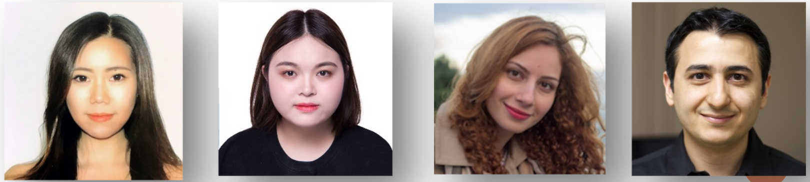
Relaxed Clustered Hawkes Process for Procrastination Modeling in MOOCs

Mengfan (Miley) Yao, Siqian Zhao, Shaghayegh (Sherry) Sahebi, Reza Feyzi Behnagh

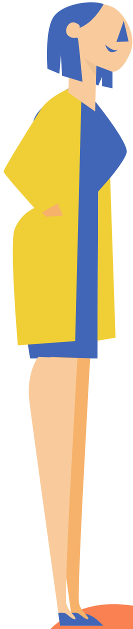


UNIVERSITY AT ALBANY
State University of New York

PersAI
onalized

NSF

Content



01

Introduction

1.1 procrastination modeling in MOOCs

MOOCs?
= Massive Open Online Courses

Procrastination?
Voluntary delay (≈ cramming behaviors)

Why?
Lack of time management, little self-regulation,…

What?
Characteristic behaviors; clusters

How?
Static measures (e.g. avg. time)

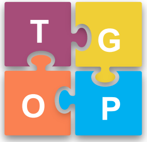
why?
Bad & prevalent
Detect & predict
regulate & prevent

1.2.1 problem

Need a procrastination model that:
- Can model temporal aspects (aot. static)
- Personalized (aot. full set of parameters)

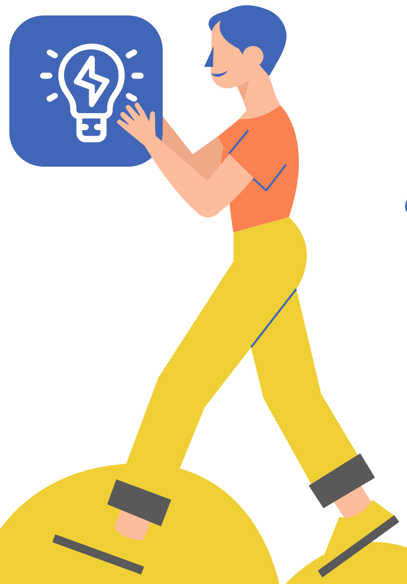
When:
- Observation is sparse
- students' group structure

1.2.2 our solution



A **personalized stochastic model** that discovers **student behavior clusters**, that can predict future activities **with missing and partial data, but without auxiliary features**

02
Model

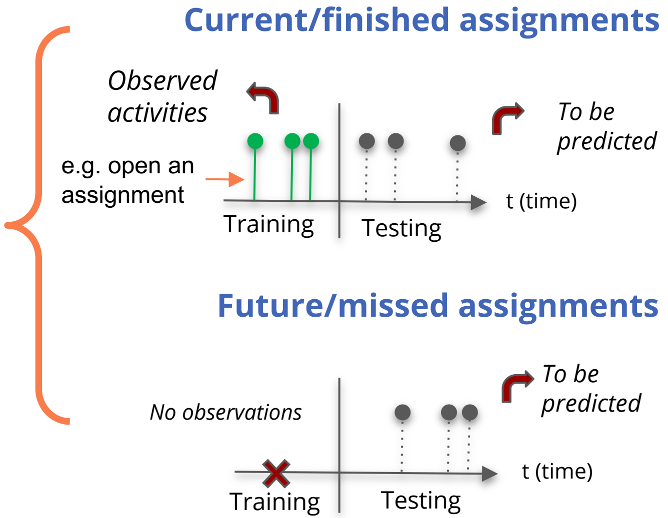


2.1 problem formulation

Consider a MOOC

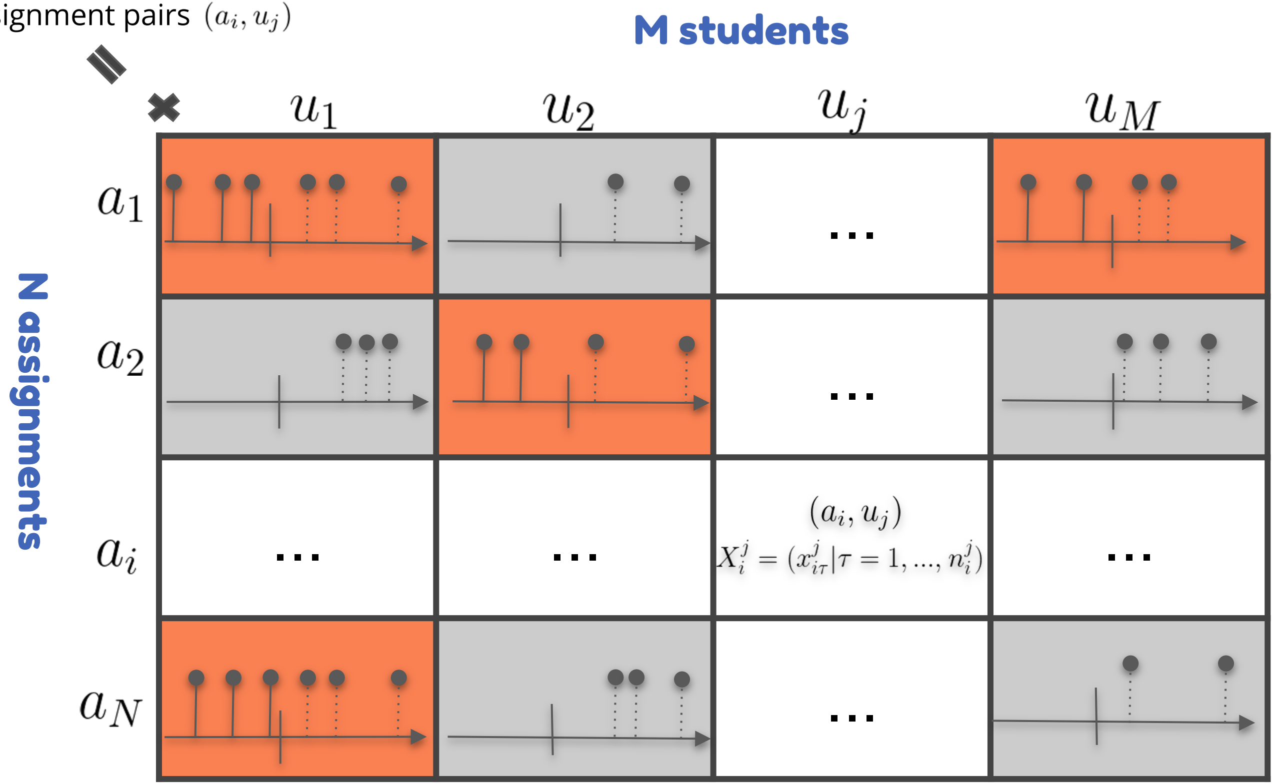
M students

N assignments



2.1 problem formulation (cont.)

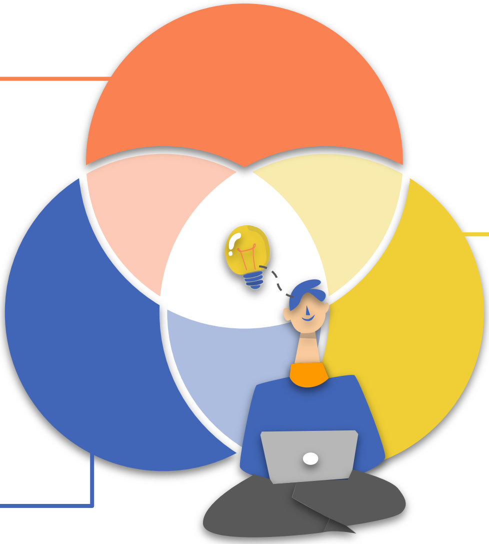
NxM student-assignment pairs (a_i, u_j)



2.2 RCHawkes-Gamma

2.2.1 Hawkes model

T *model temporal aspects*



2.2.3 Mixture Gamma

Robustness ✓ *interpretability* ✓

2.2.2 Joint relaxed Clustering

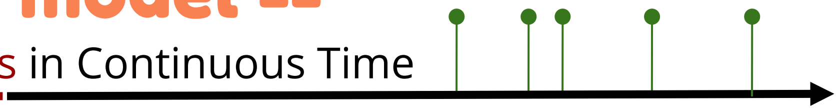
P *Personalized*

O *Observation is sparse*

G *students' group structure*

2.2.1 Hawkes model --

Model Discrete Events in Continuous Time

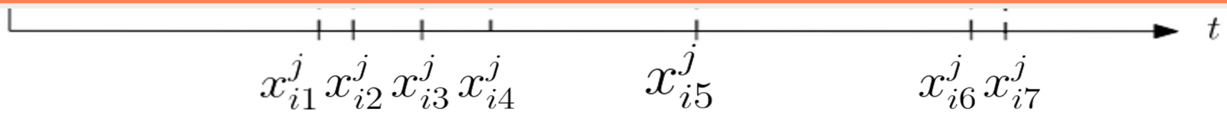


model temporal aspects

For (a_i, u_j), observe its time sequence: $X_i^j = (x_{i\tau}^j | \tau = 1, \ldots, n_i^j)$

Define its intensity function $\lambda(t)_{ij} = U_{ij} + A_{ij}\beta \sum^{n_{ij}} \exp(-\beta(t - x_{i,\tau}^j))$

But what if there is no observed activity for a sequence? And how to model more than one sequences?



2.2.2 Joint relaxed Clustering

P *Personalized*

O *Observation is sparse*

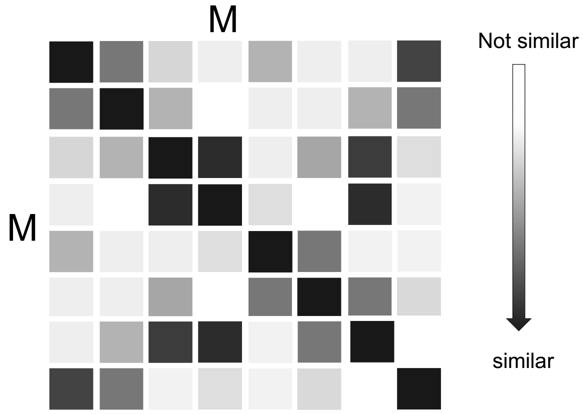
G *students' group structure*

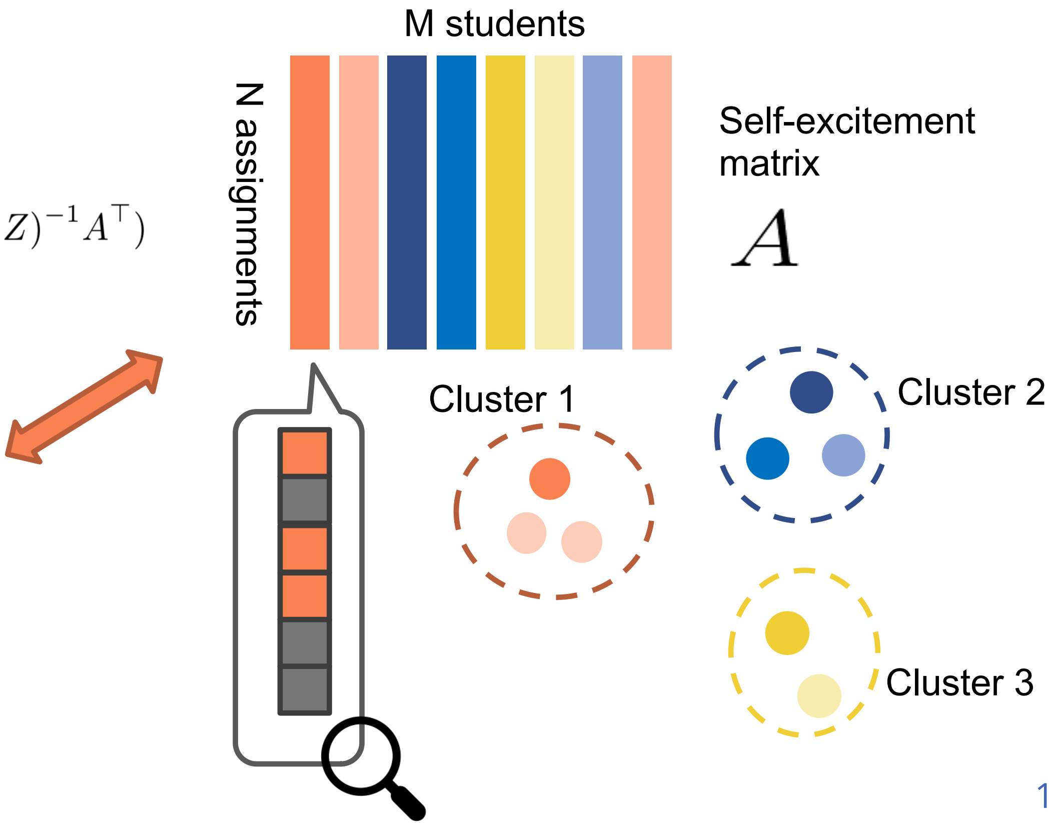
Convex relaxation:

$$\min \mathcal{L}_c(A, Z) = \min \frac{\rho_2(\rho_2 + \rho_1)}{\rho_1} tr(A(\frac{\rho_1}{\rho_2} I + Z)^{-1} A^\top)$$

$$\text{s.t. } tr(Z) = k, Z \preceq I, Z \in S_+^M.$$

Learnable Z





2.2.3 Mixture Gamma prior

Robustness ☑

interpretability ☑

The prior when student i is in m-th cluster

$$p(A_{ij};\Theta_m) = \frac{1}{\Gamma(s_m)\theta_m^{s_m}} A_{ij}^{s_m-1} \exp(-\frac{A_{ij}}{\theta_m^{s_m}})$$

$$\Theta_m = (s_m, k_m)$$

Shape
(pesudo base rate)

Scale
(pesudo #. Events
triggered internally)

Loss:

$$\mathcal{L}_g = \log p(A;\Theta_1,..,\Theta_k)$$

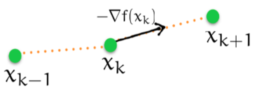
$$= \sum_{X_i^j \in \mathcal{O}} \left[\log \sum_{m=1}^{k} \frac{1}{k} \frac{1}{\Gamma(s_m)\theta_m^{s_m}} A_{ij}^{s_m-1} \exp(-\frac{A_{ij}}{\theta_m^{k_m}})\right]$$

2.2.4 Optimization

$$\begin{aligned}\min_{A_z,U_z,Z_z} \; & -L_{mle} - L_c - L_g \\ & + \|A_z - A_s\|_F^2 + \|U_z - U_s\|_F^2 + \|Z_z - Z_s\|_F^2, \\ \text{s.t. } & tr(Z_z) = k, tr(A_z) \leq c, A_z \geq 0, U_z \geq 0, \\ & Z_z \leq I, Z_z \in S_+^M\end{aligned}$$

Gradient Descent

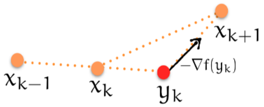
$$x_{k+1} = x_k - \epsilon \nabla f(x_k)$$



Accelerated GD

$$x_{k+1} = y_k - \epsilon \nabla f(y_k)$$

$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$$



when $\|\nabla^2 f\| \leq \frac{1}{\epsilon}$

$O\left(\frac{1}{\epsilon k}\right) \xrightarrow{\text{accelerated!}} O\left(\frac{1}{\epsilon k^2}\right)$ optimal rate Proof

source

03

Experiments

3. Experiments

3.1 Datasets

3.2 Baselines

3.3 Evaluation

3.4 Student Procrastination Analysis

3.1 datasets

- **Synthetic**
  - **5400 s-a pairs**
  - **~810 K activities**
    - 10% missing
    - 30% missing
    - 50% missing
    - 70% missing
- **real-world**
  - **CANVAS**
    - 471 students, 6 assignments, ~740K activities
  - **MORF**
    - 675 students, 8 assignments, ~102K activities

3.2 baseline approaches

Application	Model	Infer future of unseen Data	Require No External Features	Capture Temporal dependency
EDM	RCHawkes-Gamma	✓	✓	✓
EDM	RCHawkes	✓	✓	✓
EDM	EdMHP (Yao et al, 2020)	x	x	✓
Rec-Sys	HPLR (Du et al, 2015)	✓	✓	✓
ATM	ERPP (Xiao et al, 2017)	x	x	✓
Rec-Sys	HRPF (Hosseini, 2017)	✓	✓	x
Rec-Sys	DRPF (Hosseini, 2017)	✓	✓	x
Others	RMTPP (Du et al, 2016)	x	x	✓

3.3. Evaluation

Parameter Estimation

RMSE of estimated parameters

The lower the better

Synthetic ✓ real-world ✗

Time prediction

RMSE of predicted next arrival time

The lower the better

Synthetic ✓ real-world ✓

Cluster recovery

Recovered correlation matrix

The higher the within-group correlation and the lower the between-group correlation the better

Synthetic ✓ real-world ✗

Parameter Estimation

Synthetic ✓ real-world ✗

Infer unseen data ✓

Table 2: RMSE (±standard deviation) of $\hat{A}$ and $\hat{U}$ on seen and unseen data, with various missing data ratios (r)

	Model	r = 0.1		r = 0.3		r = 0.5		r = 0.7	
		seen	unseen	seen	unseen	seen	unseen	seen	unseen
RMSE for A	RCHawkes-Gamma	**0.094±0.024**	**0.102±0.037**	0.121±0.017	**0.114±0.056**	0.141±0.033	0.139±0.033	**0.136±0.077**	**0.137±0.052**
	RCHawkes	0.108±0.017	0.108±0.054	**0.115±0.024**	0.116±0.039	**0.126±0.033**	**0.136±0.033**	0.180±0.072	0.170±0.048
	HPLR	0.631±0.110	0.663±0.331	0.645±0.141	0.607±0.216	0.635±0.133	0.633±0.133	0.634±0.304	0.634±0.204
	HRPF	0.664±0.769	0.664±0.769	0.664±0.770	0.664±0.770	0.663±0.769	0.663±0.770	0.664±0.769	0.664±0.767
	DRPF	0.474±0.461	0.474±0.461	0.479±0.465	0.479±0.465	0.473±0.462	0.473±0.462	0.474±0.463	0.474±0.463
RMSE for U	RCHawkes-Gamma	0.075±0.022	0.085±0.036	**0.069±0.017**	**0.060±0.050**	**0.062±0.030**	**0.064±0.030**	0.071±0.039	0.075±0.026
	RCHawkes	0.074±0.020	0.089±0.061	0.074±0.020	0.075±0.032	0.077±0.030	0.079±0.030	**0.069±0.026**	**0.062±0.017**
	HPLR	0.110±0.082	**0.078±0.047**	0.081±0.060	0.078±0.094	0.091±0.035	0.091±0.035	0.090±0.096	0.095±0.065
	HRPF	0.105±0.055	0.311±0.055	0.119±0.068	0.183±0.068	0.141±0.071	0.142±0.071	0.179±0.068	0.120±0.070
	DRPF	**0.062±0.052**	0.300±0.035	0.088±0.049	0.165±0.045	0.121±0.051	0.121±0.050	0.167±0.053	0.102±0.054

- Lower RMSE
- Less sensitive to miss ratio

Cluster recovery

Synthetic

real-world

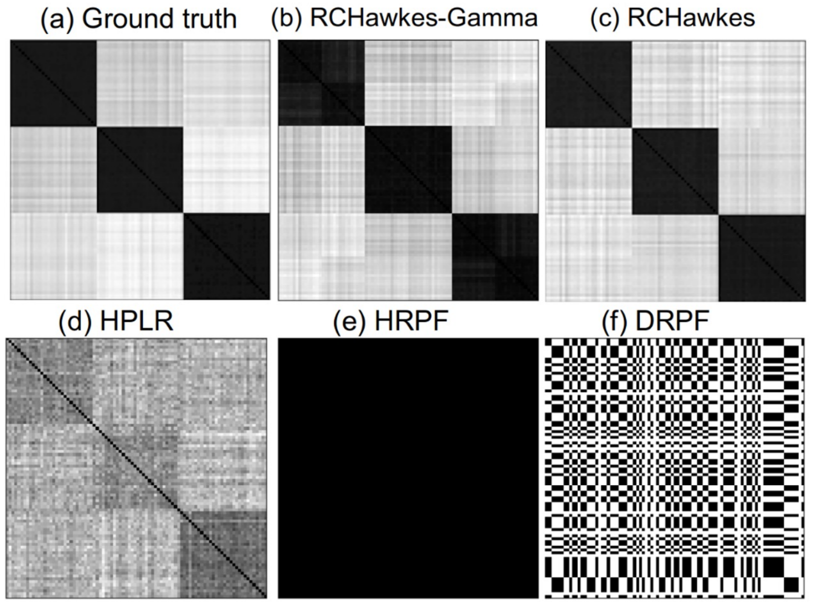


Figure 3: The ground truth of A's correlation matrix (a), and the estimated $\hat{A}$'s correlation matrix learned by each model.

- Successfully recover correlation matrix
- The closest to ground truth

Time prediction

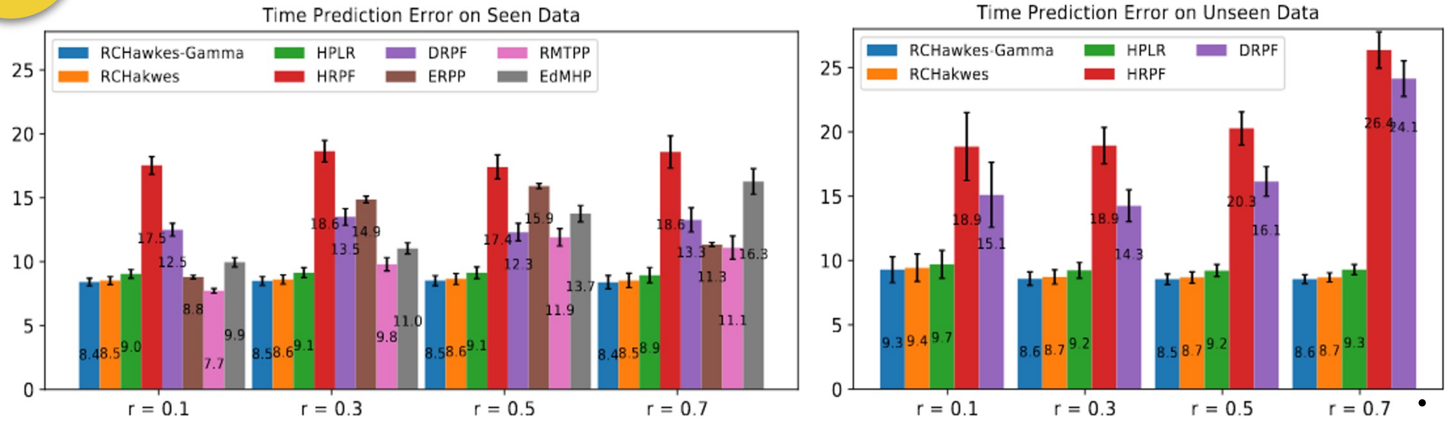


Figure 1: Time prediction error and 95% confidence interval on synthetic datasets with varying data missing ratios (r)

- Smaller RMSE in all settings
- More robust to large miss ratio

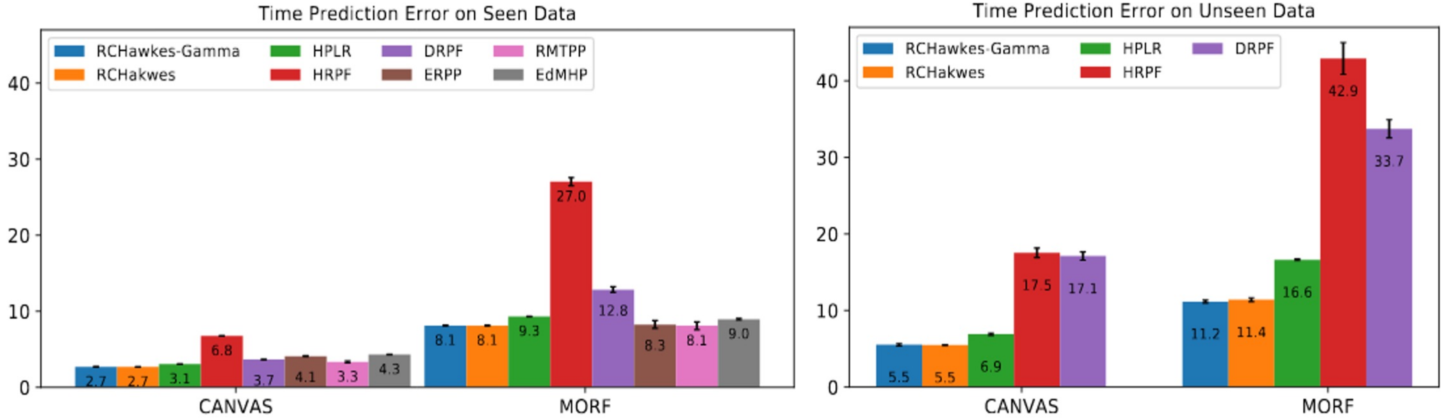


Figure 2: Time prediction error on seen and unseen data with 95% confidence interval on real-world datasets

3.4 Student Procrastination Analysis

Correlation Analysis

		A	U	delay
CANVAS	A	1		
	U	0.284***	1	
	delay	0.345***	0.144***	1
MORF	A	1		
	U	0.243***	1	
	delay	0.264***	0.412***	1

- All correlations are statistically significant with p-value<0.05
- All correlations are positive (e.g. higher delay and higher self-excitement)

3.4 Student Procrastination Analysis

Clustering Analysis - CANVAS

Assign. #.	cluster 1	cluster 2	cluster 3	cluster 4	p-value
size	81	144	207	39	-
1	0.3335	0.4583	0.6108	0.9064	1.34E-16***
2	0.6245	0.5788	0.8476	1.0854	3.59E-09***
3	0.6911	0.7143	0.8633	0.9655	4.36E-05***
4	0.6050	0.6958	0.8515	1.0717	0.0008***
5	0.5969	0.7080	0.9084	1.1217	0.0195*
6	0.5351	0.7647	0.9002	1.0970	0.0149*

Cluster 1: small, increasing then decreasing delay
Cluster 2: higher, increasing delay
Cluster 3: high, increasing delay
Cluster 4: high, consistent delay

Clustering Analysis - MORF

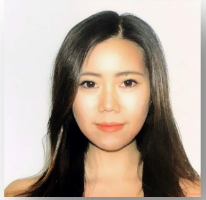
Assign. #.	cluster 1	cluster 2	cluster 3	p-value
size	573	34	68	-
1	0.4991	0.6710	0.4477	2.30E-09***
2	0.5120	0.7288	0.4855	1.90E-08***
3	0.5570	0.6904	0.6105	7.50E-05***
4	0.4699	0.6122	0.5360	0.0004***
5	0.5626	0.6358	0.6308	0.0070***
6	0.5329	0.6236	0.6642	8.56E-06***
7	0.4325	0.5598	0.7672	2.12E-20***
8	0.3974	0.5172	0.7629	3.84E-27***

Cluster 1: small, consistent delay
Cluster 2: higher, decreasing delay
Cluster 3: small, increasing delay

4. Conclusion

1. Provide our solution to the problem of modeling procrastination in MOOCs.
2. Proposed a student procrastination model: **personalized** ; capture group structure, predict future for unseen, without auxiliary features.
3. Provide a novel framework for Hawkes process modeling that jointly learn all sequences.
4. Achieve a better performance than state-of-the-art in synthetic and real datasets
5. Demonstrate the identified clusters are meaningful representations of procrastination.

Thanks!




myao@albany.edu szhao2@albany.edu ssahebi@albany.edu rfeyzibehnagh@albany.edu

This paper is based upon work supported by the National Science Foundation under Grant Number 1917949.

PersAI onalized https://persai-lab.github.io/

Questions?