

Does a Q-Learning NetLogo Extension Simplify the Development of Agent-based Simulations?



Eloísa Bazzanella

Aluna de Engenharia de Software / CEAVI
elobazzanella@gmail.com



Fernando dos Santos

Professor / CEAVI
fernando.santos@udesc.br

Introdução

- **Agentes:** autônomos em sua tomada de decisão e aptos a interação com outros agentes.
- **Simulações com agentes:** compreender comportamentos de sistemas complexos
- **Técnicas de inteligência artificial:** aperfeiçoar sua tomada de decisão
- **Q-Learning:** possibilita ao agente determinar, de maneira autônoma e interativa, uma política comportamental

Introdução

- **Extensão Q-Learning:** plataforma NetLogo
 - Oferece comandos para uso do Q-Learning
 - Trabalho de Conclusão de Curso, UDESC – CEAVI
 - Kevin Kons, 2019
- **Objetivo:** validar as melhorias proporcionadas pela extensão através de uma avaliação quantitativa
 - Simulação Cliff Walking
 - Simulação de Controle Adaptativo de Tráfego (ATSC)





Q-Learning

Q-Learning

- É um algoritmo de aprendizado por reforço
- **Objetivo:** aprender uma política, que diga a um agente que ação tomar em que circunstâncias
- **Q-Table:** Matriz de transição estado-ação
- **Elementos:** Taxa de aprendizagem, fator de desconto, recompensa, estados, ações, método de seleção da próxima ação (random-normal, e-greedy)

Q-Learning

Input: $S, A, \alpha \in (0, 1) \gamma \in (0, 1)$

foreach *episode* **do**

$s \leftarrow \text{initial_state}$

repeat

 choose an action a for state s using a selection policy (e.g., ϵ -greedy)

 perform the action a

 observe the new state s' and the reward r received

 update *Q-table*:

$$Q(a, s) \leftarrow Q(a, s) + \alpha(r + \gamma \max_a Q(a', s') - Q(a, s))$$

$s \leftarrow s'$

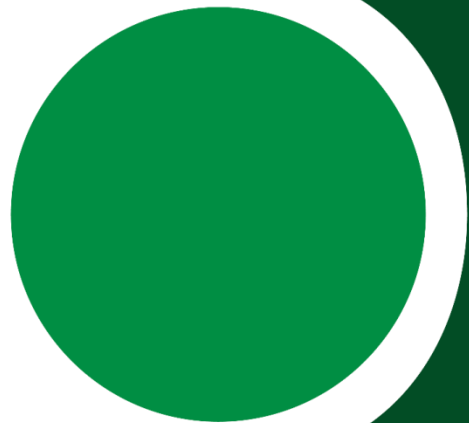
until $s \neq \text{terminal_state}$;

end

Extensão Q-Learning

- Comandos para usar Q-Learning no NetLogo

```
1  to setup
2    clear-all
3    ask Walkers [
4      qlearningextension:state-def ["xcor" "ycor"]
5      (qlearningextension:actions [goUp] [goDown] [goLeft] [goRight])
6      qlearningextension:reward [rewardFunc]
7      qlearningextension:learning-rate 0.4
8      qlearningextension:discount-factor 0.2
9      qlearningextension:action-selection "e-greedy" [0.8 0.99]
10   ]
11  end
12
13  to go
14    ask Walkers [
15      qlearningextension:act
16      qlearningextension:learn
17    ]
18  end
```



Simulações Analisadas

Simulação Cliff Walking

- **Ambiente:** Grade de células
 - Subconjunto de células representa o penhasco
- **Objetivo**
 - Sair da célula inicial **S**, chegar a célula final **G**, com o menor trajeto possível sem cair no penhasco

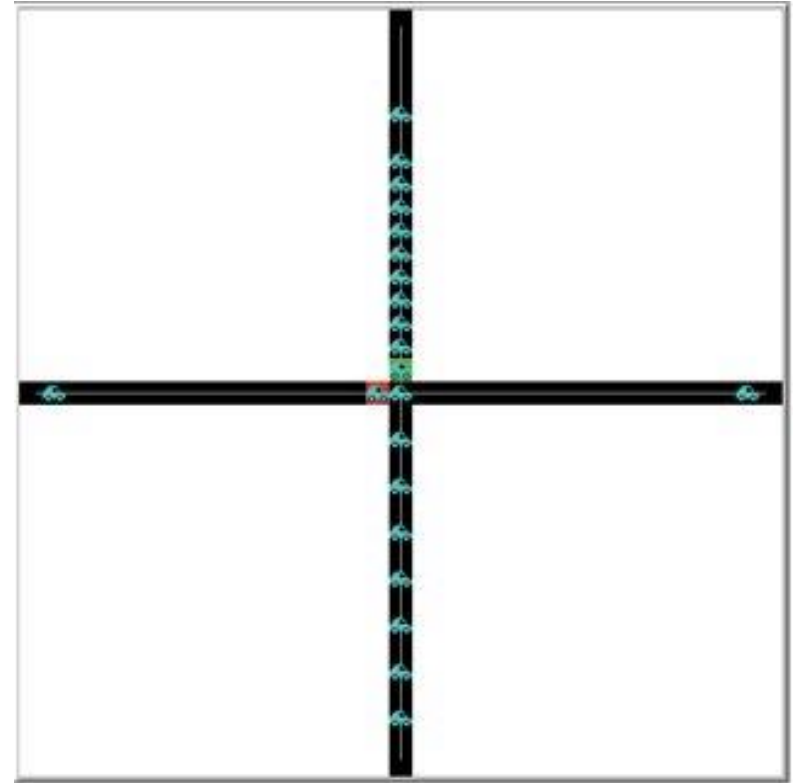
2	$r = -1$	$r = -1$	$r = -1$	$r = -1$	$r = -1$	$r = -1$
1	$r = -1$	$r = -1$	$r = -1$	$r = -1$	$r = -1$	$r = -1$
0	S $r = -1$	$r = -100$	$r = -100$	$r = -100$	$r = -100$	G $r = -1$
	0	1	2	3	4	5

Simulação de Controle Adaptativo de Tráfego (ATSC)

- **Semáforos Inteligentes**
 - Modelagem do problema *Traffic Lights Control* utilizado no artigo de Oliveira e Bazzan (2009)
- **Elementos:** a malha viária, os veículos e os semáforos

Simulação de Controle Adaptativo de Tráfego (ATSC)

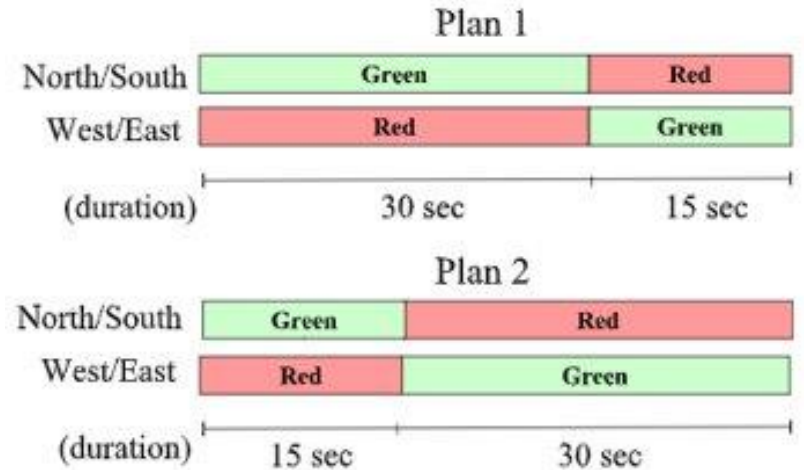
- **Malha Viária:** duas vias:
 - sentido norte-sul (vertical)
 - sentido leste-oeste (horizontal)
- **Veículos:** percorrem as vias vindos do sentido norte e oeste, de cima e do lado esquerdo



Simulação de Controle Adaptativo de Tráfego (ATSC)

- **Ações: Planos Semafóricos**

- Plano 1 = 30 seg. de sinal verde no sentido vertical e 15 seg. de sinal verde no sentido horizontal
- Plano 2 = 15 seg. de sinal verde no sentido vertical e 30 seg. de sinal verde no sentido horizontal



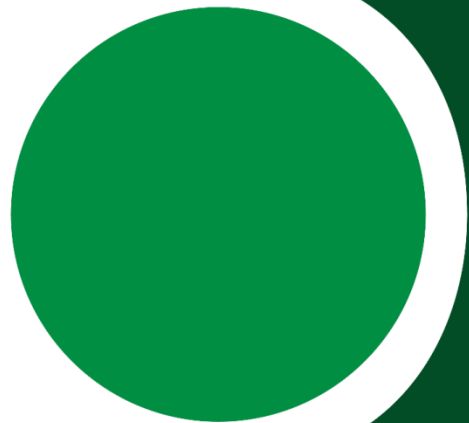
Simulação de Controle Adaptativo de Tráfego (ATSC)

- **Estados**

- quantidade de veículos parados nas vias norte e oeste

- **Recompensa**

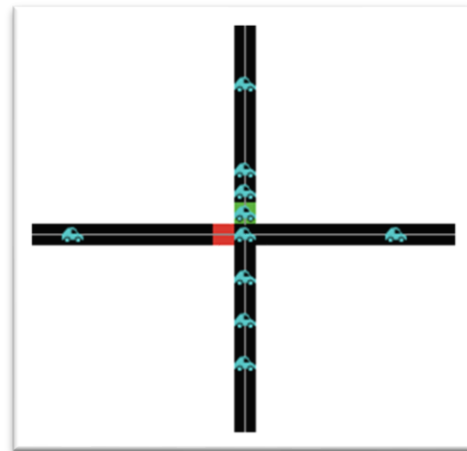
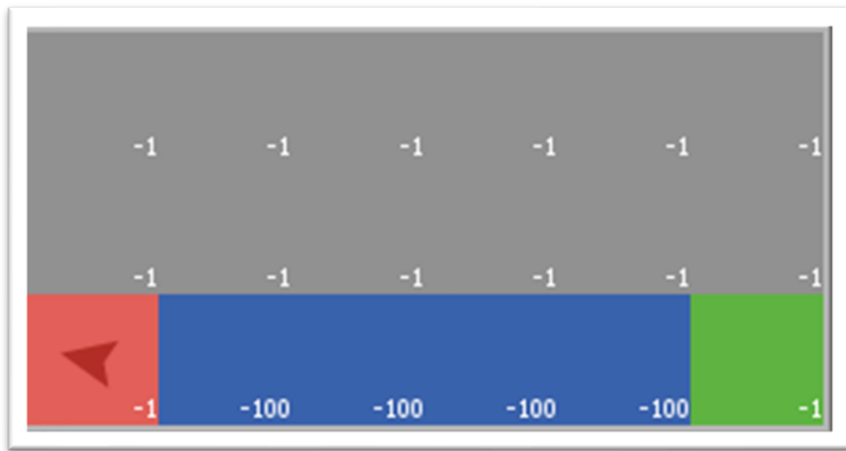
- recebe ao final de cada ciclo
- quantidade de carros parados em ambas as vias multiplicado por -1



Avaliação Quantitativa

Implementações Utilizadas

- **Duas versões das simulações**
 - Utilizando a extensão vs. codificada manualmente
 - Em ambas foram utilizados os mesmos parâmetros de aprendizagem e de ambiente para comparação de resultados



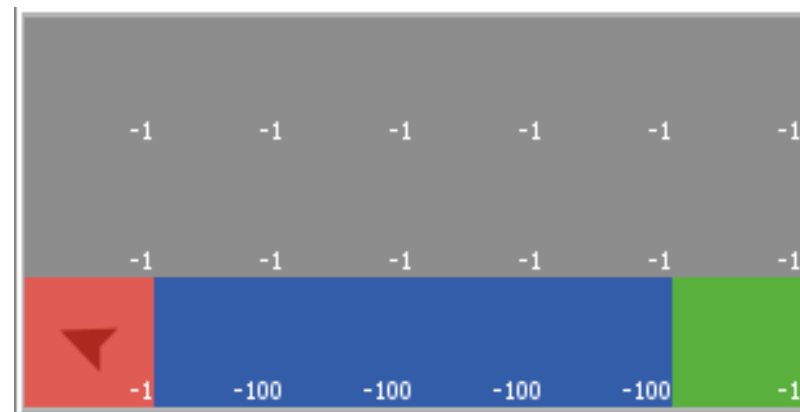
Análise de Consistência

- Verificar se a aprendizagem usando a extensão é consistente:
 - Inspeção da política aprendida após execução das simulações.
 - Uso dos mesmos parâmetros e tempo de execução.
- **ATSC:** 5 execuções de 86400 ticks
- **Cliff Walking:** 5 execuções de 350 ticks.

Análise de Consistência

Cliff Walking

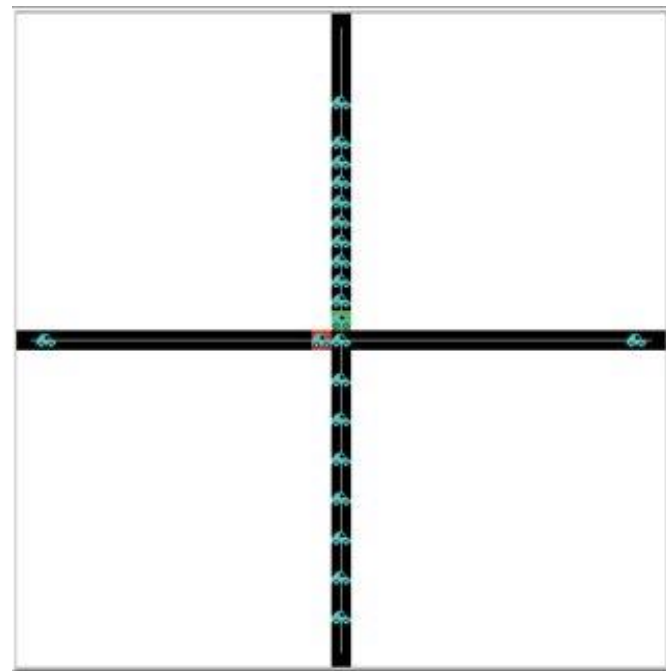
Estados (x,y)	Ações			
	<i>Direita</i>	<i>Baixo</i>	<i>Esquerda</i>	<i>Cima</i>
(0, 0)	-99,9999	-1,4281	-1,4282	-1,4277
(0, 1)	-1,4259	-1,4280	-1,4272	-1,4267
(0, 2)	-1,4248	-1,4255	-1,4254	-1,4252
(1, 1)	-1,4207	-99,8333	-1,4253	-1,4235
(1, 2)	-1,4209	-1,4217	-1,4237	-1,4214
(2, 1)	-1,4075	-99,5073	-1,4138	-1,4109
(2, 2)	-1,4123	-1,4134	-1,4174	-1,4146
(3, 1)	-1,3727	-96,9192	-1,3873	-1,3869
(3, 2)	-1,3931	-1,3941	-1,4004	-1,3991
(4, 1)	-1,2776	-91,2390	-1,3335	-1,3335
(4, 2)	-1,3505	-1,3535	-1,3697	-1,3548
(5, 1)	-1,1439	-0,9911	-1,1997	-1,2276
(5, 2)	-1,3006	-1,2678	-1,3006	-1,2976



Análise de Consistência

ATSC

Estados			Estados		
Ações			Ações		
<i>veículos parados</i>			<i>veículos parados</i>		
<i>(norte, oeste)</i>	<i>Plano 1</i>	<i>Plano 2</i>	<i>(norte, oeste)</i>	<i>Plano 1</i>	<i>Plano 2</i>
(4, 2)	-9,8587	-13,9570	(10, 2)	-9,6641	-16,1368
(5, 2)	-9,5479	-13,5261	(11, 1)	-2,5743	-4,7753
(6, 2)	-8,6826	-10,6698	(11, 2)	-9,8594	-18,6116
(7, 2)	-1,1338	-1,7662	(12, 1)	-2,1176	-5,0304
(8, 2)	-4,2193	-7,3978	(13, 1)	-7,9200	-11,5943
(9, 1)	-8,8922	-18,0331	(14, 1)	-13,7738	-21,0598
(9, 2)	-9,4045	-15,6132	(15, 1)	-14,2968	-19,0880
(10, 1)	-9,8652	-19,4982	(16, 1)	-15,6289	-21,1684



Avaliação Quantitativa: Resultados

- **Métrica:** Linhas de Código (LoC) - Mensurar o tamanho do sistema e a simplicidade para desenvolver simulações

Implementação	Tipos e Variáveis	Inicialização	Execução	Total
Controle Adaptativo de Tráfego (ATSC)				
Manualmente	27	23	119	170
Utilizando a Extensão Q-Learning	9	30	44	97
Cliff Walking				
Manualmente	15	30	51	96
Utilizando a Extensão Q-Learning	9	30	44	83

Avaliação Quantitativa: Resultados

- Redução de 42.94% de linhas de código na simulação **ATSC**, ao utilizar extensão
- Redução de 13.54% de linhas de código na simulação **Cliff Walking**, ao utilizar extensão

Implementação	Total
Controle Adaptativo de Tráfego (ATSC)	
Manualmente	170
Utilizando a Extensão Q-Learning	97
Cliff Walking	
Manualmente	96
Utilizando a Extensão Q-Learning	83

- Vantagem de utilizar a Extensão Q-Learning no desenvolvimento de agentes inteligentes em NetLogo



Conclusão

Conclusão

- Avaliação Quantitativa da Extensão Q-Learning para NetLogo
- **Resultados:** ao utilizar a Extensão Q-Learning, houve uma redução nas linhas de código
 - Simplifica o desenvolvimento de simulações
- **Trabalhos Futuros**
 - Análise de tempo de execução e escalabilidade
 - Análise com desenvolvedores e tempo de implementação

Does a Q-Learning NetLogo Extension Simplify the Development of Agent-based Simulations?



Eloísa Bazzanella

Aluna de Engenharia de Software / CEAVI
elobazzanella@gmail.com



Fernando dos Santos

Professor / CEAVI
fernando.santos@udesc.br