Multi-Objective Reinforcement Learning-based Deep Neural Networks for Cognitive Space Communications

CCAA Workshop 2017

Paulo Ferreira – *Worcester Polytechnic Institute*

Randy Paffenroth – *Worcester Polytechnic Institute*

Alexander M. Wyglinski – *Worcester Polytechnic Institute*

Timothy Hackett – *The Pennsylvania State University*

Sven Bilén – *The Pennsylvania State University*

Richard Reinhart – *NASA John H. Glenn Research Center*

Dale Mortensen – *NASA John H. Glenn Research Center*

June 28th, 2017

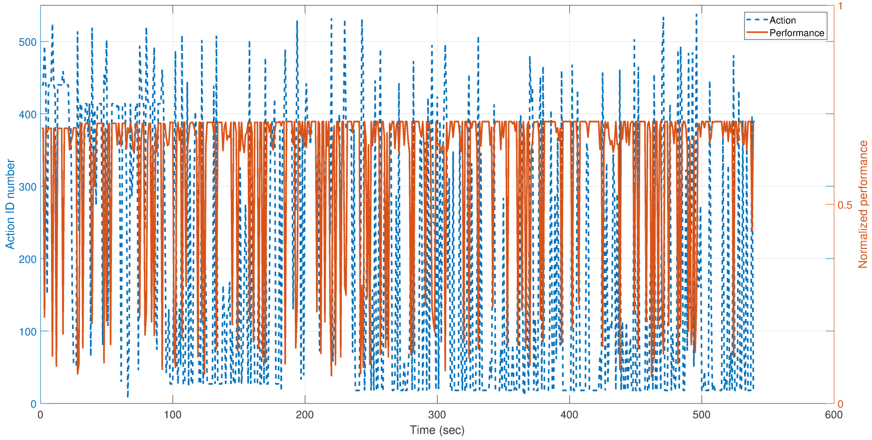
Wireless Innovation Laboratory

WPI

Acknowledgments

This work has been funded in part by the Brazilian Federal Agency CAPES through the Science without Borders scholarship program, grant number BEX 18701/12-4, and by NASA John H. Glenn Research Center, grant number NNC14AA01A.

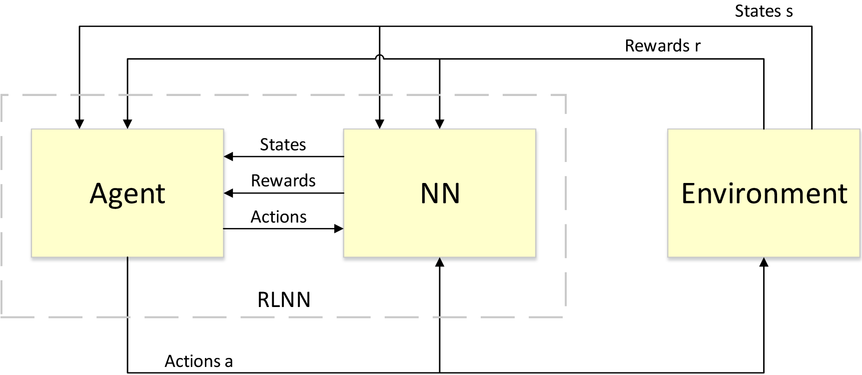
Motivation



P. V. R. Ferreira, R. Paffenroth, A. M. Wyglinski, T. M. Hackett, S. Bilén, R. Reinhart, and D. Mortensen, "Multi-Objective Reinforcement Learning for Cognitive Radio-Based Satellite Communications," in 34th AIAA International Communications Satellite Systems Conference, October 2016.

Proposed Solution

RLNN: a neural network-based reinforcement learning method



Proposed Solution

Reinforcement learning Q–function equations:

- State-Action-Reward-State-Action (SARSA)

$$Q_{k+1}(s_k, a_k) = Q_k(s_k, a_k) + \alpha[r + \gamma Q(s_{k+1}, a_{k+1}) - Q(s_k, a_k)] \quad (1)$$

- Time-Difference

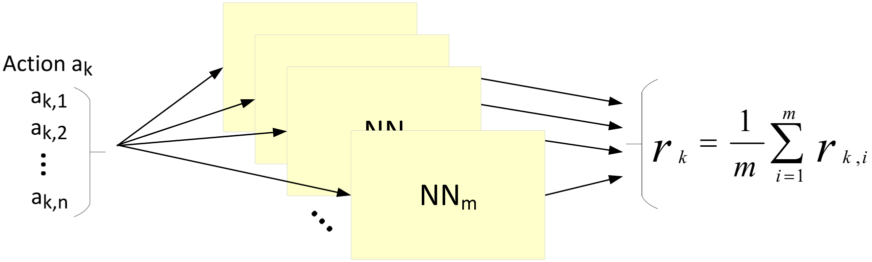
$$Q_{k+1}(s_k, a_k) = Q_k(s_k, a_k) + \alpha[r + \gamma \max_a Q_k(s_{k+1}, a) - Q_k(s_k, a_k)] \quad (2)$$

- Proposed equation for SATCOM

$$Q_{k+1}(s_k, a_k) = Q_k(s_k, a_k) + \alpha[r_k - Q_k(s_k, a_k)] \quad (3)$$

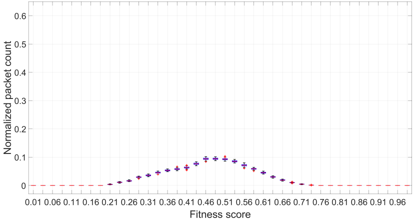
Proposed Solution

Ensemble of deep neural networks

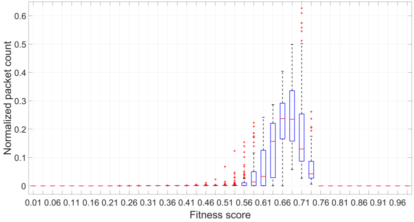


Simulation results

Exploration probability $\epsilon = 0.5$, $w_i = 1/6$



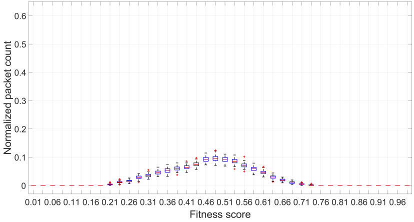
(a) Exploration OFF



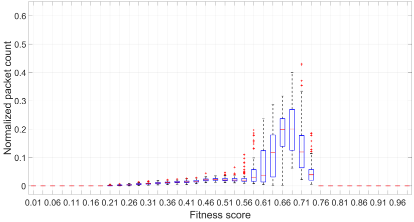
(b) Exploration ON

Simulation results

Exploration probability $\epsilon = 1/k$, $w_i = 1/6$



(a) Exploration OFF



(b) Exploration ON

Conclusions

- Hybrid ML-based multi-objective radio resource allocation – RLNN
  - Virtual exploration enables control over:
    - Performance levels while exploring actions
    - Time spent exploring very "bad" actions
- RLNN is independent of exploration probability function
- Improvements of up to $3.9\times$ on packets experiencing performance values higher than 0.55

Thank you!

Alexander Wyglinski: alexw@wpi.edu

Paulo Ferreira: paulovrf@hotmail.com

Backup

- Performance threshold
  - 95% of current maximum performance predicted by NN
- Rejection probability $= 1$

Backup

$$f_{obs}(x) = w_1 f_{\mathrm{Thrp}} + w_2 f_{\mathrm{BER}} + w_3 f_{\mathrm{BW}} + w_4 f_{\mathrm{Spc\_eff}} + w_5 f_{\mathrm{Pwr\_eff}} + w_6 f_{\mathrm{Pwr\_con}} \tag{4}$$

Throughput

$$f_{Thrp} = R_s * k * c \tag{5}$$

Bandwidth

$$f_{BW} = R_s * (1 + \beta) \tag{6}$$

Spectral efficiency

$$f_{Spc\_eff} = k * c/(1 + \beta) \tag{7}$$

Power efficiency

$$f_{Pwr\_eff} = (k * c)/((10^{(E_s/N_0)/10)}) * R_s) \tag{8}$$

Additional consumed power

$$f_{Pwr\_con} = E_s * R_s \tag{9}$$

Backup

Table 1: Adaptable parameters

Parameter	Variable	Value range
Modulation order	$\bar{M}$	$[4, 8, 16, 32]$
Bits per symbol	$\bar{k}$	$[2, 3, 4, 5]$
Encoding rate[1]	$\bar{c}$	$[1/4 - 9/10]$
Roll-off factor	$\bar{\beta}$	$[0.2, 0.3, 0.35]$
Bandwidth	$B\bar{W}$	$[0.5 - 5]$ MHz
Symbol rate	$\bar{R}_s$	$[0.41 : 0.1 : 3.7]$ MSamples/sec
Additional Tx E_s/N_0	$\bar{E}_s$	$[0 : 1 : 10]$ dB

[1] Different modulation schemes use different encoding rate sets