Machine Learning in Simulation

Thomas Newman
Supervisor: Graham Laidler

September 2, 2020

Stochastic simulation:

Is a method for analysing the performance of systems whose behaviour depends on the interaction of random processes, processes that can be fully characterised by probability models.

Reasons to use stochastic simulations:

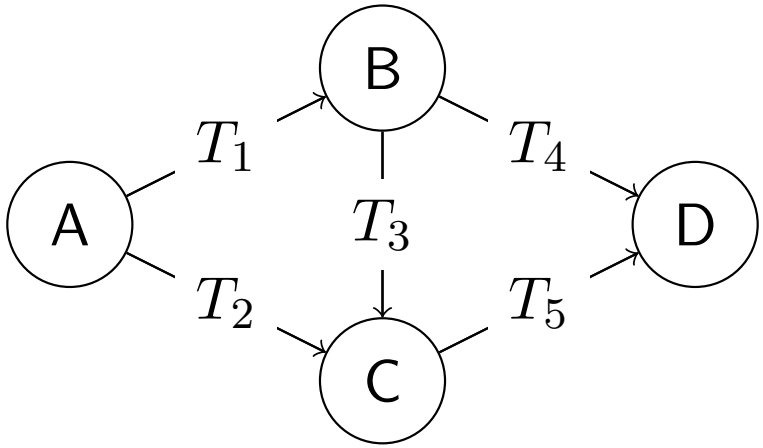
1. Feasibility: Will a project "work"?
2. Sensitivity: How important are the things we do not know?
3. Optimisation: What are the good options and how good are they?

Simulation analytics:

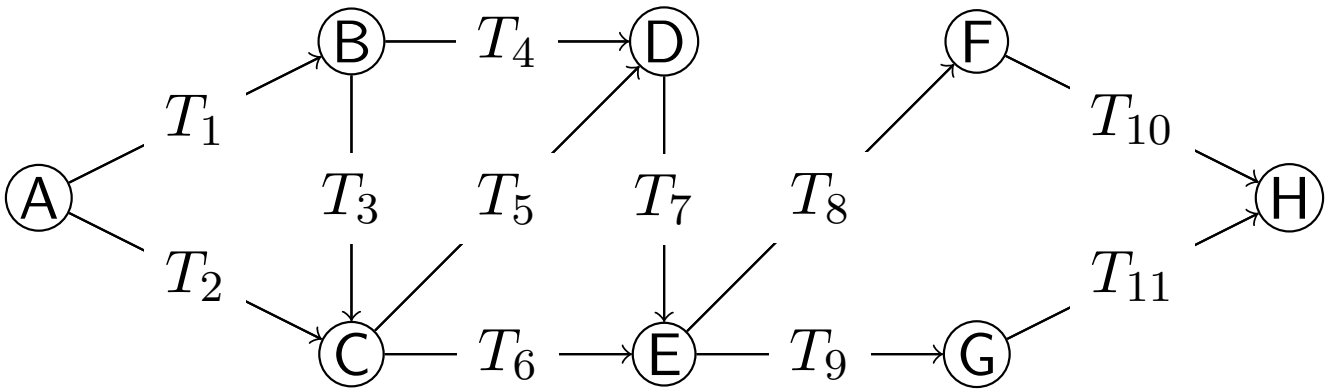
Simulation analytics refers to the methodology of applying machine learning or data analytics to the data generated by a stochastic simulation in order to understand more about how it behaves.

Reasons to use simulations analytics:

1. Understand the relationships of inputs and system state to outputs.
2. Full characterization of the observed output behaviour, marginally at a point in time, and dynamically across time.
3. Understanding about how and why alternative system designs differ, and how they will behave if implemented.
4. To generate inverse conditional statements: relationships of outputs to inputs or the system state.



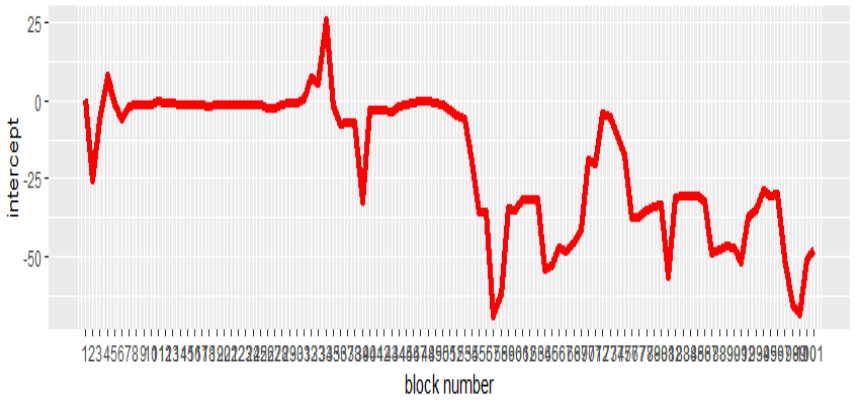
Time	Task1	Task2	Task3	Task4	Task5	Longest Path
0	0	0	0	0	0	19.32
1	1	1	0	0	0	19.32
2	1	2	0	0	0	19.32
3	2	2	1	1	0	19.32
...	...	...	...	...	...	19.32
19	2	2	2	2	1	19.32
20	2	2	2	2	2	19.32
0	0	0	0	0	0	21.07
...	...	...	...	...	...	21.07

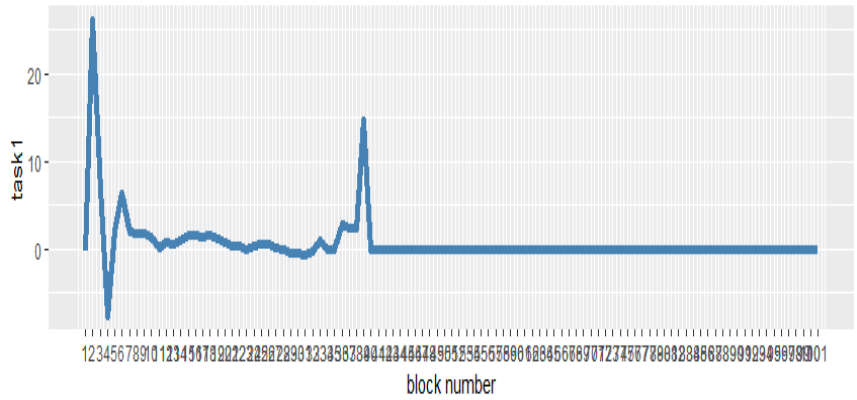


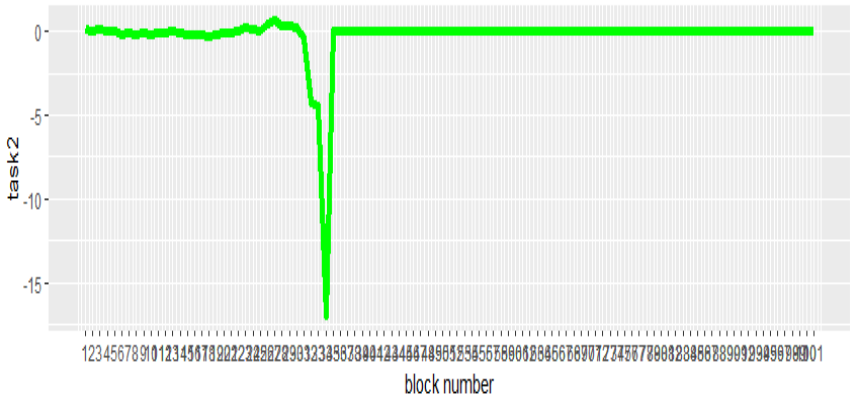
Time	Task1	Task2	Task3	...	Task11	Longest Path
8	2	2	1	...	0	60.14
55	2	2	2	...	1	59.32
13	2	2	2	...	0	57.84
3	1	1	0	...	0	61.55
...	...	...	...	...	...	...

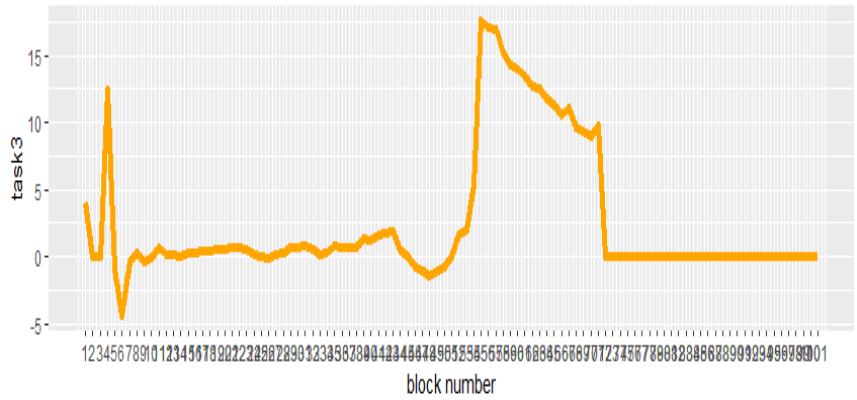
1. Split data into blocks.
2. Predict using K-fold cross validation logistic regression.
3. Plot task coefficients, correct classification rate and Kappa.

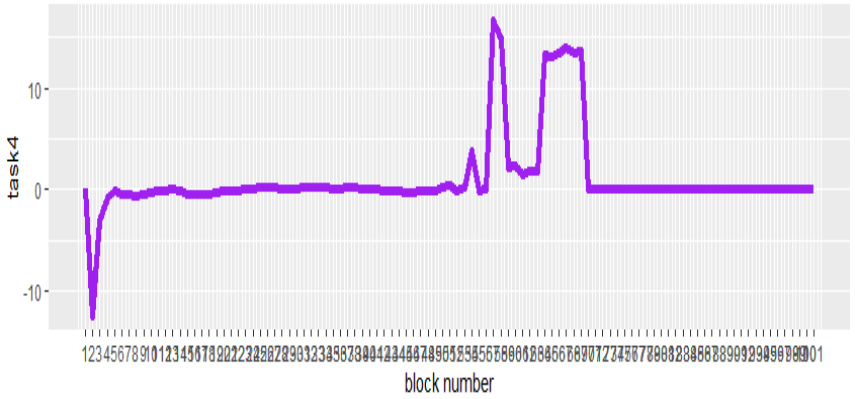
1. Creating overlap between consecutive blocks.
2. Changing the probability distributions of task durations to something more realistic.

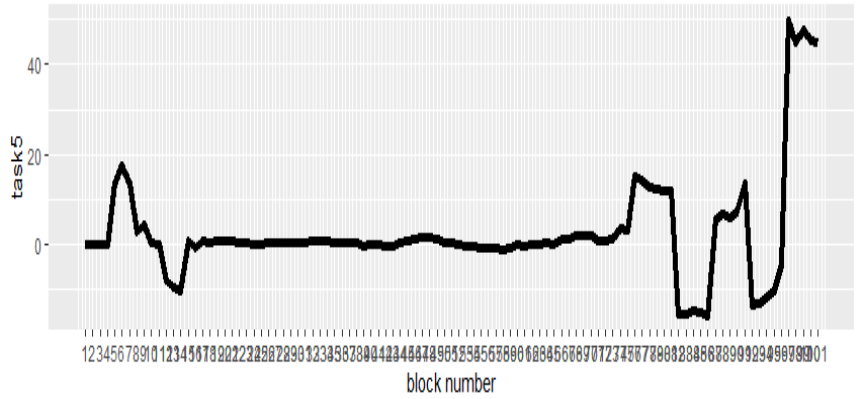

intercept
25
0
-25
-50
block number

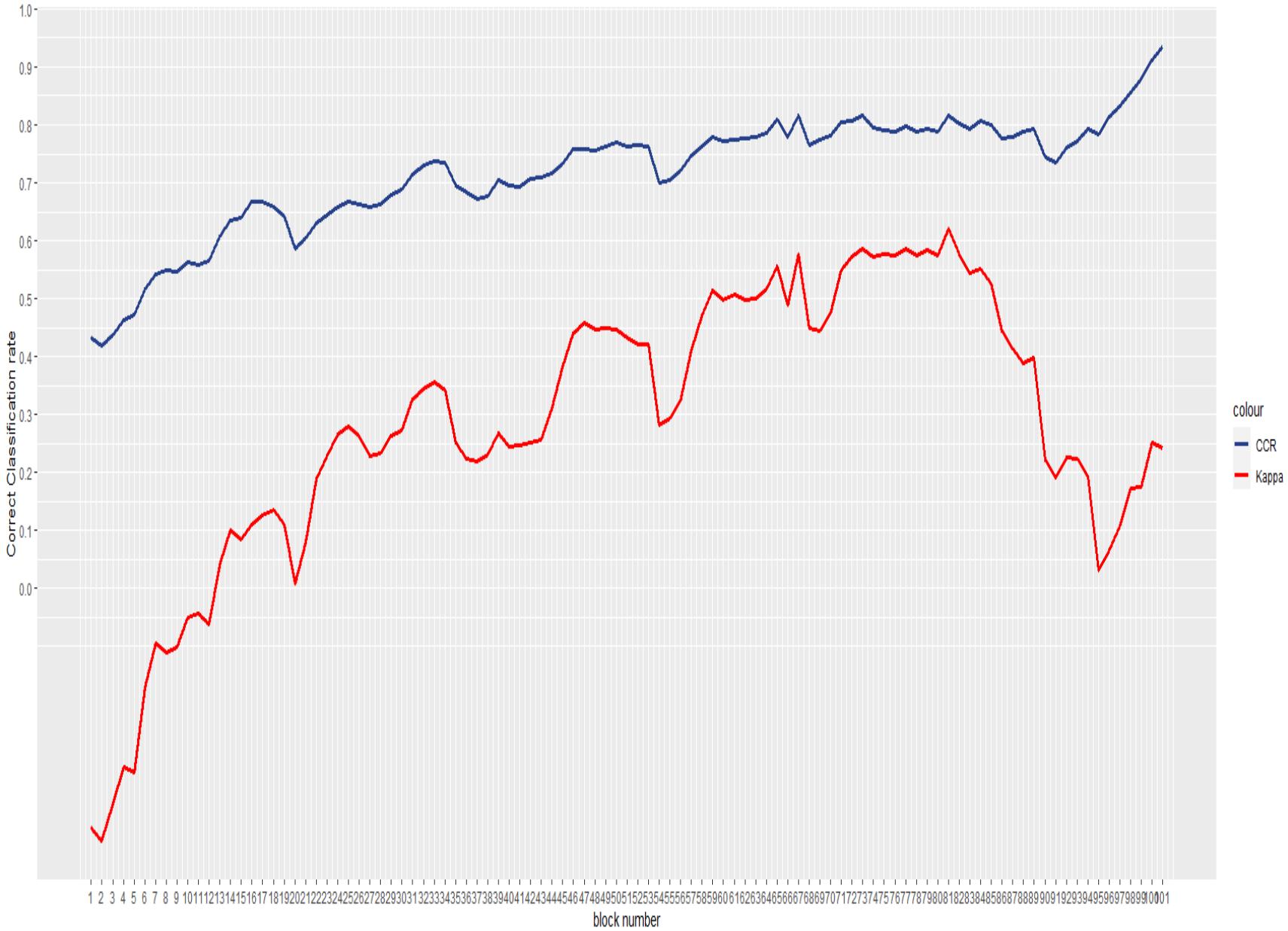

task1
20
10
0
block number


task2
0
-5
-10
-15
block number


task3
15
10
5
0
-5
block number


task4
10
0
-10
block number


task5
40
20
0
block number


Correct Classification rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
block number
colour
CCR
Kappa

What's Next?

1. Look at change points detection: univariate and multivariate
2. Outliers detection and effect of removal on increase robustness.
3. Look at the effect of changing the probability threshold of the logistic regression on the correct classification rate.

1. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

2. Nelson, B. (2013). Foundations and methods of stochastic simulation: a first course. Springer Science and Business Media.

3. Nelson, B. L. (2016). 'Some tactical problems in digital simulation'for the next 10 years. Journal of Simulation, 10(1), 2-11.


Thank you
for
listening!