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

Hongyuan Mei and Jason Eisner
Center for Language and Speech Processing, Department of Computer Science, Johns Hopkins University

Overview

Events happen at random times \(0 < t_1 < t_2 \ldots\)
At time \(t_i\), there occurs an event of type \(k_i \in \{1, 2, \ldots, K\}\)
Given past events, what might happen next, and when?
- Generative model \(P((k_1, t_1), (k_1, t_1), \ldots, (k_{i-1}, t_{i-1}))\)
- Medical: patient’s visits, tests and diagnoses
- Online shopping: purchasing and feedback
- Social media: posts, shares, comments
- Other: quantified self, news, dialogue, music, etc

Traditional model is a Hawkes process
- Each event type has an intensity \(\lambda_k(t)\)
- Each event token occurs with probability \(\lambda_k(t)dt\)
- Past events temporarily excite future events
  \[\lambda_k(t) = \mu_k + \sum_{h:t_h < t} \alpha_{h,k} \exp(-\delta_{h,k}(t - t_h))\]

Model

An event stream from a neural Hawkes process
- Continuous-time LSTM
- Hidden state \(\mathbf{h}(t) = \mathbf{o}_t \odot (2\sigma(2\mathbf{c}(t)) - 1)\)
- Cell memory \(\mathbf{c}(t) = \mathbf{c}_{i+1} + \mathbf{v}_{i+1}(t)\)
- \(\mathbf{v}_{i+1}(t) = (\mathbf{c}_{i+1} - \mathbf{c}_{i+1}) \exp(-\delta_{i+1}(t - t_i))\)
- Extra gates to compute \(\mathbf{c}_{i+1}\) and \(\delta_{i+1}\)

Neural Hawkes process: \(\lambda_k(t) = f_k(\mathbf{w}_k^T \mathbf{h}(t))\)

Train the model by maximizing log-likelihood
\[\ell = \sum_{t_i \leq T} \log \lambda_k(t_i) - \int_{t=0}^T \lambda(t)dt\]
- Total intensity \(\lambda(t) = \sum_{k=1}^K \lambda_k(t)\)
- Integral estimation by Monte Carlo simulation
- Minimum Bayes Risk prediction
- Density for \(t_i\) is \(p_t(t) = \lambda(t) \exp(-\int_{t_{i-1}}^t \lambda(s)ds)\)
- Time prediction \(t_i = \int_{t_{i-1}}^{\infty} p_t(t)dt\)
- Type prediction \(k_i = \arg \max_{k} \int_{t_{i-1}}^{\infty} p_t(t)\lambda_k(t)/\lambda(t)dt\)

Thinning algorithm for sampling sequences

Experiments (more in paper)

Experiments on artificial datasets
- Models try to fit data generated by each other
- Oracle model performance —

Experiments on real-world social media datasets
- Retweet (top): long sequences with \(K = 3\)
- MemeTrack (bottom): short sequences with \(K = 5000\)

Neural Hawkes process vs. similar work