KDD Cup 2019 Humanity RL the 5th place solution

Etsuro Minami [1][2], Akira Okada[3], Michihiro Nakamura[2], Masayuki Ishikawa[2]

NS Solutions Corporation [1]

Financial Engineering Group, Inc [2]

NS Solutions USA Corporation [3]

About US

NSSOL: NS Solutions Corporation（a Nippon Steel Group company)

・Consultation on business and information system

・Planning, design, development, implementation, operation and maintenance of information system

・Development, manufacturing and sales of software and hardware

・Provision of outsourcing services using information technology

・Providing data analysis and machine learning services as well as modelling and solving combinatorial optimization problems

FEG: Financial Engineering Group (a NS Solutions Group company)

・A consulting firm in Tokyo, Japan, specialized for data mining and modelling in financial industries.

・Providing data analysis services in various industries

Award history

・FEG won the 2nd place in KDD CUP 2009

・FEG & NSSOL team won the 2nd place in KDD CUP 2015

Problem Description

Action(ITN) -> a0

Action(IRS) -> a1

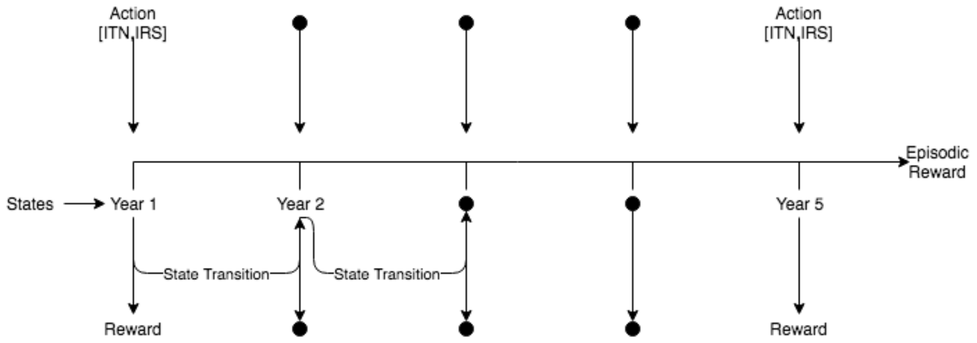
a0, a1 ∈ [0.0, 1.0]

Policy : p = $\{a_0, a_1\}$ * 5 years = $\{\{a_{00}, a_{01}\}, \{a_{10}, a_{11}\}, \{a_{20}, a_{21}\}, \{a_{30}, a_{31}\}, \{a_{40}, a_{41}\}\}$

A total of ten variables should be given respective coverage values corresponding the actions for each episode

Reward: r(p) for each episode

Objective: maximize r(p) in 20 episodes



Our Method

- Prepare a set of feature functions, F = $\{f_0, f_1, f_2, f_3 \cdots, f_n\}$, to reduce dimension, where the input of f_x , x∈{0..n}, is a set of policies, p, and the output of f_x is the value of the specified *feature* defined as a function, which is expected to reduce dimensionality.

  For example, a set of f_x consists of simple statistical functions, such as:

  the sum of $|a_0 - a_1|$ for the entire episode,

  the average of $a_0 + a_1$

  the variance of $a_0 - a_1$,

  the sum of $|a_{x0} - a_{y0}| + |a_{x1} - a_{y1}|$ where y = x − 1

  … (10 functions are defined in the submitted source code)

- Introduce a value function, g(p), which stands for ‘goodness’ of the given policy based on the feature functions above
  - g(p) is calculated from the weighted linear combination of the respective correlations between the rewards and the values of the feature functions
  - it is Higher reward could be expected for higher g(p)

$$g(p) = \sum_{x=0}^{n} \frac{f_x(p) - f_{min}(p)}{f_x(p)} \times sgn(corr) \times (e^{abs(corr)} - 1)$$

  where

  - r(p): the reward for a given policy, p.
  - g(p): ‘goodness’ of the policy, p, which is calculated from the correlation between the values of the feature functions and the actual rewards for the policies that have been already applied
  - $f_{min}(p)$: the minimum value of $f_x(p)$ for x = {0, 1, …n}

- Generate random actions for a specific year and evaluate them with g(p), pick up the best actions among them, and add the actions to the policy being generated
  - actions are evaluated incrementally, say, year by year
  - the number of trials to generate random actions increases as the number of episodes grows in order to get better g(p)
    - the entire process works like simulated annealing method (SA); explore first, exploit later
  - differentials of the feature functions are not used to optimize g(p)
    - sophisticated optimization technique to get better g(p) is not used in this solution because the available information is very limited