Louisiana State University

The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process

Hongyuan Mei   Jason Eisner

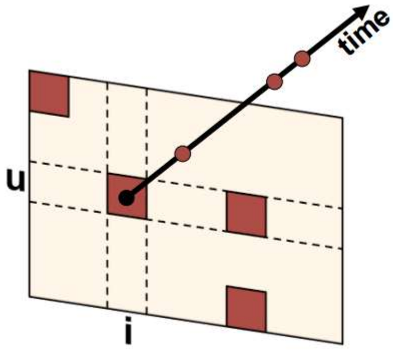
Presented by Zhuoqun Li

Contents

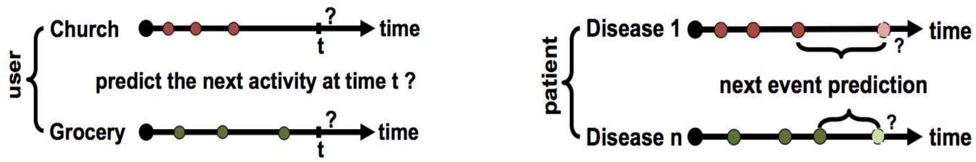
- Introduction
- Model
- Experiments
- Conclusion

Introduction

Sequences of discrete events in continuous time
- ❑ Medical events.
- ❑ Consumer behavior.
- ❑ Social media actions.
- ❑ Other event streams arise in news, animal behavior, dialogue, music, etc.



Introduction



Learning the distribution of sequences of events.
- ❑ Predict most desirable activity at a given time t for a user.

(which events are likely to happen next)
- ❑ Predict the returning time to a particular activity of a user.

(when they will happen)

Introduction

Poisson process (Palm, 1943): Assumes that events occur independently of one another.

Non-homogenous Poisson process: the probability of an event happening at time t may vary with t, but it is still independent of other events.

Hawkes process (Hawkes, 1971; Liniger, 2009): Supposes that past events can temporarily raise the probability of future events, assuming that such excitation is positive, additive over the past events, and exponentially decaying with time.

Introduction

① **Violated if one event inhibits another rather than exciting it**.
(e.g. Cookie consumption inhibits cake consumption.)

② **Violated when the combined effect of past events is not additive.**
(e.g. The new advertisement sometimes does not increase purchase rate as much as the early advertisement did, and may even drive customers away.)

③ **Violated when a past event has a delayed effect.**

Model

Hawkes Process: A Self-Exciting Multivariate Point Process (SE-MPP)

1. Non-homogeneous multivariate Poisson process

$$\lambda(t) = \sum_{k=1}^{K} \lambda_k(t)$$

2. Self-exciting multivariate point process (SE-MPP)

$$\lambda_k(t) = \mu_k + \sum_{h:t_h<t} \alpha_{k_h,k} \exp(-\delta_{k_h,k}(t - t_h))$$

Model

Neural Hawkes Process: A Neurally Self-Modulating MPP

1. Relax the positivity constraints on αj,k and μk, which allows the events inhibition (αj,k < 0) and (μk < 0).

2. Removes the restriction that the past events have independent, additive influence on λ˜k(t). In the new process, the time-decaying influences are controlled by a hidden state vector.

Model

Relax the positivity constraints on αj,k and μk

$$\lambda_k(t) = f_k(\tilde{\lambda}_k(t))$$
$$f(x) = s \log(1 + \exp(x/s))$$
$$\tilde{\lambda}_k(t) = \mu_k + \sum_{h:t_h<t} \alpha_{k_h,k} \exp(-\delta_{k_h,k}(t - t_h))$$

As t increases between events, the intensity λk(t) may both rise and fall, but eventually approaches the base rate f(μk+0), as the influence of each previous event still decays toward 0 at a rate δj,k > 0.

Model

Neural Hawkes process

At each time t > 0, it obtain the intensity λk(t)

$$\lambda_k(t) = f_k(\mathbf{w}_k^\top \mathbf{h}(t))$$

The hidden states h(t) are continually obtained from the memory cells c(t) as the cells decay:

$$\mathbf{h}(t) = \mathbf{o}_i \odot (2\sigma(2\mathbf{c}(t)) - 1) \text{ for } t \in (t_{i-1}, t_i]$$

In the new process, these dynamics are controlled by a hidden state vector h(t), which in turn depends on a vector c(t) of memory cells in a continuous-time LSTM.

Model

Neural Hawkes process

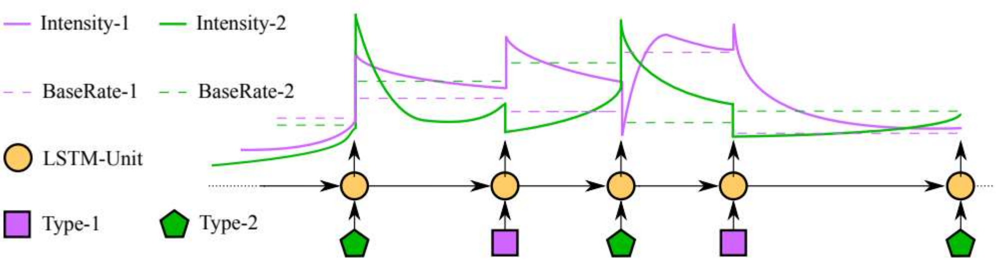
How does the continuous-time LSTM make those updates?

$$\mathbf{i}_{i+1} \leftarrow \sigma\left(\mathbf{W}_{\mathrm{i}}\mathbf{k}_i + \mathbf{U}_{\mathrm{i}}\mathbf{h}(t_i) + \mathbf{d}_{\mathrm{i}}\right) \quad (5a)$$
$$\mathbf{f}_{i+1} \leftarrow \sigma\left(\mathbf{W}_{\mathrm{f}}\mathbf{k}_i + \mathbf{U}_{\mathrm{f}}\mathbf{h}(t_i) + \mathbf{d}_{\mathrm{f}}\right) \quad (5b)$$
$$\mathbf{z}_{i+1} \leftarrow 2\sigma\left(\mathbf{W}_{\mathrm{z}}\mathbf{k}_i + \mathbf{U}_{\mathrm{z}}\mathbf{h}(t_i) + \mathbf{d}_{\mathrm{z}}\right) - 1 \quad (5c)$$
$$\mathbf{o}_{i+1} \leftarrow \sigma\left(\mathbf{W}_{\mathrm{o}}\mathbf{k}_i + \mathbf{U}_{\mathrm{o}}\mathbf{h}(t_i) + \mathbf{d}_{\mathrm{o}}\right) \quad (5d)$$

$$\mathbf{c}_{i+1} \leftarrow \mathbf{f}_{i+1} \odot \mathbf{c}(t_i) + \mathbf{i}_{i+1} \odot \mathbf{z}_{i+1} \quad (6a)$$
$$\bar{\mathbf{c}}_{i+1} \leftarrow \bar{\mathbf{f}}_{i+1} \odot \bar{\mathbf{c}}_i + \bar{\mathbf{i}}_{i+1} \odot \mathbf{z}_{i+1} \quad (6b)$$
$$\boldsymbol{\delta}_{i+1} \leftarrow f\left(\mathbf{W}_{\mathrm{d}}\mathbf{k}_i + \mathbf{U}_{\mathrm{d}}\mathbf{h}(t_i) + \mathbf{d}_{\mathrm{d}}\right) \quad (6c)$$

Model

Generalize the Hawkes process by determining the event intensities from the hidden state of a recurrent neural network.



Model

Log-likelihood

$$\ell = \sum_{i:t_i \leq T} \log \lambda_{k_i}(t_i) - \underbrace{\int_{t=0}^{T} \lambda(t)dt}$$

The sum of the log-intensities of the events that happened at the times they happened, minus an integral of the total intensities over the observation interval [0, T]

Model

Model Training

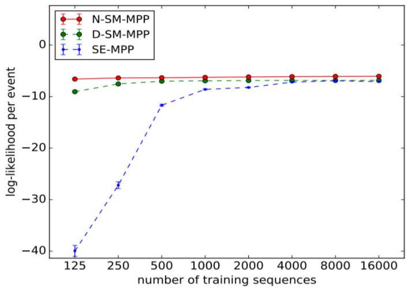
1. Using single-layer LSTM (Graves, 2012) in model. Empirically found that the model performance is robust to a set of hyperparameters.

2. Adam algorithm with its default settings (Kingma and Ba, 2015) as optimization algorithm.

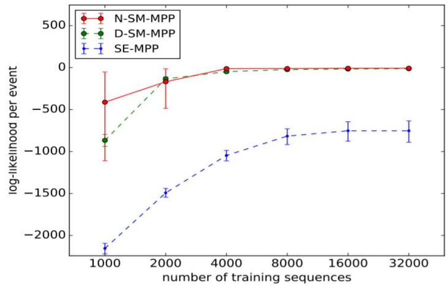
[1]Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), 2015.
[2]Alex Graves. Supervised Sequence Labelling with Recurrent Neural Networks. Springer, 2012.

Experiments

Real-World Media Datasets



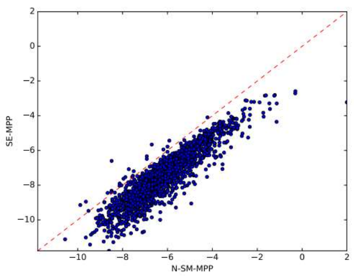
The MemeTrack contains time-stamped instances of meme use in articles and posts from 1.5 million different blogs and news sites. The event types correspond to the different websites. K=5000 .

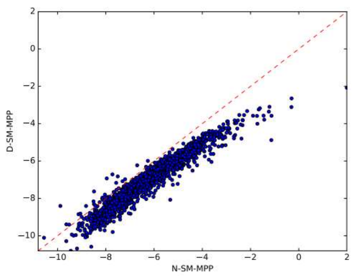


Retweets dataset: includes 166076 retweet sequences, each corresponding to some original tweet. They divide the events into K = 3 types: retweets by "small," "medium" and "large" users.

Experiments

Comparing the held-out log-likelihood of the two models





Nearly all points fall to the right of y = x, since N-SM-MPP (the neural Hawkes process) is consistently more predictive than the Hawkes process.

Conclusion

Presented a neural model of the multivariate Hawkes process.

1. Past events may now either excite or inhibit future events.
They do so by sequentially updating the state of a novel continuous-time recurrent neural network.

2. Hawkes sums the time-decaying influences of past events, they instead sum the time-decaying influences of the LSTM nodes

Louisiana State University

Thanks!