Predictive Modeling in Online Learning Environments

Forough Arabshahi* Furong Huang* Animashree Anandkumar* Carter T. Butts*

*University of California, Irvine

Massive Open Online Courses (MOOCs)

MOOCs
- Education available to everyone
- Ease of access
- Efficient learning

Challenges
- Large number of students
- Hard to track individual performance

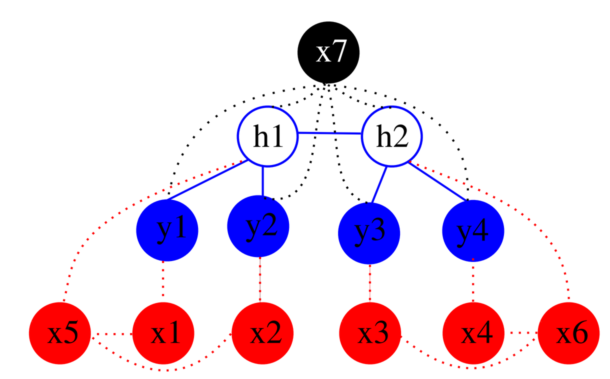
Goals
- Predict students' performance
- Analyze student performance
- Provide personalized feedback to students
- provide feedback to the course instructor

Main Idea
- Find underlying hidden groups of students
- Find predictive covariates
- Reduce dimensionality by latent variable modeling algorithms

Conditional Latent Tree Model (CLTM)

Combination of Conditional Random Fields (CRFs) and latent tree graphical models

Conditional Random Fields (CRFs)
- Relevant covariates affect student learning performance
  - Previous time observations
  - History of observations
  - Knowledge Components
  - Seasonality (day of week)



Latent Tree Graphical Models
- Shared underlying groups of students
- Students are not conditionally independent
- Latent trees are scalable and tractable

$$\Pr(Z|X,\boldsymbol{\theta}) = \exp\left(\sum_{k\in\mathcal{Z}} \phi_k(X,\boldsymbol{\theta}) z_k + \sum_{kl\in\mathcal{E}} \phi_{kl}(X,\boldsymbol{\theta}) z_k z_l - A\left(X,\boldsymbol{\theta}\right)\right)$$

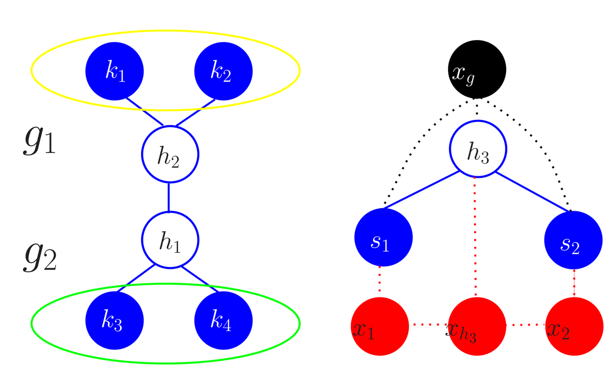
Learning CLTMs

Structure Learning
- Unsupervised learning
- Chow-Liu Recursive Grouping algorithm (CLRG)
- Consider conditional information distances in CLRG

Parameter Estimation
- Exponential family distribution with latent variables
- Expectation Maximization (EM) algorithm

Two phase model learning description for MOOCs

Steps
- Learn latent tree over knowledge components(KCs)
- Use graph partitioning on the learned latent tree to cluster KCs
- Use the cluster information as well as other relevant covariates to learn a CLTM for the students



(a) (b)

Data set Description

Student interactions with Stanford OLI in a Psychology course offered in Spring 2013

- data available on CMU Datashop
- $5,615$ students
- 266 knowledge components
- $2,493,612$ interaction records
- 60 train and 20 test data instances

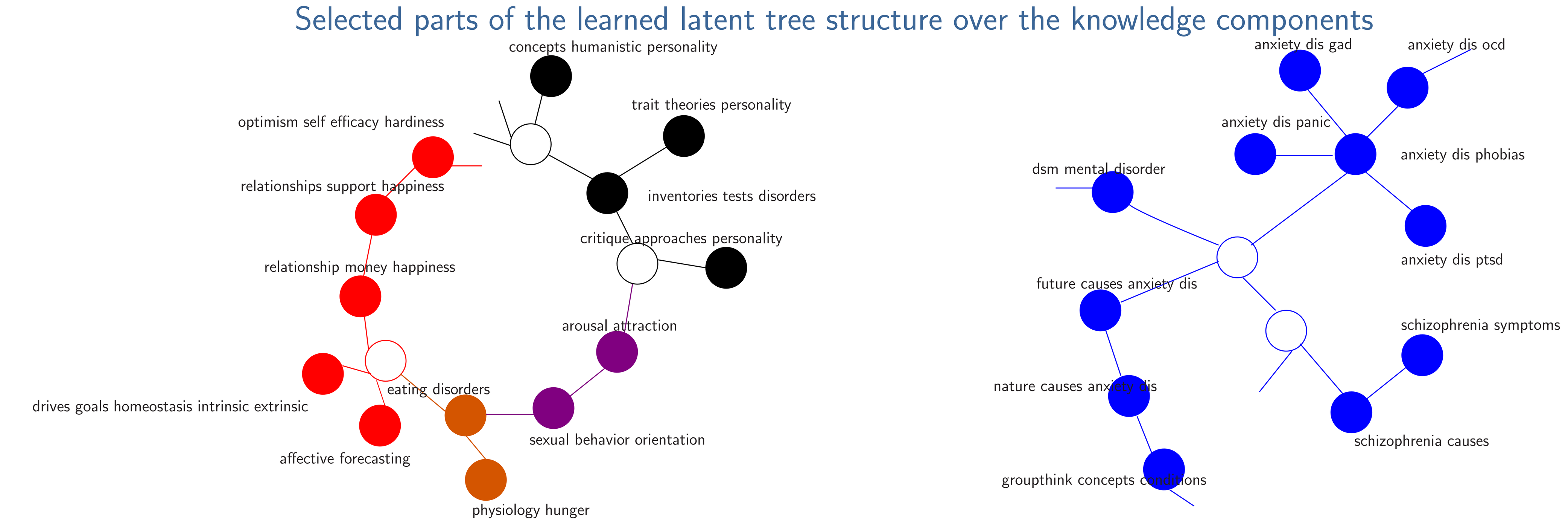
Project Page

Visit our project page:

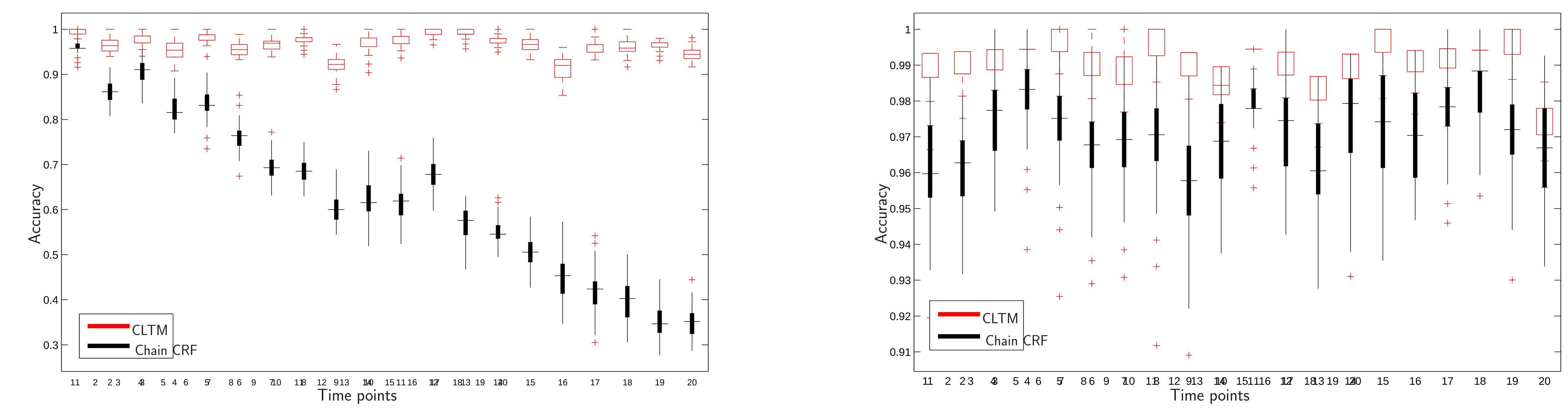
`http://newport.eecs.uci.edu/anandkumar/Lab/Lab_sub/CLTM.html`

Summary of Results

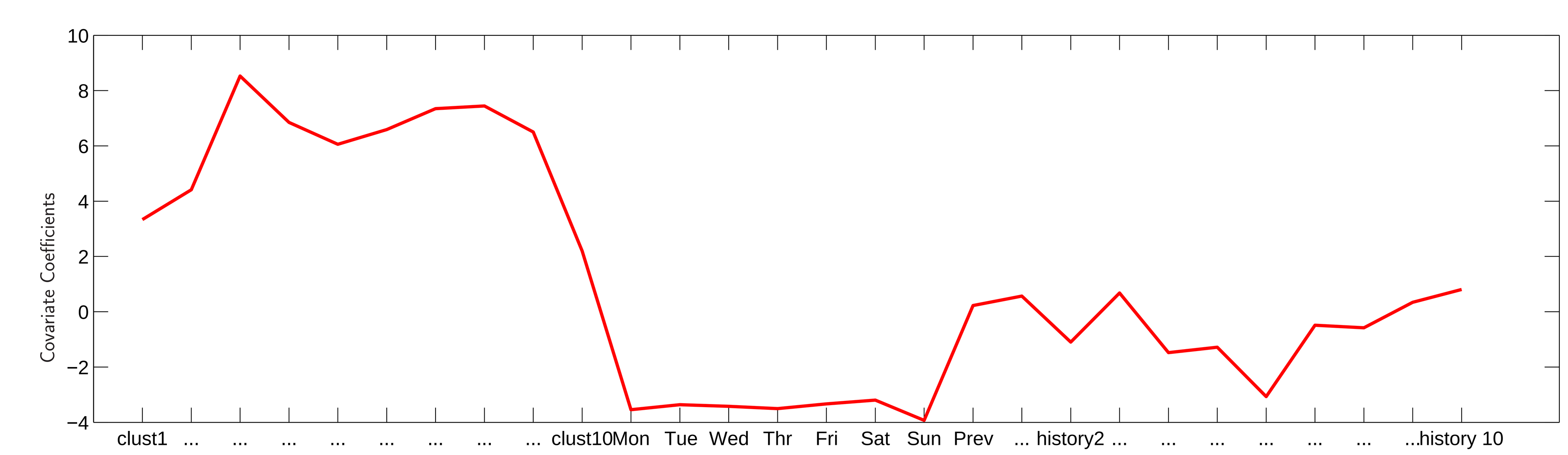
Selected parts of the learned latent tree structure over the knowledge components



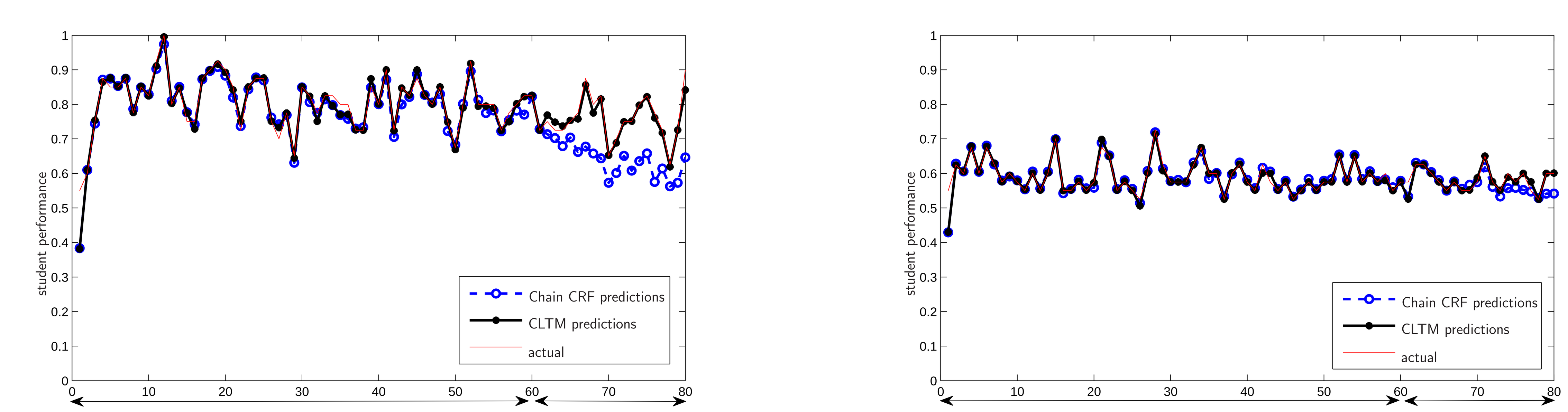
Student performance prediction



Learned Weights



Group of strong and weak learners



Conclusions
- Applicability of CLTM to student performance prediction
- Ability of giving qualitative analysis on student performance
- Groups of students learn differently