Pipeline for Expediting Learning Analytics and Student Support from Data in Social Learning

Yohan Jo, Gaurav Tomar, Oliver Ferschke, Carolyn P. Rosé, Dragan Gašević
{yohanj, gtomar, ferschke, cprose}@cs.cmu.edu, dgasevic@acm.org

Motivation

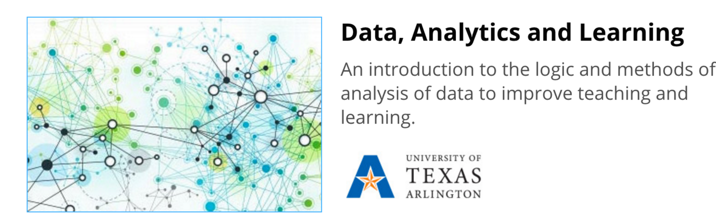
- The goal is to tighten the analytics cycle of data leading to insights on student needs and improvements in student support.
- We focus on social learning, where students learn through social interaction, e.g., via observation, help exchange, and discussion.

Contributions

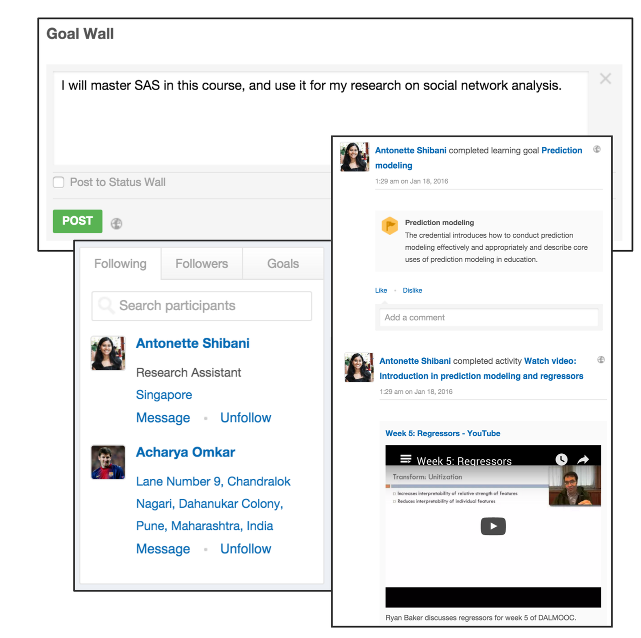
- Propose a pipeline and component models for data infrastructure, learning process analysis, and intervention.
- Demonstrate an application of the pipeline to real data to examine goal-setting behavior as qualifications of role models.

Course Context

Conventional xMOOC Platform



Self-Regulated Learning Platform



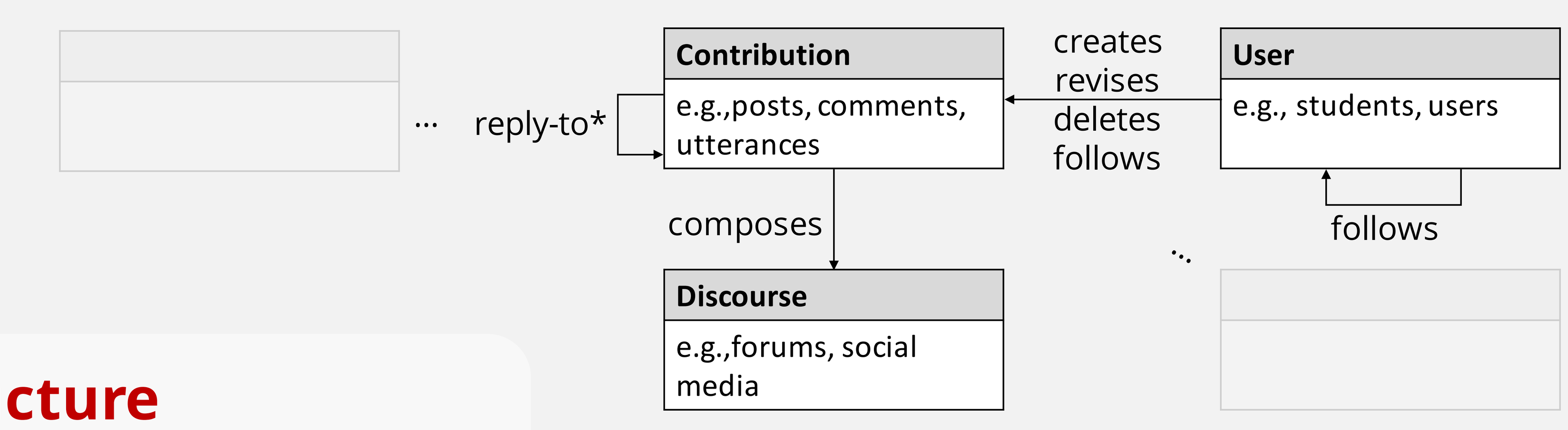
Period: Oct-Dec, 2014
Number of Students

- edX: 23,000
- ProSolo: 1,700

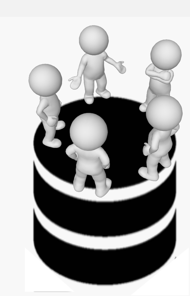
DiscourseDB (http://discoursedb.github.io)

- Maps diverse forms of textual conversations and social interactions into a common structure.
- Enables the subsequent components—learning process analysis and intervention—to apply the same tools to different data with little modification.
- Allows annotating entities and text spans manually or automatically.
- Keeps track of changes in relationships between entities and in the content of textual contributions.

Entity-Relation Model



*DiscourseDB allows defining arbitrary relations between contributions, avoiding data-specific tables.



Data Infrastructure unifies social interaction into a uniform interface

Learning Process Analysis models learner behaviors conditioned on social connection

Intervention helps students engage in beneficial social interaction

Temporal Bayesian Network

- Represents the building blocks (states) of learning process as
  - Distribution over discussion topics (θ)
  - Distribution over discussion media (ψ)
  - Transition probabilities to other states for each social connection type (π)

Discussion Media (e.g.)

- Blog, Twitter, Forum

Social Connection Types (e.g.)

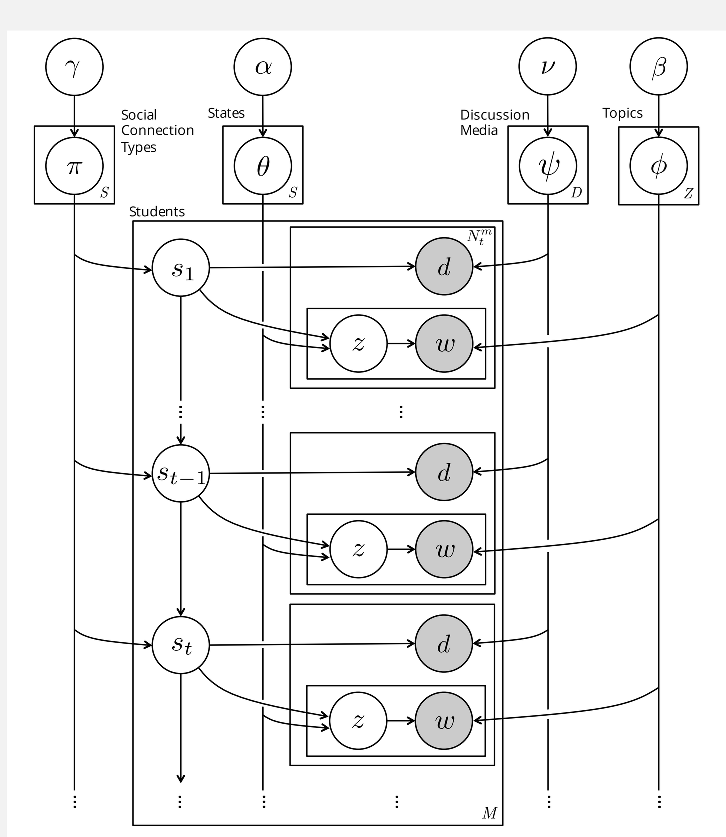
- Follows goal-setting peers
- Follows no one

Input

- Each student's discussions and social connection types over time

Output

- For each state: discussion topics and media, state transition probabilities
- For each student: state sequence



Recommender System

- Draws upon insights from the learning process analysis and aims to foster beneficial social connection among students.
- Recommends discussions to qualified students.
- Allows the discussants to interact with the qualified students.

Relevance Prediction

- Relevance between students and discussions is calculated using:
  - Students' expertise & motivation.
  - Discussion's length & popularity.

$$\underbrace{\hat{r}_{u,d}}_{\text{Relevance}} = bias + \overbrace{(P_u + \phi_u \Phi + \theta_u \Theta + \lambda_u \Lambda + \psi_u \Psi + \Gamma_\gamma)}^{\text{Student features}}{}^T \underbrace{(Q_d + \delta_d \Delta + l_d L + \frac{1}{\sqrt{|U(d)|}} \sum_{v \in U(d)} \varphi_v)}_{\text{Discussion features}}. \quad (1)$$

Constraint Filtering

- Helpful recommendation also requires constraints:
  - Every discussion has at least one qualified student.
  - A student is not overloaded.

$$\max \sum_{u,d} f_{u,d} \cdot \overbrace{r_{u,d}}^{\text{Relevance}} \overbrace{- \alpha \cdot \sum_d \sum_u \mathbb{1}(G_u \cdot f_{u,d} \geq G)(G_u - G) - \alpha \cdot \sum_d \sum_u \mathbb{1}(C_u \cdot f_{u,d} \geq C)(C_u - C)}^{\text{Penalty on workload}} \quad \text{s.t.}$$

Every discussion needs a qualified student

$$\begin{cases} \forall d \in D, \exists u \in U, G_u \cdot f_{u,d} \geq G \\ \forall d \in D, \exists u \in U, C_u \cdot f_{u,d} \geq C \end{cases} \quad (2)$$

Findings from Application

- **Learning process analysis** finds that students who follow goal-setting peers show positive learning behaviors:
  - Stay long in the course.
  - Engage in hands-on practices.
  - Revisit learning materials across the course.
- **Recommender system** finds that
  - Explicit intervention is necessary for helping students be aware of qualified students and interact with them.
  - Our algorithm effectively matches qualified students to relevant discussions while satisfying the constraints.
- **DiscourseDB** eases similar analysis and intervention on different data.

References

- **Detailed Article:** Y. Jo, G. Tomar, O. Ferschke, C. P. Rose, D. Gasevic. Expediting Support for Social Learning with Behavior Modeling, *EDM* '16.
- **Learning Process Model:** Y. Jo and C. P. Rose. Time Series Analysis of Nursing Notes for Mortality Prediction via a State Transition Topic Model. *CIKM* '15.
- **Recommendation Model:** D. Yang, D. Adamson, C. P. Rose. Question recommendation with constraints for massive open online courses. *RecSys* '14.
- **Course:** C. P. Rose, O. Ferschke, G. Tomar, D. Yang, I. Howley, V. Aleven, G. Siemens, M. Crosslin, D. Gasevic, and R. Baker. Challenges and Opportunities of Dual-Layer MOOCs: Reflections from an edX Deployment Study. *CSCL* '15.
- **More Research & Resources**: http://dance.cs.cmu.edu

Discussion Affordances for Natural Collaborative Exchange