Numerical Computing Cheatsheet

Edoardo Riggio

January 4, 2022

Numerical Computing - SA. 2021
Computer Science
Università della Svizzera Italiana, Lugano

Contents

1 PageRank Algorithm

The PageRank algorithm is entirely determined by the link structure of the World Wide Web. A page will thus have a high rank if other pages with high ranks link to it.

1.1 Random Surfer Model

This algorithm is based on the **Random Surfer Model**. Here we imagine a user going from one page to the other by randomly choosing an outgoing link from one page to the next. This process is also called **exploitation**. The problem with exploitation is that it could lead to dead-ends – i.e. a page with no outgoing links, or cycles around cliques of interconnected pages. For this reason, we sometimes choose a random page from the web to navigate to. This process is called **exploration**.

1.2 Markov Chains

The random walk generated by the combination of exploitation and exploration is known as a **Markov Chain**. A Markov Chain – or Markov Process, is a stochastic process. Differently from other stochastic processes, it has the property of being memory-less. This means that the probability of future states are not dependent upon the steps that led up to the present state.

Let W be the set of webpages that can be reached by a Markov Chain of hyperlinks, n the number of pages in W, ad G the $n \times n$ connectivity matrix of a portion of the Web. The matrix G will be composed as follows:

$$g_{ij} = \begin{cases} 1 & \text{if there is a hyperlink from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$$

From matrix G we can determine the in-degree and out-degree of a page j. This can be computed as follows:

$$r_i = \sum_j g_{ij}$$

$$c_J = \sum_i g_{ij}$$

Where r_i is the in-degree and c_j is the out-degree. Let now p be the probability that the random walk follows a link – i.e. performs exploitation. A typical value for p is 0.85. Let δ be the probability that a particular random page is chosen, and $1 - p$ is the probability that some random page is chosen – i.e. perform exploration. Then δ will have the following formulation:

$$\delta = \frac{1-p}{n}$$

Now let A be a matrix that comes from scaling G by its column sums. The elements of matrix A will be:

$$a_{ij} = \begin{cases} \rho \cdot \frac{g_{ij}}{c_j} + \delta & \text{if } c_j \neq 0 \\ \frac{1}{n} & \text{if } c_j = 0 \end{cases}$$

Matrix A is the transition probability matrix for the Markov Chain – for this reason this matrix is also known as the Markov matrix. All of its elements are strictly between 1 and 0, and its column sums are all equal to 1.

1.3 Power Method

The **power method** is an algorithm used in order to produce the dominant eigenvector of matrix A. In order to do so, we need to repeat the following computation:

$$x = G \cdot (x + e) \cdot (z + x)$$

Until x settles down to several decimal places.

1.4 Inverse Iteration

The **inverse iteration** is an algorithm which has the same goal of the power method. In this case, we need to find the dominant eigenvector of $(A - \alpha)^{-1}$ rather than of A. By doing so, the convergence is faster, but the computations are more expensive – this method solves a system of equations, while the power method simply multiplies matrices together.