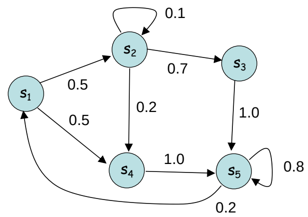
Markov Chains and PageRank

Scott Sanner
University of Toronto

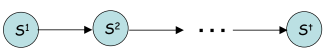
Markov Models (or Markov Chains)

- At each time step, probabilistically transition from current state to next state ($S = \{s_1, s_2, \ldots, s_n\}$)
- Finite State Machine (FSM) view for n=5:



Markov Models

- The graphical model view for t steps:

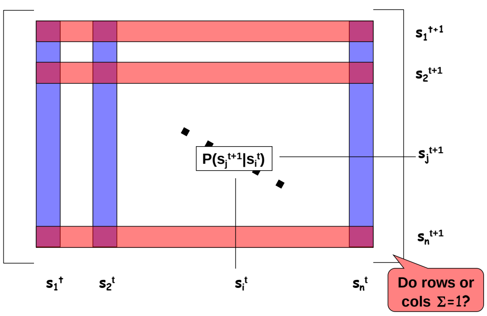


  – Note: for $t = \infty$, an infinite graphical model!
- Or assuming transition stationarity, just:

$P(S^{t+1}|S^t) = P(S^t|S^{t-1})$

Transition Matrix

- Represent $P(s^{t+1}|s^t)$ as transition matrix:

T = 

Do rows or cols Σ=1?

Transition Probabilities

- Formally
  – Define state set $S^t = \{s_1, s_2, \ldots, s_n\}$; $\forall t$
  – Define transition matrix $T_{ij}^{\,t} = P(S_i^{t+1}|S_j^{\,t})$; $\forall t$
- Properties of T_{ij}
  – *Stationary:* $T_{ij}^{\,t} = T_{ij}^{\,t-1}$ OR $P(S^{t+1}|S^t) = P(S^t|S^{t-1})$; $\forall t$
  – *Irreducible:* Possible to get from any s_i to s_j
  – *Aperiodic:* Time to return has periodicity = 1
  – *Transient:* Positive probability of not returning to state
  – *Recurrent:* Not transient
  – *Ergodic:* Aperiodic and (positive) recurrent

Examples of each?

Distribution at Time t

- Given $P(s^0)$, what is $P(s^t)$?
- Let Ps^0 & Ps^t be column vectors…
  – Then simply: $Ps^t = (T\char`^t)\, Ps^0$

Stationary Distribution

- Stationary Distribution π at t=∞
  - $\pi = (T^{\wedge\infty})\, Ps^0$
  - If T ergodic & irreducible, Ps^0 irrelevant
    - Reaches *unique* steady-state distribution: $\pi = T\pi$
    - So π=any row of T^∞
    - Can solve via eigenvector analysis (note: λ=1)
      - Related to (Krylov) iterated eigenvector computation
    - Or use fixed point to solve linear system
      - $T\pi - \pi = 0 \rightarrow \pi' T' - \pi' = 0 \rightarrow \pi' (T' - I) = 0$ s.t. constraints on π
      - Can solve linear system via matrix inversion

Why? What are they?

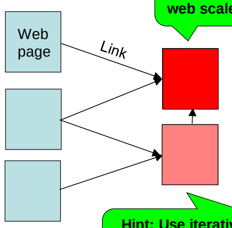
Markov Model Applications

- Simple theory, ingenious applications:
  - n^{th}-order Markov models
    - Relax Markovian assumption to previous n states
    - Used in text and speech processing
      - N-grams for predicting next word occurrence http://nbviewer.ipython.org/gist/yoavg/d76121dfde2618422139
      - Colocation identification
      - Dasher for text input, try it in your web browser
  - More generally
    - Physics (states of systems)
    - Queuing theory (random entries and exits)
    - Economics, Biology, Chemistry, etc…
    - Google!

Google PageRank Example

- Very beautiful use of Markov Models
- Model of web browsing:
  - Probabilistically take link with ~1/k chance if k links
  - Small chance of random transition
- Stationary distribution π gives PageRank!
  - Measure of "authority"

How to compute on web scale?



Hint: Use iterative method; how would you compute $T^{\wedge 256}$?