept of RL is to reinforce the decision that has led to good outcome according to the experience by increasing the chance to perform the same decision again. The same concept can be applied to the robot as the learning ability to solve the uncertainty problem of environment Patompak et al. (2017). Therefore, during the operation, the robot generates a path according to the estimated privacy area, and interact with the human at the boundary. However, with the preliminary setting of the mapping social information process, the aberration of the estimated area can occur due to the uncertainty of humans. Here, learning from the human's response take place to adjust the parameters that can make the estimated area more accurate and appropriate to the human.

In this work, the RL algorithms are applied to the problem of parameters adaptation for the human's social space or the area of privacy estimation. Starting with Q-learning, which is the famous RL technique, is the first learning technique that used to solve the problem due to the easiness of technique and does not require a model of the environment. However, Q-learning requires the discount factor that is the factor that makes the agent care the immediate reward more than the previous reward, which is difficult to choose the appropriate discount factor value to the problem. Another method is the average reward approach, R-learning. This method makes the agent care every reward by comparing with the average reward. This method is believed that suit to continuous problems that interaction between agent and environment go on and on forever without termination or start state. Both of these techniques is the value based where the agent learn a value function that maps each state-action pair to a value. From that value map, the agent can find the best action to take for each state. Next method is actor-critic which actor controls how our agent behaves, and critic measures how good of the action taken is. Finally, deep learning is applied to RL techniques to improve the RL in the large state-action space. This deep RL has used the benefit of replay buffer to store the experience and use it to improve the learning process. Therefore, this thesis studies the efficacy of each RL on the to the problem of parameters adaptation for the human's social space or the area of privacy estimation, to compare which algorithm is more suitable for this problem.

## 1.2 Research Objective

Our main goal with this work is to study the efficacy of popular reinforcement algorithms for parameters adaptation of human's social space model and way to accelerate reinforcement algorithm to meet the real-world task. In finally the robot should be able to adapt the estimate human's social space to appropriate to humans in shared-environment.

## 1.3 Research Motivation

Our motivation for working with reinforcement learning in human's social space come from the requirement of the robot to operate naturally and acceptable in the shared environment with humans. Our premise is that once the robot understands the relationship between human's social space and social factor, the robot will be able to estimate human's social space and behave to interact with the humans in the appropriate distance which does not annoy human's comfortable feeling. The reinforcement learning algorithms can be applied to adapt the parameters of the estimation function by learning from the response or feedback from the humans. The different algorithms of RL have different results to the problem which appealing to us to find the suit algorithm to use in the real-world task.

## 1.4 Thesis Outline

This thesis organizes as follows. Chapter 2 show the review of works that related to our research. It presents the summary of the social model in human-robot interaction and reinforcement learning which are necessary for understanding this thesis. Chapter 3 discuss the social model for robot navigation. In this chapter, the mathematical model of the human's social space is described. It also covers the process to map social factors to human's social space, and how to formalize our social model to reinforcement learning problem. Chapter 4 discuss the detail of reinforcement learning that we use to deal with the parameters adaptation for human's social space estimation. It also includes the accelerated version of reinforcement learning algorithm. Chapter 5 shows the comparison results of the simulation between each algorithm. The last chapter gives a concluding remark and direction for future research.

# CHAPTER 2

# LITERATURE REVIEW

This section has attempted to provide a summary of the literature relating to our work. This chapter began by briefly introduces the concept of human-robot interaction (HRI) which include the knowledge of social science that useful to human-aware navigation, and the application in robot navigation in the human shared environment. Then provide the information about reinforcement learning that will be used to determine the efficacy of its techniques to parameter adaptation.

The first section has endeavored to grasp the definition of the human's area of privacy in social science and exemplified the studies to support its definition. Then the studies that relate to robot application like human-aware navigation are present. The second section has provided the necessary information about reinforcement learning (RL) and shown the beauty of the variety of its applications that can be used in any application.

## 2.1 Human-Robot Interaction

Human-Robot Interaction (HRI) is a field of study to understanding, designing, and evaluating robotic systems for use by or with humans Sheridan (2016). Interaction can be several forms such as speaking to each other, operating in the same area, walking companion or guiding to the destination location Kirby et al. (2009); Pandey and Alami (2010). However, to design the robot system to operate naturally and acceptable in the shared environment, the roboticist should understand the social information from humans' behaviors. Therefore, the knowledge and comprehension of social science and psychology are vital to model and design human-robot interaction. The goal of this section is to present some definition of social science study that useful for human-robot interaction and discuss challenge problems that are likely to shape the human-robot interaction field in the near future.

## 2.1 Privacy area of human in Social Science

Human-Robot Interaction (HRI) is a field of study to understanding, designing, and evaluating robotic systems for use by or with humans. Interaction can be several forms such as speaking to each other or operating in the same area. However, to design the robot system to operate naturally and acceptable in the shared environment. The roboticist should understand the social information from humans' behaviors. Therefore, the knowledge and comprehension of social science and psychology are vital to model and design human-robot interaction. The goal of this section is to present some definition of social science study that useful for human-robot interaction and discuss challenge problems that are likely to shape the field in the near future.

## 2.2 Privacy and Proxemics in SocialScience

The key to formalize human social space model is to understand and accommodate human behavior. Therefore, the knowledge of social science and psychology is a vital aspect which allows the robot to better understand the behavior of the humans. When robots operate in a shared environment, the area of privacy is the crucial key for naturalness, sociability, and acceptance. Privacy was defined in human-human interaction study by Jonathan Herring Herring (2014). They defined privacy as the ability of an individual or group to separate themselves and select to share some of their information to whom they allow. The boundaries and content of what is considered private differ between cultures and individuals.

There is a lot of research that exemplifies the study of privacy of the living things. For instance, Westin Westin (1970) mentioned that most animals seek privacy either as individuals or in small groups. In this study, he reported three areas of privacy observed among animals which included: personal distance between animals, social distances between groups, and fighting distances at which an intruder cause conflicts. At the same time, animals often gather in large groups. They seem to live in a tension between privacy and sociability.

Zeeger studied human privacy in childhood Zeeger S. K. (1994). He found that 58 of 100 of three to five-year-olds said they had a special place at the daycare center which belonged only to them. Newell et al. investigated the reason why humans required privacy. By the survey, they found that most of the participants in different cultures believed that emotion like grief, fatigue or attention were the main useful sets associated with seeking privacy Newell (1998).

Another study about the spacing of human was introduced by Edward T. Hall Edward (1969). He introduced the theory of Proxemics, which describe how humans use space, and the effect that population density has on behavior. He emphasized the impact of the use of space on interpersonal communication. According to his study, Proxemics is valuable in organizing the surrounding space to interact with others. These organized spaces depend on the type of interaction and social information between individual. Therefore, the human interaction area could be organized as follows:

- **Intimate Space** is an area for intimate contact like whispering, touching or hugging with very close relationship person like wife and husband or mom and children. This area has a distance less than 0.46 meters with respect to the human's center.

- **Personal Space** is the zone for people who have close relationships. In this space, humans feel discomfort if unfamiliar being enters this area. This area has a distance greater than 0.46 meters but less than 1.22 meters.

- **Social Space** is the space that humans use to contact with new acquaintances. The distance of this area is between 1.22 and 3.70 meters.

- **Public Space** is the space that often used to interact with strangers or to give public speeches. The distance of this area is more than 3.70 meters.

The organized space for human interaction is shown in Figure 2.1. On the one hand, humans use these organized space concepts to approach others humans. For example, humans try to get near to the close friend or the member in the family to get a better quality of interaction; however, they keep the distance or space from the

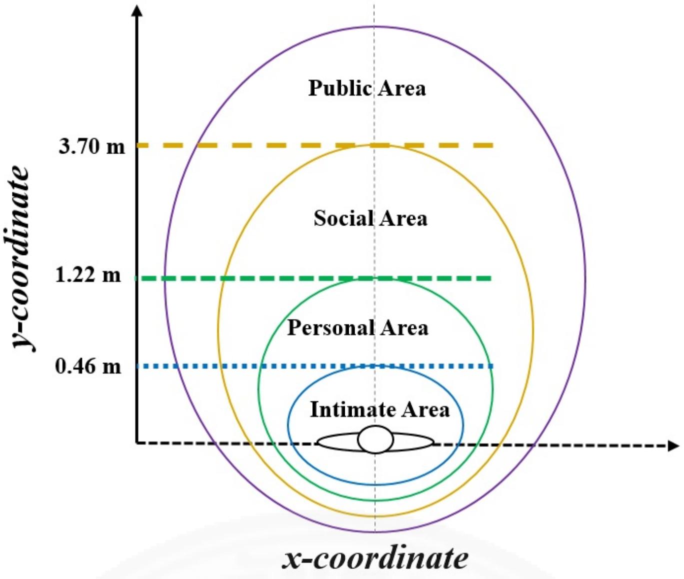


Figure 2.1 Human interaction area according to "Proxemics" theory which introduced by Edward T. Hall

strangers to increase comfortable feeling. Consequently, it is evident that closeness is paramount for good interaction, but an area of privacy should also be respected.

Furthermore, protecting one's privacy is an essential prerequisite for forming long-term, stable relationships. This concept can be applied to develop the social robot. For example, in the approaching to the human problem in Takayama and Pantofaru (2009) or the problem of path planing in crowded environment in Martinez-Garcia et al. (2005). Human-aware navigation is the topic of research that assists the robot to navigate in the human-robot shared environment. The next section will provide the studies in human-aware navigation that relate to the area of privacy of the human.

## 2.3 Human-Aware Navigation

There are different goals that human-aware navigation research follows. Most of the research attempt to minimize annoyance, stress, and discomfort, so that the robot can interact more comfortably with humans Elena Pacchierotti (2006); Martinson and Brock (2007); Pandey and Alami (2010). Other approaches focus on the robot behaving more naturally and behaving according to the social norm. All of these goals have a common theme that attempts to make the robot acceptance to humans.

However, the method may vary. Therefore, the following definition of naturalness, sociability, and comfort can be used to classify the research reviewed.

**Naturalness**

Naturalness is the similarity between robots and humans in low-level behavior patterns. This group of research strives to imitate nature, such as human motion, as the target behavior to recreate the robot's behavior. Most research in this group works well with low-level behaviors like shapes and velocities where a continuous measure can be applied between human-robot behavior.

Natural motion is another goal in human-aware navigation that attempts to make robots navigate more acceptably near humans by imitating human behaviors. The assumption of natural behavior research is if a robot behaves more similarity to humans, the interaction between them becomes easier and more intuitive for humans Kruse et al. (2013).

One aspect of natural motion is smoothness. This aspect refers to both the geometric path and the velocity of the robot. For example, Arechavaleta et al. (2008) presented that a principle of energy optimization influences human motion. They summarized that the behavior of to approach the group of humans like the speed of movement should depend on the distances between the robot and humans. The robot should slow down when getting closer to not scare anybody. Hence, it is possible to observe human motions with near minimum jerk.

Another aspect is the motion relative to other agents, like humans or robots, walking side by side or following each other. Gockley et al. Gockley et al. (2007) did the experiment to investigated two different approaches of person-following such as direction following and path following, to see which approach is more natural motion. These two different approaches have been rated by the participant in the pilot study, as shown in figure 2.2. The results showed that it was favored to make the robot share the same direction or following the direction rather than following the same path as the humans.

A behavior design of approaching a group of human and maintaining formation

Figure 2.2 The experiment to investigate the natural motion of person-following in hallway.

was presented in Althaus et al. (2004). They suggested maintaining a certain distance to the closest person, and the robot body should face the middle of the group. Another investigation focused on how different kinds of nonverbal cues were used by the robot to catch the attention of humans was presented by Saulnier et al. Saulnier et al. (2011). They designed the path planning based on the cost-map function to the robot arm. This robot arm tried to pick up an object and send to the human according to the desired path. This study shows that the navigation behavior can serve as the messages for nonverbal communication. In addition, this is another natural language that must be considered for avoiding misunderstanding.

Gracia et al. Martinez-Garcia et al. (2005) and Tamura et al. Tamura et al. (2012) used Social Force Model (SFM) as a means to guide the robot to a group of humans in a natural way. Social Force Model (SFM) represents moving agents like a robot or human as a mass under virtual force. Thus, the agent can move to its goal while being repelled by obstacles and other agents. This model can be used as the cost function and the input to the robot motion control. Another special challenge for the robot is to move into densely crowded areas. For example, the strategy of making a robot exhibit human-like behavior in the highly populated environment Müller et al.

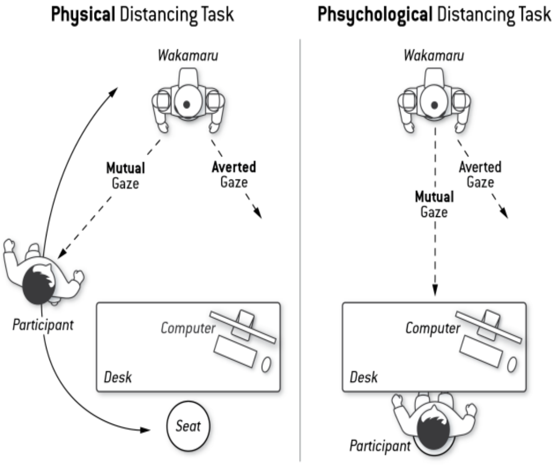


Figure 2.3 The experiment for physical distance task and psychological task that have set to explore that social norm is effected to these distances. These could be used to design proxemics behavior for the robot.

(2008); Tamura et al. (2012).

### Sociability

Sociability adheres to explicit high-level culture conventions. This group of research is concerned with the different abilities of the robot compared to humans and how not all of these abilities are suitable to transfer from humans to the robot. Therefore, it is considered that the robot are able to make the same high-level decisions like humans. Protocols consider for sociability are constraints imposed by society. For example, the rule to walk on the right-hand side in corridors, or to approach the human concerning the social relationship information.

Human-aware navigation can be improved by adding behavior that considers social protocols for behavior in a particular situation. In navigation, there are rules such as standing in queues, excusing oneself when one has to traverse a personal zone to reach a goal, and so on. Consequently, the robot should understand social rules or social signals to behave correctly to the human in human interactions. Amount of research studies improved approach direction initiate explicit interaction Althaus et al. (2004); Butler and Agah (2001); Dautenhahn et al. (2006); Jens Kessler (2011); K. L. Koay and Alami (2007); Mumm and Mutlu (2011); Takayama and Pantofaru (2009). They suggested that violation of social rules or social signal can also cause the discomfort of the humans.

In Mumm and Mutlu (2011), the experiment is constructed to explore whether the proxemics model can explain how people physical and psychological distance themselves from the robot. This also guideline how to use proxemics behavior for the robot. The experiment was set two test with two different distances. The participant asked to approach the robot to do some task and measure physical distance. Then the robot asked their personal distance to measure the psychological distance as shown in figure 2.3. The results show that the person who did not like the robot maintained the physical distance with the robot when the robot moved its gaze, and also disclosed less personal information to the robot.

Another example interesting experiment is in Dautenhahn et al. (2006). The experiment was set to investigate the direction that the robot should use to approach. The robot is controlled to approach the participants in a different direction. The results show that most of the participants did not prefer the robot to approach from the front direction. They preferred the robot to approach at the side especially the right side of them. The extended experiment from this paper includes the behavior that the robot hand the can of soft drink to the human K. L. Koay and Alami (2007). The robot handing over human' hand position had the most influence on determining from where the robot approach.

All of these research suggested that violation of social rules or social signal can also cause the discomfort of the humans.

### Comfort

Comfort is the absence of annoyance and stress for humans when interacting with robots. The comfort is different from safety. Even when the robot moves safely in human's zone, humans may still feel unsafe because of the lack of trust in technology, due to being unfamiliar with the appearance or the robot type. Therefore, the research on comfort attempts to not only make the robot move safely but also manoeuver to make human feel more relaxed.

When the robot is moving toward its destination, it can cause discomfort to humans by moving too close, too fast, too slow or getting in the way. This section presents the research that points out the causes of discomfort reactions felt by humans and how to alter the robot's behavior to reduce this discomfort. Most of the given literature, on comfort requirement, stresses the importance to the distance a robot needs to keep from humans. This distance does not serve collision avoidance but prevents the feeling of discomfort.

Edward T. Hall proposed the concept of virtual personal space around a person which others should respect called *Proxemics*. He found differences in social space that humans chose for human-human interaction depending on the relationship and intention. The idea is that when interacting with other agents, humans feel annoyed or show signs of discomfort when others get too close or too far away. This general idea of Proxemics can be applied to the appropriate social space chosen by the robot for any explicit or implicit interaction with humans. Paccheierotti Elena Pacchierotti (2006); Pacchierotti et al. (2006) presented the studies of the robot navigation in a hallway with humans walking in opposite directions, as shown as figure 2.4. In their studies, they applied a control strategy as a reaction to humans coming the other way. The robot deviated a larger lateral distance making participants feel better. However, on a few occasions, a large lateral distance was evaluated as unnatural. Butter and Agah Butler and Agah (2001) presented several experiments where a robot approached a standing person. They found that different types of the robot, such as vacuum robots or humanoid robots, caused different levels of discomfort even at the same distance. Takayama and Pantofaru Takayama and Pantofaru (2009) extracted the robot design for approaching distance and gaze depending on social contexts. They summarized that the familiarity with robots and attitude towards them should be taken into account for social space selection.

To take the aspect of comfort beyond the definition of "a distance to maintain," Martinson Martinson (2007a,b) takes into account the noise generated by the robot's motion itself and presents an approach to generate a hiding path while moving around a person. A model of human awareness to navigate the robot in a way that reduces noise discomfort is present in Martinson and Brock (2007). Another way to be

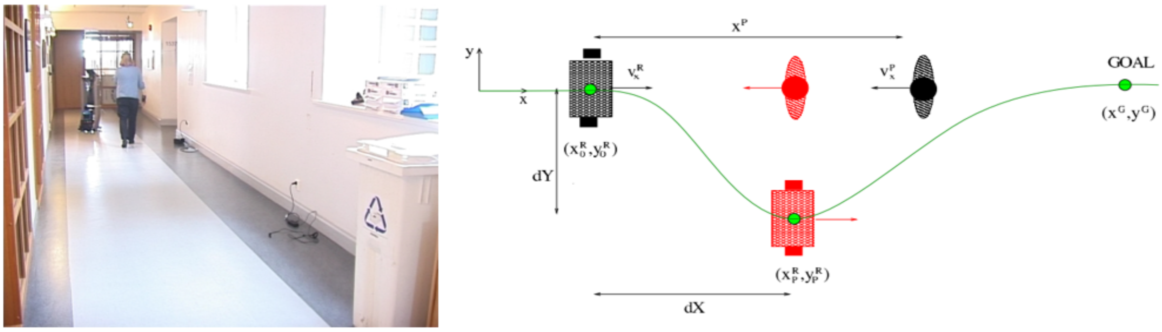


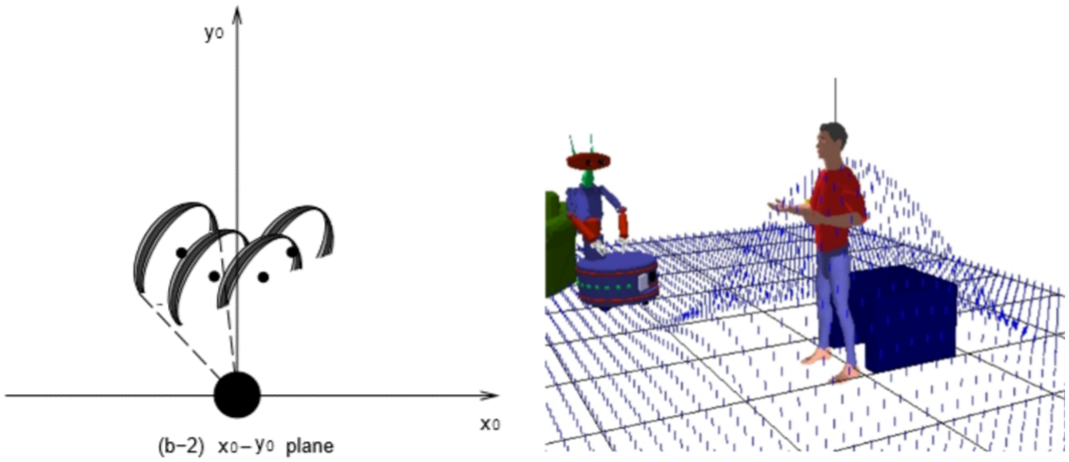
Figure 2.4 The experiment located in the hallway. Pacchierotti et al. desired the robot to move to the left side of humans by keep the large distance as possible to give the human more comfort Pacchierotti et al. (2006).

comforting to others is not to disturb them unless it is necessary. For example, Tipaldi et al.Diego and Arras (2011) have addressed this aspect by programming the robot to navigate in areas which avoid potential interference with others while performing tasks like cleaning the home. They tried to maintain a "spatial affordance map" which contains probabilities for human activities at time intervals as poison process. This map presented the location of humans from observations in their environment and allowed for the robot to reason about whether its activity was likely to occur in the human-robot shared environment.

## 2.3 Human Social Model

Human-aware navigation of mobile robot should consider two constraints. The first is the task constraints which include minimizing the distance traveled toward a goal, avoiding obstacles and keeping a safe distance from them Lam et al. (2011); Nakauchi and Simmons (2000). This task constraint is considered to be a major significance in every research of robot navigation. An additional constraint is the social constraints that include the social convention, such as comfort, naturalness, sociability Huang et al. (2010); Kirby et al. (2009),Sisbot et al. (2007). The challenge is how to formalize both constraints into mathematics model. Therefore, in this section, we show the research about human social space model that is used as the base for the robot navigation.

Using the concept of Proxemics, Lindner summarized that the interaction area could be delineated based on the geometric and potential field Lindner (2015), as



(a) (b)

Figure 2.5 Human social space can be model based on geometric 2.5a and cost function like potential field concept 2.5b. For example, the ellipse function was used to determine the personal space of human in line scenario Nakauchi and Simmons (2000), while Sisbot et al. use the potential filed concept as the cost function with depend on human posture Sisbot et al. (2007).

shown in figure 2.5. The geometric model has a clear boundary to represent a sharp transition between different zones of interaction area Pandey and Alami (2010), for example, explicit model to represent personal space in nearest histograms for local navigation in Lam et al. (2011); Tomari et al. (2012), or using the ellipse function to determine the humans' personal space in queue scenario. This help the robot to approach or avoid the person in queue Nakauchi and Simmons (2000), as shown in figure 2.5a.

Another method to describe interpersonal space is to define the cost function or potential field Hansen et al. (2009); Kirby et al. (2009); Scandolo and Fraichard (2011); Sisbot et al. (2007); Svenstrup et al. (2010). The cost function is superior to solutions defining forbidden zone around humans Huang et al. (2010) because in the limited space the cost function can be useful and necessary for the robot to move past humans. This cost function and potential field concept were used in Kirby et al. (2009); Papadakis et al. (2014) to prevent the robot from getting into the human's forbidden area. Hansen et al. employed Rapidly Exploring Random Tree (RRT) in the social map which provides the degree of comfortable of humans and gained the response from the robot dynamics and human motion prediction Hansen et al. (2009).

A model of the level of comfort in humans' field-of-view and posture was used as the cost to guide a human aware motion planner Sisbot et al. (2007), as shown in figure 2.5b. The potential field was used as the cost function to assist the robot in determining the position to approach humans Jens Kessler (2011).

## 2.4 Fuzzy Inference System

Fuzzy inference system (FIS) is the computation technique following an approach that is considered to be somewhat similar to both human reasoning and decision-making process Zadeh (1965). This thesis uses FIS to choose the appropriate Gaussian parameters value that used to estimating human social space from social factors. Therefore, this part will explain the concept of FIS that will be used in the research.

In most of the decision-making process, the capacity for addressing uncertainty and imprecision is a key that influences the quality of the decision. Precisely determine the value for the parameters that used to estimate human social space is difficult because of imprecise of the robot cognitive abilities. For example, the relationships of the robot to the human are frequently described in term of 'familiar to the human' rather than ' 80 percent of the total time with the robot'. In addition, the human social factors which are describe in chapter 3 are defined based on the linguistic variable. Therefore, imprecision between social factors and the value of parameters can be easy to determine.

In this thesis, FIS is incorporated to social space model estimation to map the social factors like genders, level of the relationship, or perception distance of the human to determine the value of the parameters that will be used to determine the human social area. The FIS process includes four parts, fuzzy membership function (MFs), fuzzy rules, Inference method, and a defuzzification method which will be described as follows:

## 2.4 Fuzzy Membership Functions

Under fuzzification, a crip value of input variable can be transformed into a fuzzy value with membership degree by their membership function (MFs). These MFs should be defined for each input and output. Let $\mathcal{U}$ represent the universe of discourse, with elements of $\mathcal{U}$ denote by $x$. A set $\mathcal{A}$ is a fuzzy subset of $\mathcal{U}$. The degree to which an element $x$ belongs to set $\mathcal{A}$, which is a real number between 0 and 1, is called a membership function (MFs) values $\mathcal{A}(x)$ in a fuzzy set $\mathcal{A}$.

The meaning of a fuzzy set $\mathcal{A}$ is characterized by an MF $\mu_{\mathcal{A}}$ that maps elements of the universe of discourse $\mathcal{U}$ to their corresponding membership value $\mathcal{A}(x)$.

$$\mathcal{A}(x) = \mu_{\mathcal{A}}(x) \in [0,1], x \in \mathcal{U} \tag{2.1}$$

The fuzzy MFs $\mu$ can be represented by a variety of shapes such as triangles, trapezoids or bell shape Gaussian function, depending on how the expert relates different domain values to belief values.

## 2.4 Fuzzy Rules

A rules-based is the collection of the domain expert knowledge and it is usually expressed as a set of 'IF-THEN' rules. These rules are used to capture the relationship between inputs and outputs. The antecedent of a fuzzy rule is a logical combination of fuzzy propositions, which is usually in the form of '$x$ is $\mathcal{A}$'. The result of fuzzy rules is calculated by the degree to which the antecedent is satisfied. In general, fuzzy rules are formulated by the domain expert based on empirical knowledge, and they can be gradually improved with further use. Fuzzy system use linguistic rules in the form " IF $variable_{input}$ IS $fuzzy_{set}$ THEN $variable_{output}$ IS $fuzzy_{set}$ " to describe the relationship between the input variable and output ones.

## 2.4 Inference Method

There are three type of fuzzy inference methods are available: Mamdani, Larsen, and Takagi-Sugeno. In this paper, the Mamdani inference method is selected to calculate the fuzzy output of each decision parameter based on the sup-min composition. The general form of multidimensional multiple fuzzy reasoning models is defined as:

$$\begin{array}{ccccc} \mathcal{A}_{11}, & \mathcal{A}_{11}, & ..., & \mathcal{A}_{1n}, & \rightarrow \; \mathcal{A}_1 \\ \mathcal{A}_{21}, & \mathcal{A}_{21}, & ..., & \mathcal{A}_{2n}, & \rightarrow \; \mathcal{A}_2 \\ \vdots & \vdots & ..., & \vdots & \vdots \\ \mathcal{A}_{m1}, & \mathcal{A}_{21}, & ..., & \mathcal{A}_{mn}, & \rightarrow \; \mathcal{A}_m \\ \mathcal{A}^*_1, & \mathcal{A}^*_2, & ..., & \mathcal{A}^*_n, & \rightarrow \; \mathcal{B}^* \end{array} \tag{2.2}$$

where $\mathcal{A}_{ij}$ and $\mathcal{A}^*_j$ are the fuzzy subset of input universe of discourse $\mathcal{U}_j$; $\mathcal{A}_{ij}$ represents the $j$th input of the $i$th fuzzy rule in a fuzzy inference model; $\mathcal{A}^*_j$ represents the $j$th input of an actual antecedent; $\mathcal{B}_{ij}$ and $\mathcal{B}^*_j$ are the fuzzy subsets of output universe of discourse $\mathcal{V}$; $\mathcal{B}_i$ represents the $j$th output of the $i$th fuzzy rule; $\mathcal{B}^*_i$ represents the composite output of an actual antecedent ($i$ = 1,2,...,m; $j$ = 1,2,...,n); $m$ is the number of fuzzy rules for a fuzzy inference model; $n$ is the number of antecedent inputs of an 'IF-THEN' fuzzy rule. The inference process is written as:

$$\begin{aligned} \mathcal{A}_1(x) &= min\mathcal{A}_{11}(x_1), \mathcal{A}_{12}(x_2), ..., \mathcal{A}_{1n}(x_n) \\ \mathcal{A}_2(x) &= min\mathcal{A}_{21}(x_1), \mathcal{A}_{22}(x_2), ..., \mathcal{A}_{2n}(x_n) \\ &\vdots \\ \mathcal{A}_m(x) &= min\mathcal{A}_{m1}(x_1), \mathcal{A}_{m2}(x_2), ..., \mathcal{A}_{mn}(x_n) \\ \mathcal{A}^*(x) &= min\mathcal{A}^*_1(x_1), \mathcal{A}^*_2(x_2), ..., \mathcal{A}^*_n(x_n) \end{aligned} \tag{2.3}$$

$$\begin{aligned} \mathcal{B}^*_1(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_1(x) \wedge \mathcal{B}_1(y)] \\ \mathcal{B}^*_2(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_2(x) \wedge \mathcal{B}_2(y)] \\ &\vdots \\ \mathcal{B}^*_m(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_m(x) \wedge \mathcal{B}_m(y)] \\ \mathcal{B}^*(y) &= \mathcal{B}^*_1(y) \vee \mathcal{B}^*_2(y) \vee ... \vee \mathcal{B}^*_m(y) \end{aligned}$$

where $x_j$ is the input value ($j$ = 1, 2, ..., n) and $\mathcal{B}_i^*(y)$ is the intermediate result of each 'IF-THEN' rule ($j$ = 1, 2, ..., n). The operators $\wedge$ and $\vee$ take the minimum and maximum values of the membership functions, respectively; $\mathcal{B}^*(y)$ represents a composite fuzzy set of output decision preferences.

## 2.4 Defuzzification Method

Because the results of the fuzzy rule set are also fuzzy subsets, these subsets must be transformed into crisp values through the defuzzification process. Here, the centroid method is used because of its accuracy. The formula is given as:

$$y_{final} = \frac{\int_{\mathcal{V}} \mathcal{B}*(y)ydy}{\int_{\mathcal{V}} \mathcal{B}*(y)dy} \tag{2.4}$$

where $y_{final}$ is a final output of fuzzy inference system.

## 2.5 Reinforcement Learning

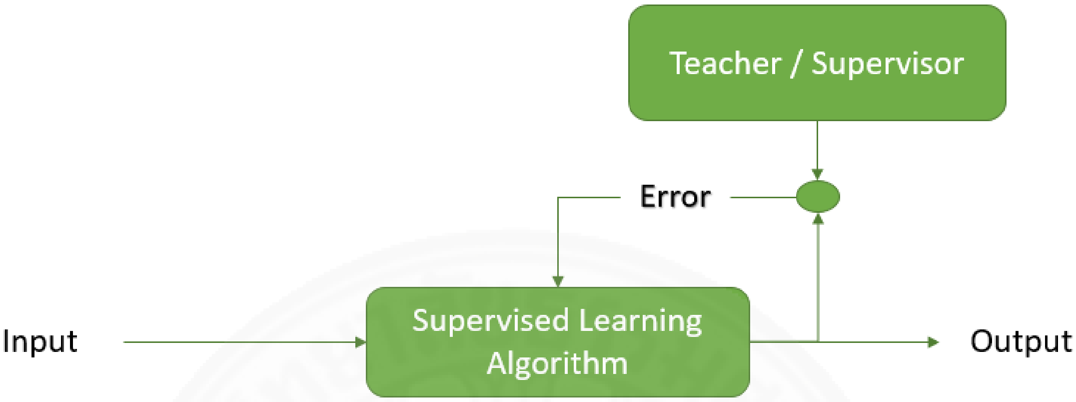
Learning is one of the abilities that substantially equivalent to understanding and adapting to the environment. These abilities should include into the human-awareness navigation. Machine learning, one of the sub-fields of artificial intelligence, is the method that enables the robot to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models. Machine learning tasks are typically classified into several broad categories: supervised learning, unsupervised learning and reinforcement learning (RL).

Supervised learning is the method that learns from the example that given by a "teacher". Its goal is to learn a general rule that maps the labeled inputs to labeled outputs. These can be seen mostly in classification problems Russell and Norvig (2009). Classification problems are the problems that ask the algorithm to predict or identify the input data as the member of a particular class or group. For example, in a training data set of animal images, that would mean each photo was pre-labeled as cat, koala or turtle. The algorithm is then evaluated by how accurately it can correctly classify new images of other koalas and turtles Kotsiantis (2007). This type of learning method is suited to the problem that input and output data sets are known and need the algorithm to sort the data.

On the other hand, sometimes the data set is not easy to label or classify, therefore, unsupervised learning is used to answer the which criteria or model that can be used to classify the data set. The well-known method in unsupervised learning is deep learning model Schuurmans and Zinkevich (2016). This model has handled a data set without explicit instructions on what to do with it. The training data set is a collection of examples without a specific desired outcome or correct answers. The neural network then attempts to automatically find structure in the data by extracting useful features and analyzing its structure. The unsupervised learning model can organize the data in different ways such as clustering, anomaly detection, association or autoencoders. Therefore, unsupervised learning is suited to the problem that labeled data is too difficult to get. Therefore, unsupervised learning rein to find patterns that can produce high-quality results.

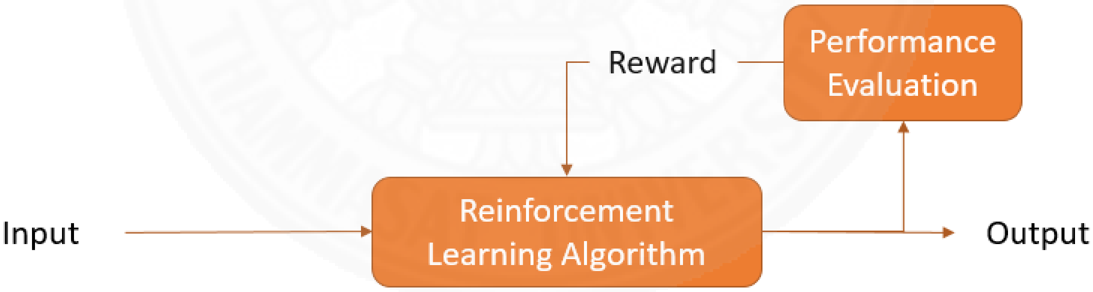
Another learning which is not in supervised or unsupervised type is the reinforcement learning (RL). RL is the process that agents or robots need to learn what to do and how to map situations and actions so that they can gain the maximum reward. The agent will try to discover by itself to choose the action that yields the reward in each situation. This action may affect not only the immediate reward but also the next situation and through that, all sub-sequence reward. RL imitate how humans and animals learn. Therefore, the machine tries a bunch of different things and is rewarded when it does something well. The structure of these three machine learning can be summarized as in the figure 2.6

In this research, robots should understand humans' information and should learn to adapt their performance according to humans' feedback. In this case, Reinforcement learning (RL), which is one of the machine learning techniques, played the role to give the robot able to learn from its experience throughout the humans' interaction. Thus, robots are able to improve their performance without explicit programming from roboticist. Before proceeding to examples of reinforcement learning



(a) Supervised Learning

(b) Unsupervised Learning



(c) Reinforcement Learning

Figure 2.6 The structure of tree type of machine learning. Supervised learning has the teacher or supervisor to help evaluate the output to the algorithm 2.6a while unsupervised learning does not need one 2.6b. Reinforcement learning is the learning algorithm that learn from the experience. It does not need the supervisor to evaluate but evaluate it self by response or feedback from the system in term of reward or punishment 2.6c.

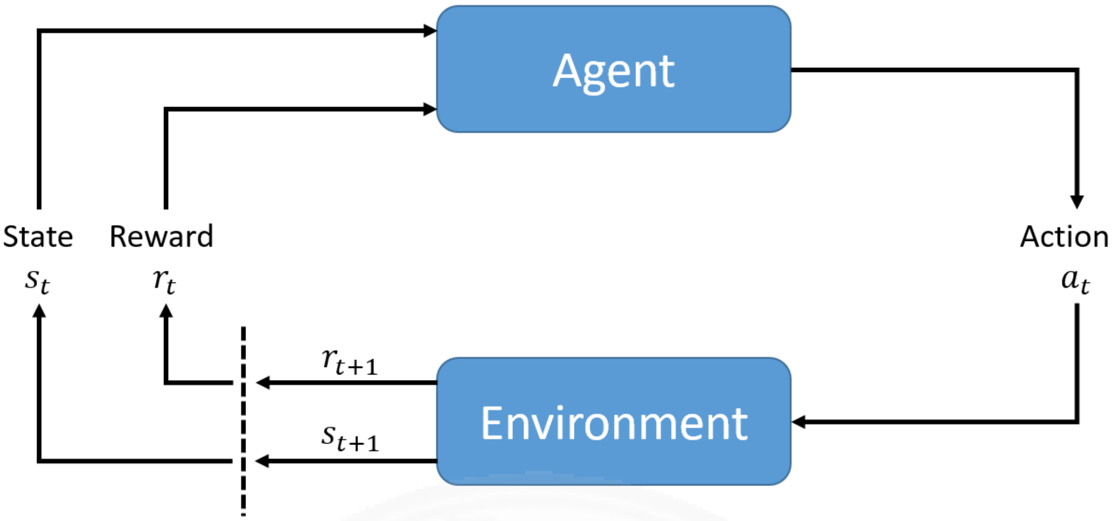


Figure 2.7 The agent-environment interaction in reinforcement learning.

research, it is vital to understand the element of reinforcement learning. To illustrate the RL concept, the process of agent-environment interaction can be shown in Figure 2.7.

The agent and the environment interact at each of a sequence of time steps. At each time step $t$, the agent receives some presentation of the environment's *state* $s_t \in S$, where $S$ is the set of possible states or situations, and on that basis selects an *action*, $a_t \in A(s_t)$, where $A(s_t)$ is the set of actions available in state $s_t$. One time step later, in part as a consequence of its action, the agent receives a numerical *reward*, $r_{t+1} \in R$, and finds itself in a new state, $s_{t+1}$.

## 2.5 Markov Decision Process

The general concept of RL can be described by the framework of Markov Decision Process (MDP). An MDP is a tuple $\langle S, A, T, R, \gamma \rangle$, of a set of states, actions, transitional probabilities, reward, and discount factor.

- $S$ is a finite set of states

- $A$ is a finite set of actions

- $T$ is a transition function that describes the probability over states.

- $R$ is a reward function representing the expected value of the next reward, given current state $s_t$, action $a_t$ and next state $s_{t+1}$.

- $\gamma$ is a discount factor that keeps the expectation finite in the case of an MDP without terminal states, where $\gamma \in [0,1]$

The idea of MDPs is that the agent chooses the action $a_t$ in the state $s_t$ and waiting for the response from the environment in the form of reward $r$ and next state $s_{t+1}$. The goal of the agent living in the environment that can be modeled as an MDP is to maximize the expected reward over time. The most common criteria are:

- Finite-horizon model: in this model, the agent tries to maximize the sum of rewards for the following $M$ step:

  $$\mathbb{E}\left\{\sum_{k=0}^{M} r_{t+k} | s_t\right\} \tag{2.5}$$

  The objective is to find the best action, considering there are only $M$ steps to collect rewards.

- Infinite-horizon discounted reward model: in this model, the agent tries to maximize the reward at the long-run but favoring short-term action:

  $$\mathbb{E}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t\right\}, \gamma \in [0,1] \tag{2.6}$$

  The discount factor $\gamma$ represent the degree of interest of the agent. a $\gamma$ close to 1 gives similar importance to short-term action and long-term action, but if $\gamma$ close to 0, it favor the short-term action.

- Average reward model: in this model, the agent tries to find the actions that maximize the average reward on the long-run:

  $$\lim_{M\to\infty} \mathbb{E}\left\{\frac{1}{M}\sum_{k=0}^{M} r_{t+k} | s_t\right\} \tag{2.7}$$

This model makes no distinction between policies which take reward in the initial phase from others that shoot for the long-run reward.

In RL, usually, the adopted model is Infinite-horizon discount reward model because of the bounds of the sum.

## 2.5 Function to Improve Behavior in RL

The agent is expected to progressive collect more rewards which in $M$ step, therefore, actively learning by reinforcement. Each state is followed by an action, which leads to another state and corresponding reward. The objective function to collect more reward or maximized the reward can be formulated as the state value or action-value function which are described as the follows:

- **State Value Function**: The state value function is defined as defined as the expected sum of future rewards following the distribution of actions given states, called policy, $\pi$. The state value can be defined as:

  $$V^{\pi}(s_t) = \mathbb{E}_{\pi}\left[G_t | s_t\right] \tag{2.8}$$

  where $G_t$ is the cumulative reward that can be written as a sum and adding the discount factor to make future reward less important and make sum finite in the continuous problem. The discount reward $G_t$ can be defined as:

  $$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{2.9}$$

  The state value function can expressed recursively with the Bellman expectation equation as:

  $$V^{\pi}(s_t) = \mathbb{E}_{\pi}\left[r_t + \gamma V^{\pi}(s_{t+1}) | s_t\right] \tag{2.10}$$

- **Action Value Function**: The action value function can be decomposed similarly as the state action value function, as shown as:

$$Q^{\pi}(s_t,a_t) = \mathbb{E}_{\pi}\left[G_t|s_t|a_t\right] \tag{2.11}$$

and to the recursive form obtained with the Bellman expectation equation can be defined as:

$$Q^{\pi}(s_t,a_t) = \mathbb{E}_{\pi}\left[r_t+\gamma Q^{\pi}(s_{t+1},a_{t+1})|s_t,a_t\right] \tag{2.12}$$

To solve the RL problem, the optimal policy that achieves a great amount of reward in the long-term should be defined. There are multiple policies to solve the same problem, some better than others but there is always at least one policy better than all the others. This is called an optimal policy denote by $\pi^*$, and it will have an optimal state value function $V^*$ which defined as:

$$V^*(s_t) = max_{\pi}V^{\pi}(s_t) \tag{2.13}$$

for all $s \in S$. An optimal policy also has an optimal action value function, denote by $Q^*$ and defined as:

$$Q^*(s_t,a_t) = max_{\pi}Q^{\pi}(s_t,a_t) \tag{2.14}$$

for all $s \in S$ and $a \in A$ This function gives the expected return for taking action $a_t$ in the state $s_t$ and thereafter following the optimal policy $\pi^*$.

## 2.5 Reinforcement Learning Related Works

Reinforcement learning is slowly gaining some stance in the agent simulation field. Its varied use in the area shows its real versatility. It can solve many learning problems and implement for learning of the high-level decision. In following works, reinforcement learning techniques are used to solve some learning problem but approaching differently. The beautiful of RL is that its framework can adapt to whatever it is we desire to model, as long as, its elements are defined for such a purpose.

Cuayahuitl et al. Cuayáhuitl et al. (2010) present an approach to include the adaptive behavior of route instructions. They proposed a two-stage approach to learn a hierarchy of wayfinding strategies using hierarchical reinforcement learning. Their experiments were based on an indoor navigation scenario for a building that is complex to navigate. Their results showed adaptation to the type of user, and the structure of the spatial environment, plus the learning speed was better than the baseline approaches they used.

Using RL to train a virtual character to move participants to a specified location was introduced by Kastanis and Slater Kastanis and Slater (2012). The states for the agent were four distances from the avatar to the participant. This states based on the *Proxemics* theory. The agent has six actions that involved working forward, backward, stay and wave. The reward function was based on the response of the human. If the person moved towards the target, the agent got a positive reward and in all other case got the punishment. The results showed that the agent did learn the rules that when the agent moves too close to the participants, they will tend to move backward.

Social Learning for the population of agent coordinate on social optimal outcome in the context of general-sum gateways proposed by Hao and Leung Hao and Leung (2013). RL is used as a learning strategy instead of evolutionary learning. The results showed that the agents were able to achieve stable coordination on socially optimal outcomes and suitable for both the settings of the symmetric and asymmetric game.

Gil et al. Martinez-Gil et al. (2012) presented a calibration method for a framework based in multi-agent RL. The agents learned to control its instantaneous velocity vector individually in scenarios with collision and frictions forces. The results indicated similarity in the learned dynamics of the agent with the real pedestrians.

The different way to use reinforcement learning can be seen in Beheshti and Sukthankar (2014). The model of the human social system was model by RL to constructing normative agent. The case study was focused on using this architecture to predict trends in smoking cessation resulting from a smoke-free campus initiative. The agents learn from their interactions with other agents, their judgment is their

**Algorithm 1** Transition-Based RRT

```
Input: the cost-space path planning problem $\boldsymbol{C}$
Parameters: the tree $\mathcal{T}$
Initialize: $\mathcal{T} \leftarrow$ initTree($node_{init}$)
while not stoppingCriteria($\mathcal{T}$) do
    $node_{rand} \leftarrow$ sampleRandomConfiguration($\boldsymbol{C}$)
    $node_{near} \leftarrow$ findNearestNeighbor($\mathcal{T}$, $node_{rand}$)
    if refinementControl($\mathcal{T}$, $node_{near}$, $node_{rand}$) then
        $node_{new} \leftarrow$ extend($node_{near}$, $node_{rand}$)
        if $node_{new} \neq$ null and transitionTest($\mathcal{T}$, cost($node_{near}$), cost($node_{new}$)) then
            addNewNode($\mathcal{T}$, $node_{new}$)
            addNewEdge($\mathcal{T}$, $node_{new}$, $node_{near}$)
        end if
    end if
end while=0
```

reward and socially correct behavior is learned.

As we can see from these works, RL has been present in many areas for some years. Its potential has already explored, but its improvement has been proved to give satisfying results. Our work presents another way of using RL to parameter adaptation of the method for estimate human social space and studying the efficacy of RL algorithms to determine the appropriate RL algorithm for the real-world task.

## 2.6 Transition Based Rapidly-Exploring Random Tree

The Transition-based RRT (T-RRT) algorithm is a sampling-based approach to a cost space path planning which has been extended from the Rapidly Random Tree (RRT). It takes advantage of two approaches. The first approach is the exploration strength of the RRT algorithm to grow random trees toward the unexplored area. The second approach is the feature to accept or reject the potential state of stochastic optimization methods. It has been successfully applied to various cost-space path planning problem in robotics Berenson et al. (2011); Iehl et al. (2012); Jaillet et al. (2010); Mainprice et al. (2011) and computation structural biology Iehl et al. (2012); Jaillet et al. (2011).

T-RRT integrates a stochastic transition test that enables it to explore the low-cost regions of the cost-space. T-RRT uses a transition test to accept or reject a new candidate configuration based on the cost variation. The pseudo of T-RRT has

**Algorithm 2** transitionTest(tree $\mathcal{T}$, cost($node_i$) $c_i$, cost($node_j$) $c_j$ )

```
Input: the maximum cost $c_{max}$
    the current temperature $T$
    the temperature adjustment factor $\alpha_T$
    the current number of consecutive rejction $nFail$
    the maximum number of consecutive rejction $nFail_{max}$
Output: true if the transition is accepted, false otherwise
if $c_j > c_{max}$ then
    return False
else if $c_j \leq c_i$ then
    return True
else if rand(0,1) < exp(- ($c_j$ - $c_i$) / $T$) then
    $T \leftarrow T/\alpha_T$
    $nFail \leftarrow 0$
    return True
else
    if $nFail > nFail_{max}$ then
        $T \leftarrow T \cdot \alpha_T$
        $nFail \leftarrow 0$
    else
        $nFail \leftarrow nFail + 1$
    end if
    return False
end if=0
```

been shown in Algorithm 1. T-RRT is extended from RRT with the addition of **transitionTest** and **refinementControl** function.

The **transitionTest** is used to accept or reject the planning path between two configurations from their cost. The level of selectivity of this function is controlled by the adaptive parameter, called temperature $T$. With low temperature, the T-RRT algorithm will limit the expansion to gentle slopes of cost space. Conversely, with high temperature, the T-RRT algorithm enable the planning path to climb steep slopes.

**Algorithm 3** refinementControl(tree $\mathcal{T}$, $node_{near}$, $node_{rand}$)

```
Input: the extesion step-size $\delta_T$
    the refinement ratio $\rho_T$
Output: true if the refinement is not too high, false otherwise
if dist($node_{near}$, $node_{rand}$) < $\delta_T$ and nbRefineNodes($\mathcal{T}$) > $\rho_T \cdot$ nbNodes($\mathcal{T}$) then
    return False
else
    return True
end if=0
```

In T-RRT, the temperature is dynamically tuned during the search process, based on the current number of consecutive rejection $nFail$ and on the maximal number of consecutive rejection $nFail_{max}$. For example, after each accepted uphill move, $T$ is decreased to avoid over-exploration on the high cost-space. More precisely, $T$ is divided by the temperature adjustment factor $alpha$, and $nFail$ is reset to 0. On the other hand, after each rejection uphill move, different actions are possible depending on the value of $nFail$. If $nFail$ is less than $nFail_{max}$, the temperature remains unchanged, but $nFail$ is incremented by 1. If $nFail$ is greater than $nFail_{max}$, $T$ is increased to facilitate exploration and avoid being trapped in local minima. More precisely, $T$ is multiplied by the temperature adjustment factor $\alpha$, and $nFail$ is reset to 0. The reansitionTest function can be seen in Algorithm 2.

The objective of the **refinementControl** function which has shown in Algorithm 3, is to limit this refinement and facilitate tree expansion toward unexplored regions of space. The idea is to reject an expansion that would lead to more refinement. If the number of refinement nodes already present in the tree is higher than a certain ratio $\rho$ of the total number of nodes, a refinement node is defined as a node whose distance to its parent is less than the expansion step-size $\delta$. Another benefit of refinement control is to reduce the computation cost of nearest-neighbor search by limit the number of nodes in the tree.

# CHAPTER 3

# HUMAN SOCIAL SPACE MODEL ESTIMATION

This chapter presents the proposed method that applies social information of each human to model individuals' social space. Many of previous work model human's social space basing on one social information like human's appearance or human's body width. However, these parameters are almost similar and assume to be the common parameters to all humans in the environment which make the robot to estimate similar social space which has the same size to everyone. This affects the performance of approaching or avoiding humans. The generated path which based on the social space cannot determine the optimal path length and optimal path cost like discomfort feeling.

Therefore, this research proposed the method that estimates the individual's social space base on their social information such as genders, perception range and relationship level between humans and the robot. The following part of this chapter moves on to describe in greater detail the estimation of the human social space model. First, the overall process of our proposed method is described to understand the whole process for human-aware navigation. Then the basic social model that uses an asymmetric Gaussian function to generate human social space with cost is introduced to use as the cost map for robot navigation. After that, human's states and social factors will be described to design the parameters for the Gaussian function. These parameters affect the size and shape of the frontal and lateral spaces of each human. Finally, the comparison results of path length that generate from the traditional human social space and our proposed method will be shown to see the effectiveness of using social factor to model human social space.

## 3.1 Overall Process of Human Aware Navigation

In human aware navigation framework, the robot has to operate in a shared environment with humans and should consider two constraints: safety, and social

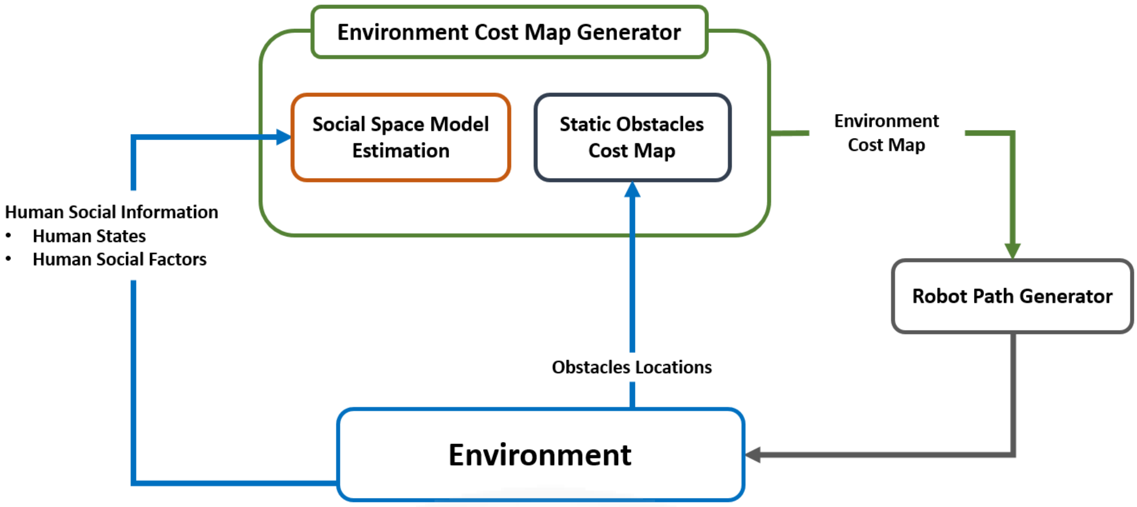


Figure 3.1 The overall process of the proposed human social space model estimation

constraint. The safety constraint is the constraint that the robot should consider to prevent itself from the harmful situation. The social constraint is the constraint that helps the robot to avoid situations that potentially make discomfort to humans. Both constraints can be defined as the cost of navigation that enables the robot to operate more natural and acceptable.

The overall view of the human aware framework can be illustrated as in Figure 3.1. Assume that the robot knows the environments such as obstacles or free space which is the static environment. From this static environment, the robot can generate the static cost-map which is the safety constraint for navigating in the environment. For the environment that humans exist, the robot considers the social constraint which navigates the robot capable of generating a socially competent path from the perceiving social information such as the human state or human social factors. The robot can generate the path based on this combination of both constraints. This generated path enable the robot to operate more natural and acceptable for humans in the environment.

However, to formalize the social constraint from the social information is the big challenge for the roboticist. Therefore, this research proposed the social space estimation method that determines the social constraints or social cost-map basing on the humans’ social factors.

## 3.2 Fixed Social Space Model

Social space model of the human is derived according to the interactive human spaces between person and person in the environment. These spaces depend on the human state and social signal, i.e., position, motion, relationship, and so on. The asymmetric Gaussian distribution function is used to integrate that information into a mathematical model and used as the cost for path planning process. The social space of human group is derived based on this function of each person.

$$F(x,y) = \sum_{i=1}^{n} f_i(x,y) \tag{3.1}$$

where $n$ is the total number of persons, $f_i$ is the repulsive force originating from the $i$-th person which can be expressed by the asymmetric Gaussian distribution function.

As shown in Figure 3.4, a person $p_i$ located at coordinate $(x_i, y_i)$, and its direction is $\theta_i$. Social area surrounding the person $p_i$ can be defined by using two dimensional Gaussian function as:

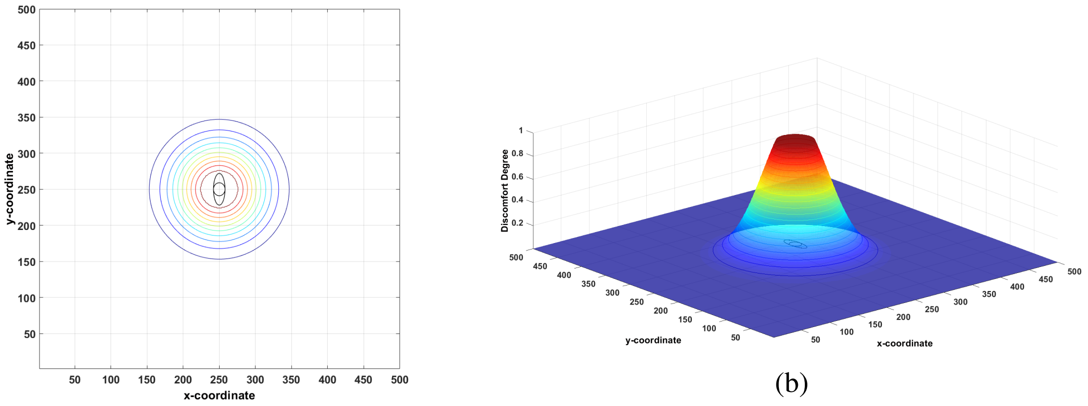
$$f_i(x,y) = A * \exp\left(-\left(\beta_{fr} - \beta_{la}\right)\right) \tag{3.2}$$

which has its maximum at the human's center $(x_i, y_i)$ and decreases when faraway from center. $A$ is the magnitude of the repulsive force which can be determined by a person's physique. $\beta_{fr}$ and $\beta_{la}$ are factor of social area in front and side of human respectively, which can be defined as:

$$\beta_{fr} = \frac{\left(d * \cos\left(\theta - \theta_i\right)\right)^2}{2 * \sigma_f^2} \tag{3.3}$$

$$\beta_{la} = \frac{\left(d * \sin\left(\theta - \theta_i\right)\right)^2}{2 * \sigma_s^2} \tag{3.4}$$

where $\theta_i$ is human direction. $\sigma_f$ and $\sigma_s$ are variances which can be modified according to human states and social factors of humans relate to the robot. Figure 3.2 shows the example social map that determined by the social model function above.



(a)

(b)

Figure 3.2 The social space model from asymmetric Gaussian function.

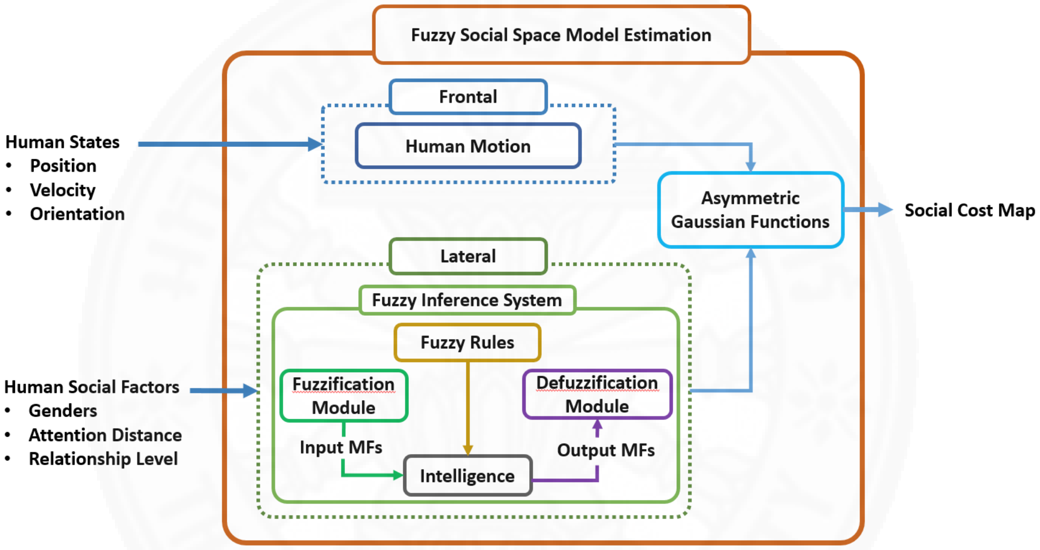


Figure 3.3 The overall process of the fuzzy social space model for human social model estimation

## 3.3 Fuzzy Social Space Model

The essential factors to develop the human social apace consist of human's states such as position and motion, and human's social characteristic. Social interaction space is the spatial space relation between humans described by Hall Edward (1969). The space around the person can be divided into four areas according to the distance to the person. These areas can be described as behavior and relationship between a person and others which depend on comfortable of the individual's sentiment. However, many of previous methods use

The structure of the proposed estimation method can be visualize as Figure 3.3.

During the human-robot interaction, the robot perceives the human states and social factors, such as positions, velocities, orientation, human-robot relationship level and human genders to design of human's social space. These human states and social factors are used to formulate the social space in frontal and lateral of the human respectively. Humans usually decide to avoid the area in the front of others depends on the speed of them. This concept can be used to define the frontal social space of humans. Social signals and cues are used to be the social factors to determine the lateral space of the humans. These social factors are the crisp set of input data which gathered for the fuzzy inference system. These crisp set are converted to a fuzzy set using fuzzy linguistic variables, fuzzy terms, and membership functions. Afterward, an inference is based on a set of fuzzy rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the output membership function, in the defuzzification module. The output from the fuzzy inference system is the parameter to determine the social space model of the human which can be calculated by the asymmetric Gaussian function. Based on this human's social model, the robot can estimate the social cost map that includes humans' social space. This social cost map will be combined with the static cost map. This combination cost map will be used in the path planning to generate the socially navigation paths to perform social interactions with humans.

## 3.3 Human States

The robot must be aware of the human position and movement in the shared work space. In reality, to avoid the collision with other persons in the environment, the human decides to move to avoid the area in the front of others depends on their speed. From this concept, the robot should avoid the frontal area of humans rely on their speed to make them more comfortable when working in the shared environment.

Let $H = (p_1, p_2, ..., p_n)$ be the number of people detected by the robot by the robot sensing and perception capacity. Human states of a person $p_i$ is represented as $p_i = (x_i, y_i, \dot{x}_i, \dot{y}_i, \theta_i)$ where the location of the human is the coordination $(x_i, y_i)$, $\dot{x}_i$ and $\dot{y}_i$ are the velocity on $x$ and $y$ direction respectively, and $\theta_i$ represent the orientation of the human reference to the world frame. This human's states can be shown in Fig.3.4.

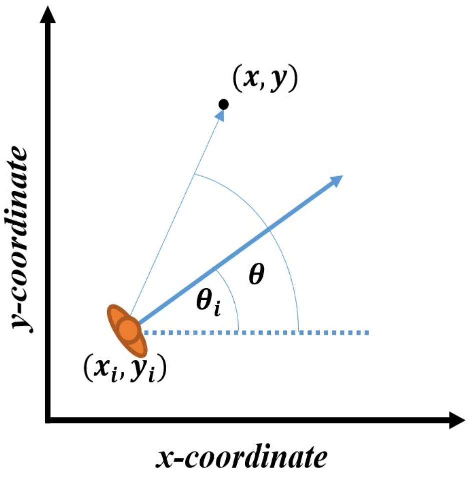


Figure 3.4 Human state which consists of position, orientation and velocity reference to the world frame

The magnitude of human motion can be computed by:

$$v_i = \sqrt{\dot{x}_i^2 + \dot{y}_i^2} \tag{3.5}$$

The value of variance ($\sigma_{fi}$) of the social space model in the frontal and rear of human can be calculated as:

$$\sigma_{fi} = \begin{cases} \sigma_{f0}, & \cos(\theta - \theta_i) \leq 0 \\ \sigma_{f0} / \left(1 + f_f v_i\right), & \text{otherwise} \end{cases} \tag{3.6}$$

where $\sigma_{f0}$ is chosen according to social space of a person. $\theta$ is the angle of the environment position $(x, y)$. $f_f$ is the normalization factor. $\cos(\theta_r - \theta_i)$ can be evaluated the area where in the front of or the rear of the person. Therefore, the social space in the front of the human will be larger than the rear.

## 3.3 Social Signals & Cue For Lateral Social Space Estimation

Human interaction and behaviors are involved, govern by not only intended goal, physical constraints, and other humans in the environment, but also social characteristics. For front and rear of the human, the social space of the human is depended on the state of the person as describe in the section above. However, for the lateral space of the human, the social space should base on the social signal and social cues between the person and others.

Social signals and cues are considering as the factors to model human lateral social space. Social cues are organized into five categories: physical appearance, gesture and posture, face and eyes, voice, and environmental space. Human social signals are defined as a communicative, informative signal or cue that, either directly or indirectly, provides information about social facts such as social interactions, a social emotion, social relationship.

In our research, we are considering the factors from human social signals and cues, such as the human's gender, the perception range of the human and the level of relationship between the human and the robot, to design the lateral space according to the Proxemics concept that introduced by Hall Edward (1969). Therefore, the reason and related work about these social factor are decribed as follows.

### Human Genders

Genders of humans are the social cue that can use to design behaviors or interactions. Various research support that humans perform difference action depending on the different genders of interacting people . Research of gender differences has been quite substantial. There are several studies demonstrate the interaction of humans pair to the interaction distances. For example, Hayduk's review counts seven studied supporting that men choose greater distances than women Hayduk (1987). In Aliakbari et al. (2011), found that female pairings preserved the smallest distance, followed by male-female pairing, and male-male pairing was most distant.

Genders of humans also influence proxemic behavior between women and men could be dominance and submission Kalbfleisch and Cody (2012). It might be that females, as well as males, maintain greater distances from males because men are labeled as more dominant, and their personal space is more respected than the personal space of women. Studies confirm that persons approaching a man stay more distant than the person who approaches a woman.

However, when humans are interacting with non-human agents such as robots and avatars in an immersive virtual reality setting, the proxemic behavior of females and males is opposite to the studies of the human-human interaction Takayama and

Pantofaru (2009). For instance, The experiment in virtual reality setting shows the result that females were less comfortable moving close to virtual agents which controlled by software, than avartar that controlled by a person, whereas males did not differentiate between both type of controllers Bailenson et al. (2003)

Therefore, in this research, genders are used to design the human social space according to the studies of human and non-human agent that mentioned above. From the above studies, females have larger social space than males because they have less comfortable to interact with non-human agents. Therefore, for the robot to approaching or navigating around humans, the robot should consider this information of humans to protect itself from interrupting human's comfort zone.

### Perception Range

The perception of the person with human's sensing organ like eyes or ears can be used to identify the locations, movement or appearance of other existences in the shared environment. The quality of human perception also depends on the distance between the person and the others Erkelens (2017). For example, when the person in the room with other two people, the person can provide the information of the nearest person more than the further person because the near distance enables the person to perceive clear information more than the far distance.

For human's emotion, humans tend to form very different mental representation on the different perception distance. As the experiment in Schiano Lomoriello et al. (2018), they implement by separate two group of participants to administer a pain decision task involving two different sizes of face stimuli painfully. These sizes manipulate the retinal size and perception distance. They observe the empathy reaction from the participants. The result shows that the empathy reaction was absent in the group of participants that receive the smaller stimuli that corresponding to the face stimuli that observed from the far distance. The experiment results show the effect of perception distance of the humans to their emotion.

The concept of perception distance can be seen in Sisbot et al. (2007). Sisbot et al. introduced a human aware navigation framework that uses the ability of humans' perception combine with safety concept to model the human's comfortable space. The concept that humans generally feel more comfortable when the robot is in their perception range, in their experiment is in the human's field of view. The results make the robot to be mostly in human's field of view during the motion.

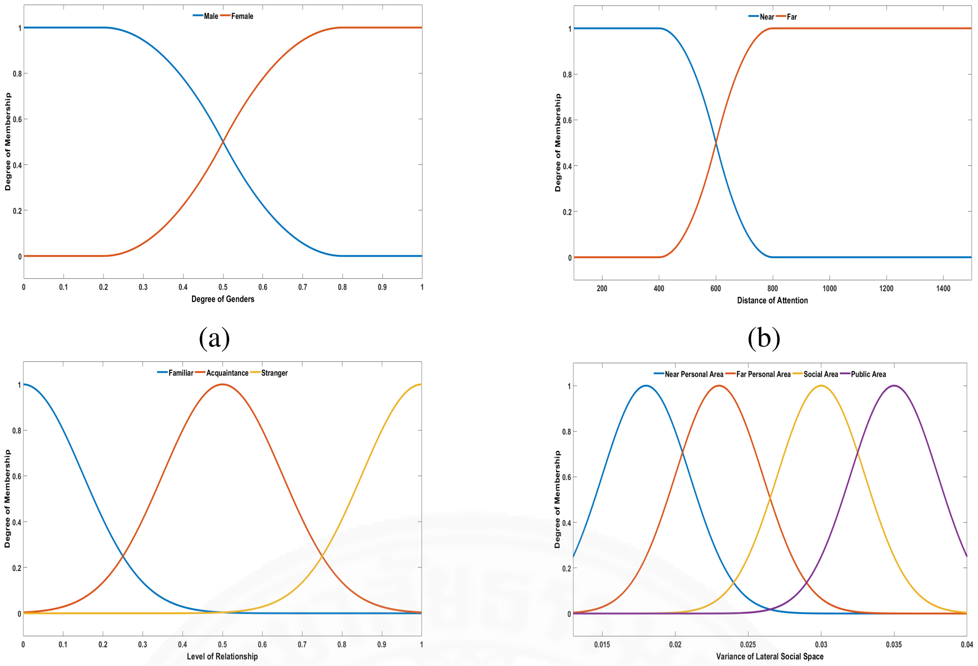
In this research, the concept of the relation between perception distance and human emotion is used to estimate the human social space. During the robot operation, if the robot moves closer to the human, the social space or the area of privacy of the human should be smaller because humans perceive more quality information of the robot that makes them more comfortable. In another word, the social space of the human will get more significantly when the robot moves far away due to less quality of information make less comfortable.

### Level of Relationships

Human behavior is mostly based on social relationships, which can be in the form of a dyadic relationship, where there is a complex peer relationship among different individuals. Edward (1969) already pointed out that how humans feel to others with a different relationship. Many studies of human-human interaction confirm that the level of relationship affects the proxemic behavior of human. For example, Hayduk (1987) confirmed that human familiar with the other person leads to smaller interpersonal distance. The signal to like other person is used to study the interaction distance between human and human Gifford (1982). Their study shows that humans use close interaction distance when they signal to like other people. They also show that humans who are familiar with other interact more proximally.

For human-robot interaction, the level of relationship can be assumed as the experience between the human and the robot. The experiment in Walters (2008) presented that humans who have prior experience with robots also tend to approach closer to it in a subsequent interaction. Takayama and Pantofaru (2009) also investigate to prove the hypothesis that experience in robotics will decrease the personal space that people maintain around the robot.

Therefore, this research takes into account of human's experience of the robot



(a) (b)

(c) (d)

Figure 3.5 Input MFs; Degree Genders(a), Perception Range(b), Level of Relationship(c). The output MFs; the variance of lateral social space(d).

to estimate human social space. The designed social model estimation enable the robot to estimate the social model of the familiar person smaller than the stranger person.

## 3.3 Mapping Social Factors to Social Space Estimation

This research incorporates a fuzzy logic system into an estimation of social interaction space framework. The fuzzy logic performs as the mapping tool to map between social factors and the variance of the lateral social interaction space. In our research, social factors, such as the gender of the human, the human perception range and the human-robot relationship level, have been considered as a linguistic variable and used as the inputs of fuzzy inference systems. The output of fuzzy logic is the value of variance for lateral social interaction space. These output sets are set according to '*Proxemics*' theory; Near personal area (NPA), Far personal area FPA, Social area (SA) and Public area (PA). According to the description of social factors and social interaction space, the fuzzy sets of inputs has been determined as follows:

- **Genders** is the genders of the human that the robot can estimate from detection

Table 3.1 Designing the social interaction space using fuzzy rules

| Input | | | Output |
|---|---|---|---|
| Gender ($\Gamma_1$) | Perception Range($\Gamma_2$) | Relationship Level.($\Gamma_3$) | Interaction Area $\mathcal{N}\left(\mu, s^2\right)$ |
| M | Near | Fam | NPA |
| M | Near | Acq | FPA |
| M | Near | Str | SA |
| M | Far | Fam | NPA |
| M | Far | Acq | FPA |
| M | Far | Str | PA |
| Fe | Near | Fam | FPA |
| Fe | Near | Acq | SA |
| Fe | Near | Str | SA |
| Fe | Far | Fam | FPA |
| Fe | Far | Acq | PA |
| Fe | Far | Str | PA |

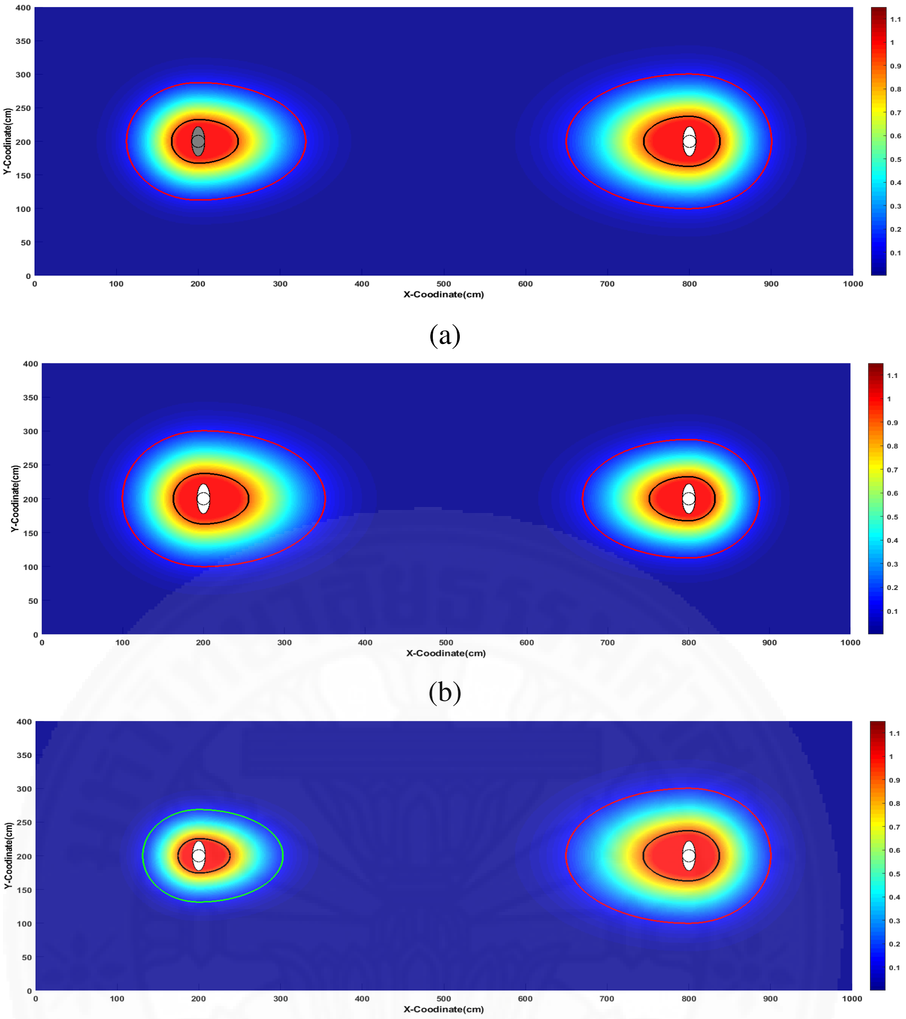
method. The input MF of gender is defined as a sigmoid function subject to male (*M*) and female (*Fe*). Let $r_g$ be the estimation value of gender input, $a_g$ the steepness of the distribution of genders, and $c_r$ the inflection point. Then the MFs of the genders is given as follows:

$$\Gamma_1(r_g; a_g, c_g) = 1/\left(1 + exp\left(-a_g * (r_g - c_g)\right)\right) \tag{3.7}$$

- **Perception Range** is the relative distance between the human and the robot which can be divided into two sets such as near (*Near*) or far (*Far*). It is represented by a sigmoid function. Let $r_d$ be the input of the relative distance, $a_d$ the steepness of the distribution of relative distance, and $c_d$ the inflection point. Then the MFs of the relative distance is given as follows:

$$\Gamma_2(r_d; a_d, c_d) = 1/(1 + exp\,(-a_d * (r_d - c_d))) \tag{3.8}$$

- **Relationship Level** describes the personal knowledge or experience with the robot which can be set by three Gaussian functions, familiar (*Fam*), acquaintance (*Acq*), and stranger (*Str*). Let $r_i$ be the relationship degree that the robot perceives



(a)

(b)

(c)

Figure 3.6 The designed human social space depends on different genders (a), perception range (b) and relationship level (c)

from people. Therefore, the relationship degree MFs are given as follows:

$$\Gamma_3\left(r_i\right)=\begin{cases}\mathcal{N}\left(\mu_{Fam},s_{Fam}^2\right) & \mathit{if\ Fam}\\ \mathcal{N}\left(\mu_{Acq},s_{Acq}^2\right) & \mathit{if\ Acq}\\ \mathcal{N}\left(\mu_{Str},s_{Str}^2\right) & \mathit{if\ Str}\end{cases} \tag{3.9}$$

For the output of the fuzzy logic, there are several ranges in the human interaction area according to the theory of Proxemics (Edward, 1969). The distance of human interpersonal space inspires us to estimate the social interaction space of the human. Therefore, the concept of different parameters in determining the different social model for each person is chosen related to these interpersonal space concept.

The fuzzy sets of outputs has been determined as follows

- **Variance of lateral space** can be separated into four sets. Therefore, four Gaussian functions are used to represent a change of variance($\sigma_{si}$) in each social interaction area which is defined as:

$$\sigma_{si} = \mathcal{N}\left(\mu, s^2\right) = \begin{cases} \mathcal{N}\left(\mu_{PA}, s_{PA}^2\right) & \textit{if PA} \\ \mathcal{N}\left(\mu_{SA}, s_{SA}^2\right) & \textit{if SA} \\ \mathcal{N}\left(\mu_{FPA}, s_{FPA}^2\right) & \textit{if FPA} \\ \mathcal{N}\left(\mu_{NPA}, s_{NPA}^2\right) & \textit{if NPA} \end{cases} \tag{3.10}$$

The input and output membership functions can be illustrated in figure 3.5. For fuzzy rules, we construct a rules base containing fuzzy rules on relationship between social factors and the lateral variance according to Proxemics theory. A detailed description of the proposed fuzzy rule is shown in Table 3.1. Combining the above-mentioned social factors, $\beta_{si}$ can be defined as follows:

$$\beta_{si} = \frac{(d * \sin(\theta - \theta_i))^2}{2 * \mathcal{N}\left(\mu, s^2\right)^2} \tag{3.11}$$

This means that, to prevent the robot from intruding onto the human social interaction space, the robot is required to delineate the dynamic boundary of interaction areas based on the human social factors.

The proposed social model of the human which using social factors can be shown in figure 3.6. For the same level of relationship, both humans are stranger (red line), and the perception range between humans and the robot are the same, male (grey body) provides smaller social space than female (white body). For the same genders of the humans, and the level of relationship between humans and the robot are the same, the robot position is set at $x = 200$ and $y = 50$. Therefore the result of the social space of near-person is larger than faraway human. For the last social factor, by the same genders and perception range, the boundary of the familiar person (green line) is smaller than the stranger person (red line)

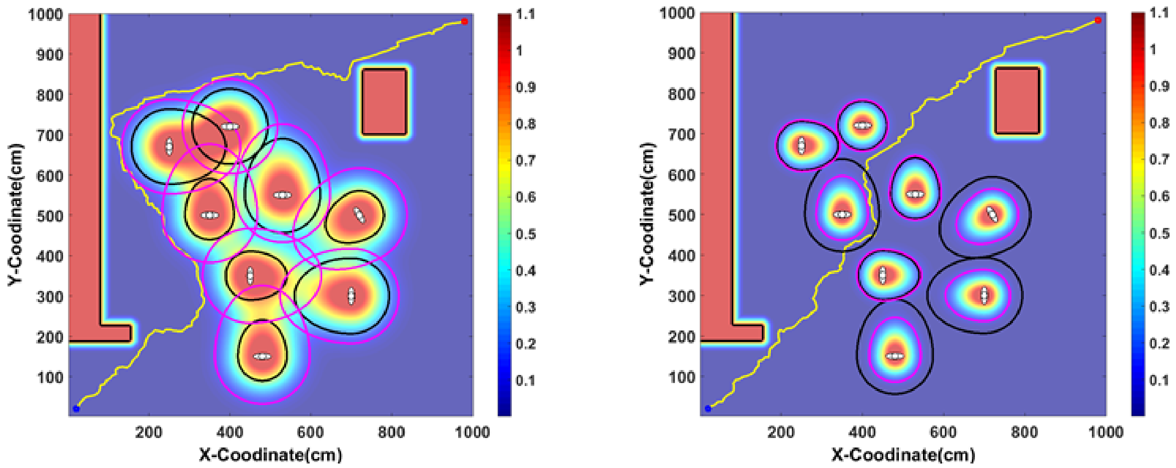


Figure 3.7 The cases of estimate social space with common parameters. The ground truth social space regions are presented by black line. The purple line represent the estimate social space with commonparmeter.

## 3.4 Effectiveness Comparison

This section shows the influence of the proposed social model to the path planning like Transition-based Rapidly Random Tree(T-RRT) which is the cost-map based path planning algorithm. T-RRT has abilities to explores and finds the optimal path cost in the large space. The scenario to simulate is that the robot tries to generate a path from the starting point at the bottom left to the ending point at the top right without colliding and intruding into humans social space. The robot will collect the path cost from human ground truth due to the position of the generated path.

The results show the impact of the proposed method compares to the common parameters method. Here the comparison results will compare the proposed method and common parameters method in different cases.

The first case is T-RRT generate path based on estimate social space that determined by the common parameters. This estimate social space is larger than ground truth social space. The second case is the estimate social space is smaller than the ground truth social space. This two cases can be visualized in Figure 3.7

In the environment, there are humans with different genders, different level of relationship. Therefore, the behavior of T-RRT based on each social factor is compared with common parameters.

The results show that the estimate social space with common parameters gives similar path length and unacceptable degree even the number of familiar person or number of males is changed. With the small estimate social space, T-RRT can provide optimal path length. However, the results of total unacceptable degree that T-RRT collect is very high. This means that the robot can generate short path but the path intrude into human's private space and make humans discomfort. However, with the large estimate social space, the collect unacceptable degree is decrease and approach to zero, but the generated path is long due to less free space. This means the generated path avoid to intrude into human personal space but it may be far from the human and difficult to interact with. The results of common parameters can be seen in Figure 3.8 and 3.9.

By using the proposed fuzzy social space model estimation, T-RRT performs adaptively to the change of environment. For example, the change of the number of the familiar person in the environment. With all stranger in environment, the estimate social space is big to avoid interrupt human's comfort. So, path length is long. However, when human get more familiar to the robot or the number of the familiar person increase. The path length is decreasing. Therefore, the fuzzy social space model can enable the robot to estimate the individual's social space that can provide more space to generate short path length while considering human's unacceptable degree, as shown in Figure 3.9 black line. The results of change of number of genders is similar, as shown as Figure 3.9.

## 3.5 Summary

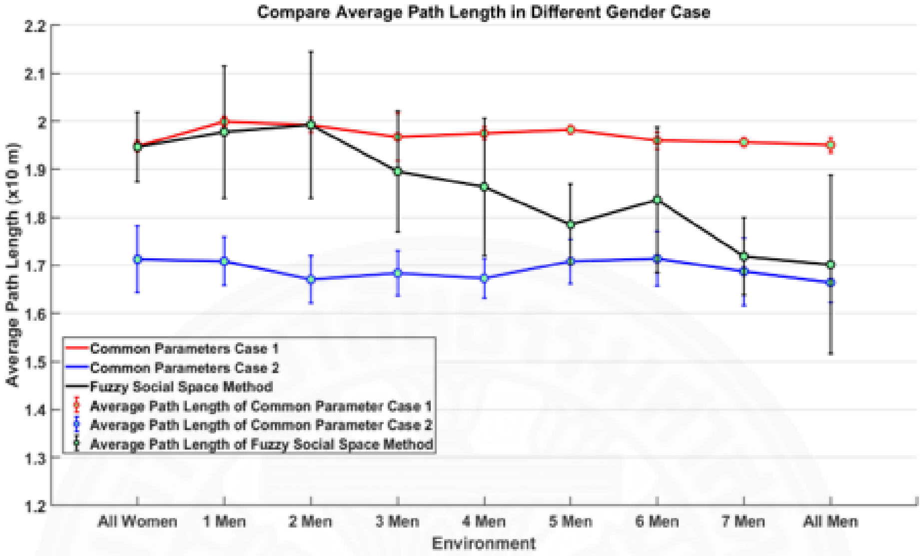
In this chapter, the designed human social space model is presented. This human model provides the size and cost that can be used to assist the navigation algorithm. The shape of the social model is not symmetric between frontal, rear and lateral. Asymmetric Gaussian function plays the role to model human social space. However, the parameters of the function depend on human factors.

The concept that human avoids to intruding into the frontal area depends on their speed. Therefore, human state, which includes position, orientation, and velocity, is used to determine the parameter for model frontal social space. The lateral space is designed based on social factors. These factors consist of social signals and cues. Social signals are defined as a communicative, informative signal or cue that, either directly or indirectly, provides information about social facts such as social emotion or social relationship. Social cues are from humans such as physical appearance, gesture, and posture. Many of previous works used only one social information and used it as the common parameters for asymmetric Gaussian function.

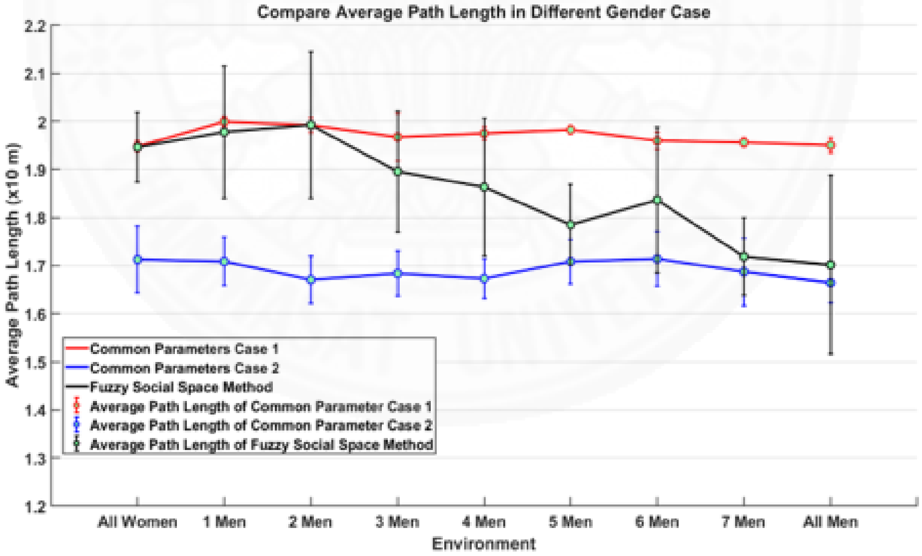
The proposed method consider three social factors of each humans to model individual's social space. However, these social factors are difficult to detect and vary depending on the variety of conditions. Therefore, a fuzzy logic approach is used to quantify these parameters.

The proposed model can be used to assist navigation algorithm like T-RRT which is cost-based navigation algorithm. The results show that our proposed model assist T-RRT to generate the optimal path length and optimal unacceptable degree of human.

Nevertheless, the designed membership functions in fuzzy inference system may not suit or satisfy to every group of humans. This problem cause the accuracy of human social model, as shown in figure 3.10. The problem can cause the robot to intrude into human privacy area or maneuver out of the interaction range. To solve this problem, the learning ability is important for the robot to learn and adjust it parameters via the the response of humans during the interaction.

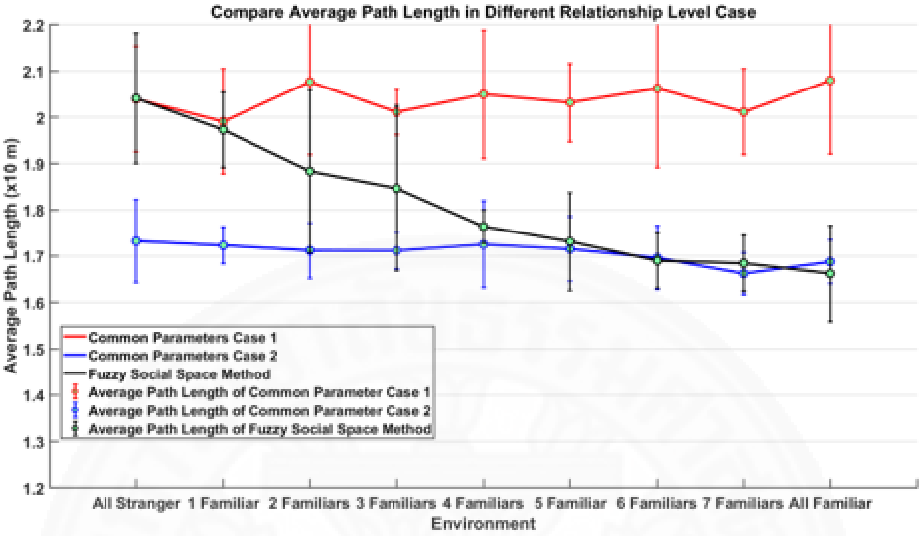


(a)

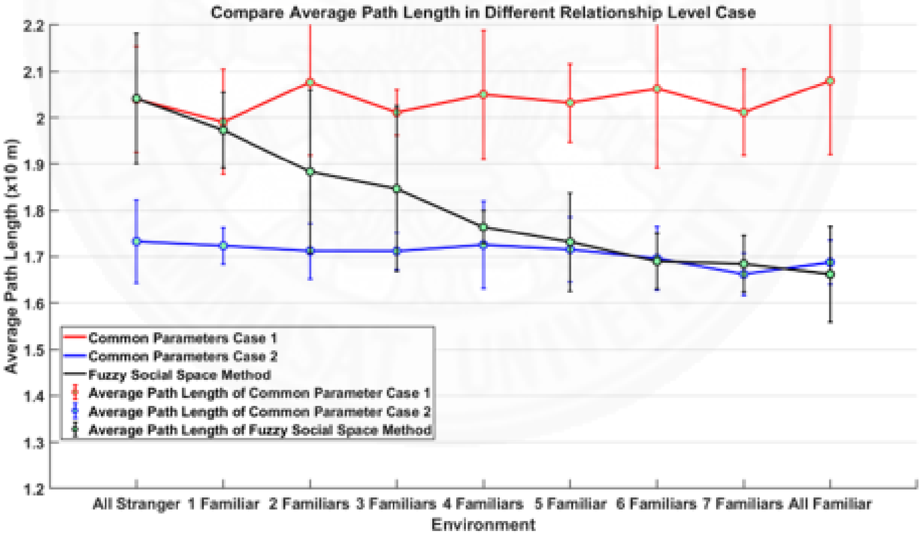


(b)

Figure 3.8 The case of the number of men changes from no men to all men in the environment



(a)



(b)

Figure 3.9 The case of the number of familiar person changes from no familiar person to all familiar person in the environment

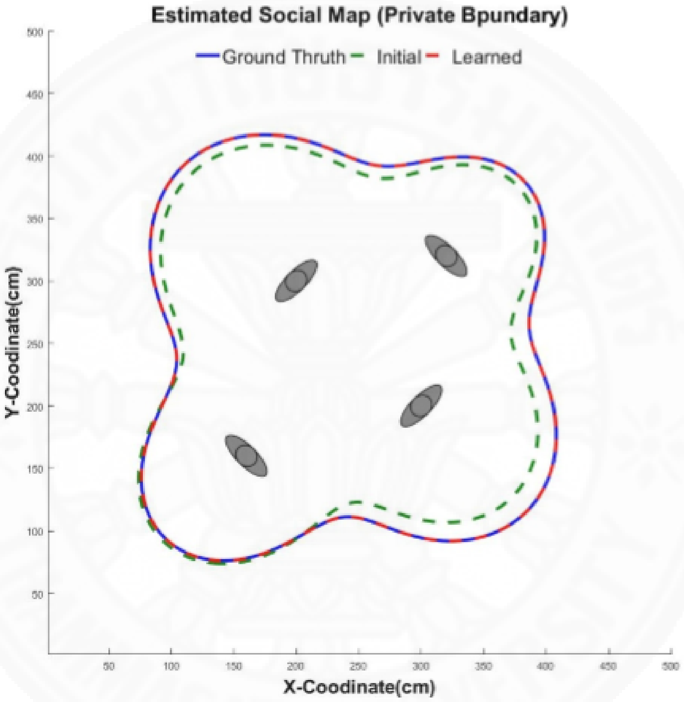


Figure 3.10 Comparison of private area boundaries. Realistic social area boundary (*blue solid line*), estimated social boundary with primary setting (*green dash line*)

# CHAPTER 4
# LEARNING FUZZY SOCIAL MODEL

From the previous chapter, the fuzzy social model uses the fuzzy inference system to maps the crisp values of detected human's social factors to the variance values of the asymmetric Gaussian function that relates to the proxemics theory. The output of the fuzzy social model is used as the cost function in the path planning algorithm. This generated path is used to guide the robot to interact with or avoid humans in a shared space. However, the design MFs, in the fuzzy inference system, are designed from the knowledge-based which may not satisfy every group of humans which have different social information. This incorrect design MFs cause the robot to behave in the way that disturbs human feeling like intruding into the privacy area or moving far away from the interaction range of humans.

Therefore, from the incorrect design mapping tool's parameter problem, the method to modify parameters should be applied to recreate social space. Carefully calibrate model parameters is one of the useful methods which enables the user to set the appropriate parameters for themselves. However, it requires the technical knowledge to modify which is difficult for the normal user. Therefore, machine learning techniques are the appropriate way to give the learning ability to the robot. The robot can learn and modify the parameters automatically. Here, reinforcement learning is chosen to apply to solve incorrect pre-design MFs automatically because it automatically learn from the experience of interacting with humans and does not require any database.

This chapter contributes the study of the reinforcement's algorithms that integrate to fuzzy inference system to adapt the parameter of MFs. By using this learning fuzzy social space, the robot is able to estimate and modify the estimate individual's social space according to an individual's response.

The comparison of RL's algorithms in term of learning time is presented to see which algorithm is suited to the problem. The results also show that the RL algorithms can modify the MFs and convert the estimate social model to similar to the real human's social model. This learning fuzzy social model allows the robot to gain the maximum reward which means it avoids to intruding the private area of human and in the interaction quality range.

## 4.1 Fuzzy Social Model Estimation To Reinforcement Learning

Our research aims to modify the MFs that used to estimate the human social model, more accurately by learning while interacting with humans in a shared environment. This learning fuzzy social model enables the robot to correctly estimate human social space that help the robot to avoid intruding into human's area of privacy and also keep distance for good quality of interaction. This research applied reinforcement learning as the tools to modify the MFs.

With the fuzzy social model estimation, the robot can estimate the human social space by estimating from the pre-designed MFs. The robot can generate the path to approach or avoid the human based on this social model. However, with pre-designed MFs, the robot may estimate human social space smaller or larger than realistic which effect to the human's feeling and emotion due to the movement of the robot. The robot receives this humans' response detection method like the verbal/non-verbal reaction or some detection method like face detection or emotion detection method **?**. This response information can be used as the reward or punishment for RL algorithms. The RL algorithms will modify the MFs by increasing or decreasing MFs value until the robot receive the maximum reward from humans. Figure 4.1 shows the overall process of learning the fuzzy social model. To express this fuzzy social model estimation into reinforcement learning framework, the element of reinforcement learning such as states, actions and reward should be defined.

## 4.1 States-Space

This work focus on modifying the pre-design MFs, especially the level of relationship MFs, to maximize the reward that given by humans. Three Gaussian functions are used to design the level of relationship MFs. The importance values of

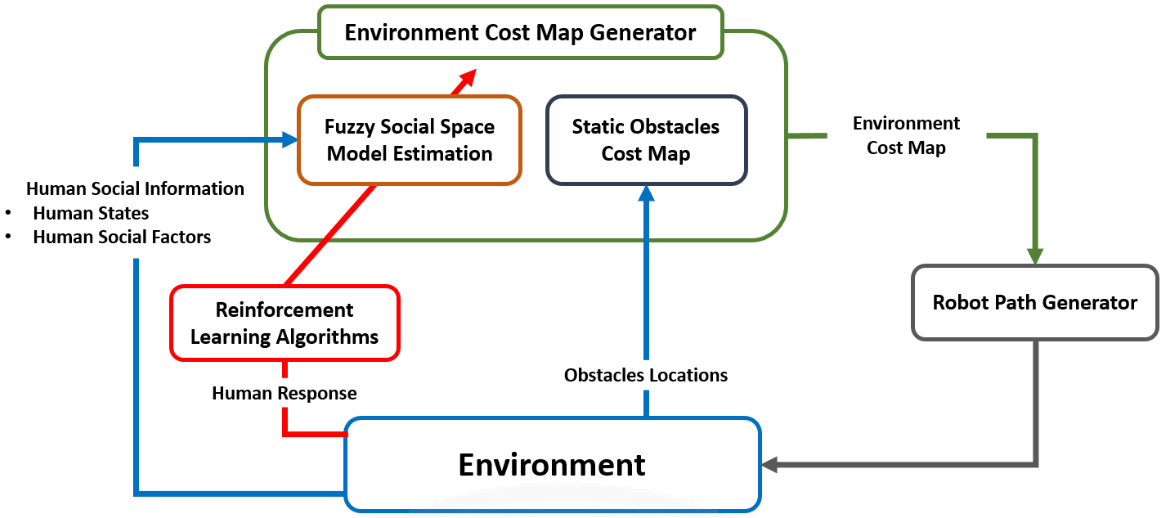


Figure 4.1 The overall process of the proposed human social space model estimation

the Gaussian function are mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$. Therefore, states $s \in S$ for our problem which consists of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ can be defined as:

$$s = \left[\mu, \sigma^2\right] \tag{4.1}$$

where $\boldsymbol{\mu} = [\mu_{Fam}, \mu_{Acq}, \mu_{Str}]$ and $\boldsymbol{\sigma^2} = [\sigma^2_{Fam}, \sigma^2_{Acq}, \sigma^2_{Str}]$. However, we make the states more simple by make $\boldsymbol{\sigma^2}$ as the constant and modify only $\boldsymbol{\mu}$. Therefore, the state of our problem can be modified as:

$$s = \left[\mu_{Fam}, \mu_{Acq}, \mu_{Str}\right] \tag{4.2}$$

## 4.1 Action-Space

The actions $a \in A$ is the set of how to adjust MFs. In this work, means of MFs are modified by increasing, decreasing or staying at the same value. Therefore, the actions $\boldsymbol{a}$ can be defined as:

$$a = \left[a_{Fam}, a_{Acq}, a_{Str}\right] \tag{4.3}$$

where $a_{Fam}$, $a_{Acq}$ and $a_{Str}$ can be set as decreasing or increasing or do noting to the mean values $\boldsymbol{\sigma^2}$.

## 4.1 Reward Function

The reward function $R(s_t, a_t, s_{t+1})$ is defined as the response from environment where performing the action $a_t$ in the st ate $s_t$ that lead to the state $s_{t+1}$. In this work after the robot first estimate the social space of the human, the robot will move to approach each human by using the estimate social space. However, because of the error of design, the estimated social space may smaller or larger than actual human's social space. Therefore, the reward is the response or feedback from humans. This response can be collected from the human's feeling which can be correctly detected by manually evaluation or emotion recognition process such as facial movement, heart rate, blood pressure, etc. However, in this thesis, we assume that the human's feeling depends on the distance from human centers.

Here, the reward is designed from two different areas that used to evaluate our social model estimation. First, the quality interaction area where humans can be engaged in high-quality interaction with the robot. At this area, interaction degree ($ID$) is designed to be the evaluation function that describes the easiness of interaction. Second, the private area where humans do not want to interfere with the robot speech or action. This area provides the degree of discomfort feeling or an unacceptable degree ($UD$) of humans that affected from the robot behavior. The $ID$ and $UD$ are increasing and decreasing respectively depending on the distance between the robot and the humans. Another factor in designing the reward is the difference of estimate path length and optical path length $\triangle l$. This factor related to the generated path that the robot generate. If the difference of path length is too large, that means robot generate path far from the human, and if the difference of path length is too small, that means robot generate path close to the humans. This different path length prevents the robot from generating path far away from humans.

The ratio $ID$ and $UD$ is used to determine the reward. Reinforcement learning techniques try to modify parameters of mapping tool in fuzzy social space model to minimize interaction degree and minimize unacceptable degree and different path length. Therefore, the estimate social space will let the robot to interact or approach human in the area of interaction quality area but out side the human's area of privacy.

The work aim to maximize the $ID$ and minimize the $UD$ and different path length $\triangle l$ while the robot is operating. Therefore, the designed reward function $R(s_t, a_t, s_{t+1})$ at state $s_t$ for human-robot interaction can be defined as:

$$R(s_t, a_t, s_{t+1}) = \frac{k_1 * ID(s_t)}{k_2 * UD(s_t) + k_3 * \triangle l(s_t)} \tag{4.4}$$

where $k_1$, $k_2$ and $k_3$ are weight for each degree. This reward correspond to the $ID$, $UD$ and $\triangle l$ at state $s_t$. To this end, the aim of our work is to determine the state that give the maximum reward:

$$s = \arg\max_s R \tag{4.5}$$

## 4.2 Reinforcement Learning Algorithms

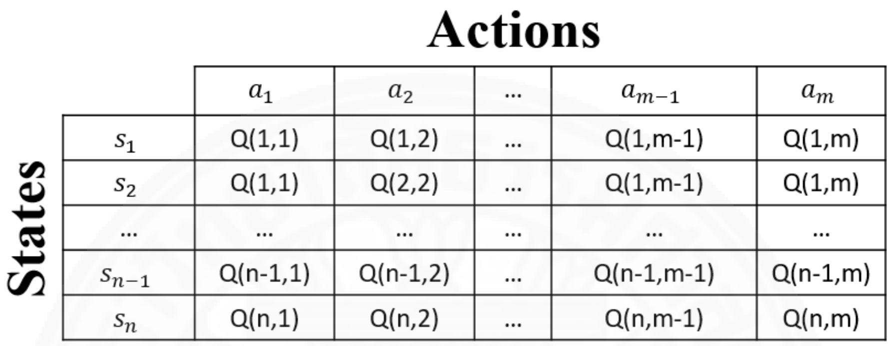
In this section, RL's algorithms that used to integrate with our fuzzy social model are described. RL algorithms are applied to determine how to adjust the MFs. This section will start with the famous and basic RL algorithm, Q-learning which is the model-free based reinforcement learning. Then the average reward learning (R-learning) is described. This R-learning is similar to Q-learning but uses the average reward instead of the discount factor. The third is the policy gradient method, Actor-Critic, which use the loss function to adjust the learning parameter. Finally, the advanced RL method that combined with the deep learning technique and experience replay buffer. This method will use the buffer memory to store the previous leaning parameters and use it to update learning parameters.

## 4.2 Q-Learning

Q-learning is a popular and widely used technique of RL which learns to optimize the state-action value (Q-value). The state-action value represents the quality of actions in the successor state and makes the task of choosing an actor easier. This means that the robot will evaluate the quality of the actions in the state. When the robot comes back to the same state, the robot will select the appropriate action based

**Algorithm 4** Q-Learning

```
Initialize: state-action value $Q(s,a)$
    Loop Process
    Choose action $a$ in $s$ using behavior policy (e.g. $\epsilon$-greedy)
    Take action $a$, observe $R$, next state $s'$
    $\delta \leftarrow R + max_{a'}Q(s',a') - Q(s,a)$
    $Q(s,a) \leftarrow Q(s,a) + \alpha\delta$
    $s \leftarrow s'$
    =0
```



**Actions**

**States**

| | $a_1$ | $a_2$ | ... | $a_{m-1}$ | $a_m$ |
|---|---|---|---|---|---|
| $s_1$ | Q(1,1) | Q(1,2) | ... | Q(1,m-1) | Q(1,m) |
| $s_2$ | Q(1,1) | Q(2,2) | ... | Q(1,m-1) | Q(1,m) |
| ... | ... | ... | ... | ... | ... |
| $s_{n-1}$ | Q(n-1,1) | Q(n-1,2) | ... | Q(n-1,m-1) | Q(n-1,m) |
| $s_n$ | Q(n,1) | Q(n,2) | ... | Q(n,m-1) | Q(n,m) |

Figure 4.2 Q-Table of $n$ states and $m$ actions

on this quality value. The Q-learning pseudo code is shown as Algorithm 4

In this thesis, the tabular Q-learning is used to determine the action to modify MFs. The tabular Q-learning is using each cell of $n$ states and $m$ actions table to store the value of quality of each pair state-action. The initial values in each cell of Q-table usually set as zero. This table is called Q-table which depicted by Figure 4.2.

By following the Algorithm 4, the beginning of the state $s_t$ is set as the location of the pre-designed MFs means $\boldsymbol{\mu}$. The action $a_t$ in this state $s_t$ is selected from the values in Q-table by $\epsilon$-greedy method which is the method to make the agent choose the random action with the probability $\epsilon$ and choose that action that has maximum Q-value for otherwise. After the action $a_t$ is taken, the robot will observe the reward $r_t$ from $R$ function which corresponding to $ID$, $UD$ and $\triangle l$ when MFs as the state $s_t$, and observe the next state $s_{t+1}$.

The Q-learning method aims to compute the optimal value $Q^*(s_t,a_t)$ for all state-action pairs, which can be done iteratively by collecting samples. Q-table is updated for every sample using the following update rules:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \left[ r_t + \gamma \max_{a'} Q_t(s_t, a') - Q_t(s_t, a_t) \right] \tag{4.6}$$

where $\alpha$ is the learning rate that allow for determination of the significance of new information in relation to the existing value. The motivation of this update equation is to calculate the difference between the predicted Q-value $r_t + \gamma \max_{a'} Q_t(s_t, a')$ and its current value $Q_t(s_t, a_t)$ for every sample.

Once a certain amount of iterations is reached, the selected action process will select the action that gives the maximum reward. It means that after a certain amount of time, most of the quality value of state-action pairs are determined. Therefore, in the state, there is the action that can gain more quality to reach the state that has a maximum reward.

## 4.2 R-Learning

The R-learning algorithm, proposed by Schwartz , is the adaptation of Q-learning to the average criterion. The average criterion mostly applies to the continuous problem. In a continuous setting, there are no terminal state or some task also no start state, and no special time step which mean cannot define the task as the episode and the interaction between agent and environment goes on and on and forever without termination state. Therefore, this R-learning, which is the average reward RL algorithm, is chosen as an interesting algorithm to solve our parameter adaptation problem. The goal of R-learning is to learn an action $a$ whose average reward $\rho$ is as close as possible to the maximal average reward $\rho^*$.

In this thesis, the tabular R-learning is also used to determine the action to modify MFs. R-learning algorithm is similar to Q-learning but use and adaptive shifting values to approximate the optimal average reward to avoid the unboundedness of Q-value in state action pairs, as shown in Algorithm 5

By following the Algorithm 5, the beginning of the state $s_t$ is set as the location of the pre-designed MFs means $\boldsymbol{\mu}$. The the action is selected from the values in

**Algorithm 5** R-Learning

```
Initialize: average reward $\rho$ and state-action value $Q(s,a)$;
LOOP Process
Choose action $a$ in $s_t$ using behavior policy (e.g. $\epsilon$-greedy)
Take action $a$, observe $r_t$, next state $s_{t+1}$
$\delta \leftarrow r_t$ - $\rho$ + $max_{a'} Q(s_{t=1}, a') - Q(s_t, a_t)$
$Q(s_t, a_t) \leftarrow Q(s_t, a_t)$ + $\alpha\delta$
if $Q(s_t, a_t) = max_{b'} Q(s_{t+1}, b')$ then
    $\rho \leftarrow \rho + \beta\delta$
end if
$s \leftarrow s'$
=0
```

Q-table by $\epsilon$-greedy method. After the action $a$ is taken, the robot will observe the reward $r_t$ from $R$ function and observe the next state $s_{t+1}$.

The R-learning method also aims to compute the optimal Q value $Q^*(s_t, a_t)$ for all state-action pairs, which can be done iteratively by collecting samples. However, instead of using discount factor for discounting the weight of the pass reward, R-learning using the average reward to update the Q value in Q-table. Therefore, Q value is updated for every sample using the following update rules:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\delta \tag{4.7}$$

$$\delta = r_t - \rho + \max_{a'} Q_t(s_t, a') - Q_t(s_t, a_t) \tag{4.8}$$

where $\rho$ is the shifting value that approximately to the averga reward. This $\rho$ will update every time that the action selection algorithm use greedy stratergy to select the action. The update of this shifting value can be determine as:

$$\rho = \rho + \beta\delta \tag{4.9}$$

where $\beta$ is the update rate of the shifting value.

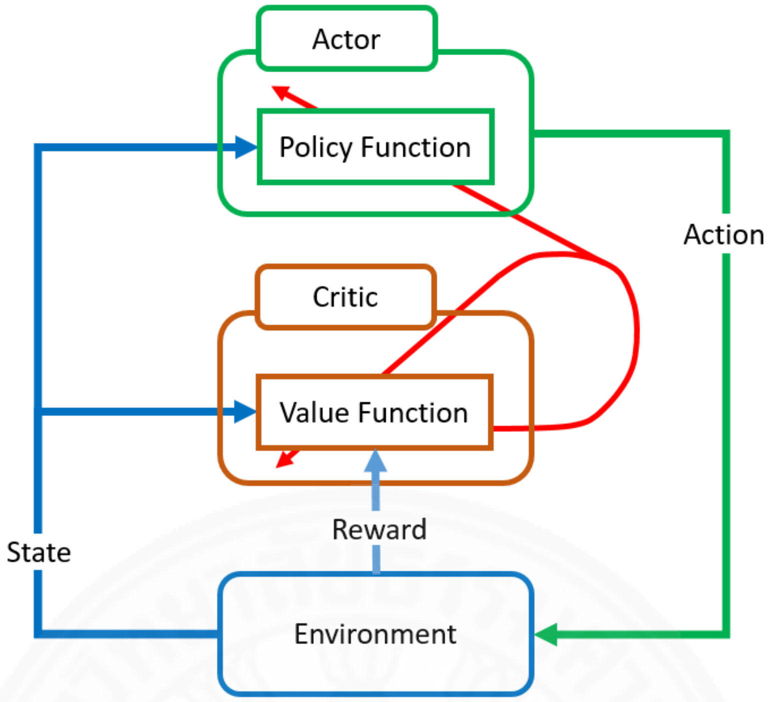


Figure 4.3 The actor-critic architecture

## 4.2 Actor-Critic

Actor-critic methods are RL methods that have a separate memory structure to represent the policy independent of the value function explicitly. The policy structure is known as the actor, because it is used to select actions, and the estimated value function is known as the critic because it criticizes the actions made by the actor. Learning is always on-policy: the critic must learn about and critique whatever policy is currently being followed by the actor. The critique takes the form of an error. This scalar signal is the sole output of the critic and drives all learning in both actor and critic, as suggested by Figure 4.3.

Actor-critic methods are the natural extension of the idea of reinforcement comparison methods to reinforcement learning problem. Typically, the critic is a state-value function. After each action selection, the critic, which is a state-value function, evaluates the new state to determine whether things have gone better or worse than expected. That evaluation is the temporal-difference (TD) error:

$$\delta_t = r_t + \gamma \hat{v}(s_{t+1}) - \hat{v}(s_t) \tag{4.10}$$

where $\hat{v}$ is the value function implemented by the critic. This TD error can be used to evaluate the action just selected, the action $a_t$ taken in state $s_t$. If the TD error is positive, it suggests that the tendency to select $a_t$ should be strengthened for the future, whereas if the TD error is negative, it suggests the tendency should be weakened.

This thesis use two neural networks as the function approximation for both actor and critic. According to algorithm 6, the action $a_t$ are selected from the policy function which generate from the neural network $\pi(a_t|s_t,\boldsymbol{\theta})$. This indicating the tendency to select each action $a_t$ when in each state $s_t$ with respect to the weight $\boldsymbol{\theta}$. Then, the agent observe the reward $r_t$ and the next state $s_{t+1}$ by taking the action $a_t$. After that, the state values $\hat{v}(s_t,\boldsymbol{w})$ and $\hat{v}(s_{t+1},\boldsymbol{w})$ of state $s_t$ and $s_{t+1}$ are defined by critic process with neural network weight $\boldsymbol{w}$. These state values are used to determine the TD-error in equation 4.10. Because AC have two neural network that must be trained, it means that two set of weights ($\boldsymbol{\theta}$ for the actor and $\boldsymbol{w}$ for critic) must be optimized separately by using back propagation process of neural network:

$$\boldsymbol{w} = \boldsymbol{w} + \alpha_{\boldsymbol{w}} \delta \bigtriangledown_{\boldsymbol{w}} \hat{v}(s_t,\boldsymbol{w}) \tag{4.11}$$

$$\boldsymbol{\theta} = \boldsymbol{\theta} + \alpha_{\boldsymbol{\theta}} \delta \bigtriangledown_{\boldsymbol{\theta}} \ln \pi(a_t|s_t,\boldsymbol{\theta}) \tag{4.12}$$

where $\alpha_{\boldsymbol{\theta}}$ and $\alpha_{\boldsymbol{w}}$ are the learning rate. The process will be repeat until the action from the actor provide optimal TD-error from critic.

## 4.2 Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) refers to the use of Deep Neural Networks (DNN) as function approximators for value functions or policy in a RL frame work. The advantage of DRL is that it can handle high dimensional of state-action space which is suitable for our parameters adaptation problem.

Q-Learning (Section 4.2.1) was initially considered that Q-learning was unstable when using with neural networks and most satisfied with small state-action

**Algorithm 6** Actor-Critic

```
Input: a differentiable policy parameterization $\pi\,(a|s,\boldsymbol{\theta})$
       a differentiable state-value parameterization $\hat{v}(s,\boldsymbol{w})$
Parameters: step size $\alpha_{\boldsymbol{\theta}} > 0$, $\alpha_{\boldsymbol{w}}$
Initialize: policy parameter $\boldsymbol{\theta}$ and state-value weights $\boldsymbol{w}$
  Loop Process:
  Choose action $a_t$ from the function $\pi\,(a|s,\boldsymbol{\theta})$ using behavior policy (e.g. $\epsilon$-greedy)
  Take action $a_t$, observe $r_t$, next state $s_{t+1}$
  $\delta \leftarrow r_t + \gamma\ \hat{v}(s_{t+1},\boldsymbol{w})$ - $\hat{v}(s_t,\boldsymbol{w})$
  $\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha_{\boldsymbol{w}}\delta\,\nabla_{\boldsymbol{w}}\,\hat{v}(s_t,\boldsymbol{w})$
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha_{\boldsymbol{\theta}}\delta\,\nabla_{\boldsymbol{\theta}}\ln\pi(a_t|s_t,\boldsymbol{\theta})$
  $s_t \leftarrow s_{t+1}$
  =0
```

**Algorithm 7** Deep Q-Learning (DQN)

```
Initialize: replay memory $D$
       state-action value function with random weight $Q(s,a,\theta)$
  Loop Process:
  Choose action $a_t$ from function $Q(s,a)$ using behavior policy (e.g. $\epsilon$-greedy)
  Take action $a_t$, observe $r_t$, next state $s_{t+1}$
  Store transition $(s_t,a_t,r_t,s_{t+1})$ in replay buffer $D$
  Sample random transitions $(s_j,a_j,r_j,s_{j+1})$ from $D$
  Calculate target for each transition $y_i$
  if $s_{j+1}$ is terminal then
      $y_j = r_j$
  else
      $y_j = r_j + \gamma \max_{a'} Q(s_{j+1},a',\theta)$
  end if
  Train the Q-network on $(y_j - Q(s_j,a_j,\theta))^2$ using Equation (4.15)
  $s_t \leftarrow s_{t+1}$ =0
```

space application in term of learning time. In Mnih et al. (2015), it was shown that Q-learning could be used with DNN and perform as the human level on seven Atari 2600 games using only raw image pixels as the input.

In Deep Q-learning (DQN), the Q-function is parameterized with a neural network with weighs $\theta$. The network is trained by minimizing a loss function at every iteration $i$, given by

$$L_i(\theta_i^Q) = \mathbb{E}_{s\sim\rho_\pi(\cdot),a\sim\pi(\cdot)}(y_i - Q(s,a,\theta)) \tag{4.13}$$

where the target at iteration $y_i$ can be defined as:

$$y_i = \mathbb{E}_{s_{t+1}\sim\varepsilon}\left[r_t + \gamma \max_{a'} Q(s_{t+1}, a', \theta) | s, a\right] \tag{4.14}$$

$\pi(s|a)$ is the behaviour policy, $\rho_p i(\cdot)$ is the distribution of states under policy $\pi(s|a)$ and $\epsilon$ refers to the environment.

To minimize the loss function, the gradient of the loss function is computed with respect to the weights and is given by:

$$\triangledown_{\theta_i^Q} L_i(\theta_i^Q) = \mathbb{E}_{s\sim\rho_\pi(\cdot), a\sim\pi(\cdot), s'\sim\varepsilon}\left[\left(r + \gamma \max_{a'} Q(s', a', \theta_{i-1}^Q) - Q(s, a, \theta_{i-1}^Q)\right) \triangledown_{\theta_i^Q} Q(s, a, \theta_{i-1}^Q)\right] \tag{4.15}$$

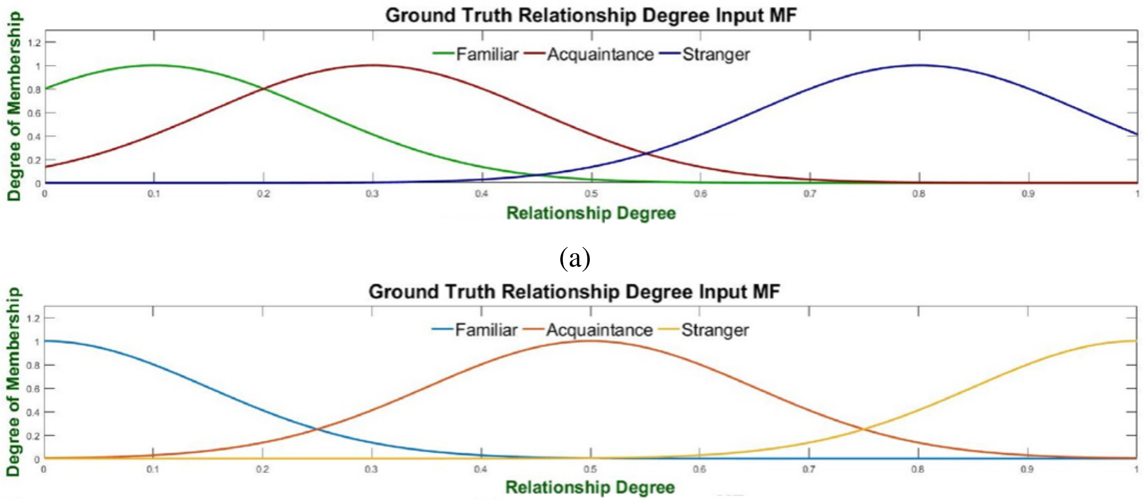
The loss function is minimized using stochastic gradient decent. The behaviour policy is an $\epsilon$-greedy policy to ensure sufficient exploration.

The key aspect which makes DQN work is the use of *experience replay*. In general, the sampled trajectories from the environment are temporally correlated and if these trajectories are used to train the network it would lead to over fitting in the network and the network would not be able to learn effectively.

In order to break the temporal correlation of data points while training, the agent's experience at each time step, $(s_t, a_t, r_t, s_{t+1})$, is saved in a replay buffer. At every instant, a fixed number of samples are extracted from the replay buffer at random and used to train the network. This makes the network to learn more efficiently. The complete algorithm of DQN is given at Algorithm 7.

## 4.3 Simulation and Results

This section shows the results of our proposed strategy to correct the human social space estimation method with a machine learning method. Here, learning fuzzy social space model is introduced to estimate human social space. By integrating the reinforcement learning with fuzzy social space model, the robot has the ability to learn and modify the pre-design MFs to match to humans in a specific group of humans in the shared environment. Therefore, the proposed model assists the robot in planning



(a)

(b)

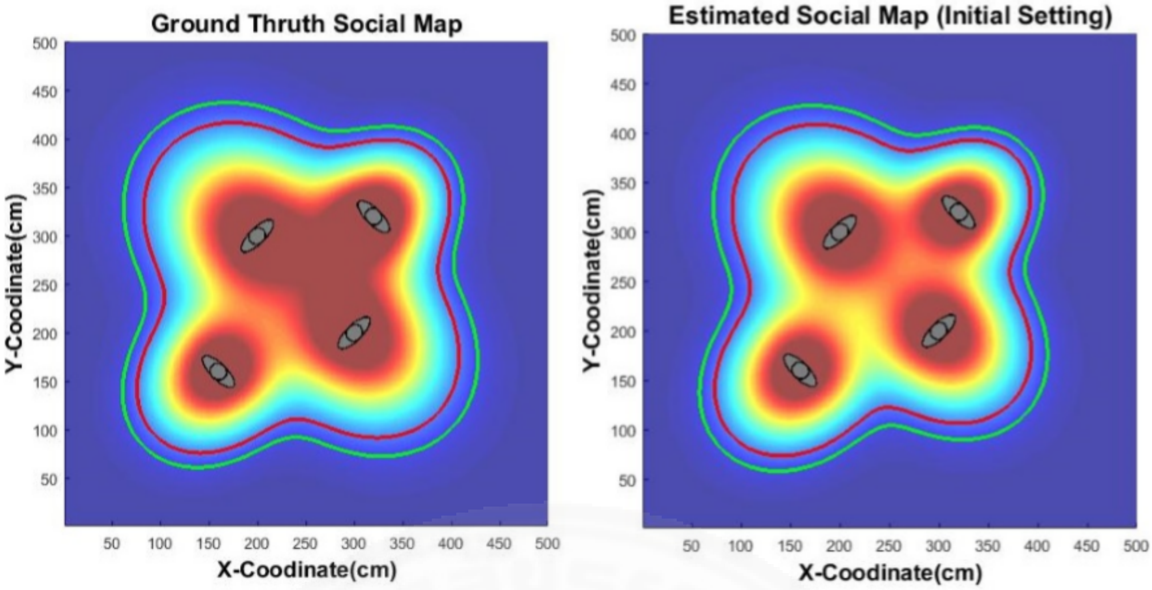
Figure 4.4 The different of ground truth relationship level MFs (4.4a) and initial setting of relationship MFs (4.4b)

paths to visit every person in the environment without trespassing on their private area, but to keep the distance from which people are able to have high-quality interactions.

## 4.3 Simulation Setup

At the beginning of this chapter, RL's algorithms are described and used as the learner for parameters adaptation problem. This section compares the efficacy of each algorithm to parameters adaptation problem in three properties. First is convergent of the algorithm that compares which algorithms can convert the social map error to a minimum. This property means that the estimated social model possible to modify and get similar to the realistic with RL algorithms. Second is the learning time which presents how fast that RL's algorithms use to modify the MFs' parameters until the social map error converts to the minimum. The third is the exploration rate of the state-action space which is the value to compare the algorithms how much state-action space has been explored. This value relates to the optimal action or best action that learning algorithm selected at each state.

The results of learning fuzzy social model that integrate RL to fuzzy social space estimation are also presented with different conditions, such as the face direction of humans or the number of humans in the environment, to show that our proposed method assist the robot to maximize interaction degree and minimize unacceptable



(a) (b)

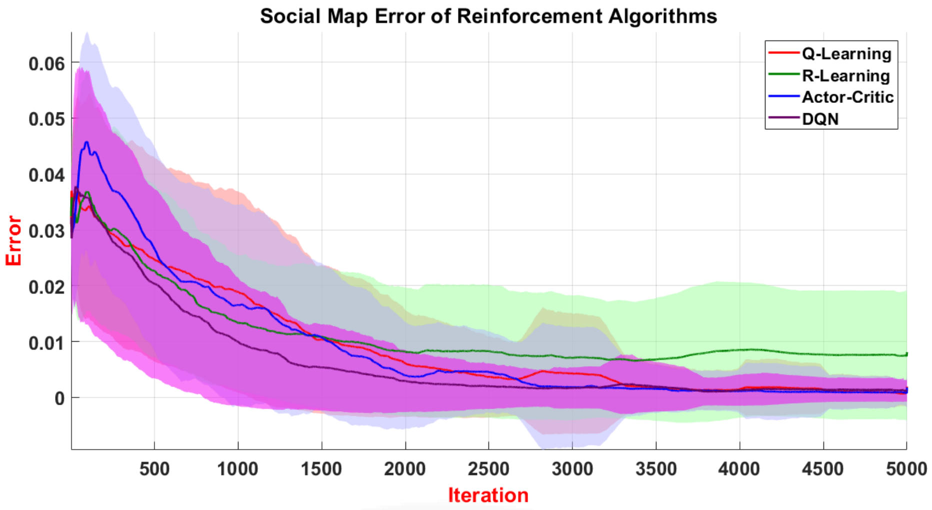
Figure 4.5 The different of ground truth social map (4.5a) and initial estimate social map from our proposed fuzzy social space (4.5b)

degree in various conditions.

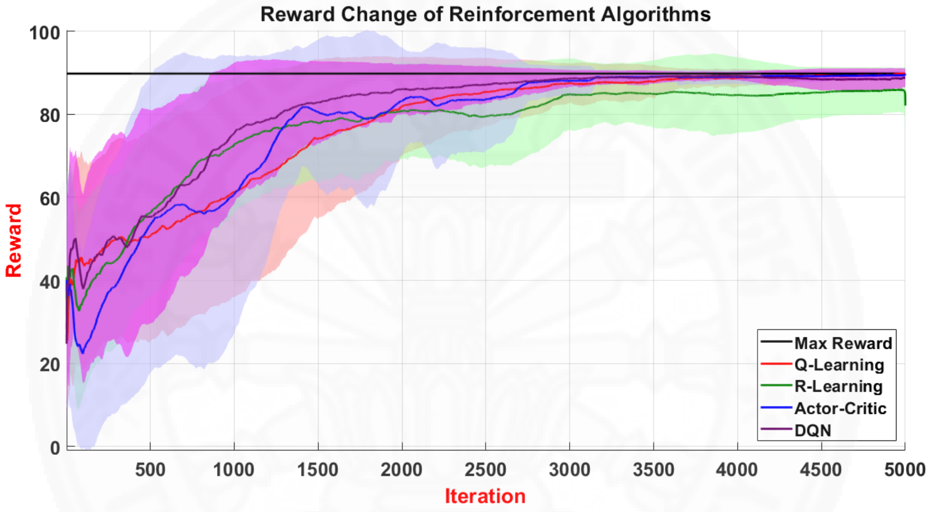
In the simulation, we assume that a geometric map is given or created by the robot. Our proposed model is to generate the social map by computing and updating social cost assigned to the geometric map. This social map is used to plan the robot navigation path in the environment. To validate the proposed model, we need to receive the reward from people. Therefore, the fuzzy social model is used with different parameters to model the ground truth social map which is used as the response or reward from humans for our proposed learning fuzzy social space model.

The parameters to model the ground truth social map are set as follows: The parameters for relationship level MFs, according to equation (3.9), $s_{Fam}$ = 0.15, $\mu_{Fam}$ = 0.1 to *Fam* set, $s_{Acq}$ = 0.15, $\mu_{Acq}$ = 0.3 to *Acq* and $s_{Str}$ = 0.15, $\mu_{Str}$ = 0.8 to *Str* set.

For our estimation method, the initial parameters of the relationship level MFs in equation (3.9) are designed as follows: $s_{Fam}$ = 0.15, $\mu_{Fam}$ = 0 to *Fam* set, $s_{Acq}$ = 0.15, $\mu_{Acq}$ = 0.5 to *Acq* set, and $s_{Str}$ = 0.15, $\mu_{Str}$ = 1 to *Str*. These parameters can be adjusted by the learning process. Likewise the perception range MFs parameters, according to equation (3.7), are designed as follows: $a_{Near}$ = -0.35, $c_{Near}$ = 300 to



(a)



(b)

Figure 4.6 The error of social map that compared between the estimated and ground truth social map (4.6a) and the reward change of each algorithm (4.6b)

*Near* set and $a_{Far} = 0.35$, $c_{Far} = 300$ to *Far* set. The relationship level MFs of ground truth and initial setting can be seen in Figure 4.4.

For the output function, the social interaction area is split into four Gaussian sets. The parameters of equation (3.10) are as follows: $\mu_{PA} = 0.035$, $s_{PA} = 0.005$, $\mu_{SA} = 0.045$, $s_{SA} = 0.005$, $\mu_{FPA} = 0.0035$, $s_{FPA} = 0.06$, $\mu_{NPA} = 0.0035$, $s_{NPA} = 0.065$. These parameters are decided based on the human interaction area concept Tomari et al. (2012) which determined the range of an individual's interpersonal space with different social factors when the robot approached the person. Reflecting their results, the parameters for the output membership functions can be determined. The different characteristic of ground truth social map and an initial estimated social map of males with different relationship levels to the robot can be seen in Figure 4.5. The figure shows that with initial setting, our fuzzy social space estimate incorrect social space of each person which effect to incorrect social map (smaller than ground truth). If the robot uses this estimated social map to planning the path, it will cause human discomfort feeling. Therefore, RL is used to modify the MFs until the estimate social map is similar to the ground truth.

## 4.3 Efficacy of RL's Algorithms

For the reinforcement learning process, the discrete states $s$ in equation 4.2 which consist of three mean values of each relationship MF are set as the value between 0 to 1, increasing by 0.1. The total number of states is $10^3$. The action set of each MFs (equation 4.3) is defined as stay, increase, decrease, *i.e.*, 0, +0.1, -0.1. The total number of the action set is $3^3$. Therefore, the total state-action space is equal to $10^3 x 3^3 = 27000$ state-action spaces. The goal to use the RLs is to modify the MFs through iterative learning processes until gaining a maximum reward signal.

The RL's algorithms that are chosen to compare in this thesis are Q-learning(Q), R-learning(R), Actor-Critic(AC) and Deep Q-Learning(DQN). The results of social error and reward value of each algorithms through the simulation are shown in Figure 4.6.To evaluate the efficacy of RL's algorithms, learning time, converge value, and exploration rate are used as the criterions.

The first criterion, the learning time is used determine the time that the algorithm requires to reach and remain within design percentage of change. In this simulation result, learning time is determined from start to the time that error of social space or reward oscillate with in five percent of minimum value. The results show that DQN is reqiure less learning time which compare to others algorithm. DQN used less learning time 17.6% compare to R-learning, 23% compare to Q-Learning, and 28.7% compare to Actor-Critc.

For the second criterion, the converge value is used to determine the final value of social map error that compare between estimated and ground truth social map,

| **RL ALgorithms** | **Learning Time (Iteration)** | **Converge Value** | | **Exploration Rate (%)** |
|---|---|---|---|---|
| | | **Social Map Error (10^{-3})** | **Reward Error** | |
| **Q-Learning (Q)** | **2390** | **2.2723** | **4.486** | **26.96** |
| **R-Learning (R)** | **2112** | **7.6772** | **8.160** | **28.09** |
| **Actor-Critic(AC)** | **2581** | **1.3968** | **4.497** | **43.05** |
| **Deep Q-Network (DQN)** | **1840** | **1.7872** | **4.755** | **29.56** |

Table 4.1 Summary Efficacy of Reinforcement Learning Algorithms

Table 4.2 Results of learning social model with the number of people

| **No. of Humans** | **Average Interaction Degree** | | **Average Unacceptable Degree** | |
|---|---|---|---|---|
| | **Initial Parameters** | **After Learning** | **Initial Parameters** | **After Learning** |
| **2** | 5.6804 | 5.7923 | 0.9742 | 0.2528 |
| **3** | 5.5694 | 7.6095 | 1.8829 | 0.2321 |
| **4** | 5.261 | 6.6644 | 3.4072 | 0.9995 |

and final value of reward. The converge value of social map error and reward can be determine at the end of learning time. The results present that Actor-Critic which is policy-gradient method converge to smallest value of social map error and approach to gain highest reward. It converge less than DQN, Q-Learning and R-Learning at 21.8%, 38.5%, 80.8% respectively.

For the last criterion, the exploration rate can be determined by the percentage of the explored state-action pair ($N_{exp}$) to total number of state-action pair ($N_{Total}$), as follows:

$$ExplorationRate = \frac{N_{exp}}{N_{Total}} x100 \qquad (4.16)$$

where $N_{Total}$ is equal to 27000 state-action pairs. The results shows that Actor-Critic explore 48% of all state-action space which mean the algorithm is possible to determine the best solution or action in each state. Actor-Critic has overcome Q-Learning, R-Learning, DQN by 59.7%, 53.2% and 45.6% respectively. The results of efficacy of reinforcement's algorithms can be summarized in Table 4.1.

Table 4.3 Results of learning social model with people facing different directions

| **Facing Direction of 4 Humans** | **Average Interaction Degree** | | **Average Unacceptable Degree** | |
|---|---|---|---|---|
| | **Initial Parameters** | **After Learning** | **Initial Parameters** | **After Learning** |
| **Into the center of group** | 5.5787 | 7.1985 | 1.7458 | 0.2129 |
| **Out of the center of group** | 5.5695 | 7.6095 | 1.8829 | 0.2321 |

## 4.3 Learning Fuzzy Social Space Model with Different Conditions

In this paper, we define the quality interaction area and the private area. Fig. 4.7 shows the interaction degree with three, four, and five subjects, respectively. The results show that our proposed method increases the interaction degree of subjects during their interaction with the robot until it suits everyone. Fig. 4.8 shows the results of the unacceptable degree. The results show that our method can reduce the unacceptable degree of subjects until they feel comfortable to interact with the robot. These results show that our proposed model outperforms the fixed-parameter for estimated the privacy area and more clearly with the number of humans in the environment. The results can be summarized in Table 4.2. We also perform the simulation with four subjects facing different directions. The results are consistent with the previous results obtained from the simulations with different numbers of subjects. The results show that our proposed method increases the quality interaction degree and reduces the unacceptable degree of the subjects, as shown in Table 4.3.

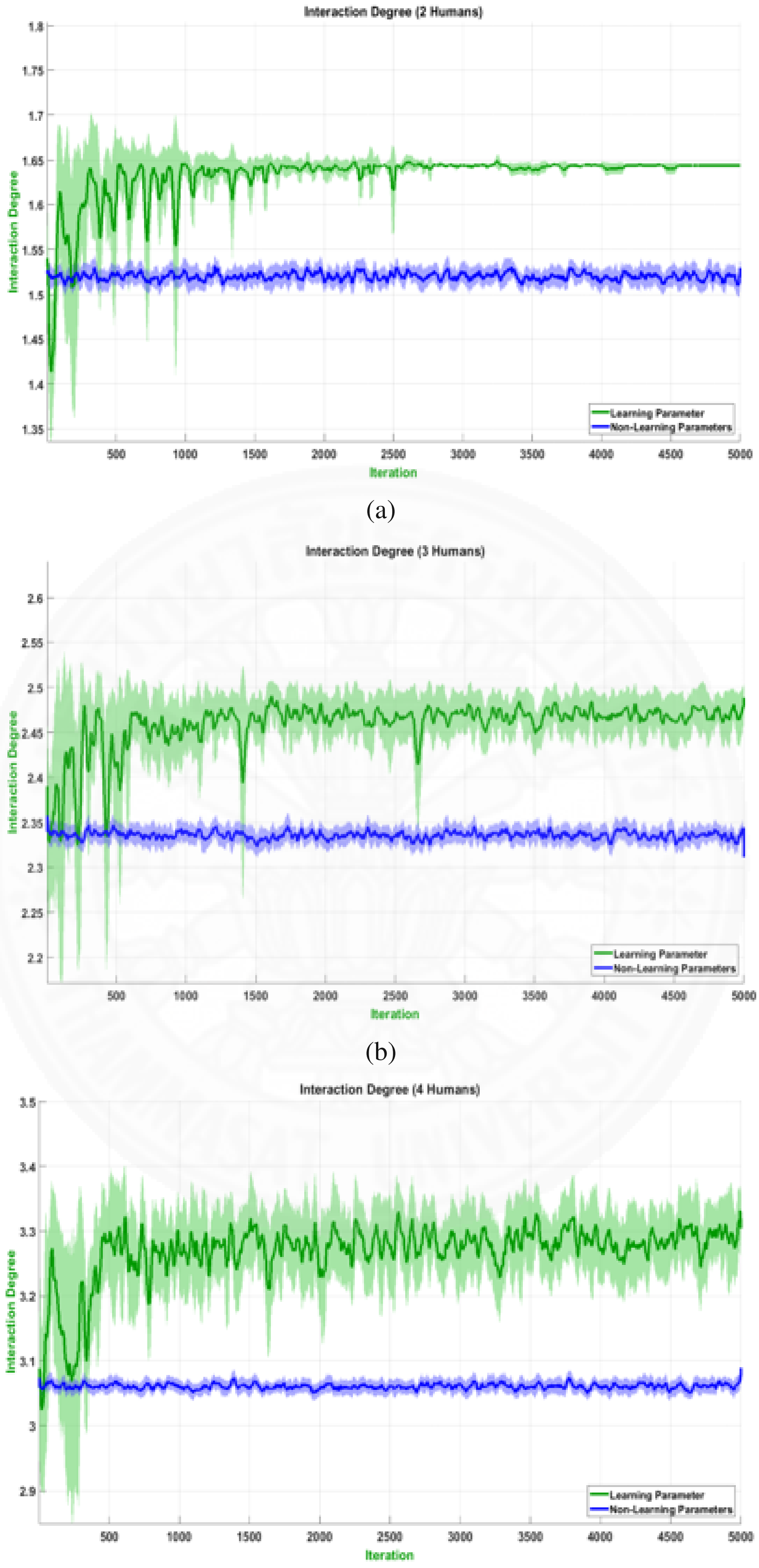
## 4.4 Summary

This section began with the problem of fuzzy social model estimation method that use to estimate human social space for a navigation task. The problem of the fuzzy social model method is that pre-designed MFs parameters cause the robot to estimate incorrect social space for humans in the environment. This problem makes the robot to perform in unacceptable ways such as intruding into human privacy area or out of interaction range. These improper performances cause humans' feeling and make them refuse to operate with the robot.

This chapter applied reinforcement learning, which is one of machine learning, to solve the problem. By giving the ability to learn from humans' responses, the robot is able to modify the parameters that estimate human social spaces suit to humans in the environment. However, there are lots of reinforcement learning algorithms; therefore, this chapter selected some algorithms to test and determine which algorithm is suited to our problem.

To find the efficacy of reinforcement learning algorithms which apply to the parameters adaptation. The convergent, learning time and exploration rate are used to evaluate algorithms. The convergent describes the possibility of algorithms to solve the parameters adaptation. The learning time determines the speed of the algorithm to solve the problems. The exploration rate is related to describe that the algorithm possible to find the best action to modify parameters in each situation.

The results show that most reinforcement algorithms can solve the problem and lead the robot to estimate perfect human social space. Most of them convert the estimate social map to approach to the ground truth or realistic social map, except R-learning that approach with small error. Deep Q-Network has the best performance to converge the estimate social map to approach to the ground truth or realistic social map faster than other due to its experience replay memory that can reuse its past experience to find the appropriate action in each state or situation. However, its exploration rate is less than Actor-Critic which means that DQN can perform to approach the optimal value but the action that is selected might not be the best action in the state.

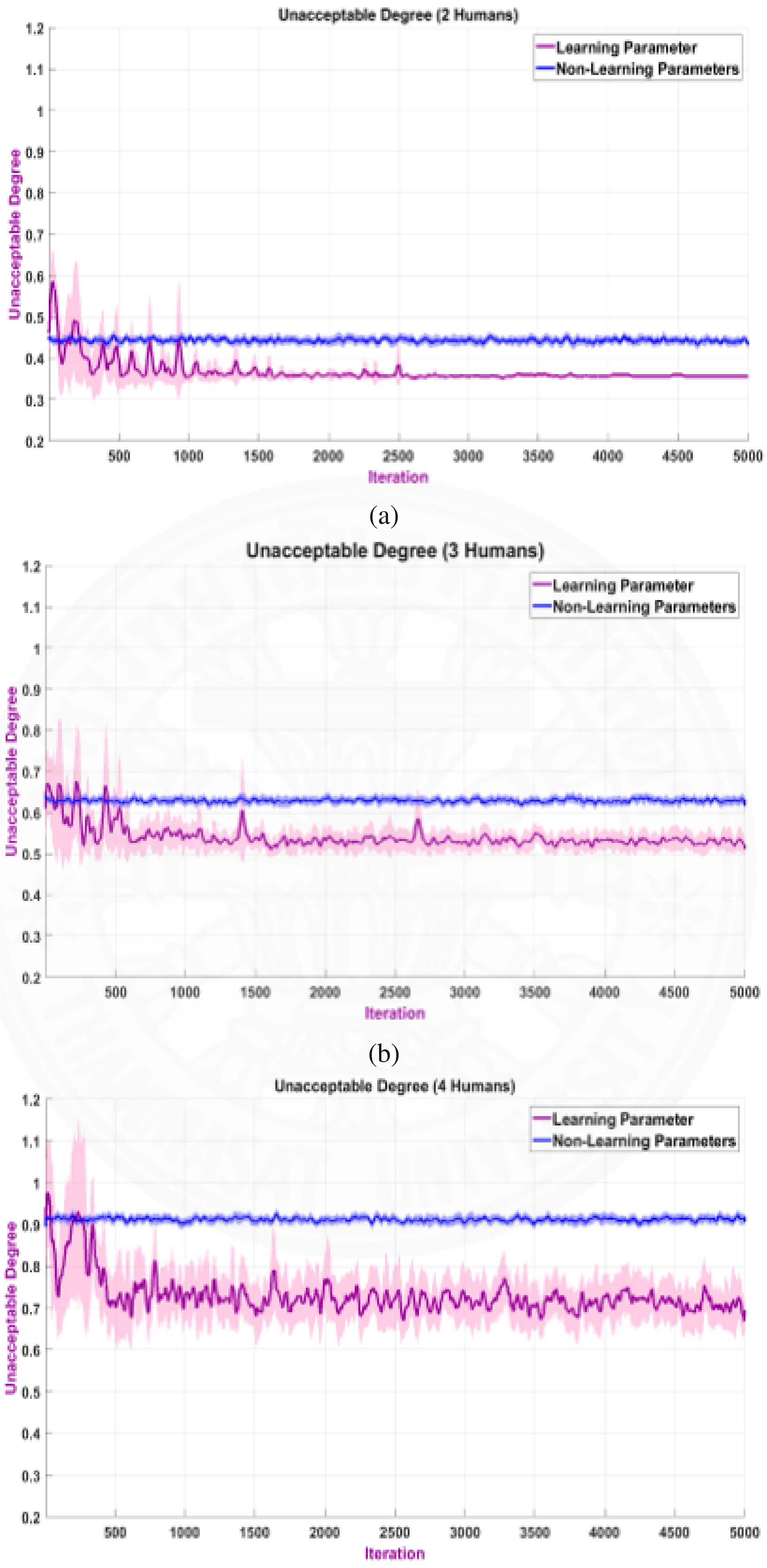


(a)

(b)

(c)

Figure 4.7 Interaction degree ($ID$) represent the acceptable or quality of interaction that the robot can receive from people along generated path. High interaction degree means that the robot approaches close enough to have interactions with humans.



(a)

(b)

(c)

Figure 4.8 Unacceptable Degree ($UD$) presents the total discomfort feeling that the robot receives from humans along the generated path. The robot should plan the path without entering the human private area.

# CHAPTER 5
# CONCLUSION AND FUTURE WORK

## 5.1 Conclusion

The mobile robots are widely used in many application, for example, health care service, customer service or event in the household. Most of the task requires robots to move and provide services in different locations. Therefore, the robot navigation is important and widely developed to enable them to operate safely. The safe navigation is the critical requirement for robot navigation which enables robots to move around the environment without harm to the surrounding environment. However, even when the robots move safely, sometimes human feel that the motion of robots is not safe. The reason for unsafe feeling is the unfamiliar or lack of trusting in the technology. Therefore, social competence which is the rules of society that humans use to act with other humans is applied to the robot navigation task. This social competence makes robots behave more naturally and acceptable for humans to feel safe and comfort while robots operate around in the shared environment.

Human-aware navigation is the challenge for human-robot symbiosis that considers both safe and socially navigation. The research of human-aware navigation can be separated into three different approaches. First is the naturalness which is the development of low-level behavior of the robot. The research in this approach strives to imitate human motion as the target behavior and recreate the robot's behavior. Second is sociability which applied high-level social conventions. The research in this approach is how to transfer the social convention into the robot. The last approach is the comfort which considers to human's feeling. The research in this area considers the motion of robots which is not only safe but also move in the way that makes human more relax.

Human-aware navigation considers both safe and social constraint. However, to formalize both constraints into the mathermatic model is still a challenge. The Proxemics theory which is a social science and psychological theory is mostly used to formalize to mathematic model. This theory describes how human use of space to different humans in the environment. Therefore, to use this theory to human-aware navigation. The researcher has to model it into mathematical. Two popular methods to model the private area or personal area according to the Proxemic theory are a geometric and cost-based method. In this thesis, the cost-based method is used to model the human's personal space.

An asymmetric Gaussian function is a famous function in the cost-based method. It provides the degree to different locations in the environment which can be used as the cost for the robot navigation. The variance parameters are important parameters to model the shape and size of human's social space. They can be determined by single social information or many social information. However, most of the research use them as the common parameters to every human which effect to path planning algorithm which cannot be optimal in term of path length and path cost (discomfort feeling).

This thesis contributes two approaches. First is the human's social space estimation that estimates an individual's social space based on an individual's social information. The second is explored the ability of reinforcement learning for adapt fuzzy inference system parameters that are used to estimate the human social model.

First, Social space estimation that use fuzzy inference system,called Fuzzy Social Space Model is proposed. It uses fuzzy inference system to map the social signals and cues, like genders, perception rage and relationship level of the human to the robot, to variance value that related to human interaction space in 'Proxemics' theory which introduced by Hall Edward (1969). These variances are used in an Asymmetric Gaussian function to determine the cost surrounding humans. This cost is used to assist the cost-based path planning algorithm that leads the robot to perform according to human's social information. However, in the fuzzy process, membership functions are design based on the knowledge of existed research which may not suit every group of humans. Therefore, learning ability is important to re-design the membership functions to suit with the specific group. This leads us to the study of reinforcement learning, which is one of machine learning tools, to learn and adjust the parameters according to the reward or punishment that received from the humans.

Therefore, reinforcement learning algorithms is integrate to fuzzy to modify membership functions of the fuzzy social model. During the operation, the robot will navigate according to the estimated social space model. During the navigation, the robot receives the reward from the human. This thesis designed the reward in term of Interaction Degree $ID$ which is the quality of interaction between humans and the robot, Unacceptable Degree $UD$ which is the discomfort feeling from the humans, and the different path length compares between generated path length and optimal path length. Therefore, reinforcement learning algorithms are used to modify parameters in the fuzzy social model by receiving the reward from humans. Finally, the robot can modify the fuzzy social model similar to real human social spaces that the robot can perceive the maximum reward.

The famous reinforcement learning algorithms, such as Q-Learning, R-Learning, Actor-Critic, and Deep Q-Network, are used to compare the efficacy in this problem. Three criteria are used to evaluate the algorithms. First is the convergent of the error between the estimated and the real social map. This convergent can describe the ability to learn to approach the realistic social space. The second is the learning period which describes how fast of algorithms to learn and converge to get the maximum value. The third is the exploration rate of algorithms which used to explain how many state-action pairs that have been explored. This exploration rate is used to describe a possibility that selected action in each state is the best action or optimal action.

The results show that most algorithms can modified the membership functions that cause the estimate human social space similar to ground truth, except R-Learning. The reason is that the adaptation of the average reward, which used as the base to compute the state-action value, is sensitive to the action selection strategy ($\epsilon$ - greedy). The R-learning used the difference between an immediate reward and the average reward to compute the state-action value. This value is used in the action selection strategy. The average reward is updated every greedy action and approach to maximum average reward very fast. Then, when computing the state-action value, the comparison between the immediate reward from a new action and average reward going to a small value which makes that action in that state has a low probability of selecting again in the same state. Therefore, the algorithm keeps choosing the action that has high state-action space which makes the robot approaches the sub-optimal state instead optimal state. For learning time, Deep Q-Network is overcome other algorithms. This because Deep Q-Network has memory to store the past experience which can be reused to learn again. However, the fast learning time may have the trade-off with the state-action space exploration rate. In this criteria, Actor-Critic can explore the state-action space better than others algorithms but has trade of to the learning time.

The limitation of this research is that, it is based on the simulation. For the real-world task, the difficulty of this approach is that the robot needs to repeatedly observe human's response with a number of interactions before obtaining enough observations which is not appropriate to practical. This is the main disadvantage of using reinforcement learning. To overcome this problem, the concept of hierarchical learning may improve learning speed of reinforcement learning.

This thesis's contributions are useful to the users who have mobile robot to service themselves like health care or house hold service. The robot is able to estimate the users' private space according to the users' gender, experience with his/her robot or the range of the robot location to themselves, and has ability to adapt its estimation according to the users' feeling. These will be process automatically by the robot. Therefore, the user will feel more relax to have the robot to service in their environment.

## 5.2 Future Work

In recognition of the never-ending nature of research, there are always some aspects of any work that can be improved and expanded upon by future research. This may be caused by the limitations of time and resources, or other unforeseen difficulties. The same is true for this thesis, where some problem remains unsolved, and new topics appear continuously. Therefore, it is the wish of the author to suggest some ideas regarding the future direction of this work.

Firstly, the simulation results presented in this thesis is based on a particular period which means this simulation in on the static environment. However, in realistic, humans are moving which means the environment is dynamics. This dynamics may cause the efficacy of reinforcement learning algorithms.

Secondly, the selected reinforcement learning algorithms in this thesis used only one action selection strategy ($\epsilon$- greedy) which may be the reason that affects the efficacy of algorithms. Therefore, the action selection strategy should be considered and investigated to find which strategy is suited to each algorithm.

Thirdly, the learning period of all algorithms is still large which might be caused by the action selection strategy or algorithms. Therefore, the development of reinforcement learning algorithms still needs to accelerate the learning period and suit our problem. Lastly, the learning fuzzy social model should experiment with the real robot to obtain the proof of validity and effectiveness of our proposed model.

# REFERENCES

Aliakbari, M., Faraji, E., and Pourshakibaee, P. (2011). Investigation of the proxemic behavior of iranian professors and university students: Effects of gender and status. *Journal of Pragmatics*, 43(5):1392 – 1402. Multilingual structures and agencies.

Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T., and Christensen, H. I. (2004). Navigation for human-robot interaction tasks. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, volume 2, pages 1894–1900 Vol.2.

Arechavaleta, G., Laumond, J.-P., Hicheur, H., and Berthoz, A. (2008). On the nonholonomic nature of human locomotion. *Autonomous Robots*, 25(1):25–35.

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833.

Beheshti, R. and Sukthankar, G. R. (2014). A normative agent-based model for predicting smoking cessation trends. In *AAMAS*.

Berenson, D., Siméon, T., and Srinivasa, S. S. (2011). Addressing cost-space chasms in manipulation planning. In *2011 IEEE International Conference on Robotics and Automation*, pages 4561–4568.

Butler, J. T. and Agah, A. (2001). Psychological effects of behavior patterns of a mobile personal robot. *Auton. Robots*, 10(2):185–202.

Cuayáhuitl, H., Dethlefs, N., Frommberger, L., Richter, K.-F., and Bateman, J. (2010). Generating adaptive route instructions using hierarchical reinforcement learning. In Hölscher, C., Shipley, T. F., Olivetti Belardinelli, M., Bateman, J. A., and Newcombe, N. S., editors, *Spatial Cognition VII*, pages 319–334, Berlin, Heidelberg. Springer Berlin Heidelberg.

Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, A., Alami, R., and Siméon, T. (2006). How may i serve you?: A robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, pages 172–179, New York, NY, USA. ACM.

Diego, G. and Arras, T. K. O. (2011). Please do not disturb! minimum interference coverage for social robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1968–1973.

Edward, H. T. (1969). *The Hidden Dimension : man's use of space in public and in private*. The Bodley Head Ltd, London, UK.

Elena Pacchierotti, Henrik I. Christensen, P. J. (2006). *Embodied Social Interaction for Service Robots in Hallway Environments*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Erkelens, C. J. (2017). Perspective Space as a Model for Distance and Size Perception. *Iperception*.

Fox, D., Burgard, W., and Thrun, S. (1997). The dynamic window approach to collision avoidance. *IEEE Robotics Automation Magazine*, 4(1):23–33.

Gifford, R. (1982). Projected interpersonal distance and orientation choices: Personality, sex, and social situation. *American Sociological Association*, 45(3):145–152.

Gockley, R., Forlizzi, J., and Simmons, R. (2007). Natural person-following behavior for social robots. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 17–24.

Hansen, S. T., Svenstrup, M., Andersen, H. J., and Bak, T. (2009). Adaptive human aware navigation based on motion pattern analysis. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 927–932, Toyama,Japan.

Hao, J. and Leung, H.-f. (2013). Achieving socially optimal outcomes in multiagent systems with reinforcement social learning. 8.

Hayduk, L. A. (1987). Personal space: An evaluative and orienting overview. *Psychological Bulletin*.

Helbing, D. and Molnár, P. (1995). Social force model for pedestrian dynamics. *Phys. Rev. E*, 51:4282–4286.

Herring, J. (2014). *Medical Law and Ethics*. Oxford University.

Huang, K., Li, J., and Fu, L. (2010). Human-oriented navigation for service providing in home environment. In *Proceedings of SICE Annual Conference 2010*, pages 1892–1897.

Iehl, R., Cortés, J., and Siméon, T. (2012). Costmap planning in high dimensional configuration spaces. In *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 166–172.

Jaillet, L., Corcho, F. J., and Perez, J. J. (2011). Randomized tree construction algorithm to explore energy landscapes. *J. Comput. Chem*, pages 3464–74.

Jaillet, L., Cortés, J., and Siméon, T. (2010). Sampling-based path planning on configuration-space costmaps. *IEEE Transactions on Robotics*, 26(4):635–646.

Jens Kessler, Christof Schroeter, H.-M. G. (2011). *Approaching a Person in a Socially Acceptable Manner Using a Fast Marching Planner*. Springer Berlin Heidelberg, Berlin, Heidelberg.

K. L. Koay, E. A. Sisbot, D. S. S. M. L. W. K. D. and Alami, R. (2007). Exploratory study of a robot approaching a person in the context of handing over an object. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, page 18==24.

Kalbfleisch, P. and Cody, M. (2012). *Gender, Power, and Communication in Human Relationships*. Routledge Communication Series. Taylor & Francis.

Kastanis, I. and Slater, M. (2012). Reinforcement learning utilizes proxemics: An avatar learns to manipulate the position of people in immersive virtual reality. *TAP*, 9:3:1–3:15.

Kirby, R., Simmons, R., and Forlizzi, J. (2009). Companion: A constraint-optimizing method for person-acceptable navigation. In *in the Proceedings of the IEEE international Symposium on Robot and Human Interactive Communication*.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. pages 3–24, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Kruse, T., Pandey, A. K., Alami, R., and Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726 – 1743.

Lam, C. P., Chou, C. T., Chiang, K. H., and Fu, L. C. (2011). Human-centered robot navigation;towards a harmoniously human; robot coexisting environment. *IEEE Transactions on Robotics*, 27(1):99–112.

Lindner, F. (2015). A conceptual model of personal space for human-aware robot activity placement. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5770–5775, Hamburg.

Mainprice, J., Sisbot, E. A., Jaillet, L., Cortés, J., Alami, R., and Siméon, T. (2011). Planning human-aware motions using a sampling-based costmap planner. In

*2011 IEEE International Conference on Robotics and Automation*, pages 5012–5017.

Martinez-Garcia, E. A., Akihisa, O., and Yuta, S. (2005). Crowding and guiding groups of humans by teams of mobile robots. In *IEEE Workshop on Advanced Robotics and its Social Impacts, 2005.*, pages 91–96.

Martinez-Gil, F., Lozano, M., and Fernández, F. (2012). Calibrating a motion model based on reinforcement learning for pedestrian simulation. In Kallmann, M. and Bekris, K., editors, *Motion in Games*, pages 302–313, Berlin, Heidelberg. Springer Berlin Heidelberg.

Martinson, E. (2007a). *Acoustical awareness for intelligent robotic action*. Atlanta, GA, USA,.

Martinson, E. (2007b). Hiding the acoustic signature of a mobile robot. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 985–990.

Martinson, E. and Brock, D. (2007). Improving human-robot interaction through adaptation to the auditory scene. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 113–120.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Müller, J., Stachniss, C., Arras, K. O., and Burgard, W. (2008). Socially inspired motion planning for mobile robots in populated environments.

Mumm, J. and Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 331–=338.

Nakauchi, Y. and Simmons, R. (2000). A social robot that stands in line. In *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)*, pages 357–364 vol.1, Takamatsu, Japan.

Newell, P. B. (1998). A cross-cultural comparison of privacy definitions and functions: A systems approach. *Journal of Environmental Psychology*, 18(4):357 – 371.

Pacchierotti, E., Christensen, H. I., and Jensfelt, P. (2006). Evaluation of passing distance for social robots. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 315–320.

Pandey, A. K. and Alami, R. (2010). A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5855–5860, Taipei.

Papadakis, P., Rives, P., and Spalanzani, A. (2014). Adaptive spacing in human-robot interactions. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2627–2632, Chicago, USA.

Patompak, P., Jeong, S., Chong, N. Y., and Nilkhamhang, I. (2016). Mobile robot navigation for human-robot social interaction. In *2016 16th International Conference on Control, Automation and Systems (ICCAS)*, pages 1298–1303.

Patompak, P., Jeong, S., Nilkhamhang, I., and Chong, N. Y. (2017). Learning social relations for culture aware interaction. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 26–31.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.

Saulnier, P., Sharlin, E., and Greenberg, S. (2011). Exploring minimal nonverbal interruption in hri. In *2011 RO-MAN*, pages 79–86.

Scandolo, L. and Fraichard, T. (2011). An anthropomorphic navigation scheme for dynamic scenarios. In *2011 IEEE International Conference on Robotics and Automation*, pages 809–814.

Schiano Lomoriello, A., Meconi, F., Rinaldi, I., and Sessa, P. (2018). Out of sight out of mind: Perceived physical distance between the observer and someone in pain shapes observer's neural empathic reactions. *ArXiv e-prints*.

Schuurmans, D. and Zinkevich, M. (2016). Deep learning games.

Sheridan, T. B. (2016). Human–robot interaction: Status and challenges. *Human Factors*, 58(4):525–532.

Sisbot, E. A., Marin-Urias, L. F., Alami, R., and Simeon, T. (2007). A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883.

Stachniss, C. and Burgard, W. (2002). An integrated approach to goal-directed obstacle avoidance under dynamic constraints for dynamic environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1, pages 508–513 vol.1.

Svenstrup, M., Bak, T., and Andersen, H. J. (2010). Trajectory planning for robots in dynamic human environments. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4293–4298, Taipei.

Takayama, L. and Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. In *Intelligent robots and systems, 2009. IROS 2009. international conference on IEEE/RSJ*, pages 5495–5502.

Tamura, Y., Le, P. D., Hitomi, K., Chandrasiri, N. P., Bando, T., Yamashita, A., and Asama, H. (2012). Development of pedestrian behavior model taking account of intention. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 382–387.

Tomari, R., Kobayashi, Y., and Kuno, Y. (2012). Empirical framework for autonomous wheelchair systems in human-shared environments. In *2012 IEEE International Conference on Mechatronics and Automation*, pages 493–498, Chengdu.

Walters, M. L. (2008). The design space for robot appearance and behaviour for social robot companions.

Westin, A. (1970). *Privacy and Freedom*. Bodley Head.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3):338 – 353.

Zeeger S. K., Readdick C.A., H.-G. S. (1994). Daycare children's establishment of territory to experience privacy. 11.

# BIOGRAPHY

| | |
|---|---|
| Name | Mr. Pakpoom Patompak |
| Date of Birth | June 15, 1987 |
| Education | 2009: Bachelor of Engineering (Telecommunications Engineering)<br>Sirindhorn International Institute of Technology<br>Thammasat University<br>2011: Master of Engineering (Information and Communication Technology for Embedded Systems)<br>Sirindhorn International Institute of Technology<br>Thammasat University |

Publications

Patompak, Pakpoom and Jeong, Sungmoon and Nilkhamhang, Itthisek and Chong, Nak. (2019). Learning Proxemics for Personalized Human–Robot Social Interaction. *International Journal of Social Robotics*, 12(1), 267-280 (2020).

Patompak, Pakpoom and Jeong, Sungmoon and Nilkhamhang, Itthisek and Chong, Nak. (2017). Learning social relations for culture aware interaction. 14th *International Conference on Ubiquitous Robots and Ambient Intelligence (URAI).* (pp. pp. 26-31). Jeju.