Recurrent Point Process

Khoa Doan[1], nik90[2], chandankreddy[2], reddy[2], and sathappan[2]

[1]Virginia Tech
[2]Affiliation not available

February 21, 2018

1 Problem Formulation

Given an observed history of a user's time-ordered trips $S_u = \{T_{u1}, T_{u2}, ..., T_{un}\}$, the task is to predict when the user would likely to make user the metro again. A trip T_{ui} is a tuple of (t_{ui}, m_{ui}, c_{ui}), where t_{ui} is the timestamp of trip, m_{ui} is information about the trip, such as the length of the trip and the stations where the user enters and exits, and c_{ui} is any contextual information around the time of the trip such as whether there is a train incident affecting the trip T_{ui}.

2 Proposed Model

2.1 Recurrent Point Process

In this section, we drop the subscript u and discuss all variables as with respect to a specific user u. We proceed by considering a commuter's historical usage as only the times when the trips happen. In other words, $S_u = \{t_1, t_2, ..., t_n\}$. We denote any variable about an observation at t_i by the subscript i; for example, x_i and h_i are the input observation at time t_i, and the hidden representation also at t_i as a result of that observation.

In the point process framework, the conditional density function at a time t given past observations can be defined as:

$$f^*(t) = \lambda^*(t) \exp(-\int_{t_n}^{t} \lambda^*(s) ds) \tag{1}$$

Given a sequence of n observations, the hidden representation at the current observation n^{th} is a function of the past hidden representation and the current input observation x_n:

$$h_n = h(h_{n-1}, x_n) = W_{hh} h_{n-1} + W_{ih} x_n \tag{2}$$

The conditional intensity function after observing an occurence at t_n is then a function of the vector reprsentation of the past activties and the current time t:

$$\lambda^*(t) = l(h_n, t - t_n, \lambda_0) = \exp(w_{ht}^T h_{t_n} + w_t(t - t_n) + \lambda_0) \tag{3}$$

Because a commuter can travel with different intensity at different days, it can be insufficient for a hidden vector after a particular observation to capture this variation. In addition, different commuters have different varations in their travel activities, which makes it inefficient to define a universal rule that can represent this cross user variation; for example, defining that all users have the conditional intensity as also a function of weekday or weekend. Therefore, it is useful if the model can automatically determine which of the previous days should contribute more to the current intensity functions. We introduce the attention mechanism by extending the conditional intensity function to include past hidden states:

$$\lambda^*(t) = l(h_{n-w}, ..., h_n, t - t_n, \lambda_0) = \exp(w_{ht}(\sum_{w=1}^{W} \alpha_{n-w} h_{n-w}) + w^t(t - \tag{4}$$

where the normalized vector α_n can be seen as the alignment of the current intensity function to the past set of hidden representations. In other words, the model chooses from which of the past activities to attend to when determining the conditional intensity at the current time step. One possible definition of α can use the softmax function as following:

$$\alpha_{n-w} = \frac{\exp(e_{n-w})}{\sum_{k=1}^{W} \exp(e_{n-k})} \tag{5}$$

where $e_{n-k} = a(h_{n-k}, x_{n-k})$ is a function of the hidden representation at time t_{n-k} and the input x_{n-k} of that time step. Essentially, the higher score e_{n-k} has, the higher influence that the observation at t_{n-k} has on the current conditional intensity. One of the hyperparameter of this approach is the window W, whose value allows us to give preference on how many previous observations we consider, with all observations prior to this window as having no influence on the current conditional intensity function.

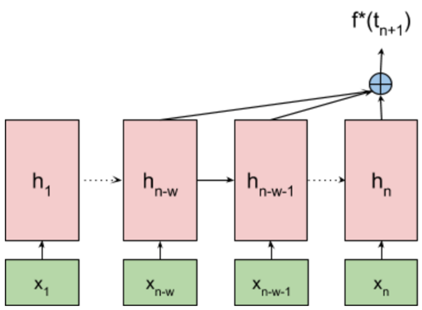

Figure 1: Recurrent point process with attention.

2.2 Temporal Input

At each trip, we observe the time when the trip occurs.