APRENDIZAGEM POR REFORÇO PARA ESPAÇOS FINITOS

APRENDIZAGEM POR REFORÇO PARA ESPAÇOS FINITOS

Considere o conjunto de políticas ótimas para um Processo de Decisão de Markov (MDP) (que será visto na Unidade II) como um espaço de estado finito S, um conjunto de ações A e matrizes de probabilidade de transição P_a para cada ação.

Russell e Norvig (2010) provaram que uma política π dada por $\pi\ (s)\ =\ a_1$ é ideal se, e somente se, para todas as outras ações , o vetor de recompensa R (que lista as recompensas para cada estado possível) satisfaz a condição:

$$(P_{a_1} - P_a)\,(I - \gamma P_{a_1})^{-1}\, R \geq 0$$

APRENDIZAGEM POR REFORÇO PARA ESPAÇOS FINITOS

A importância desse teorema é que ele mostra que podemos selecionar de forma eficiente a "melhor" função de recompensa para a qual a é política ótima, usando algoritmos de programação linear.

Apenas um estado é realmente recompensado.

Isso permitiu que diversas literaturas tratem a IRL como um problema de otimização tratável, em que se tenta otimizar as heurísticas a seguir para que a função de recompensa se "encaixe" bem nos dados do especialista.

APRENDIZAGEM POR REFORÇO PARA ESPAÇOS FINITOS

Maximize a **diferença** entre a **qualidade da ação ótima** e a **qualidade da próxima melhor ação** (sujeita a um limite na magnitude da recompensa, para evitar diferenças arbitrariamente grandes). Logo, se quer encontrar uma política ótima que se distingue claramente de outras políticas possíveis.

Minimize o tamanho das recompensas na função/vetor de recompensa. A intuição é que o uso de pequenas recompensas incentiva a função de recompensa a ser mais simples, semelhante à regularização na aprendizagem supervisionada.

Russell e Norvig (2010) escolhem a norma L_1 com um coeficiente de penalidade ajustável, que incentiva o vetor de recompensa a ser diferente de zero em alguns estados.

APRENDIZAGEM POR REFORÇO PARA ESPAÇOS FINITOS

Referências

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

Unyleya EDUCACIONAL

Obrigada!

hulianeufrn@gmail.com