Nikolas Lee

Professor Cheng Xiang Zhai

CS410

11/15/2020

Tech Review:
Deep Neural Networks for Youtube Recommendations

Introduction

In class, we have learnt briefly about recommender systems and that they generally fall into two categories: Content Based and Collaborative Filtering. The paper that I am reviewing is interesting in that the system it proposed seems to me like a hybrid between the two. With it's Deep Learning based approach, the model is able to take in features about the item and user and in a sense, learn a similarity between a given item and a user. Moreover, the approach itself can be interpreted as a form of Collaborative Filtering since transaction's of other users are used to inform the model about the relationship between items and users. While it was mentioned in class that the two approaches can be combined, I felt that an example of an application in industry (Youtube Recommendations) that most people can relate to would help give other students a clearer idea of how the strengths of both approaches can be leveraged.

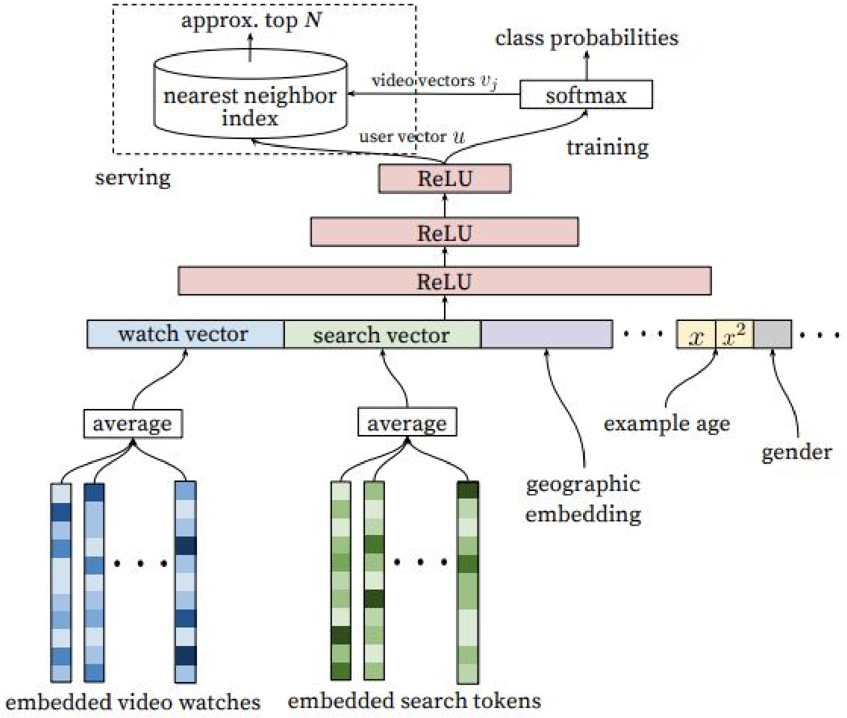
Description

The Youtube recommender system is divided into two components, Candidate Generation and Ranking. Candidate generation acts as a principled way to narrow down a pool of millions of potential videos to recommend to the user to a smaller set of about a hundred videos. From there, the ranker does a more precise estimation of the videos' relevance to the user. This is interesting since it shows a practical view of implementing recommender systems. If computation on the entire item base is infeasible, find a way to first narrow down the set of possible items.

Both components are based on a Deep Learning Approach, which uses complex neural networks to learn highly flexible functions that map a set of features to a prediction. In this application, Youtube frames the recommendations problem as a prediction task where they want to predict the probability that a user will watch a video. Due to the goals of each component, the candidate generation step and the ranking step execute this differently.

Candidate Generation

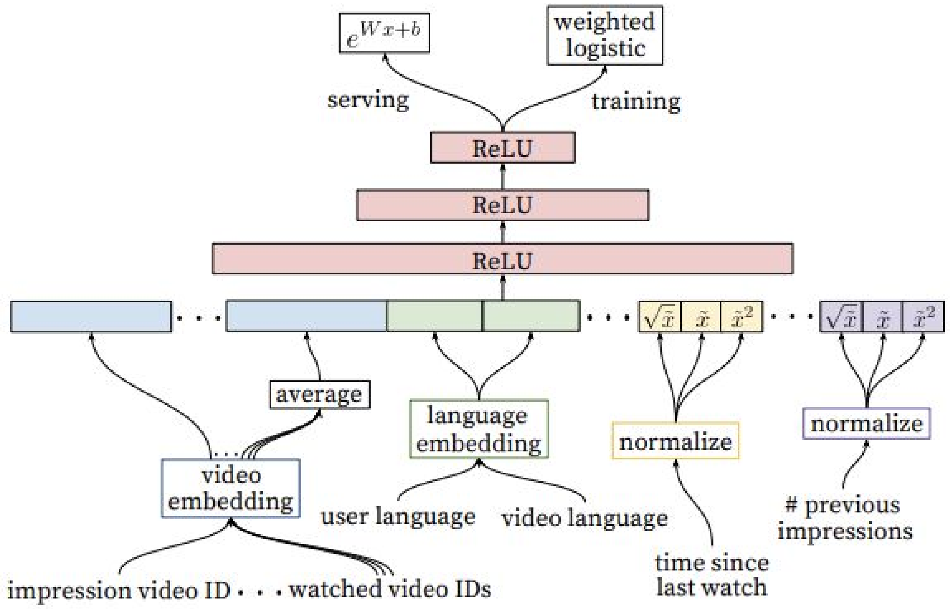
The candidate generation step largely aims to learn a model that is able to give high probabilities to items that the user will watch. At a high level, it learns flexible functions that transform a user's features into user vectors such that when computed against video vectors, the resulting probabilities are high for videos that the user has watched previously and low for those that were not watched. According to the authors, this reduces to a nearest neighbour search during prediction time. I believe this is because similar vectors will have a high dot product and hence a high probability when calculated with a softmax function. An actual softmax function isn't used during training since there are millions of videos and hence millions of classes.As such, during training negative samples are used instead. It is interesting to note the kinds of features used here.



A large component of the neural network is the use of item embeddings. This reminds me of the vector space model taught in class. Here, an item is also represented by a vector, though the method used to generate such a vector is much more complex. Using these embeddings, the network is able to use features such as video watches, search tokens as a way to describe a user.

Ranker

At the ranking stage, they are free to use a much more sophisticated approach in terms of features.



Here, they use a similar embedding approach but also for the user and video's language. This is interesting since it allows the model to learn if similarities in the user and video's language is important in recommendation instead of assuming it does like in traditional content based approaches. They also make full use of the framing of the problem by adding their business objectives into the modeling approach. Leaving it up to the model to figure out a healthy compromise between relevance and business objectives. Here, they are not only satisfied that a user clicks and watches a video, they would also like the time the user spends on the video to be long. Hence, they use a weighted logistic regression for training, giving more importance to positive examples with high watch time.

Conclusion

In summary, I think this work is interesting as it accomplishes a lot with a single approach. It is able to capture item based mechanisms as well as collaborative filtering. It can even allow for additional business objectives.

Works Cited:

https://research.google/pubs/pub45530/