Unyleya
EDUCACIONAL

APRENDIZAGEM POR REFORÇO INVERSA

APRENDIZAGEM POR REFORÇO INVERSA

Conceito

Aprendizagem por reforço, o objetivo é aprender um processo de decisão para produzir um comportamento que maximize alguma função de recompensa predefinida.

Aprendizado por Reforço Inverso (IRL), como descrito por Russell e Norvig (2010), aborda o problema e tenta extrair a função de recompensa do comportamento observado de um agente.

APRENDIZAGEM POR REFORÇO INVERSA

Exemplo - dirigir um carro autônomo

Uma abordagem simples seria criar uma função de recompensa que capte o comportamento desejado de um motorista: parar nos semáforos, ficar fora da calçada, evitar pedestres e assim por diante.

Infelizmente, isso exigiria uma lista exaustiva de todos os comportamentos que gostaríamos de considerar, além de uma lista de pesos descrevendo a importância de cada comportamento (imagine ter que decidir exatamente o quanto os pedestres são mais importantes do que os sinais de trânsito!).

APRENDIZAGEM POR REFORÇO INVERSA

Exemplo - dirigir um carro autônomo

Na estrutura da IRL, a tarefa é pegar um conjunto de dados de condução gerados pelo homem e extrair uma aproximação da função de recompensa desse humano para a tarefa.

É claro que essa aproximação necessariamente lida com um modelo simplificado de direção. Grande parte da informação necessária para resolver um problema é capturada dentro da aproximação da função de recompensa verdadeira.

Segundo Russell e Norvig (2010), a **função de recompensa**, e não a política, é a **definição mais sucinta**, **robusta** e **transferível da tarefa**, já que quantifica quão boas ou ruins certas ações são.

Uma vez que temos a função de recompensa correta, o problema é encontrar a política correta, podendo ser resolvido com métodos de aprendizado por reforço padrão

APRENDIZAGEM POR REFORÇO INVERSA

Exemplo - dirigir um carro autônomo

Neste caso, estaríamos usando dados de condução humana para aprender automaticamente os pesos de recursos certos para a recompensa.

A tarefa é descrita completamente pela função de recompensa, nem precisamos saber os detalhes da política humana, desde que tenhamos a função de recompensa correta para otimizar.

Os algoritmos que resolvem o problema da IRL podem ser vistos como um método para alavancar o conhecimento especializado para converter uma descrição de tarefa em função de recompensa compacta.

APRENDIZAGEM POR REFORÇO INVERSA

Alguns problemas

O principal deles é converter uma tarefa complexa em uma função de recompensa simples, pois determinada política pode ser ideal para muitas funções de recompensas diferentes.

Ou seja, mesmo que tenhamos as ações de um especialista, existem muitas funções diferentes de recompensas que o especialista pode estar tentando minimizar.

Algumas dessas funções são bem simples: todas as políticas são ideais para a função de recompensa que é zero em todos os lugares, portanto, essa função de recompensa é sempre uma possível solução para o problema da IRL.

Porém, é necessário que a função de recompensa capture informações significativas sobre a tarefa e seja capaz de diferenciar claramente entre políticas desejadas e não desejadas.

APRENDIZAGEM POR REFORÇO INVERSA

Diante dos problemas

Russell e Norvig (2010) formularam o aprendizado de reforço inverso como um problema de otimização, onde se quer escolher uma função de recompensa para a qual a política especializada especificada é a ideal. Mas, dada essa restrição, também se quer uma função de recompensa que maximize adicionalmente certas propriedades importantes.

APRENDIZAGEM POR REFORÇO INVERSA

Referências

RUSSELL, S.; NORVIG, P. **Artificial Intelligence:** A Modern Approach. 3.ed. New Jersey: Pearson Education, 2010.

Obrigada!

hulianeufrn@gmail.com