

6 Conclusões

A realização de testes do algoritmo de aprendizado por reforço sobre o problema de revisitação de páginas web, representadas por um conjunto de páginas da Wikipédia, demonstrou que é possível utilizar esse método para aprender como controlar o agendamento de visitas a essas páginas.

Foi possível, utilizando valores decrescente no tempo de taxas de aprendizado e exploração, aprender uma política mais eficiente que as políticas clássicas de revisitação. Isso confirma que a técnica de aprendizado por reforço é útil no problema de revisitação de páginas web.

Ficou evidente, no decorrer do trabalho, que é muito complicado encontrar boas características de modelagem do estado do sistema. E, mesmo após determinadas essas características, também é trabalhoso realizar o mapeamento da característica para um valor real que possa ser utilizado pela função de aprendizado.

Focando especificamente no problema de limitação física do sistema, que considera a quantidade máxima de páginas a ser atualizada por unidade de tempo, foi possível concluir que é melhor aumentar a quantidade de vezes que o sistema determina as páginas que devem ser atualizadas do que aumentar a quantidade de páginas que devem ser atualizadas a cada vez. Quando o intervalo entre decisões é maior, a quantidade de páginas que o sistema deve escolher para atualizar fica maior (respeitando a limitação física do sistema de N páginas por unidade de tempo). Ao escolher um grupo para ter suas páginas atualizadas, corre-se o risco de escolher um grupo com uma quantidade de páginas menor que a quantidade que o sistema pode atualizar – causando desperdício de recursos. Uma boa alternativa encontrada nesse trabalho foi aumentar a quantidade de oportunidades de tomada de decisão, e diminuir proporcionalmente a quantidade de páginas a

serem atualizadas, diminuindo assim a chance de desperdiçar capacidade computacional.

O cérebro do algoritmo é a função que representa o sistema. As características de modelagem do estado, por melhores que sejam, precisam ser bem utilizadas pela função. Isso ficou claro quando foi utilizada uma função linear para representar $Q(s)$: o aprendizado era impraticável. A função de regressão logística demonstrou ser extremamente mais poderosa para servir como abstração do comportamento do sistema.

A função de regressão logística é também conhecida como *perceptron*, ou rede neural artificial de uma camada. O fato de uma rede neural de uma camada ter sido capaz de alcançar índices comparáveis aos das políticas clássicas de revisão sugere que talvez seja possível, utilizando-se redes neurais de maior número de camadas, alcançar índices ainda melhores. É o que observamos quando vemos casos como o do TD-Gammon, apresentado em [TES].

Outra maneira de melhorar os resultados atingidos pelo algoritmo de aprendizado por reforço é utilizar um conjunto ampliado de características para representar o estado do sistema. As características adicionais podem agregar conhecimento à função, e tornar assim o algoritmo mais eficiente.