

Wap in Python to implement << avg >> word2vec

```
[12]: import numpy as np

[13]: def avg_word_2_vector(sentence,model):
        words = sentence.lower().split() #convert to Lower case and split using space
        word_vec = [model[word] for word in words if word in model]
        return np.mean(word_vec,axis=0) if word_vec else np.zeros(model.vector_size)

[ ]: from gensim.models import Word2Vec

[16]: import gensim.downloader as api # This donwLoad will help in downloading the required model :pretrained ~ 1.5+

[26]: #How to find the Model List
        infos = api.info()
        #print('Inforamations',infos)

        models_list = infos['models']
        #print('Model List',models_List)
        #print(List(models_List.keys())) # Keys with Version

[27]: model = api.load('glove-wiki-gigaword-50') #25,50,100,.... 70Mb -80 Mb
        [=====] 100.0% 66.0/66.0MB downloaded

[30]: model = api.load('glove-wiki-gigaword-100') #25,50,100,.... 70Mb -80 Mb
        [=====] 100.0% 128.1/128.1MB downloaded

[28]: sentences = "machine learning is fun and amazing"
        vector = avg_word_2_vector(sentences,model)
        print('Average Word2Vec Vector:',vector)

Average Word2Vec Vector: [ 8.31471607e-02  2.18846679e-01 -1.38463333e-01  6.13261648e-02
  3.85051489e-01  1.88059986e-01 -2.05892682e-01 -5.00615001e-01
 -3.75501625e-02  4.05635029e-01 -5.74426502e-02  6.43583313e-02
 -3.12729329e-01 -1.87067494e-01  1.14046663e-01 -5.25454991e-02
  1.75052509e-01  7.08079994e-01 -3.58426660e-01 -4.36812311e-01
  6.10968359e-02  5.28240025e-01  1.01175465e-01  1.72002330e-01
  7.43966639e-01 -1.01666319e+00 -7.79363334e-01 -4.19316739e-02
  5.74840724e-01 -4.77062345e-01  2.94435000e+00  1.60043344e-01
 -1.40605018e-01 -3.63293320e-01  4.47284915e-02  3.46259981e-01
  8.62003025e-03  4.20613319e-01 -8.16682950e-05 -9.25899968e-02
  1.70995995e-01 -5.62680066e-02 -3.76873344e-01  2.87421495e-01
  1.54005498e-01 -5.15500223e-03  3.54964972e-01 -2.78354973e-01
  3.36866714e-02  4.50016707e-01]

[31]: sentences = "machine learning is fun and amazing"
        vector = avg_word_2_vector(sentences,model)
        print('Average Word2Vec Vector:',vector)

Average Word2Vec Vector: [-1.34138837e-01  3.59145492e-01  1.50785163e-01 -2.16113344e-01
 -1.06366649e-02  2.80669838e-01  1.03497505e-01  3.35541874e-01
 -2.37133410e-02  5.02013415e-02  2.00698331e-01 -3.35729986e-01
  6.82625100e-02 -1.45879492e-01  2.16315985e-01 -4.12190072e-02
  1.38842598e-01  3.30208987e-01 -3.05715144e-01  5.79083383e-01
  5.47016673e-02  9.80758294e-02  1.34446681e-01 -3.77443314e-01
  2.47162342e-01  3.13926667e-01 -1.23676330e-01 -1.60482585e-01
 -1.42685145e-01 -1.32208660e-01 -5.41059971e-01  6.55141294e-01
 -1.21306330e-01 -1.13967180e-01  2.60186166e-01 -1.87904999e-01
 -1.54202670e-01  1.57063320e-01  2.16242984e-01 -4.83716637e-01
 -1.09215014e-01  1.85354650e-02 -6.83656558e-02 -2.44222507e-01
 -3.18887681e-01  4.34966898e-03  2.24950328e-01 -2.06503332e-01
  1.38037845e-01 -6.01214945e-01 -1.20705329e-01 -1.98849186e-01
  2.56113023e-01  9.85803366e-01  1.31813824e-01 -2.11428332e+00
  1.36009499e-01  2.30268359e-01  1.07837999e+00  3.33496600e-01
  5.54566681e-02  7.02011764e-01 -3.18614990e-01  9.37383249e-02
  4.45675641e-01  9.85273346e-02  6.20333016e-01 -4.88225073e-02
  3.83785009e-01  2.54523326e-02  4.05251682e-02 -1.60301641e-01
  8.96711648e-02 -1.36388138e-01  2.09496662e-01  3.77328306e-01
 -9.83833242e-03  3.15160491e-02 -4.19224650e-01 -1.04246633e-02
  2.75268346e-01  1.82515204e-01 -3.42304945e-01  6.78076670e-02
 -1.37777662e+00  1.37228325e-01  6.17553294e-02 -3.93926859e-01
 -1.71988323e-01 -3.32144827e-01 -3.70835274e-04  1.45283416e-02
  1.77019998e-01  3.05356830e-01 -2.65246183e-01 -2.21995354e-01
 -3.91200989e-01 -8.32891703e-01  7.32914984e-01  5.15126705e-01]

[32]: model.save('glove-average-word2vec.model')
```

[]:

