

O impacto da duração do episódio de avaliação na robótica evolutiva: uma análise qualitativa

Larissa Gremelmaier Rosa

Graduanda em Ciência da Computação
Universidade Federal de Santa
Catarina
Florianópolis, Brasil
larissa.gremelmaier.rosa@grad.ufsc.br

Vitor Hugo Homem Marzarotto

Graduando em Ciência da Computação
Universidade Federal de Santa
Catarina
Florianópolis, Brasil
vitor.h.m@grad.ufsc.br

Jônata Tyska Carvalho

Departamento de Informática e Estatística
Universidade Federal de Santa
Catarina
Florianópolis, Brasil
jonata.tyska@ufsc.br

Abstract—In evolutionary robotics, the combination of genetic algorithms, neural networks, and mobile robots is widespread. Recent studies revealed that the length of the agent evaluation episode and the manner in which it is varied during evolution directly affects the evolution performance and the behavior generated. The length of the episode influences both the accuracy of the fitness function as well as the quality of the generated behaviors. However, more investigation is still needed to better comprehend the full impact of different episode duration. The goal of this paper is to analyze and characterize the behaviors generated in robots evolved under varying the length of the evaluation episode. The agents all aim at solving locomotion problems. Results showed that reducing the episode length helps to develop initial behaviors that may perform fairly well, but are generally not fully developed. Longer evolutionary episodes tend to produce more static stay-at-a-fixed-position robotic behaviors, preventing further development. On the other hand, experiments that gradually increase episode duration over the course of evolution avoid the optimal static standing location, allowing more efficient behaviors to develop at the end of evolution.

Key Words—Artificial intelligence, Adaptive robotics, Evolutionary algorithms

I. INTRODUÇÃO

O desenvolvimento de robôs autônomos capazes de executar tarefas altamente complexas em ambientes dinâmicos e desconhecidos, com um controle externo relativamente baixo, é um objetivo altamente almejado. Tais robôs são muito úteis quando os dados da tarefa precisam de processamento em tempo real, ou quando há a necessidade de interação entre os agentes de um sistema de maneira simultânea [1].

Esses agentes autônomos capazes de apresentar comportamento adaptativo têm aplicações na indústria, nos serviços, dentre outras áreas [2]. Por exemplo, podem ser utilizados para substituir seres humanos em tarefas de alto risco, para realizar tarefas domésticas e muitas outras. No entanto, apesar da vasta gama de aplicações, a robótica autônoma e a síntese automática de comportamento adaptativo ainda enfrentam desafios no que diz respeito à sua concepção, construção e implementação [3]. Assim, o que é habitualmente visto são robôs capazes de operar dentro de um conjunto limitado de ações e condições ambientais, e com uma sequência de tarefas pré-determinadas, ou ainda, robôs parcialmente controlados por humanos [3].

Portanto, para avançar na área e ser capaz de criar agentes autônomos úteis, é necessário aperfeiçoar e desenvolver métodos que sejam capazes de superar estes desafios. Entre estes métodos, destaca-se a robótica evolutiva. A este respeito, a robótica evolutiva permite sintetizar robôs, morfologias e comportamentos, para realizar tarefas sem ter de especificar as estratégias que devem ser executadas para as cumprir [4] [5].

O campo utiliza meta-heurísticas de inspiração biológica que, normalmente, acrescentam iterativamente pequenas variações aos parâmetros de controle e, por vezes, às morfologias dos robôs, avaliam as soluções candidatas e selecionam as de melhor desempenho [3]. O processo continua até se atingir algum critério de parada, e espera-se que tenha sido encontrada uma solução satisfatória no contexto da aplicação em questão. Uma característica importante dos algoritmos evolutivos é que eles demonstram uma elevada robustez e a possibilidade de paralelizar o processo de otimização [6], tornando-o veloz quando há grande poder computacional disponível.

O processo de avaliação, que envolve a maximização da chamada função fitness, embora mais simples quando comparado com a definição manual dos comportamentos do agente, ainda apresenta os seus desafios [7]. É sabido que este processo é complexo em razão de que pequenas alterações em elementos como os parâmetros iniciais do algoritmo, os critérios de terminação, o estado inicial do agente e as características do ambiente podem conduzir a diferentes resultados que podem ter um impacto positivo ou negativo sobre o objetivo final do agente em questão [8] [9] [10].

O papel da variação ambiental para o desenvolvimento de soluções robustas, por exemplo, vem recebendo atenção na literatura nos últimos anos [8] [10] [11] [12]. Por exemplo, [8] testa diferentes formas de variar o ambiente no decurso do processo evolutivo cujo propósito é desenvolver um comportamento robótico adaptativo, procurando verificar a robustez dos agentes que evoluem sob estas condições. Para isso, evoluíram uma população de agentes usando um algoritmo de estado da arte [13]. A tarefa utilizada foi o problema do equilíbrio do Double-pole [14] em várias condições, tais como sem qualquer variação ambiental ou com diferentes níveis de variação ambiental e frequência de mudança. Com

base nos resultados obtidos, os autores concluíram que a variação ambiental é necessária e benéfica, uma vez que todas as experiências que continham variação tiveram um melhor desempenho do que as que não continham qualquer variação. Além disso, o estudo concluiu que existe uma quantidade ideal de variação ambiental que maximiza o desempenho do algoritmo, o que significa manter esse ambiente fixo durante um determinado número de gerações antes de o variar. Isto levou a uma combinação adequada de variação ambiental e estabilidade.

Recentemente, foi demonstrado em [15], que também a duração do episódio de avaliação e a forma como a duração dos episódios varia durante o curso do processo evolutivo poderia modificar drasticamente a velocidade e o resultado final da evolução. Contudo, nas análises apresentadas no estudo, foi considerada apenas o impacto na aptidão (*fitness*) dos agentes, ou seja, a capacidade de locomoção alcançada por eles em termos quantitativos. No presente trabalho, examinamos aspectos qualitativos dos comportamentos evoluídos a fim de compreender como são gerados comportamentos mais eficazes por meio de uma manipulação adequada da duração do episódio de avaliação. Assim, é possível utilizar uma análise qualitativa para caracterizar os comportamentos nas diferentes configurações experimentais apresentadas em [15]. Além disso, comparamos os comportamentos evoluídos finais em relação à distância percorrida nas tarefas de locomoção robótica com os valores de performance alcançados ao final do processo evolutivo. Os resultados demonstram que o tamanho padrão do episódio de avaliação mostra tendências de movimento estático, enquanto que as abordagens incrementais são mais susceptíveis de resultar em movimento contínuo ao longo do episódio. Além disso, percebe-se que a função *fitness*, apesar de em geral favorecer movimentos contínuos, pode melhor pontuar movimentos estáticos a depender da duração do episódio avaliativo.

O presente trabalho está organizado da seguinte forma: na seção 2, apresenta-se os métodos utilizados e critérios de avaliação definidos e utilizados para a análise proposta. Na seção 3, apresentam-se os resultados encontrados com os experimentos mencionados na segunda seção, discutindo seus aspectos relevantes. Na seção 4, são apresentadas as conclusões do trabalho, apontando trabalhos futuros e possíveis melhorias no processo avaliativo desenvolvido.

II. MÉTODO

Os experimentos foram definidos conforme os resultados obtidos por [15], que investigou as diferentes formas de variar a duração do episódio avaliativo dos robôs evolutivos. Assim, utiliza-se os problemas de locomoção do Pybullet [16] para treinar os agentes na habilidade de se locomover da maneira mais rápida possível. Cada agente tem uma morfologia única e diferentes números de juntas e segmentos, formando seus próprios desafios comportamentais.

Como os resultados da tarefa Humanoid apresentado em trabalho recente avaliando o impacto da duração dos episódios de avaliação [15] foram inconclusivos, visto que são necessários

maiores estudos com o agente, no presente trabalho, considera-se apenas os problemas Hopper, Ant e Walker2D, apresentados na Figura 1.

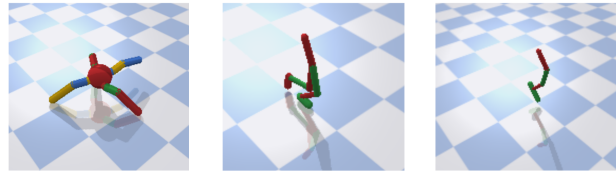


Fig. 1: A Ant, o Hopper e o Walker2D, respectivamente da esquerda para direita

Na tabela I, é possível ver o número de juntas, a morfologia e a habilidade esperada para cada um dos agentes supramencionados.

Tabela I
Número de juntas e habilidade de cada agente

Agentes	Número de juntas	Habilidade
Hopper	3	Pular com uma perna
Walker2D	6	Caminhar com duas pernas
Ant	8	Caminhar com quatro pernas

Para realizar os experimentos, foi utilizado o simulador Evorobotpy2 executando o algoritmo evolutivo OpenAI-ES [6]. A função de *fitness* utilizada para avaliar os indivíduos é definida como a soma de uma lista de componentes importantes para o movimento dos robôs: um bônus de vida, isto é, uma pontuação por permanecer levantado em equilíbrio, um bônus dado aos robôs por cada ação que fazem sem cair, o progresso que fizeram em etapas, e um componente negativo que contabiliza os custos do movimento. Os custos do movimento podem ser divididos em três: custos de electricidade, juntas a custos limite e custos de colisão dos pés. Esta função de aptidão é a implementação padrão nos ambientes de locomoção PyBullet, o que facilita uma maior comparação entre experiências semelhantes. Os demais parâmetros utilizados na montagem dos experimentos podem ser consultados em [15]. Para comparação, também utilizamos os experimentos cuja função não utiliza o bônus de vida, que de acordo com [17], funciona melhor para algoritmos evolutivos.

Durante a evolução, por padrão, a duração dos episódios de avaliação é definida para 1000 passos de simulação, em que a unidade "passos" é medida em passos de tempo discretos. As diferentes configurações relativas à duração da avaliação dos episódios utilizados para cada configuração experimental são as seguintes (N é uma variável medida em passos):

- 1) Fixo-1000: Uma condição de base na qual a duração dos episódios de avaliação é definida para 1000 passos, de

acordo com o padrão do PyBullet.

- 2) Fixo-N: Um experimento em que a duração permanece constante durante a evolução mas com valores diferentes do padrão. Foram testados valores de 100, 200, 300, 400 e 500.
- 3) Fixo-N-1000: Um experimento no qual 50 % do processo evolutivo é feito com um dos comprimentos listados no item 2 e os outros 50 % nos 1000 passos padrão
- 4) Incremental: Um experimento em que a duração do episódio avaliativo começa em 100 passos e é aumentada em 100 passos a cada 10% da evolução, até atingir 1000 passos.
- 5) V5-controle: O experimento descrito por [17] no qual retira-se o bônus de vida da função de aptidão e mantém-se a duração do episódio em 1000 passos durante toda a evolução.

Na pesquisa descrita em [15], ainda havia um último experimento definido como uma variação aleatória do tamanho dos episódios avaliativos. Os resultados encontrados, entretanto, mostraram que a variação aleatória a cada geração adiciona estocasticidade excessiva no processo evolutivo, impedindo assim a geração dos comportamentos desejados, assim, tal configuração foi excluída da análise demonstrada nesse trabalho.

Após o processo evolutivo, o desempenho dos agentes evoluídos, ou seja, a capacidade de um agente evoluído para resolver o problema em diferentes momentos de avaliação de episódios, é determinado pela pós-avaliação dos agentes para 1000 episódios em que as características do ambiente e o estado inicial do robô são definidos aleatoriamente. A duração do episódio de avaliação é sempre definida para 1000 passos que é o valor padrão do simulador.

A fim de caracterizar os comportamentos evoluídos, introduziu-se novas métricas e análises comportamentais. Primeiro, os agentes gerados tiveram os seus comportamentos analisados visualmente de uma forma qualitativa. Os primeiros cinco episódios de pós-avaliação foram analisados e classificados em um dos três tipos de situação, como se segue:

- 1) O agente não aprende a capacidade de locomoção esperada, nem demonstra o desenvolvimento de capacidades iniciais ou intermediárias para avançar, ou seja, habilidades como ficar parado ou cair imediatamente no início dos episódios de avaliação.
- 2) O agente mostra indicadores de aprendizagem para a competência esperada, mas não a desenvolve totalmente. Exemplo: Salta e cai, dá um passo e para.
- 3) O agente desenvolve a habilidade esperada e alcança alguma constância na sua utilização.

Cada configuração de experimento foi executada 10 vezes, gerando 10 agentes, portanto a classificação final de um experimento é a classificação que mais apareceu nos seus agentes.

Juntamente com a classificação, adicionou-se uma breve descrição do comportamento verificado, de modo a que se torne possível compreender como os comportamentos diferi-

ram entre os diversos experimentos, e como esta diferença qualitativa pode levar a um melhor desempenho. Depois, comparou-se a classificação e o comportamento anotado com o seu desempenho obtido, de modo que fosse possível compreender a eficácia do desempenho na pontuação dos comportamentos finais. Finalmente, comparou-se os resultados obtidos entre os experimentos com o objetivo de encontrar correlações entre os comportamentos gerados em detrimento da forma como a variação da duração do episódio foi alterada.

III. RESULTADOS E DISCUSSÃO

A duração do episódio de avaliação, quando variado de diferentes formas, influencia os comportamentos gerados, as suas classificações médias e seu desempenho. Além disso, o comportamento gerado revelou ser dependente da tarefa, a ver, enquanto o Walker evoluído com o valor fixo de 400 passos resultou em um experimento do tipo 2, o Ant nas mesmas configurações resultou em um comportamento geral do tipo 3.

Os experimentos envolvendo Hopper, quando nas configurações padrão de PyBullet (fixo-1000), tinham comportamentos gerais de tipo 1, ou seja, os agentes, na sua maioria, não desenvolveram comportamentos que mostrassem qualquer sinal de aquisição da habilidade de locomoção. Contudo, indicavam alguma habilidade que tendia ao comportamento esperado, como ficar em pé, geralmente perdendo o equilíbrio e caindo antes do final da avaliação. Tal ação pode ser explicada pelo fato de que como, por padrão, o bônus de vida dado ao agente por permanecer em equilíbrio é alto, 1000 pontos, pois o bônus de vida é igual ao tamanho máximo da duração do episódio avaliativo, a tentativa de adquirir a habilidade acaba sendo desvantajosa, já que, no começo, esse agente poderá perder pontos por tentar saltar e cair.

Para os experimentos do tipo fixo-N, onde $N! = 1000$ passos, tipicamente o agente Hopper alcançou classificações de tipo 2. Considere o Hopper evoluído com episódios de avaliação de 100 passos, o agente aprende a saltar com as suas articulações bem flexionadas, mas cai logo em seguida. O mesmo acontece para 200 passos. Novamente, a explicação para tal comportamento parece estar na pontuação inicial e na relação dessa pontuação com o episódio avaliativo. Por exemplo, o agente pode ficar em equilíbrio durante o episódio, porém só conseguirá um máximo de 100 pontos, se saltar, entretanto, poderá obter uma avaliação maior. Como o episódio é curto, pode acontecer de ele nem cair durante o processo de avaliação, não perdendo os pontos de queda. Logo, é mais vantajoso que o comportamento de dar um pulo, mesmo com o risco de cair, seja desenvolvido. De forma geral, isso indica que a mudança na duração do episódio modifica o peso dos componentes da fitness, tornando viável desenvolver comportamentos funcionais, saltar, antes inviáveis em função da existência de comportamentos subótimos vantajosos, ficar parado.

Especificamente, os indivíduos que passam metade da sua evolução com um episódio de avaliação de 100 passos e aqueles com 200 passos mantêm ambos a articulação superior

direita ligeiramente curvada para trás, para então saltar, o que aumenta a potência do salto. A articulação do pé experimenta uma maior mudança de ângulo, enquanto que a articulação do meio permanece maioritariamente dobrada. No entanto, alguns agentes ficam parados, inclinados para a frente em equilíbrio, e são classificados como tipo 1. Na configuração de 500 a 1000 passos, a maioria dos agentes apresenta este tipo de comportamento.

Para o experimento que varia incrementalmente a duração do episódio, o agente Hopper mostrou o seu melhor rendimento, no qual todos os agentes atingiram comportamentos de tipo 3. Neste caso, é interessante notar que a distância média alcançada entre saltos foi também maior do que a média dos outros experimentos. Estes resultados indicam que, embora a utilização apenas de episódios curtos conduza a comportamentos de tipo 2, e a utilização apenas de episódios longos conduza a comportamentos de tipo 1 devido aos fortes mínimos locais de imobilização, a experiência incremental permite uma transição mais eficaz de comportamentos de tipo 2 para comportamentos de tipo 3. Finalmente, a versão sem bônus de vida mostra comportamentos de salto em todos os seus agentes, o que confirma que os resultados de [17] que sustenta que a função de aptidão adaptada remove os mínimos locais que levam o agente a ficar parado, em equilíbrio. Entretanto, olhando a performance, esses agentes acabam sendo penalizados mais pelo gasto elétrico e juntas ao limite.

A Figura 2 resume a classificação obtida para todos os agentes que evoluíram no melhor caso para cada tipo de experimento e a tabela II mostra a classificação mais comum para cada experimento em comparação com o desempenho médio.

Tabela II

Classificação comportamental do Hopper para os melhores experimentos em relação a performance

Experimento	Performance média	Classificação
Fixo-1000	1013.90 +- 4.05	1
Fixo-500	1012.86 +- 621.45	1
Fixo-100-1000	1935 +- 641.91	3
Incremental	2443.17 +- 195.92	3
V5 - controle	1839.39 +- 275.87	3

Quando comparamos os tipos de comportamentos gerados com o desempenho obtido, vemos que, em geral, o desenvolvimento de comportamentos de tipo 3 é consistente com a obtenção de um alto desempenho. Contudo, o experimento Fixo-1000, ou seja, o padrão do simulador, tem um desempenho médio superior àquele em que a duração do episódio avaliativo se mantém em 500 passos, e quando comparamos os comportamentos gerados, este último apresenta um agente de tipo 1 mas tem muitos agentes com comportamentos próximos do tipo 3, ou seja, conseguem saltar durante um certo período antes de caírem. Neste sentido, os melhores desempenhos nem sempre apresentam melhores comportamentos qualitativos.

Por outro lado, os resultados obtidos com a utilização do Walker são diferentes. Por exemplo, no experimento padrão, os agentes são classificados como tipo 2, todos aprenderam a dar

Hopper: classificação comportamental

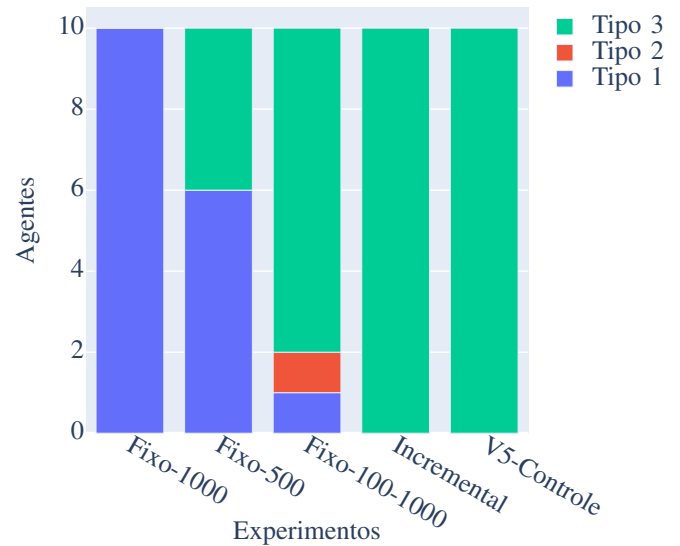


Fig. 2: A classificação do Hopper para os experimentos de [15] que tiveram os melhores resultados de performance.

um passo e depois permanecem em equilíbrio. As experiências em que toda a evolução se mantém a um valor fixo, em geral, situam-se entre 2 e 3. Damos 100 passos aos agentes que evoluíram, metade deles têm comportamentos como andar rápido e depois cair ou andar com as articulações do joelho dobradas, enquanto parte deles pode andar normalmente. A configuração a 200 passos é capaz de gerar comportamentos de tipo 3, embora eventualmente alguns agentes caiam.

Mais uma vez, para os experimentos em que 50% da evolução permanece num valor predeterminado e 50% a 1000 passos, 50% dos experimentos apresentam comportamentos 50%, do tipo 3, geralmente com os agentes a aprender a andar e eventualmente a cair. O incremental também comporta-se de forma semelhante. A versão V5-controle, por outro lado, atingem um 3 em 80% dos seus agentes, embora não de forma unânime como o Hopper Incremental.

A imagem 3 resume a classificação obtida para todos os agentes Walker2D que evoluíram sob o melhor caso para cada tipo de experiência e a tabela III mostra a classificação mais comum para determinada experiência em comparação com o desempenho médio.

Mais uma vez, pode-se perceber que os agentes com os melhores comportamentos nem sempre recebem as pontuações médias mais altas. Comparamos o caso do Walker a evoluir incrementalmente, em que a maioria dos seus agentes recebeu classificações de tipo 2 em contraste com os agentes cuja duração do episódio permaneceu fixa em 200 passos, em que a maioria dos agentes foi classificada como tipo 3, por não ter acabado por cair, os primeiros foram pontuados mais alto do que os segundos, apesar de os segundos apresentarem um

Walker: classificação comportamental

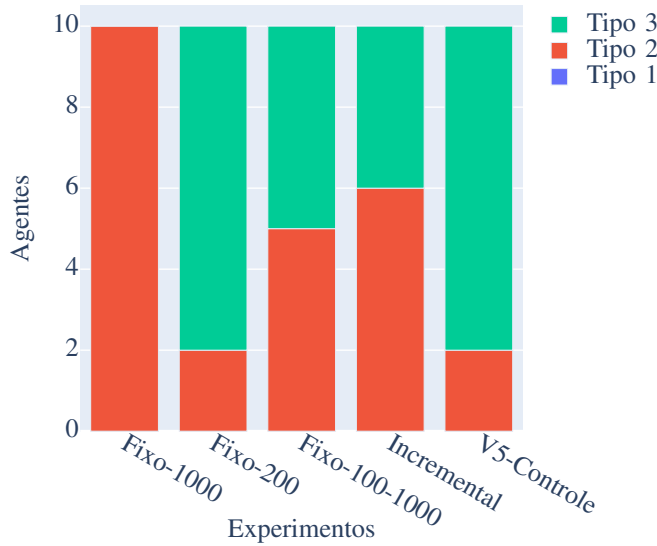


Fig. 3: A classificação do Walker para os experimentos de [15] que tiveram os melhores resultados de performance.

Tabela III

Comportamento do Walker para os melhores resultados obtidos nos experimentos em relação a performance

Experimento	Performance média	Classificação
Fixo-1000	903.19 +- 51.01	1
Fixo-200	1086.34 +- 600.79	3
Fixo-100-1000	1339.40 +- 601.91	2/3
Incremental	1425.86 +- 649.05	2
V5 - controle	940.44 +- 469.18	3

comportamento mais semelhante à habilidade desejada para o Walker.

Os experimentos com o ambiente Ant, em geral, geraram bons comportamentos em todos os casos. No padrão, metade dos agentes apresentaram a classificação tipo 1 e metade a classificação tipo 3. As configurações com outros valores fixos, e metade da evolução a um valor e metade a 1000 passos, apresentaram classificações de tipo 2 e 3, com alguns casos de tipo 1. Todas as experiências com acréscimo e bônus de vida desenvolveram, em todos os casos, agentes do tipo 3, contudo, como já mencionado no caso do Hopper, a forma como a movimentação ocorre em cada experimento varia.

A imagem 3 resume a classificação obtida para todos os agentes Ant evoluíram sob o melhor caso para cada tipo de experimento e a tabela IV mostra a classificação mais comum para determinado experimento em comparação com o desempenho médio.

Nos experimentos com agentes Ant, a performance não variava significativamente em relação ao comportamento observado, porém, experimentos do tipo V5 tiveram desempenho médio 52% inferiores em comparação com o tipo incremental.

Ant: classificação comportamental

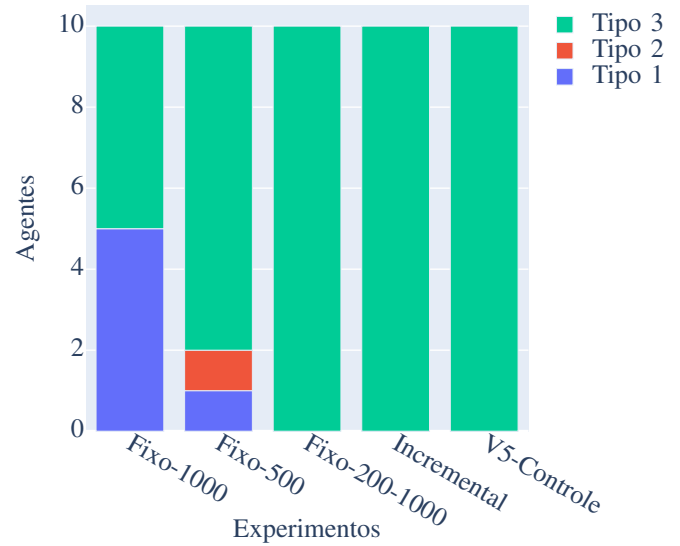


Fig. 4: A classificação do Ant para os experimentos de [15] que tiveram os melhores resultados de performance.

Tabela IV

Classificação comportamental da Ant para os melhores experimentos em relação a performance

Experimento	Performance Média	Classificação
Fixo-1000	1455.90 +- 605.96	1/3
Fixo-500	2130.23 +- 521.67	3
Fixo-200-1000	2651.85 +- 140.87	3
Incremental	2577.74 +- 140.87	3
V5 - controle	1340.42 +- 236.03	3

Isto porque a estabilidade deste tipo de agente não é tão grande como nos outros casos, ou seja, acabam por cair ou mover-se de forma a perder pontos pela movimentação excessiva de juntas, e consequente consumo de energia.

Com a análise do comportamento concluída, o próximo passo é discutir como diferentes resultados se relacionam. Por exemplo, experimentos que mantém fixo a duração do episódio possuem parte da evolução com as mesmas condições daqueles que mudam em 50% da evolução para 1000 passos. Busca-se, portanto, entender como essas etapas semelhantes auxiliam na geração de comportamentos melhores quando se aumenta a duração do episódio para os 1000 passos padrões.

Os gráficos em 9 e 6 demonstram a diferença entre os experimentos do Hopper em que a evolução é fixa em N passos em comparação com a que tem a evolução 50% em 1000 passos. Nota-se que os comportamentos do tipo 2 gerados no primeiro experimentos, geralmente são refinados para os do tipo 3 quando metade da evolução acontece com um número maior de episódios. Entretanto, quando o valor inicial já é maior logo no início, como quando a evolução começa com episódios de avaliação em 400 ou 500 passos, os resultados

acabam não alcançando os valores esperados, principalmente porque a duração inicial acaba impedindo que seja vantajoso desenvolver as habilidade iniciais de "pular e cair" nos agentes. Ou seja, eles normalmente ficam parados ou apenas aprendem a pular e permanecer no lugar.

Os gráficos 7 e 8 já mostram resultados um pouco diferente, visto que cada morfologia impacta em como os agentes percebem e agem durante o processo avaliativo. No caso do Walker, quando tem-se um valor menor e fixo na duração do episódio, percebe-se que ele aprende a dar um passo e ficar em equilíbrio, quando aumenta-se a duração para 1000 passos na segunda metade do episódio ele já varia no comportamento, podendo aprender a continuar caminhando ou a apenas dar um passo, apenas um desenvolveu a habilidade de ficar parado.

Valores maiores da duração no início do processo raramente geram resultados do tipo 3, pois eles aprendem a dar um passo muito grande e permanecer em equilíbrio, como a penalidade de cair caso tentem dar outro passo é muito grande, acabam não desenvolvendo a habilidade esperada. Dessarte, é perceptível que a estabilidade do agente frente a uma ação interfere diretamente no impacto que as diferentes configurações de experimento vão ter nos robôs gerados.

Enquanto isso, os gráficos em 9 e 10 corroboram a ideia de que aumentar o tamanho do episódio durante o processo evolutivo ajuda a transitar agentes do tipo 2 para o tipo 3. Como é fácil manter a estabilidade ao andar para o agente, é raro ter um experimento em que o resultado não seja bom.

Hopper: classificação Fixo-N-1000

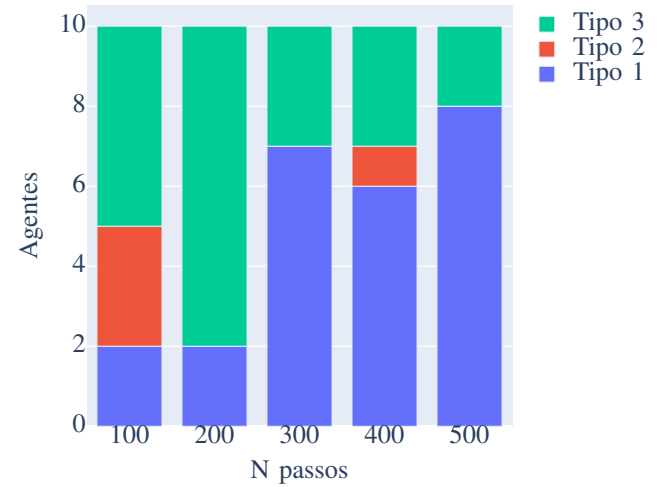


Fig. 6: A classificação do Hopper para todos os experimentos feitos com o valor da duração do episódio mudando a cada 50% da evolução

Hopper: classificação Fixo-N

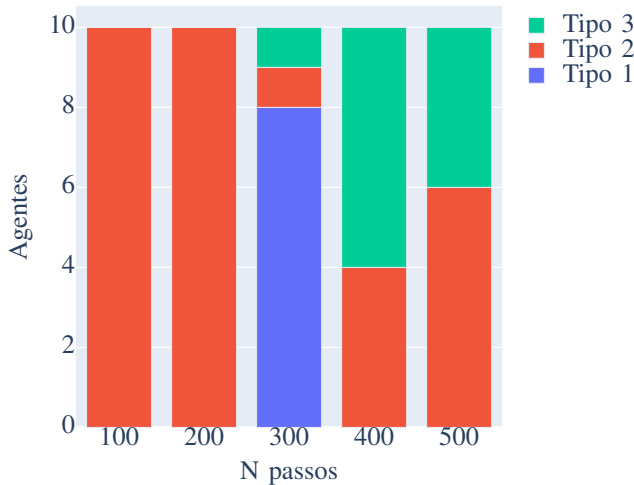


Fig. 5: A classificação do Hopper para todos os experimentos feitos com o valor da duração do episódio de avaliação fixo

Walker: classificação Fixo-N

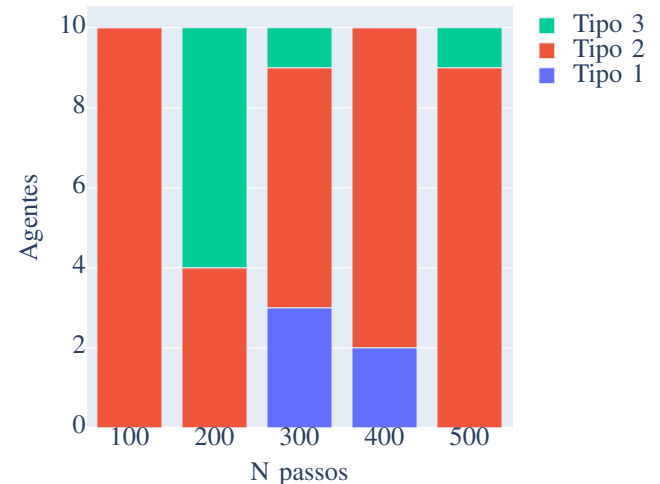


Fig. 7: A classificação do Walker para todos os experimentos feitos com o valor da duração do episódio de avaliação fixo

IV. CONCLUSÃO

O presente trabalho investigou como a duração dos episódios de avaliação de um agente impacta qualitativamente

Walker: classificação Fixo-N-1000

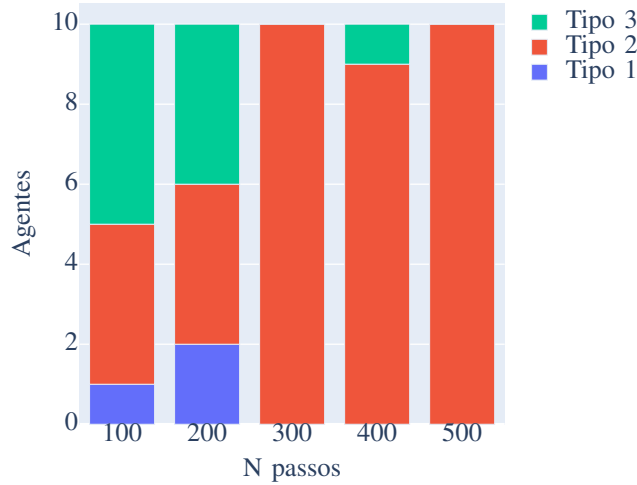


Fig. 8: A classificação do Walker para todos os experimentos feitos com o valor da duração do episódio mudando a cada 50% da evolução

Ant: classificação Fixo-N-1000

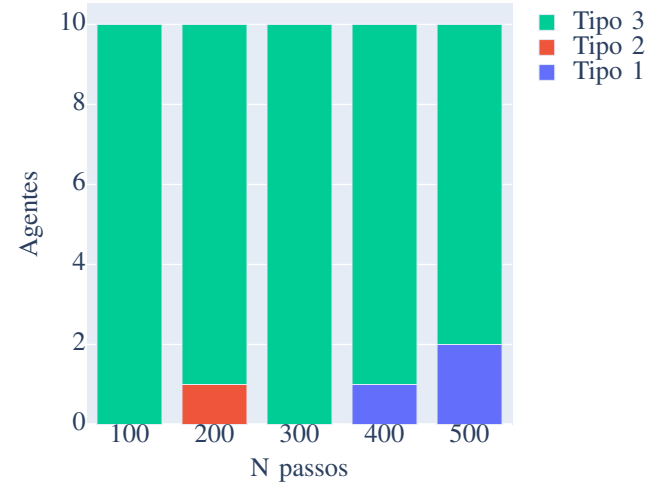


Fig. 10: A classificação do Ant para todos os experimentos feitos com o valor da duração do episódio mudando a cada 50% da evolução

Ant: classificação Fixo-N

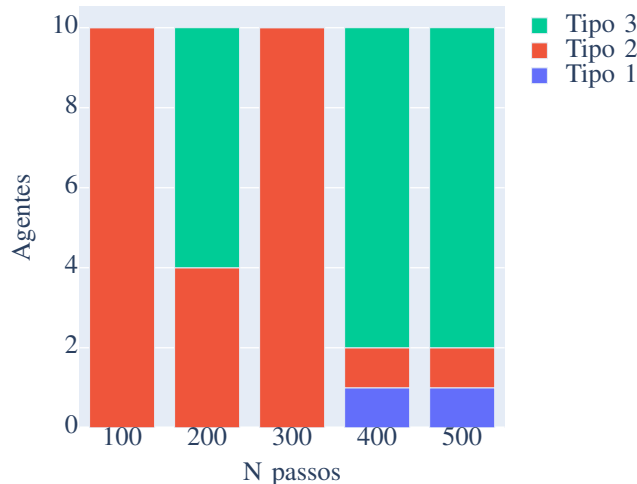


Fig. 9: A classificação do Ant para todos os experimentos feitos com o valor da duração do episódio de avaliação fixo

os comportamentos gerados, e como este impacto leva a comportamentos de locomoção mais efetivos. Os resultados obtidos apontam que os comportamentos gerados são mais ou menos efetivos a depender de quão fácil é estabilizar cada agente, de acordo com a sua morfologia, e quão complexa é a habilidade de se locomover. Como supracitado, durações pequenas geram comportamentos iniciais que indicam certo aprendizado inicial em relação a tarefa a ser aprendida, enquanto episódios maiores selecionam comportamentos de ficar parado ou de buscar posições de equilíbrio como, por exemplo, dar um passo e parar, bloqueando a evolução futura de comportamentos efetivos. Além disso, modificar a duração dos episódios de avaliação, iniciando com episódios menores e aumentando durante o processo evolutivo acaba facilitando o desenvolvimento dos comportamentos iniciais e intermediários que facilitem a posterior evolução de comportamentos efetivos.

Outro ponto importante é que a função que calcula a performance nem sempre conduz o processo evolutivo ao desenvolvimento desses comportamentos iniciais, fazendo que robôs que apresentam o comportamento de "ficar parado e ganhar" se sobressaiam sobre os que andam e caem, mesmo que o comportamento de andar e cair possa ser refinado para um comportamento mais efetivo. Isso se dá por conta do bônus de vida, retirar esse bônus, apesar de selecionar esses robôs que aprendem a andar no começo, também não seleciona estabilidade e acabam sendo muito custosos em questão energética.

Demonstra-se então que embora o resultado da função que calcula a performance, isto é, a função *fitness*, esteja longe de ser uma métrica ruim para a avaliação dos agentes, ela

não deve ser utilizada como a única métrica de avaliação. Como demonstrado anteriormente, a performance, calculada através da função *fitness*, nem sempre indica melhores comportamentos ou mesmo se os comportamentos desenvolvidos até então levariam a melhores resultados futuros. Portanto, estudos comportamentais são importantes para avaliar outras características dos experimentos realizados dentro do contexto da robótica evolutiva de forma a entender o que de fato está sendo observado, e como essas observações podem ser utilizadas para melhor compreender o próprio processo evolutivo.

Como trabalhos futuros, mostra-se como uma direção interessante investigar como o desenvolvimento do comportamento acontece durante todo o processo evolutivo, bem como quão robustas são as soluções à mudança ambiental. Ainda, percebendo quão importante é a análise comportamental para a área em questão, é necessário a definição de protocolos mais definidos para realizar tais estudos, de modo a obter um framework padrão que possa facilmente ser replicado em novas tarefas.

REFERENCES

- [1] S. Bensalem, M. Gallien, F. Ingrand, I. Kahloul, and N. Thanh-Hung, "Designing autonomous robots," *IEEE Robotics & Automation Magazine*, vol. 16, no. 1, pp. 67–77, 2009.
- [2] F. Ingrand and M. Ghallab, "Deliberation for autonomous robots: A survey," *Artificial Intelligence*, vol. 247, pp. 10–44, 2017.
- [3] J. C. Bongard, "Evolutionary robotics," *Commun. ACM*, vol. 56, no. 8, p. 74–83, aug 2013. [Online]. Available: <https://doi.org/10.1145/2493883>
- [4] S. Nolfi and D. Floreano, *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines*. MIT press, 2000.
- [5] S. Nolfi, J. Bongard, P. Husbands, and D. Floreano, "Evolutionary robotics," in *Springer handbook of robotics*. Springer, 2016, pp. 2035–2068.
- [6] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever, "Evolution strategies as a scalable alternative to reinforcement learning," *arXiv preprint arXiv:1703.03864*, 2017.
- [7] A. L. Nelson, G. J. Barlow, and L. Doitsidis, "Fitness functions in evolutionary robotics: A survey and analysis," *Robotics and Autonomous Systems*, vol. 57, no. 4, pp. 345–370, 2009.
- [8] N. Milano, J. T. Carvalho, and S. Nolfi, "Moderate environmental variation across generations promotes the evolution of robust solutions," *Artificial life*, vol. 24, no. 4, pp. 277–295, 2019.
- [9] T. Jansen, K. A. D. Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evolutionary Computation*, vol. 13, no. 4, pp. 413–440, 2005.
- [10] P. Pagliuca and S. Nolfi, "Robust optimization through neuroevolution," *PloS one*, vol. 14, no. 3, p. e0213193, 2019.
- [11] J.-B. Mouret and K. Chatzilygeroudis, "20 years of reality gap: a few thoughts about simulators in evolutionary robotics," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2017, pp. 1121–1124.
- [12] D. S. Shah, J. P. Powers, L. G. Tilton, S. Kriegman, J. Bongard, and R. Kramer-Bottiglio, "A soft robot that adapts to environments through shape change," *Nature Machine Intelligence*, vol. 3, no. 1, pp. 51–59, 2021.
- [13] H.-G. Beyer and H.-P. Schwefel, "Evolution strategies—a comprehensive introduction," *Natural computing*, vol. 1, no. 1, pp. 3–52, 2002.
- [14] A. P. Wieland, "Evolving controls for unstable systems," in *Connectionist Models*. Elsevier, 1991, pp. 91–102.
- [15] L. G. Rosa, V. H. Homem, S. Nolfi, and J. T. Carvalho, "On the impact of the duration of evaluation episodes on the evolution of adaptive robots," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2022, pp. 520–529.
- [16] Admin, "Bullet real-time physics simulation," May 2021. [Online]. Available: <https://pybullet.org/>
- [17] P. Pagliuca, N. Milano, and S. Nolfi, "Efficacy of modern neuro-evolutionary strategies for continuous control optimization," *Frontiers in Robotics and AI*, vol. 7, p. 98, 2020.