Applying theory of Markov chains to the problem of ranking

A. Govan   C. Meyer

Department of Mathematics
North Carolina State University

First ACES workshop, December 2006
Outline

Overview of the Markov Chains

Ranking with Markov Chains - Google’s PageRank

Ranking with Markov Chains - extending to Football

Summary
Markov Chains Basics

States-finite:

A. Govan, C. Meyer

Applying theory of Markov chains to the problem of ranking
Overview of the Markov Chains
Ranking with Markov Chains - Google’s PageRank
Ranking with Markov Chains - extending to Football
Summary

Markov Chains Basics

Transitioning between states:

A. Govan, C. Meyer
Applying theory of Markov chains to the problem of ranking
Markov Chains Basics

Transition probabilities:
Restrictions on the transition probabilities:

- Memoryless (Markov property)
  - \[ p_{ij} = P(X_{t+1} = S_j | X_t = S_i, X_{t-1} = S_{i_{t-1}}, \ldots, X_0 = S_{j_0}) = P(X_{t+1} = S_j | X_t = S_i) \]
Restrictions on the transition probabilities:

- Memoryless (Markov property)
  
  \[ p_{ij} = P(X_{t+1} = S_j | X_t = S_i, X_{t-1} = S_{i_{t-1}}, \ldots, X_0 = S_{j_0}) = P(X_{t+1} = S_j | X_t = S_i) \]

- Homogeneous
  
  \[ p_{ij} \text{ has no time dependence} \]
Overview of the Markov Chains

Ranking with Markov Chains - Google’s PageRank
Ranking with Markov Chains - extending to Football

Summary

Markov Chains Basics-Matrices

Transition probability matrix-stochastic matrix

\[
\begin{pmatrix}
S_1 & S_2 & \ldots & S_i & \ldots & S_n \\
S_j & \vdots & \ddots & \vdots & \ddots & \vdots \\
S_i & p_{i1} & p_{i2} & \ldots & p_{ii} & \ldots & p_{in} \\
S_i & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
S_n & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

A. Govan, C. Meyer

Applying theory of Markov chains to the problem of ranking
Markov Chains-Stochastic Matrices

- Markov Chain:
  - \( \{p(0), p(1), p(2), \ldots\} \)
  - such that \( p(i) \) is a probability distribution vector and \( p^T(i) = p^T(0)P^i \)

- Irreducible
- Primitive
  - \( \lambda = 1 \) is the only one on the spectral circle
  - can use power method to compute stationary distribution vector (eigenvector corresponding to \( \lambda = 1 \))
Markov Chains - Stochastic Matrices

Markov Chain:

\[ \{ p(0), p(1), p(2), \ldots \} \]

such that \( p(i) \) is a probability distribution vector and

\[ p^T(i) = p^T(0)P^i \]

some Stochastic Matrices are:

- Irreducible
- Primitive
- \( \lambda = 1 \) is the only one on the spectral circle
- can use power method to compute stationary distribution vector (eigenvector corresponding to \( \lambda = 1 \))

A. Govan, C. Meyer

Applying theory of Markov chains to the problem of ranking
Markov Chains-Stochastic Matrices

- Markov Chain:
  - \( \{ p(0), p(1), p(2), \ldots \} \)
  - such that \( p(i) \) is a probability distribution vector and \( p^T(i) = p^T(0)P^i \)

- some Stochastic Matrices are:
  - Irreducible
    - from any state to any state
Markov Chains-Stochastic Matrices

- Markov Chain:
  - \( \{ p(0), p(1), p(2), \ldots \} \)
  - such that \( p(i) \) is a probability distribution vector and 
    \( p^T(i) = p^T(0)P^i \)

- some Stochastic Matrices are:
  - Irreducible
    - from any state to any state
  - Primitive
    - \( \lambda = 1 \) is the only one on the spectral circle
    - can use power method to compute stationary distribution vector (eigenvector corresponding to \( \lambda = 1 \))
Making Google Matrix

▶ Webpages are states
Making Google Matrix

- Webpages are states
- Hyperlink Matrix $H$
  - $H(i, j) = \begin{cases} 
  \frac{1}{|i|} & \text{there is a link from } i \text{ to } j \\
  0 & \text{otherwise}
\end{cases}$
Making Google Matrix

- Webpages are states
- Hyperlink Matrix $H$
  - $H(i, j) = \begin{cases} 1/|i| & \text{there is a link from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$
- Stochastic matrix $S$
  - Replace the zero rows of $H$ with $(1/n)e^T$, where $e$ is a column vector of ones.
Making Google Matrix

- Webpages are states
- Hyperlink Matrix $H$
  - $H(i, j) = \begin{cases} 1/|i| & \text{there is a link from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$
- Stochastic matrix $S$
  - Replace the zero rows of $H$ with $(1/n)e^T$, where $e$ is a column vector of ones.
- Google Matrix $G$.
  - Convex combination: $G = \alpha S + (1 - \alpha)ev^T$, $\alpha \in (0, 1)$ and $v^T > 0$
  - Personalization vector $v$. 

A. Govan, C. Meyer
Applying theory of Markov chains to the problem of ranking
PageRank vector $\pi$. 

$$G = \alpha S + (1 - \alpha)ev^T$$

- $G$ is the transition probability matrix.
PageRank vector $\pi$.

$$G = \alpha S + (1 - \alpha)ev^T$$

- $G$ is the transition probability matrix.
- $G$ is irreducible (and primitive).

$$\pi^T = \pi^T G$$

A. Govan, C. Meyer
PageRank vector $\pi$.

$$G = \alpha S + (1 - \alpha)ev^T$$

- $G$ is the transition probability matrix.
- $G$ is irreducible (and primitive).

$$\pi^T = \pi^T G$$

- $\pi$ is the stationary probability distribution vector.
PageRank vector $\pi$.

$$G = \alpha S + (1 - \alpha)ev^T$$

- $G$ is the transition probability matrix.
- $G$ is irreducible (and primitive).

$$\pi^T = \pi^T G$$

- $\pi$ is the stationary probability distribution vector.
- $\pi$ is unique (up to a scalar multiple).
NFL set up.

- Each NFL team is a state.
NFL set up.

- Each NFL team is a state.
- Score differences determine transition probability
NFL Matrix

\[
\begin{pmatrix}
0 & 0 & \frac{10}{33} & \frac{20}{33} & \frac{3}{33} \\
0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
\frac{10}{13} & 0 & \frac{3}{13} & 0 & 0 \\
\frac{3}{17} & 0 & 0 & \frac{14}{17} & 0
\end{pmatrix}
\]
NFL Matrix-Stochastic

Dealing with an undefeated team:

<table>
<thead>
<tr>
<th></th>
<th>Car</th>
<th>Pit</th>
<th>Chi</th>
<th>TB</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>0</td>
<td>0</td>
<td>10/33</td>
<td>20/33</td>
<td>3/33</td>
</tr>
<tr>
<td>Pit</td>
<td>1/5</td>
<td>1/5</td>
<td>1/5</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>Chi</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TB</td>
<td>10/13</td>
<td>0</td>
<td>3/13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NO</td>
<td>3/17</td>
<td>0</td>
<td>0</td>
<td>14/17</td>
<td>0</td>
</tr>
</tbody>
</table>

A. Govan, C. Meyer

Applying theory of Markov chains to the problem of ranking
Adding stats:

\[ F = \alpha S + (1 - \alpha)es^T \]

where \( s \) is based on teams statistical data.
NFL Matrix-Irreducible and Primitive

- Adding stats:
  \[ F = \alpha S + (1 - \alpha)es^T \]
  where \( s \) is based on teams statistical data.

- Adding more stats?
  \[ F = \alpha_0 S + \alpha_1 es_1^T + \ldots + \alpha_k es_k^T \]
  where \( s_i \) is statistics based and \( \sum \alpha_i = 1 \).
Current and Future Work:

- Determining the “important” statistics vectors $\mathbf{s}_1^T, \ldots, \mathbf{s}_k^T$
- Automate the selection of the best $\alpha_i$ for a specified $\mathbf{s}_i^T$. 