scripts.glove2word2vec – Convert glove format to word2vec

This script allows to convert GloVe vectors into the word2vec. Both files are presented in text format and almost identical except that word2vec includes number of vectors and its dimension which is only difference regard to GloVe.

Notes

GloVe format (real example can be founded **on Stanford size**)

```
word1 0.123 0.134 0.532 0.152
word2 0.934 0.412 0.532 0.159
word3 0.334 0.241 0.324 0.188
...
word9 0.334 0.241 0.324 0.188
```

Word2Vec format (real example can be founded **on w2v old repository**)

```
9 4
word1 0.123 0.134 0.532 0.152
word2 0.934 0.412 0.532 0.159
word3 0.334 0.241 0.324 0.188
...
word9 0.334 0.241 0.324 0.188
```

How to use

```
>>> from gensim.test.utils import datapath, get_tmpfile
>>> from gensim.models import KeyedVectors
>>> from gensim.scripts.glove2word2vec import glove2word2vec
>>>
>>> glove_file = datapath('test_glove.txt')
>>> tmp_file = get_tmpfile("test_word2vec.txt")
>>>
>>> _ = glove2word2vec(glove_file, tmp_file)
>>>
>>> model = KeyedVectors.load_word2vec_format(tmp_file)
```

Command line arguments

```
...
  -h, --help            show this help message and exit
  -i INPUT, --input INPUT
                        Path to input file in GloVe format
  -o OUTPUT, --output OUTPUT
                        Path to output file
```

gensim.scripts.glove2word2vec.get_glove_info(*glove_file_name*)

Get number of vectors in provided *glove_file_name* and dimension of vectors.

Parameters

glove_file_name (*str*) – Path to file in GloVe format.

Returns

Number of vectors (lines) of input file and its dimension.

Return type

(int, int)

gensim.scripts.glove2word2vec.glove2word2vec(*glove_input_file*, *word2vec_output_file*)

Convert *glove_input_file* in GloVe format to word2vec format and write it to *word2vec_output_file*.

Parameters

- **glove_input_file** (*str*) – Path to file in GloVe format.

- **word2vec_output_file** (*str*) – Path to output file.

Returns

Number of vectors (lines) of input file and its dimension.

Return type

(int, int)

gensim

Support:

Stay informed via gensim mailing list:

your@email.com

Subscribe

2

Tweet @Gensim_py