Unyleya
EDUCACIONAL

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Quatro elementos importantes para que ocorra a interação dinâmica do sistema:

- Política de ação
- Função de reforço
- Função de valor
- Modelo do Ambiente

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Política de ação

Mapeia o estado percebido do ambiente pelo agente para a ação a ser executada nesse estado, maximizando a satisfação dos seus objetivos.

A seleção da ação do agente pode ser modelada como um mapa de políticas:

$$\pi: S * A \rightarrow [0,1]$$
$$\pi(a \mid s) = P(= a \vDash s)$$

O mapa de políticas fornece a probabilidade de escolher uma ação a quando estiver no estado *s*.

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Função de reforço

Mapeia estados do ambiente ou transição do ambiente de um estado para um outro para um número indicando a satisfação *imediata* dos objetivos do agente *nesse* estado ou no estado resultando da transição.

Em cada espaço de tempo o ambiente envia para o agente de aprendizagem por reforço um valor, um número correspondente a uma recompensa.

O sinal recebido de cada recompensa define quais ações são boas ou ruins para o agente.

A recompensa enviada depende da ação do agente e do estado do ambiente em que o agente se encontra.

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Função de valor

Mapeia o estado do ambiente para um número indicando a satisfação futura atingível dos objetivos do agente a partir desse estado.

As **funções de valores**, ao contrário das **funções de reforço,** indicam o que é bom para o sistema em longo prazo.

O que a função faz é garantir totalmente a recompensa que um agente espera acumular a partir daquele estado.

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Função de valor

A função de valor $V\pi$(S) é definida como um retorno esperado que começa no estado *s*, ou seja, s_0= *s*, e segue sucessivamente até a política π. Assim sendo:

$V\pi(s) = E[R] = E[\sum_{t=0}^{\infty} y^t \, r_t \mid S_0 = S]$, onde a variável randômica R, detona o retorno e é definida como o somatório que será descontado no futuro;

Sendo, $R = \sum_{t}^{\infty} y^t r_t$, onde r_t é a recompensa no passo t, $y \in [0, 1]$ é percentual de desconto.

ELEMENTOS PARA UMA APRENDIZAGEM POR REFORÇO

Modelo do ambiente

Cada sistema de aprendizagem por reforço aprende um mapeamento de ações por meio de tentativa e erro com um ambiente dinâmico.

- **Modelo perceptivo**, mapeia percepções para representação interna do estado do ambiente

Exemplo : equipamentos com sensores que fazem a leitura do ambiente, descrição de símbolos ou processos mentais, como exemplo, a sensação de estar perdido em um lugar desconhecido e pensar em como sair daquele lugar, ou seja, mapeiam as percepções internas do estado do ambiente

- **Modelo efetivo**, mapeia ação a executar para representação interna do estado do ambiente. No modelo de sistema.

Obrigada!

hulianeufrn@gmail.com