

Proposta de Pesquisa Científica - MSI I

Estruturação de Estados orientada a Game Design em Aprendizado por Reforço

Aluno

Marcelo Augusto Salomão Ganem

Universidade Federal de Minas Gerais

marceloganem@dcc.ufmg.br

Orientador

Prof. Luiz Chaimowicz

Universidade Federal de Minas Gerais

chaimo@dcc.ufmg.br

15 de abril de 2025

Resumo

A escolha de estados representativos em tarefas de aprendizado por reforço é essencial na construção de modelos com comportamentos explicáveis e previsíveis. Essa Proposta de MSI tem por pretensão investigar a qualidade de modelagens de Processos de Decisão de Markov orientadas a aspectos bem-definidos de *game design* na implementação de agentes de IA para jogos. Como justificativa e orientação para a pesquisa, as literaturas de inteligência artificial, aprendizado por reforço e design de jogos servem a base para a construção de modelos coerentes dos problemas comumente enfrentados em jogos.

1 Introdução

Agentes de aprendizado por reforço aprendem e interagem com o mundo por meio de abstrações de ambientes reais. Para o caso particular de jogos, eletrônicos ou não, é frequente a necessidade de modelar mecânicas simples que interagem entre si para produzir comportamentos complexos.

1.1 Tópico da pesquisa

A tarefa de modelar jogos a partir da definição de suas mecânicas de maneira formal foi inicialmente explorada por Dormans (2012), mas ainda não integrada a estratégias de aprendizado de máquina, no geral, ou de aprendizado por reforço, que é o escopo específico do presente trabalho.

Investigamos, portanto, a aplicabilidade de tais modelagens desenvolvidas a partir de regras estritas (ver Seção 2.1) na construção de modelos de aprendizado por reforço capazes de responder coerentemente às estruturas¹ de jogos – idealmente, de maneira generalizada.

1.2 Relevância

As perguntas aqui propostas são caminhos para avançar o conhecimento em aprendizado por reforço, especificamente no domínio da interpretabilidade por meio de modelagens bem estruturadas. Entende-se uma representação de estado em *features* que correspondam a aspectos explicitamente definidos do ambiente como uma ferramenta para produzir agentes cujo comportamento pode ser mais coerentemente explicado.

¹Estruturas lógicas referentes à economia de recursos e condições de vitória.

Ainda, tais questões são relevantes para o desenvolvimento de inteligência artificial em jogos de uma maneira geral. Partindo de um referencial centrado em *game design* – evitando, portanto, miopias particulares de uma perspectiva estritamente computacional, espera-se que o método aqui executado seja replicável na indústria de jogos eletrônicos em geral, bem como na otimização de mecânicas em jogos de tabuleiro.

1.3 Objetivos

O escopo do trabalho é dividido em dois semestres, correspondentes às disciplinas MSI I e II. Assim, definimos os objetivos da **MSI I** como:

- **Objetivo geral:**
 - Responder a *research question*: "Qual é o impacto de providenciar uma representação de estado estruturada para agentes de aprendizado por reforço na convergência e captura de padrões complexos em jogos?"
- **Objetivos específicos:**
 - Implementar simulações de diagramas *Machinations* em código aberto²;
 - Treinar agentes de aprendizado por reforço via Q-Learning e DQNs, utilizando um modelo observacional estruturado com base em grafos de *Machinations*.
 - Investigar a performance e convergência dos modelos desenvolvidos.

Observa-se que a expectativa não é contemplar todos os detalhes da sintaxe proposta por Dormans. Para a **MSI 2**, pretende-se:

- **Objetivo geral:**
 - Responder a *research question*: "A utilização de *Graph Neural Networks* em combinação com representações de estado estruturadas pode auxiliar na generalização do aprendizado de padrões em jogos?"
- **Objetivos específicos:**
 - Implementar GNNs capazes de processar o estado do ambiente formatado como um grafo.

²Disponíveis à época da publicação, hoje são propriedade intelectual da empresa de mesmo nome. O trabalho original, entretanto, está sob a *Creative Commons License*.

- Processar definições de jogos na sintaxe *Machinations* com entradas de tamanho variável.
- Aferir a capacidade de generalização de modelos de *Deep Reinforcement Learning* que interpretem diagramas *Machinations* diretamente como grafos.

2 Referencial Teórico

2.1 Engineering Emergence

A sintaxe *Machinations* (Dormans, 2012) é introduzida por Dormans como parte da sua tese de doutorado "*Engineering Emergence: Applied Theory for Game Design*". Seu elemento central é a definição de uma gramática de elementos para a composição de um grafo cuja premissa é representar o estado da economia interna de um jogo.

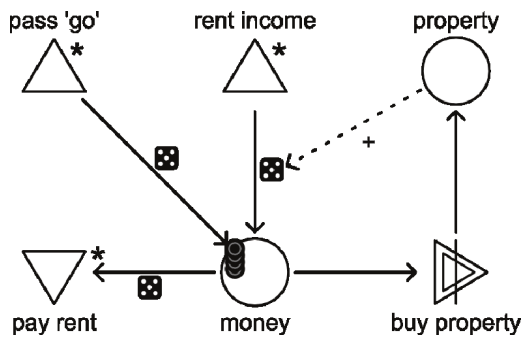


Figura 1: Representação do jogo *Monopoly* (1935) sob a sintaxe *Machinations*

Dormans define essa economia interna como a alocação dos recursos (representados na Figura 1 como fichas em cima de cada nó) entre os nós. O estado dessa economia é alterado a partir do andamento discreto do tempo e ações do jogador ou de quaisquer agentes presentes no sistema. O caráter dessas transições, i.e., o estado atingido em seguida, é ditado pelas arestas do grafo, que indicam transferência regular, condicional ou probabilística de recursos.

2.2 Reinforcement Learning

O modelo proposto por Dormans é análogo em vários aspectos a um Processo de Decisão de Markov (MPD), como introduzido por Sutton e Barto – que não inventaram o conceito, mas cuja literatura serve como referência para essa e outras definições fundamentais no contexto de aprendizado por reforço (Sutton e Barto, 2018).

Em um MDP, temos um agente que interage com um ambiente cujo estado se altera com o tempo e com as ações desse agente – ou de outros agentes, em alguns casos. O objetivo é maximizar a recompensa recebida para cada um dos estados atingidos a partir das ações tomadas. Semelhantemente, em um jogo modelado na sintaxe *Machinations*, o objetivo é direcionar a economia interna para um estado terminal onde o jogador vença o jogo ou continuamente estabeleça uma economia favorável a si.

Assumindo que a economia interna de um jogo possua a *Markov property*, i.e., que seus estados possam ser inferidos sequencialmente sem dependência de qualquer estado além do antecessor, podemos aplicar a maioria das técnicas introduzidas por Sutton e Barto em "*Reinforcement Learning: An Introduction*". Naturalmente, recorre-se à aproximação em casos práticos – assim, desde que o estado possa ser razoavelmente inferido sob o modelo de um MDP, o escopo de estratégias de *Reinforcement Learning* ainda se aplica.

2.3 Artificial Intelligence: a modern approach

Um diagrama na sintaxe *Machinations* também revela informação pertinente às principais taxonomias para desafios em inteligência artificial propostas por Russell e Norvig (2020). Uma análise estática do tipo de cada um dos nós e arestas do grafo revela se o ambiente a ser modelado é determinístico ou estocástico, diferenciando inclusive ambientes estacionários de não-estacionários. Semelhantemente, essa análise revela se o ambiente inclui outros agentes, se a interação é competitiva ou colaborativa e se o ambiente é episódico ou sequencial.

Apesar de sua implementação ser adequada para estados contínuos, Dormans decide não cobrir o caso de ambientes denominados contínuos sob a definição de Russell e Norvig. Isso se deve à limitação do trabalho à análise de mecânicas e dinâmicas discretamente representáveis em jogos – que se estende para essa MSI.

3 Metodologia

Propõe-se responder as perguntas introduzidas na Seção 1.3 pela construção de representações de estado em cenários de Aprendizado de Reforço que capturem as informações contidas em um diagrama de *Machinations*. A qualidade dessas representações, mensurada a partir da performance do agente, informa a efetividade da modelagem proposta.

3.1 Simulação da sintaxe *Machinations*

As definições em *Engineering Emergence* permitem a simulação dos diagramas como uma máquina de estado cujos comportamentos são determinados pelos tipos de arestas, tipos de nós e conexões entre eles. Semelhantemente, pode-se implementar o modelo proposto como um MDP que se altera com a passagem do tempo em intervalos discretos.

Essa etapa consiste de implementar as funcionalidades de diagramas *Machinations*, tal que o estado – i.e., o alocamento dos recursos entre os nós – se altere apropriadamente a partir de ações e tempo conforme as definições da sintaxe.

3.2 Modelagem de jogos sob a sintaxe *Machinations*

Com diagramas que podem ser consistentemente simulados, a etapa seguinte é modelar jogos como grafos na sintaxe definida – esse será o ambiente onde os agentes serão efetivamente treinados. Jogos comuns e amplamente discutidos na obra de referência são *Monopoly* (1935) e *Settlers of Catan* (1995). A coleção de jogos implementados deve ser representativa da diversidade de ambientes com os quais se espera que os agentes sejam capazes de lidar.

3.3 Baselines

Para comparação, são necessários modelos de aprendizado por reforço que não utilizem representações estruturadas de estado. Assim, essa etapa contempla a implementação de um agente de *Q-Learning* e um agente baseado em *Deep-Q Networks* que sejam *naïve* com relação à estrutura do estado.

3.4 Informed *Q-Learning* Agents

O primeiro experimento implementa agentes de *Q-Learning* que recebem o estado como uma entrada estruturadas. Nós e seus tipos, bem como arestas e seus tipos, serão representadas como *features* para o modelo. Essa solução relativamente simples deve permitir aferir com clareza as implicações da estrutura proposta para o estado.

3.5 Deep-Q Networks

Expandir as soluções de *Q-Learning* com técnicas de Aprendizado Profundo. A expectativa é que soluções como DQNs sejam capazes de capturar padrões mais complexos como *feedback*

loops, múltiplas condições de vitória e aspectos específicos a cada jogo. A exploração de outras técnicas de Aprendizado Profundo faz parte do escopo da MSI II, para o semestre seguinte.

4 Resultados Esperados

Espera-se, especialmente, que agentes de aprendizado por reforço treinados a partir da modelagem *Machinations* de um jogo sejam capazes de identificar estratégias ótimas e condições de vitória para cada jogo analisado. Ainda, com a adoção de técnicas de aprendizado profundo, espera-se que os aprendizados dos modelos sejam generalizados para jogos modelados sob a mesma sintaxe. Tais aspectos qualitativos serão inferidos a partir de métricas que incluem número de episódios até a convergência, taxa de vitória, recompensa média por episódio e recompensa acumulada, analisados comparativamente com o auxílio de visualizações.

5 Etapas e Cronograma

Semana 1 (22/04 – 28/04): Definição da metodologia, alinhamento da proposta e levantamento preliminar da literatura.

Semana 2 (29/04 – 05/05): Revisão de literatura e estudo detalhado dos diagramas *Machinations*.

Semana 3 (06/05 – 12/05): Início da implementação da simulação da sintaxe *Machinations* e desenvolvimento do modelo de máquina de estados.

Semana 4 (13/05 – 19/05): Continuação da implementação, realização de testes preliminares e ajustes na simulação.

Semana 5 (20/05 – 26/05): Modelagem dos jogos (por exemplo, *Monopoly* e *Settlers of Catan*) conforme a sintaxe proposta e integração ao ambiente de simulação.

Semana 6 (27/05 – 02/06): Desenvolvimento dos agentes de *Q-Learning* utilizando a representação de estado estruturada.

Semana 7 (03/06 – 09/06): Testes e ajustes dos agentes de *Q-Learning*; preparação para o primeiro pitch (previsto para 06/06).

Semana 8 (10/06 – 16/06): Implementação e integração de *Deep-Q Networks* (DQNs) para captura de padrões mais complexos.

Semana 9 (17/06 – 23/06): Realização do pitch final (23/06), experimentos finais e análise dos resultados obtidos.

Semana 10 (24/06 – 27/06): Redação e finalização do relatório final da pesquisa.

Referências

- Dormans, Joris (2012). *Engineering Emergence: Applied Theory for Game Design*. Amsterdam: Universiteit van Amsterdam. ISBN: 9461907524, 9789461907523. URL: <https://dare.uva.nl/document/2/102091>.
- Monopoly* (1935). Jogo de tabuleiro referenciado no estudo Engineering Emergence.
- Russell, Stuart J. e Peter Norvig (2020). *Artificial Intelligence: A Modern Approach*. 4th. Pearson. ISBN: 9780134610993.
- Sutton, Richard S. e Andrew G. Barto (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book. ISBN: 9780262039246. URL: <http://incompleteideas.net/book/the-book-2nd.html>.
- The Settlers of Catan* (1995). Jogo de tabuleiro referenciado no estudo Engineering Emergence.