

Bom dia a todos. No dia de hoje vamos falar sobre um tema que tem vindo a crescer imenso ao longo dos anos, e da logística por detrás de todo o processo inteligente e de adição humana.

Primeiramente, a escolha deste paper recai sobre o interesse mútuo da minha parte e do Rafael acerca da robótica e a tecnologia que nela está envolvida. Contudo, e a questão que mais nos puxou ao desejo de explorar este tema é: de que maneira os robôs adaptam-se ao comportamento humano?

De seguida iremos não só realçar pontos importantes para o processo de entendimento deste paper, como também explica-los devidamente para que os mesmos sejam compreendidos por vocês. Os pontos são os seguintes:

- robôs necessitam de se adaptar à preferência humana, o que implica alterações à sua função de recompensas

- um robô deve pedir por dados que explicitamente lhe ensinam sobre o que lhe falta, e não aprender implicitamente através de ações otimizadas.

- existe a introdução de um novo tipo de input humano chamado traços de recurso(feature traces)

- e é possível recuperar recompensas mais generalizáveis com menos entrada humana, ao invés de ter uma linha de aprendizagem através de demonstrações

Convenhamos que temos o exemplo de um humano a observar um robô que segurar uma caneca a uma distância da mesa muito grande(o que prova não ser muito realista quando em comparação com a realidade). Este imediatamente percebe que obtém um resultado aquém do desejável ou que esperava observar.

Desta maneira, o humano puxa o braço para baixo e trabalho antigo mostra como o robô consegue interpretar essa introdução humana e adapta o seu entendimento de recompensa em dados.

Portanto, e como vimos anteriormente, robôs possuem uma função de recompensa e esta função pode ser alterada através de correções. Neste caso, o robô pode otimizar esta nova função de recompensa para realizar a tarefa de uma maneira que se aproxima do desejo da pessoa.

Isto funciona bastante bem quando os recursos de espaço do robô carregam tudo o que a pessoa possivelmente quer saber ou deseja. Mas e se esse não for o caso, como geralmente não o é no mundo real?

Por exemplo, no caso de querermos que o robô se distancie de nós, eles não têm um recurso para isso. Podes empurrá-lo para longe de ti mas, no máximo, o robô reconhece que não tem as características necessárias para interpretar o ajuste humano.

Perante estas faltas, surgiu duas maneiras de testar este aprendizado de recursos necessários. inclui:

- Utilizadores reais que conhecem o funcionamento do robô e as suas capacidades, assim como um manipulador real de robô.

- Utilizador online que recria o ambiente de laboratório do simulador.

Consoante isto, foi possível criar-se dois grafos (seis recursos estudados):

1. Qualitativo:

- a. True(o real) e o de baixo o encenado -> olhando para ambas as imagens, podemos perceber que existe uma grande proximidade entre os resultados obtidos pelos casos reais e os simulados no que toca à recompensa . Quanto mais próximo do computador, pior é a recompensa. Quanto mais próximo da mesa, melhor a recompensa. As zonas mais clareadas representam a zona do computador(daí a recompensa ser menor). O objetivo recai então em manter distancia do computador mas próximo da mesa, causando uma maior segurança e semelhança em relação à ação.

2. Quantitativo:

- a. Para os seis recursos estudados, foi-se testando até 10 vezes o mesmo recurso até que se conclui que: com um número adequado de dados, FERL aprende muito melhor os recursos do que de forma random e torna-se menos sensível a uma intervenção/entrada humana (comparação da primeira barra laranja à última), diminuindo portanto os erros. Traços a cinzento representam uma média dos erros quando feito de forma random.