Forecasting User Demand of Uber with Wavenet

Long Chen

28/11/2018
Real-time Driver-Rider Matching

App Views

KAFKA

Stream Processing

- Driver-Rider Match
- ETA

Vehicle information
How does the Uber Matching algorithm work?

- Live traffic at specific times
- Distance to the rider
- Driver rating
- Customer rating
- Ride acceptance rates
- Destination
- Dynamic pricing
Why is the User Demand Forecasting so important?

- Intelligent Transportation System
- Surge Price Search
- Dynamic Pricing
Dataset

Over 20 million Uber pickups in NYC

Public Available at:

https://github.com/fivethirtyeight/uber-tlc-foil-response
Problem Statement: User Demand Forecasting
Frequency of surge by hours of day

Note: This figure shows the frequency of surges by hour of day and day of week for UberX. Darker rectangles identify times and days when riders are more likely to face surge pricing. Tuesday at 11am represents the time and day combination when surge pricing is least common, and Saturday at 11pm represents the time and day combination when surge pricing is most common.
Data: External Features
Current Solution: Cons

- **Classical forecasting methods**
  - ARIMA: Seasonality, Trend, Spikes
  - Exponential smoothing (Holt-Winters)

- **Cons of classical approaches**
  - Stationarity assumption
  - Accelerating trends, repeated irregular patterns
RNN

- Deep learning methods
  - RNN: Recurrent Neural Network

- Cons of RNN
  - Gradient Vanishing Issue
  - Computationally expensive
RNN
Figure 2: Visualization of a stack of causal convolutional layers.
WaveNet
Results: Experiments & Methodology

- Internal and public datasets
- Two years of data from NYC
- Target variable is completed trips.
- Records for holidays, weather
- Forecast is done one hour ahead.
- Measure SMAPE (Symmetric Mean Absolute Percent Error):

\[
SMAPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}
\]
Results: Example of a forecast (Testing)
Results: Special Event Prediction Performance (SMAPE)

<table>
<thead>
<tr>
<th>Event</th>
<th>WaveNet</th>
<th>ARIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christmas Day</td>
<td>11.1</td>
<td>29.2</td>
</tr>
<tr>
<td>MLK</td>
<td>8.7</td>
<td>20.2</td>
</tr>
<tr>
<td>Independence Day</td>
<td>2.8</td>
<td>17.6</td>
</tr>
<tr>
<td>Labor Day</td>
<td>2.9</td>
<td>6.9</td>
</tr>
<tr>
<td>New Year’s Day</td>
<td>6.8</td>
<td>7.8</td>
</tr>
<tr>
<td>Veteran’s Day</td>
<td>4.7</td>
<td>8.9</td>
</tr>
</tbody>
</table>
Resources:

- MIT Uber Study:
  https://qz.com/1222744/mits-uber-study-couldnt-possibly-have-been-right-it-was-still-important/

- WaveNet:
  https://jeddy92.github.io/JEddy92.github.io/
Thank you!

.... Questions?