

03_Doc2Vec Sample Code

April 13, 2022

```
[25]: #https://medium.com/red-buffer/  
      ↪doc2vec-computing-similarity-between-the-documents-47daf6c828cd  
import gensim  
from gensim.models.doc2vec import Doc2Vec, TaggedDocument  
from nltk.tokenize import word_tokenize  
from gensim.models.doc2vec import Doc2Vec
```

```
[2]: data = ["The process of searching for a job can be very stressful, but it_  
      ↪doesn't have to be. Start with a\  
          well-written resume that has appropriate keywords for your occupation._  
      ↪Next, conduct a targeted job search\  
          for positions that meet your needs.",  
          "Gardening in mixed beds is a great way to get the most productivity_  
      ↪from a small space. Some investment\  
          is required, to purchase materials for the beds themselves, as well as_  
      ↪soil and compost. The\  
          investment will likely pay-off in terms of increased productivity.",  
          "Looking for a job can be very stressful, but it doesn't have to be._  
      ↪Begin by writing a good resume with\  
          appropriate keywords for your occupation. Second, target your job_  
      ↪search for positions that match your\  
          needs."]
```

```
[3]: tagged_data = [TaggedDocument(words=word_tokenize(_d.lower()), tags=[str(i)])_  
      ↪for i, _d in enumerate(data)]
```

```
[29]: print (tagged_data)  
tagged_data[0]  
tagged_data[0].words  
#tagged_data[0].tags
```

```
[TaggedDocument(words=['the', 'process', 'of', 'searching', 'for', 'a', 'job',  
'can', 'be', 'very', 'stressful', ' ', 'but', 'it', 'doesn', ' ', 't', 'have',  
'to', 'be', ' ', 'start', 'with', 'a', 'well-written', 'resume', 'that', 'has',  
'appropriate', 'keywords', 'for', 'your', 'occupation', ' ', 'next', ' ',  
'conduct', 'a', 'targeted', 'job', 'search', 'for', 'positions', 'that', 'meet',  
'your', 'needs', ' '], tags=['0']), TaggedDocument(words=['gardening', 'in',
```

```
'mixed', 'beds', 'is', 'a', 'great', 'way', 'to', 'get', 'the', 'most',
'productivity', 'from', 'a', 'small', 'space', '.', 'some', 'investment', 'is',
'required', ',', 'to', 'purchase', 'materials', 'for', 'the', 'beds',
'themselves', ',', 'as', 'well', 'as', 'soil', 'and', 'compost', '.', 'the',
'investment', 'will', 'likely', 'pay-off', 'in', 'terms', 'of', 'increased',
'productivity', '.'], tags=['1']), TaggedDocument(words=['looking', 'for', 'a',
'job', 'can', 'be', 'very', 'stressful', ',', 'but', 'it', 'doesn', '', 't',
'have', 'to', 'be', '.', 'begin', 'by', 'writing', 'a', 'good', 'resume',
'with', 'appropriate', 'keywords', 'for', 'your', 'occupation', '.', 'second',
',', 'target', 'your', 'job', 'search', 'for', 'positions', 'that', 'match',
'your', 'needs', '.'], tags=['2']))]
```

```
[29]: ['the',
'process',
'of',
'searching',
'for',
'a',
'job',
'can',
'be',
'very',
'stressful',
',',
'but',
'it',
'doesn',
'',
't',
'have',
'to',
'be',
'.',
'start',
'with',
'a',
'well-written',
'resume',
'that',
'has',
'appropriate',
'keywords',
'for',
'your',
'occupation',
'.',
'next',
```

```

',',
'conduct',
'a',
'targeted',
'job',
'search',
'for',
'positions',
'that',
'meet',
'your',
'needs',
'.']

```

```

[12]: model = gensim.models.doc2vec.Doc2Vec(vector_size=30, min_count=2, epochs=80)
model.build_vocab(tagged_data)
model.train(tagged_data, total_examples=model.corpus_count, epochs=80)
model.save("d2v.model")

```

```

[24]: similar_doc = model.docvecs.most_similar('0')

```

```

print(similar_doc)

print(tagged_data[0].words)

print(tagged_data[2].words)

```

```

[('2', 0.9292380213737488), ('1', 0.8601543307304382)]
['the', 'process', 'of', 'searching', 'for', 'a', 'job', 'can', 'be', 'very',
'stressful', ',', 'but', 'it', 'doesn', '', 't', 'have', 'to', 'be', '.',
'start', 'with', 'a', 'well-written', 'resume', 'that', 'has', 'appropriate',
'keywords', 'for', 'your', 'occupation', '.', 'next', ',', 'conduct', 'a',
'targeted', 'job', 'search', 'for', 'positions', 'that', 'meet', 'your',
'needs', '.']
['looking', 'for', 'a', 'job', 'can', 'be', 'very', 'stressful', ',', 'but',
'it', 'doesn', '', 't', 'have', 'to', 'be', '.', 'begin', 'by', 'writing', 'a',
'good', 'resume', 'with', 'appropriate', 'keywords', 'for', 'your',
'occupation', '.', 'second', ',', 'target', 'your', 'job', 'search', 'for',
'positions', 'that', 'match', 'your', 'needs', '.']

```

```

[ ]: # Model Concat - #https://towardsdatascience.com/
→multi-class-text-classification-with-doc2vec-logistic-regression-9da9947b43f4#:
→~:
→text=Doc2vec%20is%20an%20NLP%20tool,advisable%20to%20understand%20word2vec%20approach.
→

```