# Visualization of Navigation Patterns on a Web Site using Model Based Clustering

by

I. Cadez, D. Heckerman, C. Meek,
P. Smyth, S. White

Proceedings of KDD-2000

Chris Williams, School of Informatics
University of Edinburgh

# Overview

- **Aim**: Cluster sequences of user navigation patterns, so as to understand users of website—exploratory data analysis
- The data
- The output
- The model—mixtures of Markov models
- Fitting the model
- Application to msnbc.com
- Summary

# The data

- Server log files have been converted into a set of sequences, one sequence for each user session
- Each sequence is an ordered list of discrete symbols
- Each symbol represents one of several possible categories of web pages requested by the user
- Example sequences

| | | | | | |
|---|---|---|---|---|---|
| frontpage | news | travel | travel | | |
| news | news | news | news | news | |
| weather | | | | | |
| news | health | health | business | business | business |

# The output

- WebCANVAS tool
- Overview screen giving all sequences in each cluster (scrollable)
- “Drill down” into a cluster by obtaining
  - marginal distribution for each cluster
  - distribution over first event
  - transition probabilities $p(i, j)$

# The model

- Mixture of Markov models

$$p(\mathbf{x}|\theta) = \sum_{i=1}^{K} \pi_k p(\mathbf{x}|\theta_k)$$

$$p(\mathbf{x}|\theta_k) = p(x_i|\theta_k^I) \prod_{i=2}^{L} p(x_i|x_{i-1}, \theta_k^T)$$

- $\theta_k^I$ is probability of the initial symbol in the sequence (multinomial)
- $\theta_k^T$ is the transition probability from $x_{i-1}$ to $x_i$; each row is a multinomial
- Can also use a zeroth-order Markov model (unigram model) $p(\mathbf{x}|\theta_k) = \prod_{i=1}^{L} p(x_i|\theta_k^U)$

# Fitting the model

- Use EM (penalized maximum likelihood)
- Initialize $\pi$'s all equal
- Initialize $\theta$'s by fitting a single Markov model, then perturbing these parameters in each component
- Do 20 restarts for each $K$, choose model with highest posterior probability
- Choose $K$ using log likelihood of hold-out data

# A small problem, and a solution

- Two or more clusters can be encoded by a single Markov model
- Example: start at $a$ then choose between $b$ and $c$, or start at $d$ then choose between $e$ and $f$
- This problem occurred frequently
- Solved by allowing only one non-zero probability start state

# Application to msnbc.com

- 100,023 training sequences, 98,687 validation seq
- Found that EM scaled linearly with $N$ (number of sequences) and $K$
- Best first-order model has 40 components
- Chose constrained model with 100 components (of course constrained model needs more components)

## Summary

- Mixture of first-order Markov models
- WebCANVAS tool to visualize the clustered data and models
- Found that this clustering has revealed numerous interesting insights