PREDICTIVE MODELING IN ONLINE LEARNING ENVIRONMENTS

UCI

Animashree Anandkumar, joint work with Forough Arabshahi, Furong Huang and Carter. T. Butts

Massive Open Online Courses

- Education available to everyone
- Ease of access
- Course access in non-sequential ways
- Greater learning gains

- Continued success in the domain relies on good feedback generation

Challenges and Goals

- Challenges:
  - Large number of students
  - Hard to track individual performance
  - Large number of courses
- Goals:
  - Analyze course content access
  - Predict student performance
  - Provide personalized feedback to students
  - Provide feedback to the instructor

Main Idea

- Knowledge components are being accessed by the students
  - Different components are related to one another and hence, pattern of their access correlated
  - Discover these correlations efficiently to predict student performance

- Students belong to latent or hidden groups
  - Similar performance within a group
  - Learn these groups to predict student access and performance

Conditional Latent Tree Models

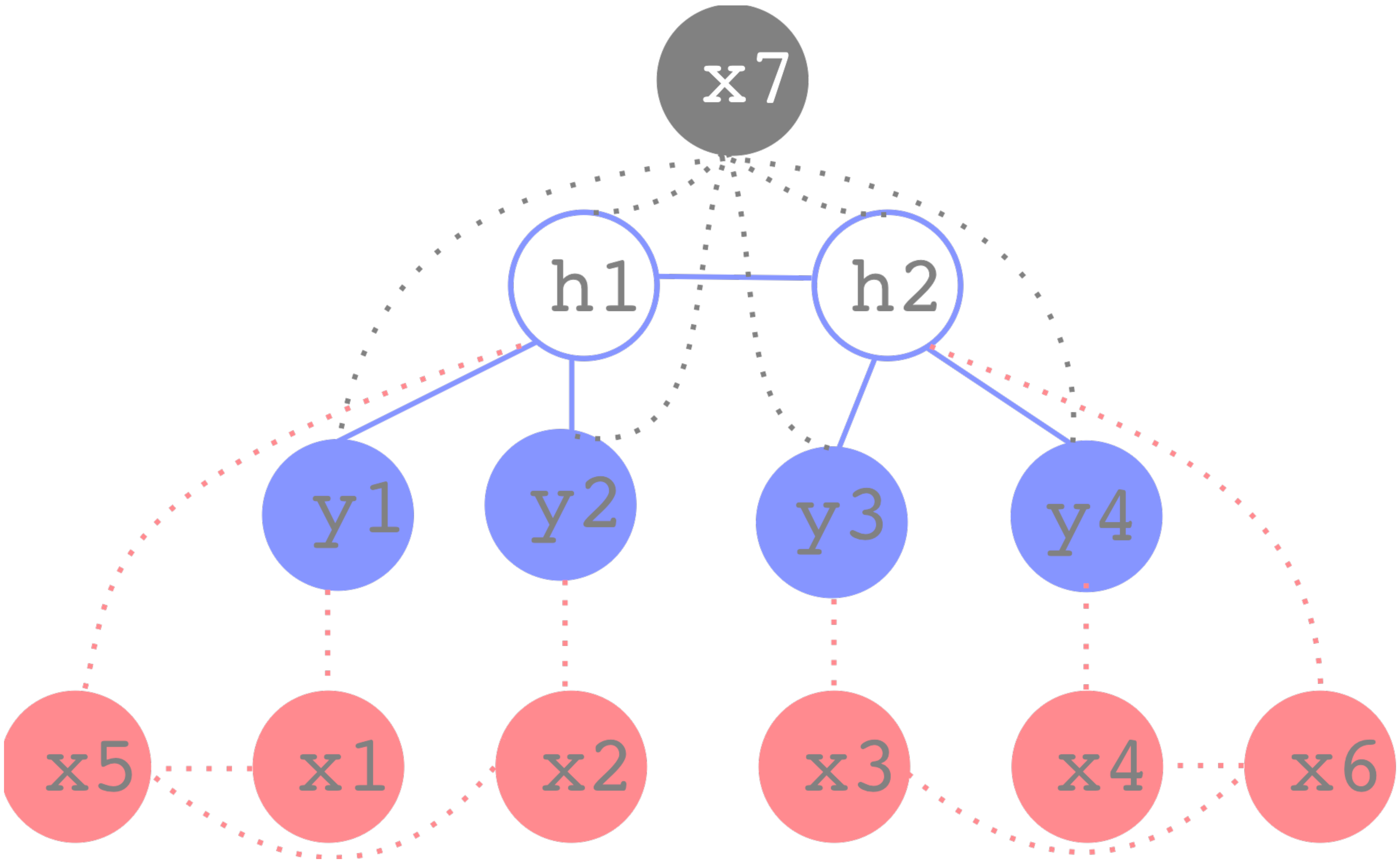
- Conditional Random Fields
  - Relevant covariates affect performance

- Probabilistic Graphical Models
  - Knowledge concepts have shared grouping
  - Students have shared grouping
  - Latent tree models are scalable and tractable

→ Conditional Latent Tree Model (CLTM)

Conditional Latent Tree Model

Toy structure over random variables. Blank nodes are hidden and blue nodes are observed random variables. Black and peach nodes are global and local covariates, respectively.



CLTM (Cont.)

- Covariates for grouping knowledge components:
  - Seasonality
  - Previous access of content
- Covariates for grouping students
  - History of performance and access
  - Knowledge component groups
- Underlying group structure
  - Latent tree graph
- Distribution
  - Exponential family distribution over the tree

CLTM (Cont.)

- Model distribution

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

- Z : Observed and hidden variables
- X : Covariates
- θ : Models parameters
- A(.) : Partition function
- $\phi_k, \phi_{k,l}$: Node and edge potential functions

Model Learning

- Structure learning
  - Unsupervised learning
  - Conditional latent tree learning algorithm : CLRG

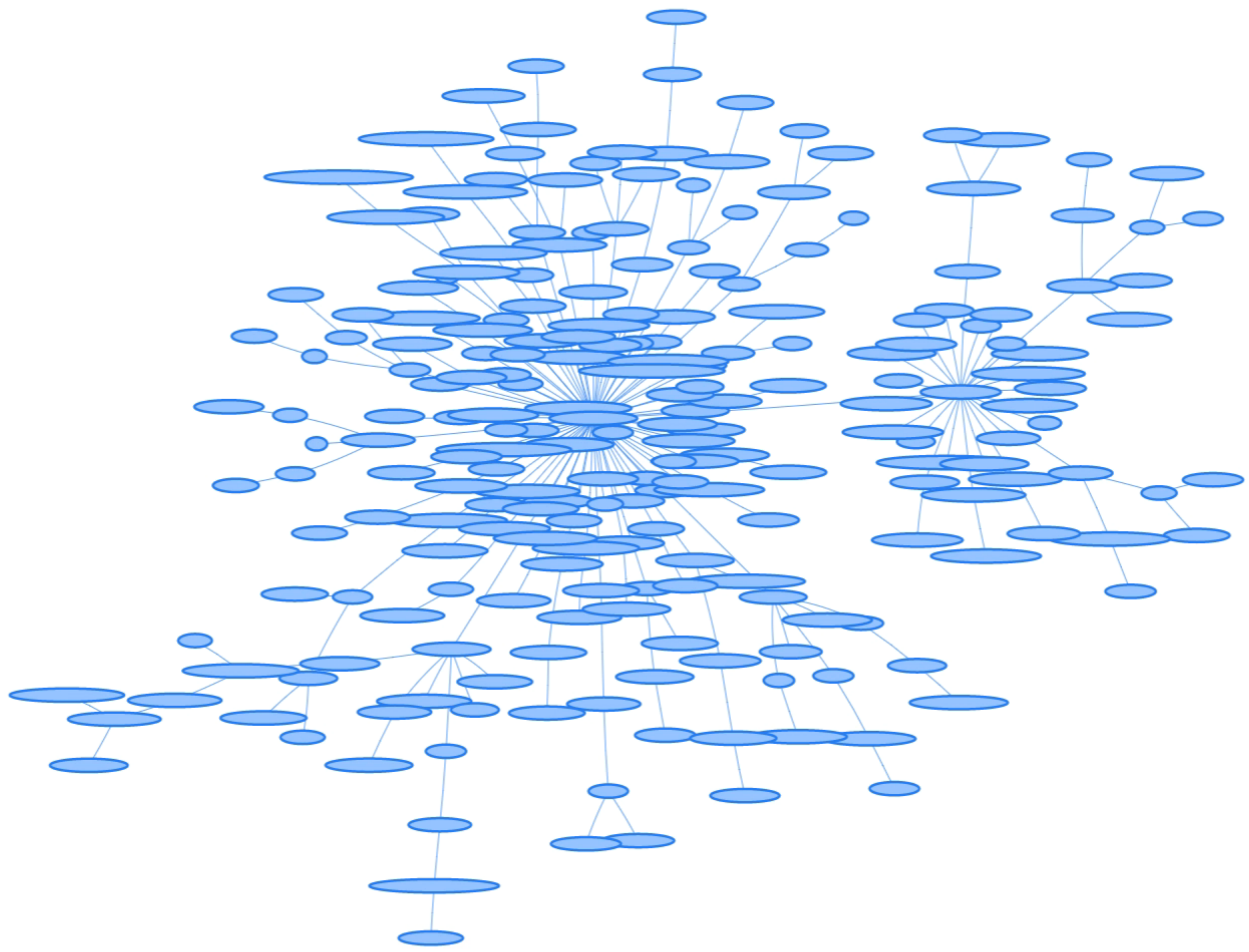
- Parameter estimation
  - Latent variables
  - EM algorithm

Data

- Psychology course on CMU datashop
- Spring 2013
- 5,616 students
- 266 knowledge components
- 2,493,612 interactions recorded
- 60 train and 20 test data instances (binned data)

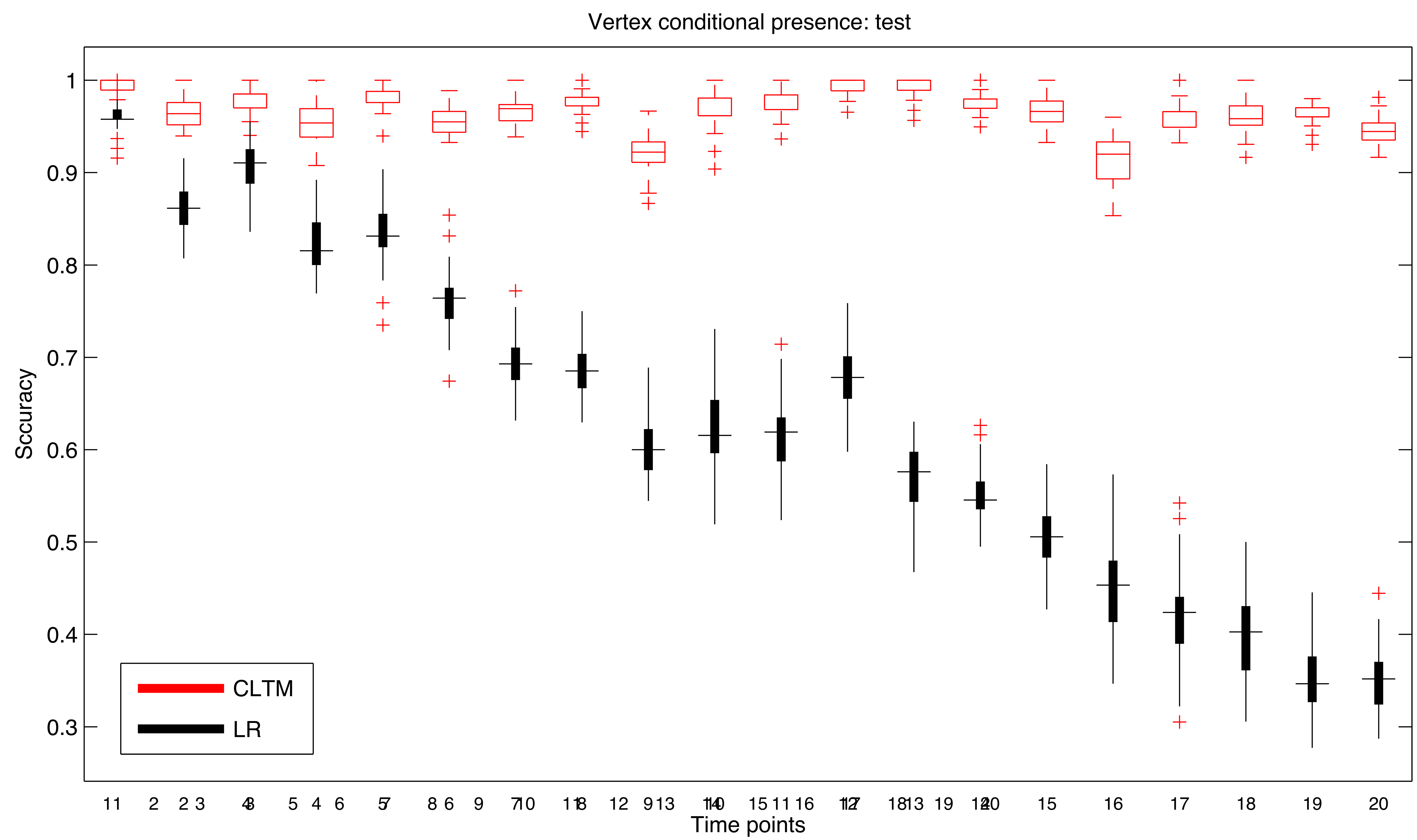
Knowledge Components

Graph learned over the knowledge components



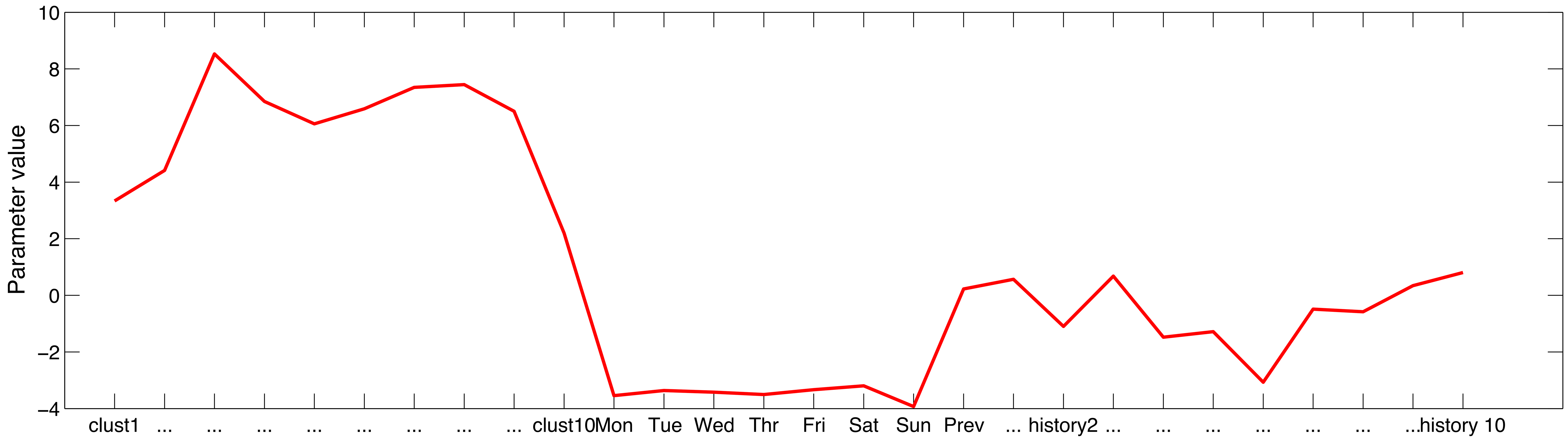
Student performance prediction

Comparison of prediction accuracy (recall) between CLTM (red) and chain CRF (black) on test data



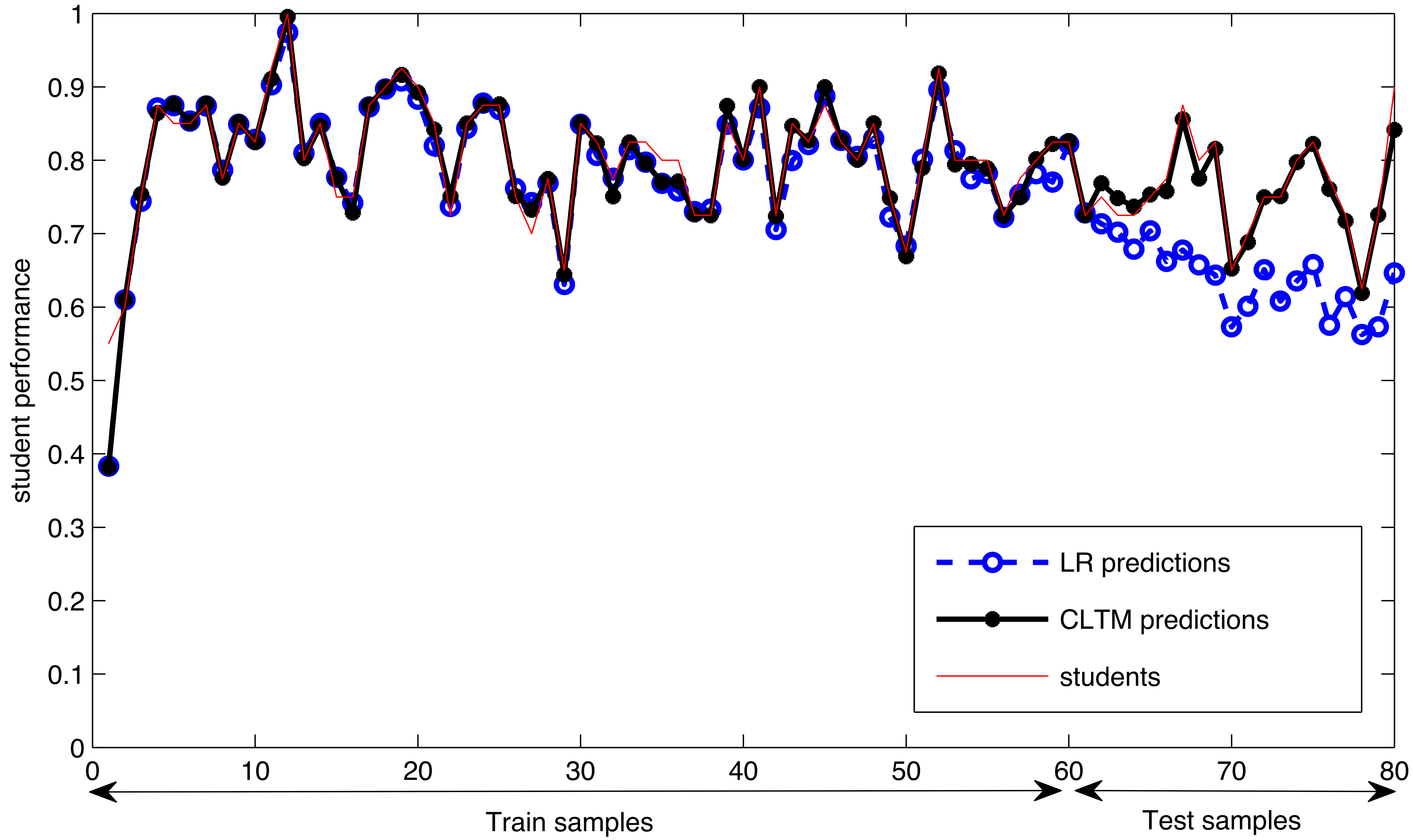
Model Weights

Learned covariate coefficients



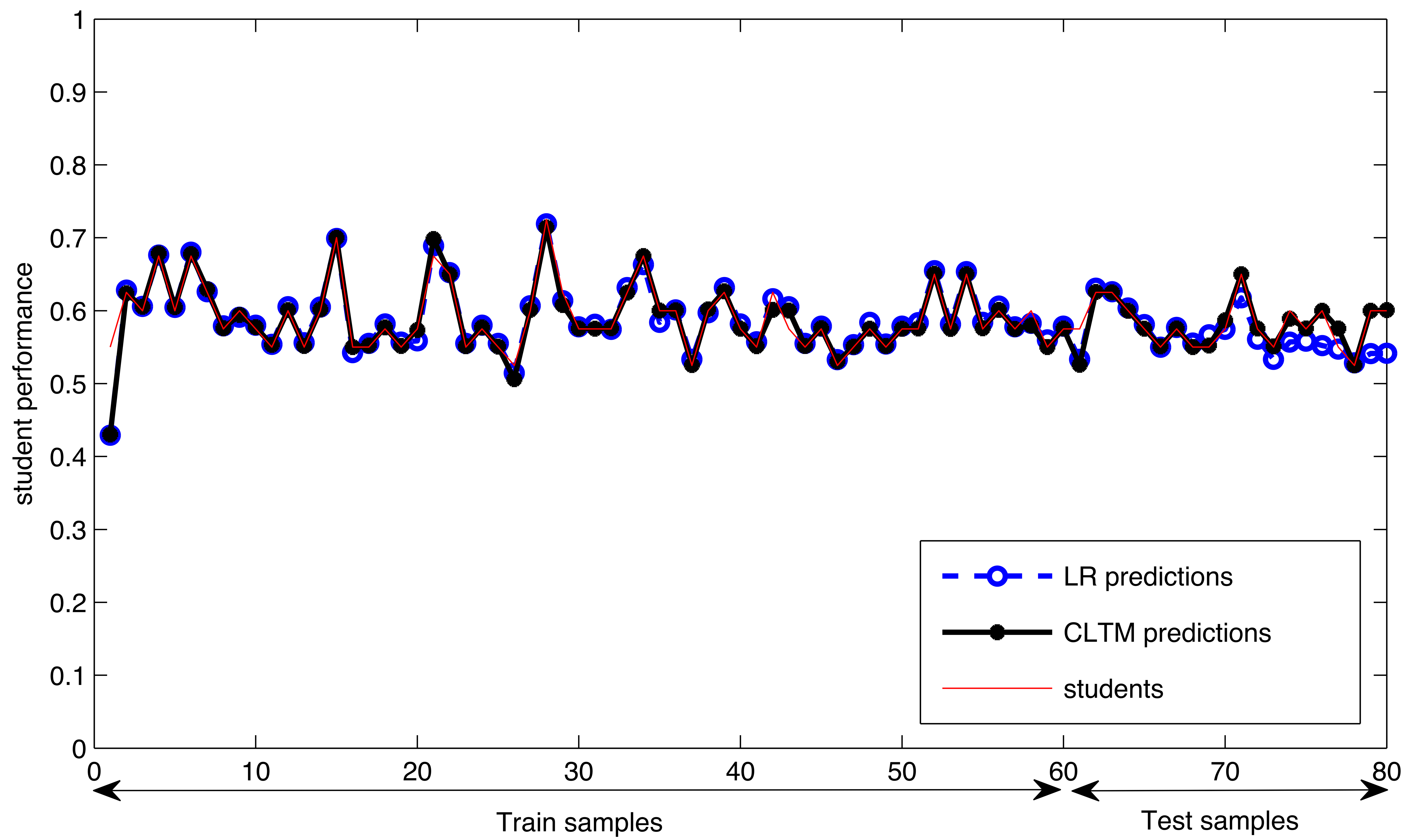
Group of Strong Learners

Group of Strong learners gathered from a neighborhood of students in the learned latent tree structure



Group of Weak Learners

Group of weak learners gathered from a neighborhood of students in the learned latent tree structure



Conclusions

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