Technical University of Munich
Department of Informatics
Data Analytics and Machine Learning

Uncertainty on Asynchronous Time Event Prediction

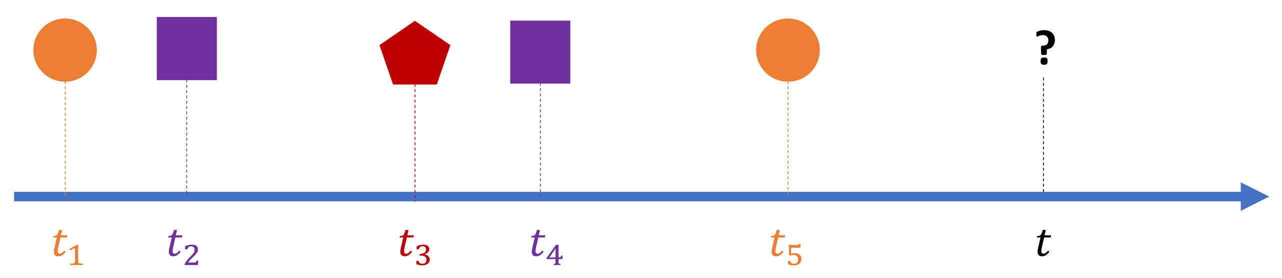
Marin Biloš*, Bertrand Charpentier*, Stephan Günnemann

tl;dr

- ❑ **Adaptive prediction** for the next event
- ❑ Real time **uncertainty estimation** in asynchronous event prediction

Asynchronous events

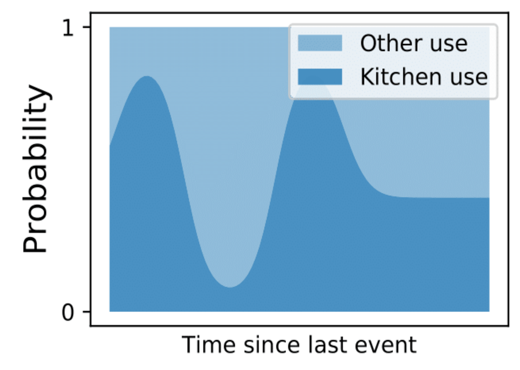
- ❑ Discrete events occurring irregularly over time
- ❑ At a given time, the most probable event can change



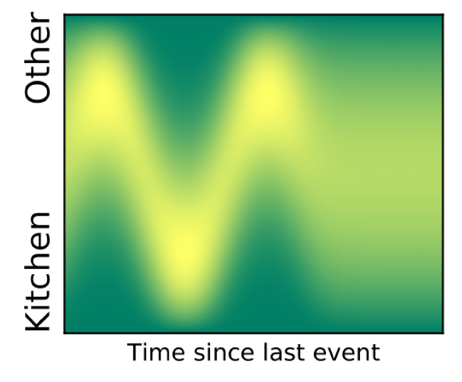
Research questions

(1) Prediction: At a given time, what event is likely to occur?

(2) Uncertainty estimation: How sure are we in our prediction?



(1) Prediction

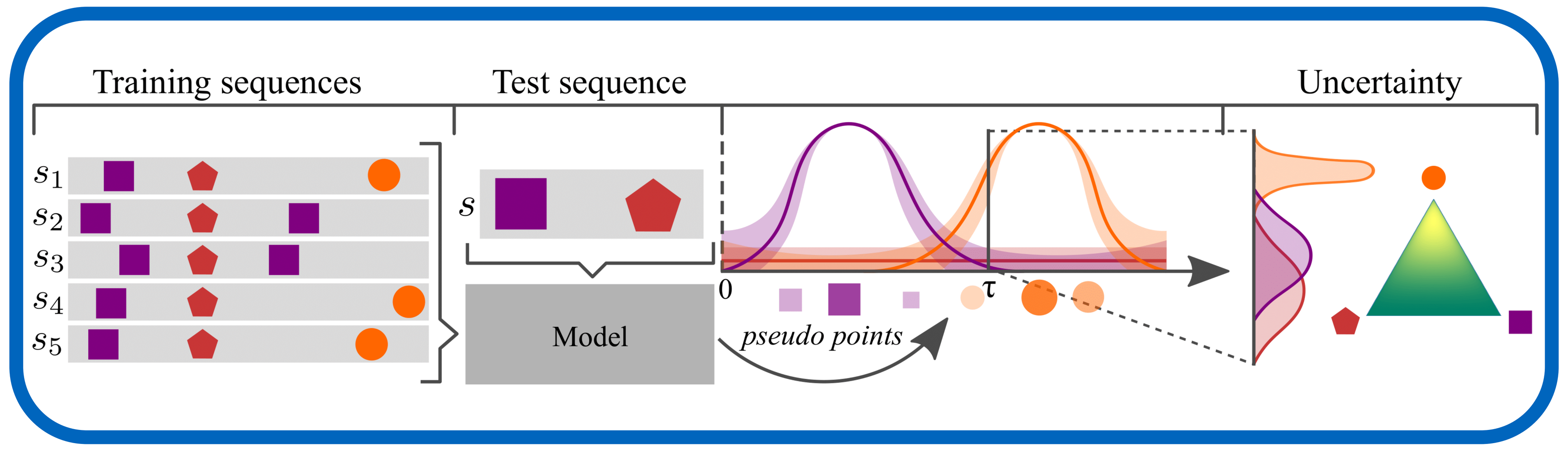


(2) Uncertainty estimation

Model overview

1. For each class c, generate M pseudo points $(w_j^{(c)}, \tau_j^{(c)}, y_j^{(c)})$ based on a hidden state of an RNN whose input is a sequence.

2. Fit a complex evolution of the predicted categorical distribution $\boldsymbol{p}(\tau)\sim P(\theta(\tau))$. The distribution over the simplex $P(\theta(\tau))$ gives certainty localized around pseudo points.



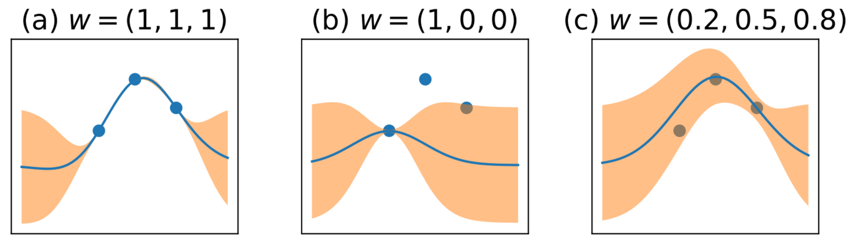
*Equal contribution

Model 1: Logistic Normal via Weighted Gaussian Process

- ❑ The pseudo points $(\tau_j^{(c)}, y_j^{(c)})$ weighted with $w_j^{(c)}$ are used to generate a Weighted Gaussian Process in the logit-space. The weigths $w_j^{(c)}$ allow to discard unecessary training points in Gaussian Process.

(Kernel)

$$k'(\tau_1,\tau_2) = min(w_1,w_2)k(\tau_1,\tau_2)$$



- ❑ The distribution $\boldsymbol{p}_c(\tau)\sim logistic\ N(\tau_c(\tau), \sigma_c^2(\tau))$ follows a Logistic Normal distribution.

Model 2: Dirichlet via Function Decomposition

- ❑ The pseudo points $(w_j^{(c)}, \tau_j^{(c)}, \sigma_j^{(c)})$ are used to describe the evolution of the Dirichlet parameters over time via function decomposition.

$$\underbrace{\log\alpha_c(\tau)}_{\text{Effective number of observations of class c at time }\tau} = \underbrace{\sum_{j=1}^{M} w_j^{(c)}\cdot N(\tau|\tau_j^{(c)}, \sigma_j^{(c)})}_{\text{Function decomposition}} + \underbrace{\vartheta}_{\text{Prior}}$$

- ❑ The distribution $\boldsymbol{p}\ (\tau)\sim Dir(\alpha_c(\tau))$ follows a Dirichlet dsitribution.

Uncertainty Cross Entropy

- ❑ The classic Cross Entropy loss looks only at the mean of the distribution $P(\theta(\tau))$:

$$\mathcal{L}_i^{UCE} = \mathrm{H}[\boldsymbol{p}_i^*, \bar{\boldsymbol{p}}_i]$$

where $\bar{\boldsymbol{p}}_i(\tau) = \mathbb{E}_{\boldsymbol{p}_i\sim P_i(\theta(\tau_i))}[\boldsymbol{p}_i]$

Not suited for uncertainty estimation

- ❑ The Uncertainty Cross entropy loss looks at the distribution $P(\theta(\tau))$:

$$\mathcal{L}_i^{UCE} = \mathbb{E}_{\boldsymbol{p}_i\sim P_i(\theta(\tau_i^*))}[\mathrm{H}[\boldsymbol{p}_i^*, \boldsymbol{p}_i]]$$

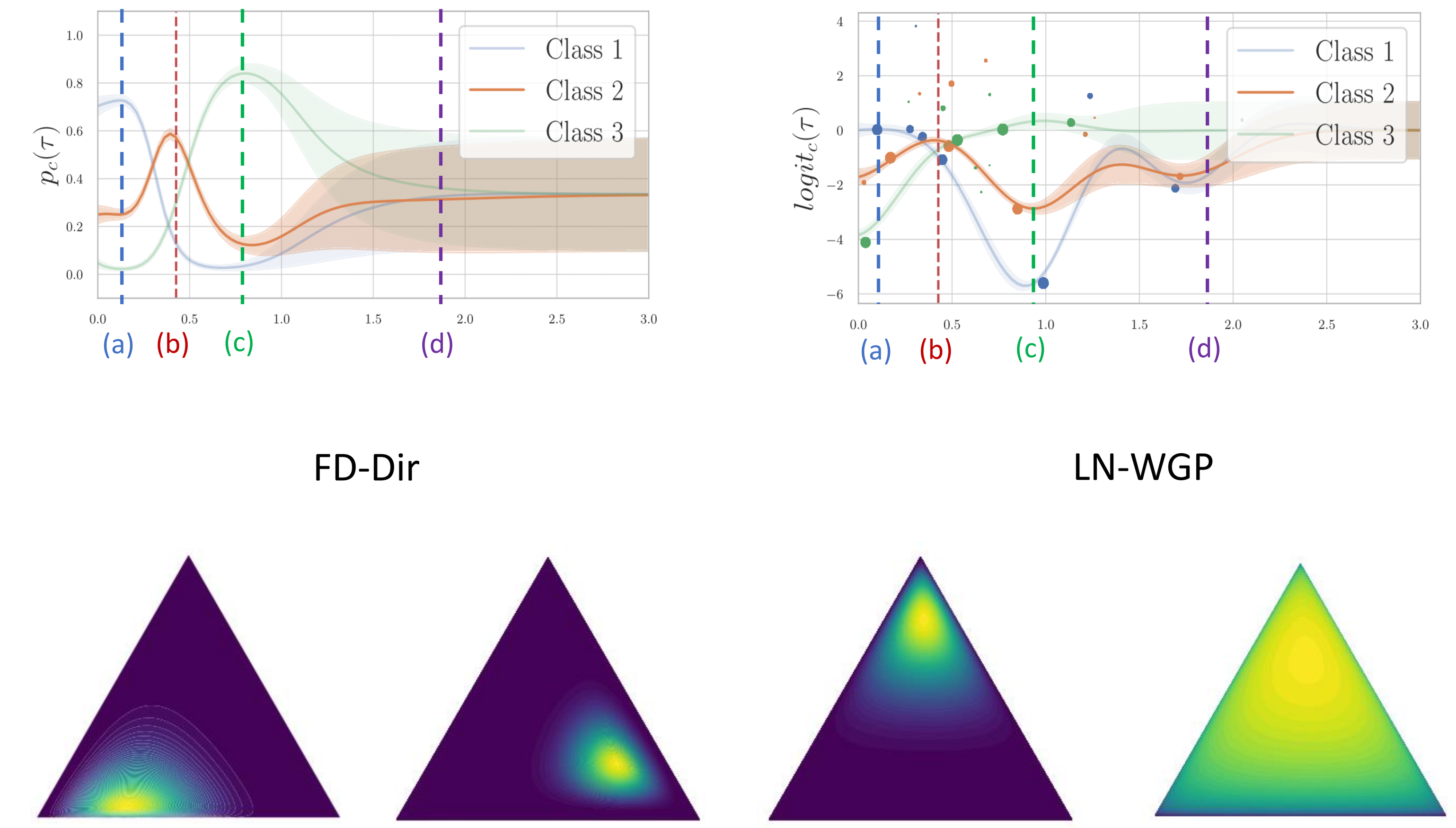
Take care of uncertainty estimation

- ❑ Regularizer adjust uncertainty estimation of out-of-distribution data:

$$r_c = \alpha\underbrace{\int_0^T(\mu_c)^2 d\tau}_{\text{Pushes mean to 0}} + \beta\underbrace{\int_0^T(\vartheta - \sigma_c^2)d\tau}_{\text{Pushes variance to }\vartheta}$$

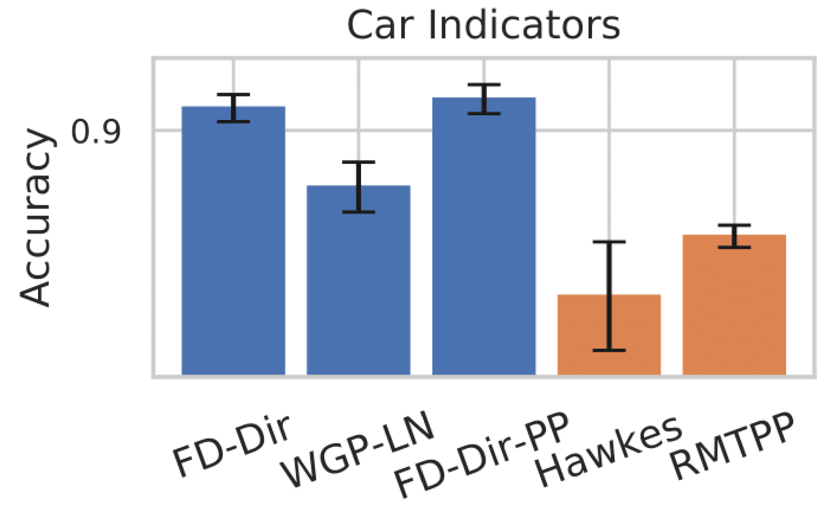
Computed by sampling

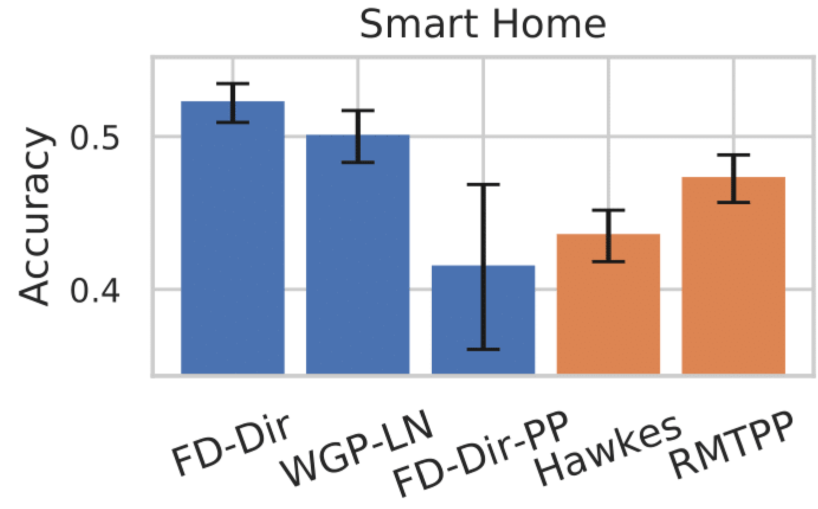
Experiment: Visualization



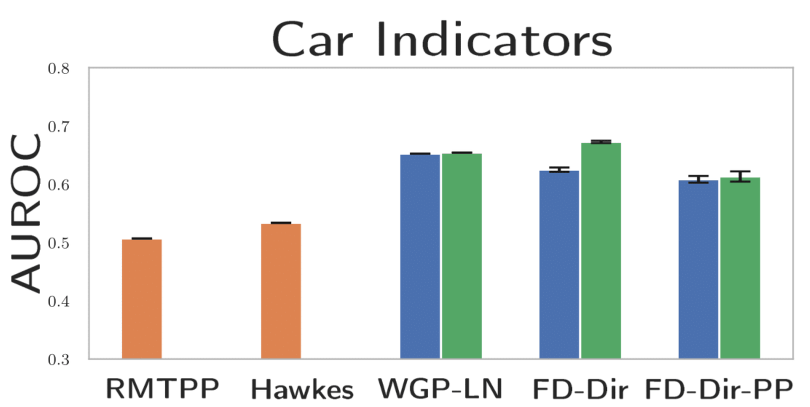
(a) (b) (c) (d)

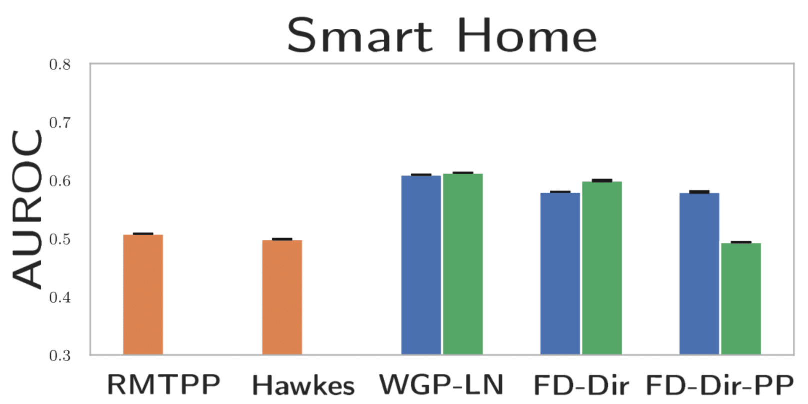
Experiment: Prediction

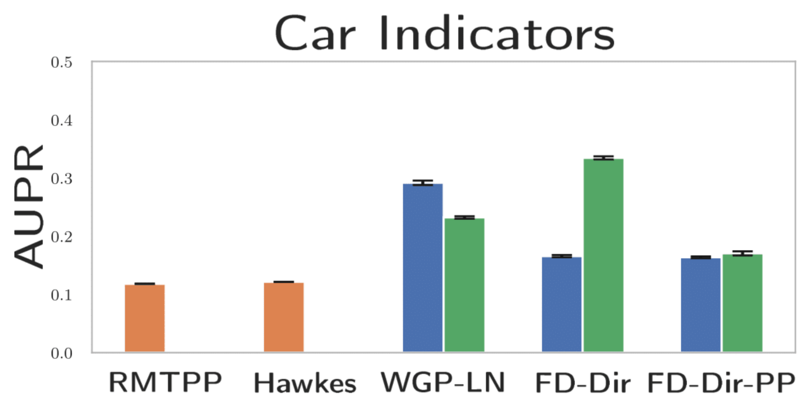


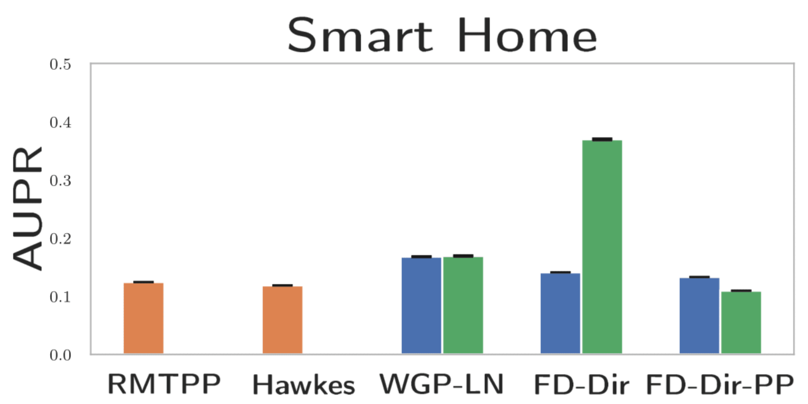


Experiment: Uncertainty









NeurIPS 2019

December 9-14, 2019
Vancouver, Canada

github.com/sharpenb/Uncertainty-Event-Prediction