

# Google Youtube Multi-View Deep Learning Approach

Kirti Magadum  
Department of Computer Science  
UIUC Champaign, IL  
magadum2@illinois.edu

## ABSTRACT

Now a days online services are mostly based on automatic personalization to recommend content to user, based on user search history. One of the widely used online service which use personalization concept is Youtube video recommendation system. Youtube also placing advertise based on user search history. On Youtube, people are uploading 500 hours of video every minute and there are approximately 2 billion of users. What makes Youtube to recommend relevant content to users in this huge dataset? At the startup, Youtube was ranking content based on clicks but that doesn't sound good idea as user can click and close a video. Then to overcome this issue Youtube introduced watch time concept where users watch time added more values to recommendation. With increasing number of users and modern technology Youtube introduced Deep Neural Network and machine learning concept to recommend videos to user. So to support millions of product and billions of users multi view Deep Neural Network is very helpful.

## INTRODUCTION

Recommendation system use collaborative filtering, content based filtering or combine both to recommend data to user. Content based filtering use item features liked by user and recommend similar kind of items to

user. Collaborative filtering recommend items to user based on user similarity or item similarity. Collaborative filtering has problem of cold start if enough user and item feature data is not available. To overcome this problem system started building user and item features space based on rich features from user search history by assuming that users online search represents user preference and background.

Deep structured semantic models maps users and items in shared semantic space and recommend items that have maximum similarity. For example, if user search for cricket world cup, then recommend cricket news, cricket games to user. Single Deep neural network is restricted to single domain where as Multi view domain learn features of items from different domain. Multi-View DNN learn more from domain which don't share common feature space. In multi view instead mapping user and features to specific domain, single mapping is done for user features in latent space such that it optimised with features of items from all domains. In this way multi view approach helps to share data across the domain and recommend more useful data to user which resolve problem of data sparsity. We will look into following concept in more detail

1. Multi-View Neural Network
2. System overview for Youtube system
3. Feedback for recommendation system
4. Feature Engineering
5. Normalizing feature

## Multi-View Deep Neural Network

Multi view neural network combines features from different domain with same set of users to recommend different data to users. For example, user watching videos related to specific game can get news, online games recommendation. Here system is trying to maximize similarity between user and item feature view. Objective of multi view domain is to transform user feature into space that matches all different item user liked in different space. This kind of data sharing allows domain that don't have enough information to learn good mapping from other domain. All this work based on

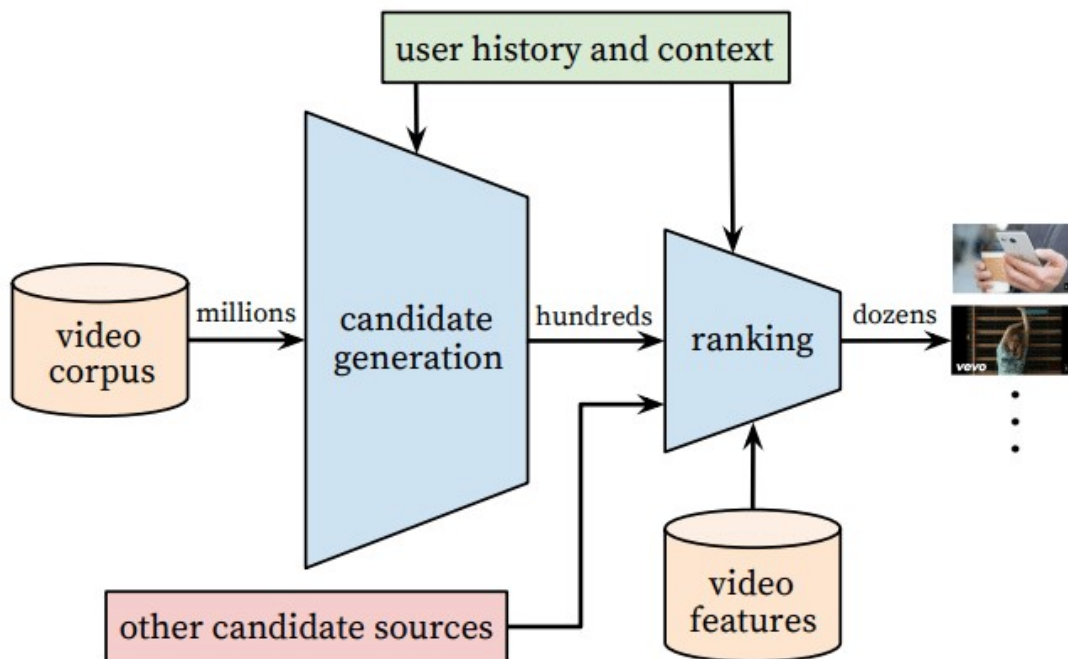
assumption that user has same taste in other domain as well.

### K-Means Algorithm:

K-means clustering technique used to creating number of cluster and sum of distance between each point and its nearest cluster is minimized. Algorithm is used to group similar features to same cluster.

### System Overview

Overall structure of youtube recommendation is illustrated in following figure.



**Figure 2: Recommendation system architecture demonstrating the “funnel” where candidate videos are retrieved and ranked before presenting only a few to the user.**

(Source: Deep Neural Networks for YouTube Recommendations, 2016)

Recommendation system architecture apply two layer of neural network to provide fine tuning of recommendation to user. Candidate Generation neural network is helpful in retrieving hundreds of videos from large video corpus. Candidate generation use Collaborative filtering to create small subset of data from large corpus. In this step, algorithm uses users history and other parameters like video id, demographic information etc. Also it will consider user and video language. In this stage precision has higher weight as compared to recall so that we can pull all relevant videos for user.

Once system extracted small subset of videos from large corpus, system has to fine tune the data such that only few relevant videos are available to user. Ranking is most important steps to provide score to videos as per rich user feature set. Recall has more weight in this state to provide high ranked videos to user.

Input:

Embedded video watch, Embedded search token, geographic information of user, User profile information

Output:

Top N video suggestion to user.

Recommendation system use ReLU neural networks as a hidden layers to provide weight on videos so that system can generate top N videos based on inputs.

### **Feedback For Recommendation System**

Youtube has explicit and implicit feedback mechanism.

Explicit feedback: user like, dislike videos.

Implicit feedback: system monitors watch time for videos. If user complete video that's positive feedback for video.

If user doesn't watch particular video then youtube has to remodel the data to decrease score for that particular video watch

### **Feature Engineering**

Feature engineering transforms user and video data into features. Recommendation algorithm looks for user interaction with item and other similar items. For example, it considers user past history for particular channel and last time when user watch particular video related to that channel.

Feature engineering is important to filtering objects from huge data set so it is more important to select rich feature set.

User language and video language features are added to network to train model which are considered as categorical features for user. Also normalized last time watch and previous impression are added as input to embedded layer.

### **Normalizing Features**

Day by day, data added to youtube is scaling at huge quantity and hence lead into increasing feature set which creating problem for convergence. It is more important to optimise features with increasing data set. For optimising features, system can pick top k most frequent features with high probability.

### **REFERENCES**

Deep Neural Networks for YouTube Recommendations by Paul Covington, Jay Adams, Emre Sargin Google Mountain View, CA

<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/45530.pdf>

A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems Ali Elkahky \*  
Department of Computer Science Columbia University New York, NY 10027, USA  
[ame2154@columbia.edu](mailto:ame2154@columbia.edu) Yang Song,  
Xiaodong He Microsoft Research One Microsoft Way Redmond, WA 98052, USA  
{yangsong, [xiaohe](mailto:xiaohe@microsoft.com)}@microsoft.com