PageRank

Tan Yan Daniel Bae

December 11, 2018

Abstract

Enter a short summary here. What topic do you want to investigate and why? What experiment did you perform? What were your main results and conclusion?

1 Introduction

2 Definitions

Definition 1. A stochastic matrix, probability matrix, or Markov transition matrix is ... A column-stochastic or left-stochastic matrix is ...

Definition 2. A positive matrix is a matrix with all positive entries

Definition 3. An Markov transition matrix is said to be irreducible if ... A positive stochastic matrix is necessarily irreducible

Definition 4. Let G be a directed graph of the webpages, with each webpage as a node and links between the webpages as edges ... such that $G_{ij} = 1$ if ... also let n denote the number of nodes ... let $L(i)$ denote the number of outgoing edges ...

Definition 5. A dangling node is a node that has no outgoing edges

Definition 6. Let P be the probablistic matrix

Definition 7. Let d denote the "damping factor" ... representing the probability of a web surfer randomly traveling from one page to any other page. $0 \leq d < 1$.

Definition 8. Let $\mathbf{1}$

Definition 9. Let M denote the "PageRank Matrix" or "Google Matrix" defined by Page and Brin. Define

Definition 10. Let λ denote an eigenvalue of the square matrix A and let $\mathbf{v}$ denote its corresponding eigenvector, such that $A\mathbf{v} = \lambda\mathbf{v}$. A probablistic eigenvector is...

3 Main Ideas

We want to find the PageRank vector, which encodes the probabilities of landing on a particular page provided the network of links between the pages. This is equivalent to finding the stationary distribution of the Markov chain that encodes the link transitions between pages.

In order to prove the existance of the PageRank vector, we first on the Perron-Frobenius Theorem.

Theorem 1. (Perron-Frobenius Theorem) Consider a $n \times n$ positive column-stochastic matrix M. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the eigenvalues of M sorted in decreasing order of magnitude. Then 1 is an eigenvalue with multiplicity 1 such that $\lambda_1 = 1$ and the corresponding eigenvector is probabilistic.

Since the Google matrix M is constructed to be a postitive column-stochastic matrix, by Perron-Frobenius theorem we know that there exists a probabilistic eigenvector $\mathbf{v}^*$ corresponding to the eigenvalue 1.

The next theorem uses the fact that all of the eigenvectors λ_i for $1 < i \leq n$ of M are less than 1 to show that the PageRank vector is equal to $\mathbf{v}^*$.

Theorem 2. (Power Method Convergence Theorem) Let M be a $n \times n$ positive column-stochastic matrix, with $\mathbf{v}^*$ denosting the probabilstic vector corresponding to 1. Let z be the column vector with all entries equal to $\frac{1}{n}$. Then the sequence $z, Mz, \cdots, M^k z$ as $k \to \infty$ converges to $\mathbf{v}^*$.

Assuming that the random surfer initially has an equal probability of visiting a given page, traversing the network using the information from the Google matrix leads us to the PageRank vector. This implies that after infinitely many random walks, the web surfer is most likely to be on the webpage corresponding to the most links.

```
def build_prob_matrix(adj_list):
# number of nodes
n = len(adj_list)
P = np.zeros((n,n), dtype=float)
for (j, connected_nodes) in enumerate(adj_list):
    if connected_nodes: # non-empty
        P[connected_nodes, j] = 1.0 / len(connected_nodes)
    else: # dangling node
        P[:, j] = 1.0 / n
return P
```

4 Conclusion

References

[1]