Controle de Variabilidade Estrutural em Grafos para Ambientes de Aprendizado por Reforço

Ismael S.

8 de julho de 2025

Resumo

Este documento detalha o desafio de gerar ambientes procedurais baseados em grafos para agentes de Aprendizado por Reforço (RL) com memória. O objetivo central é desenvolver um gerador de mapas que possua *variabilidade estrutural controlada*, evitando que os agentes explorem padrões previsíveis no design do ambiente, em vez de desenvolverem estratégias de aprendizado generalizáveis. Apresentamos a definição do problema, exemplos de cenários indesejados e o estado final almejado, que consiste em um sistema de geração de grafos com parâmetros ajustáveis para a frequência e distribuição de pontos de decisão (bifurcações) e corredores (afunilamentos).

1 Contexto do Projeto

O projeto consiste no desenvolvimento de um **ambiente de simulação** destinado a servir como *benchmark* para agentes de Inteligência Artificial treinados via **Aprendizado por Reforço (RL)**, com foco em agentes dotados de memória.

- Agentes de RL aprendem por meio de tentativa e erro, otimizando seu comportamento com base em um sinal de recompensa.
- Um *benchmark* robusto deve avaliar a capacidade **genuína** de aprendizado e adaptação do agente, e não sua habilidade de explorar falhas ou padrões previsíveis no design do ambiente.

2 O Ambiente: Mapas como Grafos Direcionados

O ambiente é constituído por **mapas de salas** gerados proceduralmente a cada nova sessão de treinamento. Essa geração procedural garante que o agente não possa simplesmente memorizar um mapa estático. A estrutura do mapa é formalmente um **Grafo Acíclico Dirigido (DAG)**.

- Cada **sala** é um **vértice** (nó) do grafo.
- Os **caminhos** entre as salas são as **arestas** (conexões) direcionadas.
- A progressão do agente é sempre unidirecional (sem ciclos), caracterizando uma estrutura *top-down*.

Com base na estrutura de saídas de cada nó, definimos dois tipos de vértices cruciais para a topologia do grafo:

- **Bifurcação:** Vértice com grau de saída (out-degree) maior que 1. Representa um ponto de decisão para o agente.
- **Afunilamento:** Vértice com grau de saída igual a 1. Representa um caminho único, sem escolha de rota.

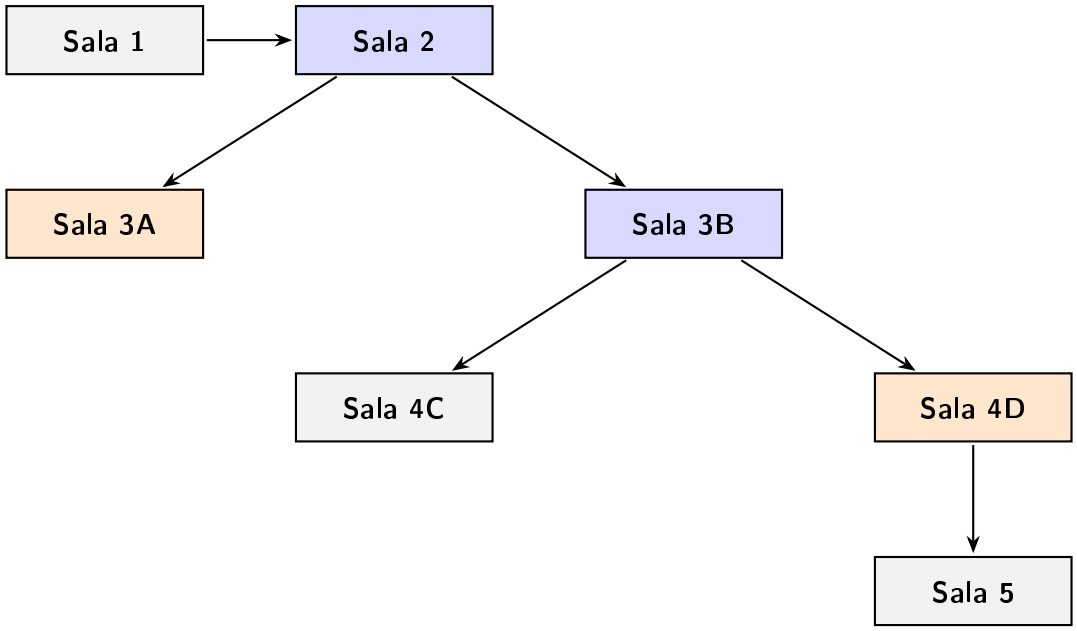


Figura 1: Exemplo da estrutura de um mapa, ilustrando **bifurcações** (nós azuis) e **afunilamentos** (nós laranjas).

3 O Problema: Previsibilidade Estrutural

Agentes de RL são proficientes em encontrar e explorar padrões. Se a geração procedural dos mapas resultar em grafos com uma **estrutura topológica previsível**, o agente pode desenvolver "atalhos"estatísticos, em vez de aprender uma política de decisão robusta.

Isso compromete a validade do *benchmark*, pois o agente estaria explorando uma falha no gerador do ambiente, e não demonstrando inteligência generalizável.

3.1 Cenário Problemático: Padrões Fixos

Considere um gerador que, embora produza grafos diferentes, sempre posiciona os afunilamentos em níveis de profundidade específicos. O agente poderia aprender a explorar essa regularidade, por exemplo, antecipando a ausência de escolhas em determinados pontos da jornada.

4 Objetivo: Variabilidade Estrutural Controlada

O objetivo é projetar um gerador de grafos que permita um **controle preciso sobre a variabilidade estrutural** do ambiente. A meta não é a aleatoriedade total, que poderia resultar em mapas insolúveis ou triviais, mas sim uma **imprevisibilidade gerenciável**.

Busca-se um sistema onde seja possível parametrizar a geração do grafo, ajustando, por exemplo:

- A **probabilidade** de um novo nó ser uma bifurcação ou um afunilamento.
- A **distribuição** desses tipos de nós ao longo do grafo (ex: mais bifurcações no início e mais afunilamentos no final).
- A **dependência contextual**, onde a probabilidade de criar um afunilamento pode depender do tipo do nó anterior.

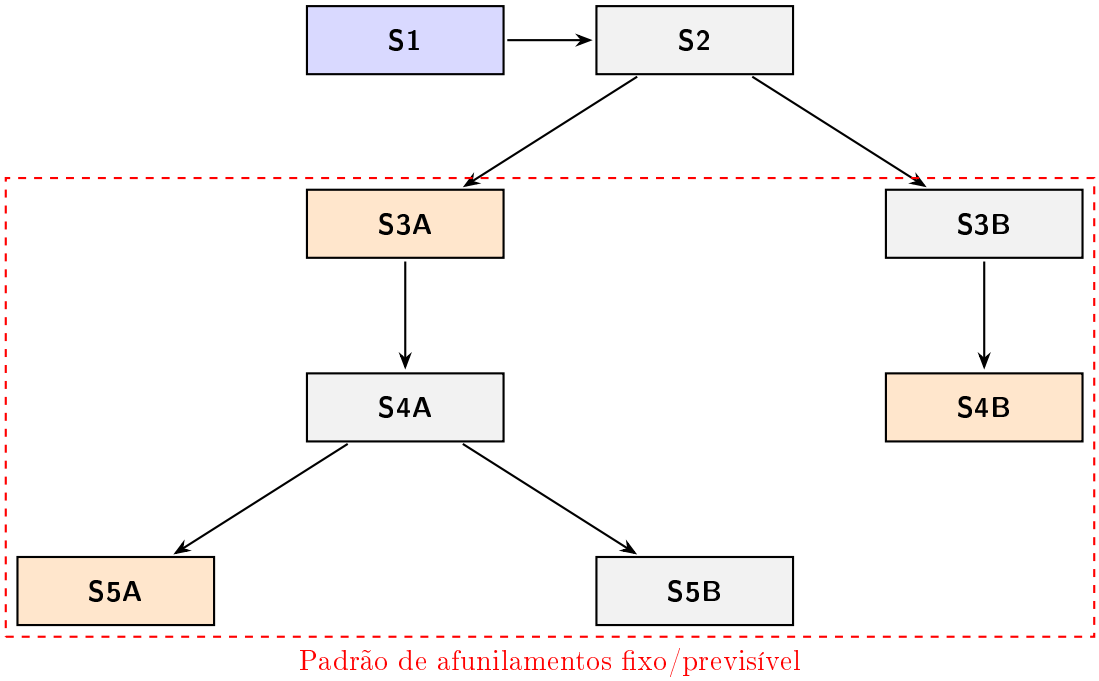
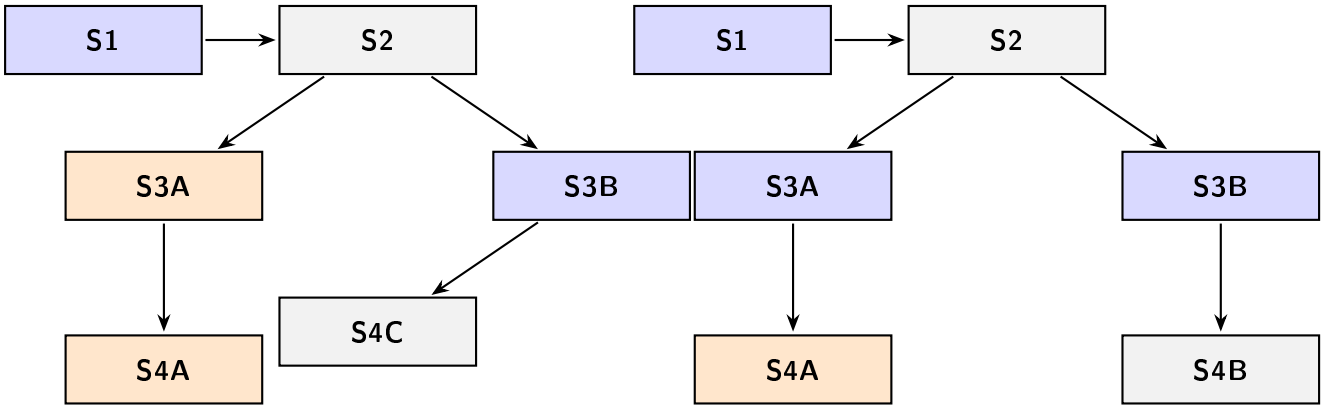


Figura 2: Cenário indesejado onde a localização dos afunilamentos segue um padrão previsível, permitindo a exploração pelo agente.

4.1 Cenário Desejado: Gerações Diversificadas

O gerador ideal deve ser capaz de produzir, a partir do mesmo conjunto de parâmetros probabilísticos, uma ampla gama de topologias de mapa. Isso força o agente a desenvolver estratégias que se adaptem à estrutura específica de cada mapa gerado, em vez de memorizar um padrão global.



(a) Geração 1: Menos bifurcações. (b) Geração 2: Mais bifurcações.

Figura 3: Cenário desejado: O gerador produz mapas com topologias distintas, forçando o agente a se adaptar em vez de memorizar.

5 Conclusão

A questão central de pesquisa é, portanto, a formulação de um algoritmo de geração procedural de grafos que equilibre ordem e caos. O sistema deve prover um nível de **imprevisibilidade estrutural** suficiente para impedir a exploração de padrões, mas com **controle paramétrico**

para garantir que os ambientes gerados permaneçam solucionáveis e propícios ao aprendizado. Acredito que uma abordagem baseada em probabilidades condicionais e distribuições ajustáveis seja um caminho promissor para alcançar este objetivo.

Atenciosamente,
Ismael S.