Hawke's neural process

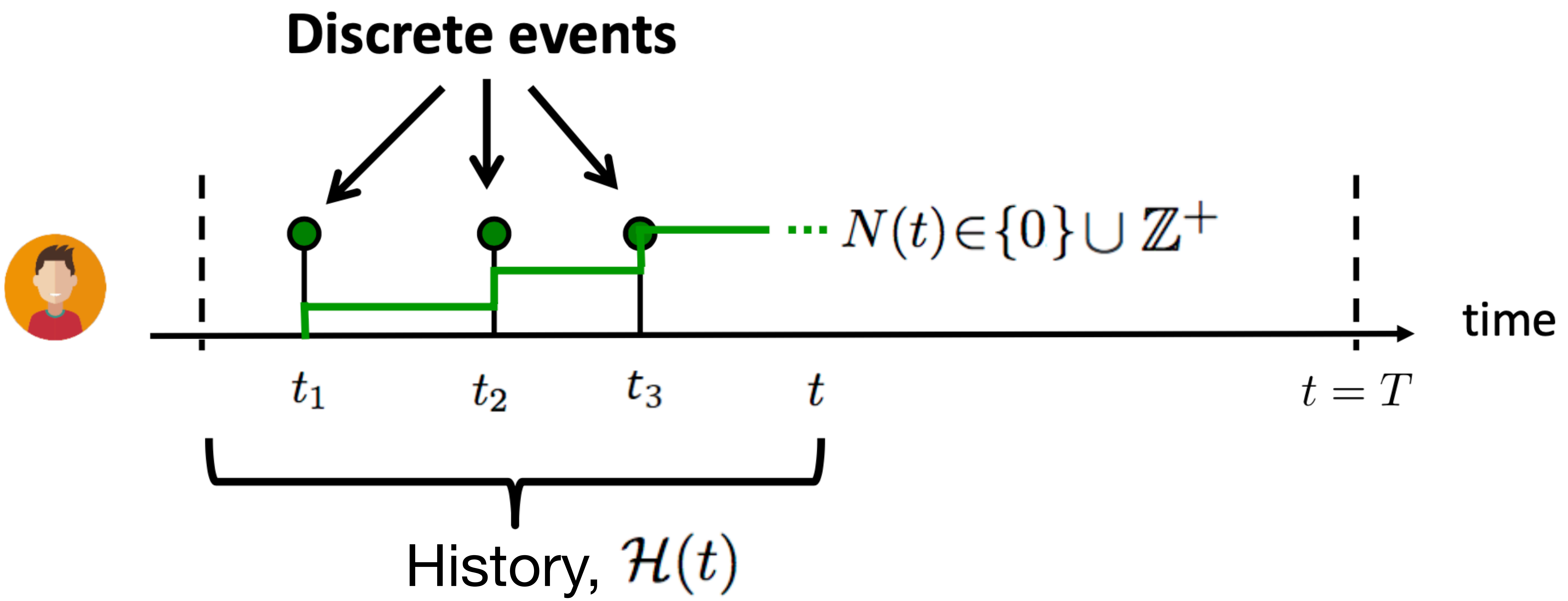
Skoltech

Temporal Point Processes: Recap

Skoltech

Temporal point processes

- ***Temporal point process*** — a random process whose realization consists of discrete events localized in time $\mathcal{H} = \{t_i\}$

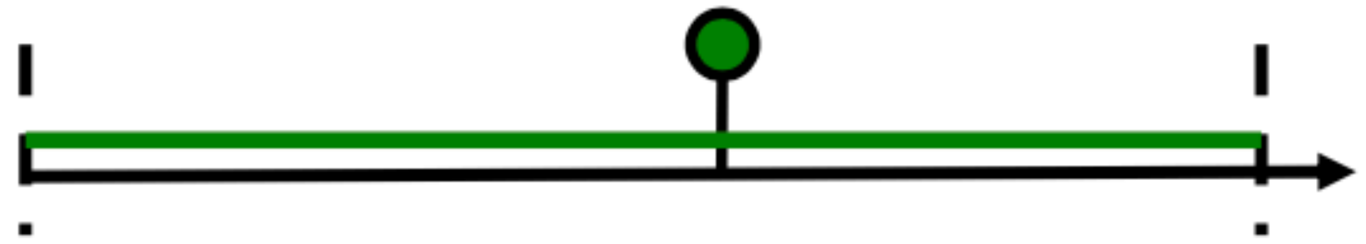


Skoltech

Building blocks to represent different dynamic processes

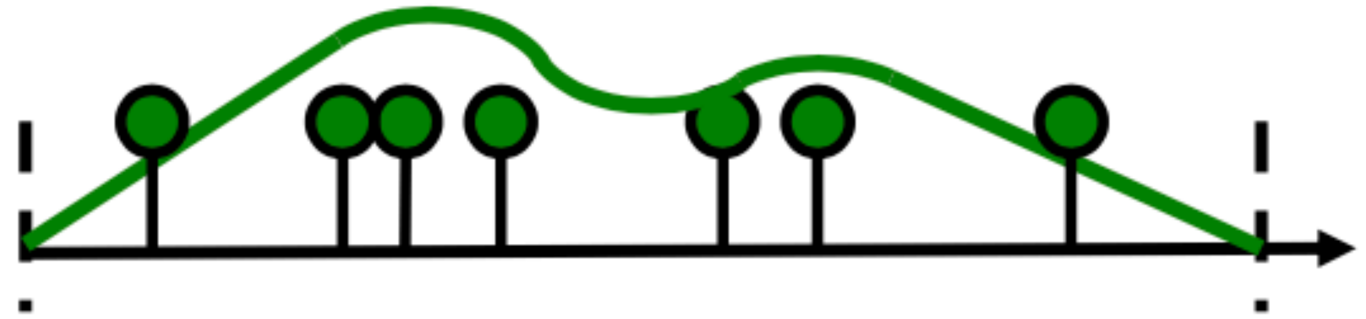
Poisson processes:

$$\lambda^*(t) = \lambda$$



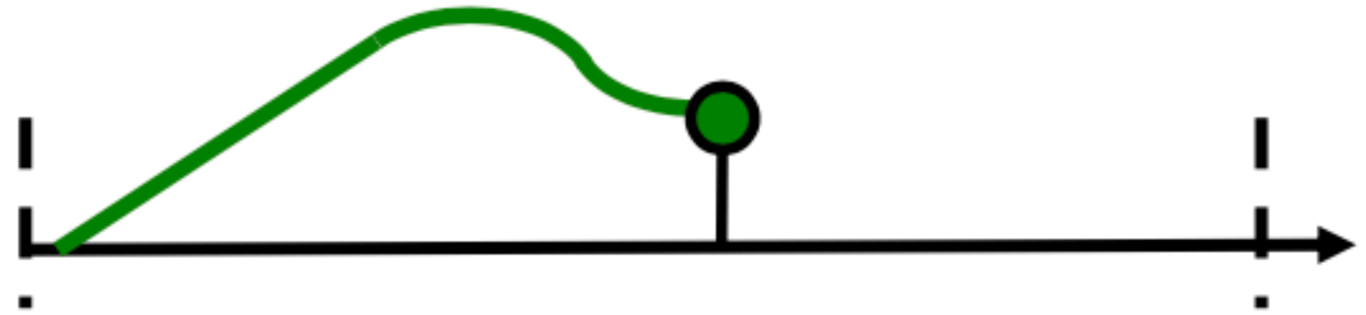
Inhomogeneous Poisson processes:

$$\lambda^*(t) = g(t)$$



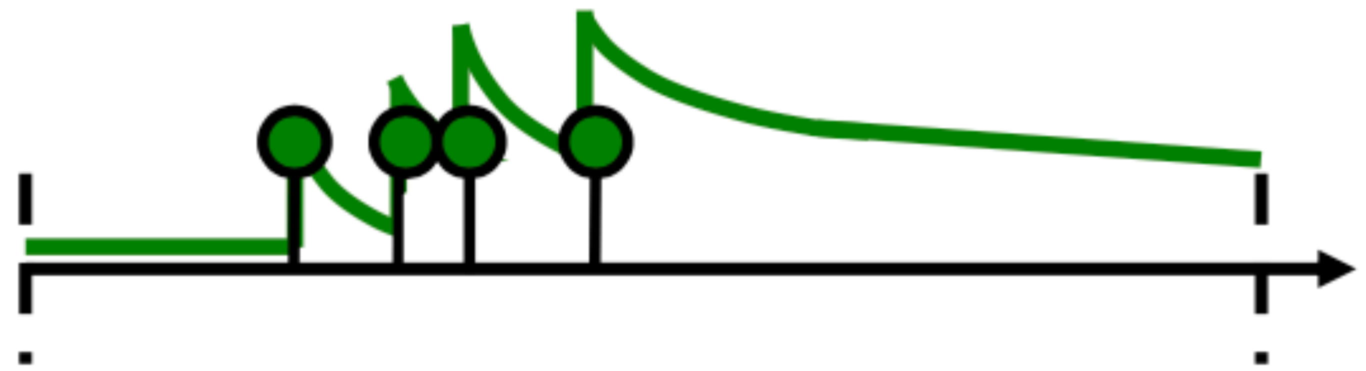
Terminating point processes:

$$\lambda^*(t) = g^*(t)(1 - N(t))$$



Self-exciting point processes:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$



Vanilla Hawkes process drawbacks

- Past events can temporarily raise the probability of future events
- *Real world application*: the flexibility of vanilla Hawkes process is very limited to approximate many sections of the problems

$$\Downarrow$$

- Need a better model with least assumptions to approximate the conditional intensity function

Skoltech

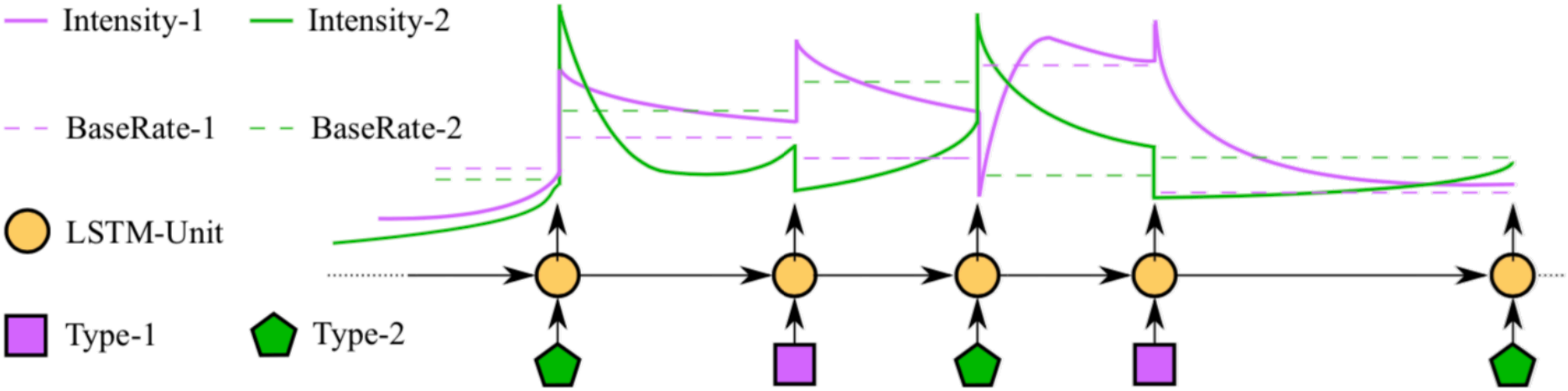
Solution: Neural Hawkes process

Neural Hawkes process

Skoltech

Neural Hawkes process advantages

- History effect does not need to be additive
- Allows for complex memory effects (such as delays)



Source: Mei & Eisner, NIPS 2017

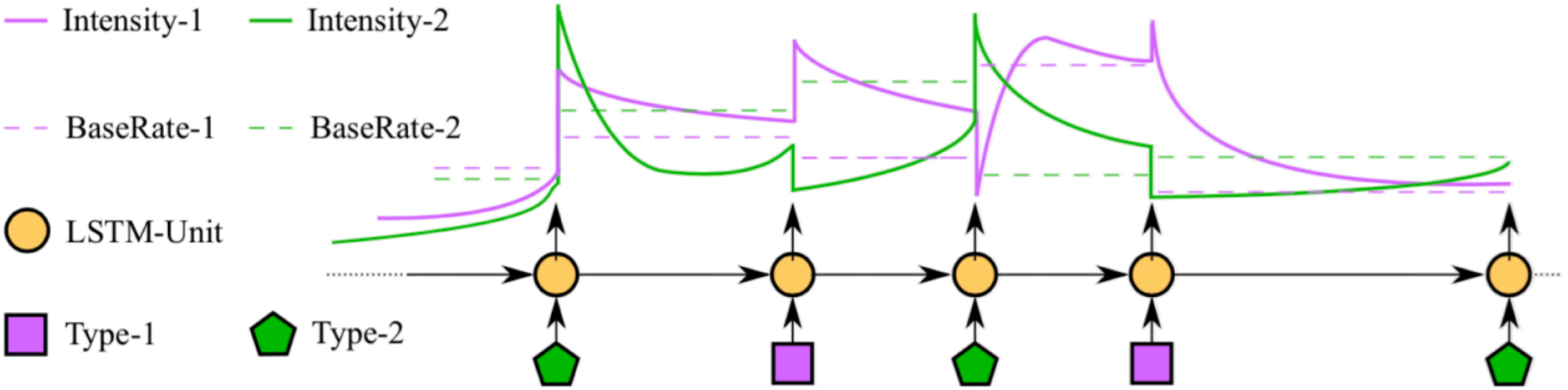
Skoltech

Neural Hawkes process: RNNs

- Memory via the continuous-time LSTM
- Excitation & inhibition via activation function

$$\mathbf{h}(t) = \mathrm{RNN}(\mathcal{H}(t))$$

$$\lambda_u(t) = f_u(\mathbf{w}_u^\top \mathbf{h}(t))$$



Source: Mei & Eisner, NIPS 2017

Skoltech

Event stream data

- Financial transactions
- Medical notes
- …

Stock trading

Flu spreadin

Reviews and sales in Amazon

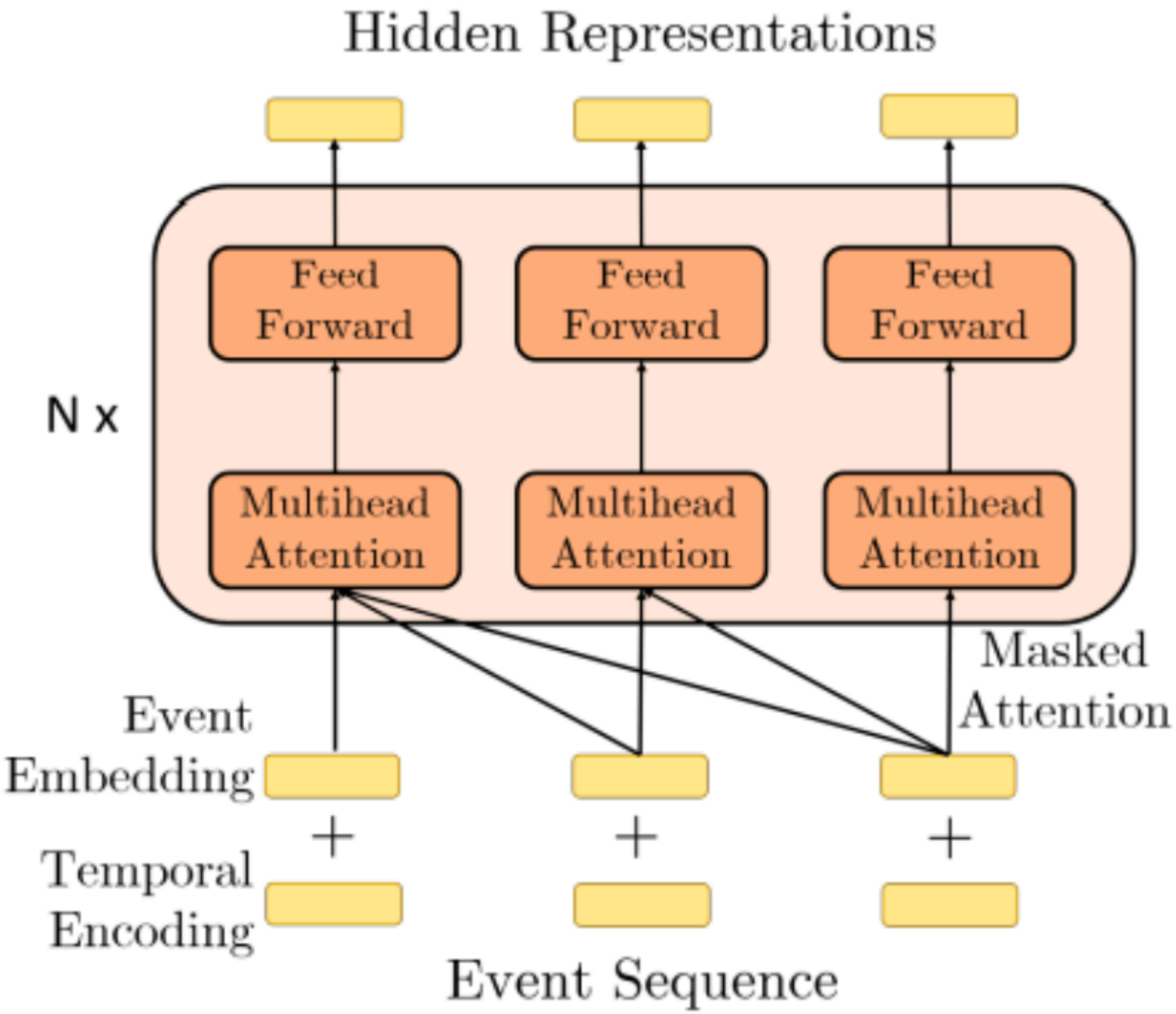
Long-term dependencies issues

- Financial transactions: policy issues
- Medical notes: chronic diseases

News spread in Twitter

Skoltech

Neural Hawkes process: Transformers



Source: Zuo, Simiao, et al. "Transformer Hawkes process." ICML. 2020

Skoltech

Neural Hawkes process: Transformers

$$\underbrace{\lambda_k(t|\mathcal{H}_t)}_{\text{History up to time } t} = f_k\Big(\underbrace{\alpha_k \frac{t - t_j}{t_j}}_{\textit{current}} + \underbrace{\mathbf{w}_k^\top \mathbf{h}(t_j)}_{\textit{history}} + \underbrace{b_k}_{\textit{base}}\Big)$$

History up to time t

Interpolation between two observed time stamps

$$p(t|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)\exp\left(-\int_{t_j}^{t} \lambda(\tau|\mathcal{H}_\tau)d\tau\right),$$

$$\widehat{t}_{j+1} = \int_{t_j}^{\infty} t \cdot p(t|\mathcal{H}_t)dt,$$

$$\widehat{k}_{j+1} = \underset{k}{\operatorname{argmax}} \frac{\lambda_k(t_{j+1}|\mathcal{H}_{j+1})}{\lambda(t_{j+1}|\mathcal{H}_{j+1})}.$$

Skoltech

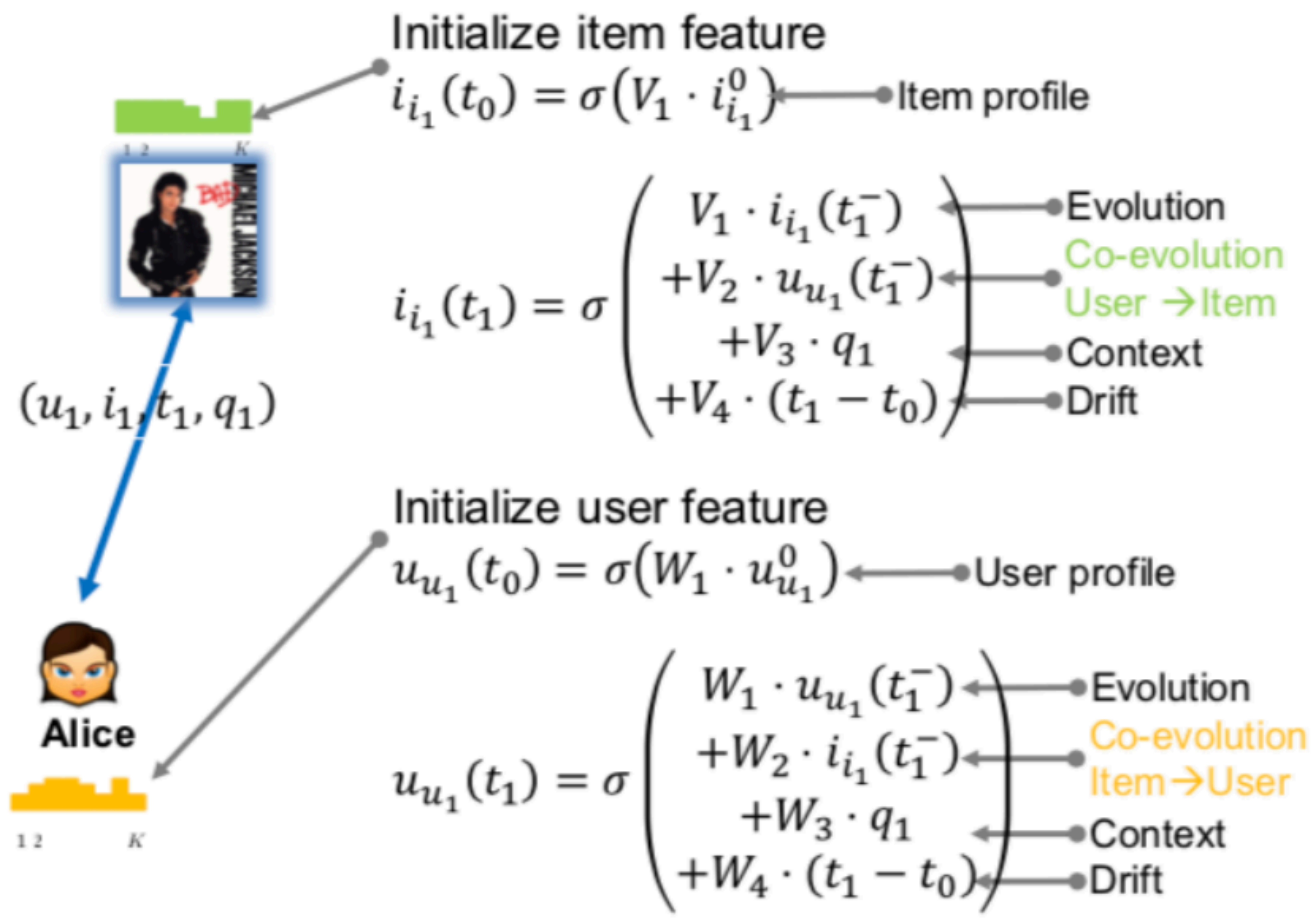
Applications

Skoltech

Predictive Models

Know-Evolve, Trivedi et al. (2017)

Coevolutionary Embedding, Dai et al. (2017)

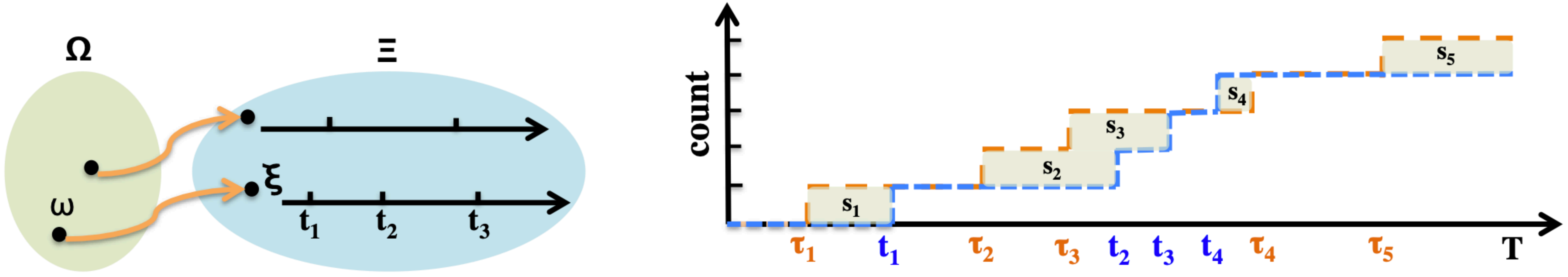


Skoltech

Generative Models

- Key idea: Intensity- and likelihood-free models

a) The outcome of the random experiment ω is mapped to a point in space of count measures ξ;
b) Distance between two sequences ξ = {t1,t2,...} and ρ = {τ1,τ2,...}



a) Point process probability space b) $\|\cdot\|_\star$ distance between sequences

Source: Shuai Xiao et al. Wasserstein Learning of Deep Generative Point Process Models

Skoltech

Conclusion

- A type of sequential data observed in many applied problems:
  - Client’s actions
  - Abnormal Events
  - Diseasespread
- Various models for intensity include LSTMs, Transformers

Skoltech