# Modifications to the Programming Assignment

A. S. M. Ahsan-Ul-Haque
ahsanhaquetarique@gmail.com

**Modifications:**

Here are the steps I have followed so far:

(1) Make a YouTube search using "sad songs" (I have used YouTube API for python3). Then, make a list of all the video songs returned by the search results.

(2) Iterate through the list of songs and get the set of words from each video's comment section. I have taken only first N most frequent words, as mentioned earlier, which consist of at least 3 letters.

(3) Each set of words from each video's comment section is now considered the document set which is fed into the Topic Modeling algorithm. Here I have used Topic Modeling using Gibb's Sampler.

(4) I have included the word "sad" in every document so that at least one instance of this word exists in each document. This is a crucial step and later used in step 6.

(5) Then, I have set the 3 hyperparameters of the algorithm. What the algorithm essentially does is it groups the words most likely to be relevant to the same topic.

(6) After running the algorithm, I have searched for the word "sad" in each of the topics in each document. The algorithm has already grouped together similar words like {sad, sorrow, unhappy} etc.

So, then, it can be easily figured out which topic refers most to the sad category.

(7) After getting the index of the desired topic, the documents are sorted according to the distribution of that topic in descending order. Now that the documents are sorted, and it is known which document corresponds to which video, we finally get a sorted list of the videos which were found in step 1. We then pick the desired number of unique videos from this list (according to the given assignment, I have selected the first 15 videos).

**Results:**

Here are the results of 3 sample runs. I have used eta = 0.1 and alpha = 50/K in all three runs.

Sample run 1: (Using N = 5, K = 2)

http://www.youtube.com/watch?v=8xOMlART0XA

http://www.youtube.com/watch?v=sHMYSyIpgnY

http://www.youtube.com/watch?v=HTXx5siaRcA

http://www.youtube.com/watch?v=GKSRyLdjsPA

http://www.youtube.com/watch?v=DDWKuo3gXMQ

http://www.youtube.com/watch?v=CpB_O0uocF8

http://www.youtube.com/watch?v=sC2nElyx7Ds

http://www.youtube.com/watch?v=ElMj9a06yZQ

http://www.youtube.com/watch?v=F_UiE7VioVo

http://www.youtube.com/watch?v=ij_0p_6qTss

http://www.youtube.com/watch?v=s1tAYmMjLdY

http://www.youtube.com/watch?v=dGR65RWwzg8

http://www.youtube.com/watch?v=0G3_kG5FFfQ

http://www.youtube.com/watch?v=3lvZeFtgscA

http://www.youtube.com/watch?v=QifhnI0r7bg

Sample run 2: (Using N = 10, K= 5)

http://www.youtube.com/watch?v=HTXx5siaRcA

http://www.youtube.com/watch?v=F_UiE7VioVo

http://www.youtube.com/watch?v=ij_0p_6qTss

http://www.youtube.com/watch?v=hLQl3WQQoQ0

http://www.youtube.com/watch?v=QifhnI0r7bg

http://www.youtube.com/watch?v=YQHsXMglC9A

http://www.youtube.com/watch?v=YZz0PNcS95U

http://www.youtube.com/watch?v=bC3WAxiLnDY

http://www.youtube.com/watch?v=wr_1N59KHdQ

http://www.youtube.com/watch?v=eG3RRQigDPs

http://www.youtube.com/watch?v=bIjVwEduD68

http://www.youtube.com/watch?v=ElMj9a06yZQ

http://www.youtube.com/watch?v=s1tAYmMjLdY

http://www.youtube.com/watch?v=8xOMlART0XA

http://www.youtube.com/watch?v=sHMYSyIpgnY

Sample run 3: (Using N = 10, K = 10)

http://www.youtube.com/watch?v=QifhnI0r7bg

http://www.youtube.com/watch?v=YQHsXMglC9A

http://www.youtube.com/watch?v=hLQl3WQQoQ0

http://www.youtube.com/watch?v=bIjVwEduD68

http://www.youtube.com/watch?v=GKSRyLdjsPA

http://www.youtube.com/watch?v=s1tAYmMjLdY

http://www.youtube.com/watch?v=8xOMlART0XA

http://www.youtube.com/watch?v=sHMYSyIpgnY

http://www.youtube.com/watch?v=eG3RRQigDPs

http://www.youtube.com/watch?v=HTXx5siaRcA

http://www.youtube.com/watch?v=DDWKuo3gXMQ

http://www.youtube.com/watch?v=CpB_O0uocF8

http://www.youtube.com/watch?v=sC2nElyx7Ds

http://www.youtube.com/watch?v=ElMj9a06yZQ

http://www.youtube.com/watch?v=F_UiE7VioVo

**Remarks**

If we create a customized application, more user data will be available to us (for example: user's age, gender, nationality, liked videos, shared videos, user's comments and so on). Then, we can use a well-known learning model (such as Neural Network, Support Vector Machine, Logistic Regression etc.) to train with this extended set of features, and get a more personalized result.