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
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.
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**

  \[
  Q_{k+1}(s_k, a_k) = Q_k(s_k, a_k) + \alpha [r_k - Q_k(s_k, a_k)] \quad (3)
  \]
Ensemble of deep neural networks

Proposed Solution

\[ r_k = \frac{1}{m} \sum_{i=1}^{m} r_{k,i} \]
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

\[ f_{obs}(x) = w_1 f_{Thrp} + w_2 f_{BER} + w_3 f_{BW} + w_4 f_{Spc\_eff} + w_5 f_{Pwr\_eff} + w_6 f_{Pwr\_con} \]  \hspace{1cm} (4)

**Throughput**

\[ f_{Thrp} = R_s \times k \times c \]  \hspace{1cm} (5)

**Bandwidth**

\[ f_{BW} = R_s \times (1 + \beta) \]  \hspace{1cm} (6)

**Spectral efficiency**

\[ f_{Spc\_eff} = k \times c / (1 + \beta) \]  \hspace{1cm} (7)

**Power efficiency**

\[ f_{Pwr\_eff} = (k \times c) / ((10(E_s/N_0)^{10})) \times R_s \]  \hspace{1cm} (8)

**Additional consumed power**

\[ f_{Pwr\_con} = E_s \times R_s \]  \hspace{1cm} (9)
Table 1: Adaptable parameters

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

$^1$Different modulation schemes use different encoding rate sets