ImportanceIteration Algorithm Explanation

Yirong Chen, Leo Luo, Yanzhen Shen

5/6/2022

Note: All indexes in this file are starting from 0.

Correctness

We will first show that function `PageRank` in `src/importance.cpp` act as a single time of matrix multplication.

Consider the procedure SINGLE-ITERATION.

SINGLE-ITERATION(G, v)

1 let w be a vector with $|V|$ elements and initialized to 0.
2 **for** $i = 0$ **to** $|V| - 1$
3 $\quad p = \frac{v[i]}{|G.Adj[i]|}$
4 $\quad$ **for** each $j \in G.adj[i]$
5 $\quad\quad w[j] = w[j] + p$

Note that SINGLE-ITERATION is the pseudocode of `PageRank`.

Lemma 1. Let $G = (V, E)$ be a directed, strongly connected, simple graph. Let A be the adjcency matrix [1] of G Let B be a left stochastic matrix [2] from the result of normalizing A^T. Then $Bv =$ SINGLE-ITERATION(G, v).

Proof. Let $w = Bv$. Then

$$w[i] = \sum_{j=0}^{|V|-1} B[i][j]v[j] \quad .$$

[1] $A[i][j] \neq 0$ iff $(i, j) \in E$ iff there is an edge from i to j

[2] a real square matrix, with each column summing to 1

Let $f(j)$ maps j to the number of non-zero entries in column j of B. Since G is strongly connected, $f(j) > 0$. Since G is simple, all elements in $G.Adj[j]$ are unique (we can treat $G.Adj[j]$ as a set); hence $|G.Adj[j]| = f(j)$. Let

$$S_i = \{j \in \{0, \cdots, |V|-1\} | i \in G.Adj[j]\} \quad .$$

Since

$$B[i][j] = \frac{1}{f(j)} \neq 0 \Longleftrightarrow (j,i) \in E \Longleftrightarrow i \in G.Adj[j] \Longleftrightarrow j \in S_i \quad ,$$

we have

$$w[i] = \sum_{j=0}^{|V|-1} B[i][j]v[j] = \sum_{j \in S_i} \frac{1}{f(j)} v[j] = \sum_{j \in S_i} \frac{v[j]}{|G.Adj[j]|} \quad . \tag{1}$$

SINGLE-ITERATION-LOOP-INVERSION(G, v)

1 let w be a vector with $|V|$ elements and initialized to 0.
2 **for** $i = 0$ **to** $|V| - 1$
3 **for** each $j \in S_i$
4 $w[i] = w[i] + \frac{v[j]}{|G.Adj[j]|}$

SINGLE-ITERATION-LOOP-INVERSION is a loop inversion version of SINGLE-ITERATION. Obviously, it produces the result of equation (1) for all $i \in \{0, \cdots, |V|-1\}$. □

Actually, the lemma is enough to show that `ImportanceIteration` produces a same result as power iteration method.

The following theorem is cited from MATH 257:

Theorem 2. Let A be an $n \times n$-left stochastic matrix with only positive entries and let $z \in \mathbb{R}^n$ be a probability vector. Then

$$z_\infty := \lim_{k \to \infty} A^k z$$

exists, and z_∞ is a stationary probability vector of A (i.e. $Az_\infty = z_\infty$).

Theorem 3. The function `ImportanceIteration` in `src/importance.cpp` correctly computes PageRank vector.

Proof. Immediately from 1 and Theorem 2. □

Complexity Analysis

Let α be the number of iteration. Because of the usage of adjacency list, each invocation of `PageRank` or SINGLE-ITERATION takes $\Theta(V + E)$ time. Therefore, the time complexity of `ImportanceIteration` is $\Theta(\alpha(V + E))$, which is better than the time complexity of power iteration, $\Theta(\alpha V^2)$ ($|V| + |E| = O(V^2)$ in a simple graph).

Acknowledge

This file uses macro package “clrscode3e”.