The 26th ACM International Conference on
Information and Knowledge Management (CIKM)
2017/11/06-2017/11/10, Singapore, Singapore

Tracking Knowledge Proficiency of Students with Educational Priors

Yuying Chen[1], Qi Liu[1], Zhenya Huang[1], Le Wu[2], Enhong Chen[1*],
Runze Wu[1], Yu Su[3], Guoping Hu[4]

[1] University of Science and Technology of China
[2] Hefei University of Technology
[3]Anhui University [4]iFLYTEK Research

Outline

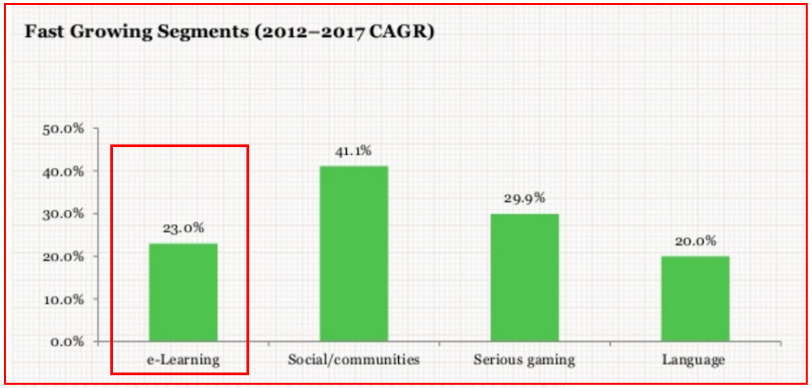
- **Backgroud and Related Work**
- Problem Statement
- Methodology
- Experiments
- Conclusion

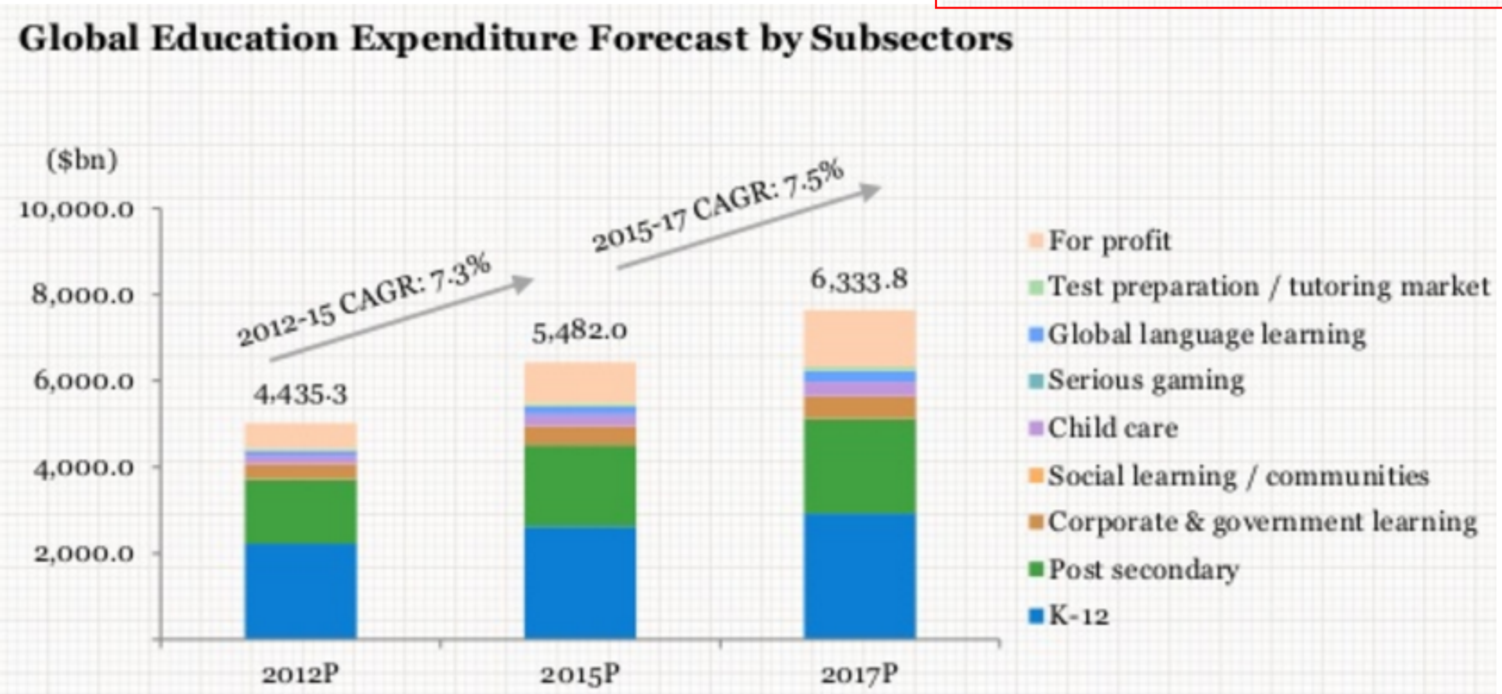
Background

- Traditional teaching method
  - Classroom Teaching
    - The teacher's energy is limited.
    - The same learning strategy, same exercises, impersonality.
  - Extracurricular Tutorials
    - Teaching quality is difficult to guarantee
    - A higher cost

Background

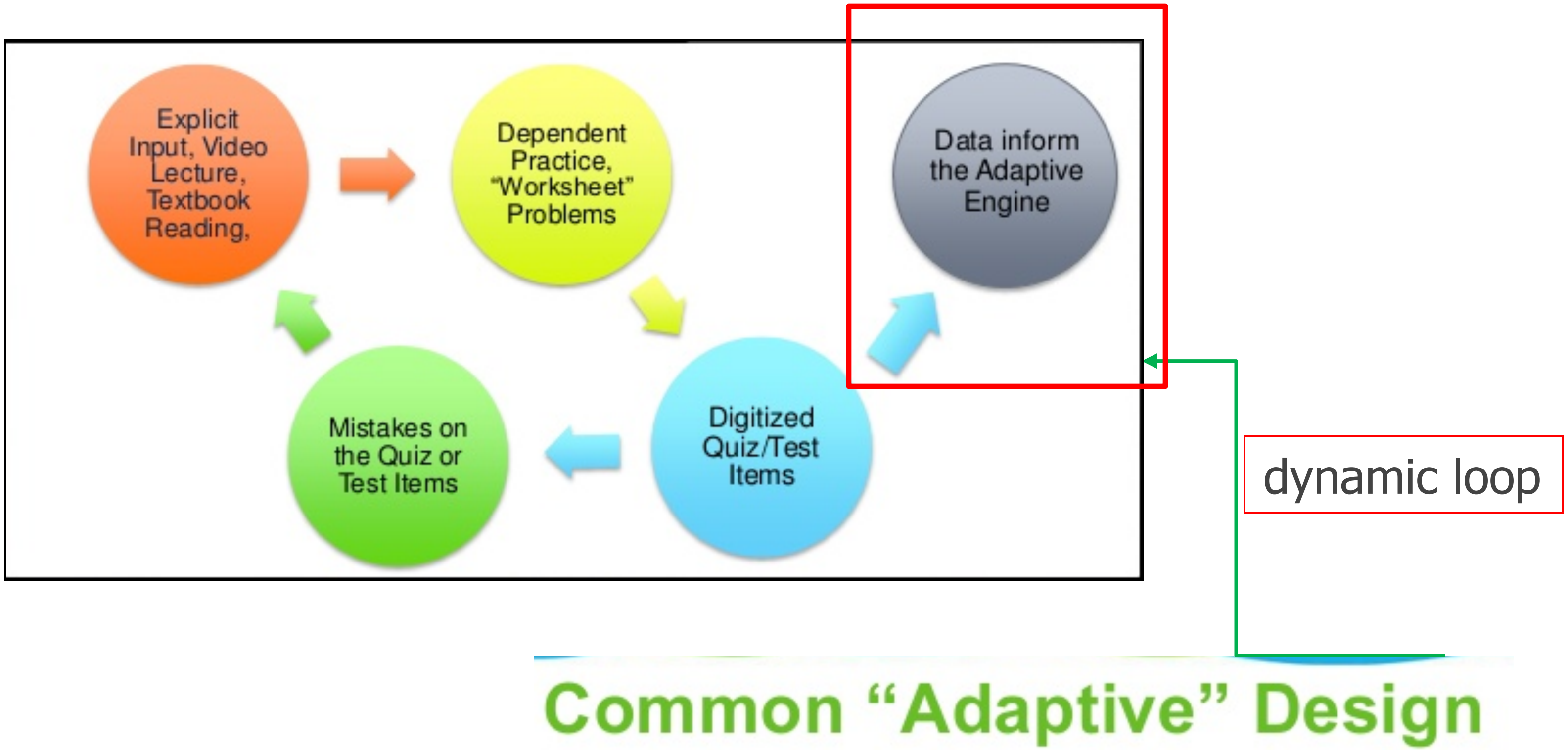
- E-Learning(Online learning)
  - **Knewton**
  - **Cognitive Tutor**
  - **etc**





Background

- Education Service Systems
  - Various online tutoring systems allow students to learn and do exercises individually.



Related work-static

- IRT

$$P(X_{ij}=1|\theta_j)=c_i+\frac{1-c_i}{1+\exp[-1.7a_i(\theta_j-b_i)]}$$

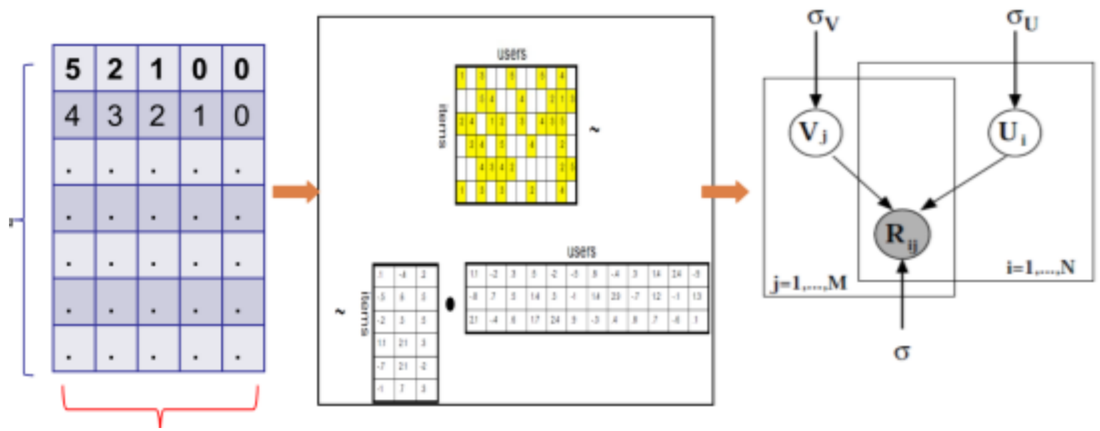
- DINA

$$P_j(\boldsymbol{\alpha}_i)=P(X_{ij}=1|\boldsymbol{\alpha}_i)=g_j^{1-\eta_{ij}}(1-s_j)^{\eta_{ij}}.$$

- PMF

$$p(R|U,V,\sigma^2)=\prod_{i=1}^{N}\prod_{j=1}^{M}\left[\mathcal{N}(R_{ij}|U_i^TV_j,\sigma^2)\right]^{I_{ij}}$$

they are only good at predicting student's proficiency from a ***static perspective.***



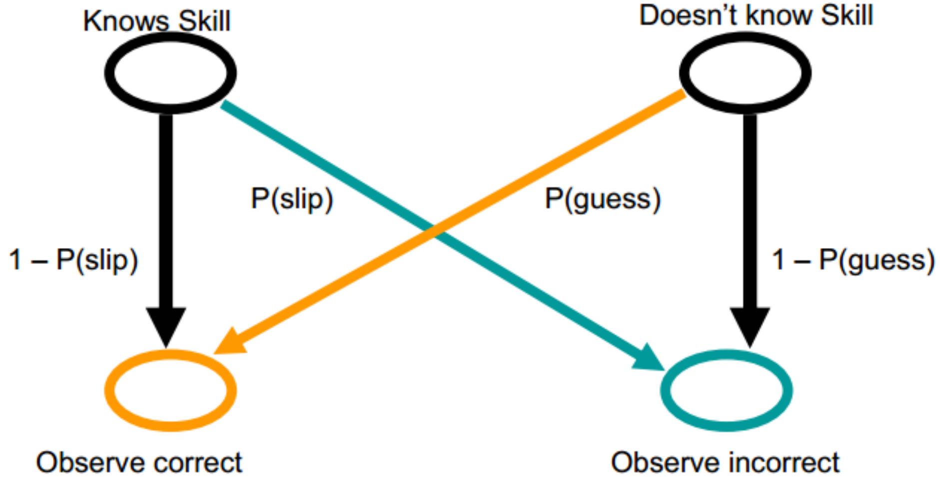
Related work-dynamic

- LFA - one-dimensional

$$m(i, j \in KCs, n) = \alpha_i + \sum_{j \in KCs} (\beta_j + \gamma_j n_{i,j})$$

$$p(m) = \frac{1}{1+e^{-m}}$$

- BKT- binary entities

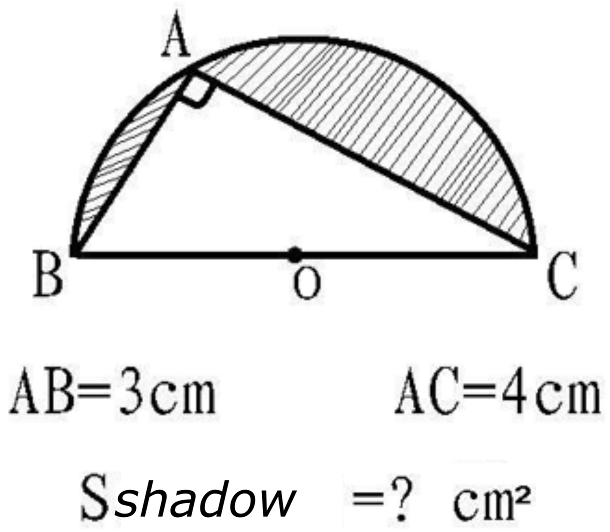


Outline

- Backgroud and Related Work
- **Problem Statement**
- Methodology
- Experiments
- Conclusion

Motivation

- Problem: How to **track** students' **knowledge proficiency** over time. (TKP task)?
- Opportunity
  - Widely use of **Intelligent tutoring system**
  - Record exercises logs and Q-matrix
  - **Educational Priors**
- Focus on Math problem

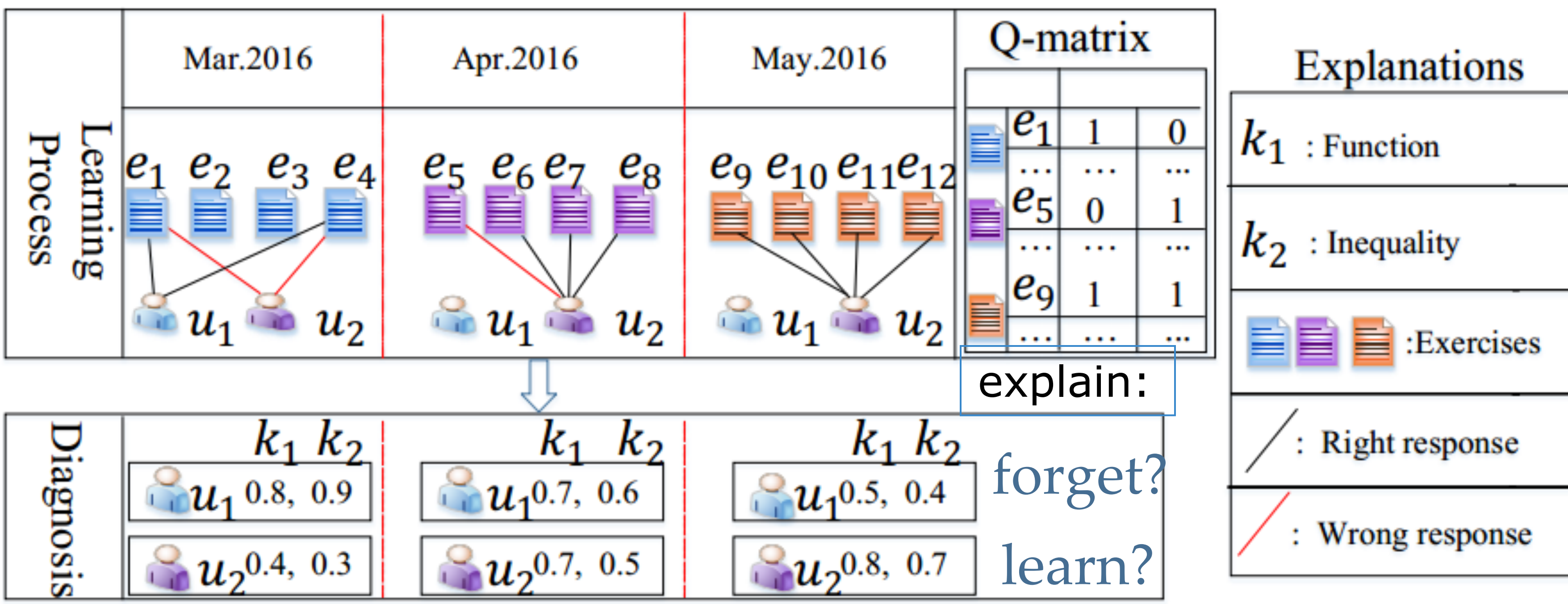


Problem Statement

- Given the students' response tensor R and Q-matrix labelled by educational experts
- our goal is two-fold:
  - modeling the change of students' knowledge proficiency from time 1 to T .
  - predicting students' knowledge proficiency and responses in time T + 1.
- Challenge:
  - 1. How to get a student's knowledge proficiency?
  - 2. How to explain the change of knowledge proficiency over time?

A toy example

- A showcase of KPD task on mathematical exercises related to the knowledge points of Function and Inequality

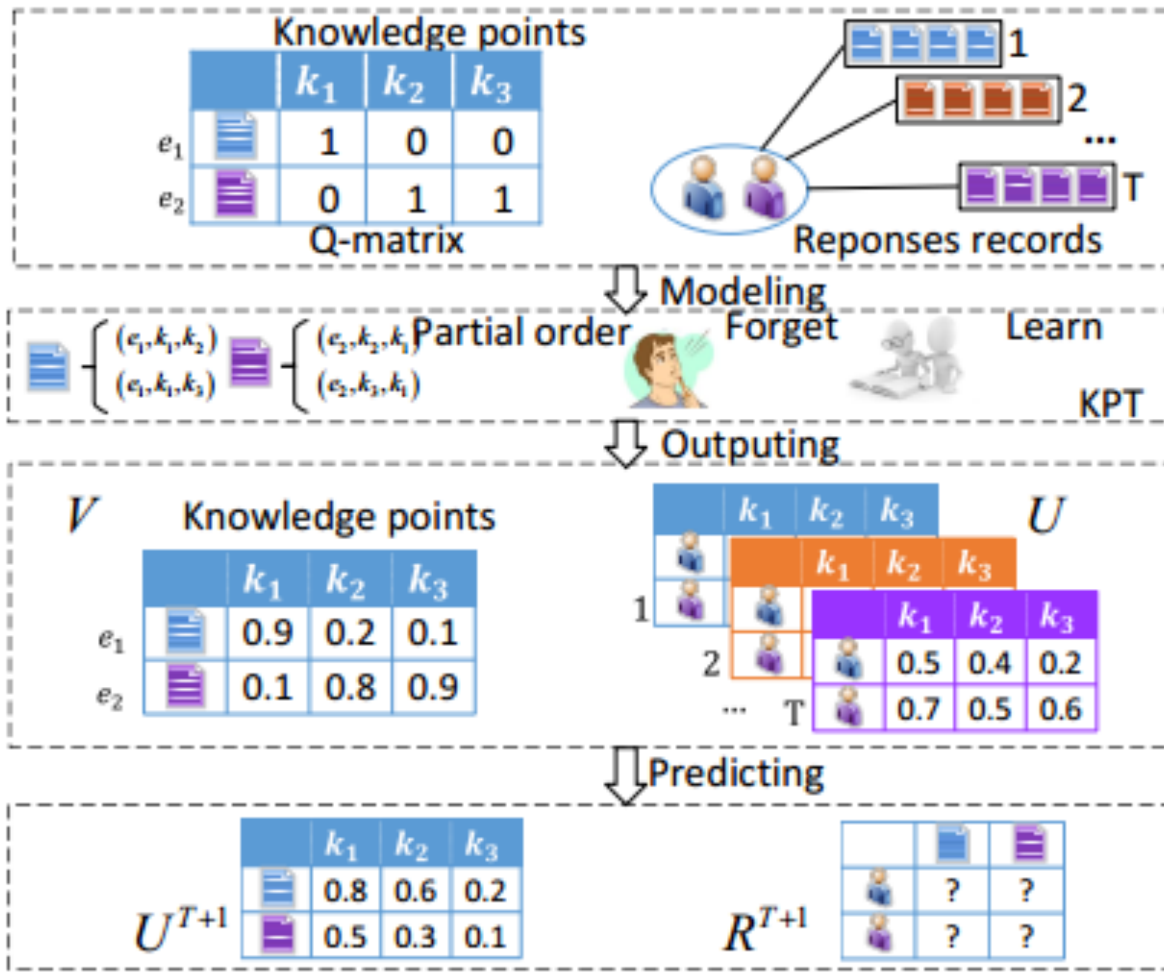


Outline

- Backgroud and Related Work
- Problem Statement
- **Methodology**
  - **Probabilistic Modeling with Priors**
  - **Model Learning and Prediction**
- Experiments
- Conclusion

Framework

- KPT: a supervised way
  - Modeling Stage
    - Partial order
    - Forget
    - Learn
  - Predicting Stage
    - Input: U^{1-T}, V
    - Output: U^{T+1}, R^{T+1}



KPT model

- **Probabilistic Modeling with Priors**
  - for each student and each exercise, we model the response tensor R as:

$$p(R|U,V,b) = \prod_{t=1}^{T}\prod_{i=1}^{N}\prod_{j=1}^{M}[\mathcal{N}(R_{ij}^{t}|\langle U_i^t, V_j\rangle - b_j, \sigma_R^2)]^{I_{ij}^{t}}, \quad (1)$$

    - $U_i^t \in \mathbb{R}^{K\times 1}$ is the knowledge proficiency of student i
    - $V \in \mathbb{R}^{M\times K}$ denotes the relationship between exercises and knowledge points
  - How to establish the corresponding relationship between students, exercises and knowledge points?

Modeling V with the Q-matrix prior

- Q-matrix
  - depicts the knowledge points of the exercises
  - each row denotes an exercise
  - each column stands for a knowledge point.

	Function	Solid Geometry	Arithmetic Progression	Inequation
exercise1	1	0	0	0
exercise2	1	1	0	1
exercise3	1	0	1	0
exercise4	0	1	0	0

  - The sparsity with the binary entities does not fit probabilistic modeling well.

Modeling V with the Q-matrix prior

- for exercise j, if a knowledge point q is marked as 1, then we assume that q is more relevant to exercise j than p with mark 0

$$q >_j^+ p, \text{ if } Q_{jq} = 1 \text{ and } Q_{jp} = 0.$$

Partial order

- After that, we can transform the original Q-matrix into a set of comparability $D_T \in R^{M \times K \times K}$ by: $D_T = \{(j, q, p) | q >_j^+ p\}.$

exercises	knowledge			
	K1	K2	K3	K4
E1	1	0	1	0
E2	0	1	0	0
E3	0	0	0	1

Q-matrix

(E1,K1,K2)
(E1,K1,K4)
(E1,K3,K2)
(E1,K3,K2)
...
(E3,K4,K1)
(E3,K4,K2)
(E3,K4,K3)

Modeling V with the Q-matrix prior

- we define the probability that exercise j is more relevant to knowledge point q than knowledge point p as:

$$p(q >_j^+ p|V_j) = \frac{1}{1 + \mathrm{e}^{-(V_{jq} - V_{jp})}}. \qquad (6)$$

- the log of the posterior distribution

$$\begin{aligned}\ln p(V|D_T) &= \ln \prod_{(j,q,p)\in D_T} p(>_j^+ |V)p(V)\\ &= \sum_{j=1}^{M}\sum_{q=1}^{K}\sum_{p=1}^{K} I\left(q >_j^+ p\right) \ln \frac{1}{1 + \mathrm{e}^{-\left(V_{jq} - V_{jp}\right)}}\\ &- \frac{1}{2\sigma_V^2}||V||_F^2.\end{aligned} \qquad (7)$$

Modeling U with learning theories.

- we assume a student's current knowledge proficiency is mainly influenced by two underlying reasons:
  - She forgets her previous knowledge proficiency over time.
  - The more exercises she does, the higher level of related knowledge proficiency she will get.
  - We model the two effects of each student's knowledge proficiency in time window t = 2; 3; :::; T as:

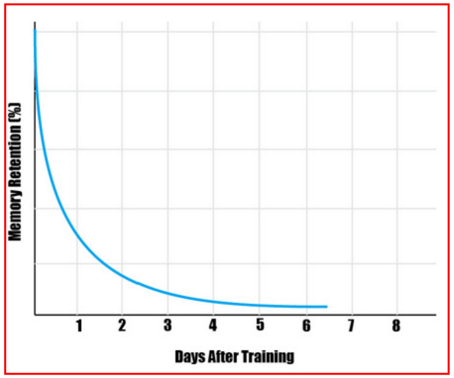
$$p(U_i^t) = \mathcal{N}(U_i^t | \bar{U}_i^t, \sigma_U^2 \mathbf{I}), \text{ where } \quad \bar{U}_i^t = \{\bar{U}_{i1}^t, \bar{U}_{i2}^t \ldots \bar{U}_{iK}^t\}$$

$$\bar{U}_{ik}^t = (1 - \alpha_i) \underbrace{\boxed{f^t(*)}}_{\text{forgetting}} + \alpha_i \underbrace{\boxed{l^t(*)}}_{\text{learning}}, \; s.t. \quad 0 \leq \alpha_i \leq 1, \qquad (8)$$

Modeling U with learning theories.

- $f^t(*)$depicts the decline of knowledge over time:

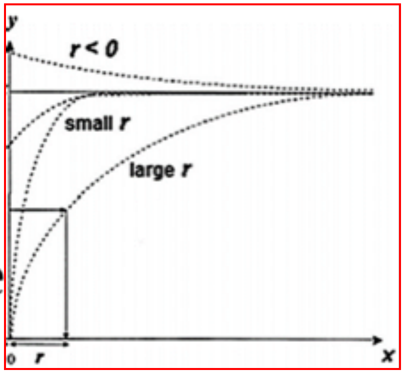
$$f^t(*) = U_{ik}^{t-1} e^{-\frac{\Delta t}{S}}$$



  - Δt is the time interval
  - S denotes the strength of memory.

- $l^t(*)$captures the growth of knowledge with exercises:

$$l^t(*) = U_{ik}^{t-1} \frac{D f_k^t}{f_k^t + r}$$



  - f_k^t denotes the frequency of knowledge k
  - r and D control the magnitude and multiplie of growth respectively.

Model Learning and Prediction

- graphical representation of the proposed latent model

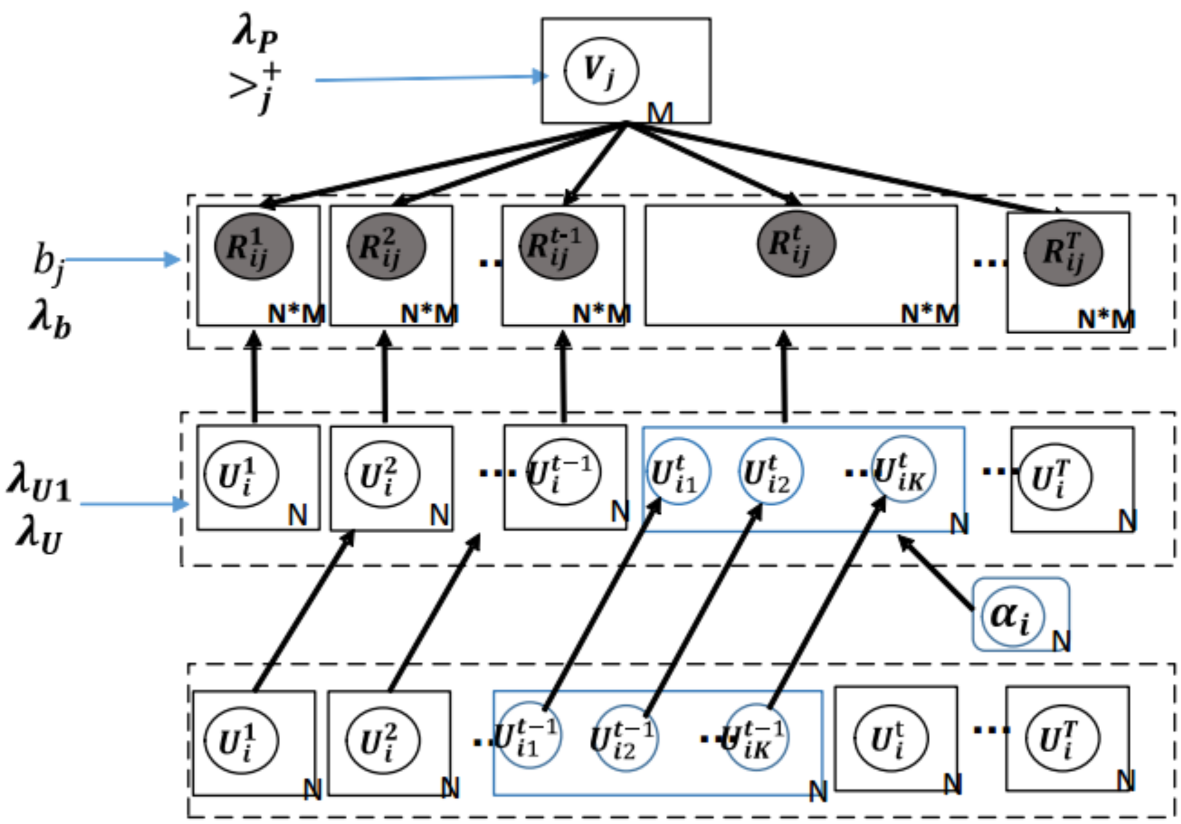


Figure 3: Graphical representation of KPT.

Model Learning and Prediction

- our goal is to learn the parameters Φ= [U,V, α, b]
  - Particularly, the posterior distribution over Φ is:

$$p(U, V, \alpha, b | R, D_T) \propto p(R | U, V, \alpha, b) \times p(U | \sigma_U^2, \sigma_{U1}^2) \times p(V | D_T).$$

  - Maximizing the log posterior of the above equation is equivalent to minimize the following objective:

$$\begin{aligned}
\min_{\Phi} \mathcal{E}(\Phi) &= \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij}^t [\hat{R}_{ij}^t - R_{ij}^t]^2 \\
&- \boxed{\lambda_P \sum_{j=1}^{M} \sum_{q=1}^{K} \sum_{p=1}^{K} I\left(q >_j^+ p\right) \ln \frac{1}{1 + e^{-(V_{jq} - V_{jp})}}} + \frac{\lambda_V}{2} \sum_{i=1}^{M} ||V_i||_F^2 \\
&+ \boxed{\frac{\lambda_U}{2} \sum_{t=2}^{T} \sum_{i=1}^{N} ||\overline{U_i^t} - U_i^t||_F^2} + \frac{\lambda_{U1}}{2} \sum_{i=1}^{N} ||U_i^1||_F^2, \qquad (13)
\end{aligned}$$

Model Learning and Prediction

- With students' knowledge proficiency $U^1, U^2, \ldots, U^T$ and related parameters, students' responses and knowledge proficiency in the next time can be calculated as:

$$U_i^{(T+1)} = \left\{U_{i1}^{(T+1)}, U_{i2}^{(T+1)}, \ldots, U_{iK}^{(T+1)}\right\},$$

$$U_{ik}^{(T+1)} \approx (1-\alpha_i)U_{ik}^T e^{-\frac{\Delta(T)}{S}} + \alpha_i U_{ik}^T \frac{M f_k^{T+1}}{f_k^{T+1}+r},$$

$$\hat{R}_{ij}^{(T+1)} \approx \langle U_i^{(T+1)}, V_j\rangle - b_j .$$

Outline

- Backgroud and Related Work
- Problem Statement
- Methodology
- **Experiments**
- Conclusion

Experiments

- Dataset
  - *Two private datasets* which are collected from daily exercise records of high school students
  - *ASSIST* is a public dataset *Assistments*[1] 2009-2010 "Non-skill builder"

Table 4: The statistics of the three datasets.

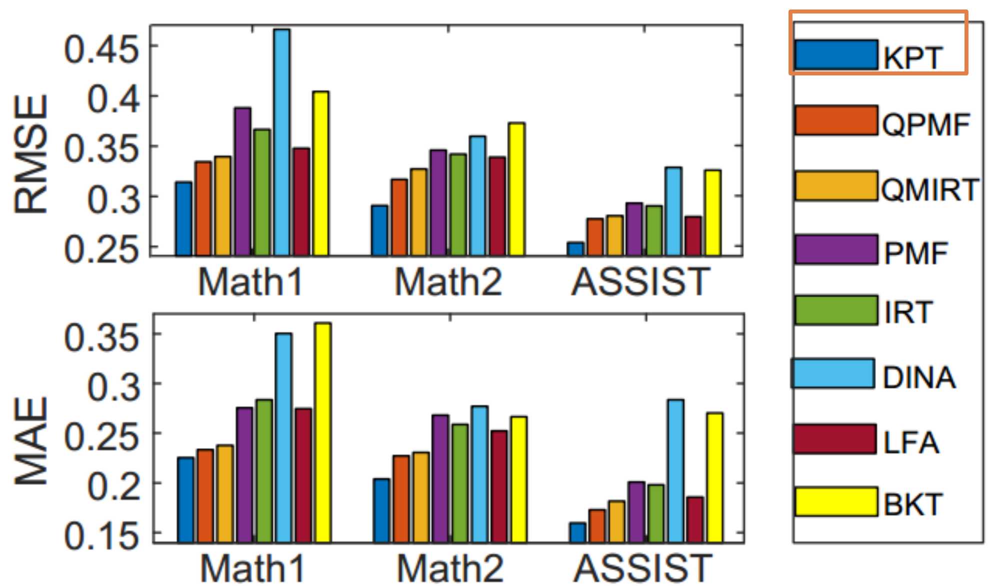
Dataset	Math1	Math2	ASSIST
Training scores logs	521,248	347,424	13,443
Testing scores logs	74,464	18,312	1,822
Students	9,308	1,306	215
Exercises	64	280	71
Time windows	4	10	4
Knowledge points	12	13	7
Average knowledge points of each exercise	1.15	1.3215	1.02

Evaluations

- Two evaluations:
  - evaluate on Students' Responses Prediction.
    - proved the **rationality of three priors for** prediction accuracy
  - evaluate on Knowledge Proficiency Diagnosis.
    - proved that the **effectiveness** of associating each exercise and student with **a knowledge vector** in the same knowledge space .

Evaluations on Students' Responses Prediction.

- Evaluation Metrics
  - For Scores prediction task performance
    - RMSE、 MAE
  - baselines:
    - *IRT*
    - *DINA*
    - *PMF*
    - *LFA*
    - *BKT*
    - *QMIRT* (MIRT+partial order)
    - *QPMF* (PMF+Partial order)



- KPT performs best on all three datasets.

Evaluations on Knowledge Proficiency Diagnosis

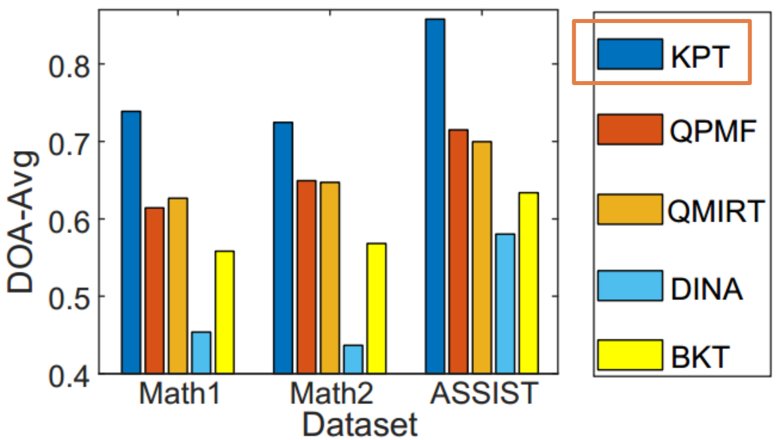
- For Knowledge Proficiency Diagnosis
  - DOA-of each specific knowledge point k

$$DOA(k) = \sum_{j=1}^{M} I\left(Q_{jk}=1\right) \sum_{u1=1}^{N} \sum_{u2=1}^{N} \frac{\delta\left(U_{u1k}^{T}-U_{u2k}^{T}\right) \cap \delta\left(R_{u2j}^{T}-R_{u1j}^{T}\right)}{\delta\left(U_{u1k}^{T}-U_{u1k}^{T}\right)},$$

  - DOA-average of all knowledge points

$$DOA-Avg = \frac{1}{K}\sum_{k=1}^{K} DOA(k)$$

- baselines:
  - *DINA*
  - *BKT*
  - *QMIRT* (MIRT+partial order)
  - *QPMF* (PMF+Partial order)



Evaluations on Knowledge Proficiency Diagnosis

- KPT performs best on KPD task for all knowledge points, followed by QPMF and QIRT, which indicates that the educational prior of Q-matrix does effectively.

(a) Math1

K	Baselines				
	KPT	QPMF	QMIRT	DINA	BKT
K1	**0.798**	0.565	0.595	0.524	0.558
K2	**0.733**	0.576	0.621	0.473	0.623
K3	**0.827**	0.614	0.629	0.497	0.523
K4	**0.752**	0.581	0.675	0.486	0.565
K5	**0.791**	0.559	0.723	0.476	0.578
K6	**0.838**	0.730	0.766	0.485	0.628
K7	**0.842**	0.697	0.634	0.520	0.697
K8	**0.784**	0.699	0.657	0.498	0.617
K9	**0.771**	0.609	0.712	0.501	0.645
K10	**0.834**	0.597	0.515	0.489	0.503
K11	**0.786**	0.608	0.631	0.478	0.617
K12	**0.842**	0.532	0.641	0.523	0.645

(b) Math2

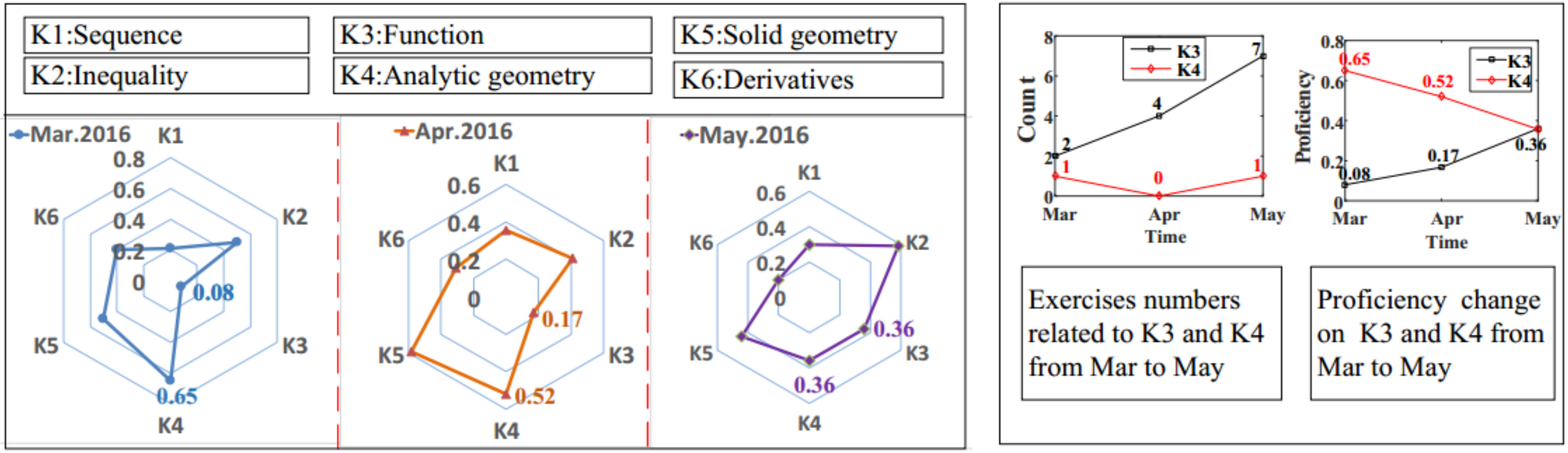
K	Baselines				
	KPT	QPMF	QMIRT	DINA	BKT
K1	**0.804**	0.743	0.754	0.517	0.568
K2	**0.757**	0.632	0.659	0.534	0.753
K3	**0.818**	0.761	0.723	0.510	0.669
K4	0.688	0.733	**0.734**	0.534	0.711
K5	**0.891**	0.703	0.668	0.474	0.553
K6	**0.699**	0.547	0.653	0.489	0.644
K7	**0.791**	0.677	0.722	0.483	0.730
K8	**0.726**	0.722	0.659	0.523	0.668
K9	**0.736**	0.558	0.541	0.507	0.567
K10	**0.652**	0.639	0.650	0.511	0.614
K11	**0.888**	0.836	0.692	0.522	0.630
K12	**0.798**	0.737	0.794	0.498	0.528
K13	**0.813**	0.797	0.804	0.453	0.633

(c) ASSIST

K	Baselines				
	KPT	QPMF	QMIRT	DINA	BKT
K1	**0.793**	0.747	0.716	0.605	0.592
K2	**0.823**	0.653	0.673	0.593	0.672
K3	**0.887**	0.852	0.671	0.631	0.577
K4	**0.792**	0.598	0.755	0.525	0.569
K5	**0.891**	0.576	0.672	0.511	0.624
K6	**0.871**	0.647	0.657	0.628	0.604
K7	**0.901**	0.793	0.654	0.573	0.796

Case study

- The diagnosis results of a student on six knowledge points at three particular time in Math2
- It clearly demonstrated the explanatory power of our proposed KPT model



(a) Diagnosis (b) Analysis

Outline

- Backgroud and Related Work
- Problem Statement
- Methodology
- Experiments
- **Conclusion**

Conclusion

- **Problem**: track students' knowledge proficiency mastery over time
- **Method**: probabilistic model with three educational priors
- **Contributions**:
  - We designed an *explanatory* probabilistic KPT model for solving the TKP task
  - We associated each exercise with a knowledge vector with the *Q-matrix* prior.
  - we embedded the *Learning curve* and *Forgetting curve* as priors to capture the change of each student's proficiency over time.

Future Work

- First, we will consider to combine more kinds' of users' behaviors (e.g., reading records) for the TKP task.

- Second, as students may learn difficult knowledge points (e.g., Function) after some basic ones (e.g., Set), it is interesting to take this kind of knowledge relationship into account for TKP

Acknowledgements

- We thanks for:
  - the SIGIR Travel Award
    - url: http://sigir.org/general-information/travel-grants/
  - the SIGWEB and US NSF Travel Award
    - url: https://cmt3.research.microsoft.com/CIKMTA2017

Q & A

Thanks !

Reporter: Yuying Chen

cyy33222@mail.ustc.edu.cn