The Theory behind PageRank

Mauro Sozio

Telecom ParisTech

September 26, 2016

The PageRank algorithm

Input: A directed graph G with N nodes (Web pages), $0 < \beta < 1, \epsilon > 0$.
Output: The PageRank vector π of the web pages in G.

```
Remove dead ends iteratively from G;
Build the stochastic matrix M_G (M for short);
Let π^(0) = [1/N, ... 1/N]^T
while (true) do
    n = n + 1;
    π^(n) = β M π^(n-1) + [(1-β)/N]_N;
    If ||π^(n) - π^(n-1)||_1 < ε break;
end while
return π^(n).
```

Remove *dead ends* iteratively from G;
Build the stochastic matrix M_G (M for short);
Let $\pi^{(0)} = [\frac{1}{N}, \ldots \frac{1}{N}]^T$
while (true) **do**
  $n = n + 1$;
  $\pi^{(n)} = \beta M \pi^{(n-1)} + [\frac{1-\beta}{N}]_N$;
  If $||\pi^{(n)} - \pi^{(n-1)}||_1 < \epsilon$ **break**;
end while
return $\pi^{(n)}$.

Random Surfer Interpretation

Given a directed graph G the random surfer starts visiting one page chosen uniformly at random.

At each time step n, let v be web page the random surfer is currently visiting. The next page is chosen as follows:

- with prob. $\frac{\beta}{\text{num. of successors of } v}$ he/she visits a random successor of v.
- with probability $\frac{1-\beta}{N}$ he/she visits one page uniformly at random.

Random Surfer Interpretation

Given a directed graph G the random surfer starts visiting one page chosen uniformly at random.

At each time step n, let v be web page the random surfer is currently visiting. The next page is chosen as follows:

- with prob. $\frac{\beta}{\text{num. of successors of } v}$ he/she visits a random successor of v.
- with probability $\frac{1-\beta}{N}$ he/she visits one page uniformly at random.

What is the interpretation for $\pi^{(n)}$?

Events and Probability

Consider a stochastic process (e.g. throw a dice, pick a card from a deck)

- Each possible outcome is a *simple event*.
- The sample space Ω is the set of all possible simple events.
- An event is a set of simple events (a subset of the sample space).
- With each simple event E we associate a real number $0 \leq P(E) \leq 1$, which is the probability that event E happens.

Probability Space

Definition 1

A *probability space* has three components:

- A *sample space* Ω, which is the set of all possible outcomes of the random process modeled by the probability space;
- A family of sets $\mathcal{F}$ representing the allowable events, where each set in $\mathcal{F}$ is a subset of the sample space in Ω;
- a *probability function* $P : \mathcal{F} \to \mathbb{R}$, satisfying the definition below (next slide).

Probability Function

Definition 2

A *probability function* is any function $P : \mathcal{F} \rightarrow \mathbb{R}$ that satisfies the following conditions:

- for any event E, $0 \leq P(E) \leq 1$;
- $P(\Omega) = 1$;
- for any finite or countably infinite sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$

$$P\left(\cup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i). \tag{1}$$

The Union Bound

Theorem 3

$$P(\cup_{i=1}^{n} E_i \leq \sum_{i=1}^{n} P(E_i)). \tag{2}$$

Example: roll a dice:

- let $E_1 =$ “result is odd”
- let $E_2 =$ “result is ≤ 2”

Independent Events

Definition 4

Two events E_1 and E_2 are *independent* if and only if

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \tag{3}$$

Conditional Probability: Example

What is the probability that a random student at Telecom ParisTech was born in Paris?
E_1 = the event "born in Paris".
E_2 = the event "student at Telecom ParisTech".
The conditional probability that a a student at Telecom ParisTech was born in Paris is written:

$$P(E_1|E_2).$$

Conditional Probability: Definition

Definition 5

The *conditional probability* that event E_1 occurs given that event E_2 occurs is

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \tag{4}$$

The conditional probability is only well-defined if $P(E_2) > 0$.

By conditioning on E_2 we restrict the sample space to the set E_2.

Law of Total Probability

Theorem 6

Let $B_1, \ldots, B_k$ be a partition of the sample space Ω, with $P(B_i) > 0$, $i = 1, \ldots, k$. Then, for any event $A \subseteq \Omega$:

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i). \tag{5}$$

Random Variable

Definition 7

A *random variable* X on a sample space Ω is a function on Ω; that is, $X : \Omega \to \mathbb{R}$.

A *discrete random variable* is a random variable that takes on only a finite number of values.

Examples

In practice, a random variable is some random quantity that we are interested in:

- I roll a die, $X =$ result. E.g. $X = 6$.
- I pick a card, $X = 1$ if card is an Ace, 0 otherwise.
- I roll a dice two times. $X_1 =$ result of the first experiment, $X_2 =$ result of the second experiment. What is $P(X_1 + X_2 = 2)$?

Stochastic Processes

Definition 8

A stochastic process in discrete time $n \in \mathbb{N}$ is a sequence of random variables $X_0, X_1, X_2 \dots$ denoted by $\mathbf{X} = \{X_n\}$.

We refer to the value X_n as the *state* of the process at time n, with X_0 denoting the initial state.

The set of possible values that each random variable can take is denoted by S. Here, we shall assume that S is finite and $S \subseteq \mathbb{N}$.

Markov Chains

Definition 9

A *stochastic process* $\{X_n\}$ is called a *Markov chain* if for any $n \geq 0$ and any value $j_0, j_1, \ldots, i, j \in S$,

$$P(X_{n+1} = i | X_n = j, X_{n-1} = j_{n-1}, \ldots, X_0 = j_0) = P(X_{n+1} = i | X_n = j),$$

which we denote by P_{ij}.

This can be stated as *the future is independent of the past given the present state*. In other words, the probability of moving to the next state does not depend on what happened in the past. Note that $P_{ij} \neq P_{ji}$.

One-step Transition Matrix

P_{ij} denotes the probability that the chain, whenever in state j, moves next into state i.

The square matrix $\mathbf{P} = (P_{ij})$, $i, j \in S$, is called the *one-step transition matrix*. Note that for each $j \in S$ we have:

$$\sum_{i \in S} P_{ij} = 1. \tag{6}$$

n-step Transition Matrix

The *n-step transition matrix* $\mathbf{P}^{(n)}$, $n \geq 1$, where

$$P_{ij}^{n} = P(X_n = i | X_0 = j) = P(X_{m+n} = i | X_m = j), \quad \forall m \tag{7}$$

denotes the probability that n steps later the Markov chain will be in state i given that at step m is in state j.

Theorem 10

$$\mathbf{P}^{(n)} = \mathbf{P}^{n} = \mathbf{P} \times \mathbf{P} \times \cdots \times \mathbf{P}, n \geq 1.$$

Stationary Distribution

Definition 11

A probability distribution π over the states of the Markov chain ($\sum_{j \in S} \pi_j = 1$) is called a *stationary distribution*[a] if

$$\pi = P\pi. \tag{8}$$

[a] in literature the transpose of P is often used, in that case $\pi = \pi P$.

Irreducible Markov Chains

Definition 12

A Markov chain is called *irreducible*[a] iff for any $i, j \in S$, there is $n \geq 1$ such that:

$$P_{ij}^n > 0. \tag{9}$$

[a] definition is different when S is not finite.

In other words, the chain is able to move from any state i to any state j (in one or more steps). As a result, if a Markov chain is irreducible then there must be n such that $P_{ii}^n > 0$.

Aperiodic Markov Chains

A state i has period k if any return to i occurs at step $k \cdot l$, for some $l > 0$. Formally,

$$k = gcd\{n : P(X_n = i | x_0 = i) > 0\} \tag{10}$$

where gcd denotes the *greatest common divisor*. If $k = 1$ then state i is said to be *aperiodic*.

Definition 13

A Markov chain is called *aperiodic* if every state is aperiodic.

Main Theorem

Theorem 14

If a Markov chain is irreducible and aperiodic[a]*, then a stationary distribution* π *exists and is unique. Moreover, the Markov chain converges to its stationary distribution, that is,*

$$\pi_j = \lim_{n\to\infty} P(X_n = j) = \lim_{n\to\infty} P(X_n = j | X_0 = i), \quad \forall i, j \in S. \tag{11}$$

[a] in this case the Markov chain is called *ergodic*

Note: Equation (11) holds for any initial state i of the Markov chain.

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible?

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic?

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.
- If there are spider traps or dead ends?

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.
- If there are spider traps or dead ends? It is not irreducible.

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.
- If there are spider traps or dead ends? It is not irreducible.
- What if we add random jumps?

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.
- If there are spider traps or dead ends? It is not irreducible.
- What if we add random jumps? It is both irreducible and aperiodic, which implies that there exists a stationary distribution π.

Markov chains and the Random Surfer

Consider the Markov chain (MC) of the random surfer, no rand. jumps.

- Is it irreducible? Not necessarily.
- Is it aperiodic? Not necessarily.
- If there are spider traps or dead ends? It is not irreducible.
- What if we add random jumps? It is both irreducible and aperiodic, which implies that there exists a stationary distribution π.
- Moreover the vector computed by PageRank converges to π...

The Random Surfer and its stationary distribution

Observation 1

Let $\pi^{(0)}$ be a probability distribution over the states of the Markov chain with $\pi_j^{(0)} = P(X_0 = j)$. Let $\pi^{(n)} = P^{(n)}\pi^{(0)}$. From the law of total probability (Thm. 6), and Thm. 10 it follows that $\pi_j^{(n)} = P(X_n = j)$, $\forall j$.

The Random Surfer and its stationary distribution

Observation 1

Let $\pi^{(0)}$ be a probability distribution over the states of the Markov chain with $\pi_j^{(0)} = P(X_0 = j)$. Let $\pi^{(n)} = P^{(n)}\pi^{(0)}$. From the law of total probability (Thm. 6), and Thm. 10 it follows that $\pi_j^{(n)} = P(X_n = j)$, $\forall j$.

The stationary distribution π of the MC is the PageRank vector!

The Random Surfer and its stationary distribution

Observation 1

Let $\pi^{(0)}$ be a probability distribution over the states of the Markov chain with $\pi_j^{(0)} = P(X_0 = j)$. Let $\pi^{(n)} = P^{(n)}\pi^{(0)}$. From the law of total probability (Thm. 6), and Thm. 10 it follows that $\pi_j^{(n)} = P(X_n = j)$, $\forall j$.

The stationary distribution π of the MC is the PageRank vector!

Sketch: At each step n of PageRank we compute $\pi^{(n)} = P\pi^{(n-1)} = P^{(n)}\pi^{(0)}$. From Observation 1, it follows that $\pi_j^{(n)} = P(X_n = j)$ which converges to π (Theorem 14).