Deep Learning Models for Restaurant Choice

Evan Munro
munro@stanford.edu

Summary

- Conditional logit models and matrix factorization approaches dominate consumer choice prediction.
- Neural network estimation for these problems is challenging due to sparsity of data and size of input.
- I have a large dataset on consumer choice of lunch restaurant inferred from mobile phone location data in the Bay Area.
- Goal is to predict consumer's choice of restaurant out of sample.
- I explore basic neural network, neural network with embeddings, and recurrent neural network performance compared to baseline.
- The best model gives 9% accuracy, nearly 3x improvement over baseline.

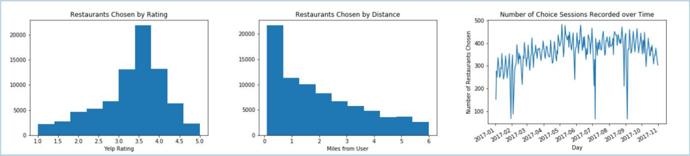
Data

- 10GB TSV file on 78,524 choices sessions from 8,552 users choosing between over 4,921 restaurants for lunch in the Bay Area.
- Derived from mobile location data. Each choice session involves data on a user who chooses a restaurant for lunch, as well as all restaurants within 6 miles of that user.
- Data includes distance from user to each restaurant and Yelp data on the restaurant.

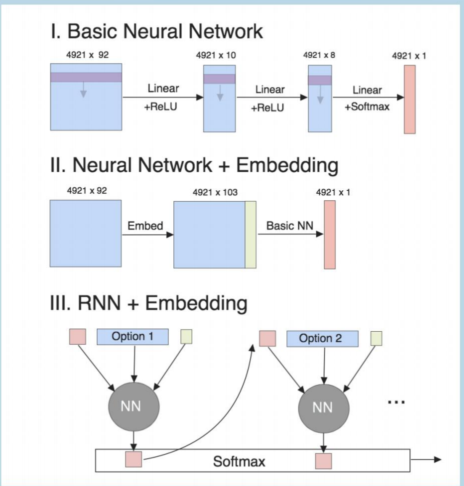
Features

- 92 input features on restaurant type and ratings, and distance between user and restaurant.
- Restaurant, user and choice session ID, used to create embeddings.
- For models with embeddings, additionally learn 5-dimensional user-specific vector, 5-dimensional restaurant specific vector and their dot product for 103 total input features.
- Single output feature: if the restaurant was chosen or not for the choice session.

Choice Sessions



Model Architecture



Results

Model	Layers	Train Loss	Train Accuracy	Test Loss	Test Accuracy
Cond. Logit	1	5.67	3.10%	5.63	3.30%
NN	3	5.42	4.70%	5.40	4.75%
NN + Embed	4	4.99	8.94%	5.04	8.95%
RNN + Embed	4	6.18	3.55%	6.20	3.60%

- Loss used for training and reported is cross-entropy loss

Discussion

- Adding non-linear interactions between features through a basic neural network results in significant improvement in accuracy compared to the baseline model.
- Adding restaurant and user specific embeddings to the basic neural network results in the best performance. The embeddings successfully capture unobserved attributes of users and restaurants that influence choice.
- The RNN model does not perform well as constructed.
- Using interpretable conditional logit models only is at a significant cost to accuracy.

Future

- Investigation: why is RNN performance so poor?
- Additional refinements on number of layers and hidden units, embedding sizes, dropout and learning algorithm.
- New architecture: Bayesian neural network allowing frequently observed restaurants and users to influence parameters for infrequently observed restaurants and users.

References

[1] S. Athey, D. Blei, R. Donnelly, F. Ruiz and T. Schmidt. *Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data.* AEA Papers and Proceedings Vol 108, May 2018.

[2] A. Mottini and R. Acuna-Agost. *Deep Choice Models Using Pointer Networks for Airline Itinerary Prediction.* 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.

[3] M. Otsuka and Takayuki Osogami. *A Deep Choice Model.* Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[4] B. Sifringer, V. Lurkin, and A. Alahi. *Enhancing Discrete Choice Models with Neural Networks.* 18th Swiss Transport Research Conference, 2018.