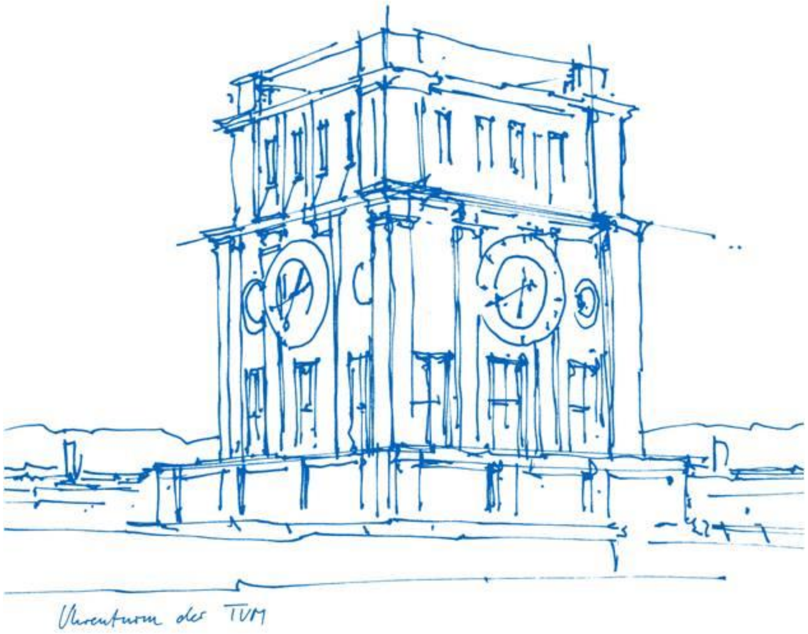
Temporal Point Processes

Johanna Maria Sommer

Advisor: Marin Biloš

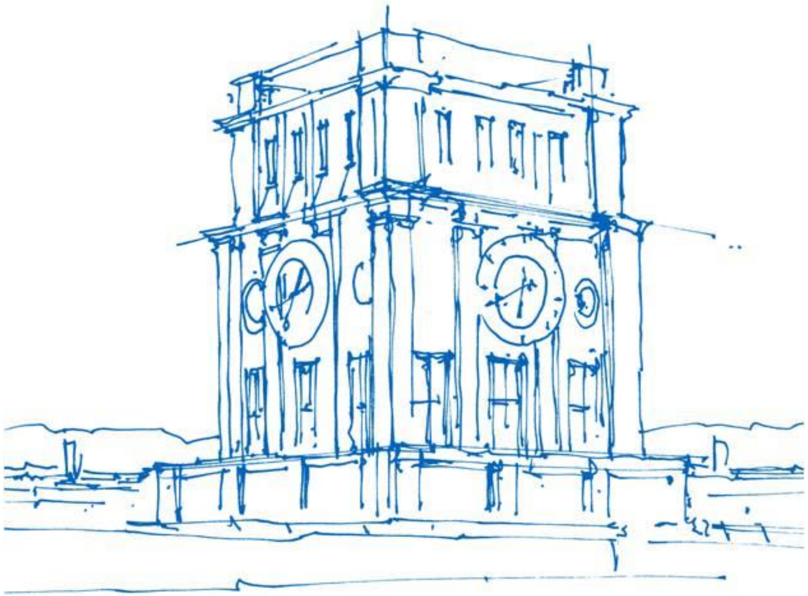
July 24th 2020



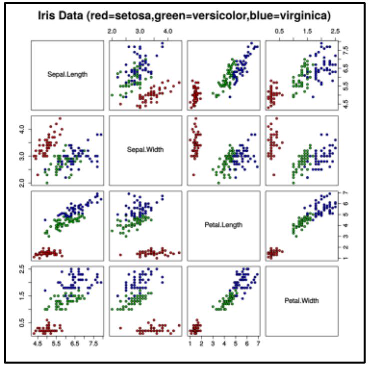
Overview

1. Introduction
2. Classical Models
3. Advanced Models
4. Related Research & Outlook

Introduction

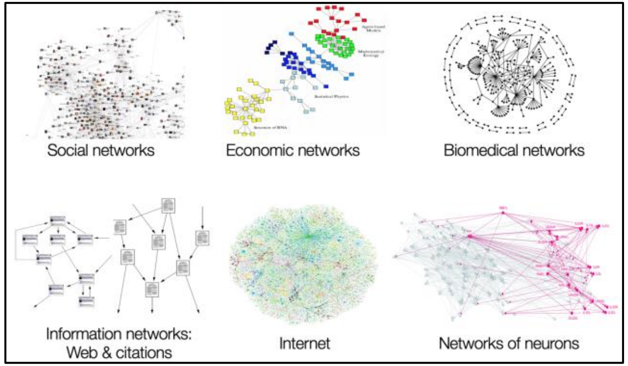


Uhrenturm der TUM



tabular data

image data



graph data

Temporal event data

- discrete events in continuous time
- time is not an index, but a random variable

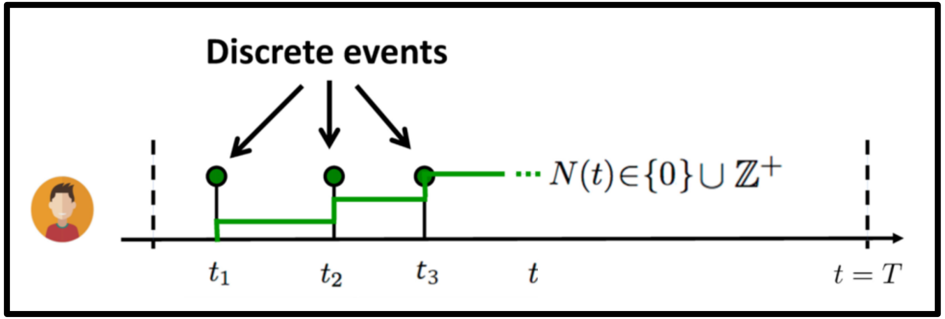


Figure from De, Upadhyay & Gomez-Rodriguez: Temporal Point Processes

Conditional Density Function

$f(t_{i+1}|H_{t_i})$

probability of next event t_{i+1} occurring in the time interval $[t, t + dt)$

Conditional Intensity Function

$$\lambda(t\,|H_{t_i}) = \frac{f(t_i|H_{t_i})}{1 - F(t_i|H_{t_i})}$$

Likelihood Function

$$L = \left(\prod_{i=1} \lambda^*(t_i)\right) \cdot \exp\big(-\Lambda^*(T)\big) \quad \text{with} \quad \Lambda^*(t) = \int_0^t \lambda^*(s)\,ds$$

Conditional Density Function

$f(t_{i+1}|H_{t_i})$ probability of next event t_{i+1} occurring in the time interval $[t, t + dt)$

Conditional Intensity Function

$\lambda^*(t)$ rate at which events come in

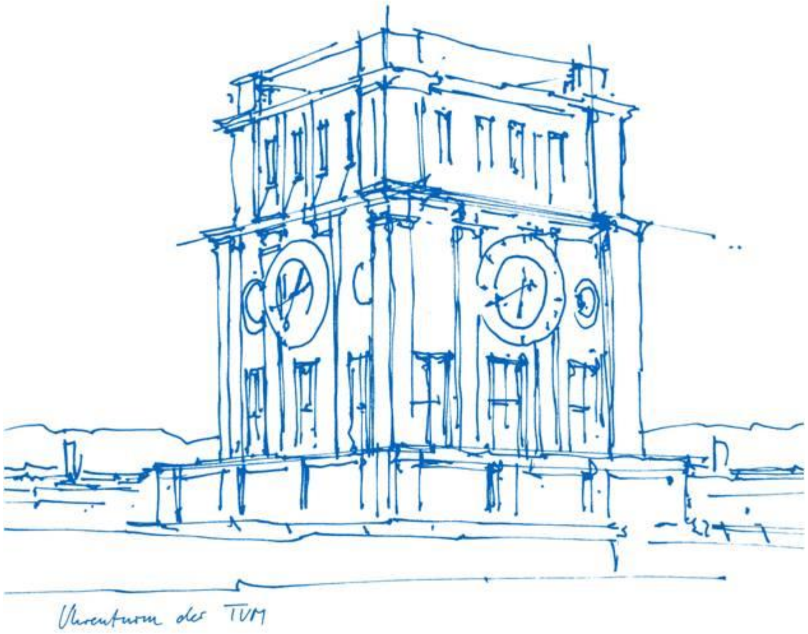
Likelihood Function

$$L = \left(\prod_{i=1} \lambda^*(t_i)\right) \cdot \exp\left(-\Lambda^*(T)\right) \quad \text{with} \quad \Lambda^*(t) = \int_0^t \lambda^*(s)\, ds$$

Marked TPPs

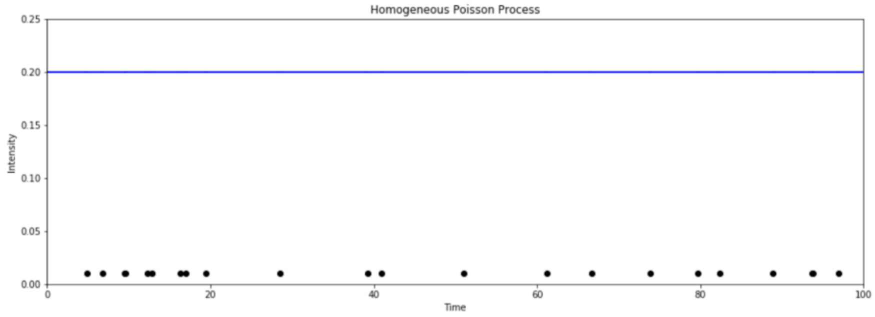
- additional discrete markers that belong to a specific event in time
  $\lambda^*(t, \kappa) = \lambda^*(t) * f\left(\kappa, H_{t_{i-1}}\right)$
- the domain of the marks is application dependent
- the history includes the timings as well as the markers

Classical Models



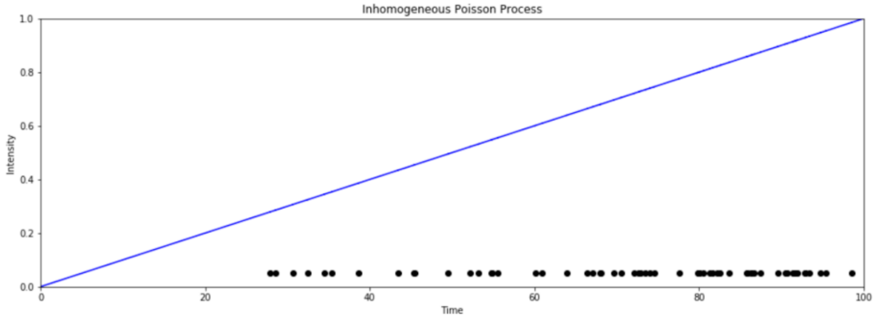
Homogeneous Poisson Process

$$\lambda_{\mu}^{*}(t) = \mu \geq 0$$



Inhomogeneous Poisson Process

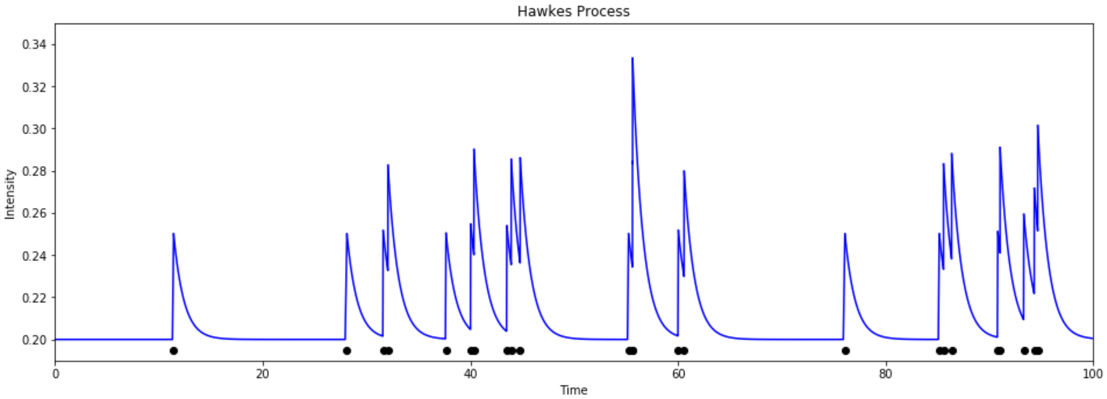
$$\lambda_{\theta}^{*}(t) = g_{\theta}(t) \geq 0$$



Hawkes Process

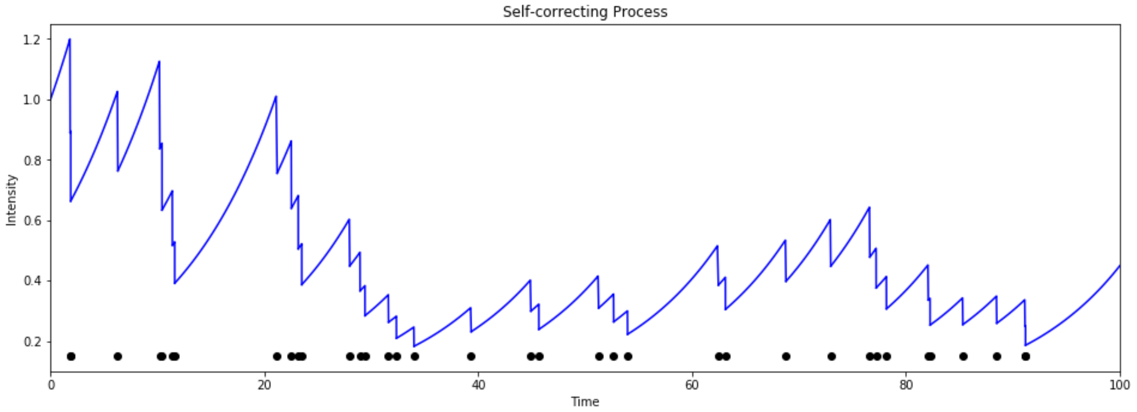
self-exciting process

$$\lambda^*(t) = \lambda_0(t) + \sum_{t_i \in H_t} \phi(t - t_i)$$



Self-correcting Process

$$\lambda^*(t) = \exp\left(\mu t - \sum_{t_i \in H_t} \alpha\right)$$



What we have seen so far

Poisson Process

- memoryless
- intensity constant or evolving according to a function
- used when events are independent of each other

Hawkes/Self-exciting Process

- intensity increases when an event occurs and then continuously decays
- leads to a clustered pattern, used in e.g. seismology

Self-correcting Process

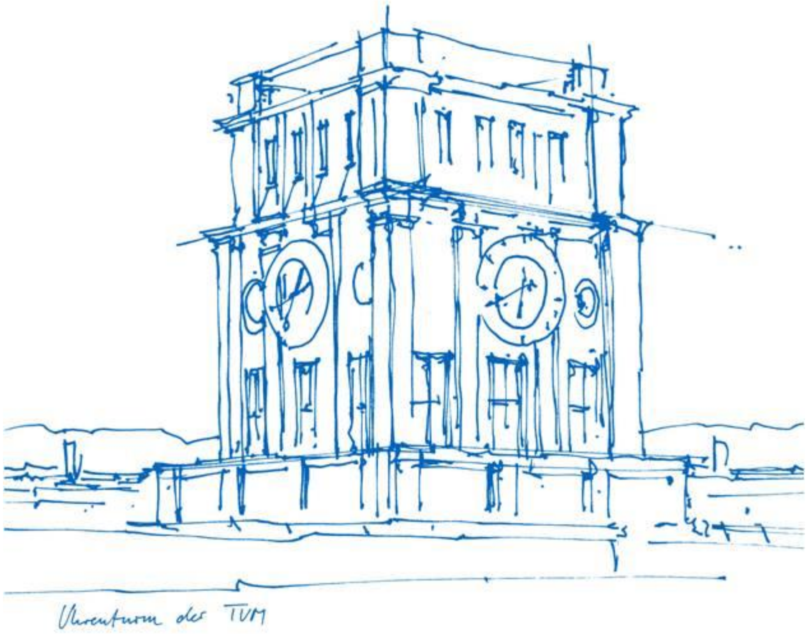
- intensity decreases when an event occurs and then continuously increases
- leads to regular events

Problems

- making assumptions about the models

  functional form, model misspecification
- hard to generally specify a process
- intractable integral

Advanced Models

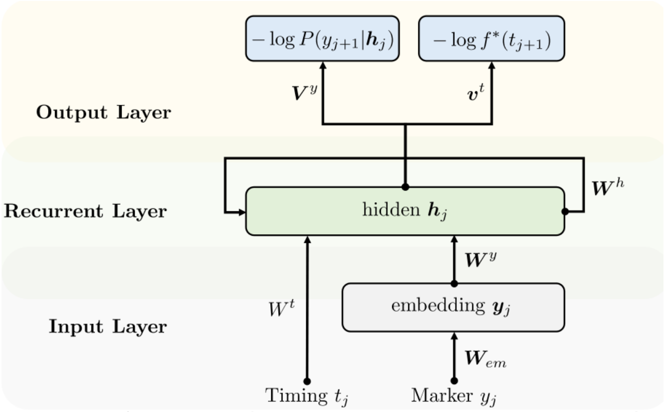


RMTPP

Recurrent Marked Temporal Point Processes: Embedding Event History to Vector
Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, Le Song | KDD 2016

- jointly model time and marker distribution
- history is encoded by hidden cell in a recurrent layer
- learning by backpropagation through the network

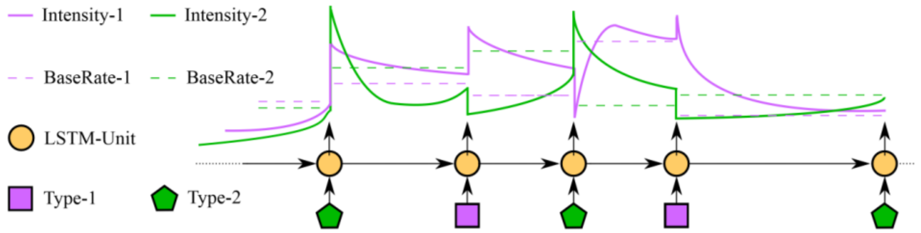
$$\lambda^*(t) = \exp\left(\boldsymbol{v}^{t^\top} * \boldsymbol{h}_j + w^t(t - t_j) + b^t\right)$$



Neural Hawkes

The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process
Hongyuan Mei, Jason Eisner | NeurIPS 2017

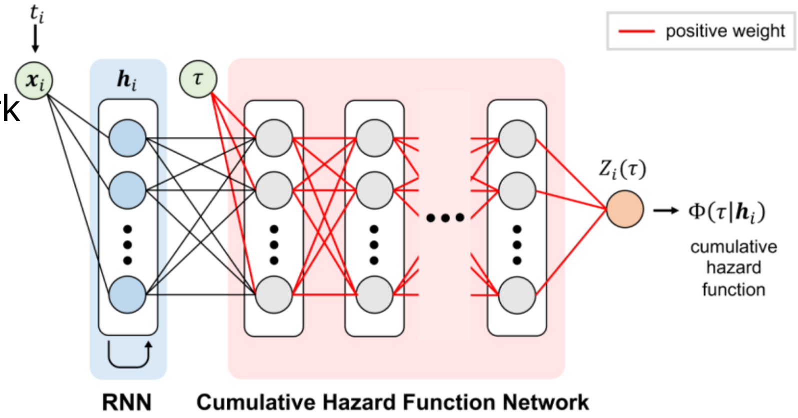
- intensity evolves according to continuous-time LSTM
- learning by backpropagation



FullyNN

Fully Neural Network based Model for General Temporal Point Processes
Takahiro Omi, Naonori Ueda, Kazuyuki Aihara | NeurIPS 2019

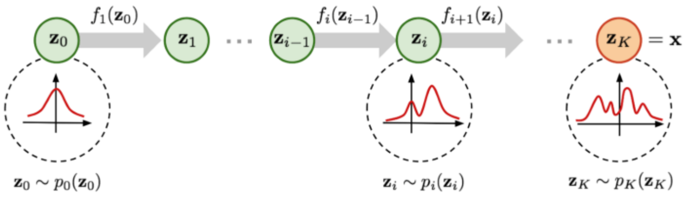
- cumulative intensity function is the output of the network
- conditional intensity can be obtained by differentiation
- generalize RNN-approaches to represent intensity in a general functional form

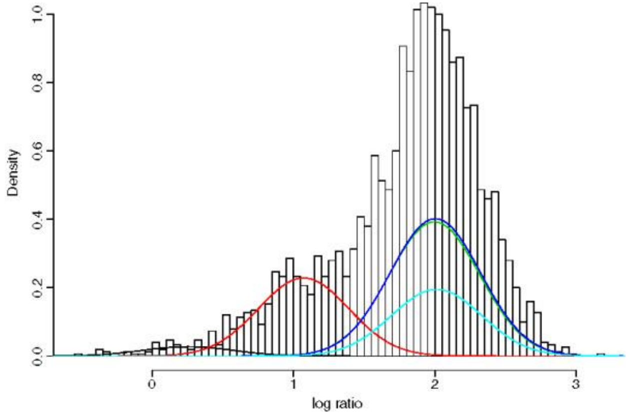


DSFlow & LogNormMix

Intensity-Free Learning of Temporal Point Processes
Oleksandr Shchur, Marin Biloš, Stephan Günnemann | ICLR 2020

- model the **conditional density function**
- model probability distribution via Normalizing Flows or a mixture model





Discussion

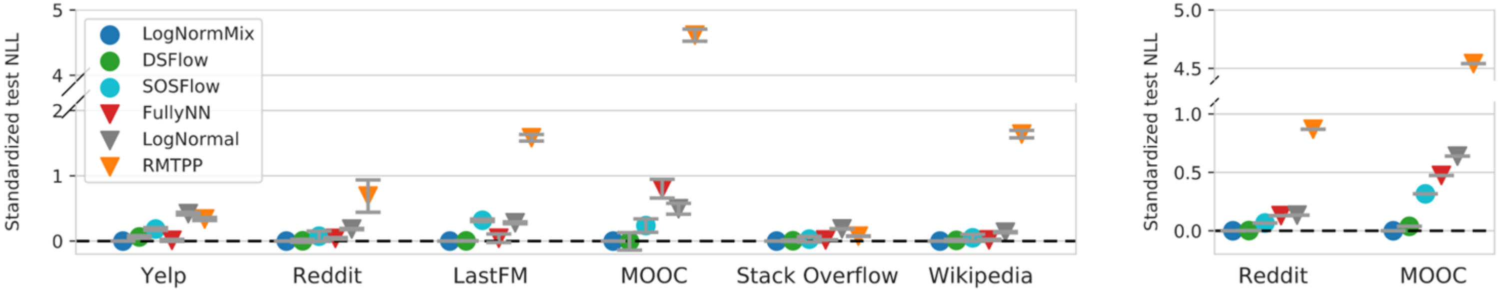


Figure from Intensity-Free Learning of Temporal Point Processes by Oleksandr Shchur, Marin Biloš, Stephan Günnemann

- FullyNN yields consistently good results, but is outperformed by LogNormMix and DSFlow
- RMTPP and Neural Hawkes fall off, most likely due to their choice of functional form

- Missing: extensive complexity and runtime analysis

To model the intensity or not to model the intensity

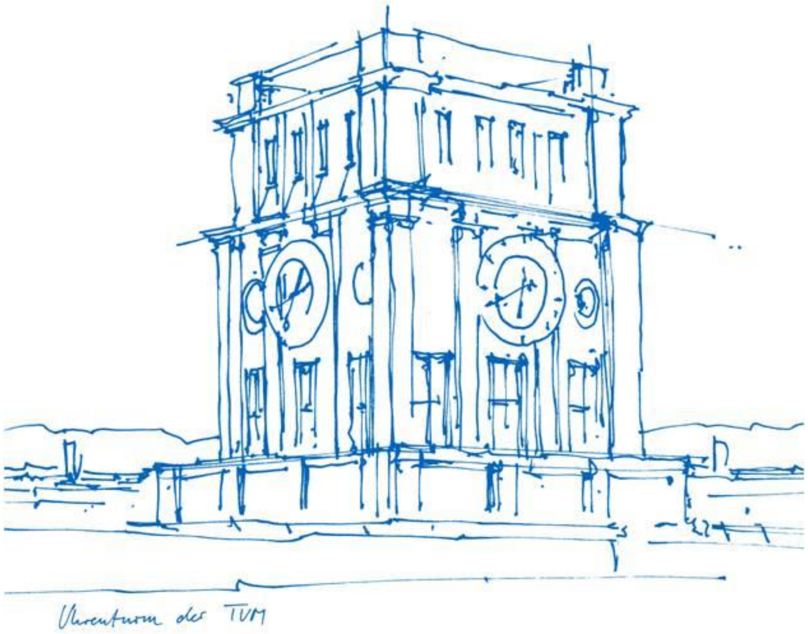
Pro

- the conditional intensity provides some intuition and interpretability
- easier to specify because it has less constraints

Con

- with recurrent formulations, the intensity function is not intuitive anymore
- it seems the conditional density function is easier to learn

Related Research & Outlook



Related Research

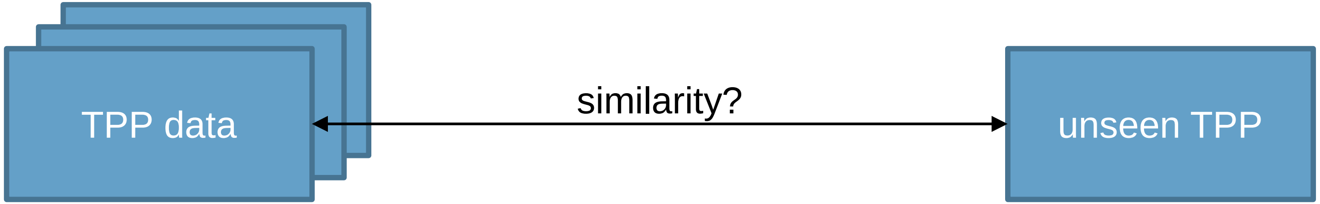
- Problem: intractable integral in the likelihood function, MLE learning
  - learn integral by a set of parameters
  - Wasserstein GAN for conditional probability distribution modelling
  - Reinforcement learning for TPPs, learning the intensity = learning the policy

- Problem: interactions between events of different types, multivariate point processes
  - model mutual excitation between marks
  - exploit sparsity inherent in multivariate point processes to increase efficiency

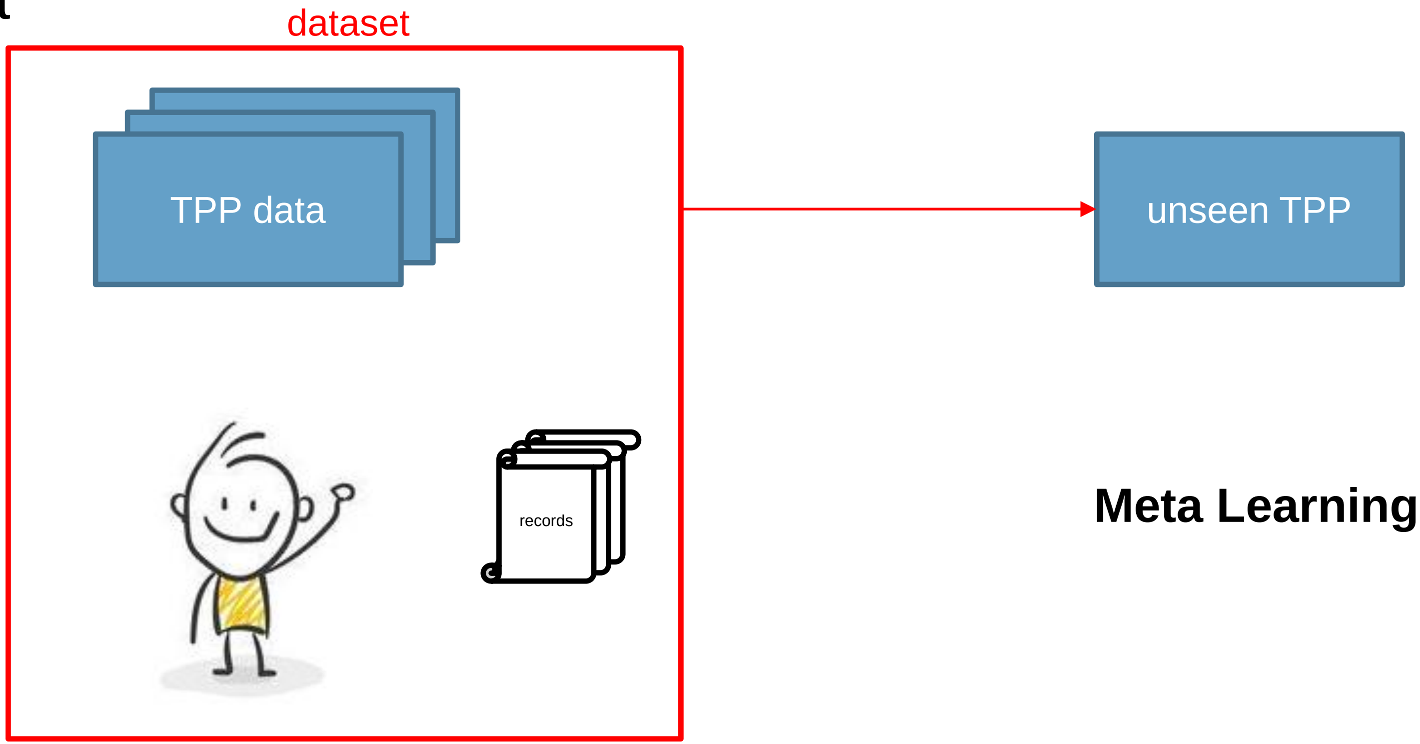
Future Research

- Multivariate point processes
- Alternatives to RNNs, e.g. self-attention
- Intensity-free learning
- Reinforcement learning for TPPs

Idea



Idea



Meta Learning

Thank you for your attention.

Temporal Point Processes: A Survey

Johanna Sommer