

Refatoração da Extensão NetLogo de Aprendizagem por Reforço para Integração com a Biblioteca BURLAP

Eloísa Bazzanella, Matheus M. Barros, Fernando Santos

Universidade do Estado de Santa Catarina (UDESC)

{elobazzanella, matheusmbarros01}@gmail.com, fernando.santos@gmail.com



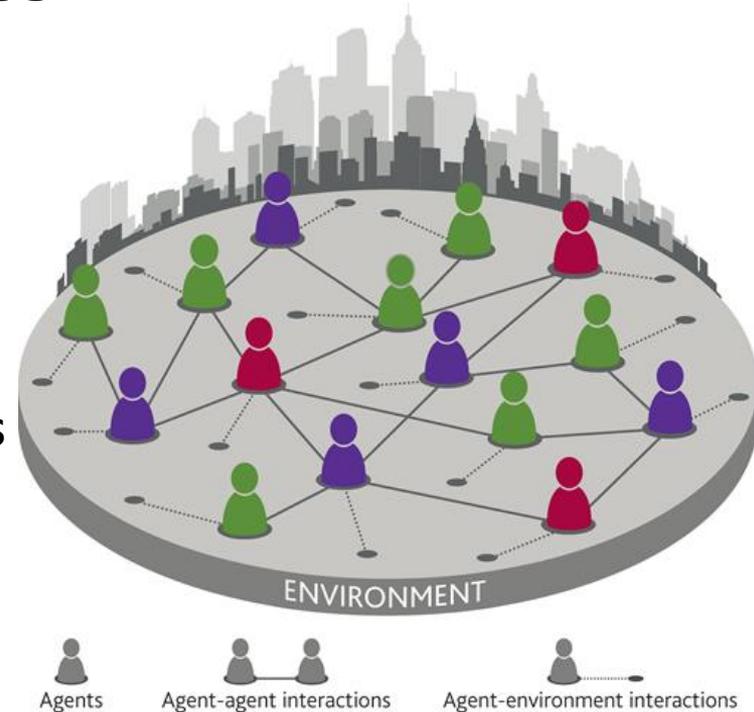
18th Workshop-School on Agents, Environments and Applications
Brasilia, DF, Brazil

August 14 - 16, 2024

Introdução

Simulações baseadas em agentes (SBAs)

- Reproduzir e investigar fenômenos
- **Agentes**
 - Entidades autônomas e heterogêneas
- **Ambiente**
 - Espaço onde os agente estão situados
- **Interações**
 - Contatos e relações existentes



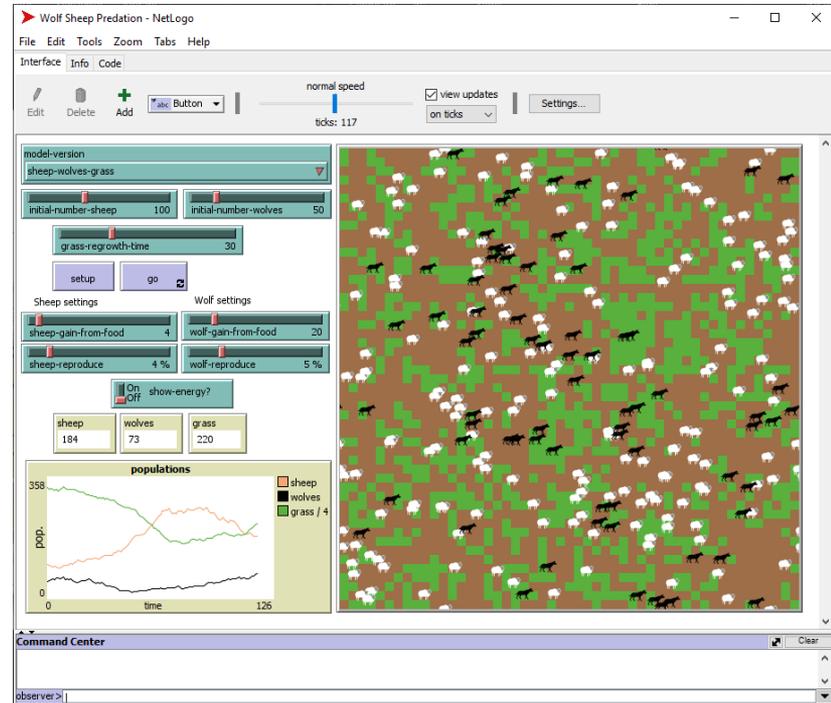
<https://www.bankofengland.co.uk/quarterly-bulletin/2016/q4/agent-based-models-understanding-the-economy-from-the-bottom-up>

SBA: como desenvolver?

- Ambientes **específicos** facilitam o desenvolvimento
- **NetLogo**: ambiente popular

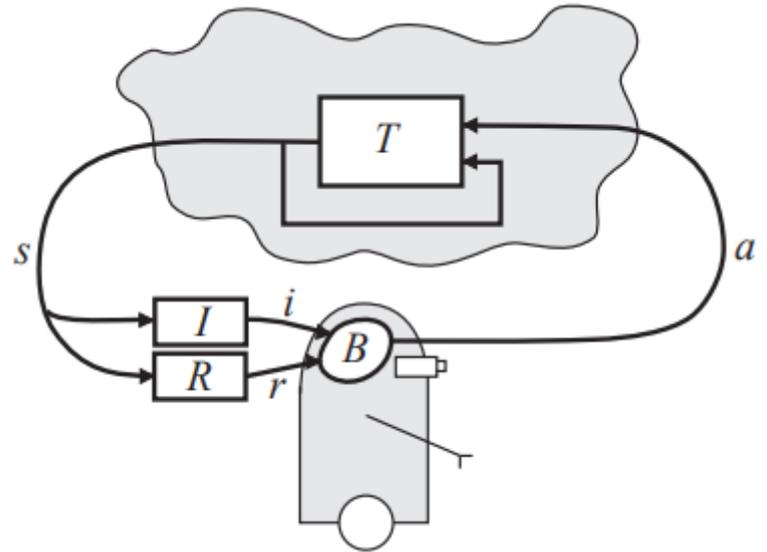


ccl.northwestern.edu/netlogo



SBAs: agentes inteligentes

- **Técnicas de IA** aprimoram a tomada de decisão dos agentes
- **Aprendizagem por reforço**
- **Implementar** a aprendizagem pode ser uma barreira



Extensão Q-Learning

- Extensão para o NetLogo com **comandos** para utilizar Q-Learning
 - Kons, K. (2019). *Biblioteca Q-Learning para desenvolvimento de simulações com agentes na plataforma NetLogo*. Trabalho de conclusão de curso, UDESC
- **Limitação:** apenas Q-Learning

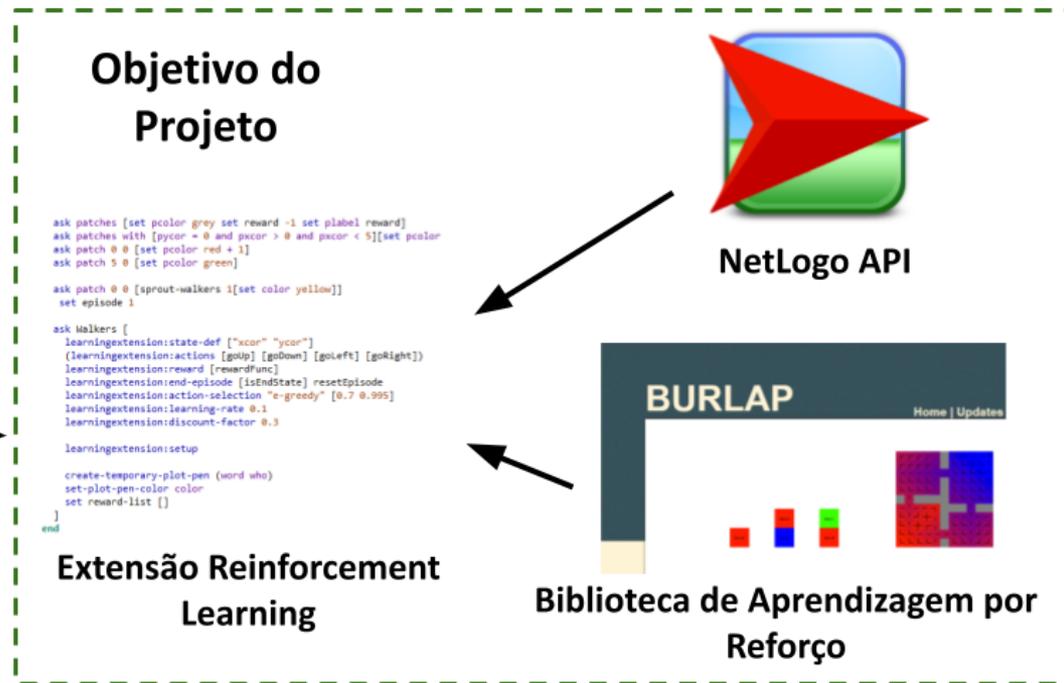
```
to setup
  clear-all
  ask Walkers [
    qlearningextension:state-def ["xcor" "ycor"]
    (qlearningextension:actions [goUp] [goDown] [goLeft] [goRight])
    qlearningextension:reward [rewardFunc]
    qlearningextension:learning-rate 0.4
    qlearningextension:discount-factor 0.2
    qlearningextension:action-selection "e-greedy" [0.8 0.99]
  ]
end

to go
  ask Walkers [
    qlearningextension:act
    qlearningextension:learn
  ]
end
```



Objetivo

- Refatorar a extensão NetLogo para incorporar outros algoritmos de aprendizagem por reforço
 - SARSA (λ) e Actor-Critic



Fundamentação Teórica

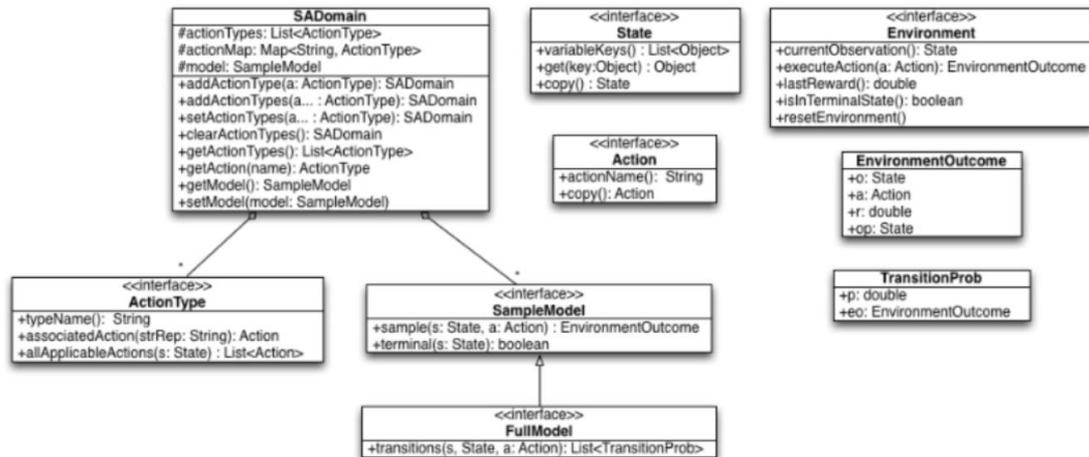
BURLAP

- **B**rown-**U**MBC **R**einforcement **L**earning and **P**lanning
- Biblioteca Java com algoritmos de aprendizagem por reforço
- Madura e amplamente testada
- <http://burlap.cs.brown.edu/>

jmacglashan/
burlap

Repository for the ongoing development of the
Brown-UMBC Reinforcement Learning And
Planning (BURLAP) java library

👤 12 Contributors 📦 184 Used by ⭐ 269 Stars 🍴 164 Forks



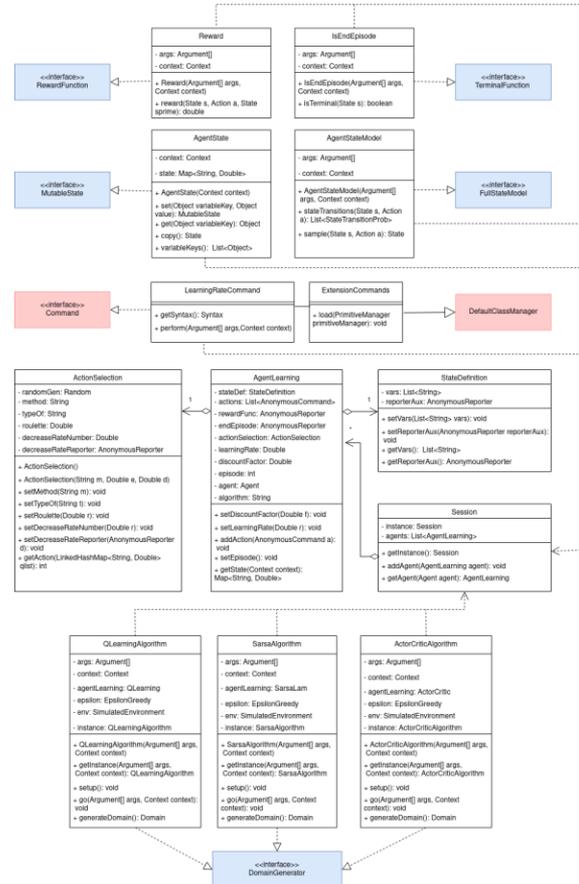
Q-Learning, SARSA (λ) e Actor-Critic

- Algoritmos de aprendizagem por reforço
- **Q-Learning**
 - *off-policy*: sem política pré-definida; livre para experimentar
- **SARSA (λ)**
 - *on-policy*: experiência com base em uma política pré-definida
 - λ : traço de elegibilidade e fator de atenuação
- **Actor-Critic**
 - *off-policy* + *on-policy*: etapas de avaliação e melhoria de política

Refatoração da Extensão

Projeto da Integração: BURLAP/NetLogo

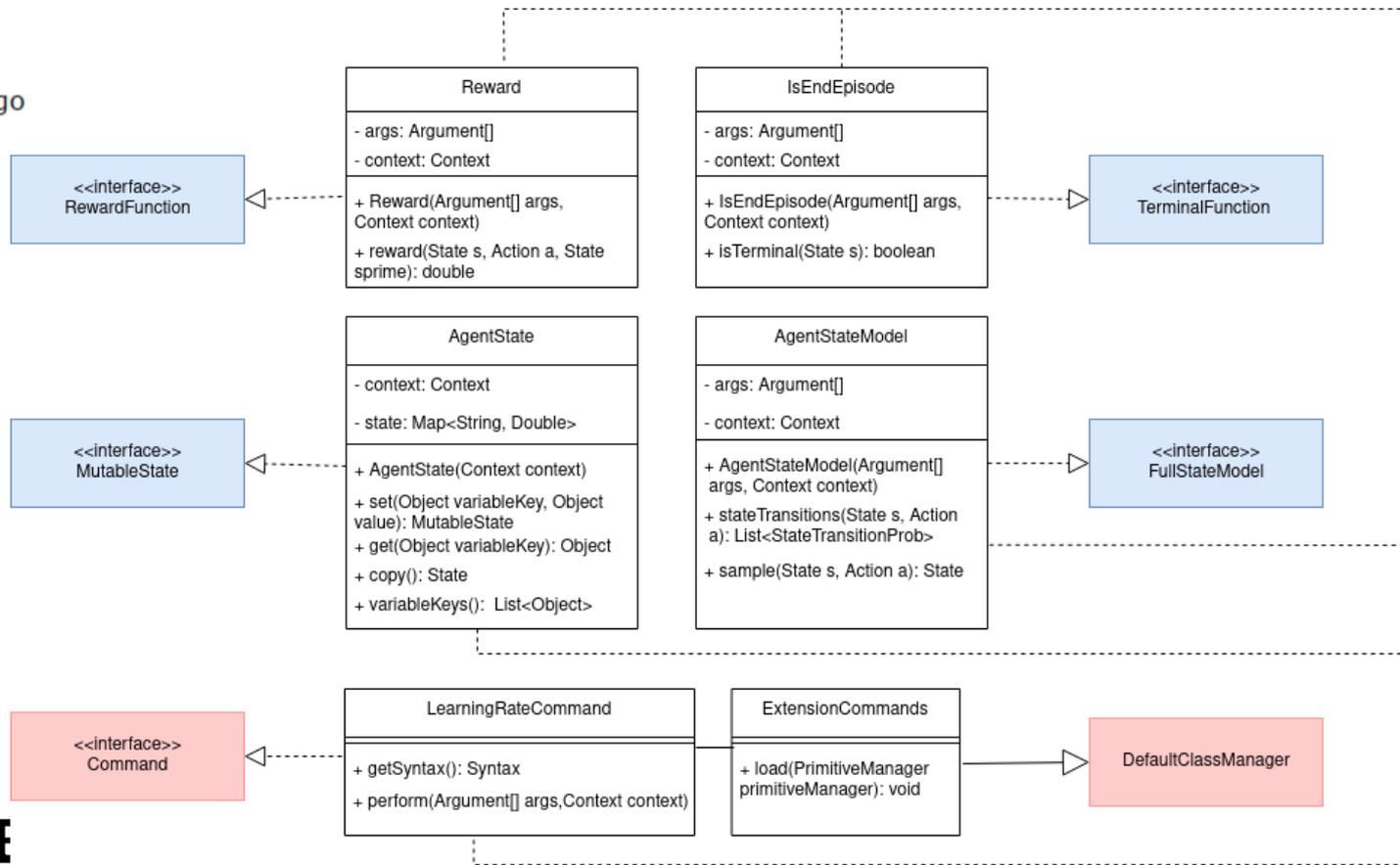
- Projeto Orientado a Objetos



Projeto da Integração: BURLAP/NetLogo

BURLAP

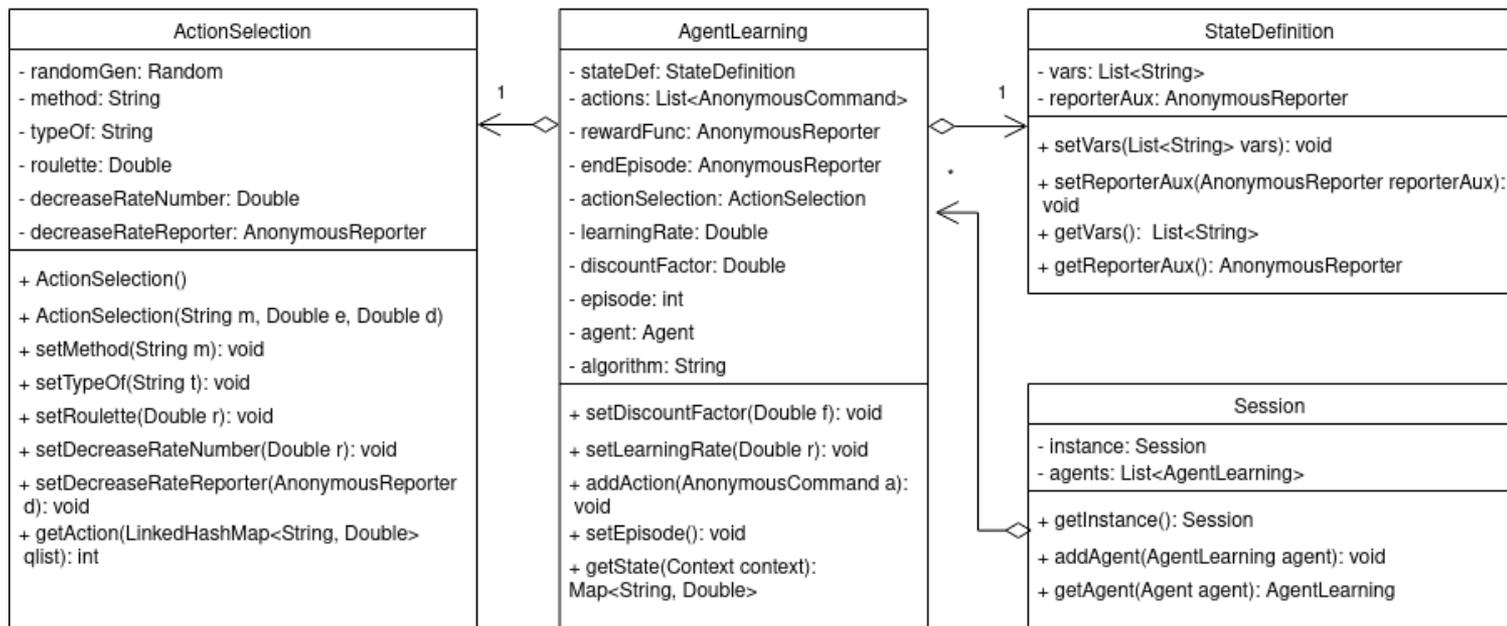
API NetLogo



Projeto da Integração: BURLAP/NetLogo

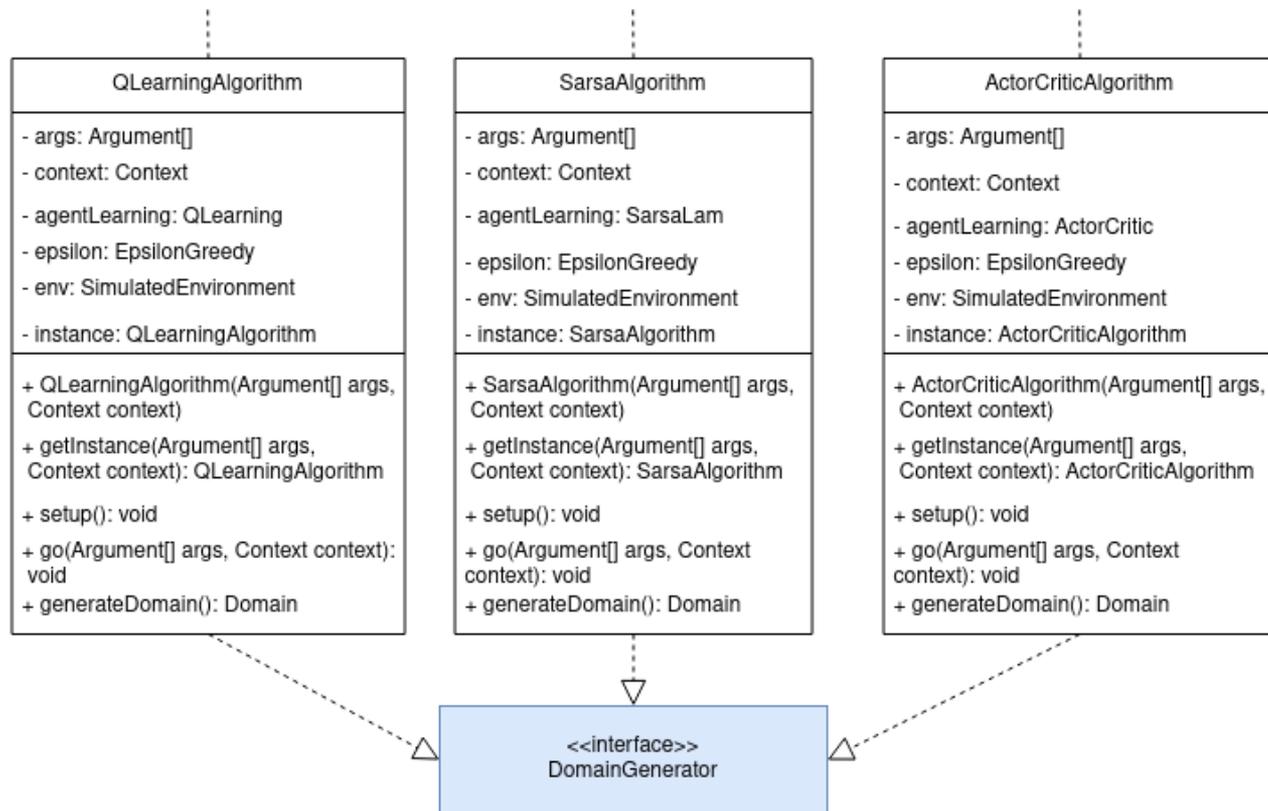
BURLAP

API NetLogo



Projeto da Integração: BURLAP/NetLogo

- BURLAP
- API NetLogo



Extensão Refatorada: exemplo

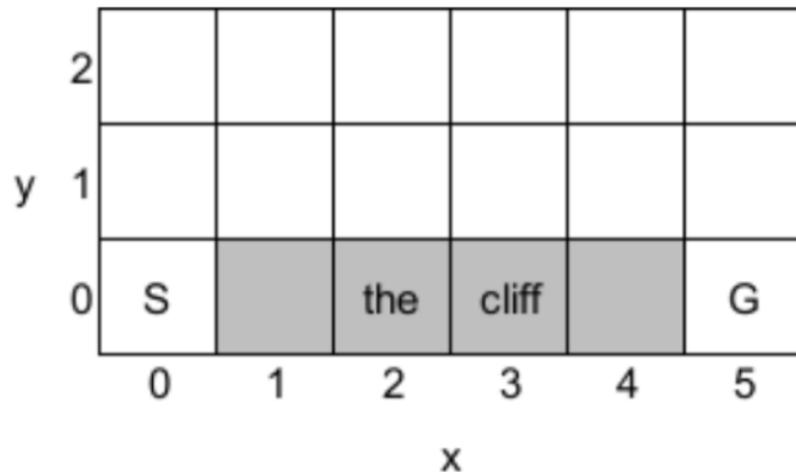
```
to setup
  ask turtles [
    learningextension:state-def ["xcor" "ycor"]
    (learningextension:actions [goUp] [goDown] [goLeft] [goRight])
    learningextension:reward [rewardFunc]
    learningextension:end-episode [isEndState] resetEpisode
    learningextension:action-selection "e-greedy" [0.7 0.995]
    learningextension:learning-rate 0.5
    learningextension:discount-factor 0.5
    learningextension:lambda 1
    learningextension:define-algorithm "sarsa-lambda"
    learningextension:setup
  ]
end
```

```
to go
  ask turtles [
    learningextension:learning
  ]
end
```

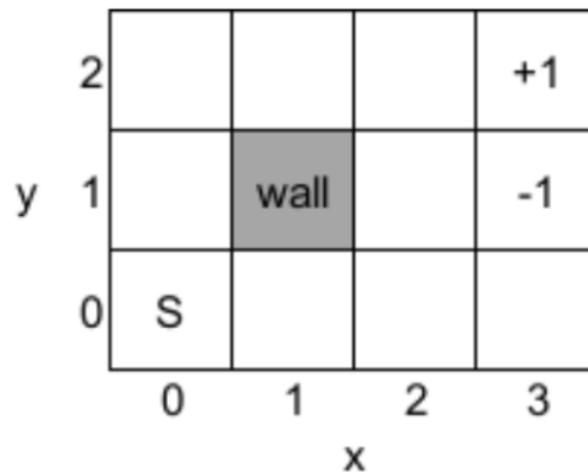
Avaliação da Refatoração

Avaliação da Refatoração

- Implementação de duas SBAs
- Verificar as políticas aprendidas
 - Detalhes da modelagem e parâmetros dos algoritmos descritos no artigo
 - 5 execuções por algoritmo



(a) Cliff Walking



(b) Mundo 4x3

Resultados: Cliff Walking

Estado (x,y)	Q-Learning				SARSA(λ)				Actor-Critic Utilidade
	Cima	Baixo	Esquerda	Direita	Cima	Baixo	Esquerda	Direita	
(0,0)	-1,42826	-1,42837	-1,42833	-99,88934	-1,99636	-12,05645	-5,23845	-100,00000	-34,44597
(0,1)	-1,42772	-1,42802	-1,42790	-1,42753	-1,99270	-2,51677	-2,24678	-15,01358	-8,00000
(0,2)	-1,42694	-1,42694	-1,42690	-1,42684	-2,30813	-2,32880	-2,01280	-1,98531	-7,00000
(1,1)	-1,42611	-98,51834	-1,42668	-1,42510	-1,76315	-100,0000	-7,02749	-2,57476	-36,99150
(1,2)	-1,42530	-1,42520	-1,42526	-1,42512	-3,20844	-8,05625	-2,33142	-1,98297	-6,00000
(2,1)	-1,42172	-96,09352	-1,42283	-1,41700	-1,98662	-99,99982	-12,74462	-19,45179	-47,34018
(2,2)	-1,42094	-1,42080	-1,42122	-1,42067	-2,77747	-3,91706	-2,12292	-1,94073	-5,00000
(3,1)	-1,40693	-91,22813	-1,40938	-1,39000	-1,97260	-98,12012	-9,42074	-22,32461	-37,33405
(3,2)	-1,40961	-1,40914	-1,40976	-1,40914	-2,18766	-11,85778	-2,09680	-1,88126	-4,00000
(4,1)	-1,36837	-88,26094	-1,37723	-1,30000	-1,90400	-99,32617	-10,46501	-1,50151	-34,25671
(4,2)	-1,38097	-1,37830	-1,38035	-1,37827	-1,93889	-9,60968	-1,96275	-1,76251	-3,00000
(5,1)	-1,26880	-1,00000	-1,28172	-1,19679	-1,78433	-1,00000	-7,38230	-1,50117	-1,00000
(5,2)	-1,30143	-1,29590	-1,30105	-1,30987	-1,96631	-1,50002	-1,90512	-1,80811	-2,00000

Tabela 1. Resultados na SBA *Cliff Walking*



Resultados: Mundo 4x3

Estado (x,y)	<i>Q-Learning</i>				<i>SARSA(λ)</i>				<i>Actor-Critic</i> Utilidade
	Cima	Baixo	Esquerda	Direita	Cima	Baixo	Esquerda	Direita	
(0,0)	0.840	0.800	0.800	0.840	0.790	0.631	0.556	0.359	0,840
(0,1)	0.880	0.800	0.840	0.840	0.841	0.526	0.683	0.644	0,488
(0,2)	0.880	0.840	0.880	0.920	0.541	0.589	0.670	0.910	0,952
(1,0)	0.840	0.840	0.800	0.880	-0.078	-0.434	0.691	-0.001	0,488
(1,2)	0.920	0.920	0.880	0.960	0.794	0.777	0.662	0.952	0,976
(2,0)	0.920	0.880	0.840	0.840	-0.812	-0.252	0.652	-0.486	-0,256
(2,1)	0.960	0.880	0.920	-1.000	0.925	0.297	0.184	-0.800	0,208
(2,2)	0.960	0.920	0.920	1.000	0.960	0.865	0.821	1.000	1,000
(3,0)	-1.000	0.840	0.880	0.840	-1.000	-0.424	0.298	-0.305	-0,216

Tabela 2. Resultados na SBA *Mundo 4x3*



Conclusão

Conclusão

- Refatoração da extensão NetLogo com BURLAP bem sucedida
- Sem inconsistências introduzidas na extensão
- Disponibilidade de 3 algoritmos de aprendizagem:
 - Q-Learning, SARSA (λ) e Actor-Critic
- **Trabalhos Futuros**
 - Avaliar em outras SBAs
 - Incorporar na extensão mais algoritmos da BURLAP

Refatoração da Extensão NetLogo de Aprendizagem por Reforço para Integração com a Biblioteca BURLAP

Eloísa Bazzanella, Matheus M. Barros, Fernando Santos

Universidade do Estado de Santa Catarina (UDESC)

{elobazzanella, matheusmbarros01}@gmail.com, fernando.santos@gmail.com



18th Workshop-School on Agents, Environments and Applications
Brasilia, DF, Brazil

August 14 - 16, 2024