Point Process Project

Description

With the development of deep learning, the neural temporal point process is becoming more and more popular. Neural point process model is good at prediction, while it is weak on interpretability. The project is to build a Interpretable Neural Temporal Point Process model, which is good at prediction and interpretability simultaneously. The goal of this project is to strengthen the understanding of neural point process and bridge between statistical point process and neural point process. The experiments will be executed on synthetic sequences which are generated in homework1 and ATMs dataset.

Background

Recurrent Temporal Point Process (RMTPP)

Traditional models either suffer from model misspecification if the chosen parametric intensity function does not fit with the real behavior of the event data, or the learning algorithm can be mathematically very complex for nonparametric models. With the fast development of deep learning theory and techniques, especially for recurrent neural network models, there is a trend for adapting neural networks to temporal point process learning. Neural point process frees the need for selecting explicit parametric intensify form, thus the neural point process models shows high model capacity for learning arbitrary and unknown distributions. Recurrent temporal point process (RMTPP) as the first work on neural point process has a great impact on the further research.

RMTPP takes the timestamp and event type as input, and then predict the timestamp and event type of the next event. The model architecture is shown as Figure 1. Base on hidden output h_j, we can reconstrcut the conditional intensity function:

$$\lambda^*(t) = \exp\left(\underbrace{\boldsymbol{v}^{t\top} \cdot \boldsymbol{h}_j}_{\substack{\text{past}\\ \text{influence}}} + \underbrace{w^t(t - t_j)}_{\substack{\text{current}\\ \text{influence}}} + \underbrace{b^t}_{\substack{\text{base}\\ \text{intensity}}}\right),$$

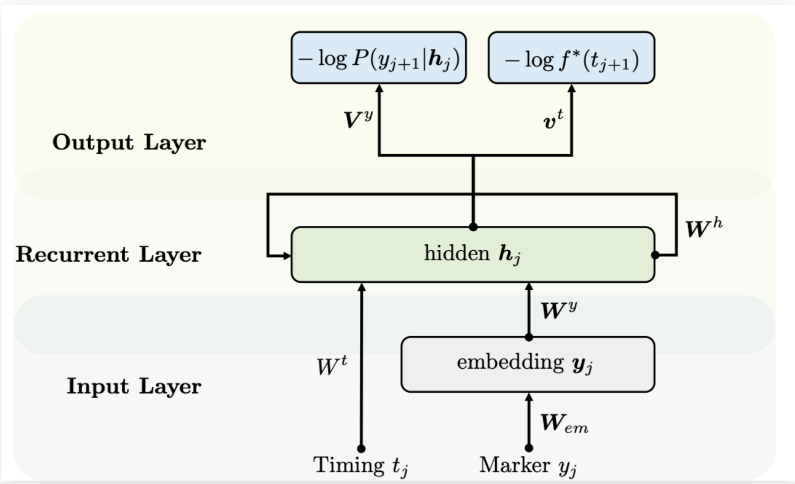
Then we can derive the likelihood that the next event will occur at the time t given the history by the following equation:

$$
\begin{aligned}
f^*(t) &= \lambda^*(t)\exp\left(-\int_{t_j}^{t}\lambda^*(\tau)d\tau\right) \\
&= \exp\left\{\boldsymbol{v}^{t\top}\cdot\boldsymbol{h}_j + w^t(t-t_j) + b^t + \frac{1}{w}\exp(\boldsymbol{v}^{t\top}\cdot\boldsymbol{h}_j + b^t) \right. \\
&\left. -\frac{1}{w}\exp(\boldsymbol{v}^{t\top}\cdot\boldsymbol{h}_j + w^t(t-t_j) + b^t)\right\}. \qquad (12)
\end{aligned}
$$

In training stage, we use cross entropy loss and negative log-likelihood loss for event types prediction and event timestamp prediction respectively. While in testing stage, , we can estimate the timing for the next event using the expectation:

$$\hat{t}_{j+1} = \int_{t_j}^{\infty} t \cdot f^*(t)dt.$$

Figure 1 RMTPP model architecture



Interpretable Neural Temporal Point Process

Through neural temporal point process models shows great potential for their high model capacity and flexibility, these networks often lack clear interpretability enjoyed by the traditional parametric models such as Hawkes process. Interpretable neural temporal point process (INTPP) bridges between neural temporal point process and traditional point process. The form of INTPP's intensity function is as follow:

$$\lambda^*(t) = \exp\left(l_j + w(t - t_j)\right) + c$$

Then the conditional density function for next event can be written as:

$$f^*(t) = \exp\left(l_j + w(t - t_j) + c\right) \cdot \exp\Bigg(\frac{1}{w}\exp(l_j) - \frac{1}{w}\exp\big(l_j + w(t - t_j)\big) - c(t - t_j)\Bigg).$$

While the event types is regarded as a multinomial distribution:

$$f^*_{\text{mark}}(y_{j+1} = k) = \frac{\exp\left(\mathbf{V}^y_{k,:}\mathbf{h}'_j + b^y_k\right)}{\sum_{k=1}^{K}\exp\left(\mathbf{V}^y_{k,:}\mathbf{h}'_j + b^y_k\right)}$$

In traning stage, the timestamp prediction loss and event types loss are as follow:

$$L_{time} = \sum_i \sum_j \log(f^*_{\text{time}}(t^i_{j+1})),$$

$$L_{mark} = \sum_i \sum_j \log(f^*_{\text{mark}}(y^i_{j+1} | \mathbf{h}_j, t^i_{j+1}))$$

Once the model training is completed, INTPP with its intensity can be interpreted in the view of Hawkes process. The intensity of Hawkes process can be written exactly as:

$$\lambda^*(t) = \mu + (\alpha \sum_{k=1}^{j} e^{\beta \cdot t_k})e^{-\beta t},$$

By comparing the intensity function of Hawkes process and INTPP, there are three connections between Hawkes process and INTPP's hidden outputs:

$$\begin{cases} c = \mu, w = -\beta, \\ L_j = \exp\left(l_j - w \cdot t_j\right) = \alpha \sum_{k=1}^{j} e^{\beta \cdot t_k}, \end{cases}$$

Then we can estimate Hawkes parameters as follow:

$$\begin{cases} \hat{\mu} = c, \hat{\beta} = -w, \\ \hat{\alpha} = \frac{1}{n} \sum_j \frac{\exp(l_j)}{\sum_{k=1}^{j} \exp(w \cdot (t_j - t_k))}, \end{cases}$$

Moreover, this method can be easily extended to Multi-dimensional Hawkes.

Task

You should implement INTPP model and experiment on synthetic dataset in homework1 and ATMs dataset. In synthetic dataset, you should simulate 3000 event sequences for training, and then estimate the parameter of Multi-dimensional Hawkes process. In ATMs dataset, you should train INTPP in training data, and then predict event timestamp and event types in testing data.

More implementation details can be accessed in the paper "Improving Interpretability and Predictive Performance of Neural Temporal Point Processes".

In report, you should describe the software anguage, key libraries, and other things necessary to run the program. Then you should explain what kind of problems are encountered and how to solve them. As for the results of model, you should you should display the actual results of the program and analyze them.

Dataset

- Synthetic dataset in homework1, 3000 event sequences.
- ATMs dataset.

Metrics

In synthetic dataset:

- Mean Absolute Error of hawkes parameters: $\boldsymbol{\mu}$ and $\boldsymbol{A}$. Taking a_{ij} of $\boldsymbol{A}$ as an example, the error of one parameter is as follow:

$$\text{error}(a_{ij}) = \begin{cases} \dfrac{|a_{ij} - a_{ij}^*|}{a_{ij}} & a_{ij} \neq 0 \\ |a_{ij} - a_{ij}^*| & a_{ij} = 0 \end{cases}$$

In ATMs dataset:

- Event Type Prediction：Precision, Recall, macro-F1 Score over 7 event types under. Note all these metrics are computed for each type, and then averaged over all types.
- Event Time Prediction: Mean Absolute Error (MAE) which measures the absolute difference between the predicted time point and the actual one.

Submission

You should submit a .zip file with all source code and the report.

Grade

Source code(50%): model results, code norm, program running

Report(50%): well-organized and analytical

Note: new creative exploration can be regarded as a bonus.

Reference

1. Du N, Dai H, Trivedi R, et al. Recurrent marked temporal point processes: Embedding event history to vector[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1555-1564.
2. Improving Interpretability and Predictive Performance of Neural Temporal Point Processes

Academic misconduct and handling

We will review all source code and analyze the plagiarism of source code and reports, therefore, don't try to test its reliability. For self-completed student, there will be reward points. For plagiarized student, it will be awarded zero points. If you are convicted of plagiarism, you can ask TAs to make a defense.