

Αλγόριθμος Pagerank με παράλληλη υλοποίηση της μεθόδου Gauss-Seidel

Νικόλαος Κατωμέρης

AEM: 8551

ngkatomer@auth.gr

Παράλληλα και διανεμημένα συστήματα.

Τμήμα ηλεκτρολόγων μηχανικών και μηχανικών υπολογιστών, ΑΠΘ

Περίληψη

Στην παρούσα αναφορά περιγράφεται η διαδικασία παραλληλοποίησης του αλγορίθμου Gauss-Seidel για τον υπολογισμό του διανύσματος Pagerank γράφων του διαδικτύου. Επίσης, παρουσιάζονται αποτελέσματα σε ενδεικτικούς γράφους καθώς και αποδείξεις ορθότητας του υλοποιημένου προγράμματος.

Pagerank

Η βασική ιδέα του αλγορίθμου Pagerank είναι η αξιολόγηση της κάθε σελίδας του δικτύου με βάση τον αριθμό των συνδέσμων προς αυτήν και την αξιολόγηση των σελίδων που τους περιέχουν.

Το αποτέλεσμα της παραπάνω αξιολόγησης ταυτίζεται με το ποσοστό του χρόνου στον οποίο ένας τυχαίος περιπατητής θα βρισκόταν στην εκάστοτε σελίδα, αν ξεκινώντας από κάποια τυχαία σελίδα, ακολουθούσε τυχαία κάποιον από τους συνδέσμους της, στη συνέχεια της επόμενης κι ούτω καθ' εξής.

Η παραπάνω διαδικασία μπορεί να προσομοιωθεί με μία αλυσίδα Markov.

Έστω A , ο $N \times N$ πίνακας μεταβάσεων της αλυσίδας για N σελίδες του διαδικτύου. Αν $j \rightarrow i$ σημαίνει ότι η σελίδα j περιέχει σύνδεσμο προς τη σελίδα i , ο A ορίζεται αρχικά ως εξής:

$$A(i, j) = \begin{cases} 1, & \text{αν } j \rightarrow i \\ 0, & \text{αλλιώς} \end{cases}$$

Επειδή όμως ο παραπάνω πίνακας θέλουμε να εκφράζει την πιθανότητα ο περιπατητής να πάει από τη σελίδα j στη σελίδα i σε ένα βήμα, είναι απαραίτητο το άθροισμα των στοιχείων

κάθε στήλης του πίνακα να ισούται με 1.

Οπότε, αφού λάβουμε τον πίνακα με τους συνδέσμους των σελίδων, διαιρούμε όλα του τα στοιχεία με το άθροισμα των στοιχείων της εκάστοτε στήλης j , έστω L_j .

Επιπλέον, καθώς ο περιπατητής ξεκινά από μία σελίδα, είναι πολύ πιθανό, ακολουθώντας συνεχώς συνδέσμους να είναι αδύνατο να μεταβεί σε όλες τις σελίδες του γράφου. Το γεγονός αυτό, συν το γεγονός ότι ένας χρήστης του διαδικτύου δεν ακολουθεί μόνο συνδέσμους, αλλά «τηλεμεταφέρεται» κιόλας σε άλλες σελίδες, καθιστά τον παραπάνω πίνακα ανεπαρκή για την προσομοίωση της τυχαίας διαδρομής ενός χρήστη του διαδικτύου.

Έτσι, εισάγεται στον αλγόριθμο μια σταθερά «τηλεμεταφοράς» που δηλώνει την πιθανότητα ο περιπατητής να μεταβεί σε κάποια τυχαία σελίδα του διαδικτύου χωρίς να υπάρχει κάποιος σύνδεσμος προς αυτήν στη σελίδα που βρίσκεται.

Η σταθερά αυτή, έστω b , μπορεί να επιλεγεί ως όρισμα στο πρόγραμμα υπολογισμού του διανύσματος pagerank που υλοποιήθηκε. Στην περίπτωση που δεν δοθεί, θα έχει την τιμή 0.15 που έχει δειχθεί ότι αποτελεί καλή επιλογή [1].

Με την εισαγωγή της παραπάνω σταθεράς, ο

πίνακας μεταβάσεων πλέον θα είναι:

$$A(i, j) = \begin{cases} \frac{1-b}{L_j}, & \text{αν } j \rightarrow i \\ \frac{b}{L_j}, & \text{αλλιώς} \end{cases}$$

Με τη χρήση του παραπάνω πίνακα είναι δυνατό να υπολογιστεί το διάνυσμα $N \times 1$ με την αξιολόγηση της κάθε σελίδας, πολλαπλασιάζοντάς τον με ένα αρχικό διάνυσμα -εδώ ένα ομοιόμορφο- αρκετές φορές, έως ότου το αποτέλεσμα συγκλίνει. Το διάνυσμα αυτό αποτελεί ιδιοδιάνυσμα του πίνακα A για την ιδιοτιμή 1.

Για να συγκλίνει με βεβαιότητα το διάνυσμα απαιτείται ακόμη ένα βήμα, θα πρέπει, αν ο περιπατητής βρεθεί σε κάποια σελίδα που δεν περιέχει καμία σύνδεση, να τηλεμεταφερθεί σε μία τυχαία σελίδα με βάση ένα διάνυσμα πιθανοτήτων. Στην υλοποίηση αυτή, ο περιπατητής έχει την ίδια πιθανότητα να τηλεμεταφερθεί σε οποιαδήποτε σελίδα.

Με όλα τα παραπάνω δεδομένα, αποδεικνύεται ότι ο τελικός πίνακας θα έχει μία ιδιοτιμή ένα και το πρόβλημα πλέον έγκειται στην εύρεση του ιδιοδιανύσματος της.

Για την εύρεση του ιδιοδιανύσματος, όπως ζητείται, χρησιμοποιείται η μέθοδος Gauss-Seidel.

Gauss Seidel

Η μέθοδος Gauss-Seidel είναι μια αναδρομική μέθοδος επίλυσης συστημάτων. ...

Hypothesis 1 *A hypothesis is likely to be false if it is longer than this line*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

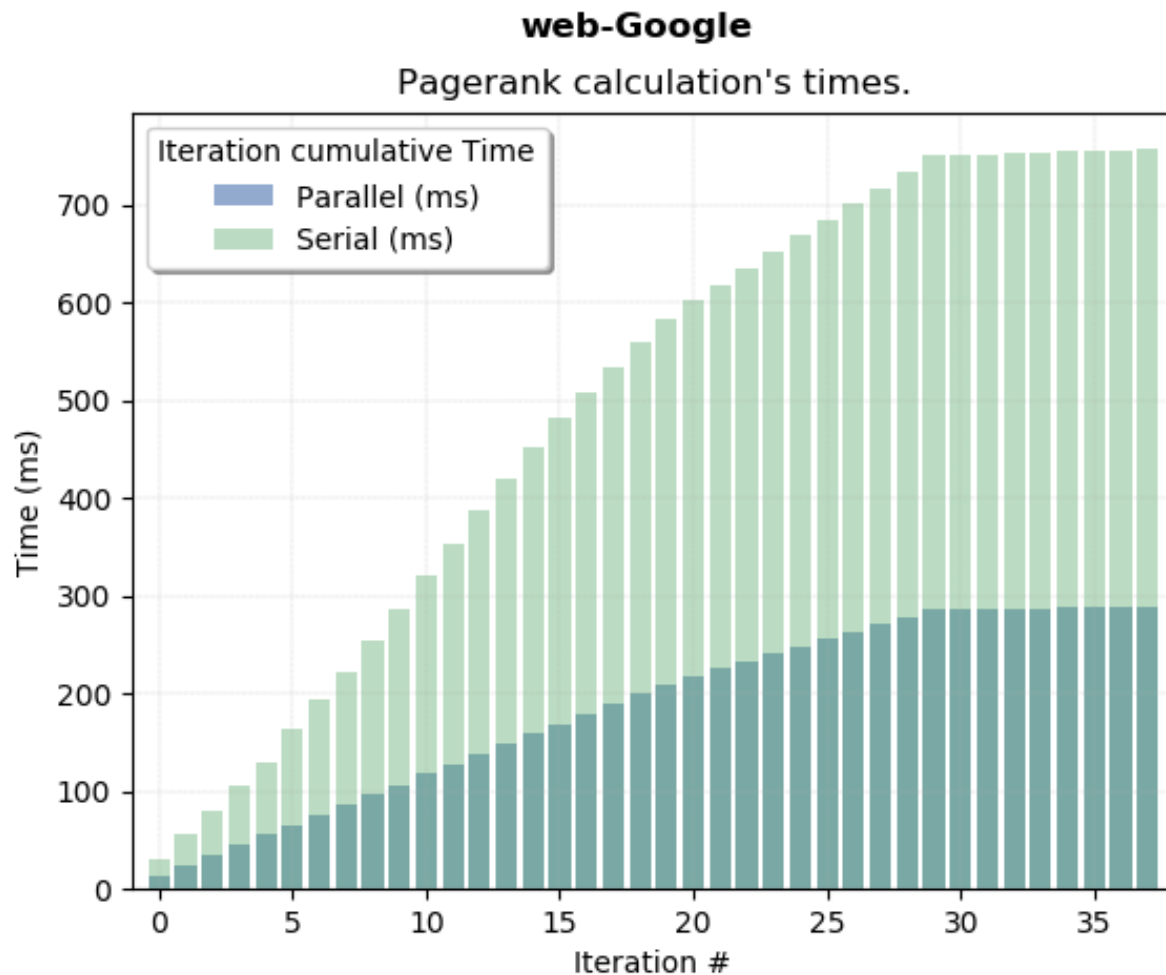
Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

text	text	text	text	text	text
1	84	46	4	3	5
2	67	24	5	1	5
3	54	26	4	2	5
4	44	16	5	2	5
5	53	21	6	2	5

Πίνακας 1: An example single column table

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lorem ipsum dolor sit amet, consectetur



Σχήμα 1: A two column figure (indicate with figure*)

adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus.

Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Αναφορές

- [1] Sepandar Kamvar, Taher Haveliwala και Gene Golub. «Adaptive methods for the computation of PageRank». Στο: *Linear Algebra and its Applications* 386 (2004), σσ. 51–65.