Artificial Intelligence: Probabilistic & learning techniques

Chris Pal

Project (INF8225)

Farnoush Farhadi & Juliette Tibayrenc

21 Apr. 2016

POLYTECHNIQUE
MONTRÉAL

LE GÉNIE
EN PREMIÈRE CLASSE

Contents

Introduction **2**

1. Reference article **3**

1.1 Literature review 3

1.2 Article summary 3

2. Theoretical elements **4**

2.1 Value iterating & policy iterating vs. Q-learning 4

2.1.1 Value iteration 4

2.1.2 Policy iteration 4

2.1.3 Classic Q-learning 4

2.2 Reward matrix 5

3. Comparing algorithms **6**

3.1 Classic Q-learning 6

3.2 Speedy Q-learning 6

3.3 Additional remarks 6

Further work and conclusion **7**

Bibliography **8**

Introduction

Our AI course has allowed us to discover Markov Decision Processes, which we decided to further explore in our project. Looking for an application for them, we stumbled upon Mastronarde & Van der Schaar's article, 'Fast Reinforcement Learning for Energy-Efficient Wireless Communications'[1], and we chose it as the basis for our work.

The article presents the two researchers' work in developing an algorithm that would enable systems to adopt the best policy to quickly send data packets in the most efficient way possible, an important goal to achieve since the majority of today's surfing is done on smartphones whose battery does not last longer than a day (at best).

In this report, we will first introduce the article, with a quick review of the state of the art at the time it was written (2011), then summarize it. We will continue with a presentation of some theoretical elements, highlighting the differences between what we saw in class and the processes on which the article and our implementations are based, and the way the reward matrix used is calculated (a process that does not appear in the article). Finally we will talk about our own implementation of classic Q-learning and one of its variants, speedy Q-learning, and compare the results we obtain with the results obtained with the authors'algorithm, plus the results we get by using value iteration and policy iteration.

1. Reference article

1.1 Literature review

1.2 Article summary

2. Theoretical elements

2.1 Value iterating & policy iterating vs. Q-learning

In our project, we compare the performance of several algorithms against that of the authors. These algorithms have to be evaluated along several criteria, among which the elements that need to be known for them to work, the speed with which they are executed (partly dependent on the efficency of the implementation and the language used) and the number of iterations their internal loop is executed for.

These algorithms are value iteration, policy iteration, classic Q-learning and speedy Q-learning.

2.1.1 Value iteration

2.1.2 Policy iteration

2.1.3 Classic Q-learning

Q-learning is part of a second category of algorithms that do not presuppose the full knowledge of the system. Contrary to value iteration and policy iteration, the environment does not need to be known for the algorithm to work. The algorithm presented by the authors of our reference article is similar in its requirements, though it goes further and actually uses the knowledge we do have about the environment, which makes it more efficient.

Q-learning helps us find the optimal policy by using the interactions between the system and the environment. In it, we compute an estimate of the cumulative discounted reward for each action executed in each state.

The algorithm for one step is as follows (from Artificial Intelligence, 3rd edition, French version[2]:

Algorithm 1 Q-learning algorithm for one step

```
Data: $s'$ /* current state */, $r'$ /* reward signal */
Result: $Q$ /* action values table indexed by states and actions */, $T$ /* frequency
        of state-action couples */, $s, a, r$ /*previous state, action and reward */
if $s$ final then
    $Q(s, Empty) = r'$
end
if $s \neq 0$ then
    $T(s, a) + +$
    $Q(s, a) + = \alpha \cdot T(s, a)(r + \gamma \max_{a'} Q(s', a') - Q(s, a))$
end
$s, a, r = s',_{a'} f(Q(s', a'), T(s', a')), r'$
return $a$
```

2.2 Reward matrix

3. Comparing algorithms

3.1 Classic Q-learning

3.2 Speedy Q-learning

3.3 Additional remarks

Further work & conclusion

Bibliography

[1] N. Mastronarde and M. Van der Schaar, “Fast reinforcement learning for energy-efficient wireless communication,” *IEEE Transactions on Signal Processing*, vol. 59, 12 2011.

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence, Prentice Hall, 3rd ed., 2010.