ON THE EXPLOITATION OF UNCERTAINTY TO IMPROVE MAXIMUM EXPECTED VALUE ESTIMATE AND EXPLORATION IN REINFORCEMENT LEARNING

Doctoral Dissertation of:
Carlo D'Eramo

Supervisor:
Prof. Marcello Restelli

Tutor:
Prof. Andrea Bonarini

The Chair of the Doctoral Program:
Prof. Marcello Restelli

2019 – XXXI cycle

Abstract

ABSTRACT goes here.

Summary

SUMMARY goes here.

Contents

List of Figures

List of Tables

Introduction

E VERYONE experiences the process of taking decisions during his life. As a matter of fact, drastically the life of an individual can be synthesized in its *perception* of the world and its *interaction* with it. The concepts of perception and interaction might seem quite straightforward to understand: for a human the perception of the world comes from its senses and the interaction comes from its possibility to change its surroundings. On the contrary, these concepts are actually absolutely hard to define and aroused, during the centuries, a strong debate between scientists, biologists, and even philosophers.

1.1 Perception and interaction

We start from the assumption that, by definition, an individual perceive the environment around it and acts on it in order to achieve goals expressed by its will. In other words, all the actions made by the individual are done to satisfy its will to obtain something from the world it lives in. This task is naturally performed by humans, but it implies some challenging problems that are hard, or unfeasible, to solve. One of them comes from the intrinsic *uncertainty* of the perception we have of the world around us. The perception of the world around an individual consists in its interpretation of the information provided by its senses, but the process of information retrieval by senses and the mental processes to understand them, inevitably introduce a certain level of noise that distorts the original true information. On the other hand, the interaction with the world deals with the will of the individual to perform actions to change the environment around it, but this apparently simple operation involves complex biologic mechanisms to coordinate the body according to the will and the difficulties in the perception of the consequences of the interaction. Moreover, the concept of goal can be unclear and the

individual may result in performing actions without being sure of what it want. It is arguable that discussing about the concept of true information and the concept of will requires strong theoretical considerations since they are both hardly definable concepts. For many centuries scientists and philosophers debated about these topics, in particular trying to solve complex problems like the real nature of perceivable things and the concept of free will. However, to make the discussion of these concepts suitable for our purposes throughout all this thesis, we lighten the definition of them to those provided by common sense.

1.2 Learn how to act with Reinforcement Learning

Reinforcement Learning (RL) [1] is a subfield of Machine Learning (ML) which aims to realize autonomous *agents* able to learn how to act in a certain *environment* in order to maximize an objective function; this is achieved providing the agent with the perception of its *state* in the environment and making it learn the appropriate *action* to perform, where an action can be seen as an atomic operation that brings the agent from a state to another one. The objective function represents a measure of how well the agent is accomplishing its task in the environment and it is usually formalized by a discounted sum of *rewards* obtained after each action. The sum is discounted to give more importance to the most recent rewards w.r.t. the ones further in the future. The reward function, i.e. the function returning the reward after each action, is not a concrete part of the environment, but it is often formalized by a human which decides which actions have to be reinforced (returning a positive reward) or inhibited (returning a negative reward).

1.2.1 Uncertainty in Reinforcement Learning

The major challenge of RL is represented by the uncertainty. In fact, initially, the agent is not provided with the knowledge of how the environment will react to its action, thus it does not know whether an action would be good or not to maximize its objective function. Indeed, before trying an action, it does not know if that action will get a positive or a negative reward, and it does not know if that action will let it go to the desired state or not. Thus, the former problem can be seen as uncertainty in the reward function and the latter as uncertainty in the transition (i.e. model) function. In some cases, also the uncertainty in the perception of the current state of the agent is considered making the problem more complex.

The uncertainty issue results in the need of the agent to try actions in order to improve its knowledge of the environment and exploit it to maximize its objective function. Moreover, since the objective function is a sum of discounted rewards where later rewards worth less than recent ones, the agent needs to learn fast. The need to explore and the need to *exploit* the actions believed to be good introduces an important problem known as exploration-exploitation dilemma.

Preliminaries

CHAPTER 3

Maximum Expected Value estimate

CHAPTER *4*

Exploration

Deep

CHAPTER 6

Mushroom

CHAPTER 7

Conclusion

Bibliography

[1] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*. MIT press, 1998.