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

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.

2. The Contextual Multi-Armed Bandit

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. Structure 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

bandit <- BasicBandit$new()

bandit$set_weights(matrix(c(0.1, 0.9, 0.1, 0.5, 0.1, 0.1), 3, 3))

policy <- EpsilonGreedyPolicy$new() agent <- Agent$new(policy, bandit) simulation <- Simulator$new(agent, horizon = 30L, simulations = 30L, worker_max = 1) context <- bandit$get_context() history <- simulation$run())

5. Special features

For instance, quantifying variance..

6. 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

7. 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.

8. 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.

9. 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