Events and Tickets Representation Learning and Personalized Recommendation

MIT Business Analytics Capstone Project 2019

San Francisco, CA


Team: Andrew Huang Iker Wang

Advisor: Prof. Georgia Perakis

PhD Mentor: Dr. Max Biggs and Dr. Rim Hariss

StubHub: Dr. Corey Reed, Camila Metello

1. Problem Statement

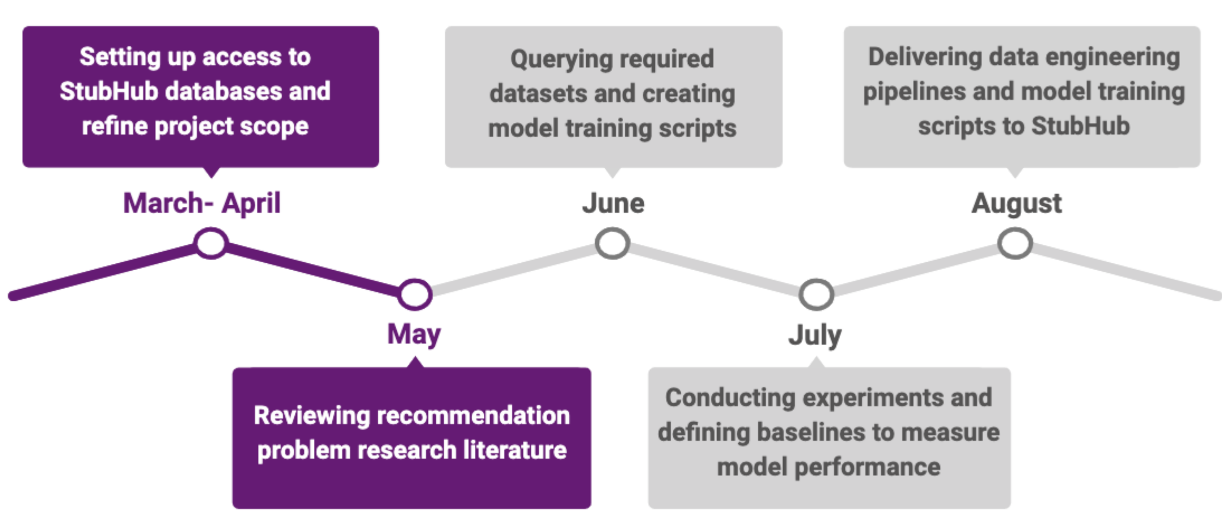
What events and tickets should we recommend to StubHub users to help them explore more options and quickly decide on a purchase?

2. Data Description

The dataset used for this project include user click stream of events on StubHub platform and tickets click history for LA Dodgers home games in 2018. The clicks are grouped into browsing sessions which consist of clicks or user interactions within a continuous period of time. We obtained 17 million event browsing and 370 thousand ticket browsing sessions.

Session	Browsed Events Sequence
Session_1	[Event_1, Event_2, Event_3, Event_5, Event_6]
Session_2	[Event_3, Event_4, Event_2, Event_6]
Session_3	[Event_4, Event_3, Event_5, Event_1]

3. Project Timeline

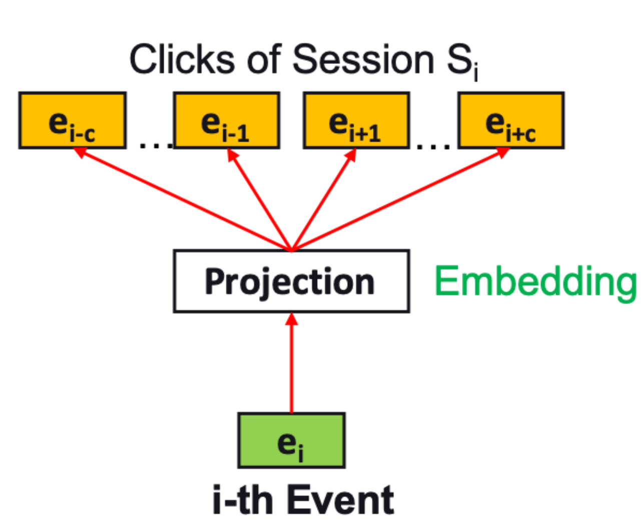


4. Methodology

Prod2Vec Embedding Model

Assumptions:

Items (events and tickets) that are viewed next to each other in the same session can reflect represent users' preferences

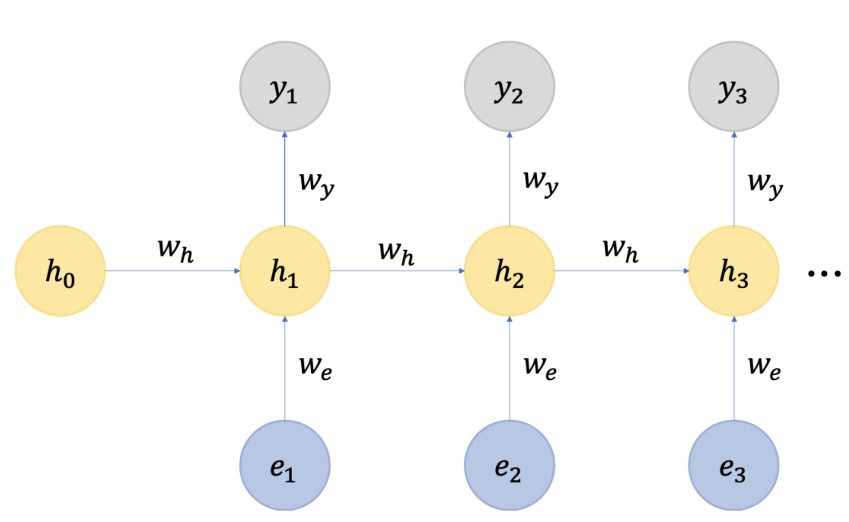


The Prod2Vec model involves learning **vector representations** of products from browsing history data by using a notion of a click sequence as a "sentence" and products within the sequence as "words", borrowing the terminology from the NLP domain.

Recurrent Neural Network (RNN) Model

Assumptions:

Sequential viewing patterns can help infer users' future actions that represent their purchase intentions



The RNN model takes a variable-length, one-hot encoding vector to represent the current state of the session (events and tickets viewed so far) and outputs **a probability estimation** of the next item being viewed using SoftMax function.

5. Model Evaluation and Results

a) Recall and Mean Reciprocal Rank (MRR) @K

The most direct measure for the recommendation system is to see if the recommended items correspond to what the users actually clicked. Since both models can output a variable number (K) of recommendations, we use recall @K, which equals to the proportion of viewed items found in the top-k recommendations and MRR @ K, which also takes the ranking of the output into account to evaluate the model performances.

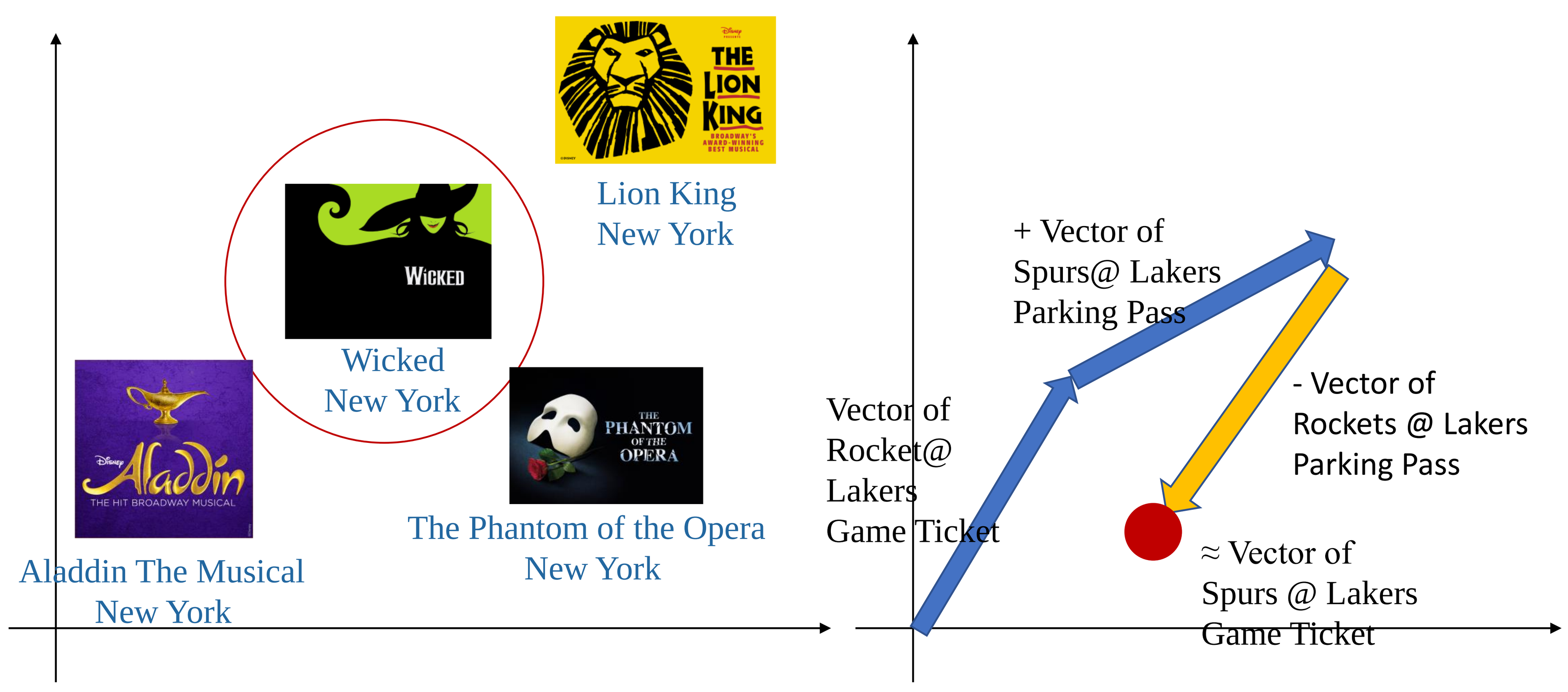
Seat Recommendation Results

Method	Recall@5	MRR@5
POP	0.005	0.002
K-NN	0.080	**0.050**
MF	0.010	0.005
Prod2Vec	0.070	0.039
RNN	**0.110**	**0.050**

Event Recommendation Result

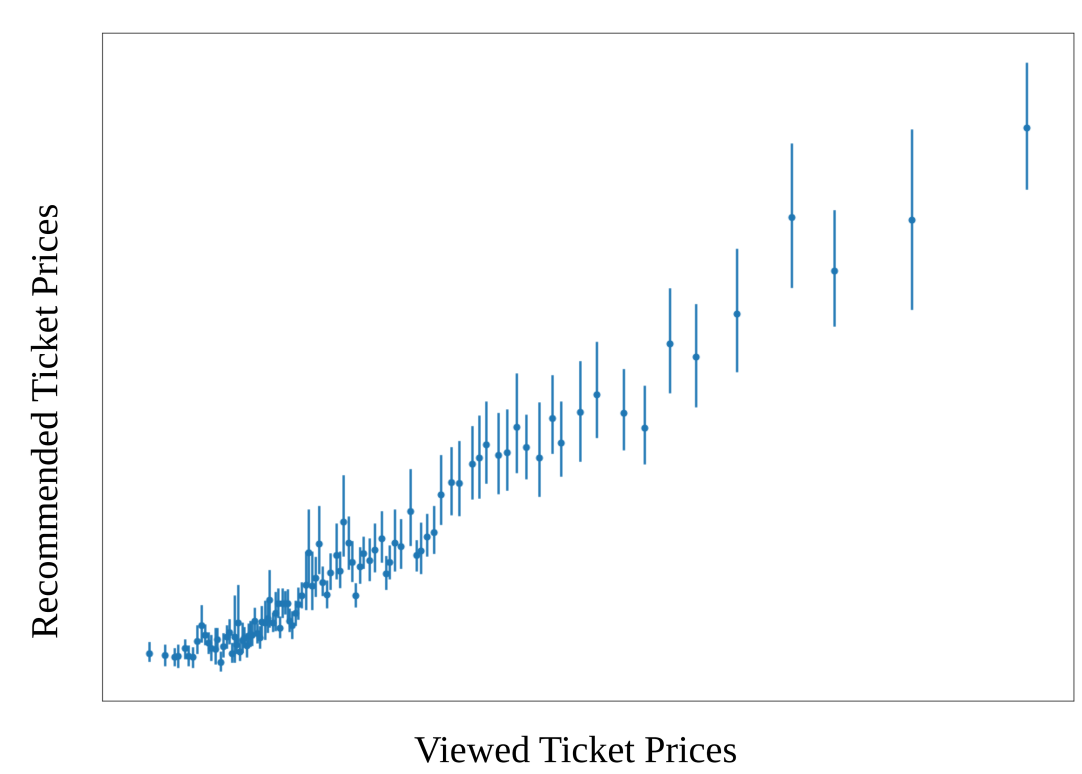
Method	Recall@5	MRR@5
Prod2Vec	0.203	0.118

b) Use Embeddings to Find Similar Events



c) Deviation between Predicted and Viewed Ticket Prices

Since price is a critical factor that influence StubHub users' purchase decisions, we are interested in understanding how well the ticket recommendation model was able to provide tickets that of similar prices to what the users actually viewed. The median percentage difference is **29%**.



d) Using Embeddings as Features

Since the embedding models create vector representation of items, we use them as side information to matrix completion models to calculate users' preference to events in terms of number of clicks and achieved results better than most matrix completion methods

Matrix Completion Model Results

Method	MAE
K-NN	0.385
NMF	0.382
SVD	0.351
SVD++	0.344
StaTNA	**0.336**

6. Next Steps

a. Explore methods to learn feature representation of users in the same space as events/tickets to handle cold start scenarios

b. Incorporate user information and additional event/ticket features explicitly when training recurrent neural network models

c. Introduce trained event and ticket embeddings as features to feed other models within StubHub to improve performance