วารสารคณิตศาสตร์ MJ-MATh 61(690) Sep–Dec 2016
โดย สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์
http://MathThai.Org MathThaiOrg@gmail.com

การประยุกต์พีชคณิตเชิงเส้นในขั้นตอนวิธีเพจแรงค์ของกูเกิ้ล

An Application of Linear Algebra in Google's PageRank Algorithm

อเนกวิทย์ บุญเกษม
Anakewit Boonkasame

Department of Applied Mathematics and Informatics
Prince of Songkla University, Surat Thani Campus, Surat Thani 84000

Email: anakewit.b@psu.ac.th

บทคัดย่อ

พีชคณิตเชิงเส้นเป็นวิชาคณิตศาสตร์พื้นฐานที่สำคัญ เนื่องจากมีการประยุกต์ที่หลากหลายในด้านต่างๆ ผู้เขียนจะอธิบายการประยุกต์พีชคณิตเชิงเส้นในขั้นตอนวิธีเพจแรงค์ของกูเกิ้ล ซึ่งเป็นเครื่องมือค้นหาเว็บเพจที่ได้รับความนิยมสูงสุด เพื่อเน้นย้ำความสำคัญของพีชคณิตเชิงเส้น

คำสำคัญ: เครื่องมือค้นหา การวิเคราะห์ลิงค์ ปัญหาลักษณะเฉพาะ วิธียกกำลัง

ABSTRACT

Linear algebra is an important subject in introductory mathematics as it has many applications in various fields. To emphasize the importance of linear algebra, this article discusses its application in the PageRank algorithm used by Google, which is by far the most popular search engine.

Keywords: search engine, link analysis, eigenproblem, power method

1. บทนำ

พีชคณิตเชิงเส้นเป็นวิชาคณิตศาสตร์พื้นฐานสำหรับนักศึกษาระดับปริญญาตรีในสาขาต่างๆ เนื่องจากมีการประยุกต์ที่หลากหลายทั้งในวิทยาศาสตร์ วิศวกรรมศาสตร์ และสังคมศาสตร์ อันเป็นผลมาจากการที่เราสามารถลดรูปปัญหาต่างๆ ทั้งแบบต่อเนื่องหรือวิยุต (continuous or discrete) และทั้งแบบเชิงเส้นหรือไม่เชิงเส้น (linear or nonlinear) ให้เป็นระบบเชิงเส้น (linear system) และหาผลเฉลยแบบแม่นตรงหรือโดยประมาณ (exact or approximate solution) โดยใช้เทคนิคของพีชคณิตเชิงเส้นได้

ผู้เขียนเขียนบทความทางวิชาการนี้ขึ้นเพื่อเน้นย้ำความสำคัญของพีชคณิตเชิงเส้น โดยยกตัวอย่างการประยุกต์พีชคณิตเชิงเส้นในกระบวนการคำนวณค่าเพจแรงค์ (PageRank) เพื่อใช้ในการจัดลำดับเว็บเพจของกูเกิ้ล (Google) ซึ่งเป็นเครื่องมือค้นหา (search engine) เว็บเพจที่ได้รับความนิยมสูงสุด ผู้เขียนเรียบเรียงเนื้อหาส่วนใหญ่ในบทความนี้จาก [1] ถ้าผู้เขียนไม่ได้ระบุเอกสารอ้างอิงสำหรับข้อมูลใด ขอให้ผู้อ่านเข้าใจว่าข้อมูลนั้นมีที่มาจาก [1]

2. การจัดลำดับเว็บเพจของกูเกิ้ล

เซอร์เกย์ บริน (Sergey Brin) และ ลอว์เรนซ์ เพจ (Lawrence Page) ผู้ก่อตั้งกูเกิ้ล คิดวิธีการจัดลำดับเว็บเพจ [2] ในระหว่างศึกษาระดับปริญญาเอกที่มหาวิทยาลัยสแตนฟอร์ด (Stanford University) ประเทศสหรัฐอเมริกา

การจัดลำดับเว็บเพจเริ่มต้นด้วยการใช้โปรแกรม “แมงมุม” (spider) เพื่อเก็บข้อมูลของเว็บเพจแต่ละเว็บเพจ โดยข้อมูลที่เก็บนั้นมีสองประเภทหลัก คือ

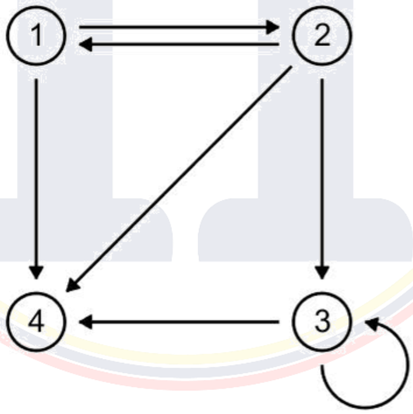
- เนื้อหา (content) ได้แก่ ชื่อของเว็บเพจ คำสำคัญ (keyword) และความถี่ของคำสำคัญเหล่านั้น
- โครงสร้าง (structure) ได้แก่ ไฮเปอร์ลิงค์ หรือ ลิงค์ (hyperlink or link) ที่ออกจากเว็บเพจนั้นไปยังเว็บเพจอื่นๆ

เมื่อแมงมุมเก็บข้อมูลของเว็บเพจหนึ่งเสร็จแล้ว ก็จะ “ไต่” ลิงค์ออกจากเว็บเพจนั้นไปยังเว็บเพจอื่นๆ เพื่อเก็บข้อมูลต่อไป การเก็บข้อมูลของเว็บเพจโดยแมงมุมนี้เกิดขึ้นก่อนที่ผู้ใช้จะใช้เครื่องมือค้นหาเว็บเพจ

การจัดลำดับเว็บเพจใช้ผลรวมของคะแนนที่ได้มาจากข้อมูลทั้งสองประเภท คือ

- คะแนนเนื้อหา (content score) ซึ่งคำนวณขึ้นหลังจากผู้ใช้ป้อนคำสำคัญให้กับเครื่องมือค้นหาเว็บเพจ โดยการวิเคราะห์ความสอดคล้องระหว่างคำสำคัญที่ผู้ใช้ป้อน และคำสำคัญของเว็บเพจต่างๆ
- คะแนนนิยม (popularity score) หรือ ค่าเพจแรงค์ ซึ่งคำนวณขึ้นก่อนที่ผู้ใช้จะใช้เครื่องมือค้นหาเว็บเพจ โดยการวิเคราะห์ลิงค์ (link analysis) ระหว่างเว็บเพจต่างๆ

ในบทความนี้ ผู้เขียนจะอธิบายกระบวนการคำนวณค่าเพจแรงค์เท่านั้น โดยยกตัวอย่างกรณีที่มีเว็บเพจเพียงสี่เว็บเพจดังรูปที่ 1 เพื่อให้ผู้อ่านสามารถเข้าใจกระบวนการดังกล่าวได้อย่างเป็นรูปธรรม



รูปที่ 1 แผนภาพแสดงเว็บเพจและลิงค์ระหว่างเว็บเพจสี่เว็บเพจ

ในรูปที่ 1 วงกลมแต่ละวงแทนเว็บเพจแต่ละเว็บเพจ และลูกศรออกจากวงกลมที่ i ไปยังวงกลมที่ j แทนลิงค์ออกจากเว็บเพจที่ i ไปยังเว็บเพจที่ j โดยมีข้อสังเกตจากรูปที่ 1 ดังนี้

- การที่แมงมุมสามารถเก็บข้อมูลของเว็บเพจแต่ละเว็บเพจได้หมายความว่าจะต้องมีลิงค์มายังเว็บเพจเหล่านั้น (อาจจะยกเว้นเว็บเพจตั้งต้นที่แมงมุมเก็บข้อมูล แต่เว็บเพจตั้งต้นมักจะเป็นเว็บเพจที่เป็นที่รู้จัก และมักจะมีลิงค์มายังเว็บเพจดังกล่าว) เมื่อเป็นเช่นนี้จึงไม่มีเว็บเพจที่โดดเดี่ยว
- อาจจะเป็นไปได้ว่ามีลิงค์มากกว่าหนึ่งลิงค์ออกจากเว็บเพจที่ i ไปยังเว็บเพจที่ j แต่เราใช้ลูกศรออกจากวงกลมที่ i ไปยังวงกลมที่ j เพียงอันเดียว
- เว็บเพจบางเว็บเพจมีลิงค์กลับมาที่เว็บเพจนั้นเอง เช่น เว็บเพจที่ 3

3. การคำนวณค่าเพจแรงค์

สมมติว่ามีเว็บเพจทั้งหมด N เว็บเพจ เราสามารถเก็บข้อมูลโครงสร้างของเว็บเพจต่างๆ โดยใช้เมทริกซ์ประชิด (adjacency matrix) A ขนาด $N \times N$ โดยที่ $A_{ij} = 1$ ถ้ามีลิงค์ออกจากเว็บเพจที่ i ไปยังเว็บเพจที่ j และสมาชิกตัวอื่นของ A เป็น 0 ทั้งหมด จากรูปที่ 1 (ซึ่ง $N = 4$) จะได้

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

กำหนดให้ r_i แทนค่าเพจแรงค์ของเว็บเพจที่ i โดยที่ $1 \le i \le N$ การคำนวณค่าเพจแรงค์เริ่มต้นด้วยหลักการอย่างง่ายสองหลักการ คือ

- ค่าเพจแรงค์ของเว็บเพจหนึ่งต้องได้รับอิทธิพลจากค่าเพจแรงค์ของเว็บเพจอื่นที่มีลิงค์มายังเว็บเพจนั้น
- ถ้ามีลิงค์ออกจากเว็บเพจหนึ่งไปยังเว็บเพจอื่น แล้วอิทธิพลจากค่าเพจแรงค์ของเว็บเพจนั้นต่อเว็บเพจอื่นต้องแปรผกผันกับจำนวนของลิงค์ออกจากเว็บเพจนั้น

เราสามารถสร้างระบบสมการเชิงเส้นที่มี $r_1, r_2, \ldots, r_N$ เป็นตัวแปรให้สอดคล้องกับหลักการอย่างง่ายสองหลักการนี้ สำหรับ $1 \le i \le N$ ดังนี้

$$r_i = \sum_j \frac{r_j}{n_j}$$

โดยหาผลรวมนี้สำหรับทุกค่า j ที่มีลิงค์ออกจากเว็บเพจที่ j มายังเว็บเพจที่ i และ n_j คือ จำนวนของลิงค์ออกจากเว็บเพจที่ j (ให้สังเกตว่า $n_j \neq 0$ โดยนิยามของ j)

จากรูปที่ 1 เราสามารถเขียนระบบสมการ (2) ได้ดังนี้

$$\begin{aligned} r_1 &= \frac{r_2}{3} \\ r_2 &= \frac{r_1}{2} \\ r_3 &= \frac{r_2}{3} + \frac{r_3}{2} \\ r_4 &= \frac{r_1}{2} + \frac{r_2}{3} + \frac{r_3}{2} \end{aligned} \tag{3}$$

ซึ่งสามารถเขียนในรูปสมการเมทริกซ์ได้เป็น

$$r = rH \tag{4}$$

โดยที่

$$r = \begin{bmatrix} r_1 & r_2 & r_3 & r_4 \end{bmatrix} \tag{5}$$

และ

$$H = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

ให้สังเกตว่าเมทริกซ์ H (ซึ่งย่อมาจาก hyperlink matrix) ได้จากการหารแต่ละแถวของเมทริกซ์ A ด้วยจำนวนของ 1 ในแถวนั้น (ถ้าสมาชิกในแถวนั้นไม่ได้เป็น 0 ทั้งหมด)

เมื่อแก้สมการ (4) จะพบว่า r เป็นเวกเตอร์ศูนย์ซึ่งเป็นคำตอบที่ไม่มีประโยชน์ เพราะไม่ช่วยในการจัดลำดับเว็บเพจ เมื่อเป็นเช่นนี้จึงต้องดัดแปลงเมทริกซ์ H ด้วยหลักการเสริมสองหลักการซึ่งสอดคล้องกับพฤติกรรมการเข้าชมเว็บเพจของผู้ใช้ ดังนี้

- เมื่อผู้ใช้เข้าชมเว็บเพจที่ไม่มีลิงค์ไปยังเว็บเพจอื่น (เช่น เว็บเพจที่ 4) ผู้ใช้ยังสามารถเข้าชมเว็บเพจอื่นต่อได้โดยการพิมพ์ที่อยู่ของเว็บเพจอื่นโดยตรง ดังนั้น เราจะพิจารณาเว็บเพจที่ไม่มีลิงค์ไปยังเว็บเพจอื่นเสมือนว่าเป็นเว็บเพจที่มีลิงค์ไปยังเว็บเพจอื่นทุกเว็บเพจ (รวมทั้งเว็บเพจนั้นเองด้วย) เมื่อเป็นเช่นนี้เราจะสร้างเมทริกซ์ใหม่ S ด้วยการแทนที่แถวของเมทริกซ์ H ที่มีสมาชิกทุกตัวเป็น 0 ด้วยแถวที่มีสมาชิกทุกตัวเป็น $1/N$ จากสมการ (6) จะได้

$$S = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

  เมทริกซ์ S ที่ได้จากการดัดแปลงเมทริกซ์ H ด้วยหลักการเสริมนี้จะเป็นเมทริกซ์สโตแคสติก (stochastic matrix) เนื่องจากสมาชิกทุกตัวไม่เป็นลบ และผลรวมของสมาชิกในแต่ละแถวเท่ากับ 1
- ผู้เข้าชมเว็บเพจแต่ละเว็บเพจอาจไม่ได้ใช้ลิงค์ออกจากเว็บเพจนั้นเพื่อเข้าชมเว็บเพจอื่นเสมอไป โดยอาจใช้การพิมพ์ที่อยู่ของเว็บเพจอื่นโดยตรง เมื่อเป็นเช่นนี้เราจะสร้างเมทริกซ์ใหม่ G (ซึ่งย่อมาจาก Google matrix)

ให้เป็นผลรวมของเมทริกซ์ S (ซึ่งสอดคล้องกับวิธีใช้ลิงค์) และเมทริกซ์ที่มีสมาชิกทุกตัวเป็น $1/N$ (ซึ่งสอดคล้องกับวิธีพิมพ์ที่อยู่โดยตรง) โดยถ่วงน้ำหนักเมทริกซ์ทั้งสองด้วยความน่าจะเป็นที่ผู้เข้าชมเว็บเพจจะใช้วิธีแต่ละวิธี จากสมการ (7) จะได้

$$G = \alpha \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} + (1-\alpha) \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad (8)$$

จากการสำรวจพฤติกรรมการเข้าชมเว็บเพจของผู้ใช้พบว่า α มีค่าประมาณ 0.85 (หรือ 17/20) ดังนั้น

$$G = \begin{bmatrix} \frac{3}{80} & \frac{37}{80} & \frac{3}{80} & \frac{37}{80} \\ \frac{77}{240} & \frac{3}{80} & \frac{77}{240} & \frac{77}{240} \\ \frac{3}{80} & \frac{3}{80} & \frac{37}{80} & \frac{37}{80} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad (9)$$

ให้สังเกตว่านอกจาก G จะเป็นเมทริกซ์สโตแคสติกแล้ว G ยังมีสมาชิกทุกตัวเป็นบวกด้วย สมบัติดังกล่าวของ G ทำให้สามารถประยุกต์ทฤษฎีของเพอร์รอน-โฟรบีเนียส (Perron-Frobenius Theorem) ใช้กับ G ได้ข้อสรุป ดังนี้ [3]

- G มีค่าลักษณะเฉพาะ (eigenvalue) เท่ากับ 1 ซึ่งเป็นค่าลักษณะเฉพาะที่มีค่าสัมบูรณ์มากที่สุดเพียงค่าเดียวของ G

- G มีเวกเตอร์ลักษณะเฉพาะ (eigenvector) สำหรับค่าลักษณะเฉพาะ 1 ที่มีสมาชิกทุกตัวเป็นบวก และปริภูมิลักษณะเฉพาะ (eigenspace) สำหรับค่าลักษณะเฉพาะ 1 เป็นปริภูมิหนึ่งมิติ

เมื่อเป็นเช่นนี้จึงมั่นใจได้ว่าสมการ

$$r = rG \tag{10}$$

ซึ่งได้มาจากการแทนที่เมทริกซ์ H ในสมการ (4) ด้วยเมทริกซ์ G ต้องมีคำตอบเป็นเวกเตอร์ r ที่มีสมาชิกทุกตัวเป็นบวก และถ้ามีเงื่อนไขเพิ่มเติมว่า

$$\sum_{i=1}^{N} r_i = 1 \tag{11}$$

เพื่อให้ r เป็นเวกเตอร์สโตแคสติก แล้วจะมีเวกเตอร์ r ดังกล่าวเพียงเวกเตอร์เดียวเท่านั้น เมื่อแก้สมการ (10) และ (11) จะได้

$$r = \begin{bmatrix} 0.166 & 0.185 & 0.289 & 0.360 \end{bmatrix} \tag{12}$$

การที่ r เป็นเวกเตอร์สโตแคสติกและเวกเตอร์ลักษณะเฉพาะสำหรับค่าลักษณะเฉพาะ 1 ของเมทริกซ์ G ซึ่งเป็นเมทริกซ์สโตแคสติก กล่าวคือ r เป็นเวกเตอร์นิ่ง (stationary vector) ของเมทริกซ์ G ทำให้เราสามารถตีความค่าเพจแรงค์โดยนัยของความน่าจะเป็นได้ เช่น ความน่าจะเป็นที่ผู้ใช้จะเข้าชมเว็บเพจที่ 4 เท่ากับ 0.360 (ค่าประมาณถูกต้องถึงทศนิยมตำแหน่งที่สาม)

4. จากทฤษฎีสู่ปฏิบัติ

การคำนวณเวกเตอร์ r จากสมการ (10) และ (11) โดยวิธีตรง (direct method) เช่น การกำจัดแบบเกาส์เซียน (Gaussian elimination) สามารถทำได้

กับเมทริกซ์ขนาดไม่ใหญ่มาก ($N \approx 10^3$) เนื่องจากความซับซ้อนเชิงคณนา (computational complexity) ของวิธีดังกล่าวเท่ากับ $O\left(N^3\right)$

ในทางปฏิบัติซึ่ง $N \approx 10^9$ การคำนวณเวกเตอร์ r โดยประมาณใช้วิธีอ้อม (indirect method) เช่น วิธียกกำลัง (power method) ซึ่งเริ่มต้นด้วยเวกเตอร์สโตแคสติก $r^{(0)}$ (เช่น เวกเตอร์ที่มีสมาชิกทุกตัวเป็น $1/N$) และสำหรับ $k \geq 1$ กำหนดให้

$$r^{(k)} = r^{(k-1)}G \tag{13}$$

การที่ 1 เป็นค่าลักษณะเฉพาะที่มีค่าสัมบูรณ์มากที่สุดเพียงค่าเดียวของ G ส่งผลให้ลำดับของเวกเตอร์สโตแคสติก $r^{(0)}, r^{(1)}, r^{(2)}, \ldots$ ลู่เข้าหาเวกเตอร์ r เสมอ (ถ้าเวกเตอร์ $r^{(0)}$ ไม่ตั้งฉากกับเวกเตอร์ r) กล่าวคือ [3]

$$\lim_{k\to\infty} r^{(k)} = r \tag{14}$$

มีข้อสังเกตสองประการเกี่ยวกับประสิทธิภาพของวิธียกกำลังในการหาเวกเตอร์ r ดังนี้

- อัตราที่เวกเตอร์ $r^{(k)}$ ลู่เข้าหาเวกเตอร์ r ใกล้เคียงกับอัตราที่ α^k ลู่เข้าหาศูนย์ กล่าวคือ

$$\lim_{k\to\infty} \frac{\left\| r^{(k)} - r \right\|}{\left\| r^{(k-1)} - r \right\|} \approx \alpha \tag{15}$$

เนื่องจากค่าลักษณะเฉพาะที่มีค่าสัมบูรณ์มากเป็นอันดับที่สองของ G มีค่าสัมบูรณ์ประมาณ α

สมการ (15) หมายความว่าผลต่างระหว่างเวกเตอร์ $r^{(k)}$ และเวกเตอร์ r ลดลงแบบเรขาคณิตเมื่อ k มีค่ามาก (โดยหาผลต่างดังกล่าวด้วยนอร์มใดก็ได้ เพื่อให้ง่ายต่อการทำความเข้าใจอาจจะใช้นอร์มแบบ L^{∞}) ดังนั้น

$$\left\|r^{(k)}-r\right\| \approx \alpha^{k}\left\|r^{(0)}-r\right\| \tag{16}$$

ถ้าผลต่าง $\left\|r^{(0)}-r\right\|$ มีค่าประมาณ 1 (ให้สังเกตว่าถ้าใช้นอร์มแบบ L^{∞} แล้วผลต่างดังกล่าวมีค่าขอบเขตบนเท่ากับ 1) จะได้

$$\left\|r^{(k)}-r\right\| \approx \alpha^{k} \tag{17}$$

ถ้าเราตั้งเงื่อนไขให้ค่าประมาณของเพจแรงค์ถูกต้องถึงทศนิยมตำแหน่งที่สามโดยการกำหนดให้ $\left\|r^{(k)}-r\right\|$ มีค่าประมาณ 0.0005 แล้ว

$$k \approx \log_{\alpha} 0.0005 \approx 47 \tag{18}$$

ซึ่งหมายความว่าต้องใช้วิธียกกำลังประมาณ 50 ครั้ง ให้สังเกตว่าค่าของ k ดังกล่าวเป็นเพียงค่าประมาณอย่างหยาบ ๆ เท่านั้น สิ่งที่สำคัญคือการที่ค่าของ k ไม่ขึ้นกับค่าของ N กล่าวคือ $k=O(1)$

- การคำนวณเวกเตอร์ $r^{(k)}$ โดยการคูณเวกเตอร์ $r^{(k-1)}$ ด้วยเมทริกซ์ G ดังสมการ (13) มีความซับซ้อนเชิงคณนาเท่ากับ $O\left(N^{2}\right)$ แต่เราสามารถลดความซับซ้อนเชิงคณนาดังกล่าวโดยสังเกตว่า

$$\begin{aligned} G &= \alpha S+\frac{1}{N}(1-\alpha)u^{T}u \\ &= \alpha\left(H+\frac{1}{N}v^{T}u\right)+\frac{1}{N}(1-\alpha)u^{T}u \\ &= \alpha H+\frac{1}{N}\left(\alpha v^{T}+(1-\alpha)u^{T}\right)u \end{aligned} \tag{19}$$

เมื่อ u คือ เวกเตอร์แถวที่มีสมาชิกทุกตัวเป็น 1 และ v คือ เวกเตอร์แถวที่มีสมาชิกตัวที่ i เป็น 1 ถ้าไม่มีลิงค์ออกจากเว็บเพจที่ i และมีสมาชิกตัวอื่นเป็น 0 ทั้งหมด ให้สังเกตว่า H เป็นเมทริกซ์ที่มีสมาชิกส่วนใหญ่เป็น 0 เพราะมีลิงค์ออกจากเว็บเพจส่วนใหญ่ประมาณ 10 ลิงค์เท่านั้น (กล่าวคือ H มีสมาชิกที่ไม่เป็น 0 ประมาณ $10N$ ตัว) เมื่อเป็นเช่นนี้การคูณ $r^{(k-1)}$ ด้วย αH จึงมีความซับซ้อนเชิงคณนาเท่ากับ $O(N)$ เท่านั้น และถ้ากำหนดให้

$$w = r^{(k-1)}\left(\frac{1}{N}\left(\alpha v^T + (1-\alpha)u^T\right)u\right) \tag{20}$$

แล้วจะเห็นว่า w เป็นเวกเตอร์ที่คำนวณได้ง่าย เนื่องจากสมาชิกทุกตัวเท่ากัน กล่าวคือ

$$w = \frac{1}{N}\left(\alpha\sum_{i=1}^{N}\left(r_i^{(k-1)}v_i\right)+1-\alpha\right)u \tag{21}$$

ดังนั้น การคำนวณ w จึงมีความซับซ้อนเชิงคณนาเท่ากับ $O(N)$ เมื่อเป็นเช่นนี้การคำนวณเวกเตอร์ $r^{(k)}$ จากเวกเตอร์ $r^{(k-1)}$ จึงมีความซับซ้อนเชิงคณนาเท่ากับ $O(N)$ เช่นกัน

จากข้อสังเกตทั้งสองนี้ เราจึงสรุปได้ว่าการคำนวณเวกเตอร์ r โดยวิธียกกำลังมีความซับซ้อนเชิงคณนาเท่ากับ $O(N)$ ซึ่งมีประสิทธิภาพสูง

5. บทสรุป

ผู้เขียนได้อธิบายกระบวนการคำนวณค่าเพจแรงค์ของเว็บเพจต่างๆ แต่ยังมีรายละเอียดปลีกย่อยอื่นๆ ซึ่งผู้เขียนไม่สามารถกล่าวถึงได้หมดในโอกาสนี้ เช่น ถ้ามีเว็บเพจเก่าถูกลบออกไปหรือมีเว็บเพจใหม่เพิ่มเข้ามา เราสามารถคำนวณค่าเพจแรงค์ใหม่ได้ โดยไม่จำเป็นต้องเริ่มกระบวนการนี้ใหม่ตั้งแต่ต้น เป็นต้น

ผู้อ่านได้เห็นแล้วว่าพีชคณิตเชิงเส้นมีส่วนเกี่ยวข้องกับกระบวนการคำนวณค่าเพจแรงค์ในเกือบจะทุกขั้นตอน ตั้งแต่การสร้างระบบสมการที่เหมาะสมไปจนถึงการคำนวณค่าเพจแรงค์โดยประมาณด้วยวิธียกกำลัง ผู้เขียนหวังเป็นอย่างยิ่งว่าผู้อ่านจะได้ใช้ตัวอย่างของการคำนวณค่าเพจแรงค์เพื่อเป็นตัวอย่างหนึ่ง (จากหลายๆ ตัวอย่าง) ของความสำคัญของพีชคณิตเชิงเส้น เพื่อผลักดันให้เกิดความตื่นตัวในการเรียนการสอนพีชคณิตเชิงเส้นมากขึ้นต่อไป

เอกสารอ้างอิง

[1] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond:The Science of Search Engine Rankings*, Princeton, USA: Princeton University Press, 2006.

[2] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, vol. 33, pp. 107-117, 1998.

[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge University Press, 2012.