University College London

Doctoral Thesis

Modeling Temporal point processes using Recurrent Neural Nets

Author:
Badrul Alom

Supervisor:
Dr. Mark Herbster
Advised by:
Pedro Mediano (Emotech Ltd.)

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Research Group Name
Department of Computer Science

July 27, 2017

Declaration of Authorship

I, Badrul ALOM, declare that this thesis titled, "Modeling Temporal point processes using Recurrent Neural Nets" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

University College London

Abstract

Faculty Name
Department of Computer Science

Master of Science

Modeling Temporal point processes using Recurrent Neural Nets

by Badrul ALOM

tbc

Chapter 1

Introduction

Event sequences are common in areas such as retail, finance, and utilities. For example, the times at which a customer makes a purchase from an online retailer, the bidding and asking orders in the stock market, and the times at which utility service customers make high use of gas and electricity. These are a particular category of event sequences in which the temporal factor could be said to be a key dimension in predicting the next occurence. In such cases, understanding and predicting user behaviors, based on purely the temporal aspect of their occurence, are of great practical, economic, and societal interest.

1.1 Motivation

The impetus for this research was to estimate the liklihood of a person being interested in listening to music in the current time-period based only on their listening history. The raw data is a series of timestamps denoting when songs were played. The implicit assumption is that a persons playlist history contains a temporal pattern such as a combination of daily and weekly schedule, that can be modelled. This concept can also be applied to other areas where a temporal pattern is thought to exist at an individual user level, such as the repeat purchase of household products, or the sleeping and eating habits of a person.

The objective of the research is to evaluate the effectiveness of several different techniques for determining the probability of a user listening to music in a given period. One such application of this would be home audio devices which could anticipate when a user would like to listen to music and play without user intervention.

As this is research is in the context of Data Science, it will focus on contrasting simple methods that can be easily explained and implemented, with more advanced methods.

The research was guided by Emotech Ltd. the creators of Olly [2].

1.2 Point Processes

One way of modeling the problem is as series of events and non-events known as a temporal point process. This has a rich history of methods as outlined in the literature review.

1.3 The dataset

The dataset being used in this analysis is the LastFM1k dataset containing the listening history of a thousand LastFM listeners.

The dataset contains the timestamp, userId, and trackId of users listening habits over a number of years (2005-2009).

1.4 Structure of the report

tbc

Chapter 2

Literature Review

Within literature the most common way of tackling the problem is through Point Processes. More recently both Gaussian Processes and Deep Learning have been applied.

2.1 Point Processes

A temporal point process is a sequence of events t_i with t being a sequence of a fixed period inteveral with $t_i \epsilon R + and i \epsilon Z+$. It can be modelled as a series of inter-event times (time until next event) or the number of events occurring in the interval. Examples of point processes are the occurence of natural disasters, machine failure, or when a customer engages with a company.

At its simplest a point process can be represented as:

$$\xi = \sum_{i=1}^{n} \delta_{X_i},$$

where δ denotes the Dirac measure, a probability measure of whether a set contains point x or not.

Point processes seek to model the probability of an event happening at time t, based on the event history upto, but not including, time t. Each point can either be seen as i.i.d. or as with the Wold process, dependent on the previouse inter-event time. There are different ways of representing point process data as shown in figure 2.1, with the inter-event time being the most common. In this form, the Poisson distribution is the most obvious choice for a conditional intensity function.

Conditional Intensity Function

A condtional intensity function $\lambda(t)$ based on parametric statistics, such as a Poisson process for inter-event times, is used to determine the probability.

Expectation measure

The expectation measure E for a point process ξ is the number of events p expected for every Borel subset of the event space S.

$$E\xi(B) := E(\xi(B)) \quad \text{for every } B \in \mathcal{B}$$

2.2 Gaussian Processes

2.3 Deep Learning

How it's referred to in lit: * Event sequences

Methods:

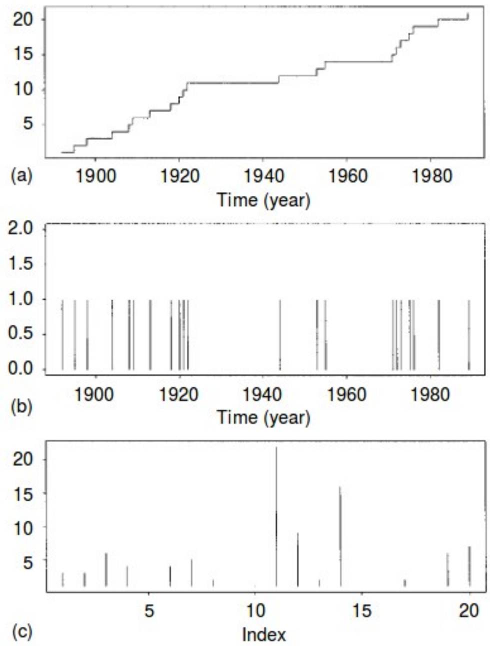

FIGURE 2.1: Three different representations of the same point-process a) cumulative count b) date of occurence c) interval time between floods

Process Models Temporal Points Processes The point process representation of sequence data is fundamentally different from the discrete time representation typically used in time series analysis. It directly models the time period between events as random variables, and allows temporal events to be modeled accurately, without requiring the choice of a time window to aggregate events, which may cause discretization errors. Moreover, it has a remarkably extensive theoretical foundation [6]. However, conventional point process models often make strong unrealistic assumptions about the generative processes of the event sequences. In fact, a point process is characterized by its conditional intensity function – a stochastic model for the time of the next event given all the times of previous events. The functional form of the intensity is often designed to capture the phenomena of interests. Some examples are homogeneous and non-homogeneous Poisson processes [7], self-exciting point processes [8], self-correcting point process models [9], and survival processes [6]. Unfortunately,they make various parametric assumptions about the latent dynamics governing the generation of the observed point patterns. As a consequence, model misspecification can cause significantly degraded performance using point process models, which is also shown by our experimental results later.

To address the aforementioned problem, the authors in [10] propose to learn a general representation of the underlying dynamics from the event history without assuming a fixed parametric form in advance. The intensity function of the temporal point process is viewed as a nonlinear function of the history of the process and is parameterized using a recurrent neural network. Apparently this line of work still relies on explicit modeling of the intensity function. However, in many tasks such as data generation or event prediction, knowledge of the whole intensity function is unnecessary. we are able to demonstrate that Wasserstein distance training of RNN point process models outperforms the same architecture trained using MLE.i)

We propose the first intensity-free generative model for point processes and introduce the first (to our best knowledge) likelihood-free corresponding learning methods; ii)-3mm We extend WGAN for point processes with Recurrent Neural Network architecture for sequence generation learning; iii) In contrast to the usual subjective measures of evaluating GANs we use a statistical and a quantitative measure to compare the performance of the model to the conventional ones. iv) Extensive experiments involving various types of point processes on both synthetic and real datasets show the promising performance of our approach. (Wasserstein Learning of Deep Generative Point)

Chapter 3

Methodology

The methodology employed is typical of that within the data science community, and as seen in online Kaggle competitions. Namely to start with preliminary analysis that will help understand the data better, then devise a test and training datasets that can be used across multiple models, with the main performance criteria beiong on how well the models perform on test data.

Python, via Jupyter notebooks was the primary source of analysis with some SQL as the data was manipualated and stored in a SqlLite3 database first.

Note: In the interest of time, analysis was carried out on 381 of the full 1000 user dataset.

3.1 Preliminary analysis

3.1.1 Context

Let us first consider the real-world aspect of the data we have - the timestamps on which users played a song. This does not necessarily mean they played the song in its entirety. Indeed initial analysis shows plenty of cases where a song was started, seconds after the previous one, suggesting that the dataset contains both tracks that were played and tracks that were skipped. For our purposes we can consider both these to be the same as they both indicate that the user was interested in playing music at time t.

We can also assume that the song plays are not i.i.d, in that the probability of a play event at time t+1 is significantly higher if there was an event at time t.

3.1.2 Basic statistics

We start with some basic information about the raw data file that was recieved. Num. of rows: 7,500,000 Num. of users: 381 Num. of unique timestamps: 7,226,934 earliest timestamp: 2005-02-14 00:02:10 latest timestamp: 2009-06-19 17:37:23

What does this tell us? Well we can deduce there are approximately 19.6k timestamps per user on average. This feels like a high number. Fig 2.1

3.1.3 Outlier analysis

3.2 Bayesian Inference

Here we apply a counting process

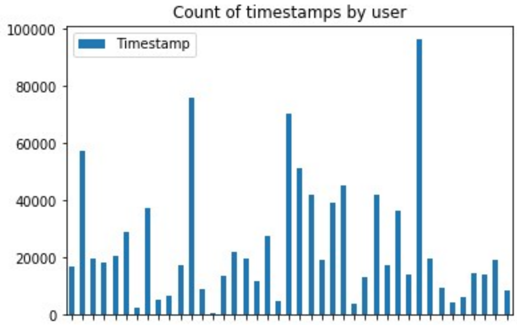

FIGURE 3.1

3.3 How to cite references in Latex

[1]

This document is an example of `thebibliography` environment using in bibliography management. Three items are cited: *The LaTeX Companion* book [**latexcompanion**], the Einstein journal paper [**einstein**], and the Donald Knuth's website [**knuthwebsite**]. The LaTeX related items are [**latexcompanion**, **knuthwebsite**].

Chapter 4

Chapter Title Here

4.1 Main Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

4.1.1 Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

4.1.2 Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

4.2 Main Section 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:
`\hypersetup{urlcolor=red}`, or
`\hypersetup{citecolor=green}`, or
`\hypersetup{allcolor=blue}`.
If you want to completely hide the links, you can use:
`\hypersetup{allcolors=.}`, or even better:
`\hypersetup{hidelinks}`.
If you want to have obvious links in the PDF but not the printed text, use:
`\hypersetup{colorlinks=false}`.

Bibliography

[1] C. J. Hawthorn, K. P. Weber, and R. E. Scholten. "Littrow Configuration Tunable External Cavity Diode Laser with Fixed Direction Output Beam". In: *Review of Scientific Instruments* 72.12 (Dec. 2001), pp. 4477–4479. URL: http://link.aip.org/link/?RSI/72/4477/1.

[2] Emotech Ltd. *Your robot with personality*. URL: https://www.heyolly.com/.