1)**Distributionally Similar Category.**

a)Get top 5 distributionally similar Nouns, Verbs, Adjectives

( using tool http://www.linguatools.de/disco/disco\_en.html)

b)Calculate Cosine similarity separately for Nouns, Adjectives, Verbs using stemmed words.

**2)LIWC Category**

Get Gay marriage text from CreateDebate,ConvinceMe, forums(excluding dialog corpus discussions)

Extract top 20 Nouns,

top 20 Verbs [ for verbs, if it is in nltk stop word list, do not add it in this top 20]

top 20 Adjectives

( From these 60 remove certain too common verbs such as) LIWC\_StopList=[*"'s"*,*"think"*,*"say"*,*"said"*,*"get"*,*"'ve"*,*"believe"*,*"saying"*].

Get LIWC categories for the above set.

From this LIWC category set, remove categories that look too general NotToUseCategory=[*"Unique Words"*,*"Words Per Sentence"*,*"Word Count"*,*"Dictionary Words"*,*"Common Verbs"*, *"Six Letter Words"*, *"Present Tense"*,*"Total Function Words"*, *"Word Count"*, *"Total Pronouns"*,*"Present Tense"*,*"Adverbs"*, *"Impersonal Pronouns"*, *"Sentences"*,*"Space"*,*"Past Tense"*, *"Auxiliary Verbs",Quantifiers"*,*"Dash"*,*"Number"*,*"All Punctuation"*,*" Impersonal Pronouns"*]

Find LIWC CategoryOverlap count across pairs for each category, as a feature

**3) Difference**  in Number of words

**4) Rouge Scores.**

Rouge gave various measures. As of now I used f score for all these measures

*"[rouge\_1\_f\_score"]*

[*rouge\_2\_f\_score"*]

[*"rouge\_3\_f\_score"*]

[*"rouge\_4\_f\_score"*]

[*"rouge\_l\_f\_score"*]

[*"rouge\_s\*\_f\_score"*]

[*"rouge\_su\*\_f\_score"*]

[*"rouge\_w\_1.2\_f\_score"*]

5)Count of **unigram,bigram, trigram overlap.**

6)***LevDistance*** *Edit Distance( keep only Nouns, Verbs, Adjectives,Adverb)*

*7)****Noun Synset path distance***

*Synset Path distance using nltk path distance metric.*

*Get stanford collapsed dependencies.*

*For Relation=["nsubj","nsubjpass","dobj","iobj","pobj"]*

*get all “Dep” for each of these above relations for both the strings*

*Take cross product of these dep.*

*Get synset using nltk WSD*

*Calculate path distance*

*Take sum of these as a feature.*

8 *Verb Synset Path distance*

*A similar thing as no 7 by taking “GOV” of the relation ["nsubj"]*

*and “Dep” of “ROOT” relation.*

1. *UMBC*