Achieving personalization with LSTMs

Ankit Jain
Sr. Data Scientist, Uber AI Labs

Sept 2018

UBER

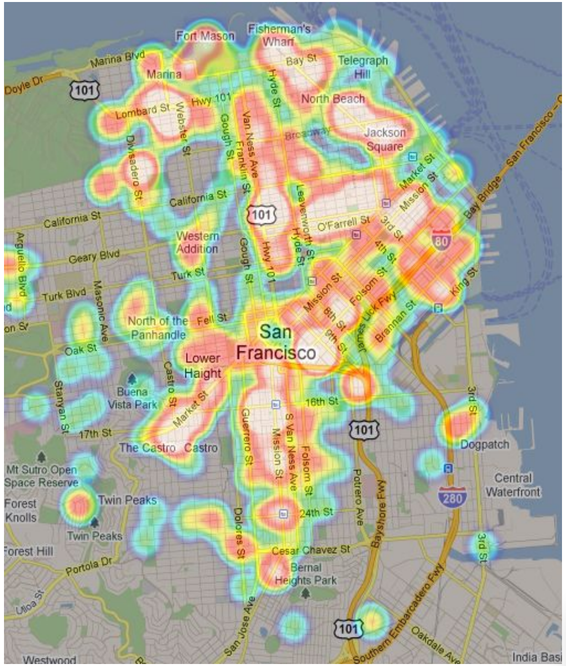
Uber heavily leverages ML for our business strategy and finance operations

Agenda

- Forecasting@Uber
- Problem Formulation
- Modeling
- Results

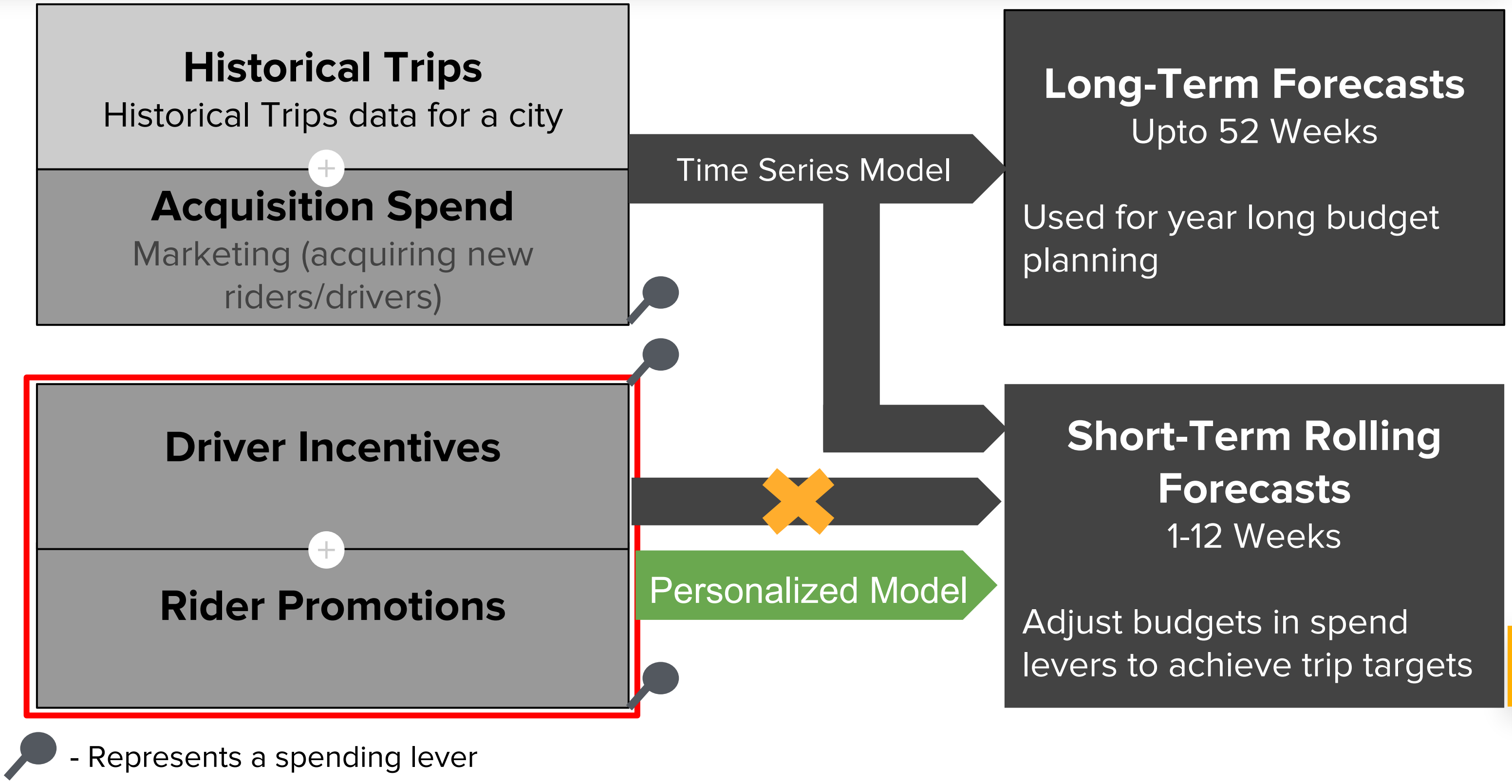
Forecasting@Uber

- 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.



Map Data@2018 Google at Uber

Trip Forecasting@Uber Finance



Business Requirements

Scenario Generation

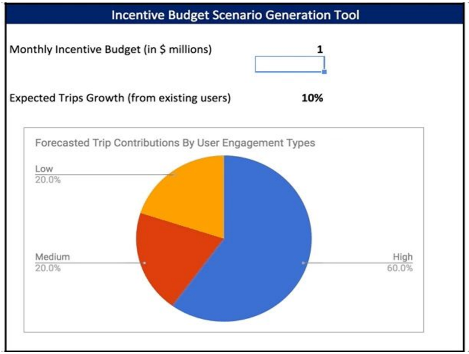
Weekly/monthly Incentive planning integrated with trip forecasting

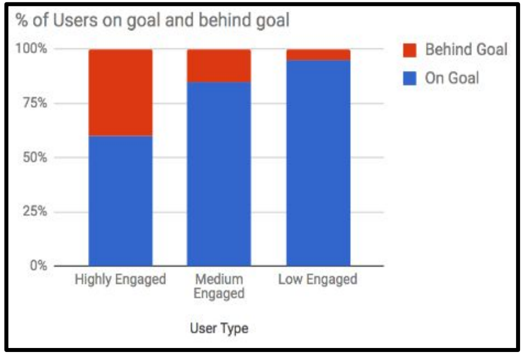
Deviation from Forecast

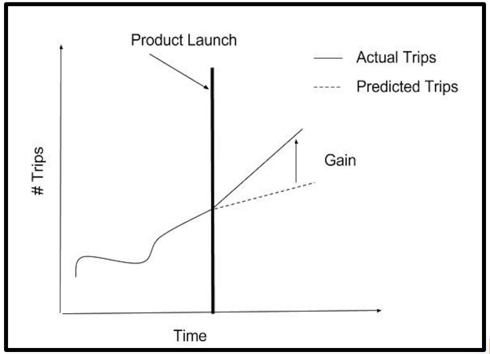
Which subset of users should we focus on to meet goal?

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

$$g(\mathbf{I}, f_t, Trips_t) = \mathbf{T}$$

Where,

f_t = vector of features from 0 to current time t

$Trips_t$ = vector of trips from 0 to current time t

$\mathbf{I}$ = Budgeted incentive spend vector for next *N* weeks

$\mathbf{T}$ = Forecasted trips vector for next *N* weeks

Agenda

- Forecasting@Uber
- Problem Formulation
- Modeling
- Results

Data (Feature Vector)

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

Data Granularity: Each row of data set corresponds to weekly features of a driver and N target variables (one for each of future weeks)

Modeling Parameters

Parameters	Value
Model Granularity	Separate model for each city
Evaluation Metric	Mean Absolute Error (MAE) of each driver for week 1-N MAPE at City level to check model if the model is unbiased
Baseline	Driver - use last week's trips as prediction for next 1-N weeks
Best Model	**LSTMs with Zero Inflated Poisson Loss**

LSTMs vs Classical ML Models

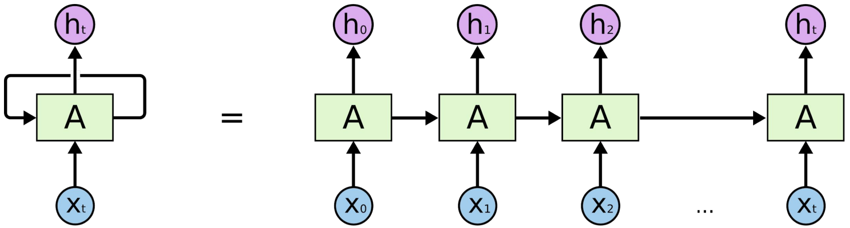


Fig: An unrolled recurrent neural network

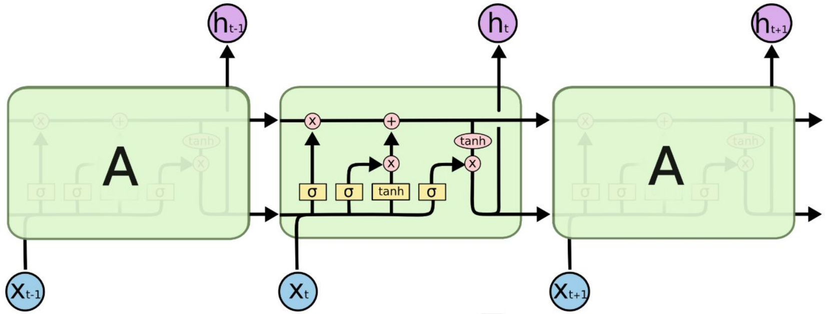


Fig: Repeating Module of LSTM

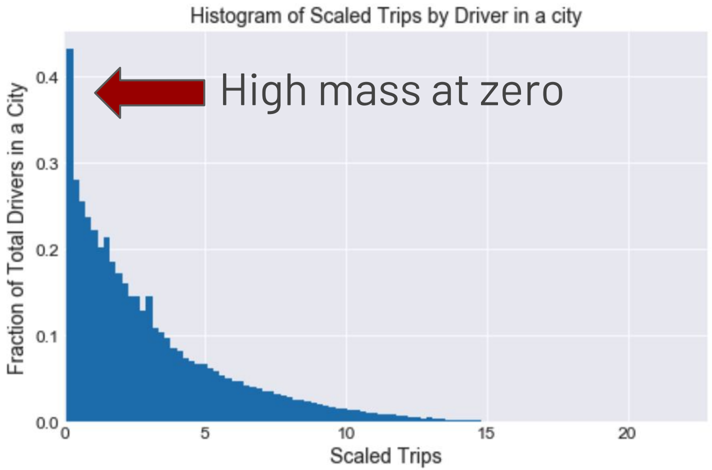
- The abundance of Uber's data serves very well for training data hungry deep learning models like LSTMs
- Time dependencies are well captured
- LSTMs model hidden non-linear interactions

Images Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

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):

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

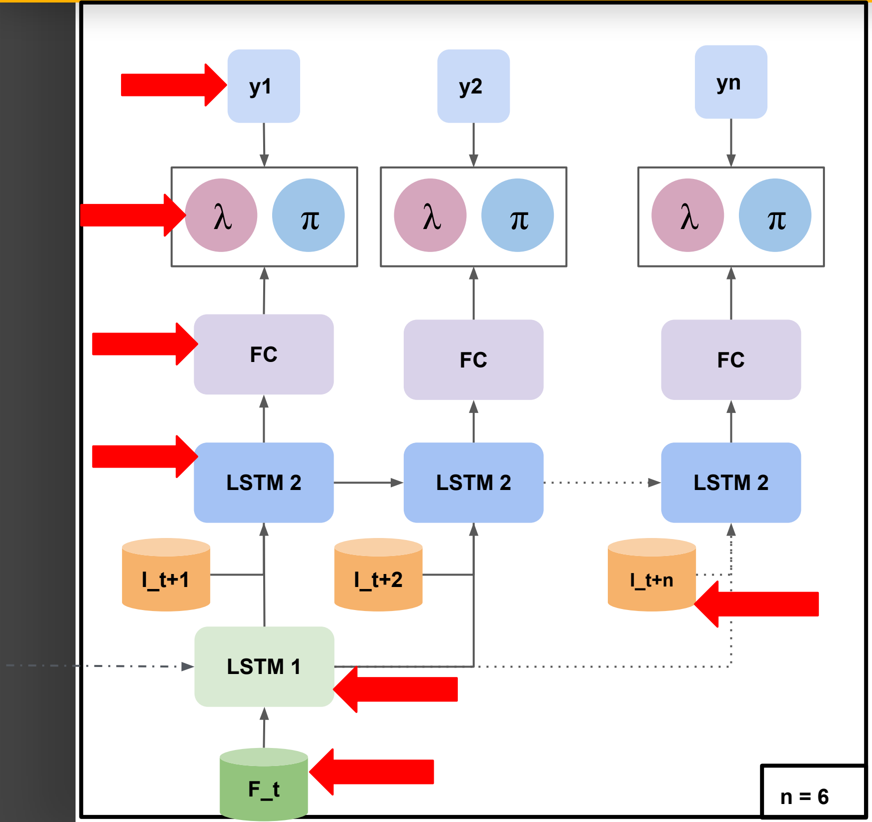
$$\mathbb{P}(y=h) = (1-\pi)\frac{\lambda^h e^{-\lambda}}{h!}, h \geq 1$$

Where,

λ is the expected Poisson count for the individual;

π is the probability of extra zeros

LSTM Model Architecture



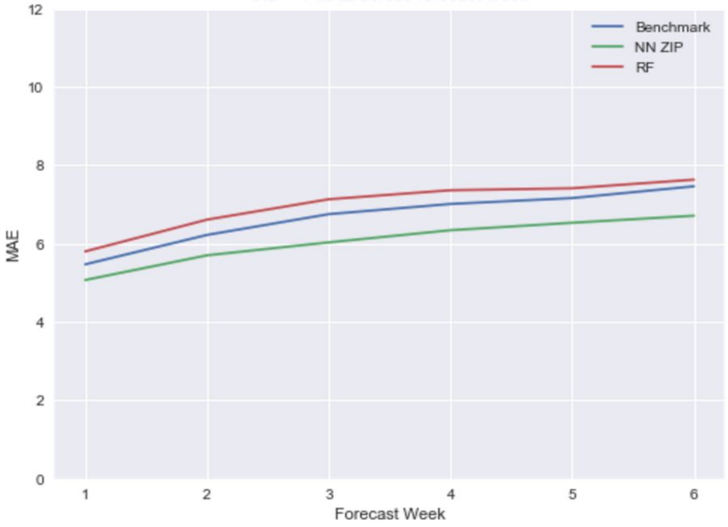
Variable	Description
F_t	Static, behavioral features at time t for a driver
LSTM 1	**Output:** Feature embeddings per driver
LSTM 2	**Input:** Feature embeddings and future incentives (I_t+1….. I_t+12) **Output:** Updated feature embeddings at a given forecasting period
Lambda, Pi	Parameters of ZIP distribution per driver per week
FC	Fully connected layer to map the feature embeddings to two outputs
y1… yn	Final trip predictions as expectation of ZIP

Agenda

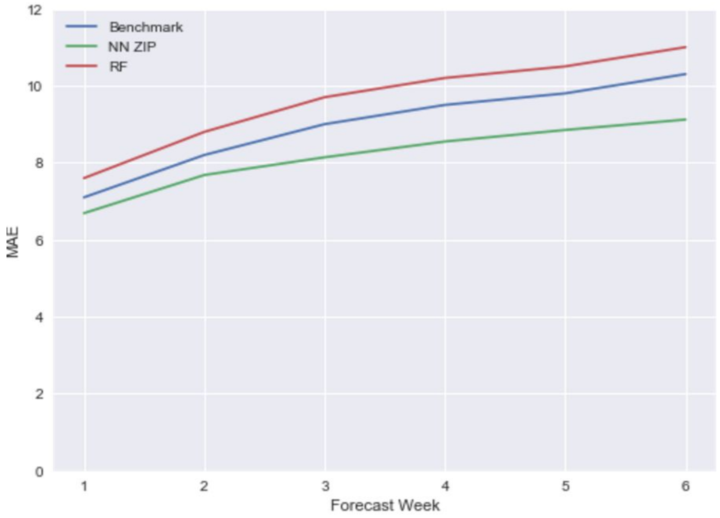
- Forecasting@Uber
- Problem Formulation
- Modeling
- Results

Mean Absolute Error (MAE) Comparison

City A



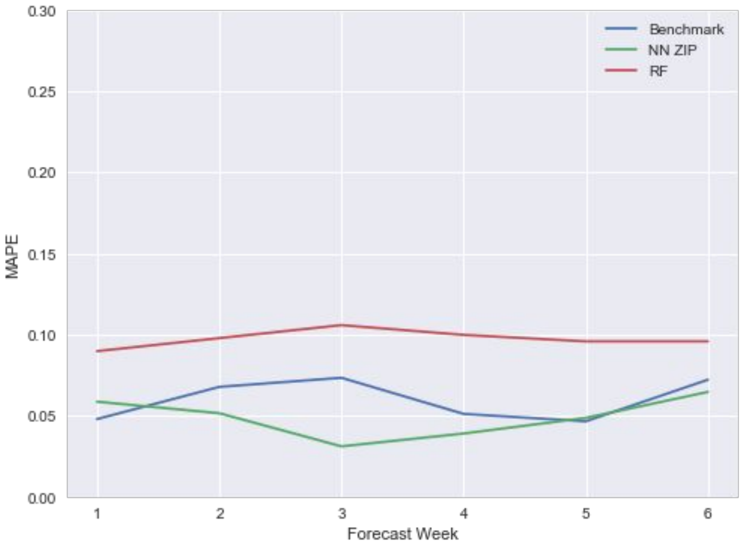
City B



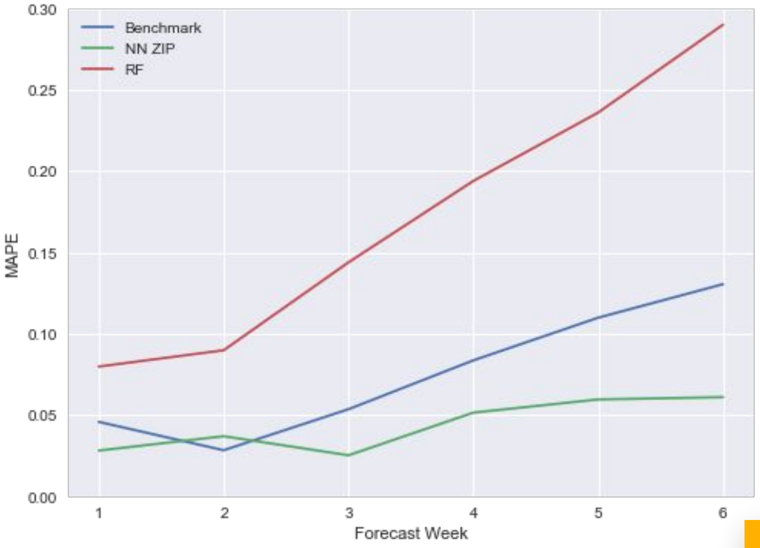
LSTM ZIP outperforms the basic models (~30% improvement)

Mean Absolute Percentage Error (MAPE)

City A

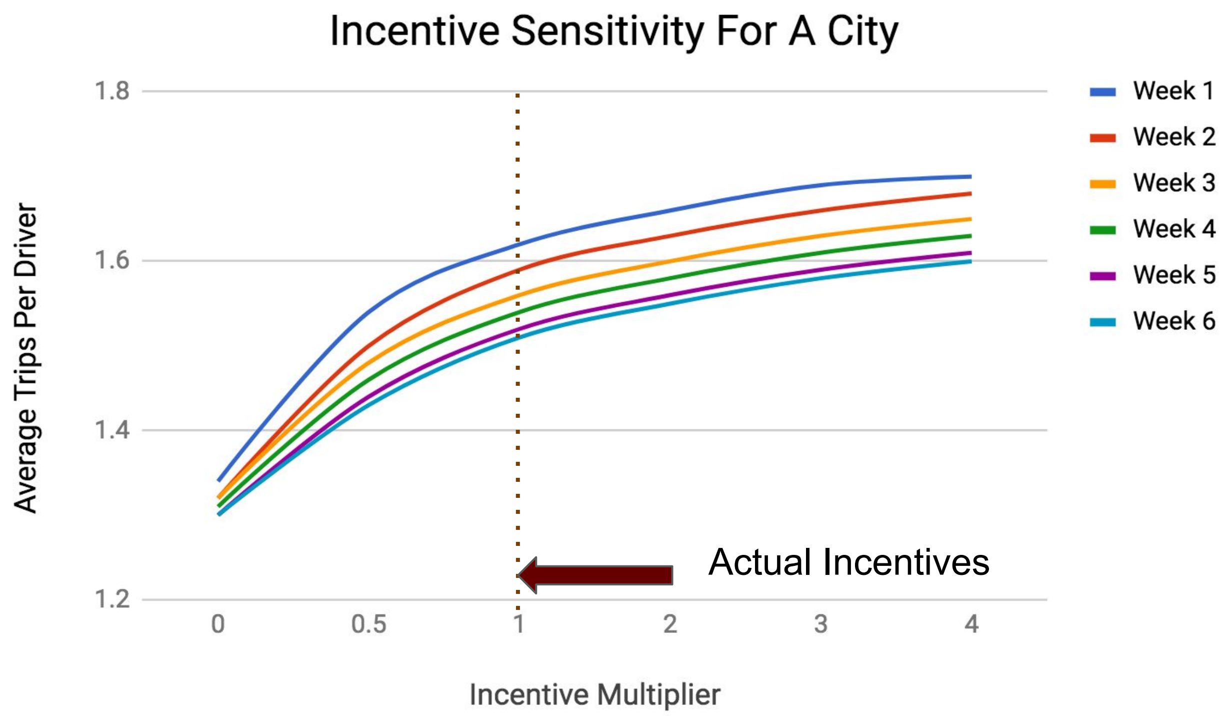


City B



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

Incentive Sensitivity Curves



*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

UBER

UBER