



NLP项目

gensim工具

gensim

- 💡 Gensim 属于一款开源的第三方Python工具包，用于从原始的非结构化的文本数据中，无监督的学习到文本隐层的主题向量表达形式，也即是词向量/文本向量转换；并且在向量转换的基础上提供了相似度计算、信息检索等功能的API。
 - 💡 语料(Corpus)：一组原始文本集合，用于无监督训练文本主题的数据。在gensim中，corpus通常是一个可迭代对象；
 - 💡 向量(Vector)：由一组文本特征构成的列表，是一段文本在Gensim中的内部表达；
 - 💡 稀疏向量(SparseVector)：略去向量中多余的0元素。此时，向量中的每一个元素是一个(key, value)的元组；
 - 💡 模型(Model)：定义了两个向量空间的转换。也就是将向量表达形式从某一种变量变成另外一种变量表达。

gensim



gensim
topic modelling for humans

💡 gensim参考网站:

- 💡 <https://radimrehurek.com/gensim/index.html>
- 💡 <https://github.com/RaRe-Technologies/gensim>

💡 主要模型:

- 💡 [models.Idamodel](#)
- 💡 [models.nmf](#)
- 💡 [models.tfidfmodel](#)
- 💡 [models.word2vec](#)
- 💡 [models.doc2vec](#)
- 💡 [models.fasttext](#)

- [models.ldamodel – Latent Dirichlet Allocation](#)
 - [Usage examples](#)
- [models.ldanmulticore – parallelized Latent Dirichlet Allocation](#)
 - [Usage examples](#)
- [models.nmf – Non-Negative Matrix factorization](#)
- [models.lsimodel – Latent Semantic Indexing](#)
- [models.ldaseqmodel – Dynamic Topic Modeling in Python](#)
- [models.tfidfmodel – TF-IDF model](#)
 - [models.word2vec – Word2vec embeddings](#)
 - [Other embeddings](#)
 - [Usage examples](#)
- [models.keyedvectors – Store and query word vectors](#)
 - [Why use KeyedVectors instead of a full model?](#)
 - [How to obtain word vectors?](#)
 - [What can I do with word vectors?](#)
- [models.doc2vec – Doc2vec paragraph embeddings](#)
 - [Usage examples](#)
- [models.fasttext – FastText model](#)
 - [Usage examples](#)
 - [Implementation Notes](#)

gensim安装

```
(base) PS C:\Users\gerry_17578261252713> pip install gensim
Looking in indexes: https://mirrors.aliyun.com/pypi/simple
Collecting gensim
  Downloading https://mirrors.aliyun.com/pypi/packages/8a/6f/a690547cb7089d4014/gensim-4.2.0-cp37-cp37m-win_amd64.whl (24.0MB)
    |████████████████████████████████| 24.0MB
Requirement already satisfied: numpy>=1.17.0 in d:\anaconda3\lib\site-packages
Collecting smart-open>=1.8.1 (from gensim)
  Downloading https://mirrors.aliyun.com/pypi/packages/09/db/fab79b619923e26ce0/smart_open-6.0.0-py3-none-any.whl (58kB)
    |████████████████████████████████| 61kB 2
Collecting Cython==0.29.28 (from gensim)
  Downloading https://mirrors.aliyun.com/pypi/packages/9f/79/311cfbca90332ab37b/Cython-0.29.28-py2.py3-none-any.whl (983kB)
    |████████████████████████████████| 993kB
Requirement already satisfied: scipy>=0.18.1 in d:\anaconda3\lib\site-packages
Installing collected packages: smart-open, Cython, gensim
  Found existing installation: Cython 0.29.13
    Uninstalling Cython-0.29.13:
      Successfully uninstalled Cython-0.29.13
  Successfully installed Cython-0.29.28 gensim-4.2.0 smart-open-6.0.0
```

```
(base) PS C:\Users\gerry_17578261252713> python
Python 3.7.4 (default, Aug 9 2019, 18:34:13) [MSC v.1915 64 bit]
Type "help", "copyright", "credits" or "license" for more information
>>> import gensim
>>> gensim.__version__
'4.2.0'
>>>
```

THANKS!