



**Data Science
Academy**

www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Algoritmo HITS

O algoritmo Hyperlink-Induced Topic Search, também conhecido como “hubs e autoridades” ou HITS, é outro algoritmo influente na análise de links. O HITS difere do PageRank de várias maneiras. Primeiro, é uma medida dependente de consulta: classifica páginas de acordo com uma consulta. Isso significa que deve ser calculado de novo para cada consulta — um fardo computacional que a maioria dos mecanismos de busca tem optado por não assumir. Dada uma consulta, o HITS primeiro encontra um conjunto de páginas que são relevantes para a consulta, pela interseção de listas de hits de palavras de consulta e, em seguida, adiciona as páginas nos links de vizinhança dessas páginas — páginas que apontam para/são apontadas de uma das páginas do conjunto original relevante.

```
função HITS (consulta) retorna páginas com hub e números de autoridade
    páginas ← EXPANDIR-PÁGINAS (PÁGINAS-RELEVANTES (consulta))
    para cada  $p$  em páginas faça
         $p.AUTORIDADE \leftarrow 1$ 
         $p.HUB \leftarrow 1$ 
    repetir até a convergência faça
        para cada  $p$  em páginas faça
             $p.AUTORIDADE \leftarrow \sum_i \text{INLINK}_i(p).HUB$ 
             $p.HUB \leftarrow \sum_i \text{OUTLINK}_i(p).AUTHORITY$ 
        NORMALIZAR (páginas)
    retornar páginas
```

Cada página desse conjunto é considerada uma **autoridade** sobre a consulta na medida em que outras páginas do conjunto relevante apontem para ela. A página é considerada um **hub** na medida em que aponta para outras páginas de autoridade no conjunto relevante. Assim como com o PageRank, não queremos apenas contar o número de links; queremos dar mais valor aos hubs e autoridades de alta qualidade. Assim, como com o PageRank, iteramos um processo que atualiza a pontuação de autoridade de uma página para ser a soma das pontuações do hub das páginas que apontam para ela, e a pontuação do hub para ser a soma das pontuações de autoridade das páginas para as quais ele aponta. Se normalizarmos as pontuações e repetirmos k vezes, o processo convergirá.

O PageRank e o HITS desempenharam papéis importantes no desenvolvimento de nossa compreensão da recuperação de informação da Web. Esses algoritmos e suas extensões são usados em bilhões de classificação de consultas diariamente à medida que os mecanismos de busca constantemente desenvolvem melhores formas de extrair sinais ainda mais finos de relevância da busca.

Referências:

Livro: Inteligência Artificial
Autor: Peter Norvig