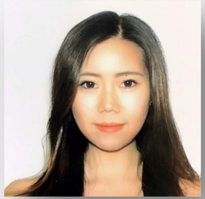
Stimuli-Sensitive Hawkes Processes for Personalized Student Procrastination Modeling



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

Start now!

1.Introduction

2.Model

3.Experiments

4.Conclusion

1.1 Procrastination Modeling in MOOCs

MOOCs?

= Massive Open Online Courses

Procrastination?

Voluntary delay (≈ cramming behaviors)

Why Important?

Bad & prevalent
Detect & predict
regulate & prevent

1.2.1 Limitations

- Static methods
- Not personalized
- discarded missing sequences
- failed to capture an important factor: triggers of procrastination

1.2.2 Proposed solutions

- We propose a novel temporal point process model by collaboratively modelling all student sequences together including the missing ones, which captures 3 dynamic types of external triggering stimuli of procrastination.

2.1.Formulation

2.2. Intensity Function

2.3.Objective

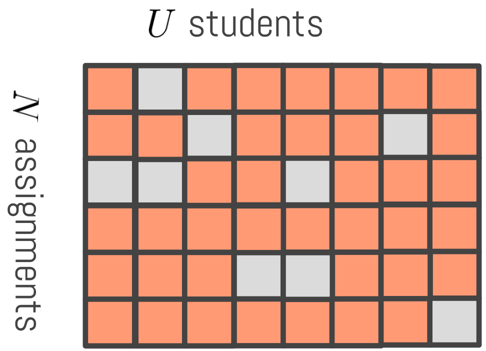
2.4.Optimization

2. Model

In this section, we introduce our model: Stimuli-Sensitive Hawkes Process model (SSHP).

2.1. Formulation

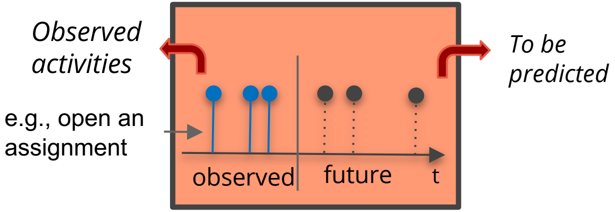
NxU student-assignment pairs



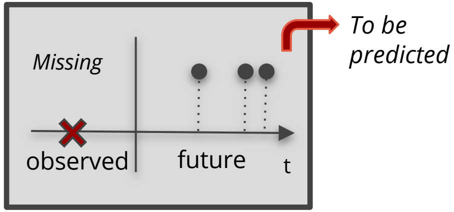
(u_i, a_j) $Xij = \{x_{ij}^{\tau} | \tau = 1, ..., K_{ij}\}$

Step index

Current/finished assignments

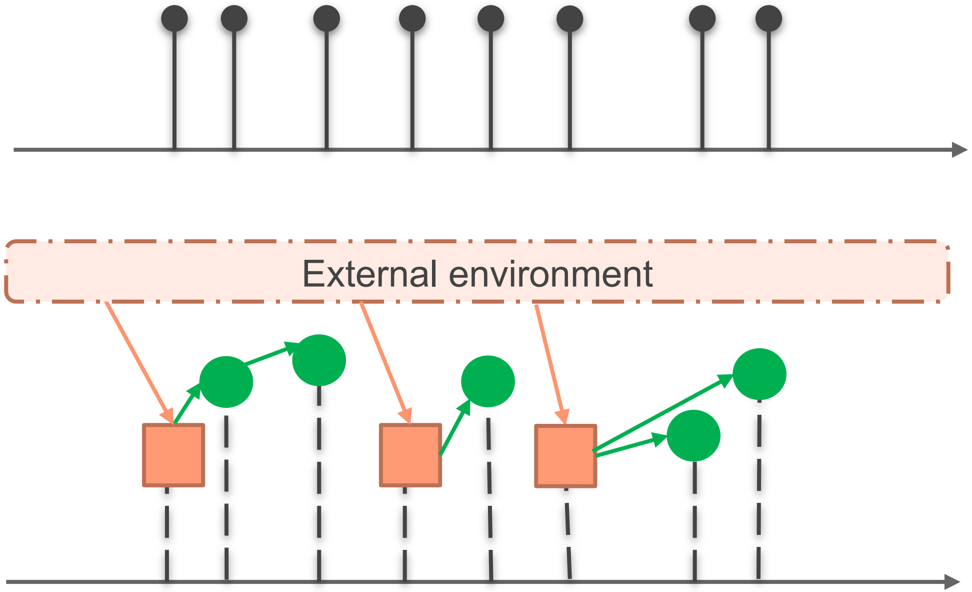


Future/missed assignments



2.1. Formulation (cont.)

- Furthermore, we assume that there are two types of triggers: internal stimuli ( ) and external ones ( ).



2.2. Intensity Function

- Typically, a point process is defined by the intensity function that describes the number of activities as a function of time, conditioning on the observed history.
- In this work we parameterize each student-assignment pair's intensity via the following:

$$
\begin{aligned}
\lambda(t) &= \mu(t) + s(t) \\
&= \underbrace{\gamma^d \mu^d(t)}_{\text{deadline}} + \underbrace{\gamma^o \mu^o(t)}_{\text{opening}} + \underbrace{\gamma^h \mu^h(t)}_{\text{habit}} + s(t) \\
&= \gamma^h\left(\sin\left(\frac{2\pi}{s}(t+p)\right) + c\right) + \gamma^o b^{t/s} + \gamma^d\left(\frac{1}{\sqrt{2\pi v}(d-m-t/s)} e^{-\frac{(\ln(d-m-t/s))^2}{v}}\right) + \sum_{x^\tau < t} \alpha\beta e^{-\beta(t-x^\tau)}.
\end{aligned}
$$

2.3. Objective Function

- MLE of observing the history of *one* student-assignment pair:

$$l(X;\theta) = \log L(\theta) = \sum_{\tau=1}^{K} \log(\lambda(x^{\tau}) - \int_0^{x^K} \lambda(u)du$$

- Total Loss of observing *all* sequences in observed sequence set $\mathcal{O}$:

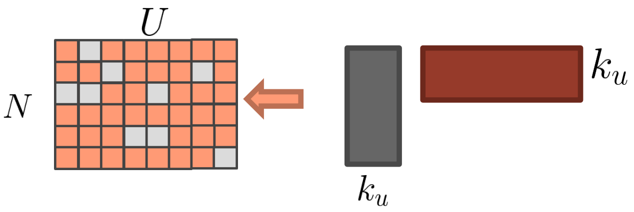
$$\min_{\Theta,\phi} \mathcal{L} = -\frac{1}{|\mathcal{O}|} \sum_{X_{ij}\in\mathcal{O}} l(X_{ij};\Theta_{ij},\phi_i)$$

$$\text{s.t. } \mathbf{A} \geq 0, \mathbf{\Gamma}_d \geq 0, \mathbf{\Gamma}_o \geq 0, \mathbf{\Gamma}_h \geq 0, \mathbf{c} \geq 1, \mathbf{v} > 0, 1 > \mathbf{b} > 0$$

$$tr(\theta_u) \leq k_u, \text{ for } \theta_u \in \Theta_{ij}.$$

$$\Theta = (A, M, \Gamma^h, \Gamma^o, \Gamma^d)$$

Low rank: analytically solved by SVD



$$\phi = \{\mathbf{c}, \mathbf{p}, \mathbf{b}, \mathbf{v}\}$$ Shared across assignments

2.4. Optimization

- We adopt Accelerated Gradient Method (AGM) framework for the inference of parameters

To compute the proximal operator with matrix format:

$$\Theta = (A, M, \Gamma^h, \Gamma^o, \Gamma^d)$$

$$\begin{aligned}\theta_u^* &= \mathrm{argmin}_{\theta_u} \mathcal{M}_{\gamma,\theta_u^S}(\theta_u) \\ &= \mathrm{argmin}_{\theta_u} \frac{\gamma}{2} \|\theta_u - P_{\theta_u}(\theta_u^S - \frac{1}{\gamma} \nabla_{\theta_u} \mathcal{L})\|_F^2.\end{aligned}$$

To compute the proximal operator with vector format:

$$\phi = \{\mathbf{c}, \mathbf{p}, \mathbf{b}, \mathbf{v}\}$$

$$\begin{aligned}\phi_u^* &= \mathrm{argmin}_{\phi_u} \mathcal{M}_{\phi_u^S,\gamma}(\phi_u) \\ &= \mathrm{argmin}_{\phi_u} \frac{\gamma}{2} \|\phi_u - P_{\phi_u}(\phi_u^S - \frac{1}{\gamma} \nabla_{\phi_u} \mathcal{L})\|_F^2.\end{aligned}$$

3. Experiments

3.1. Baselines & Datasets

3.2. Experiment Setup

3.3. Evaluation of SSHP

3.4. Ablation Study

3.5. Procrastination Analysis

3.1.1. Baselines

We consider the following baselines from these 4 aspects:

Model	Self-exciting	Non-constant base of time	Infer completely missing seq.	Application in Education
Poisson	✗	✗	✗	✗
HRPF	✗	✗	✓	✗
RMTPP	✓	✓	✗	✗
ERPP	✓	✓	✗	✗
DHPR	✓	✗	✓	✗
HPLR	✓	✗	✓	✗
EdMPH	✓	✗	✗	✓
SSHP	✓	✓	✓	✓

3.1.2. Datasets

Synthetic datasets

We simulated 500 students 20 assignments and sampled ~100 activities per student-assignment pair via Ogata Thining. Then we created:

Syn-10

Randomly selected **10%** pairs and their activities to be entirele missing

Syn-90

Randomly selected **90%** pairs and their activities to be entirele missing

Real-world datasets

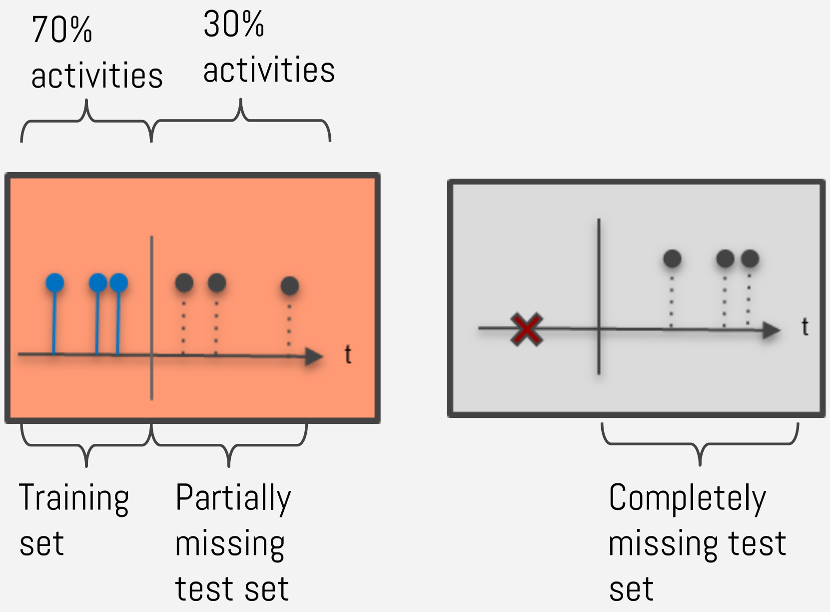
CANVAS

It is from the Canvas Network online platform; we extracted **~729K** timestamps from **384** and **6** graded quiz-style assignment

MORF

It is collected from an 8-week Big Data in Education course on the Coursera platform; we extracted **~52K** timestamps from **246** students and **8** assignments.

3.2. Experiment Setup



Train Test Ratio

Syn-10

90% : 10%

Syn-90

10% : 90%

Real:

80% : 20%

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

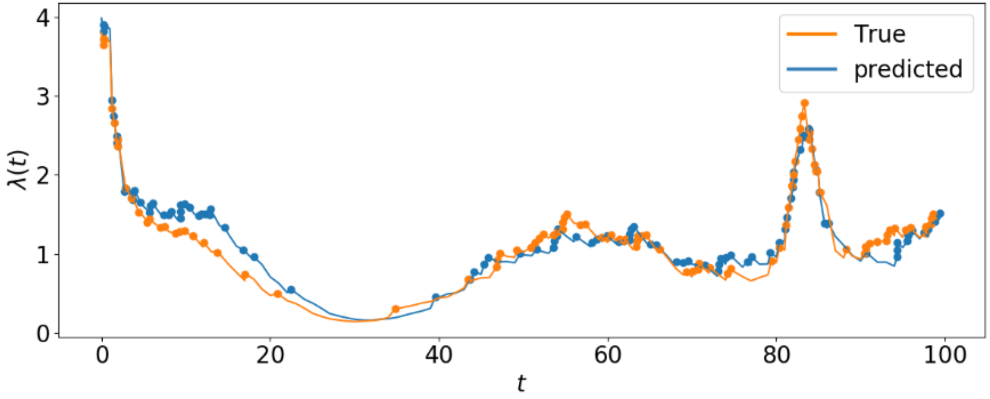
Synthetic ✓ Real-world ✓

3.2 SSHP Evaluation - Parameter Estimation

- RMSE of parameters learned by SSHP in synthetic datasets.

Datasets		v	b	p	c	A	M	Γ^d	Γ^o	Γ^h
Syn-10	part. miss.	1.33	0.1	1.33	0.09	0.05	2.64	1.65	1.08	0.16
	compl. miss.	1.23	0.12	1.39	0.16	0.13	2.60	2	1.54	0.13
Syn-90	part. miss.	1.34	0.10	1.33	0.09	0.06	2.39	1.80	1.14	0.18
	compl. miss.	1.31	0.12	1.38	0.16	0.12	2.61	1.97	1.51	0.17

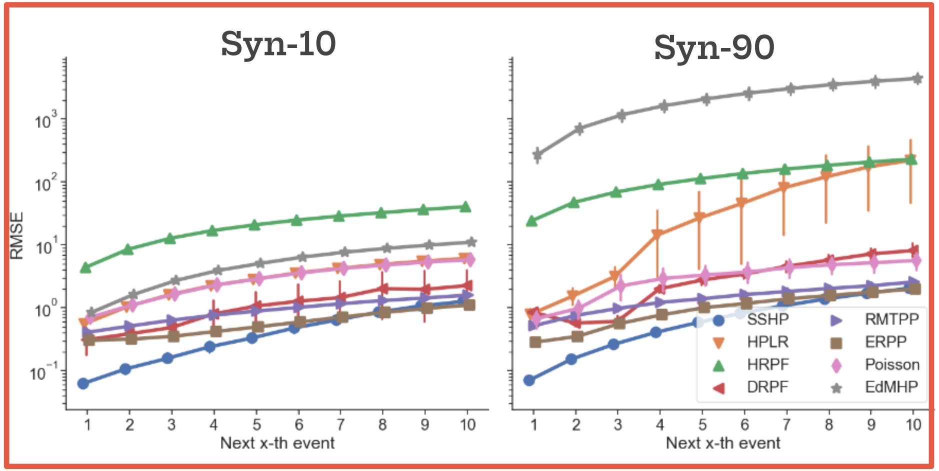
- Visualization of the predicted Intensity of a synthetic sequence.



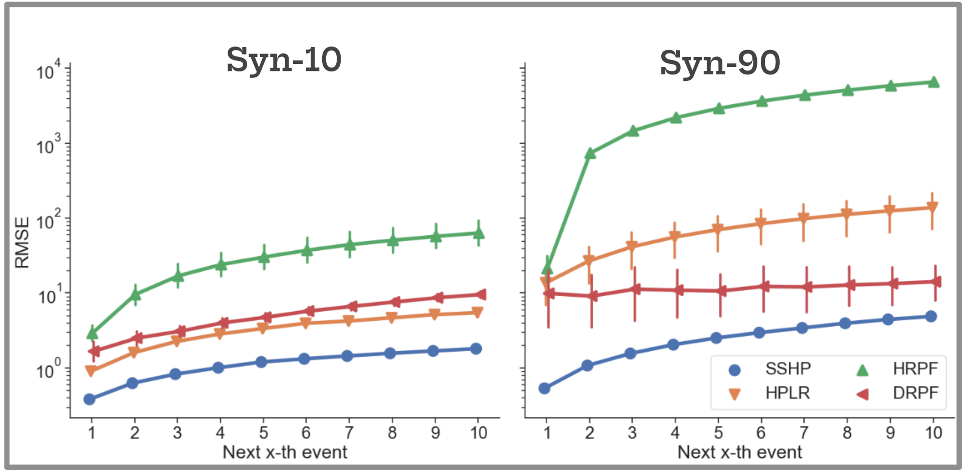
- Marginally higher RMSE in Syn-90 than Syn-10: **robust to missing ratio**
- The figure demonstrates model's ability in **accurately** capturing the dynamics of the sequence.

3.2. SSHP Evaluation - Time prediction (synthetic)

Partially missing test set



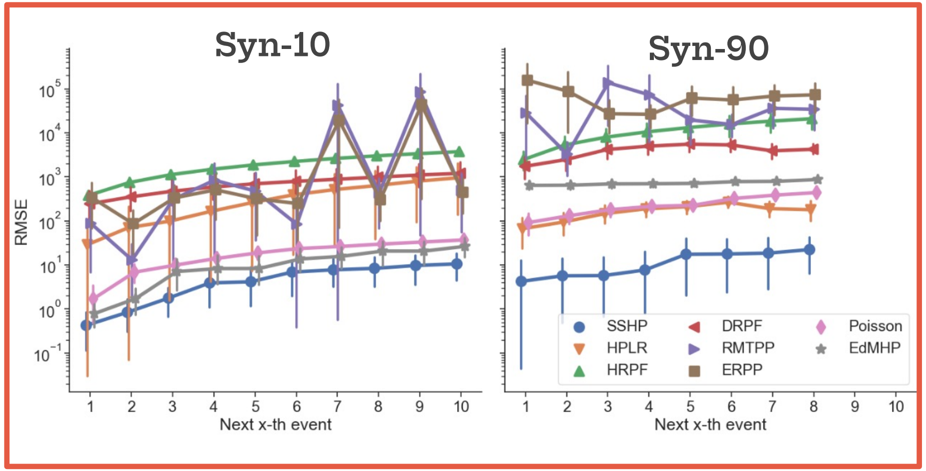
Completely missing test set



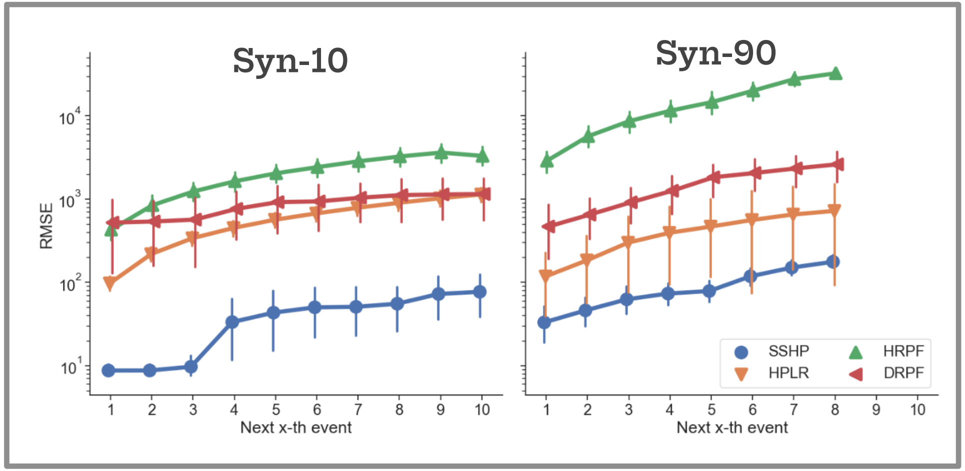
- Smaller RMSE than baselines in all settings
- Smaller RMSE increase in Syn-90 comparing with Syn-10
- Similarly, smaller RMSE increases in completely missing test set

3.2. SSHP Evaluation –Time prediction (real)

Partially missing test set



Completely missing test set

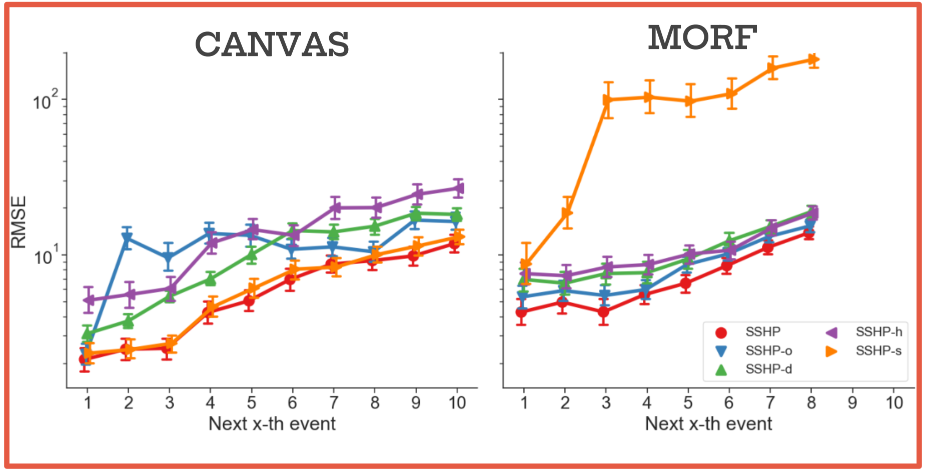


- Consistent performance of SSHP with Syn-10 and Syn-90: i.e. smaller RMSE, more robust to missing ratio and missing history.

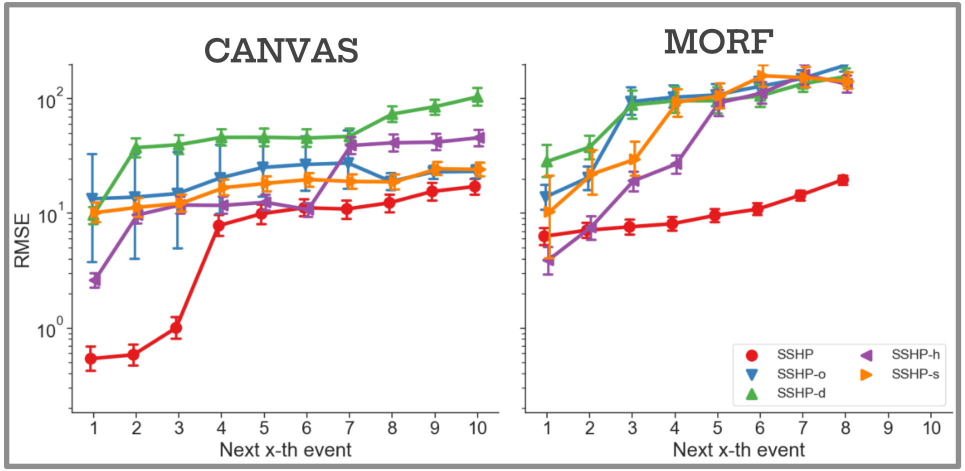
3.3. Ablation Study

- To verify each component's importance in the intensity function, we compare SSHP to its variations SSHP-s (internal self-excitement), SSHP-o (assignment opening), SSHP-h (habit) and SSHP-d (deadline).

Partially missing test set



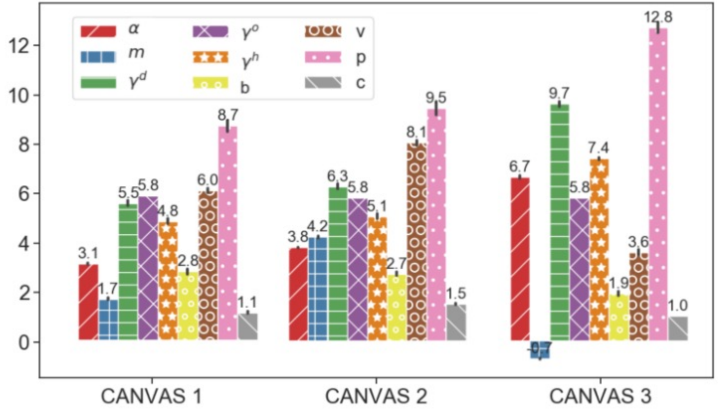
Completely missing test set

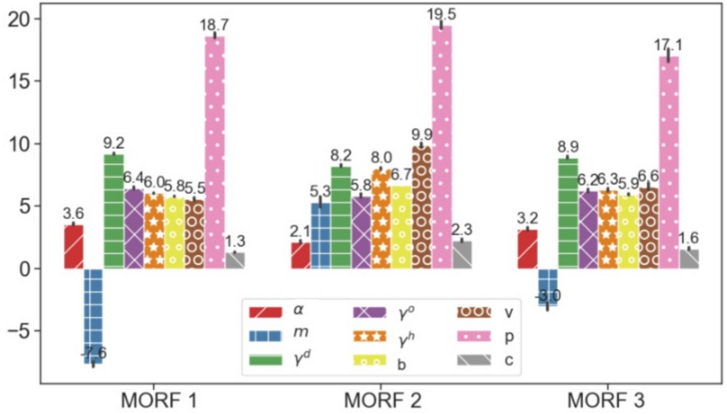


- Smaller RMSE of SSHP: importance of modeling each type of stimuli
- In partially missing set: internal triggering is more important in MORF (more bursty)
- In completely missing set: deadline is the most important factor, especially in CANVAS

3.4. Procrastination Analysis – clustering analysis

- To see if the learned parameters can describe students' cramming and procrastination behaviors, we cluster all student-assignment pairs via K-Means (via elbow method).





Normal paced (somewhere in between)

Early birds:
- Finish earlier, more sensitive the assignment opening
- less bursty behaviors

Procrastinating-like behaviors:
- less sensitivity to the deadline and the assignment opening
- more bursty and intense behaviors

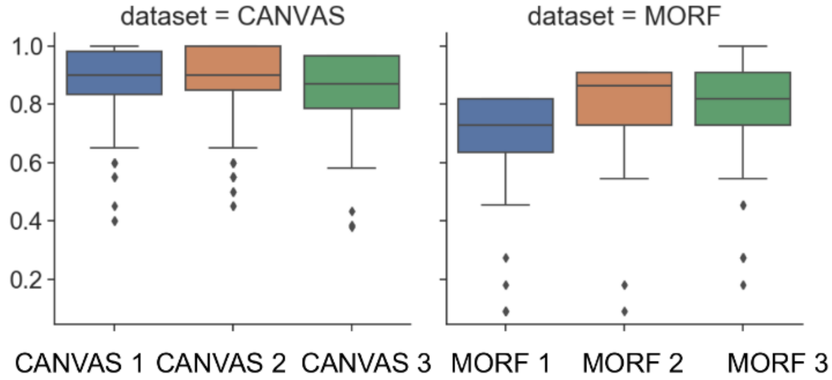
Extreme procrastinating -like behaviors

Early birds

Procrastinating -like behaviors

3.4. Procrastination Analysis – association with grades

- We check the distribution of grades of these clusters and run Kruskal-Wallis to check the significance of differences between each pairs of clusters.



- P-values of KS <<0.05: significant differences in grade distributions among clusters
- Procrastination like behaviors (MORF 1 and CANVAS 3) are associated with lower grades
- SSHP captures meaningful underlying procrastination patterns and performances

4.Conclusions

We propose our novel Stimuli-sensitive Hawkes Process model that **address the limitations** of literature (static, not personalized, discards missing sequences, not sensitive to external stimuli).

SSHP achieves **better performances** than state-of-the-art in synthetic and real datasets

SSHP successfully captures **3 types of procrastination external stimuli** (deadline, assignment opening and student habits), and each of them has shown to be **important** in ablation study.

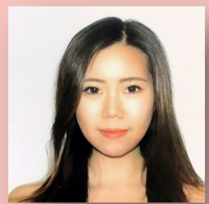
We provide **an effective solution** to the problem of procrastination modeling, and identify **meaningful** types of procrastination behaviors **associated with grades** (low-performing extreme procrastinators)

Thank you!

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

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

Feel free to let us know if you have questions!



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