contextual: Simulating Contextual Multi-Armed Bandit Problems in R

Robin van Emden
JADS

Eric Postma
Tilburg University

Maurits Kaptein
Tilburg University

Abstract

A large number of statistical decision problems in the social sciences and beyond can be framed as a (contextual) multi-armed bandit problem.

Keywords: contextual multi-armed bandits, simulation, sequential experimentation, R.

A vignette for the van Emden, Kaptein, and Postma (2018) paper.

1. Introduction

The need to make decisions permeates our lives. We are continuously making them in order to affect the world around us. Decisions range from the trivial, like 'what should I have for breakfast?', to the much more serious, like 'how should a power station be controlled safely?'. Many of these decisions are now made on our behalf by automated systems. From automated stock trading systems, to cars that can navigate and drive, there are an increasing number of automated decision-making systems that already, or soon will, have an increasing effect on our lives. With the increased ubiquity of computing devices and sensors, the importance of automating decision making based on this influx of data becomes of greater interest. How should we make systems that can make these decisions for us?

The first attempts at making decision-making expert systems concentrated on the use of logic and formal reasoning. Such systems were brittle since feeding the system with sufficient ground truths is difficult and time consuming, and more importantly not all outcomes are entailed by a purely logic system especially when they are probabilistic or uncertain. More recently progress has been made towards more robust systems that take a different approach, which instead learn how to make decisions through experience. This thesis is interested in this aspect of decision-making systems, how a decision-making agent should make decisions and interact with an environment, improving their behaviour through experience. The specific decision-making problems with which we are concerned can be explained by way of examples.

— adapt the above rigorously —

2. Contextual Multi-Armed Bandits

In the canonical multi-armed bandit (MAB) problem a gambler faces a number of slot machines, each with a potentially different payoff. It is the gamblers goal to make as much profit (or, in the case of gambling, as little loss) as possible by sequentially choosing which machine to play, learning from the observations as she goes along.

3. Implementation of the contextual R package

In the canonical multi-armed bandit (MAB) problem a gambler faces a number of slot machines, each with a potentially different payoff. It is the gamblers goal to make as much profit (or, in the case of gambling, as little loss) as possible by sequentially choosing which machine to play, learning from the observations as she goes along.

4. A basic example

```
library("contextual")

bandit      <- BasicBandit$new()
bandit$set_weights(c(0.1, 0.9))

policy      <- EpsilonGreedyPolicy$new()
agent       <- Agent$new(policy, bandit)
simulation  <- Simulator$new(agent, horizon = 100L, simulations = 100L)
history     <- simulation$run()

Plot$new()$grid(history)
```

For results, see Figure 1 on page 3.

5. Object orientation: extending contextual

The R6 package allows the creation of classes with reference semantics, similar to R's built-in reference classes. Compared to reference classes, R6 classes are simpler and lighter-weight, and they are not built on S4 classes so they do not require the methods package. These classes allow public and private members, and they support inheritance, even when the classes are defined in different packages.

One R6 class can inherit from another. In other words, you can have super- and sub-classes.

Subclasses can have additional methods, and they can also have methods that override the superclass methods. In this example of a custom **contextual** bandit, we'll extend BasicBandit and override the initialize() method..

6. Special features

For instance, quantifying variance..

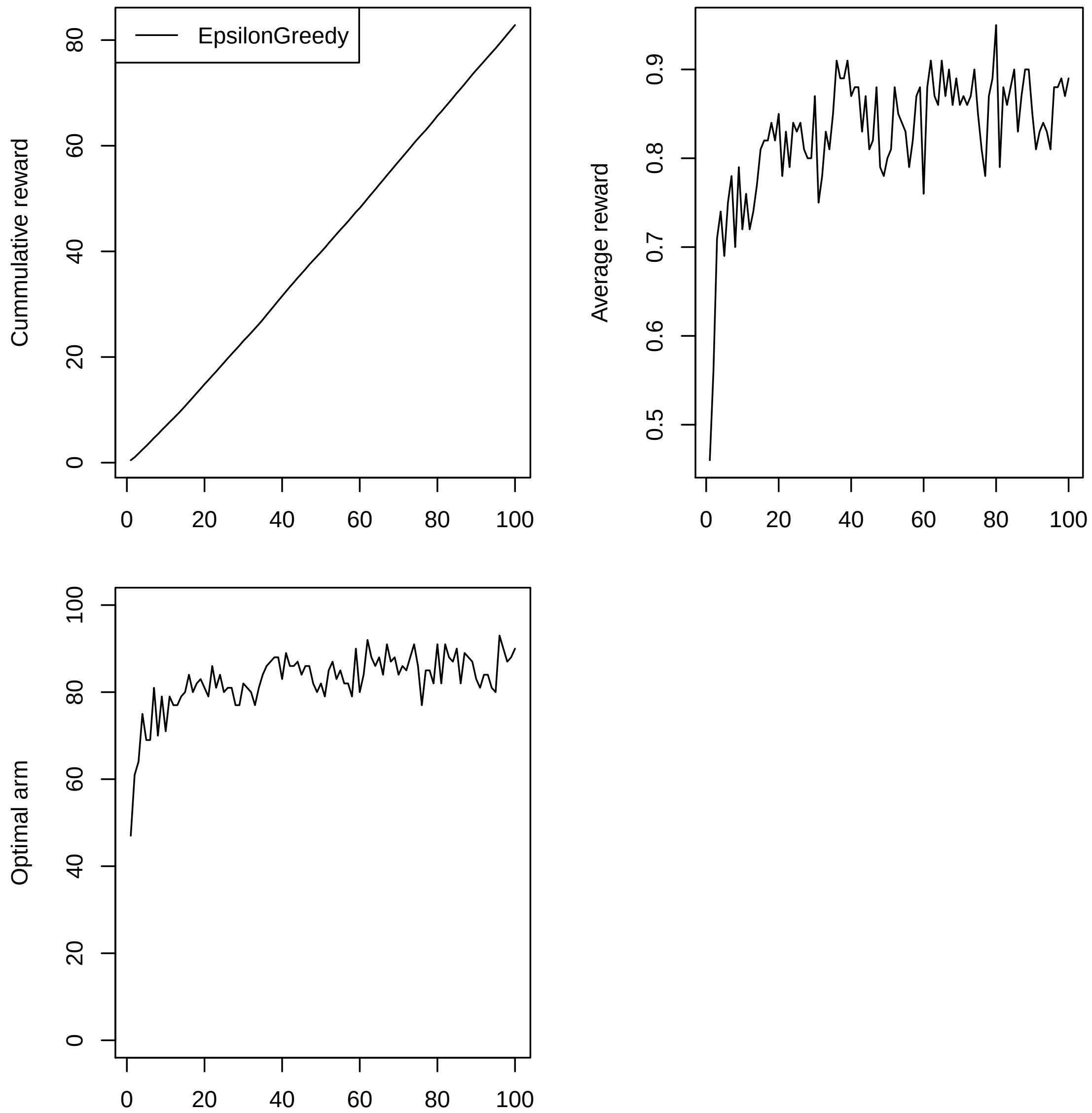

Figure 1: Epsilon Greedy

7. The art of optimal parallelisation

There is a very intersting trade of between the amount of parallelisation (how many cores, nodes used) the resources needed to compute a certain model, and the amount of data going to and fro the cores.

PERFORMANCE DATA ————————————————————

on 58 cores: k3*d3 * 5 policies * 300 * 10000 –> 132 seconds

on 120 cores: k3*d3 * 5 policies * 300 * 10000 –> 390 seconds

—

on 58 cores: k3*d3 * 5 policies * 3000 * 10000 –> 930 seconds

on 120 cores: k3*d3 * 5 policies * 3000 * 10000 –> 691 seconds

8. Extra greedy UCB

In the canonical multi-armed bandit (MAB) problem a gambler faces a number of slot machines, each with a potentially different payoff. It is the gamblers goal to make as much profit (or, in the case of gambling, as little loss) as possible by sequentially choosing which machine to play, learning from the observations as she goes along.

9. Conclusions

The goal of a data analysis is not only to answer a research question based on data but also to collect findings that support that answer. These findings usually take the form of a table, plot or regression/classification model and are usually presented in articles or reports.

10. Acknowledgments

Thanks go to CCC.

References

van Emden R, Kaptein M, Postma E (2018). *contextual: Simulating Contextual Multi-Armed Bandit Problems in R.* Jheronimus Academy of Data Science.

Affiliation:

Robin van Emden
Jheronimus Academy of Data Science
Den Bosch, the Netherlands
E-mail: robin@pwy.nl
URL: pavlov.tech

Eric O. Postma
Tilburg University
Communication and Information Sciences
Tilburg, the Netherlands
E-mail: e.o.postma@tilburguniversity.edu

Maurits C. Kaptein
Tilburg University
Statistics and Research Methods
Tilburg, the Netherlands
E-mail: m.c.kaptein@uvt.nl
URL: www.mauritskaptein.com