Assignment-V (Based on Lab Session-VII)

Word Vectors

Q1: Code a function display_pca_similar_display(model,word) that takes word vector model and a word as an input and displays 10 most similar words for the given word (in green color) and 5 most dissimilar words to the given word (in red color) in a two dimensional space (dimensionality reduced using PCA).

Use the following Word Vector models present in genism:

word2vec-google-news-300
glove-wiki-gigaword-300
glove-twitter-200
fasttext-wiki-news-subwords-300

Q2: Compare the performance of the following word vector models:

word2vec-google-news-300
glove-wiki-gigaword-300
glove-twitter-200
fasttext-wiki-news-subwords-300

on (i) Word Similarity task using WordSim353 and Rare Words dataset

(ii) Word Analogy task using Mixed Dataset

Dataset Link:

WordSim353: https://drive.google.com/file/d/1XPI1vJLu-8opoacQJJMJU5AaZjGBa8-7/view?usp=sharing

Rare Words:
https://drive.google.com/file/d/1K2PmINOjIQDR0sdhPrvziNmGAork49uA/view?usp=sharing

Mixed Dataset

https://drive.google.com/file/d/1D5PiDQ-uZEUUiKLUmr0HPRdAj_GokOa5/view?usp=sharing

Q3: For the movie_genre_classification experiment discussed in class improve the accuracy by trying following options:

(i) Pre-Processing

(ii) Word Vector Model

(iii) Classifier Model

(iv) Dimensionality of word Vector