



<b>Evento</b>	Salão UFRGS 2020: SIC - XXXII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2020
<b>Local</b>	Virtual
<b>Título</b>	Modelos Composicionais de Predição em Aprendizado por Reforço
<b>Autor</b>	GIOVANNA LAZZARI MIOTTO
<b>Orientador</b>	BRUNO CASTRO DA SILVA

## **Modelos Composicionais de Predição em Aprendizado por Reforço**

Giovanna Lazzari Miotto (Prof. Bruno Castro da Silva)

Universidade Federal do Rio Grande do Sul

Aprendizado por reforço (RL) é um paradigma de aprendizado de máquina que permite com que agentes aprendam através de interações com um ambiente, corrigindo seus comportamentos com base em sinais de recompensa. É útil que tais agentes consigam prever a dinâmica de aspectos diversos de seus ambientes. Isso pode ser feito através de General Value Functions (GVFs): modelos especializados em predições sobre aspectos particulares do mundo. Ao se combinar GVFs, obtém-se modelos que descrevem como o ambiente, de forma integral, responde a ações. Considere um robô que ora caminha na grama, ora no gelo. Alguns elementos do ambiente (gravidade) são idênticos; outros (e.g., atrito) são diferentes e devem ter suas respectivas GVFs retreinadas. Redes de GVFs facilitam a transição de aprendizado entre ambientes, aproveitando o conhecimento já adquirido. Propomos, aqui, instanciar famílias de GVFs para descrever ambientes com características que mudem temporalmente. Esse modelo deve conseguir identificar conjuntos de GVFs que passem a fazer predições inválidas. Desenvolvemos inicialmente um simulador realístico de um robô multidirecional operando em um cenário contínuo com obstáculos. Seu objetivo é atingir um alvo não-estacionário. Utilizamos o simulador de física Box2D, o qual modela realisticamente características tais como torques e inércia. Criamos também módulos para otimizar o comportamento do robô. Tais modelos também permitem que sejam modificados, ao longo do tempo, aspectos físicos da simulação, como massas e coeficientes de atrito. Isso permite simular a ocorrência de defeitos—características temporalmente variáveis do ambiente. Os resultados obtidos mostram que o sistema consegue aprender comportamentos ótimos em ambientes diversos. Como próximo passo, iremos inserir falhas no robô e estender o sistema para que substitua GVFs cujas predições estejam em falha. Espera-se, com isso, construir um modelo que se adapte mais rapidamente a novos ambientes, comparado com sistemas (usuais na literatura) que modelam o ambiente de forma monolítica.