Conditional Generative Modeling is All You Need for Marked Temporal Point Processes

Ke Wan

September 20, 2023

Introduction to Key Concepts

1. Conditional Generative Models (CGM)

- Definition: A generative model conditioned on some observed variables to produce data samples.
- Application: Utilized to produce data that fits specific contexts or scenarios.
- Key Feature: Offers the flexibility of capturing conditional distributions.

2. Marked Temporal Point Processes (MTPP)

- Definition: A random process where each event happens at a particular time and has an associated mark (or label).
- Application: Suitable for modeling complex events in time, such as social network actions or system logs.
- Key Feature: Captures both the timing and type (mark) of events, offering rich information about sequences.

Motivation

Shortcomings of Intensity-Based Neural Temporal Point Processes

1. Likelihood functions contain an integral term over the intensity function, which can be computationally challenging.
2. For point process data with high-dimensional marks (e.g., images, text), modeling becomes infeasible.
3. The thinning algorithm introduces high computational complexity for sampling.

Motivation: Advantages of CGM

Benefits of Conditional Generative Modeling

1. More efficient in generating samples.
2. No parametric assumptions, leading to stronger representative power.
3. No need for integral computations, making optimization more convenient.

Components of the Generative Model

Two Key Components

This generative modeling framework comprises two main elements:

1. **Conditional Event Generator**:
   - Dynamically generates events based on given conditions.
   - Leverages deep neural networks for flexible event generation.
2. **Conditional Probability Estimation Model**:
   - Estimates the conditional probabilities for generated events.
   - Ensures generated events align with the true underlying event distribution.

Harmonious Integration

Together, these two components ensure that the generative model can produce events that are both diverse and representative of the actual data distribution.

Model Methodology

Conditional Event Generator

Let us define the conditional event generator that satisfies the following equation:

$$g_{(z,h_{i-1})} : \mathbb{R}^{r+p} \to (0,\infty) \times M$$

Where:

- $g()$ is the conditional generator.
- $h_i = \phi(x_i, h_{i-1})$ with $h_0 = 0$.
- h follows an RNN structure.
- An event x_i is given by $x_i = (t_{i-1} + \Delta t, m_i)$.
- $z \in \mathbb{R}$ and follows $\mathcal{N}(0, I)$.

Advantages of the Generative Model

1. Efficiency

Compared to the thinning algorithm with time complexity $O(N^d + N_T^d)$, this approach has a lower complexity of $O(N_T)$.

2. Strong Representational Power

The generation process relies entirely on the training of neural networks without any parametric assumptions, ensuring high representational power.

3. High Stability

Predictions can be made more stable by taking the average of multiple sampling results.

Learning the Generative Model

Learning Objective

Our goal is to optimize the following loss function:

$$l(\theta) = \frac{1}{K} \int \log f_\theta(x|H)\, dN_k(x)$$

Notations

- $l(\theta)$: The objective function to be optimized.
- $f_\theta(x|H)$: Conditional probability of event x given history H.
- $N_k(x)$: The counting process of the k th event sequence.
- K: Total number of observed events.

Estimation Method 1: Kernel Density Estimation

Introduction to Kernel Density Estimation

Kernel density estimation (KDE) is a non-parametric method to estimate the probability density function of a continuous random variable. It works by placing a kernel on each data point and then summing the results.

Mathematical Formulation

Given the set of observed samples $x_1, x_2, \ldots, x_L$, the kernel density estimate is given by:

$$f_\theta(x|H) = \frac{1}{L}\sum_{i=1}^{L} \kappa_\delta(x - x_i)$$

Where κ_δ is a kernel function with bandwidth δ.

Key Points: Kernel Density Estimation

Kernel Choice

- Various types of kernels available: Gaussian, Epanechnikov, uniform, etc.
- The choice affects smoothness and shape of the estimated density.

Advantages and Limitations

- No prior assumptions about the form of the distribution.
- Adaptable to multimodal distributions.
- Requires storage of all data points (can be computationally expensive).
- Curse of dimensionality: KDE becomes less reliable in high-dimensional spaces.

CVAE with RNN Embedding

Conditional Variational Autoencoder (CVAE)

- An extension of Variational Autoencoder (VAE).
- Allows conditional generation by conditioning on certain variables.
- More flexible in modeling intricate data distributions.

Embedding in RNN

- RNN captures temporal dependencies in sequential data.
- Embedding layer in RNN transforms discrete input data into dense vectors.
- This dense vector (embedding) serves as the conditioning variable in CVAE.

Model Training via ELBO

ELBO for CVAE (Evidence Lower BOund)

- The objective in training a CVAE.
- Balances reconstruction accuracy against regularization of the latent space, with an additional conditioning term.
- Given by:
  $\mathrm{ELBO} = \mathbb{E}_{q_\phi(z|x,h)}[\log p_\theta(x|z,h)] - \mathrm{KL}(q_\phi(z|x,h)||p(z|h))$

Training the Model

- Optimize the ELBO using gradient-based methods.
- This ensures the training of both the inference (encoder) and generative (decoder) components of the CVAE.
- The embedding from the RNN serves as a context (y) in the CVAE, enhancing its generative capacity.

Value Proposition of the Model

Simplicity and Power

- Elegance in approach: Achieves complexity through simple means.
- Effective for generating point processes in intricate label spaces.

Potential in the Realm of Large Models

- Scales gracefully: Minimal overheads increase potential applicability in expansive architectures.
- Fosters a foundation: Encourages further exploration and developments in complex model domains.

Thank You for Watching!

Any Questions?