UNIVERSITY AT ALBANY
State University of New York

Rank-Based Tensor Factorization for Predicting Student Performance

Thanh-Nam Doan, Sherry Sahebi

SUNY, Albany

Partially supported by the National Science Foundation, Grant No. 1755910

EDM 2019

Introduction

UNIVERSITY AT ALBANY
State University of New York

- Motivation: Online learning services are popular nowadays
  - Coursera: 33 million registered users, 2400 courses (June 2018)
  - Udacity: 1.6 million users (2014)
- Predicting students' performance is an essential problem in these systems
  - Early detect high-risk students that may quit or fail classes
  - Class evaluation
  - Course planning activities
  - Learning materials recommendation to students

Introduction

UNIVERSITY AT ALBANY
State University of New York

- Research question: How can we predict students’ performance
  - Do not require domain knowledge of the courses
  - Students freely select their own learning trajectory
  - Capture the gradual knowledge gain of students
    - Sometimes forget the concepts
    - Personalized learning rates

- Contributions: We propose Rank-based Tensor Factorization (RBTF) model that considers the above requirements

Related Works

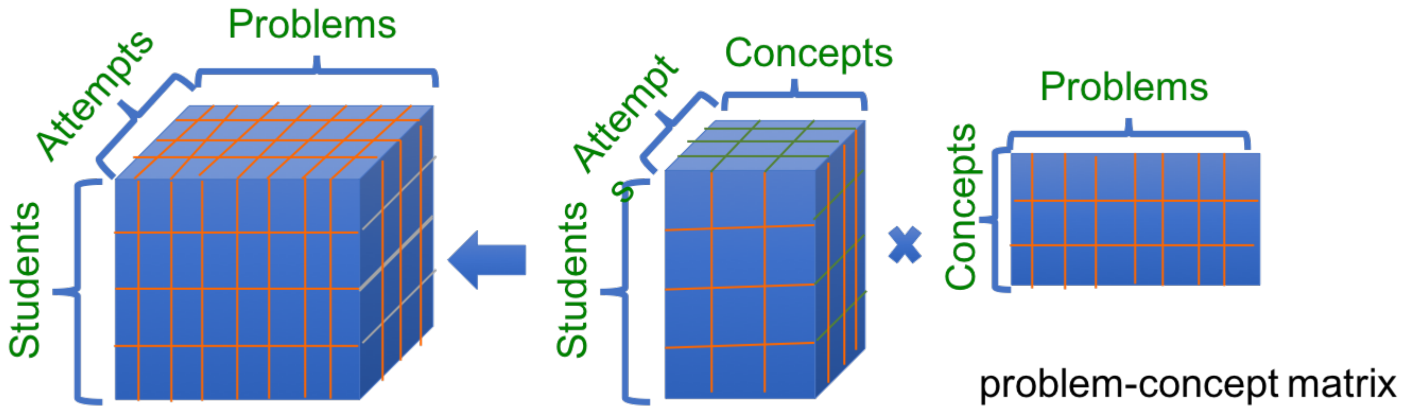
- Needing a predefined domain model
  - BKT, PFA, FAST, etc
- Recommender Systems - inspired
  - Apply recommender system techniques to educational data
  - Do not tailor for education data, or consider the sequence of student activities

Proposed model: student performance

UNIVERSITY AT ALBANY
State University of New York

- Student score tensor **Y** is factorized into student knowledge in concept $t_{a,s}$ and problem's latent concept vector q_p:

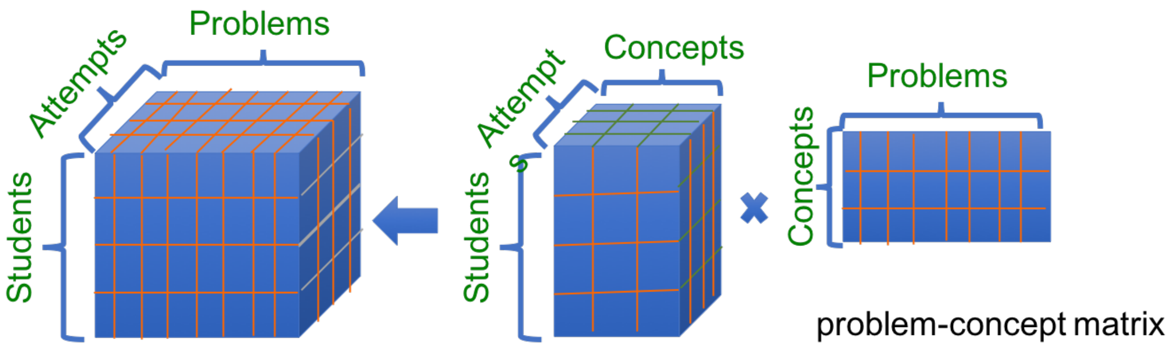
$$\hat{y}_{a,s,p} \approx t_{a,s}\, q_p + \underbrace{b_s + b_p + b_a + \mu}_{\text{biases}}$$



Proposed model: student performance

- Student score tensor **Y** is factorized into student knowledge in concept $t_{a,s}$ and problem's latent concept vector q_p:

$$\hat{y}_{a,s,p} \approx t_{a,s}\, q_p + \underbrace{b_s + b_p + b_a + \mu}_{\text{biases}}$$



- Learning parameters is the optimization problem by minimizing L_1

$$L_1 = \sum_{a,s,p}\left(\hat{y}_{a,s,p} - y_{a,s,p}\right)^2 + regularization$$

Proposed model: gradual knowledge gain

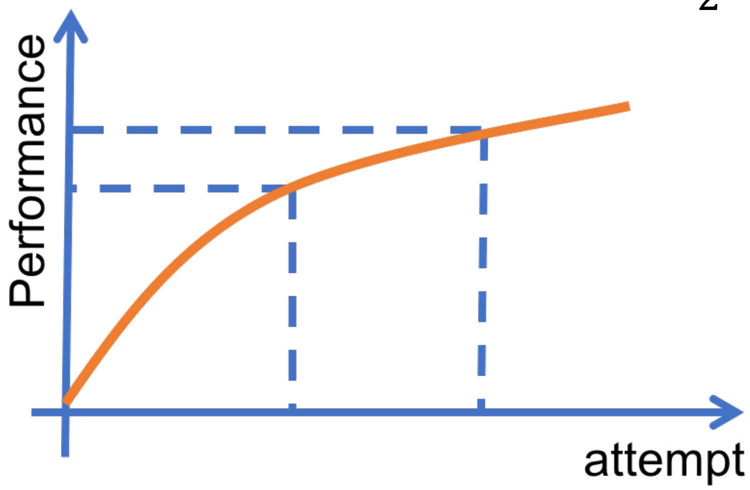
UNIVERSITY AT ALBANY
State University of New York

- To capture the gradual learning, we assume that a student knowledge increases over time

$$t_{a,s}q_p - t_{a-1,s}q_p \geq 0$$

- For attempt a of student s, the ranking of s's score at a is higher than the one of s at j with $j < a$

$$L_2 = \sum_{j=1}^{a} \sum_{s} \sum_{p} \log(\sigma(t_{a,s}q_p - t_{j,s}q_p))$$



Proposed model: gradual knowledge gain

UNIVERSITY AT ALBANY
State University of New York

- To capture the gradual learning, we assume that a student knowledge increases over time

$$t_{a,s}q_p - t_{a-1,s}q_p \geq 0$$

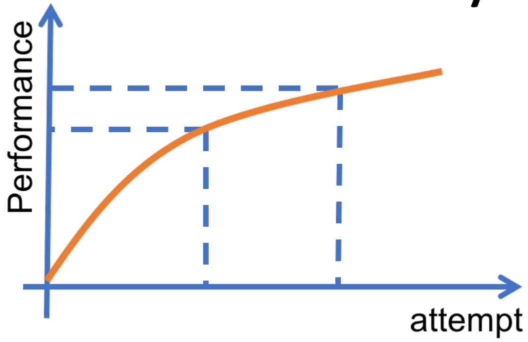
- For attempt a of student s, the ranking of s's score at a is higher than the one of s at j with $j < a$

$$L_2 = \sum_{j=1}^{a} \sum_{s} \sum_{p} \log(\sigma(t_{a,s}q_p - t_{j,s}q_p))$$

- We embed the gradual knowledge gain into tensor factorization by minimizing L

$$L = L_1 - \omega L_2$$

- ω controls the contribution of gradual learning



Dataset & Experiment Setup

UNIVERSITY AT ALBANY
State University of New York

- Canvas network data
- 80% data is for training, 20% is for testing

Dataset	#students	#problems	#attempts	Avg. attempts
Course 1	531	91	87	29.92
Course 2	2597	32	30	12.73

Baselines & Metrics

UNIVERSITY AT ALBANY
State University of New York

- Baselines
  - Feedback-Driven Tensor Factorization (FDTF): It has “hard” constraint on gradual knowledge gain [Sahebi et al., 2016]
  - SPARse Factor Analysis (SPARFA): It calculates the probability of students’ correct response [Lan et al., 2014]

- Metrics
  - Root Mean Squared Error (RMSE)
  - Accuracy

Student Performance Prediction

UNIVERSITY AT ALBANY
State University of New York

- RBTF and FDTF is better than SPARFA → the importance of considering student sequence
- RBTF is better than FDTF → gradual knowledge gains should be model flexibly and allow for occasional forgetting of concepts

Dataset	RMSE			Accuracy		
	RBTF	FDTF	SPARFA	RBTF	FDTF	SPARFA
Course 1	**0.12**	0.27	0.59	**92.5%**	85.2%	81.7%
Course 2	**0.2056**	0.2116	0.567	**95.24%**	92.8%	87.41%

Hyper-parameter Sensitivity Analysis

UNIVERSITY AT ALBANY
State University of New York

- Sensitivity to ω:
  - ω controls the trade off between having accurate estimation of student performance and constraint of knowledge increase
  - Larger ω: more emphasis on knowledge increase
  - Tune value of ω from 0 to 1 and measure RMSE of model
- Result:
  - $\omega = 0.5$ has the best performance in both dataset
  - Course 2 is more sensitive due to being more sparse

	ω				
Dataset	0.01	0.25	0.5	0.75	1.0
Course 1	0.191	0.128	**0.12**	0.137	0.141
Course 2	0.233	0.2064	**0.2056**	0.2154	0.2224

Hyper-parameter Sensitivity Analysis

UNIVERSITY AT ALBANY
State University of New York

- Sensitivity to k:
  - *k* is the number of concepts
  - The larger *k*, the larger the latent space of students and questions
  - We tune value of *k* and measure RMSE
- Result
  - Increasing *k* makes RBTF performs slightly worse
  - RBTF is robust since the the increase in error is mirror

Dataset	k = 3	5	10	15
Course 1	**0.12**	0.122	0.127	0.128
Course 2	**0.2056**	0.206	0.2065	0.2065

Conclusion

- We proposed a novel rank-based tensor factorization (RBTF)
  - RBTF considers the sequence of student activities
  - RBTF considers the gradual knowledge gains, allowing for occasional forget
  - RBTF does not require the prior knowledge of courses

- We evaluate RBTF on the task of students’ performance prediction

UNIVERSITY AT ALBANY
State University of New York

Thank you
Q&A

ssahebi@albany.edu