VOZES DIVERSAS
DIFERENTES SABERES

CONHECIMENTO FORMAÇÃO INOVAÇÃO
Salão UFRGS 2018

SALÃO DE
INICIACAO CIENTIFICA
XXX SIC

15 A 19
OUTUBRO
CAMPUS DO VALE

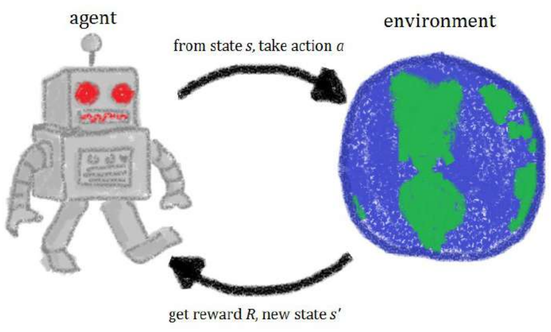
UFRGS
PROPESQ

# Blind Agent Playing Games Through Deep Reinforcement Learning

Aluno: Leonardo B. Duarte
Orientador: Luís C. Lamb

## Introdução

O aumento do poder computacional aliado a enorme quantidade de dados disponíveis nos últimos anos permitiram que algoritmos baseados em aprendizado de máquina fossem aplicados à tarefas antes consideradas complexas computacionalmente. Classificação de imagens, geração de textos e tradução de idiomas são alguns exemplos. A principal técnica utilizada nessas aplicações é a aprendizagem profunda (do inglês Deep Learning), que consiste em criar abstrações em alto nível dos dados se inspirando na interpretação do processamento de informações e padrões de comunicação de um sistema nervoso. Para construir um algoritmo de inteligência artificial geral, capaz de realizar múltiplas tarefas (como o cérebro humano), a aprendizagem profunda é utilizada dentro de um contexto de aprendizagem por reforço (do inglês Reinforcement Learning). Aprendizado por reforço é um paradigma de aprendizado de máquina no qual o agente (algoritmo) interage com um ambiente tentando maximizar um sinal numérico.




*Esquema representando a ideia do aprendizado por reforço.*

## Problema

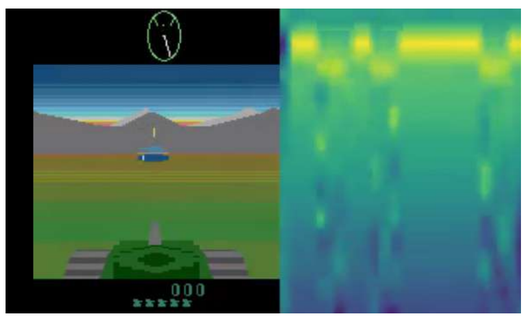
Parte da limitação de algoritmos atuais de inteligência artificial reside na grande dimensionalidade dos dados, exigindo um poder computacional cada vez mais alto para ser possível realizar tarefas complexas. O principal meio de avaliação de algoritmos de aprendizado por reforço utilizado atualmente são jogos de videogame, que alimentam o algoritmo com a tela em cada passo. Em jogos atuais, essa tela consiste em 1920x1080 pixels onde cada pixel possui uma codificação RGB indicando sua cor. Essa quantidade de dados inviabiliza qualquer tipo de teste. Contudo, nem todos os jogos precisam da tela para se obter uma boa pontuação. Muitos jogos atuais são jogados por pessoas que possuem deficiência visual e que se utilizam apenas do áudio como informação.

## Objetivo

Este trabalho tem como objetivo implementar algoritmos recentes de aprendizagem profunda no contexto de aprendizado por reforço, modificando-o para que aceite áudio como entrada.

## Metodologia

Para a realização do trabalho, foi necessário buscar ferramentas que permitissem aplicar o mesmo algoritmo em jogos diferentes além de uma fácil integração de áudio. O Arcade Learning Environment (ALE) é uma plataforma usada por pesquisas da área e que integra jogos do console Atari 2600, lançado em 1977. Considerando a capacidade limitada de áudio do console da plataforma disponível, Battlezone (lançado em 1980) foi o jogo escolhido para o teste. O jogo consiste em um ambiente 3D no qual o jogador controla um tanque de guerra e, baseado no radar, deve guiá-lo destruindo tanques inimigos e desviando dos seus tiros. Os sinais sonoros envolvem um apito toda vez que o radar passa por um inimigo, o som dos tiros (tanto do jogador quanto do inimigo) e o som quando um tanque é destruído.



*Na esquerda, a tela do jogo e na direita, um espectrograma do sinal do áudio.*

## Algoritmo

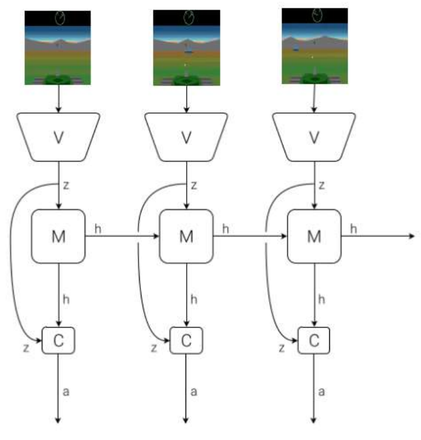
O algoritmo implementado inicialmente possui 3 módulos, que podem ser treinados individualmente utilizando observações geradas por um comportamento aleatório durante o jogo. O primeiro módulo V é um Variational Autoencoder (VAE) implementado através de redes neurais convolucionais. A ideia é criar uma representação da imagem que contenha apenas os detalhes necessários, assim reduzindo a complexidade do algoritmo. O VAE é treinado utilizando apenas observações do jogo.

*A saída do VAE devolve uma imagem que abstrai detalhes desnecessários, capturando apenas a essência da imagem.*

O segundo módulo M consiste em uma rede neural recorrente com uma arquitetura LSTM (Long Short-term Memory) e é responsável por criar um modelo preditivo do jogo. Dada uma observação, uma ação e a sua memória interna, o modelo M devolve uma distribuição de probabilidade da próxima observação. Dessa forma, mesmo sem ter chegado em estados futuros, a memória possui informações sobre possíveis observações futuras. Este módulo é treinado utilizando sequências de observações e ações geradas por um comportamento aleatório do jogo.

Por último, o módulo C consiste em um modelo linear simpels responsável por escolher a ação a ser tomada dada a observação atual e a memória interna. Isso significa que a complexidade do modelo está principalmente nos módulos V e M enquanto que C possui poucos parâmetros possíveis.

Essa simplicidade permite utilizar métodos não convencionais para treinar C. O método utilizado foi o CMA-ES (Covariance Matrix Evolution Strategy) que consiste em uma estratégia evolutiva que seleciona os melhores parâmetros a cada geração e gera novas soluções a partir desses parâmetros.




*Esquema que representa os três módulos e como eles interagem entre si.*

## Áudio

Para incorporar o som no modelo atual, substituímos uma das entradas no módulo M. Retiramos a observação da tela e adicionamos o áudio da tela atual. Assim, a cada passo, o modelo terá que realizar a sua predição do estado futuro utilizando o sinal de audio.

## Resultados

Os resultados até o momento sugerem que o modelo não irá convergir em uma boa pontuação. As limitações de áudio do console resultam em um sinal muito esparso e de difícil assimilação por parte do modelo. Outro fator que pode ser o problema é a grande complexidade do jogo. Sem muitas ferramentas para a integração de áudio, a escolha de ambientes para teste acabam limitadas a poucos jogos. Uma grande dificuldade é o poder computacional necessário para rodar os experimentos. Além de ser um processo lento, uma grande quantidade de memória é necessária para gerar os dados utilizados durante o treinamento.

## Principais Referências

David Ha and Jürgen Schmidhuber. World models, CoRR, abs/1803.10122, 2018.

Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. Nature, 521(7553):436–444, 2015.

Marlos C. Machado, et al. Revisiting the arcade learning environment: Evaluation protocols and open problems for general agents, CoRR, abs/1709.06009, 2017.

Volodymyr Mnih, et al. Playing atari with deep reinforcement learning, arxiv:1312.560 NIPS Deep Learning Workshop, 2013.

Volodymyr Mnih, et al. Human-level control through deep reinforcement learning. Nature, 518(7540):529–53, 2015.

Paul J Werbos. Backpropagation through time: what it does and how to do it, Proceedings of the IEEE, 78(10):1550–1560, 1990.