Teorema do ponto fixo de Banach e o buscador do Google

Luísa Andrade Martins, Ana Paula Tremura Galves (orientadora)

Universidade Federal de Uberlândia

Objetivo

A empresa Google foi fundada em 1998, por dois amigos Lawrence Page e Segey Brin, ambos colegas de doutorado. O problema da tese de Page era estudar a estrutura de links inerente à internet. Essas pesquisas deram frutos e em 1996, Page e Brin lançaram o algoritmo PageRank que permite ranquear as páginas da internet em termos de "importância" de cada uma delas.

Assim nasceu o Google, e o que permitiu a empresa se destacar tão rapidamente de seus concorrentes foi a Matemática, mais especificamente, o Teorema do ponto fixo de Banach, o qual enunciaremos a seguir e utilizaremos para explicar o funcionamento do buscador do Google.

Conceitos preliminares

Apresentaremos as definições e os resultados básicos de espaços métricos que serão adotados ao longo do texto.

Definição

Dado um conjunto $M \neq \emptyset$, seja $d : M \times M \longrightarrow \mathbb{R}_+$ e indiquemos por $d(x, y)$ a imagem de um par genérico $(x, y) \in M \times M$, através da função d. Dizemos que d é **métrica** sobre M se as seguintes condições se verificam para quaisquer $x, y, z \in M$

i) $d(x, y) = 0 \iff x = y$

ii) $d(x, y) = d(y, x)$

iii) $d(x, y) \leq d(x, z) + d(z, y)$ (desigualdade triangular).

Ao par (M, d) chamamos **espaço métrico**.

Definição

Uma sequência (x_n) é de Cauchy num espaço métrico (M, d) quando qualquer que seja $\varepsilon > 0$, é possível encontrar $n_0 \in \mathbb{N}$ tal que, escolhidos quaisquer dois índices a partir de n_0, digamos n e m, teremos sempre $d(x_n, x_m) < \varepsilon$, ou seja, $m, n > n_0 \Rightarrow d(x_m, x_n) < \varepsilon$.

Por exemplo, a sequência $\left(\frac{1}{n}\right)$ é de Cauchy.

De fato, dado $\varepsilon > 0$ existe $n_1 \in \mathbb{N}$ tal que $n_1.\varepsilon > 2$. Daí,

$$m, n > n_1 \Rightarrow |x_m - x_n| = \left|\frac{1}{m} - \frac{1}{n}\right| \leq \left|\frac{1}{m}\right| + \left|\frac{1}{n}\right| = \frac{1}{m} + \frac{1}{n}.$$

Mas, para $m > n_1$, temos $\frac{1}{m} < \frac{1}{n_1}$ e para $n > n_1$ temos $\frac{1}{n} < \frac{1}{n_1}$. Logo,

$$m, n > n_1 \Rightarrow |x_m - x_n| < \frac{1}{n_1} + \frac{1}{n_1} = \frac{2}{n_1},$$

donde segue que $m, n > n_1 \Rightarrow |x_m - x_n| < \varepsilon$.

Portanto, $\left(\frac{1}{n}\right)$ é de Cauchy.

Definição

Um espaço métrico (M, d) é dito ser completo se toda sequência de Cauchy, tomada em M, convergir em M.

Observe que a reta é um exemplo de espaço métrico completo.

Definição

Sejam M, N espaços métricos. Chamamos de contração uma aplicação $f : M \longrightarrow N$ quando existe uma constante c, com $0 \leq c < 1$, tal que $d(f(x), f(y)) \leq c.d(x, y)$, para quaisquer $x, y \in M$.

Teorema do Ponto Fixo de Banach

Nesta seção será apresentado o Teorema do ponto fixo de Banach e sua demonstração, o qual será ultilizado para mostrar o funcionamento do buscador do Google na seção seguinte.

Definição

Seja M um espaço métrico. Um ponto fixo de uma aplicação $f : M \longrightarrow M$ é um ponto $x \in M$ tal que $f(x) = x$.

Note que a aplicação $f : \mathbb{R} \longrightarrow \mathbb{R}$, dada por $f(x) = x^2$, tem dois pontos fixos, que são os pontos 0 e 1. Pois, $f(0) = 0^2 = 0$ e $f(1) = 1^2 = 1$.

Teorema

Se M é um espaço métrico completo, toda contração $f : M \longrightarrow M$ possui um único ponto fixo em M.

Demonstração: Ver [1], Teorema 2, Capítulo VII.

Funcionamento do Buscador do Google

O funcionamento de um buscador é baseado essencialmente em dois passos:

1) Matching $\Longrightarrow$ Busca

2) Ranking $\Longrightarrow$ Seleciona e Ordena

A forma de ordenar as páginas encontradas é o segredo do sucesso do Google (Ver [2]).

Sabemos que as páginas se conectam através de links e podemos pensá-las como nós de um grafo direcionado cuja arestas são dadas pelos links. Seja G um grafo direcionado, com nós $1, 2, \ldots, n$. Nosso objetivo é, para cada nó i, atribuir um valor real x_i que traduza a relevância do nó i.

Quando j é um nó que aponta para i, chamaremos de **link** de i para j e denotaremos $j \to i$. O número de links saindo de j será denotado por l_j. Assim, definiremos a relevância de um nó ou site da seguinte forma:

$$x_i = \sum_{j \to i} \frac{x_j}{l_j}.$$

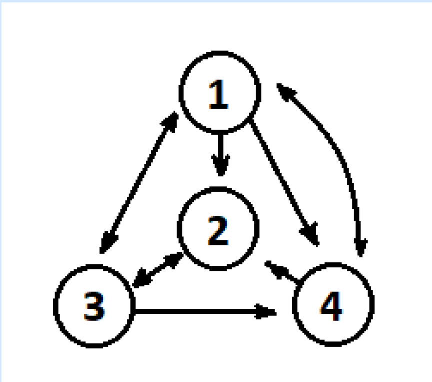
Desse modo, definiremos m_{ij} como o número de links de $j \to i$, que pode inclusive ser zero. Com isso, temos o sistema linear

$$\begin{cases} x_1 = \frac{m_{11}}{l_1}x_1 + \frac{m_{12}}{l_2}x_2 + \cdots + \frac{m_{1n}}{l_n}x_n \\ x_2 = \frac{m_{21}}{l_1}x_1 + \frac{m_{22}}{l_2}x_2 + \cdots + \frac{m_{2n}}{l_n}x_n \\ \vdots \\ x_n = \frac{m_{n1}}{l_1}x_1 + \frac{m_{n2}}{l_2}x_2 + \cdots + \frac{m_{nn}}{l_n}x_n \end{cases}$$

Nesses termos, se $A = (a_{ij})$, onde $a_{ij} = \frac{m_{ij}}{l_j}$, então podemos ver $A : \mathbb{R}^n \to \mathbb{R}^n$ como uma transformação linear e a relevância $x = [x_1, x_2, \ldots, x_n]$ como um autovetor de autovalor unitário. Ou equivalente, um ponto fixo de A, isto é, $Ax = x$.

Exemplo

Observe o grafo dado abaixo.



Neste caso, $l_1 = 4, l_2 = 1, l_3 = 3$ e $l_4 = 2$. O sistema linear fica:

$$\begin{cases} x_1 = \frac{0}{4}x_1 + \frac{0}{1}x_2 + \frac{1}{3}x_3 + \frac{1}{2}x_4 \\ x_2 = \frac{1}{4}x_1 + \frac{0}{1}x_2 + \frac{1}{3}x_3 + \frac{1}{2}x_4 \\ x_3 = \frac{1}{4}x_1 + \frac{1}{1}x_2 + \frac{0}{3}x_3 + \frac{0}{2}x_4 \\ x_4 = \frac{2}{4}x_1 + \frac{0}{1}x_2 + \frac{1}{3}x_3 + \frac{0}{2}x_4 \end{cases}$$

Uma possível solução desse sistema é dada pelo vetor $[4, 5, 6, 4]$, que indicam, respectivamente, a relevância das páginas $1, 2, 3$ e 4. Portanto, a página 3 é a que possui maior relevância.

Interpretemos a_{ij} como sendo a probabilidade de, saindo do vértice j, chegar ao vértice i. Para entendermos melhor esse funcionamento suponha que um internauta, aleatoriamente, escolhe uma das n páginas, por exemplo, $v_0 = [1, 0, \ldots, 0]$. O vetor v_1, obtido na equação $Av_0 = v_1$, indica a probabilidade do internauta se encontrar na página i após um click, partindo de v_0. Continuando assim, sucessivamente, após n clicks a probabilidade de o internauta se encontrar na página i é dado pela equação $Av_{n-1} = v_n$. Esse modelo seria ideal se o internauta sempre a partir de um link de uma página encontrasse outro de outra página, porém isso não ocorre. A partir daí, a ideia de Page e Brin foi introduzir um fator probabilístico p de começar tudo de novo e, evidentemente, $1 - p$ de continuar nos links. Desse modo, a aplicação que indica o percurso aleatório do internauta num grafo de n vértice é:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \mapsto p. \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix} + (1-p) \begin{bmatrix} \frac{m_{11}}{l_1} & \frac{m_{12}}{l_2} & \cdots & \frac{m_{1n}}{l_n} \\ \frac{m_{21}}{l_1} & \frac{m_{22}}{l_2} & \cdots & \frac{m_{2n}}{l_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{m_{n1}}{l_1} & \frac{m_{n2}}{l_2} & \cdots & \frac{m_{nn}}{l_n} \end{bmatrix} . \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Podemos escrever, de forma mais simples, como $T : \mathbb{R}^n \to \mathbb{R}^n$, onde $y \mapsto pe + (1-p)Ay$, com

$$e = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix}.$$

Note que a função T é contínua. Se mostrarmos que T é uma contração e, lembrando que $\mathbb{R}^n$ é um espaço métrico completo, o Teorema do ponto fixo de Banach nos garante que T possui um único ponto fixo, ou seja, que o internauta chega sempre à página desejada. Para tanto, vejamos que, dados $y, z \in \mathbb{R}^n$ e, notando que

$$\sum_{i=1}^{n} a_{ij} = 1$$

temos

$$\|T(y) - T(z)\| = \|(1-p)A(y-z)\| = (1-p)\sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}|y_j - z_j|\right) =$$

$$= (1-p)\sum_{j=1}^{n}|y_j - z_j| = (1-p)\|y - z\|.$$

Já que $1 - p$ é menor que 1, temos que T é uma contração. Finalmente, por a aplicação possuir um único ponto fixo, o Teorema do ponto fixo de Banach garante que tudo isso funciona e indicando que a relevância de cada página está bem definida.

Bibliografia

1. DOMINGUES, Hygino Hugueros. *Espaços Métricos e Introdução à Topologia.* São Paulo: Atual Editora, 1982.
2. BARROS, Cícero Demétrio Vieira. *O Teorema do Ponto Fixo de Banach e algumas Aplicações.* 2013. 46 p. Dissertação de Mestrado (PROFMAT) - Universidade Federal da Paraíba, João Pessoa, 2013.

luisaamartins10@gmail.com, ana.galves@ufu.br