JORNAL DE ENGENHARIA, TECNOLOGIA E MEIO AMBIENTE

ESTUDO QUANTITATIVO SOBRE APRENDIZAGEM POR REFORÇO UTILIZANDO UM AGENTE INTELIGENTE NA RESOLUÇÃO DO JOGO DA FORCA

Fábio Oliveira[1]; Otavio Folador[2]; Fernando Eduardo Resende Mattioli[3,4]; Eduardo Fernandes Saad[5]

[1,2,3,5] Faculdade de Talentos Humanos - FACTHUS, Uberaba (MG), Brasil
[4]Universidade Federal do Triângulo Mineiro – UFTM, Uberaba (MG), Brasil

fabio.erickson@hotmail.com, otaviofolador@gmail.com, fernando.mattioli@facthus.edu.br, eduardo.saad@facthus.edu.br

RESUMO: O uso de técnicas de Inteligência Artificial tem apresentado um significativo aumento nos últimos anos, sendo a mesma aplicada na resolução de vários tipos de problema incluindo a predição de doenças, veículos autônomos e recomendações de compras online. Dentre os métodos utilizados no domínio da Inteligência Artificial, a aprendizagem por reforço destaca-se por utilizar um mecanismo de recompensas ao treinar um agente para uma tarefa específica. O presente estudo tem como objetivo demonstrar o uso de um agente treinado por aprendizagem por reforço na resolução do jogo da forca, um jogo simples e muito conhecido. Os experimentos apresentados neste trabalho foram desenvolvidos utilizando a biblioteca *ReactNative* e o editor de texto *Visual Studio Code*. Para se avaliar o agente inteligente, o desempenho do mesmo foi comparado ao de um agente puramente estatístico, sendo constatada a superioridade do agente inteligente no conjunto de dados analisado.

PALAVRAS CHAVE: Inteligência Artificial, Aprendizagem por reforço, Jogo da forca.

A QUANTITATIVE STUDY ON REINFORCEMENT LEARNING USING AN INTELLIGENT AGENT IN THE HANGMAN GAME

ABSTRACT: *The use of Artificial Intelligence techniques have been presenting a significant increase in recent years, given its application in there solution of many classes of problems including disease prediction, autonomous vehicles and online shopping recommendation. Amongst the methods used in the Artificial Intelligence domain, reinforcement learning is worth mentioning given its characteristic of presenting rewards during agent training for a specific task. This study hás the objective of demonstrating the use of a reinforcement learning-based agent applied to the hangman game, a simple and widely known game. The experiments presented in this work were developed using the React Native library and the Visual Studio Code text editor. Aiming at evaluating the intelligenta gent, its performance was compared to a purely statistics-based agent, which was outperformed by the intelligent agent in the analyzed dataset.*

KEYWORDS: *Artificial Intelligence, Reinforcement Learning, Hangman game.*

INTRODUÇÃO

A inteligência artificial é definida como "uma capacidade do sistema para interpretar corretamente dados externos, aprender a partir desses dados e utilizar essa aprendizagem para atingir objetivos e tarefas específicos" (KAPLAN; HAENLEINB, 2018). O principal objetivo dos sistemas de Inteligência Artificial (IA) é executar funções que seriam consideradas inteligentes. É um conceito que recebe várias definições, apresentando como características básicas: capacidades de raciocínio (aplicar regras lógicas a um conjunto de dados disponíveis para chegar a uma conclusão), aprendizagem (aprender com os erros e acertos de forma que no futuro possam agir de maneira mais eficaz), reconhecimento de padrões (tanto padrões visuais e sensoriais, como também padrões de comportamento) e inferência (capacidade de conseguir aplicar o raciocínio nas situações do nosso cotidiano) (VASCONCELOS; MARTINS JUNIOR, 2004).

Na área da IA, agentes inteligentes são treinados durante o processo de aprendizagem. No entanto, várias técnicas diferentes estão disponíveis para o treinamento desses agentes, dentre as quais destaca-se a aprendizagem por reforço. Nesta técnica, o agente toma suas decisões a partir de uma probabilidade de exploração (escolha de uma ação aleatória) ou usufruto (decisão com base no conhecimento adquirido). Quando a ação escolhida leva a um resultado positivo, o agente é recompensado por esta decisão, o que aumenta a probabilidade de a mesma ser escolhida futuramente. Em contrapartida, quando uma ação produz um efeito indesejável, esta recebe uma punição, o que diminui a chance de escolha futura desta mesma ação. Após um extenso período de treinamento, o agente adquire um conhecimento razoável, que permite ao mesmo a escolha das ações com melhor potencial para um resultado satisfatório (RUSSEL; NORVIG, 2013).

O presente estudo avalia o treinamento de um agente inteligente com aplicação no jogo da forca. Para tal, foram conduzidos vários experimentos, avaliando dois métodos distintos de treinamento: no primeiro, o agente é treinado com um conjunto de palavras escolhidas aleatoriamente, possivelmente repetindo-se algumas

palavras ao longo do treinamento. No segundo método, o agente é treinado com um conjunto de palavras sem repetição, ou seja, cada palavra será visualizada pelo agente uma única vez. Finalmente, o desempenho do agente inteligente será comparado ao de um agente puramente estatístico, que faz suas escolhas com base na frequência absoluta de cada letra, no dicionário de treinamento. Para tal, um conjunto de palavras distintas das utilizadas em treinamento foi criado (conjunto de validação).

MATERIAIS E MÉTODOS

Para o desenvolvimento do agente utilizou-se a linguagem *Typescript* que é um superconjunto de *JavaScript* desenvolvido pela Microsoft que adiciona tipagem e alguns outros recursos à linguagem (FOLEY, 2012). Também se empregou juntamente com a linguagem *Typescript* a biblioteca *ReactNative*, que foi criada pelo facebook e é utilizada para criar aplicações para Android e IOS.

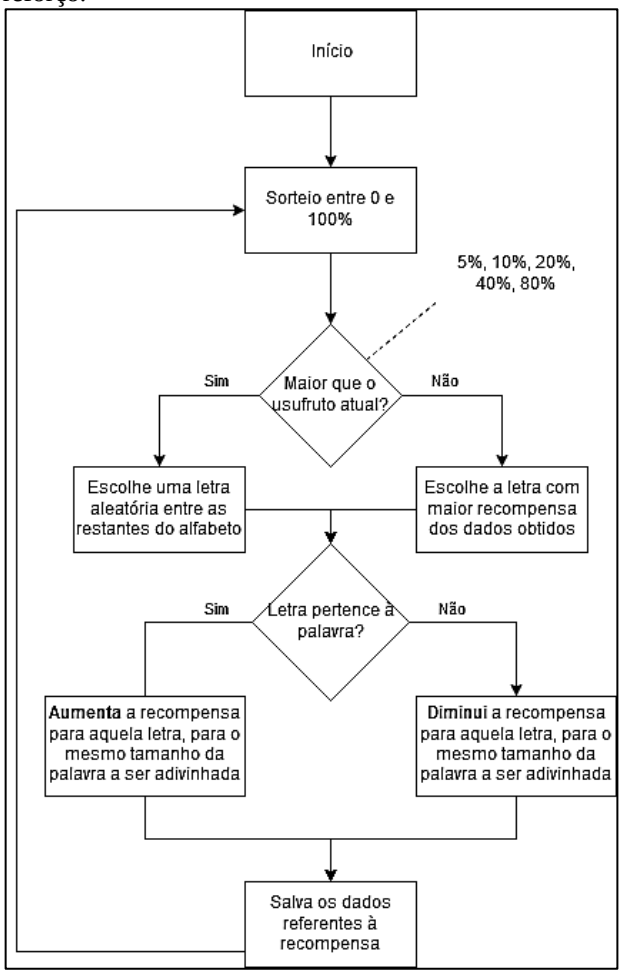
Foi desenvolvido um aplicativo para realizar e demonstrar o uso do agente inteligente no jogo da forca. Para tal, utilizou-se um dispositivo Android da marca Xiaomi, modelo Mi 9T Pro com Processador Qualcomm Snapdragon 855 e 6GB de memória RAM.

Para aplicação ao jogo da forca, o agente de aprendizagem por reforço foi treinado utilizando-se o seguinte procedimento: é selecionada uma palavra a partir de um dicionário de palavras destinado ao treinamento do agente. O processo de aprendizagem consiste no preenchimento de um mapa de recompensas, no qual uma recompensa é atribuída a cada letra do alfabeto. As letras com maiores recompensas possuem maior probabilidade de acerto, enquanto que as letras com menores recompensas possuem maior probabilidade de erro. A partir do conjunto de palavras de treinamento, foi criado um mapa de recompensas para cada tamanho de palavra.

A cada rodada, o agente então escolhe uma letra, executando a exploração ou o usufruto do mapa de recompensas segundo uma probabilidade pré-definida de usufruto. Na exploração, uma letra é escolhida aleatoriamente, a partir do conjunto de letras disponíveis (descartando-se as letras já utilizadas). No usufruto, o agente selecionará a letra com maior recompensa, dentre as letras disponíveis. Após a escolha, caso a letra esteja presente na palavra em questão, a mesma recebe uma recompensa positiva no mapa. Caso a letra não faça parte da palavra chave, a mesma recebe uma recompensa negativa (ou punição). Uma vez finalizado o jogo (pela descoberta da palavra chave ou pelo esgotamento do número de tentativas), uma nova palavra é escolhida e o jogo é reiniciado, mantendo-se porém, o estado atual do mapa de recompensas para os próximos jogos. A Fig. 1 apresenta resumidamente este processo de treinamento.

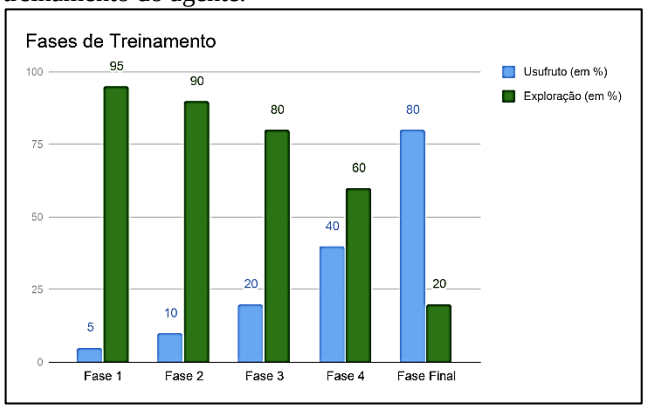
O treinamento do agente foi realizado em 3 experimentos distintos. Em um primeiro experimento, foram utilizadas 4000 palavras em treinamento. No segundo, foram utilizadas 10000 palavras para o treinamento. Finalmente, 14623 palavras foram utilizadas para treinar o agente. Para este trabalho, foi utilizado um dicionário composto apenas por palavras que contenham entre 4 e 8 letras. Um conjunto de 100 palavras - diferentes das palavras contidas no dicionário de treinamento - foi utilizado para validação dos agentes treinados.

Figura 1: Treinamento do agente de aprendizagem por reforço.



Fonte: Os autores, 2019.

Figura 2: Probabilidades de exploração e usufruto durante o treinamento do agente.

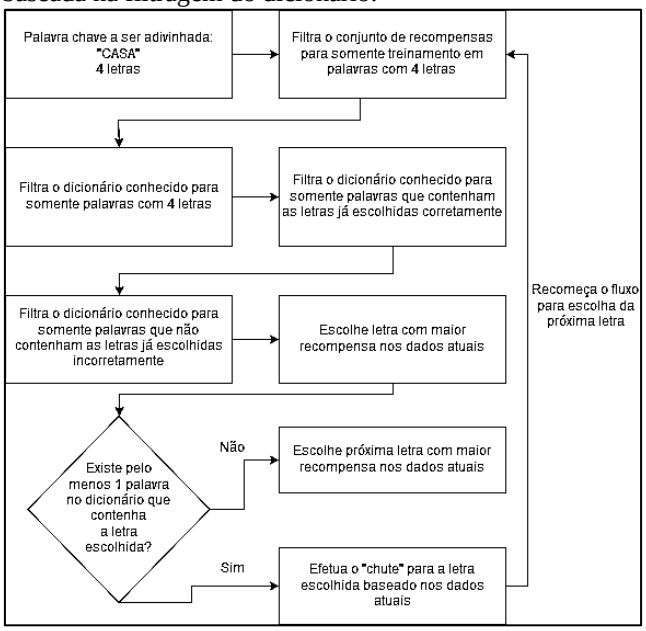


Fonte: Os autores, 2019.

O treinamento do agente foi dividido em 5 fases distintas, cada uma contendo uma probabilidade diferente de usufruto. Partindo-se do valor de 5%, a probabilidade de usufruto foi dobrada a cada fase, obtendo-se o valor de 80% na quinta fase do treinamento (fase final). A Fig. 2 apresenta estas probabilidades.

A escolha de uma letra pelo agente é direcionada por três aspectos: o tamanho da palavra, as palavras candidatas (excluindo-se aquelas que possuem letras que já foram testadas e não estão presentes na palavra chave) e o mapa de recompensas. Desta forma, uma letra que apresente uma recompensa alta mas não esteja presente nas palavras candidatas será ignorada. Cada um destes aspectos atua como um filtro na escolha das possíveis candidatas para a palavra chave. A Fig. 3 apresenta resumidamente este procedimento.

Figura 3: Fluxograma explicativo para escolha da letra baseada na filtragem do dicionário.



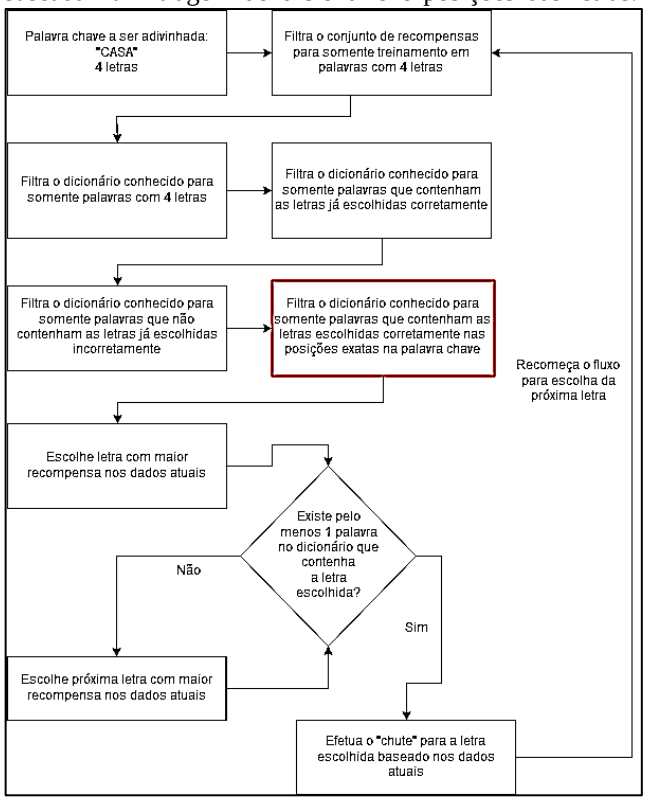
Fonte: Os autores, 2019.

No exemplo apresentado na Fig. 3, a palavra a ser adivinhada é “CASA”. O agente inteligente irá escolher uma letra baseando-se primeiramente no tamanho da palavra, neste caso 4 letras, utilizando apenas o mapa correspondente a este tamanho de palavra.

Supondo que o agente escolha a letra “A” nesta rodada, e escolha será bem sucedida e a palavra a ser adivinhada estará na seguinte máscara: _A_A. O agente então, para a escolha da próxima letra, irá basear-se somente nas palavras que contenham “A”, independentemente das posições da letra na palavra.

Supondo que agora o agente escolha a letra “P”. Este não obterá sucesso e a máscara da palavra a ser adivinhada não sofrerá alteração. No entanto, as palavras que contenham a letra “P” são agora excluídas do conjunto de palavras candidatas, uma vez que esta letra foi rejeitada. Sendo assim, na próxima rodada o agente escolherá a letra baseando-se em palavras que contenham a letra “A” e não contenham a letra “P”. Uma pequena variação deste procedimento foi testada neste trabalho. Nesta variação, a filtragem das palavras considera não somente a presença das letras, mas também a posição das mesmas na palavra chave. Os demais passos do procedimento não foram alterados, com o objetivo de se isolar a influência desta única alteração no desempenho do agente. O procedimento modificado é apresentado na Fig. 4.

Figura 4 - Fluxograma explicativo para escolha da letra baseada na filtragem do dicionário e posições das letras.



Fonte: Os autores, 2019.

Como exemplo da aplicação do procedimento apresentado na Fig. 4, a filtragem se daria na seguinte forma: considerando-se novamente a palavra chave “CASA”, o agente já havia escolhido anteriormente a letra “A”. Sendo assim, a palavra a ser adivinhada estaria na máscara: _A_A. O novo filtro para as palavras candidatas consideraria somente palavras que contenham a letra “A” nas respectivas posições, como “FACA” ou “CAMA” por exemplo. Este processo de escolha é análogo ao pensamento humano, tendo em mente palavras conhecidas, letras já escolhidas, letras ainda não escolhidas e o tamanho da palavra a ser adivinhada.

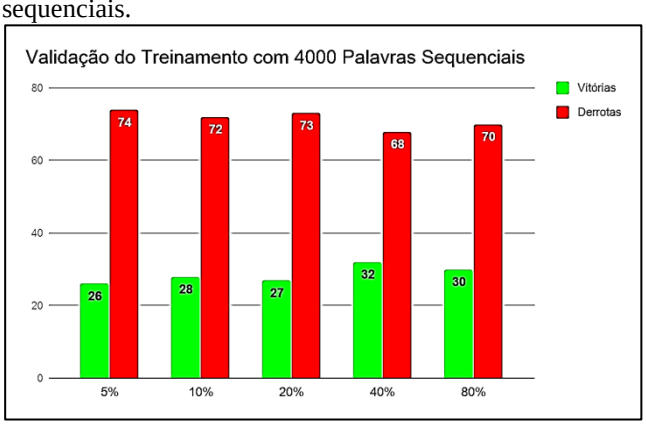
Para se avaliar o desempenho do agente, o mesmo foi configurado para 100% de usufruto, processando o conjunto de validação anteriormente criado (100 palavras desconhecidas em treinamento). Foram medidos o número de vitórias e derrotas neste conjunto de 100 amostras. Para comparação dos resultados, o desempenho de referência foi

utilizado configurando um agente estatístico, que escolhe as letras baseando-se unicamente na distribuição estatística destas no conjunto de treinamento.

RESULTADOS E DISCUSSÃO

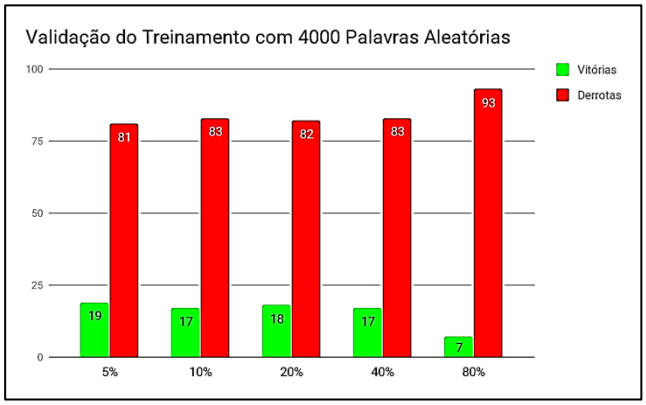
As Fig. 5 e 6 apresentam o número de vitórias e derrotas obtidos ao processar o conjunto de validação, após o treinamento com, respectivamente, 4000 palavras sequenciais (sem repetição) e 4000 palavras aleatórias (com repetição).

Figura 5: Desempenho em validação para 4000 palavras sequenciais.



Fonte: Os autores, 2019.

Figura 6: Desempenho em validação para 4000 palavras aleatórias.

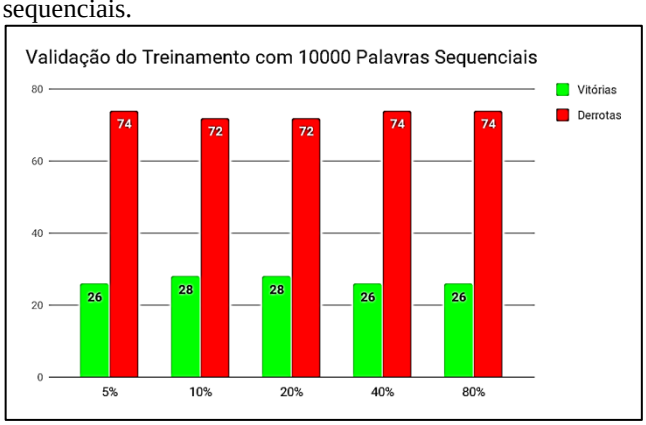


Fonte: Os autores, 2019.

Observa-se que o procedimento de treinamento utilizando palavras sequenciais (sem repetição de palavras) apresentou resultados melhores, para todos os valores de usufruto. Não foram observadas diferenças significativas entre valores distintos de usufruto, exceto pela queda observada ao se utilizar palavras aleatórias com 80% de usufruto. Nestas condições, percebe-se que o reduzido número de amostras utilizadas na exploração (possivelmente repetidos inclusive) afeta negativamente o desempenho do agente. Nestes testes, o melhor resultado observado foi de 32% de vitórias, após o treinamento com 40% de usufruto e o conjunto sequencial de palavras. As Fig. 7 e 8 apresentam os resultados para o treinamento com 10000 palavras.

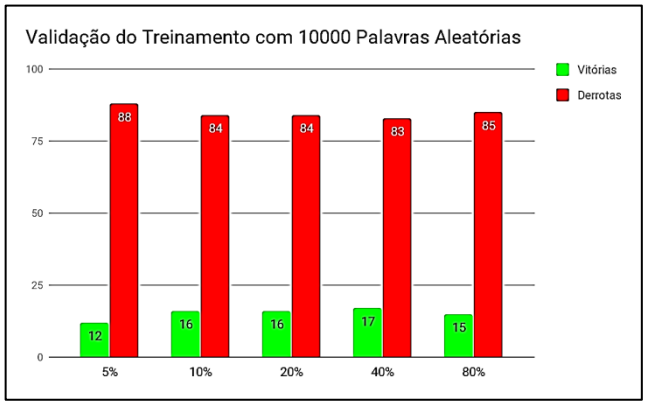
Assim como no caso anterior, observa-se, pelas Fig. 7 e 8 o melhor desempenho do agente treinado com o conjunto sequencial de palavras. Ao se comparar estes gráficos com os obtidos para 4000 palavras, percebe-se uma piora no desempenho, para ambos os casos (de 32% para 28% no conjunto sequencial, e de 19% para 17% no conjunto aleatório). Este fato pode ser justificado pelo maior distanciamento das amostras de treinamento do conjunto de validação, uma vez que mais palavras estão disponíveis aumentando os ruídos e oscilações no mapa de recompensas. As Fig. 9 e 10 apresentam os resultados obtidos no processamento do conjunto de validação, utilizando agora todo o conjunto disponível de palavras (totalizando 14623 palavras).

Figura 7 - Desempenho em validação para 10000 palavras sequenciais.



Fonte: Os autores, 2019.

Figura 8 - Desempenho em validação para 10000 palavras aleatórias.



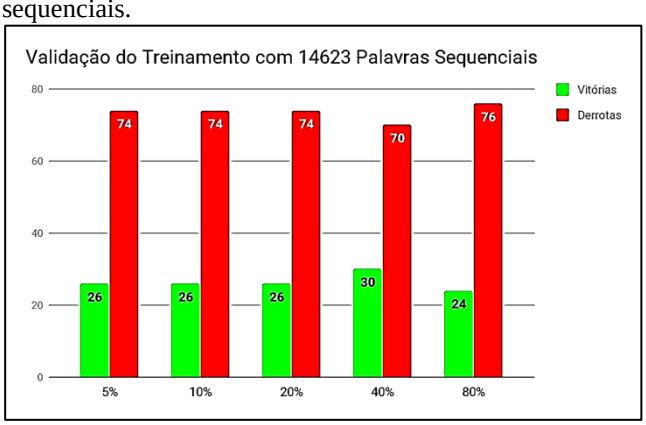
Fonte: Os autores, 2019.

Analisando os resultados apresentados pelas Fig. 9 e 10, percebe-se o mesmo comportamento obtido no experimento anterior, com exceção de um moderado aumento no número de vitórias (possivelmente ocasionado pela presença de mais palavras próximas ao conjunto de validação). O desempenho nos dois testes ainda é inferior àquele obtido com 4000 palavras, sendo o treinamento no

conjunto sequencial mais eficaz em comparação ao conjunto aleatório.

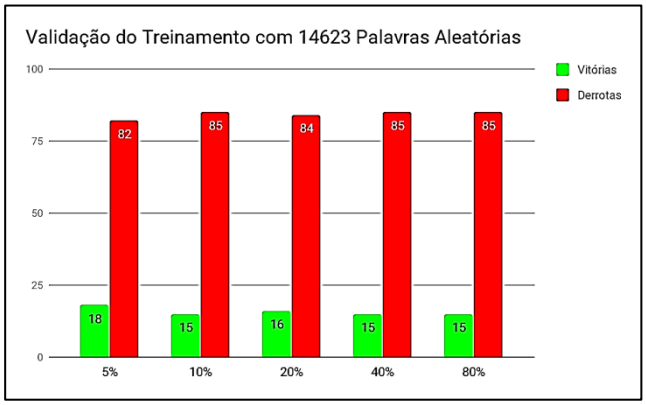
Pode-se observar, comparando os gráficos apresentados, um aumento em média de 12% na quantidade de acertos quando compara-se o treinamento com palavras sequenciais ao treinamento com palavras aleatórias. Deve-se ressaltar, no entanto, que estes experimentos foram conduzidos sem a filtragem de palavras. Desta forma, o agente sempre escolhe as letras com maior recompensa, independentemente de seus acertos ou erros.

Figura 9: Desempenho em validação para 14623 palavras sequenciais.



Fonte: Os autores, 2019.

Figura 10: Desempenho em validação para 14623 palavras aleatórias.



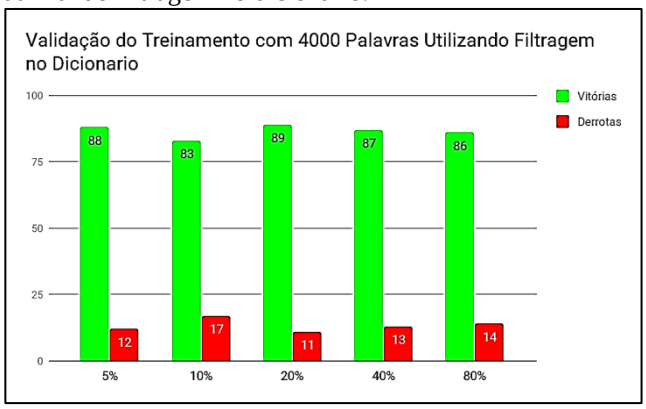
Fonte: Os autores, 2019.

Com o objetivo de avaliar o desempenho do agente em uma aplicação mais próxima da realidade, foi configurada a filtragem das palavras candidatas, conforme apresentado na seção “Materiais e Métodos”. Para tal, foi utilizado o treinamento com o conjunto sequencial de palavras, uma vez que este apresentou desempenho superior nos testes realizados. Os resultados destes experimentos são apresentados nas Fig. 11, 12 e 13.

Observa-se, pelo gráfico apresentado na Fig. 11 uma significativa melhora no desempenho do agente, quando comparado aos resultados obtidos sem a filtragem.

De fato, uma vez que a filtragem reduz a probabilidade de o agente escolher uma letra incorreta, este resultado condiz com o esperado, assemelhando o comportamento do agente ao de um humano, que também eliminaria algumas escolhas de acordo com o histórico do jogo em andamento. O melhor resultado foi obtido utilizando-se 20% de usufruto, apresentando um total de 89 vitórias.

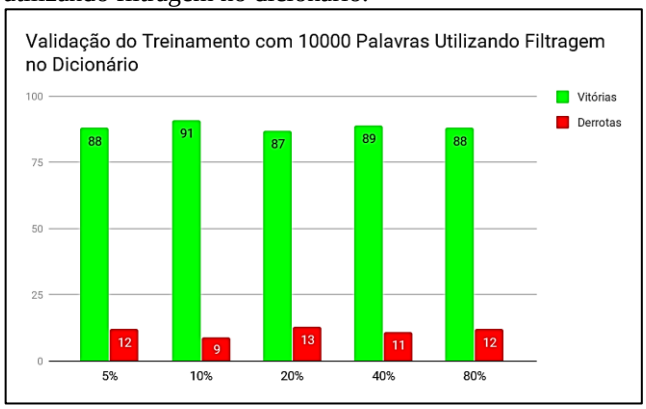
Figura 11: Desempenho em validação para 4000 palavras utilizando filtragem no dicionário.



Fonte: Os autores, 2019.

Os resultados apresentados na Fig. 12 revelam uma melhora em relação ao experimento com 4000 palavras (91 vitórias, obtidas com 10% de usufruto, contra 89 vitórias). Novamente, os resultados obtidos demonstram a significativa melhora do desempenho graças ao uso da filtragem. Estas características são confirmadas pelos resultados apresentados na Fig. 13, na qual um desempenho de 90 vitórias foi observado utilizando probabilidade de usufruto de 20 e 80%. A melhora observada em relação ao treinamento com 4000 palavras revela uma maior robustez do agente, resultante também da filtragem aplicada no conjunto de palavras candidatas. Tal fato se justifica pela característica da filtragem de eliminar palavras que estejam distantes das possibilidades, permanecendo apenas aquelas que se enquadrem perfeitamente na máscara de palavra atual.

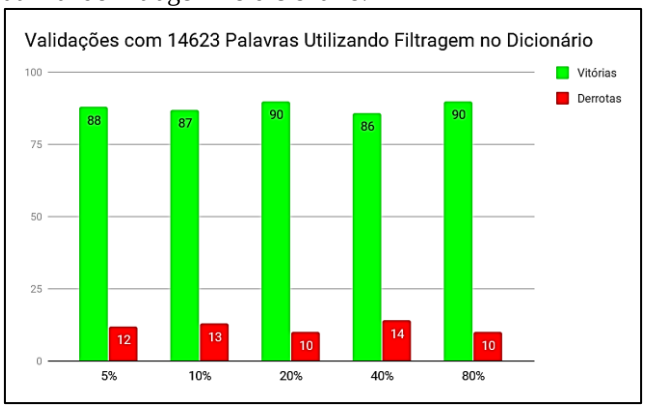
Figura 12: Desempenho em validação para 10000 palavras utilizando filtragem no dicionário.



Fonte: Os autores, 2019.

Finalmente, com o objetivo de se comparar o desempenho do agente inteligente com um valor de referência, foi avaliado o desempenho de um agente estatístico, cuja escolha das letras depende exclusivamente da frequência destas no conjunto de treinamento. O resultado desta avaliação é apresentado na Fig. 14.

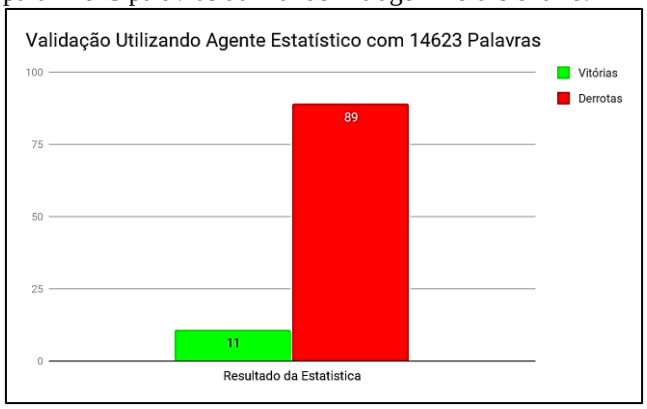
Figura 13: Desempenho em validação para 14623 palavras utilizando filtragem no dicionário.



Fonte: Os autores, 2019.

Observa-se, pela Fig. 14, que ainda que o agente estatístico utilize o mecanismo de filtragem, seu desempenho é bem inferior ao do agente inteligente (11 vitórias do agente estatístico contra 90 vitórias do agente inteligente). Este comportamento é justificado pelo fato de o agente inteligente sofrer punições quando uma letra não é encontrada na palavra chave, ainda que a mesma esteja presente na maioria das outras palavras. Este comportamento faz com que o agente inteligente opere em uma distribuição relativa das letras, enquanto o agente estatístico ignora os erros cometidos.

Figura 14: Desempenho em validação do agente estatístico para 14623 palavras utilizando filtragem no dicionário.



Fonte: Os autores, 2019.

A aleatoriedade das palavras no jogo da forca apresenta um grande desafio, seja para jogadores humanos ou para agentes computacionais. Os jogadores tentarão escolher as letras de acordo com seu conhecimento prévio (mapa de recompensas) e memória (conjunto de palavras candidatas). O código fonte utilizado na condução dos experimentos apresentados neste trabalho está disponível em https://gitlab.com/fabioerickson1/jogo-da-forca/.

CONCLUSÃO

Os experimentos conduzidos neste trabalho demonstraram a aplicabilidade das técnicas de aprendizagem por reforço no jogo da forca. Foi constatada a importância da filtragem no conjunto de palavras candidatas, responsável por um aumento significativo no desempenho do agente. Não foram observadas diferenças relevantes ao se experimentar taxas de usufruto distintas, no entanto sua exploração constitui um importante tema para trabalhos futuros. O agente inteligente desenvolvido apresentou desempenho superior ao de um agente puramente estatístico (melhoria de quase 80%), uma vez que as punições e recompensas obtidas modulam as probabilidades de escolha de cada letra no problema do jogo da forca.

Como possibilidade de trabalhos futuros destaca-se a investigação do comportamento do agente em outros problemas de categorias similares, tais como Sudoku ou Palavras Cruzadas. A habilidade do agente de continuar aprendendo após o treinamento também não foi explorada no presente trabalho, constituindo um importante tópico para pesquisas futuras.

REFERÊNCIAS

FOLEY, Mary Jo. **Microsoft takes the wraps off TypeScript, a superset of JavaScript**. [*S. l.*], 1 out. 2012. Disponível em: https://www.zdnet.com/article/microsoft-takes-the-wraps-off-typescript-a-superset-of-javascript/. Acesso em: 22 nov. 2019.

KAPLAN, Andreas; HAENLEINB, Michael. Siri, Siri, in myhand: Who'sthefairest in theland? Ontheinterpretations, illustrations, andimplicationsof artificial intelligence. [*S. l.*]: **ScienceDirec**t, 6 nov. 2018. Disponível em: https://www.sciencedirect.com/science/article/pii/S0007681318301393. Acesso em: 31 out. 2019.

RUSSELL, Stuart; NORVIG, Peter. Inteligência Artificial. [*S. l.*]: **Elsevier**, 2013.

VASCONCELOS, V.V.; MARTINS JUNIOR, P.P. **Protótipo de Sistema Especialista em Direito Ambiental para Auxílio à decisão em Situações de Desmatamento Rural**. NT-27. CETEC-MG. 2004. 80p.