

# Predictive Discrete Latent Factor Models for Large Scale Dyadic Data

Deepak Agarwal, Srujana Merugu  
Yahoo! Research

Matt Rodriguez

UCSD

March 5, 2009

# Outline

1 Motivation

2 Background

3 Algorithm

4 Analysis

# Recommender System Problem

$$\begin{array}{c} \cdot \quad i_2 \quad i_3 \dots i_m \\ U_2 \\ U_3 \\ \dots \\ U_n \end{array} \left| \right.$$

## Properties

- Rows represent users
- Columns represent items
- Matrix is sparse, most elements are unknown
- How to estimate unknown elements

## Problems

- Sparsity
- Signal to Noise Ratio is very low
- Mean centric approaches do not work very well
- Overfitting

## PDLF Algorithm

Captures global and local structure by simultaneously using a supervised and unsupervised approach.

# Capturing Global Structure

## Supervised Approach Example

- Multinomial logistic regression
- Select  $\beta$  features for each class
- Fails to capture local interaction between dyads

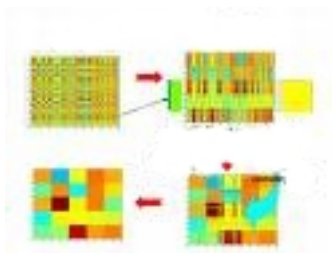
## PDLF

The PDLF algorithm captures global structure through the weights associated with the  $\beta$  features.

# Capturing Local Structure

## Unsupervised approach

- Hard Assignment Co-Clustering
- Cluster  $n$  rows into  $k$  row clusters
- Cluster  $m$  rows into  $l$  column clusters
- Minimize the Mutual Information Loss



# Exponential Family Distribution

## Model

The algorithm trains a model from an exponential family for each dyad

$$f(x; \theta) = \exp(\theta t(x) - \psi(\theta)) p_0(x)$$

## Terms

- $\psi$ , cumulant generating function
- $\theta$ , natural parameter
- $t(x)$ , sufficient statistic
- $p_0(x)$ , normalizing term

# General Linear Models

## GLM

- Maps the  $\beta$  features to the model parameter in the exponential family.
- The response variable  $Y$  is linear combination of input random variables  $\{X_1, X_2, \dots, X_n\}$
- The link function  $g(\theta) = \beta^t x$
- The response function  $f(\beta^t x) = \theta$

## Relation to PDLF

The GLM's response function maps  $\beta^t x + \delta$  to the model parameter of the exponential family distribution.



# Hard Assignment PDLF

- $\beta$ : Features capture global structure
- $\Delta$ : Interaction Effects capture local structure
- $\psi$ : cumulant generating function of the exponential family
- $\rho$ : row cluster assignments for each user
- $\gamma$ : column cluster assignments for each item/movie
- $x_{ij}$ : feature values of a given dyad
- $y_{ij}$ : response variable of the dyad (rating)
- $k$ : number of row clusters
- $l$ : number of column clusters

# Hard Assignment PDLF

Generalized M-Step

Step 1. Update Interaction Effects

$$\delta_{I,J} \leftarrow \arg \max_{\delta} \sum_{i \in I, j \in J} y_{ij} \delta - \psi(\beta^t x_{ij} + \delta)$$

Step 2. Update Regression Coefficients

$$\beta \leftarrow \arg \max_{\beta} \sum_{ij} y_{ij} \beta^t x_{ij} - \psi(\beta^t x_{ij} + \delta)$$

Generalized E-Step

Step 3. Update Row Cluster Assignments

$$\rho(i) \leftarrow \arg \max_I \sum_j (y_{ij} \delta_{I\gamma(j)} - \psi(\beta^t x_{ij} + \delta_{I\gamma(j)}))$$

Step 4. Update Column Cluster Assignments

$$\gamma(j) \leftarrow \arg \max_J \sum_I (y_{ij} \delta_{\rho(i)J} - \psi(\beta^t x_{ij} + \delta_{\rho(i)J}))$$

Repeat until convergence

# Algorithm Analysis

## Algorithm Analysis

- Algorithm returns  $\beta, \Delta, \rho, \gamma$
- Prediction for a dyad:  $f_{\psi}(:, B^t x_{ij} + \delta_{\rho(i)} \gamma(j))$
- Algorithm runs in  $O(Nkl)$

# Questions?

## References

- Predictive Discrete Latent Factor Models for Large Scale Dyadic Data ,Deepak Agarwal, Srujana Merugu, KDD 2007
- Clustering with Bregman Divergences, Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, Joydeep Ghosh, JMLR 2005
- Information-Theoretic co-clustering, I. Dhillon, S. Mallela, D. Hodha KDD 2003
- [http://videlectures.net/kdd07\\_agarwal\\_pdfm/](http://videlectures.net/kdd07_agarwal_pdfm/)