Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

PageRank da Matriz *Google* Através do Método da Potência

Matheus M. Nery [1]
Bacharel em Matemática/UFF, Volta Redonda, RJ
Marina S. D. de Freitas[2]
VMA/UFF, Volta Redonda, RJ

O método da potência é um algoritmo iterativo utilizado para encontrar o autovalor de maior magnitude (em valor absoluto) e seu autovetor associado de uma matriz quadrada A. Ele funciona tomando algum vetor arbitrário de norma 1 e multiplicando esse vetor por uma matriz A, obtendo, assim, um novo vetor, o qual será normalizado. O processo é repetido até que essa sequência de vetores gerados convirja para um autovetor correspondente ao maior autovalor de A [3]. Na Figura 1, temos o algoritmo desse método.

Método da potência
A uma matriz $n \times n$
$v_0 =$ algum vetor com $\|v_0\| = 1$
$k = 1$
repita
  $w_k = Av_{k-1}$
  $v_k = \frac{w_k}{\|w_k\|}$ (autovetor aproximado)
  $\lambda_k = (v_k)^T Av_k$ (autovalor aproximado)
  $k = k + 1$
até a convergência.

Figura 1: algoritmo da iteração de potência. Fonte: Adaptado de [1].

O método da potência é eficiente para matrizes grandes e esparsas, pois dispensa a fatoração completa da matriz. Uma das aplicações mais conhecidas é o cálculo do *PageRank*, uma métrica criada por Larry Page e Sergey Brin para avaliar a importância de páginas na *web*.

O *PageRank* é um autovetor da matriz do *Google*, que representa um grafo onde os vértices são páginas *web* e as arestas são os *links* entre elas. A relevância de cada página é determinada com base na quantidade e qualidade dos *links* que apontam para ela. Este conceito fundamenta-se na premissa de que a quantidade de *links* direcionados para uma página e oriundos dela fornece informações relevantes sobre a sua importância [2].

No cálculo do *PageRank*, as páginas da *web* são ordenadas de 1 a n, e a matriz Q é definida como:

$$Q_{ij} = \begin{cases} \frac{1}{N_j}, & \text{se há um } \textit{link} \text{ de } j \text{ para } i \\ 0, & \text{caso contrário,} \end{cases} \tag{1}$$

onde N_j é o número de *links* de saída da página j. O *rank* de i, denotado por r_i, é definido de tal forma que, se uma página j, altamente classificada, tiver um *link* de saída para i, isto aumenta a

[1] matheus.mneryy@gmail.com
[2] msdias@id.uff.br

importância de i. A definição preliminar do *PageRank* é:

$$r_i = \sum_{j \in I_i} \frac{r_j}{N_j}, \tag{2}$$

onde I_i é o conjunto de páginas que possuem um *link* para a página i. Esta definição é recursiva, portanto o *PageRank* não pode ser calculado diretamente. Assim, dado $\mathbf{r}_0$ o vetor com o *rank* inicial de cada página, após $k+1$ iterações, temos:

$$r_i^{(k+1)} = \sum_{j \in I_i} \frac{r_j^{(k)}}{N_j}, \text{ para } k = 0, 1, \ldots \tag{3}$$

Esse processo reflete a importância relativa de cada página, com base nos *links* recebidos.

O *PageRank* permite filtrar páginas irrelevantes em pesquisas, priorizando aquelas com maior relevância. O cálculo iterativo, iniciado com um vetor $\mathbf{r}_0$, converge para o autovetor que representa a importância de cada página, demonstrando a eficácia do método da potência em problemas de grande escala. Pode-se observar que a definição (2) é equivalente a multiplicar a linha i da matriz Q pelo vetor $\mathbf{r}$, que contém o *rank* de todas as páginas. Assim, pode-se escrever a equação na forma matricial $\lambda\mathbf{r} = Q\mathbf{r}$, $\lambda = 1$. Portanto, a iteração (3) é equivalente a $\mathbf{r}^{k+1} = Q\mathbf{r}^k$, $k = 0, 1, \ldots$, que corresponde ao método da potência para calcular o autovetor. Assim, queremos resolver o problema de autovalores $A\mathbf{r} = \mathbf{r}$, onde $\|\mathbf{r}\| = 1$. Note que a matriz A é esparsa e sua dimensão é da ordem de bilhões. Para realizar este cálculo, propomos o método da potência.

Para garantir que o *PageRank* esteja bem definido e que existe um autovalor igual a 1, precisamos fazer algumas modificações. A matriz Q é modificada de modo que, se uma página não possui *link* para nenhuma outra, então a coluna de Q composta de zeros é modificada por um valor constante em cada posição. Isto significa que existe igual probabilidade de ir para qualquer página na rede. Assim, Q passa a ter a forma: $P = Q + \frac{1}{n}\mathbf{e}\mathbf{d}^T$, onde $\mathbf{e}, \mathbf{d} \in \mathbb{R}^n$, $\mathbf{e}^T = (1, 1, \ldots, 1)$ e $d_j = 1$, se $N_j = 0$ e 0 caso contrário. A matriz P obtida é uma matriz coluna-estocástica, isto é, ela possui apenas elementos não negativos e os elementos de cada coluna somam 1. Reescrevendo $\lambda\mathbf{r} = Q\mathbf{r}$, obtemos $P\mathbf{r} = \mathbf{r}$. Além disso, para garantir que exista um *link* de toda página da *web* para outra, é feita uma combinação convexa de P e uma matriz de posto 1: $A = \alpha P + (1-\alpha)\frac{1}{n}\mathbf{e}\mathbf{e}^T$, para algum α satisfazendo $0 \leq \alpha \leq 1$. Para a convergência do algoritmo, devemos conhecer como os autovalores de P se alteram após sua modificação, o que é mostrado no Teorema 1.

Teorema 1. *Suponha que os autovalores da matriz coluna-estocástica P sejam $\{1, \lambda_2, \lambda_3, \ldots, \lambda_n\}$. Então os autovalores de $A = \alpha P + (1-\alpha)\frac{1}{n}\mathbf{e}\mathbf{e}^T$ são $\{1, \alpha\lambda_2, \alpha\lambda_3, \ldots, \alpha\lambda_n\}$.*

O Teorema 1 implica que mesmo que P tenha múltiplos autovalores iguais a 1, o que é verdadeiro para a matriz *Google*, o segundo maior autovalor em magnitude de A é sempre igual a α. Uma vantagem adicional é que não precisamos utilizar o vetor $\mathbf{d}$ em nenhum momento, o que significa que não é necessário saber quais páginas não possuem *links* de saída.

Referências

[1] J. W. Demmel. **Applied Numerical Linear Algebra**. 1a. ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1997. ISBN: 9780898713893.

[2] L. Eldén. "Numerical Linear Algebra in Data Mining". Em: **Acta Numerica** 15 (2006), pp. 327–384. DOI: 10.1017/S0962492906240017.

[3] L. N. Trefethen e D. Bau. **Numerical Linear Algebra**. 1a. ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1997. ISBN: 9780898713619.