Glasgow College, UESTC

电子科技大学
格拉斯哥学院
Glasgow College, UESTC

Project Specifications and Preliminary Report on UESTC4006P(BEng) Final Year Project

Please start by saving this file with the name: GUID_Surname_UESTC4006P_Preliminary_year

***** Please add appropriate course code*

Student Name	Chen Yu
Student Matriculation Number	2288975Y
UESTC Student Number	2016200104015
Degree programme	Electronics and Electrical Engineering (UESTC) BEng
Academic year	4

Placement Company (if appropriate)	--
Working Title of Project	Reinforcement Learning Policy-Search Algorithms for a Walking Robot
Name of First Supervisor	Andre Rosendo
Name of Second Supervisor	Faisal Tariq
Declaration of Originality and Submission Information	*I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements* Signed (Student) :

Your report should be NO more than 6 pages in length and include the below subject headings and incorporated within this document:

Project Description (no more than half a page)
Technical Background (no more than one page)
Main tasks and targets (no more than half a page)
Measureable Outcomes (no more than half a page): Tangible outcomes (Hardware, Software, Hardware & Software, Theoretical research)
Project outline (no more than one page)
Work plan (no more than one page)

Resources: Complete the component request form and email the form to your 1st supervisor separately
Rist Assessment Form: You may be asked to submit a Risk Assessment Form if your project includes hazardous activities. Please check with your supervisor if you need to submit it.

Deadlines for submission of this report: Please upload this report via the Moodle page by the deadline mentioned in Table 1 of your project handbook.

Comments from your Second Supervisor will be made via Moodle or via email.

Project Specifications and Preliminary Report on UESTC4006P(BEng) Final Year Project

Feedback from Second Supervisors: Second supervisors may provide their feedback by adding comments directly on Moodle taking into account the questionnaire below **or** by filling out the below form and uploading it to Moodle.

Name of Second Supervisor	

Was the report satisfactory?	Yes	☐	No	☐
Is project suitable for an BEng project?	Yes	☐	No	☐
Is the project plan feasible?	Yes	☐	No	☐
Would you like to give any suggestions/recommendations?	Yes	☐	No	☐

Please write your comments in the space provided below:

Signature: Date:

Project Description

Legged robot platforms offer many advantages over traditional wheeled robots. In addition to their ability to traverse a wide variety of terrain, walking robots are basically a requirement for performing useful tasks in our human-centric world [1]. However, despite these advantages, walking is also one of the fundamental challenges for legged robots. Optimizing a robot's gait is not a simple control problem since designing agile locomotion for quadruped robots often requires extensive expertise and tedious manual tuning [2]. In this paper, I would present a system to automate this process by leveraging reinforcement learning (RL) techniques. The system would learn a walking strategy from scratch using an algorithm that combines reinforcement learning and Bayesian optimization (BO) or other state-of-the-art optimization methods.

The main body of the project would be a legged robot that could move forward without previous knowledge of "how to move the legs", and it would also adjust its gait and hence adapt to different environments automatically. This feature could be useful in practice since a real-world environment is typically unstructured, unknown and can exhibit obstacles which have to be negotiated. The meaning of the project is that the algorithm could then be used for industrial inspection, mining, agriculture, search and rescue missions, and other areas in the future. These environments require robots that can cope with obstacles, rough outdoor terrain and steps, which is still challenging for both industry and academe nowadays.

Technical Background

As a joint field of artificial intelligence and modern statistics, machine learning is concerned with the design and development of algorithms and techniques that allow computers to automatically extract information and "learn" structure from data. The learned structure can be described by a statistical model that compactly represents the data. As a branch of machine learning, reinforcement learning is a computational approach to learning from interactions with the surrounding world and concerned with sequential decision making in unknown environments to achieve high-level goals. The basic method of this project would be reinforcement learning, and a new algorithm would be proposed combining Bayesian learning to find a successful policy to solve the existing problems.

Reinforcement Learning. RL considers a learning agent and its interactions with the environment [3]. In RL, the agent would try to maximize its accumulated reward over its lifetime. It enables robotic systems to improve continuously through experience and hence has been applied to various robotic manipulation tasks, such as folding clothes [4], pouring water [5] and so on. However, to date, RL still often suffers from being data inefficient, i.e., RL requires too many trials to learn a particular task [6]. For example, learning one of the simplest RL tasks, the mountain-car, often requires tens if not hundreds or thousands of trials— independent of whether policy iteration, value iteration or policy search methods are used. In this project, I would adopt Gaussian processes (GPs) and Bayesian optimization (BO) for policy search to overcome these problems.

Gaussian Processes. GPs are a kind of stochastic processes, which is a powerful tool in the machine learning toolbox. They allow us to make predictions about our data by incorporating prior knowledge. Their most obvious area of application is fitting a function to the data. GPs have been used before to learn models of robotic systems, see [7, 8], for instance. However, in these methods either expert knowledge was given in the form of a teacher, by demonstrations, or a vague dynamics model or the GP models were not used for long-term planning/policy evaluation.

Bayesian Optimization. BO is a novel strategy to black-box global optimization, especially when the objective function is expensive to evaluate [9]. It uses Gaussian Processes to approximate the (true) objective function and an acquisition function to promote a balance between exploitation and exploration in the selection of the next iteration [1].

In [10, 11], and [12] BO is applied in a 2-dimensional bipedal locomotion problem, where [10] is used in simulations and [11, 12] being applied to a real bipedal robot. In [13] the authors show that a hexapod robot could use BO to adapt their controllers to damage.

Probabilistic Inference for Learning Control. Probabilistic Inference for Learning Control (PILCO) is a policy search approach [6] that to some extent, combines reinforcement learning and Bayesian optimization to speed up the efficiency of the training process. PILCO reduces model bias, one of the key problems of model-based reinforcement learning, in a principled way. By learning a probabilistic dynamics model and explicitly incorporating model uncertainty into long-term planning, PILCO can cope with very little data and facilitates learning from scratch in only a few trials.

Random Search. Random Search (RS) belongs to the fields of Stochastic Optimization and Global Optimization. Random Search is a direct search method as it does not require derivatives to search a continuous domain. For a better understanding of the problem to optimize, a Random Search method would also be performed for a better understanding of the task from a real-world perspective.

Main Tasks And Targets

The eventual task of the project is to build a walking robot that moves forward as far as possible. Overall, the targets of the robot would be:

1. The robot would learn to walk by itself without prior expert knowledge.
2. The robot would move forward as far and as fast as possible.
3. The robot would adapt to different environments and learn new gait by itself.

To achieve this, there are two sub-tasks in the form of hardware and software. First, a physical robot platform should be built. The objectives of this platform are as followed:

1. It should be a legged walking robot, moving in the 2D or 3D space, controlled by mbed or other microcontrollers.
2. It should have a robust structure, for multi-iteration experiment and practical application.
3. The appearance of the robot should be as enchanting as possible, with a sense of “the future is coming”.

Then an appropriate algorithm (software) should be designed, with the goals of:

1. The training process should be as fast as possible (with least iterations of training).
2. The code should be effective and clear.
3. The algorithm could run on a PC or be embedded in the robot.

In the experiment, we may use the evaluation criteria for the vertical distance y (the vertical component of the initial and end positions’ displacement) that the robot walked forward within a fixed period of time, and if it goes backwards, the distance y would be negative.

Measurable Outcomes

Overall, the main outcome of the project is a physical legged robot that could learn to walk forward by itself. In addition, the robot could try to adjust its gait and hence adapt to different environments. This outcome consists of two parts: hardware and software.

In term of the hardware, the basic mechanical structure of the robot is a legged platform, which is designed on the computer and 3D printed. The platform could move in 3-Dimension space, with the help of the servo motor on each joint of the robot. There would also be a microcontroller on the moving robot platform, playing a role of the brain of the robot. Signals from the microcontroller would control the servo motors on the robot and hence move it forward.

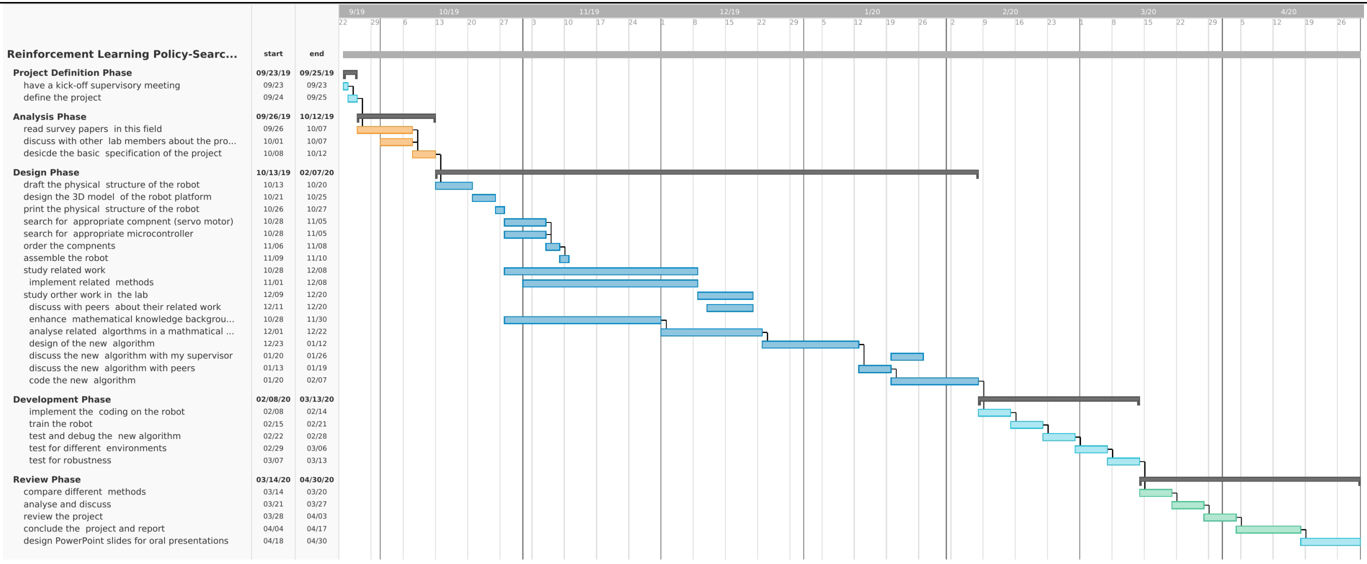
As for the software, the controller is based on the reinforcement learning algorithm, a branch of machine learning methods. The input of the controller is the state of the environment, i.e., the current condition of the robot, while the output is the action of the agent, i.e., the controlling signals from the microcontroller. To overcome the problem of traditional reinforcement learning, the algorithm used in this project should be a combination of reinforcement learning and some state-of-the-art optimization methods.

Project Outline

1 Introduction
2 Related work
  2.1 Reinforcement learning
  2.2 Gaussian Processes
  2.3 Bayesian Optimization
  2.4 Probabilistic Inference for Learning Control
  2.5 Other related work
3 The hardware part of the robot
  3.1 Design of the physical structure
  3.2 Oscillators or other motors on the robot.
  3.3 Set up of the microprocessor
4 The software part of the robot
  4.1 Implementation of some basic reinforcement learning algorithms
    4.1.1 Introduction of some related work
    4.1.2 Implementation of some previous methods in similar contexts
  4.2 Implementation of some basic bayesian optimization algorithms
    4.2.1 Introduction some related work and related mathematic theories
    4.2.2 Implementation of some previous methods
  4.3 Design of the new combined algorithm appropriate for the targets.
    4.3.1 Introduction some related work, materials and theories
    4.3.2 Implementation of the method on the robot
5 Experiment and results
  5.1 The experiment of the new algorithm on the robot
  5.2 The experiment of other algorithms on the robot
  5.3 Comparison of different methods
  5.4 Visualisation of the results
  5.5 Analysis of the results
6 Discussion and further work
7 Conclusion

Work Plan

A Gantt chart of the project for work plan is shown as followed.



Reference

[1] D. J. Lizotte, T. Wang, M. Bowling, and D. Schuurmans, "Automatic gait optimization with gaussian process regression," in *IJCAI*, 2007.

[2] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," *Robotics: Science and Systems XIV*, Jun 2018.

[3] M. Deisenroth, G. Neumann, and J. Peters, *A Survey on Policy Search for Robotics*, vol. 2. 08 2013.

[4] J. Matas, S. James, and A. J. Davison, "Sim-to-real reinforcement learning for deformable object manipulation," *CoRR*, vol. abs/1806.07851, 2018.

[5] C. Schenck and D. Fox, "Visual closed-loop control for pouring liquids," *CoRR*, vol. abs/1610.02610, 2016.

[6] M. P. Deisenroth and C. E. Rasmussen, "Pilco: A model-based and data-efficient approach to policy search," in *International Conference on International Conference on Machine Learning*, 2011.

[7] J. Ko, D. J. Klein, D. Fox, and D. Haehnel, "Gaussian processes and reinforcement learning for identification and control of an autonomous blimp," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 742–747, April 2007.

[8] D. Nguyen-Tuong, M. Seeger, and J. Peters, "Local gaussian process regression for real time online model learning and control," in *Advances in neural information processing systems 21*, (Red Hook, NY, USA), pp. 1193–1200, Max-Planck-Gesellschaft, Curran, June 2009.

[9] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2010.

[10] K. A. Saar, A. Rosendo, and F. Llda, "Bayesian optimization of gaits on a bipedal slip model," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1812–1817, Dec 2017.

[11] R. Calandra, A. Seyfarth, J. Peters, and M. P. Deisenroth, "An experimental comparison of bayesian optimization for bipedal locomotion," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1951–1958, May 2014.

[12] A. Rai, R. Antonova, S. Song, W. Martin, H. Geyer, and C. Atkeson, "Bayesian optimization using domain knowledge on the atrias biped," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[13] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, p. 503- 507, May 2015.