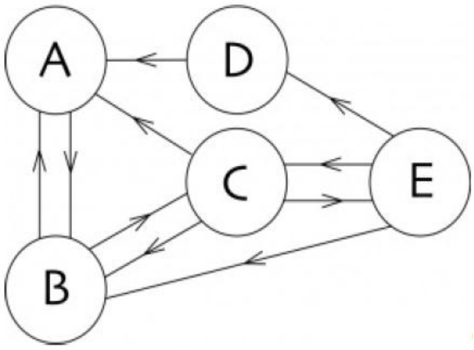
អត្ថបទដកស្រង់ចេញពី

Klein Project Blog

បែបបទដែល Google ធ្វើការ: ចំណងទំនាក់ទំនង Markov និងតម្លៃ eigenvalues



រៀបរៀងដោយអ្នកនិពន្ធ: Christiane Rousseau។

តាំងពីការចាប់ផ្តើមដំបូងបំផុតរបស់ Google, Google បានក្លាយជាឧបករណ៍ស្រាវជ្រាវដំបូងបំផុត ។ ការតម្រៀបទិន្នន័យទាំងនេះតាម algorithm (The PageRank algorithm) ជាមួយនឹងទិន្នន័យជាច្រើនទំព័រនៅលើ World-Wide-Web ហើយការស្រាវជ្រាវជាច្រើន បញ្ចប់ទៅដោយលទ្ធផលជាច្រើនលានលទ្ធផល។ ប្រសិនបើការរាវរកមិនបានត្រឹមត្រូវតាមលំដាប់លំដោយ នោះការស្វែងរករបស់យើងត្រូវបរាជ័យ និងគ្មាននរណាម្នាក់អាចស្រាវជ្រាវរកឃើញទិន្នន័យដែលខ្លួនចង់បាននៅក្នុងចំណោមទិន្នន័យរាប់សិបលានដែលមាននៅលើ world-wide-Web។

តើ PageRank algorithm ធ្វើការរបៀបម៉េច?

យើងនឹងពន្យល់អំពីបញ្ហាទាំងអស់នេះ។ ប៉ុន្តែមុនដំបូង យើងបង្ហាញការស្រាវជ្រាវមួយនៅលើGoogle នៅថ្ងៃទី ៤ ខែ មិថុនា ឆ្នាំ ២០១០ ដែលទទួលបាន ១៦,៣០០,០០០ លទ្ធផល សម្រាប់ Klein project ដែលនេះគ្រាន់តែជាការចាប់ផ្តើមលើកដំបូងតែប៉ុណ្ណោះរបស់គម្រោងនេះ។

ជាដំបូង យើងអាចភ្ជាប់ចូលទៅកាន់:

http://www.mathunion.org/icmi/other-activities/klein-project/introduction/

ហើយក៏អាចចូលទៅកាន់:

http://www.kleinproject.org/

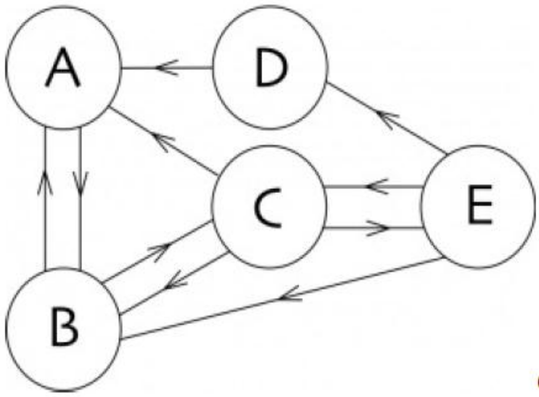
វេបសាយចម្បងគេនៅលើ Internet គឺ URL (URL= Uniform Resource Location) ដែលបង្ហាញនៅលើវេបសាយសហគមន៍គណិតវិទ្យាអន្តរជាតិ (The International Mathematical Union) : http://www.mathunion.org។ ពីព្រោះសហគមន៍គណិតវិទ្យាអន្តរជាតិ គឺជាចំណុចសំខាន់នៅលើវេបសាយពេលដែលយើងបង្កើតការស្រាវជ្រាវ “ International Mathematical Union ” ម្យ៉ាងវិញទៀតវាធ្វើការ ទំនាក់ទំនងទៅលើផ្នែកសំខាន់ៗជាច្រើននៃទំព័រវេបសាយ ។ ហើយទំព័រមួយក្នុងចំណោមទំព័រទាំងនោះគឺ

http://www.mathunion.org/icmi/other-activities/klein-project/introduction/

ចាប់ពីពេលនេះ បន្តទៅមុខក្នុងកំឡុងពេល ២ ទៅ ៣ខែ ឬ ២ ទៅ ៣ឆ្នាំ យើងអាចជឿជាក់ថាទំព័រវេបសាយ http://www.kleinproject.org/ នឹងបង្ហាញដំបូងបង្អស់សម្រាប់ការស្រាវជ្រាវ klein project ។

ដើម្បីពន្យល់អំពី algorithm យើងយកគំរូតាមក្រាបទិសដៅដែលបានបង្ហាញខាងលើ ។ យើងកំណត់យកផ្ទាំងមូលៗតំណាងអោយទំព័រដែលស្ថិតនៅលើវេបសាយ និងសញ្ញាបញ្ជាក់ទិសដៅតំណាងអោយការភ្ជាប់ទំនាក់ទំនងពីទំព័រមួយទៅទំព័រមួយ ។ ដូច្នេះទំព័រវេបសាយនីមួយៗមានទំនាក់ទំនងគ្នាជាមួយទំព័រដ៏ទៃទៀត នៅលើប្រព័ន្ធវេបសាយទាំងនេះ ។ ក៏ប៉ុន្តែភាគច្រើនទំព័រដែលយើងត្រូវការតាម algorithm នឹងបង្ហាញ ទំព័រណាដែលល្អ ហើយមានសារៈសំខាន់បំផុត ។

ឧទាហរណ៍ៈ



ពួកយើងក្រឡេកមើលវេបសាយសាមញ្ញមួយ ដែលបានបង្ហាញនូវខាងលើនេះដែលមានឈ្មោះសម្រាប់ ៥ទំព័រគឺ A, B, C, D និង E ។ នៅលើវេបសាយមួយ វាអាចភ្ជាប់ជាមួយទំព័រផ្សេងទៀតបាន ២ រឺ ៣ទំព័រ ។ ប្រសិនបើយើងបើកទំព័រ Aមួយ នោះយើងមានតែមួយទំព័រគត់ដែលអាចបង្ហាញបាន នោះគឺទំព័រ B ភ្លាមនោះដែរ ប្រសិនបើយើងបើកទំព័រ C យើងនឹងបានទំព័របីផ្សេងទៀត ដែលមានទំនាក់ទំនងជាមួយវា ហើយយើងអាចផ្លាស់ប្តូរទៅទំព័រផ្សេងទៀតបានដោយងាយ ដូចជាទៅទំព័រ A រឺ B រឺ E ជាដើម ។ ដូចនេះយើងត្រូវចងចាំថាយ៉ាងហោចណាស់ក៏មាន មួយទំព័រយ៉ាងតិចនៅពេលដែលយើងបានបើកចូលទំព័រណាមួយហើយនោះ ។

យើងចាប់ផ្តើមសាកល្បងលេងល្បែងនេះ ជ្រើសរើសយកមួយទំព័រដោយចៃដន្យណាមួយក៏បាន បន្ទាប់មក បន្តអនុវត្តការភ្ជាប់ទៅទំនាក់ទំនាក់ចំពោះទំព័រដ៏ទៃទៀត ។ ឧទាហរណ៍ៈ ប្រសិនបើយើងចាប់ផ្តើមនៅលើទំព័រ B នោះបន្ទាប់មកយើងអាចធ្វើដំណើរបន្តទៅទំព័រ A រឺ C ផ្សេងទៀត ជាមួយប្រូបាប 1/2 សម្រាប់ករណីនីមួយៗ ។ ប៉ុន្តែប្រសិនបើយើង ចាប់ផ្តើមនៅលើទំព័រ D នោះយើងអាចធ្វើដំណើរទៅទំព័រ A តែមួយគត់ ជាមួយនឹងប្រូបាបស្មើ 1។ តាមរយៈល្បែងនេះ

តើយើងនឹងបានទំព័រណា បន្ទាប់ពីយើងធ្វើដំណើរបាន n ជំហាន?

យើងសង្ខេបទិន្នន័យទាំងនេះ ក្នុងម៉ាទ្រីស P ដែលយកធាតុនៃម៉ាទ្រីសតាមជួរដេក និង ជួរឈរ ជាតម្លៃប្រូបាបនៃករណីនីមួយៗ ។ ទិន្នន័យដែលយើងទទួលបានគឺ

$$
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
P = \begin{pmatrix} 0 & \tfrac{1}{2} & \tfrac{1}{3} & 1 & 0 \\ 1 & 0 & \tfrac{1}{3} & 0 & \tfrac{1}{3} \\ 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{3} \\ 0 & 0 & 0 & 0 & \tfrac{1}{3} \\ 0 & 0 & \tfrac{1}{3} & 0 & 0 \end{pmatrix} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}
\end{array}
$$

យើងសម្គាល់ឃើញថា ផលបូកតម្លៃតាមជួរឈរនីមួយៗមានតម្លៃស្មើនឹង ១ ហើយធាតុទាំងអស់នោះមានតម្លៃធំជាង ឬ ស្មើនឹង ០ ។ ម៉ាទ្រីសនេះមានពីរលក្ខណៈសំខាន់គឺ ម៉ាទ្រីសនីមួយៗ គឺជាម៉ាទ្រីសនៃ Markov chain process ដែលគេអាចហៅថា Markov transition matrix។ វាតែងតែមានតម្លៃស្មើនឹង ១ ជានិច្ច វាដូចគ្នាទៅនឹងតម្លៃ eigenvalue ដែលតម្លៃនេះមានចំពោះវ៉ិចទ័រ eigenvector ជាមួយតម្លៃ eigenvalue ស្មើនឹង១ ហើយធាតុនីមួយៗរបស់ម៉ាទ្រីសមានតម្លៃនៅចន្លោះពី [0,1] និងផលបូកជួរឈរមានតម្លៃស្មើនឹង ១ ។ ប៉ុន្តែ មុននឹងរំលឹកឡើងវិញនូវនិយមន័យរបស់តម្លៃ eigenvalue និងវ៉ិចទ័រ eigenvector យើងធ្វើការស្រាវជ្រាវពី ប្រយោជន៍នៃម៉ាទ្រីសដែលបានបង្ហាញតាមរយៈក្រាហ្វនៃវេបសាយ ។

យើងគិតពិចារណាទៅលើអត្រាបម្រែបម្រួលចៃដន្យនៃ X_n ជាមួយនឹងតម្លៃនៅក្នុងគេហទំព័រទាំងនេះ $\{A,B,C,D,E\}$ ដែលមាន N ទំព័រ (ពេលនេះ N=5) ។ X_n តំណាងអោយទំព័រដែលយើងបន្ទាប់ពី n ជំហាន ។ យើងកំណត់ p_{ij} ជាធាតុនៃម៉ាទ្រីស P ដែល i ជាជួរដេក ហើយ j ជាជួរឈរ , ម្យ៉ាងទៀត p_{ij} គឺជា ប្រូបាបមានលក្ខខណ្ឌ ដែលទំព័រទី i គឺជាតួទី n+1 ហើយ ទំព័រទី j ជាតួទី n ។

$$P_{ij} = \Pr\mathrm{ob}(X_{n+1} = i | X_n = i)$$

ចាំថា ប្រូបាបនេះ គឺមានករណីអាចស្មើនឹង n! ដែលយើងអាចនិយាយបានថា Markov chain process ពុំមានការចងចាំនៅក្នុងអតីតកាលឡើយ ។ វាមិនពិបាកក្នុងការគូសបន្ទាត់តំណាងប្រូបាបបន្ទាប់ពីធ្វើបានពីរជំហាននោះទេ ដែលយើងអាចសង្ខេបវាបាននៅក្នុងម៉ាទ្រីស P^2 ។

យើងបង្ហាញទ្រឹស្តីបទប្រូបាប ដែលមានប្រយោជន៍មួយចំនួនៈ

$$\Pr\mathrm{ob}(X_{n+2} = i | X_n = j) = \sum_{k=1}^{N} \Pr\mathrm{ob}(X_{n+2} = i \ \ and \ \ X_{n+1} = k | X_n = j)$$

និយមន័យប្រូបាបមានលក្ខខណ្ឌគឺ

$$\Pr\mathrm{ob}(X_{n+2} = i | X_n = j) = \sum_{k=1}^{N} \frac{\Pr\mathrm{ob}\left(X_{n+2} = i \ \ and \ \ X_{n+1} = k \ \ and \ \ X_n = j\right)}{\Pr\mathrm{ob}\left(X_n = j\right)}$$

យើងប្រើវិធីដែលធ្លាប់ប្រើកន្លងមកៈ គុណ ចែក នឹងចំនួនតែមួយៈ

$$\Pr\mathrm{ob}(X_{n+2} = i | X_n = j) = \sum_{k=1}^{N} \frac{\Pr\mathrm{ob}\left(X_{n+2} = i \ \ and \ \ X_{n+1} = k \ \ and \ \ X_n = j\right)}{\Pr\mathrm{ob}\left(X_{n+1} = k \ \ and \ \ X_n = j\right)} \frac{\Pr\mathrm{ob}\left(X_{n+1} = k \ \ and \ \ X_n = j\right)}{\Pr\mathrm{ob}\left(X_n = j\right)}$$

ផលចែកដំបូងស្មើនឹង:

$$\Pr ob\left(X_{n+2}=i\middle|X_{n+1}=k\ \ and\ \ X_n=j\right)=\Pr ob\left(X_{n+2}=i\middle|X_{n+1}=k\right)$$

ចាប់តាំងពី Markov chain process មិនអាចចងចាំនៃជំហានពីមុន ដូចនេះ

$$\Pr ob(X_{n+2}=i|X_n=j)=\sum_{k=1}^{N}\Pr ob(X_{n+2}=i\ \ and\ \ X_{n+1}=k|X_n=k)\Pr ob(X_{n+1}=k|X_n=j)$$

$$=\sum_{k=1}^{N}P_{ik}P_{kj}$$

$$=\left(P^2\right)i_j$$

នៅក្នុងឧទាហរណ៍របស់យើងគឺ

$$\begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ P^2= & \begin{pmatrix} 1/2 & 1/6 & 1/6 & 0 & 11/18 \\ 0 & 2/3 & 4/9 & 1 & 1/9 \\ 1/2 & 0 & 5/18 & 0 & 1/6 \\ 0 & 0 & 1/9 & 0 & 0 \\ 0 & 1/6 & 0 & 0 & 1/9 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}$$

គិតសាឡើងវិញពីគំនិតទាំងនេះ វាច្បាស់ណាស់ថាធាតុ $(P^m)_{ij}$ នៃម៉ាទ្រីស P^m ដែលបានរៀបរាប់ ពីប្រូបាប $\Pr ob\left(X_{n+m}=i\middle|X_n=j\right)$ តួយ៉ាងដូចជា

$$\begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ P^{32}= & \begin{pmatrix} 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}$$

គ្រប់ជួរឈរនៃ P^{32} គឺដូចគ្នាបេះបិទ ប្រសិនបើយើងធ្វើការជ្រើសរើសផ្នែកទសភាគ៣ខ្ទង់ ហើយគ្រប់ ជួរឈរទាំងអស់នៃ P^n ពេលដែល n>32 ។ បើសិនជាយើងជ្រើសរើសតម្លៃដែលត្រឹមត្រូវខ្ពស់ជាងនេះ យើងសង្កេតឃើញតម្លៃនេះមិនប្រែប្រួលទេ ទោះបី n ធំជាង ៣២ ក៏ដោយ ។ ដូចនេះបន្ទាប់ពី n ជំហានមក តម្លៃដែលយើងត្រូវយកគឺវាធំល្មមគ្រប់គ្រាន់ ហើយតម្លៃប្រូបាបនៃទំនាក់ទំនងទំព័រនីមួយៗមិនជាប់គ្នាទៅនឹងទំព័រ ដែលយើងបានចាប់ផ្តើមដំបូងនោះទេ ។

យើងពិចារណាទៅលើវ៉ិចទ័រ π

$$\pi^t=(0.293,0.390,0.220,0.024,0.073)$$

π គឺជាវ៉ិចទ័រតំណាងអោយជួរឈរ ហើយ π^t គឺវ៉ិចទ័រតំណាងអោយជួរដេក ។ វាងាយស្រួលនឹងពិនិត្យឃើញថា $P\pi=\pi$ ។ បើសិនជាយើងគិតលើ កូអរដោនេទី i នៃវ៉ិចទ័រ π^t គឺវាដូចគ្នានឹងតម្លៃប្រូបាបនៃទំព័រទី i នៅខណៈ n ដូច្នេះ π^t គឺជាតម្លៃប្រូបាបនៃទំព័រដែលបានរៀបចំហើយនៅខណៈ n ហើយវាក៏ជាប្រូបាបនៅខណៈ

n+1 ផងដែរ ។ គេអាចហៅ វ៉ិចទ័រ π ថាជា តម្លៃរបាយមិនប្រែប្រួល ។ តម្លៃរបាយមិនប្រែប្រួលនេះ គឺវារៀបតាមលំដាប់នៃទំព័រទាំងនោះ ។ ឧទាហរណ៍ៈ យើងតម្រៀបទំព័រតាមលំដាប់ B,A,C,E,D ហើយយើងបញ្ជាក់ថាទំព័រ B ជាទំព័រសំខាន់ជាងគេបំផុត ។

ករណីទូទៅៈ

ករណីទូទៅអាចក្លាយជាការអនុវត្តយ៉ាងពិតប្រាកដនៅក្នុងឧទាហរណ៍របស់យើង ។ យើងតាងគេហទំព័រទាំងនោះដោយ ក្រាហ្វទិសដៅមួយ ដែលមាន N ជាជួរឈរ ក្នុងចំនួន N ទំព័រនៃគេហទំព័រនេះ ហើយគែម ទិសដៅតំណាងអោយការតភ្ជាប់គ្នាពីទំព័រមួយទៅទំព័រមួយ ។ យើងអាចសង្ខេបទិន្នន័យទាំងនេះនៅក្នុងម៉ាទ្រីសលំដាប់ $N \times N$ នៃម៉ាទ្រីស P ជាមួយនឹងជួរឈរ j ជាផ្នែកទី j នៃទំព័រដែលចាកចេញ និង ជួរដេក i ជាផ្នែកទី i នៃទំព័រដែលត្រឡប់មកដល់ ។ នៅក្នុងឧទាហរណ៍យើង យើងបានរកឃើញវ៉ិចទ័រ π បំពេញដោយ $P\pi = \pi$ ។ វ៉ិចទ័រនេះជាវ៉ិចទ័រ eigenvector ដែលមានតម្លៃ eigenvalue ស្មើនឹង ១ ។ ពួកយើងរំលឹកឡើងវិញនូវនិយមន័យ eigenvector និង eigenvalue ៖

និយមន័យៈ គេអោយ P ជាម៉ាទ្រីសលំដាប់ $N \times N$ ។ $\lambda \in \mathbb{C}$ គឺជាតម្លៃ eigenvalue នៃម៉ាទ្រីស P ប្រសិនបើ X មិនមែនជាវ៉ិចទ័រសូន្យ ដែល $X \in \mathbb{C}^N$ នោះគេបាន $PX = \lambda X$ ។ ហើយវ៉ិចទ័រ X គេហៅថាវ៉ិចទ័រ eigenvector នៃម៉ាទ្រីស P ។

យើងអាចរំលឹកពីរបៀបរកតម្លៃ eigenvalue និងវ៉ិចទ័រ eigenvector ផងដែរៈ

អំណះអំណាងៈ គេអោយ P គឺជាម៉ាទ្រីសលំដាប់ $N \times N$ ។ តម្លៃ eigenvalue នៃម៉ាទ្រីស P គឺជាឫសនៃកន្សោមពហុធា $\det(\lambda I - P) = 0$ ដែល I ជាវ៉ិចទ័រឯកតាលំដាប់ $N \times N$ ហើយវ៉ិចទ័រ eigenvector នៃតម្លៃ eigenvalue λ គឺជាចម្លើយមិនសូន្យ នៃប្រព័ន្ធសមីការលីនេអ៊ែរ $(\lambda I - P)X = 0$ ។

តាមទ្រឹស្តីបទ Frobenius ធានាថា សម្រាប់ម៉ាទ្រីសដែលជាប់ទាក់ទងនឹងក្រាហ្វវេបសាយ យើងតែងតែរកឃើញការដោះស្រាយមិនប្រែប្រួលជានិច្ច។

ទ្រឹស្តីបទៈ (Frobenius theorem) យើងពិនិត្យមើលម៉ាទ្រីស Markov លំដាប់ $N \times N$ ក្លាយទៅជាម៉ាទ្រីស $P = (p_{ij})$ (ដែល $p_{ij} \in [0,1]$ ហើយគ្រប់ i, j និងផលបូកនៃតម្លៃតាមជួរដេកនីមួយៗ គឺស្មើនឹង i ដែល $\sum_{i=1}^{N} p_{ij=1}$)។ ហើយបន្ទាប់មក

១) $\lambda = 1$ គឺជាមួយតម្លៃ eigenvalue នៃម៉ាទ្រីស P

២) តម្លៃ eigenvalue λ នីមួយៗ នៃម៉ាទ្រីស P , $|\lambda| \le 1$

៣)	តម្លៃ eigenvalue នៃវ៉ិចទ័រ eigenvector ស្មើនឹង 1 ហើយគ្រប់កូអរដោនេនីមួយៗ មានតម្លៃធំជាង រឺស្មើនឹង 0 ។ ដោយមិនគិតពីឱកាសដែលបាត់បង់ យើងអាចគិតស្មានទុកជាមុនថា ផលបូកលេខនៃកូអរដោនេទាំងនោះស្មើនឹង 1 ។

ពេលនេះយើងស្គាល់អំពីផលវិជ្ជមានរបស់ទ្រឹស្តីបទនេះ។ គោលបំណងរបស់យើង យើងនឹងបង្កើតសម្មតិកម្មមួយដែលថា ម៉ាទ្រីស P មានវ៉ិចទ័រមូលដ្ឋាន eigenvector $B=\{v_1,v_2,...,v_n\}$ និងយើងគិតថា v_1 គឺជាវ៉ិចទ័រ π នៃទ្រឹស្តីបទ Frobenius។ សម្រាប់វ៉ិចទ័រ v_i នីមួយៗត្រូវមានតម្លៃ λ_i ដែល $Pv_i=\lambda_i v_i$ ។ យើងយកវ៉ិចទ័រមិនសូន្យណាមួយនៃ X ហើយ $X^t=(x_1,...,x_N)$ ដែល $x_i\in[0,1]$ និង $\sum_{i=1}^{N}x_i=1$ ។ យើងបំបែកជាតុ X ក្នុង B:

$$X=\sum_{i=1}^{N}a_i v_i$$

បច្ចេកទេសក្នុងការដោះស្រាយដែលយើងបានរំលងចោលនៅក្នុងចំណុចមុន វាអាចទាញបញ្ជាក់ថា $a_1=1$ ។ ឥឡូវយើងចាប់ផ្តើមគណនា PX

$$PX=\sum_{i=1}^{N}a_i P v_i=\sum_{i=1}^{N}a_i\lambda_i v_i$$

v_i គឺជាវ៉ិចទ័រ eigenvector នៃតម្លៃ eigenvalue λ_i ។ ហើយបើសិនជាយើងពិនិត្យឡើងវិញ យើងទទួលបាន

$$P^n X=\sum_{i=1}^{n}a_i\lambda_i^n v_i$$

ប្រសិនបើតម្លៃ λ_i សម្រាប់ i>1 ដែល $|\lambda|<1$ នោះ $\lim_{x\to\infty}P^n X=a_1v_1=\pi$

នេះបញ្ជាក់ថានឹងមានបញ្ហាអ្វីកើតឡើង ចំពោះឧទាហរណ៍របស់យើង!!

ទោះជាយ៉ាងណាក៏ដោយ ទ្រឹស្តីបទមិនបានធានាថា ម៉ាទ្រីស P នីមួយៗអាចបំពេញតម្រូវការតាមការប៉ានស្មានរបស់ទ្រឹស្តីបទឡើយ។ យើងនឹងរៀបរាប់អំពីបញ្ហាដែលអាចនឹងកើតមានឡើង និង វិធីទប់ស្កាត់ ។

បញ្ហាដែលអាចកើតមានឡើង:

- តម្លៃ eigenvalue 1 ប្រហែលជាស្មើនឹងផលគុណរឹសនៃកន្សោមពហុធារបស់ $\det(\lambda I-A)=0$
- ម៉ាទ្រីស P ប្រហែលជាអាចមានតម្លៃ eigenvalue λ ផ្សេងទៀត មានតម្លៃធំជាង 1 ជាមួយម៉ូឌុយឡូស្មើនឹង 1 ។

តើត្រូវធ្វើដូចម្តេចក្នុងករណីនេះ?

យើងប្រើអន្តរម៉ាទ្រីស Markov ដើម្បីបញ្ចៀសបញ្ហាទាំងឡាយដែលអាចកើតមាន ដែលគេអោយឈ្មោះម៉ាទ្រីស Markov តាមរយៈ

និយមន័យ: អន្តរម៉ាទ្រីស Markov នេះមានភាពទៀងទាត់ប្រសិនបើ

- តម្លៃ eigenvalue 1 គឺជារឹសធម្មតានៃផលគុណពហុធារបស់ $\det(\lambda I-A)=0$

- គ្រប់តម្លៃ eigenvalue λ ទាំងអស់នៃម៉ាទ្រីស P ធំជាង 1 ដែលមានម៉ូឌុយឡូតូចជាង 1។

កត់សម្គាល់ថា ម៉ាទ្រីស P ភាគច្រើនគឺមានភាពទៀងទាត់ណាស់ ។ ដូច្នេះ ប្រសិនបើមានម៉ាទ្រីសណាមិនប្រក្រតី យើងត្រូវរិះរកយុទ្ធសាស្រ្តណាមួយដើម្បីបង្កើតអោយមានម៉ាទ្រីសដែលមានភាពទៀងទាត់មួយយ៉ាងតិច។

វិធីទប់ស្កាត់: យើងពិចារណាទៅលើម៉ាទ្រីស $Q=(q_{ij})$ លំដាប់ $N\times N$ ដែល $q_{ij}=\dfrac{1}{N}$ ចំពោះគ្រប់ ចំនួនគត់ i, j យើងធ្វើការផ្លាស់ប្តូរម៉ាទ្រីស P នៃគេហទំព័រ ទៅជាម៉ាទ្រីសដូចខាងក្រោម:

$$(1) \qquad P_\beta=(1-\beta)P+\beta Q$$

សម្រាប់តម្លៃ β តូចមួយ ដែល $\beta\in[0,1]$ (តម្លៃ $\beta=0.15$ ធ្លាប់បានប្រើនៅក្នុង Google) បញ្ជាក់ថា ម៉ាទ្រីស P_β នៅតែមានធាតុធំជាងរឺស្មើ 0 ហើយផលបូកធាតុនៃជួរឈរនីមួយៗ មានតម្លៃស្មើ 1 ដែលវានៅតែជាអន្តរម៉ាទ្រីស Markov ។ តាមទ្រឹស្តីបទ បញ្ជាក់ថាមានតម្លៃតូចមួយ β ណាមួយដែលធ្វើអោយបញ្ហា ត្រូវកើតមានបិតនៅនឹងថ្កល់មួយកន្លែង ។

ទ្រឹស្តីបទ: គេអោយ អន្តរម៉ាទ្រីស Markov នីមួយៗមានតម្លៃវិជ្ជមាន β តូចមួយតាមដែលយើងចង់បាន ដូចនេះ ម៉ាទ្រីស P_β គឺមិនប្រែប្រួល ។ អោយ π ជាវ៉ិចទ័រ eigenvector ដែលមានតម្លៃ eigenvalue 1 សម្រាប់ម៉ាទ្រីស P_β ហើយធ្វើអោយផលបូកនៃកូអរដោនេមានតម្លៃស្មើ 1។ សម្រាប់ម៉ាទ្រីស P_β អោយវ៉ិចទ័រ មិនសូន្យ X ដែល $X^t=(p_1,...,p_N)$ ជាមួយនឹង $p_i\in[0,1]$ និង $\sum_{i=1}^N p_i=1$ ហើយ $\lim\limits_{n\to\infty}(P_\beta)^n X=\pi$ ។

ការភ្ជាប់ជាមួយនឹងទ្រឹស្តីចំនុចថេរ Banach:

ការយកអត្ថបទខ្លីៗសំខាន់ៗចំពោះទ្រឹស្តីបទចំណុចថេរ Banach ទ្រឹស្តីបទខាងលើ អាចមានលក្ខខណ្ឌពិសេសសម្រាប់ការអនុវត្តនេះ ។ យើងអាចរំលងផ្នែកនេះប្រសិនបើអ្នកមិនបានអានអត្ថបទខ្លីៗដែលមានសារៈសំខាន់ៗដ៏ទៃទៀត ។

ទ្រឹស្តីបទ: អោយម៉ាទ្រីស P ជាអន្តរម៉ាទ្រីស។ យើងគិតពិចារណា $S=\left\{X/X^t=(p_1,...,p_N),p_i\in[0,1],\sum_{i=1}^N p_i=1\right\}$ ជាមួយនឹងចម្ងាយគ្រប់គ្រាន់ d(X,Y) នៅចន្លោះចំណុចជាច្រើន (ចំណុចនេះអាស្រ័យនឹងម៉ាទ្រីស P) ។ នៅលើ S យើងគិតថាលីនេអ៊ែរនៃ $L:S\to S$ ចង្អុលបង្ហាញតាមរយៈ L(X)=PX ។ ការធ្វើដំណើរការរបស់ L បានបង្រួមទៅជា S ដែលមានតម្លៃ $c\in[0,1]$ហើយដែល គ្រប់ $X,Y\in S$, $d(L(X),L(Y))\le cd(X,Y)$ ។

មានវ៉ិចទ័រតែមួយគត់ $\pi\in S$ ដែល $L(\pi)=\pi$

ម៉្យាងវិញទៀត តម្លៃ $X_0\in S$ យើងអាចចង្អុលបង្ហាញស្វ៊ីត $\{X_n\}$ តាមរយៈការណែនាំ ដែល $X_{n+1}=L(X_n)$ បន្ទាប់មក $\lim\limits_{n\to\infty}X_n=\pi$ ។

និយមន័យចម្ងាយ d: និយមន័យចម្ងាយ d គឺជាចំណុចប្រសព្វ និង រំលងទៅកន្លែងផ្សេងទៀត ។ យើងអាចបង្រួមវាបានដោយគ្រប់គ្រាន់សម្រាប់ការអាន សម្រាប់អ្នកដែលត្រូវការ ។ យើងកំណត់ដោយខ្លួនឯង ទៅ

តាមករណីដែលម៉ាទ្រីស P គឺអាចមានលទ្ធភាពបញ្ចៀស ។ អោយ $B=\{v_1=\pi,v_2,...,v_N\}$ ដែលវាជា មូលដ្ឋាន នៃវ៉ិចទ័រ eigenvector ។ វ៉ិចទ័រ $X,Y\in S$ អាចសរសេរជា វ៉ិចទ័រ B

$$X=\sum_{i=1}^{N}a_iv_i, Y=\sum_{i=1}^{N}b_iv_i$$

ដែល $a_i,b_i\in\mathbb{C}$ បន្ទាប់មកយើងអាចទាញបានថា:

$$d(X,Y)=\sum_{i=1}^{n}|a_i-b_i|$$

ជាមួយនឹងចម្ងាយនេះ S គឺអាចបំពេញកន្លែងម៉ាទ្រីស ដូចជាគ្រប់ស៊ីត Cauchy រួម ។

ទ្រឹស្តីបទនេះ មិនអាចបំពេញគ្រប់គ្រាន់នៃ វ៉ិចទ័រ π ប៉ុន្តែអោយវិធីមួយដើម្បីកសាងវាឡើងវិញ ដូចជាការកំណត់ស៊ីត $\{X_n\}$។ យើងធ្លាប់ឃើញការបកស្រាយនៃការបង្រួមនៅក្នុងឧទាហរណ៍របស់យើង ។ តាមពិតជួរឈរទី j នៃម៉ាទ្រីស P^n គឺជាវ៉ិចទ័រ P^ne_j ដែល $e_j^t=(0,...,0,1,0,...,0)$ គឺជាវ៉ិចទ័រទី j នៃ ការបែងចែកជាមូលដ្ឋានគ្រឹះកាណូនិច ។

នៅក្នុងឧទាហរណ៍របស់យើង យើងអាចរកឃើញវ៉ិចទ័រ π តាមរយៈការដោះស្រាយប្រព័ន្ធ (I-P)X=0 ជាមួយម៉ាទ្រីស

$$I-P=\begin{pmatrix}1 & -\frac{1}{2} & -\frac{1}{3} & -1 & 0\\ -1 & -1 & -\frac{1}{3} & 0 & -\frac{1}{3}\\ 0 & -\frac{1}{2} & 1 & 0 & -\frac{1}{3}\\ 0 & 0 & 0 & 1 & -\frac{1}{3}\\ 0 & 0 & -\frac{1}{3} & 0 & 1\end{pmatrix}$$

យើងស្វែងរកថា គ្រប់ដំណោះស្រាយគឺជាទម្រង់ $(4S,\frac{16}{3}S,3S,\frac{1}{3}S,S)^t$ សម្រាប់ $S\in\mathbb{R}$។ លទ្ធិផលនៃផលបូកកូអរដោនេ ស្មើ 1 ដូចនេះ π គឺ $\pi^t=(\frac{12}{41},\frac{16}{41},\frac{9}{41},\frac{1}{41},\frac{3}{41})$

ការបែងចែកការគណនាដើម្បីអោយមានភាពងាយស្រួលៈ

យើងមានគំនិតត្រឹមត្រូវតិចតួចសម្រាប់ algorithm ។ ទោះជាយ៉ាងណាក៏ដោយ ការស្វែងរកវ៉ិចទ័រ π ដូចជា វ៉ិចទ័រ eigenvector នៃតម្លៃ eigenvalue 1 សម្រាប់ម៉ាទ្រីស P_β ក្នុង (1) គឺវាមិនមែនជាការងារតូចតាចនោះទេ ពេលដែលម៉ាទ្រីសមានរាប់រយពាន់លានជួរឈរទាំងពីរ ទាំងការគណនាតាមម៉ាស៊ីន និង ការបង្ហាញស្មារតីបកស្រាយពីបញ្ហាពិតទាំងនេះ ។ ជាធម្មតា វិធីបំបាត់ Gauss គឺវាត្រូវការដើម្បីប្រែក្លាយ មេគុណអោយនៅតូច ។ ការដំណើរការ algorithm បង្កើតអោយមានការប្រើប្រាស់ លក្ខណៈ(២) (មើល[LM])។ ទាំងនេះអាចបង្កើតការតភ្ជាប់ជាមួយអត្ថបទទ្រឹស្តីបទខ្លីៗរបស់ទ្រឹស្តីបទចំណុចថេរ Banach ដែលនឹងពន្យល់ បានថា ទ្រឹស្តីបទចំណុចថេរ Banach បានផ្តល់អោយ algorithm អាចសាងសង់ចំណុចថេរមួយ ។

ជាការពិតណាស់ យើងអាចចាប់ផ្តើមជាមួយ X_0 ដូចខាងក្រោម:

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, ..., \frac{1}{N})$$

ហើយយើងត្រូវការការគណនា $\lim_{n\to\infty}(P_\beta)^n X_0$ ជាធម្មតា $P_\beta^n X_0$ ដែល n នៅចន្លោះ ៥០ ទៅ ១០០ អោយតម្លៃប្រហែលដ៏ល្អមួយនៃ π។ ការណែនាំអោយយើងគណនា $X_{n+1} = P_\beta X_n$ ការគណនារបៀបនេះពិតជាវែងឆ្ងាយខ្លាំងណាស់ ។ តាមពិត ដោយសារតែការបង្កើតម៉ាទ្រីស P_β ក្នុង(១) មិនមានធាតុ 0 ទេ។ នៅលើធាតុដ៏ទៃទៀតភាគច្រើននៃធាតុទាំងនោះគឺ 0។ ដូច្នេះ យើងត្រូវបំបែកការគណនាលេខដោយយកតែចំនុចសំខាន់ៗតែប៉ុណ្ណោះដែលៈ

$$X_{n+1} = (1-\beta)PX_n + \beta QX_n$$

ពីព្រោះទម្រង់ពិសេសនៃ Q គឺវាងាយស្រាយបំភ្លឺថា បើសិន X គឺជាកត្តានៃផលបូកធាតុនីមួយៗស្មើនឹង 1 ពេលដែល $QX = X_0$។ ដូចនេះ វាគ្រប់គ្រាន់សម្រាប់ការគណនាស្ទីតនេះ

$$X_{n+1} = (1-\beta)PX_n + \beta X_0$$

សន្និដ្ឋានៈ

យើងបានបង្ហាញជាសាធារណៈ នៃចំណែកមួយរបស់ Google's PageRank algorithm។ អ្នកអាចមានបទពិសោធន៍គ្រប់គ្រាន់ ជាមួយនឹងគេហទំព័រធម្មតា និងការស្វែងរកល្បិចដើម្បីធ្វើអោយប្រសើរឡើងក្នុងការតម្រៀបទិន្នន័យនៃទំព័ររបស់អ្នក តាមរយៈការបន្ថែមការទំនាក់ទំនងទាំងខាងក្នុង និងខាងក្រៅ។ ការងារខ្លះ មានភាពឥតខ្ចោះ ដែលវារួមចំណែកក្នុងការអភិវឌ្ឍន៍ជាខ្លាំង ។ ពួកវាខ្លះកើតឡើង ដែលបានផ្លាស់ប្តូរ " neutral " នៃម៉ាទ្រីស Q ក្នុង (១) តាមរយៈម៉ាទ្រីសជាច្រើនដែលមាននៅលើគេហទំព័រ ។ ការធានាផ្សេងទៀតបញ្ជាក់ថា ការតម្រៀបតម្លៃនេះ គឺមិនអាចអោយដឹងគ្រប់គ្រាន់ដែលជាហេតុនាំទៅរកនរណាម្នាក់អាចមានលទ្ធភាពធ្វើអោយប្រសើរឡើងចំពោះការរៀបចំគេហទំព័ររបស់ពួកគេ ។

សេចក្តីសន្និដ្ឋានទូទៅៈ តើយើងបានធ្វើការអង្កេតលើអ្វីខ្លះ? ធម្មតាគំនិតវ៉ៃឆ្លាតវានាំទៅរកការរកឃើញរបកគំហើញដ៏ធំមួយដែលគ្រប់គ្រាន់ក្នុងការស្វែងរកដោយម៉ាស៊ីន ហើយវាជាកំណើតនៃប្រភពពាណិជ្ជកម្ម។ ប្រសិនបើការគណនាដោយខ្លួនវាមានភាពមិនត្រឹមត្រូវ ដូច្នេះគំនិតដំបូងត្រូវការ "ធាតុ" នៃគណិតវិទ្យា ដែល អោយឈ្មោះថា ពីជគណិតលីនេអ៊ែរ និងទ្រឹស្តីបទប្រូបាប ។ ទំនាក់ទំនងជាស្តង់ដានៃទ្រឹស្តីគណិតវិទ្យាក្នុងអង្កត់ទ្រូងពិសេសនៃម៉ាទ្រីស ដែលវាមានប្រសិទ្ធភាពនៅពេលដែលគេយកវាទៅប្រើនៅខាងក្រៅនៃបរិបទធម្មតារបស់ពួកគេមួយ។ យើងមានផ្នែកសំខាន់ផងដែរ ដែលរួមផ្សំគំនិតនៅក្នុងវិទ្យាសាស្រ្ត ជាមួយនឹងទ្រឹស្តីបទចំណុចថេរ Banach ដែលបានអនុវត្តន៍ ដូច្នេះវាដាច់ស្រឡះពីទម្រង់ដើមរបស់វា ។

ឯកសារយោង:

[E] M. Eisermann, Comment Google classe les pages webb, http://images.math.cnrs.fr/Comment-Google-classe-les-pages.html, 2009.

[LM] A. N. Langville and C. D. Meyer, A Survey of Eigenvector Nethods for Web Information Retrieval, SIAM Review, Volume 47, Issue 1, (2005), pp. 135-161.

[RS] C. Rousseau and Y. Saint-Aubin, Mathematics and technology, SUMAT Series, Springer-Verlag, 2008 (A French version of the book exists, published in the same series).