NeurIPS 2022

HYPRO: A Hybridly Normalized Probabilistic Model for Long-Horizon Prediction of Event Sequence

Siqiao Xue [1] , Xiaoming Shi[1] , James Y Zhang[1], Hongyuan Mei[2]

1：Ant Group

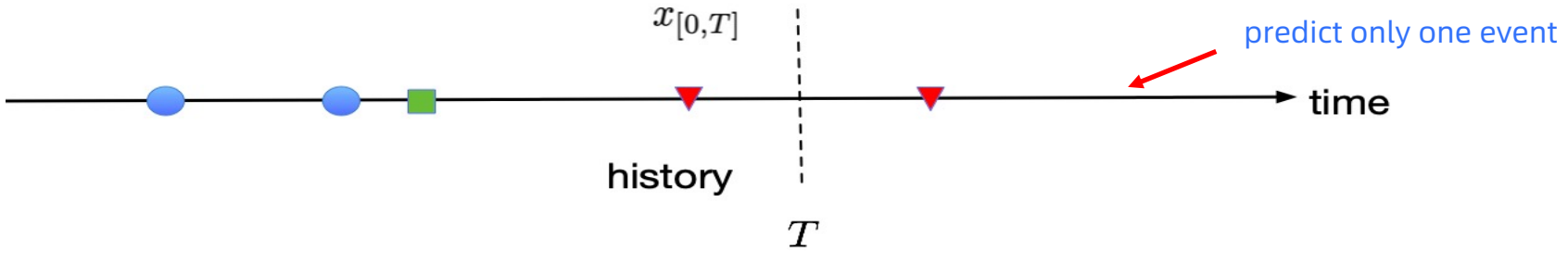
2：Toyota Technological Institute at Chicago

2022.10

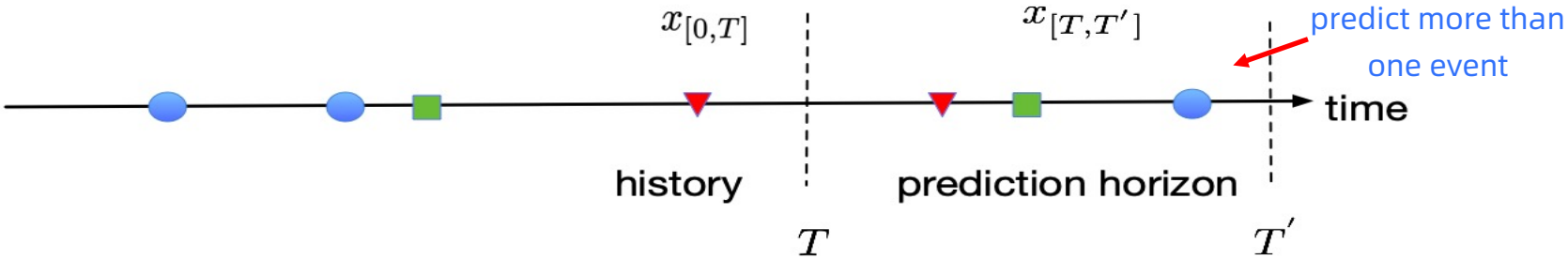
Problem

Given an event sequence $x_{[0,T]} = \{(t_0, k_0), (t_1, k_1), \ldots, (t_n, k_n)\}, t_i \in R, k_i \in \mathrm{N}$

The typical problem: predict next event



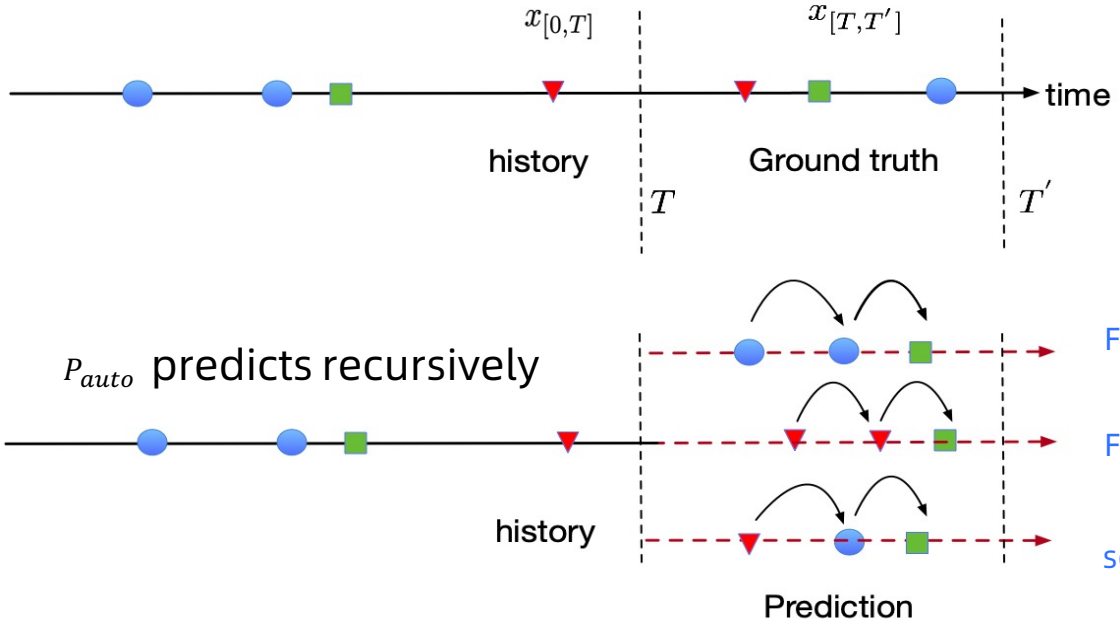
Our problem: predict **next multiple events over a long future horizon** $[T, T']$.



Challenge: Cascading Error and Local Normalization

A typical solution is **autoregressive** model P_{auto}

$$P(x_{i+k},...,x_{i+1}|x_i,...,x_0) = \prod P(x_{i+k}|\hat{x}_{i+k-1},...,x_{\leq i})...P(x_{i+2}|\hat{x}_{i+1},x_{\leq i})P(x_{i+1}|x_{\leq i})$$



First type prediction incorrect

First time prediction incorrect

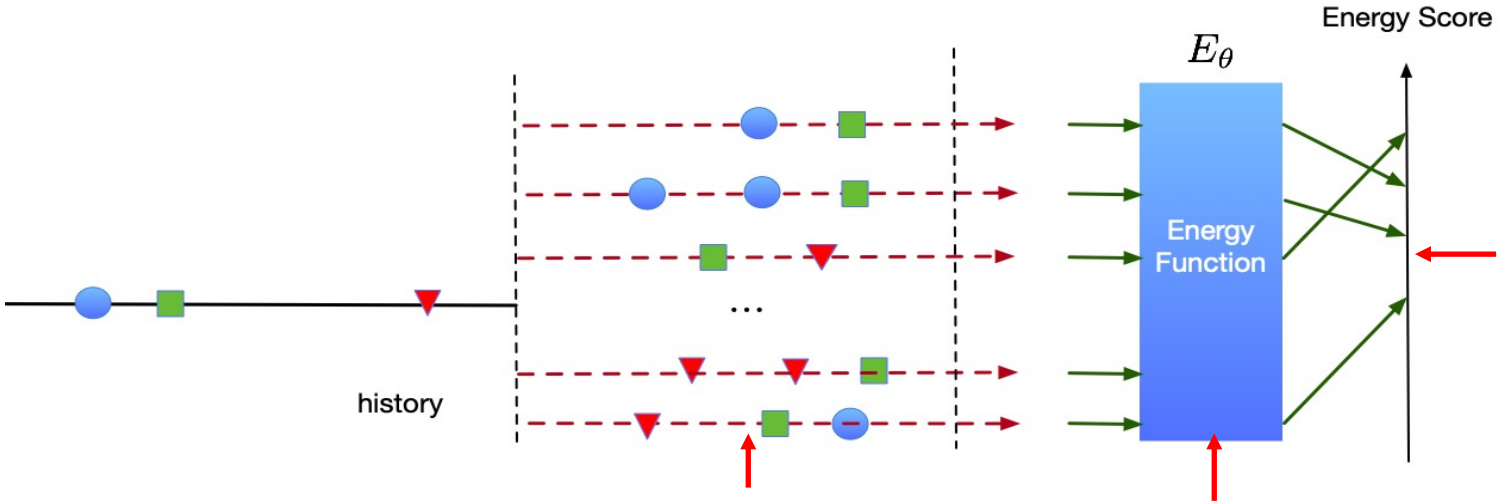
The first is correct while the second is incorrect: the errors are still cascaded to the subsequent events

Whenever you make mistakes, you'll never have a chance to correct it !

Our Key Idea: Using Energy-based Model

$$p_{\text{HYPRO}}\left(x_{(T,T']} \mid x_{[0,T]}\right) = p_{\text{auto}}\left(x_{(T,T']} \mid x_{[0,T]}\right) \frac{\exp\left(-E_\theta(x_{[0,T']})\right)}{Z_\theta\left(x_{[0,T]}\right)}$$

Hybridly normalized | Locally normalized | Globally normalized



The more correct sample has lower energy so the overall prob is correctly scaled

Using P_{auto} to generate many sequence samples

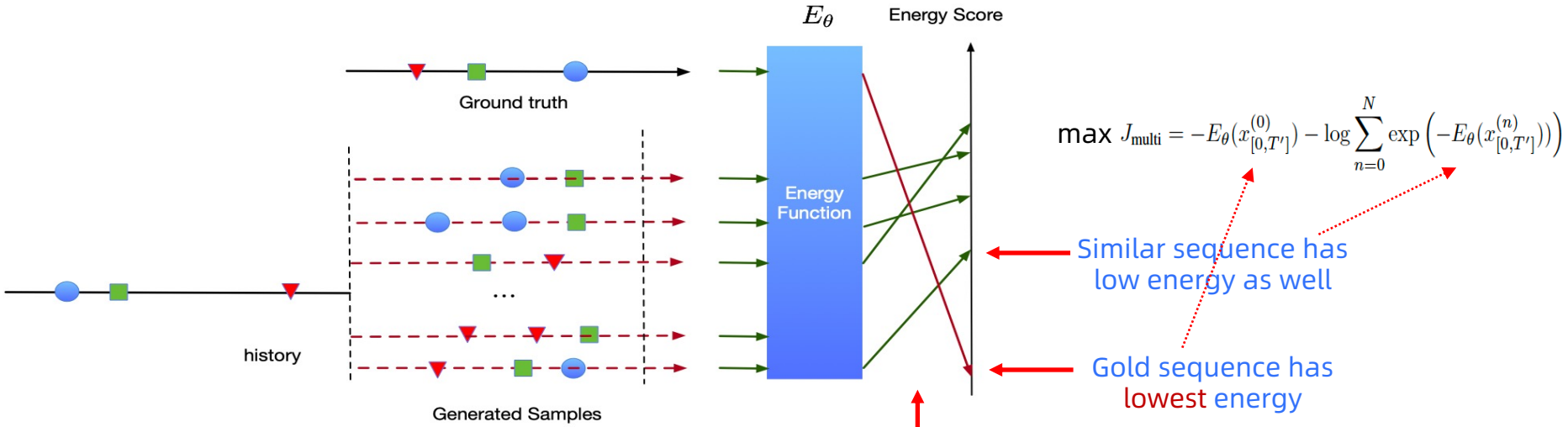
E_θ looks at the entire sequence of each sample and compute its energy

Our energy function looks at each entire sequence, so it has a chance to correct any earlier errors!

Model Training: Noise-Contrastive Learning

$$p_{\text{HYPRO}}\left(x_{(T,T']} \mid x_{[0,T]}\right) = p_{\text{auto}}\left(x_{(T,T']} \mid x_{[0,T]}\right) \frac{\exp\left(-E_\theta(x_{[0,T']})\right)}{Z_\theta\left(x_{[0,T]}\right)}$$

Hybridly normalized | Locally normalized | Globally normalized



$$\max J_{\text{multi}} = -E_\theta(x^{(0)}_{[0,T']}) - \log \sum_{n=0}^{N} \exp\left(-E_\theta(x^{(n)}_{[0,T']})\right)$$

Similar sequence has low energy as well

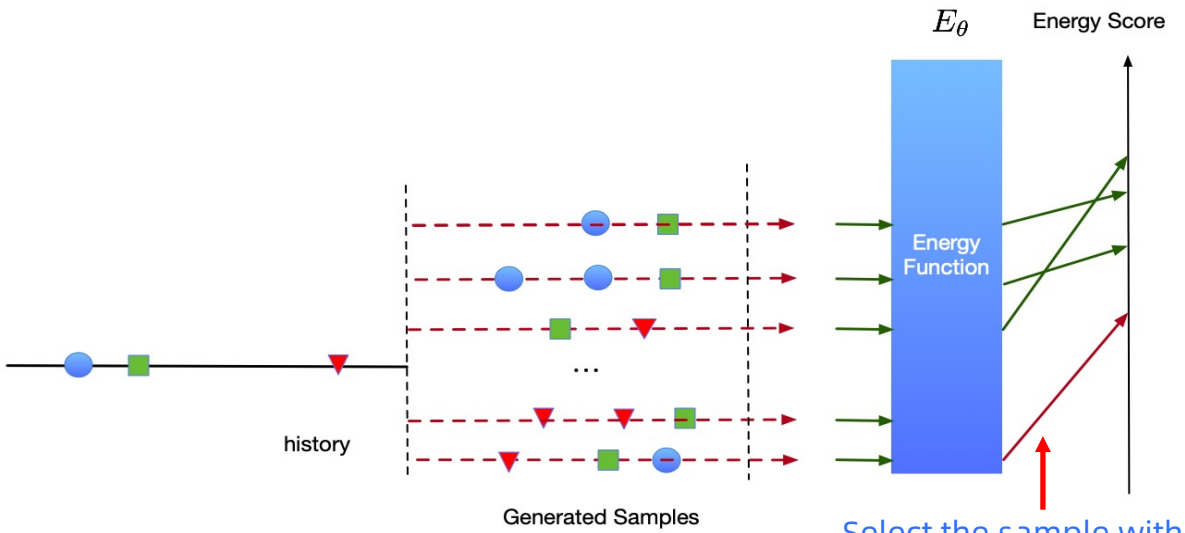
Gold sequence has lowest energy

Learn to discriminate between the gold and generated samples via NCE

Model Inference：Normalized Importance Sampling

$$p_{\text{HYPRO}}\left(x_{(T,T']} \mid x_{[0,T]}\right) = p_{\text{auto}}\left(x_{(T,T']} \mid x_{[0,T]}\right) \frac{\exp\left(-E_\theta(x_{[0,T']})\right)}{Z_\theta\left(x_{[0,T]}\right)}$$

Hybridly normalized　Locally normalized　Globally normalized



Select the sample with highest model probability
(consider both P_{auto} and energy function E_θ)

See Algorithm 2 in the paper for details

Looking forward to seeing you at our poster and we can discuss

model details

training details

experimental results

paper can be downloaded from