Achieving personalization with LSTMs

Ankit Jain
Sr. Data Scientist, Uber AI Labs

Sept 2018
Uber heavily leverages ML for our business strategy and finance operations
Agenda

● Forecasting@Uber
● Problem Formulation
● Modeling
● Results
Forecasting@Uber

- Forecasting trips, gross bookings etc. are paramount to Uber

**Time Horizon:**
- Time horizon for forecasting varies from few minutes to a year depending on the application

**Space:**
- City
- Neighborhood
- Country etc.
Trip Forecasting@Uber Finance

**Historical Trips**
- Historical Trips data for a city

**Acquisition Spend**
- Marketing (acquiring new riders/drivers)

**Driver Incentives**

**Rider Promotions**

**Long-Term Forecasts**
- Upto 52 Weeks
- Used for year long budget planning

**Short-Term Rolling Forecasts**
- 1-12 Weeks
- Adjust budgets in spend levers to achieve trip targets

- Represents a spending lever
Business Requirements

Scenario Generation
Weekly/monthly Incentive planning integrated with trip forecasting

Deviation from Forecast
Which subset of users should we focus on to meet goal?

Impact of Product Launches
Quantifiable uplift in business after a product launch through counterfactuals

Combining insights like these can help Uber adjust its budget in the short term.
Agenda

- Forecasting@Uber
- Problem Formulation
- Modeling
- Results
Problem Formulation (Incentives)

Functional Form

Learn a function $g$ to establish the following relationship

$$g(I, f_t, Trips_t) = T$$

Where,

- $f_t = \text{vector of features from 0 to current time } t$
- $Trips_t = \text{vector of trips from 0 to current time } t$
- $I = \text{Budgeted incentive spend vector for next } N \text{ weeks}$
- $T = \text{Forecasted trips vector for next } N \text{ weeks}$
Agenda

- Forecasting@Uber
- Problem Formulation
- Modeling
- Results
## Data (Feature Vector)

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Driver Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static</strong></td>
<td>City, Acquisition channel, Acquired month etc.</td>
</tr>
<tr>
<td><strong>Behavioral</strong></td>
<td>Trips, Supply hours, Open Supply Hours, Accepts, Rejects, Cancellations, Earnings, Referrals, %Surge Trips etc.</td>
</tr>
<tr>
<td><strong>Incentives</strong></td>
<td>• Incentive features like DxGy, Guaranteed Surge etc.</td>
</tr>
<tr>
<td></td>
<td>• Look ahead features of budgeted incentives for the future</td>
</tr>
</tbody>
</table>

**Data Granularity:** Each row of data set corresponds to weekly features of a driver and N target variables (one for each of future weeks)
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Granularity</td>
<td>Separate model for each city</td>
</tr>
<tr>
<td>Evaluation Metric</td>
<td>Mean Absolute Error (MAE) of each driver for week 1-N MAPE at City level to check model if the model is unbiased</td>
</tr>
<tr>
<td>Baseline</td>
<td>Driver - use last week’s trips as prediction for next 1-N weeks</td>
</tr>
<tr>
<td>Best Model</td>
<td>LSTMs with Zero Inflated Poisson Loss</td>
</tr>
</tbody>
</table>
LSTMs vs Classical ML Models

- The abundance of Uber’s data serves very well for training data hungry deep learning models like LSTMs
- Time dependencies are well captured
- LSTMs model hidden non-linear interactions

Images Source: [http://colah.github.io/posts/2015-08-Understanding-LSTMs/](http://colah.github.io/posts/2015-08-Understanding-LSTMs/)
Zero Inflated Poisson (ZIP) Loss

Weekly trips distribution (i.e., the y variable) has the following properties:

- Non-negativity: Trips is a count data
- Lots of structural zeros in the system

Hence zero-inflated-poisson (ZIP):

\[
\mathbb{P}(y = 0) = \pi + (1 - \pi)e^{-\lambda}
\]

\[
\mathbb{P}(y = h) = (1 - \pi)\frac{\lambda^h e^{-\lambda}}{h!}, h \geq 1
\]

Where,

* $\lambda$ is the expected Poisson count for the individual;
* $\pi$ is the probability of extra zeros
### LSTM Model Architecture

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_t$</td>
<td>Static, behavioral features at time $t$ for a driver</td>
</tr>
<tr>
<td><strong>LSTM 1</strong></td>
<td><strong>Output</strong>: Feature embeddings per driver</td>
</tr>
</tbody>
</table>
| **LSTM 2** | **Input**: Feature embeddings and future incentives ($I_{t+1} \ldots I_{t+n}$)  
**Output**: Updated feature embeddings at a given forecasting period |
| $\lambda, \pi$ | Parameters of ZIP distribution per driver per week |
| **FC** | Fully connected layer to map the feature embeddings to two outputs |
| $y_1 \ldots y_n$ | Final trip predictions as expectation of ZIP |

### Diagram Explanation

- **$F_t$** represents static, behavioral features at time $t$ for a driver.
- **LSTM 1** outputs feature embeddings per driver.
- **LSTM 2** takes as input feature embeddings and future incentives ($I_{t+1} \ldots I_{t+n}$) and outputs updated feature embeddings at a given forecasting period.
- $\lambda, \pi$ are the parameters of the ZIP distribution per driver per week.
- **FC** maps the feature embeddings to two outputs.
- $y_1 \ldots y_n$ are the final trip predictions as expectation of ZIP.

The diagram illustrates the flow of data through the model, with arrows indicating the direction of information flow from input to output.
Agenda

- Forecasting@Uber
- Problem Formulation
- Modeling
- Results
Mean Absolute Error (MAE) Comparison

LSTM ZIP outperforms the basic models (~30% improvement)
Mean Absolute Percentage Error (MAPE)

City A

City B

MAPE is similar to the benchmark indicating Neural Network model is not biased
Incentive Sensitivity Curves

Incentive Sensitivity For A City

*Note that values are rescaled*
Thank you

Ankit Jain | ankit.jain@uber.com

Data Science | Uber AI Labs

We are hiring!
Machine Learning Engineers
Software Engineers
Data Scientists

https://www.uber.com/careers