Controle e Simulação de Sistemas Sujeitos a Saltos Markovianos

Oscar Neiva E. Neto

***orientador:* Marcos G. Todorov, D. Sc.**

Laboratório Nacional de Computação Científica - LNCC

Faculdade de Educação Tecnológica do Estado do Rio de Janeiro - FAETERJ

oscarnen@lncc.br   todorov@lncc.br

1 Introdução

Este trabalho é dedicado ao estudo de sistemas sujeitos a saltos Markovianos e questões relacionadas a cadeias de Markov. A pesquisa desenvolvida visa o aprendizado de aspectos teóricos fundamentais relacionados a cadeias de Markov, teoria de controle, e simulação de problemas distribuídos.

2 Sistemas Lineares com Saltos Markovianos

Considere $\theta = \{\theta(k), k = 0, 1, \ldots\}$ uma cadeia de Markov homogênea no espaço de estados discreto $\mathbf{N} = \{1, 2, ..., M\}$, isto é, tal que:

$$P\big(\theta(k+1) = j \mid \theta(k) = i, \theta(k-1), \ldots, \theta(0)\big) = P\big(\theta(k+1) = j \mid \theta(k) = i\big), \tag{1}$$

que é denominada a propriedade de Markov. Tal processo estocástico tem aplicação em diversas áreas da ciência como, por exemplo, computação, engenharia, economia e biologia [1].

O interesse neste trabalho é no estudo de sistemas dinâmicos governados pela seguinte equação:

$$x(k+1) = A_{\theta(k)}x(k), \qquad k = 0, 1, 2, \ldots, \tag{2}$$

onde x é o estado do sistema. Existe uma vasta literatura dedicada ao estudo desses sistemas, que sofrem variações abruptas em sua estrutura. Uma amostra significativa de avanços recentes pode ser encontrada no livro [3].

3 PageRank

O *PageRank* foi criado em 1998 para atribuir valores para páginas de um conjunto de documentos interligados [2]. A ideia por trás do algoritmo é de atribuir pesos às páginas da *world wide web*, em que recebem os maiores pesos as mais visitadas.

3.1 Definição

A análise do PageRank tipicamente modela a web como um grafo orientado contendo N nós, que representam as páginas, e uma coleção de arestas $\mathcal{E}$ que representam os links correspondentes. A navegação é então representada através de uma cadeia de Markov com espaço de estados discreto de dimensão N. A hipótese usual é de que todas as arestas que saem de um dado nó têm o mesmo peso, e portanto a matriz de transição da cadeia de Markov é a seguinte:

$$p_{ij} = \begin{cases} \dfrac{1}{N_i}, & \text{caso } (i,j) \in \mathcal{E}, \\ 0, & \text{caso contrário,} \end{cases} \tag{3}$$

onde N_i é o número de *links* que saem da página i.

O PageRank de um conjunto de N páginas da web é definido como o vetor $x \in \mathbb{R}^{N\times 1}$ que satisfaz as seguintes equações:

$$x = Ax, \quad x \geq 0, \quad \sum_{j=1}^{N} x_j = 1, \tag{4}$$

onde $A \in \mathbb{R}^{N\times N}$ é a transposta da matriz de transição da cadeia de Markov.

Um método usual para obtenção do PageRank é o chamado *Power Method* [4], que busca aproximar o PageRank através da seguinte iteração: $x(k+1) = Ax(k)$, $k \geq 0$, com $x(0) = x_0$, onde $x_0 \in \mathbb{R}^{N\times 1}$ é uma condição inicial positiva de soma igual a um.

3.2 Teleportation Model

Muito embora a simplicidade do *Power Method* o torne atraente, em geral ele não fornece uma garantia de convergência. O *Teleportation Model* é uma estratégia reconhecida para que, através de uma pequena modificação na matriz A, o método convirja globalmente para o PageRank. Matematicamente o *teleportation model* é representado como uma combinação convexa de duas matrizes. Em que, m é um parâmetro, tal que $m \in (0,1)$, e a matriz link modificada é dada por $M \in \mathbb{R}^{N\times N}$, definida por, $M = (1-m)A + \frac{m}{n}11^T$, onde $11^T \in \mathbb{R}^{N\times N}$ é a matriz cujas entradas são todas iguais a um.

3.3 Algoritmos Distribuídos

A fim de tornar o cálculo do PageRank menos custoso, e melhor explorar os recursos computacionais dos servidores disponíveis na web, uma alternativa é o emprego de algoritmos distribuídos. A agregação de tais experimentos aleatórios faz então com que o *power method* se torne um sistema linear com saltos Markovianos do seguinte tipo:

$$x(k+1) = A_{\theta(k)}x(k), \quad k \geq 0, \quad \text{com} \quad x(0) = x_0. \tag{5}$$

Para as matrizes links distribuídas $A_i \in \mathbb{R}^{N\times N}$, $i = 1, 2, ..., n$, tem-se que elas são definidas da seguinte forma: a i-ésima coluna de A_i coincide com a i-ésima coluna de A, a j-ésima entrada diagonal de A_i é igual a um, para $j = 1, ..., n$, $j \neq i$, todas as outras entradas a_{ij} são iguais a zero.

De forma similar ao que ocorre no *power method*, este método pode apresentar problemas de convergência, que são resolvidos pela seguinte modificação no *teleportation model*,

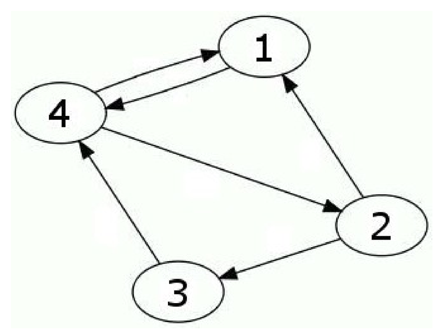
$$x(k+1) = (1-\hat{m})A_{\theta(k)}x(k) + \frac{\hat{m}}{N}11^T, \quad k \geq 0, \quad \text{com} \quad x(0) = x_0, \tag{6}$$

onde $\hat{m}$ é definido em através da seguinte fórmula: $\hat{m} = \frac{2m}{n-m(n-2)}$.

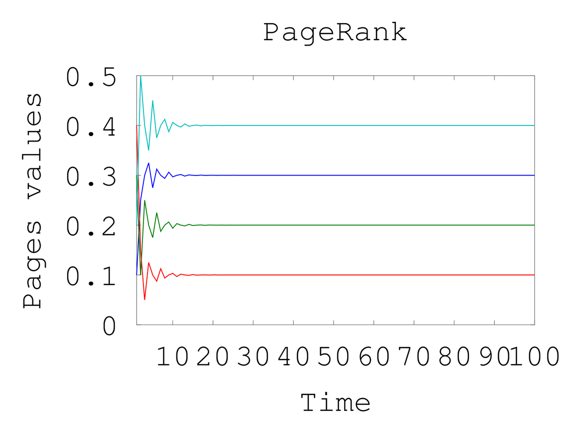
Esse algoritmo converge, conforme mostrado em [4], no sentido da média quadrática dada por: $\lim_{x\to\infty} \mathbb{E}[\| \ y(k) - x \ \|^2] = 0$, onde $y(k)$ é a média do conjunto de amostras $x(0), ..., x(k)$ definida como: $y(k) = \frac{1}{k+1}\sum_{l=0}^{k} x(l)$. Essa média é conhecida como média de Polyak ou média de Cèsaro e sua forma recursiva é dada por: $y(k+1) = \frac{(k+1)}{(k+2)}y(k) + \frac{1}{(k+2)}x(k+1)$.

4 Simulações

As simulações foram feitas considerando-se um conjunto de quatro páginas, ou seja, uma cadeia de Markov com quatro estados. Foram feitas simulações dos três modelos descritos anteriormente, e uma simulação usando-se algoritmo de Monte Carlo.

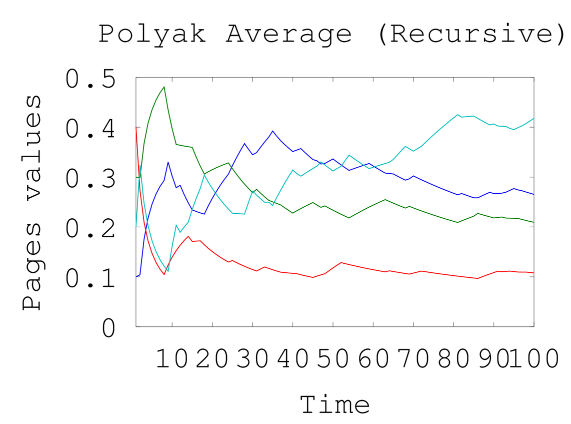


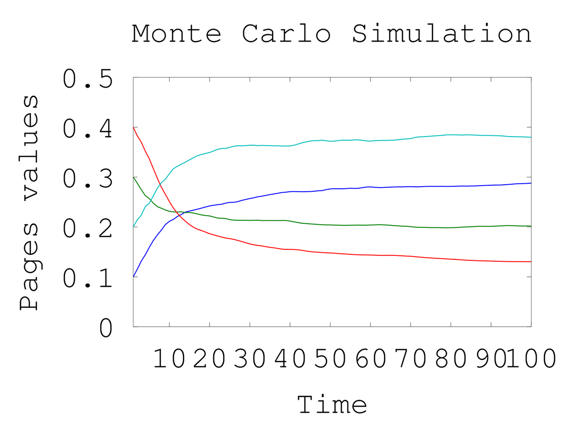
Na simulação do modelo do PageRank em sua forma mais simples, observa-se uma rápida convergência para cada um dos valores do vetor de estados, isso ocorre devido a dimensão do problema usado na simulação. Nesta simulação as linhas do gráfico estão associadas às páginas da seguinte forma: página 1 ao azul escuro, página 2 ao verde, página 3 ao vermelho e página 4 ao azul claro. Assim pode-se observar que a página de número quatro foi a que recebeu um maior peso neste conjunto de links. Ao mesmo tempo que é a página que está associada a um maior número de links, tanto de entrada quanto de saída.



Na simulação do *Teleportation Model*, obteve-se uma grande semelhança com os resultados da simulação anterior, isso ocorre pois o exemplo simulado é bem comportado.

No modelo distribuído observa-se uma oscilação dos estados sem que ao longo do tempo estacionem em algum valor, mas feita a média dos valores da cadeia no modelo distribuído, chega-se a valores finais próximos aos anteriormente encontrados para cada um dos estados. Ao final fez-se uma simulação usando método de Monte Carlo, de forma que é feita uma média dos estados obtidos na simulação do modelo distribuído.





5 Perspectivas Futuras

Neste projeto pretende-se ainda trabalhar com outros métodos de controle e simulação de sistemas estocásticos, que também sejam válidos na simulação do PageRank. Simulações essas como as de modelos agregados de cadeias de Markov e dentro do contexto do PageRank abordagens também de problemas de consenso.

Bibliografia

[1] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, volume 31 of *Texts in Applied Mathematics*. Springer, New York, 1999.

[2] K. Bryan and T. Leise. The $25,000,000,000 eigenvector: The linear algebra behind Google. *SIAM Rev.*, 48(3):569–581, 2006.

[3] O. L. V. Costa, M. D. Fragoso, and M. G. Todorov. *Continuous-Time Markov Jump Linear Systems*. Probability and Its Applications. Springer-Verlag, Heidelberg, 2013.

[4] H. Ishii and R. Tempo. The PageRank problem, multiagent consensus, and web aggregation: a systems and control viewpoint. *IEEE Control Syst. Mag.*, 34(3):34–53, 2014.

Processo CNPq: 153399/2015-5