

```
In [2]: import fasttext
```

```
In [16]: cbowmodel = fasttext.train_unsupervised('wiki_seg.txt', model='cbow', epoch=
```

```
Read 213M words
Number of words: 1003967
Number of labels: 0
Progress: 100.0% words/sec/thread: 97134 lr: 0.000000 avg.loss: 0.432436
ETA: 0h 0m 0s 10.3% words/sec/thread: 96787 lr: 0.044862 avg.loss: 1.1
67565 ETA: 0h47m12s70.3% words/sec/thread: 97334 lr: 0.014844 avg.loss:
0.529561 ETA: 0h15m31s 71.2% words/sec/thread: 97329 lr: 0.014417 avg.l
oss: 0.526149 ETA: 0h15m 5sh12m48s
```

```
In [17]: cbowmodel.save_model("newfasttext_model.bin")
```

```
In [18]: len(cbowmodel.words)
```

```
Out[18]: 1003967
```

```
In [19]: print (cbowmodel.words[:20])
```

```
['年', '中', '日', '小行星', '</s>', '月', '中國', '日本', '一個', '時', '香港',
'美國', '陳', '臺灣', '站', 'zh', 'px', '研究', '新', '電影']
```

```
In [20]: print (cbowmodel['數學家'])
```

```
[ -0.14894783 -1.1552591 -3.3536594 0.7525944 -0.83275825 0.89067185
 2.156817 0.0651063 -1.279353 0.07770692 0.19925724 0.16411196
-1.6754671 0.47449932 -0.37663874 -1.842189 1.2138776 -4.0351872
 0.97010946 0.16074318 -0.35382205 -1.9478725 -0.569569 -1.5937629
-0.37494224 0.9866736 0.5704824 -0.82443655 0.03014347 -1.7206781
 1.1712375 1.6310178 0.25829548 2.0266414 -0.38489696 -0.41668504
 0.33555797 3.1481738 1.5572802 -1.976383 1.2734907 0.05798674
 0.5227537 0.11597324 -0.46256745 -2.6796567 1.1094084 3.0837615
-1.6403348 0.5381183 0.9071227 -0.39464036 -2.0591626 0.8431229
 1.7413149 0.539336 -1.1917838 -1.2384293 -0.4398976 -1.6367152
-1.4840738 -0.627899 -0.42461908 -2.658467 -0.72780097 -1.6159099
-1.5800112 -0.11509487 0.70857453 0.04585471 0.6500463 -0.49665284
 1.4485593 0.63753265 -0.32634377 1.2377899 0.60685885 1.5268406
 0.0185772 -1.6065822 -1.953324 -1.3696564 0.51373667 -1.1270518
-1.3647636 -2.259235 -0.59813976 2.0049155 1.2963461 -0.19201212
-0.01453091 1.4761349 -1.4668998 0.8621714 0.1406255 0.9118301
 1.6768305 0.8456932 1.0728552 3.8460395 -1.042035 -0.9609054
-1.6206812 -1.6337742 1.514563 0.46056068 0.58360505 -0.6286432
-0.48522118 1.0121975 0.5529992 0.11600122 0.16986687 0.29172108
-0.4958315 -0.8328327 -0.5060014 -0.06495776 0.9188782 1.1617895
 0.30785733 0.24330127 -1.9496747 0.16663222 -0.9334646 -0.73748285
 0.07386508 3.8498254 -0.24335375 -0.00465591 -0.62488025 -0.01229596
-2.898337 0.31089088 0.2880013 1.2045447 0.33621657 2.5970545
 0.6993257 0.18010518 0.06332541 -0.96654975 2.0137794 0.566853
 2.0639496 0.7666444 -2.4160697 -1.6420139 -0.43061116 -0.26309547
-0.6882184 0.22019571 0.7990433 -1.999571 1.4254533 0.23616247
 1.521013 1.3552173 1.1136663 -0.03954174 0.1312558 -0.2654455
 4.011512 -0.4787183 0.08465659 0.7792323 1.06037 0.30956754
 0.30938998 -0.529427 -0.4078776 -1.2024792 -2.3178 -0.11798641
-1.2008193 -0.14821827 0.5779223 1.6811693 3.1992242 -0.6381721
-1.5767831 0.3728946 -0.41529638 0.7604324 0.7289358 -0.3689356
-0.6683996 -1.0957711 -0.98804307 1.0106868 1.3775774 0.5767862
 2.4132223 -0.15895717 0.6187072 0.54956925 1.1977834 -0.74189883
-0.01634382 0.8567224 0.97765553 1.2571958 1.4862683 -1.8481039
 2.69007 1.1754425 2.2660055 1.5544916 -1.086744 0.08545934
 0.8512228 -2.7127392 -1.4737937 -0.50650424 -1.5138463 -1.3879019
 0.7369825 0.01800244 1.5621387 1.379208 -0.53531 1.844126
-0.41027883 0.6954171 0.6671407 1.0174766 0.60396457 1.853182
 0.32635555 -2.7269647 0.34600484 1.9888045 1.3891473 0.5277364
-0.44188654 -0.0885203 -0.35421345 0.27191442 0.24312034 0.04790611
 2.2092607 -0.12326286 -0.8810332 0.24440219 0.3841828 0.8861823
-0.10411951 2.4050674 -0.85098165 0.48661023]
```

```
In [21]: cbowmodel.get_nearest_neighbors("數學家")
```

```
Out[21]: [(0.8219419121742249, '幾何學家'),
 (0.7725763320922852, '光學家'),
 (0.7702062129974365, '聲學家'),
 (0.7606274485588074, '物理學家'),
 (0.7409933805465698, '邏輯學家'),
 (0.7294262051582336, '地學家'),
 (0.7277169823646545, '道學家'),
 (0.7199861407279968, '修辭學家'),
 (0.7162450551986694, '啟智班'),
 (0.7153647541999817, '文法學家')]
```

```
In [22]: cbowmodel.get_nearest_neighbors("飲料")
# model.wv.most_similar("飲料", topn=10)
```

```
Out[22]: [(0.8788970708847046, '輝劍'),
(0.803864061832428, '名松'),
(0.80056232213974, '飲料類'),
(0.7759467959403992, '飲料機'),
(0.7476385831832886, '飲品'),
(0.7389094829559326, '飲料罐'),
(0.7211691737174988, '軟飲料'),
(0.6965533494949341, '茶飲料'),
(0.6946642398834229, '經米濱'),
(0.6877132058143616, '乳酸飲料')]
```

```
In [23]: cbowmodel.get_nearest_neighbors("car")
```

```
Out[23]: [(0.7211601734161377, 'boxcar'),
(0.6991508603096008, 'motorcar'),
(0.6957062482833862, 'carcar'),
(0.6935480237007141, 'cars'),
(0.6817461252212524, 'epcar'),
(0.681530237197876, 'stcar'),
(0.6799877285957336, 'ncar'),
(0.6782801747322083, 'zipcar'),
(0.6744192838668823, 'ccar'),
(0.6733412146568298, 'necar')]
```

```
In [24]: cbowmodel.get_nearest_neighbors("facebook")
```

```
Out[24]: [(0.8913792371749878, 'youtubefacebook'),
(0.8290598392486572, 'thefacebook'),
(0.7335013747215271, 'instagram'),
(0.7030842900276184, 'youtube'),
(0.6779652833938599, '臉書'),
(0.6687554121017456, '網站'),
(0.6591519713401794, 'lnstagram'),
(0.6555401682853699, '專頁'),
(0.6479989886283875, 'twitteryoutube'),
(0.6423900723457336, 'youtubers')]
```

```
In [25]: cbowmodel.get_nearest_neighbors("詐欺")
```

```
Out[25]: [(0.7401989102363586, '中境'),
(0.7256292700767517, '賈邱'),
(0.6625373959541321, '欺詐'),
(0.6587205529212952, '詐欺罪'),
(0.6556805968284607, '越中境'),
(0.6005792021751404, '赤坑鎮'),
(0.581577479839325, '詐騙'),
(0.5316701531410217, '詐騙者'),
(0.5108726024627686, '聚少'),
(0.49943235516548157, '竊盜')]
```

```
In [26]: cbowmodel.get_nearest_neighbors("合約", k=10)
```

```
Out[26]: [(0.7632400393486023, '合同'),
(0.724915623664856, '德康'),
(0.7028045654296875, '簽約'),
(0.7019715309143066, '續約'),
(0.6697090864181519, '綠蠅'),
(0.6487407684326172, '合同額'),
(0.6303138732910156, '租約'),
(0.6205465197563171, '合同商'),
(0.6112428307533264, '合同期'),
(0.6108682155609131, '合同條款')]
```

```
In [27]: model = fasttext.load_model('newfasttext_model.bin')
```

```
Warning : `load_model` does not return WordVectorModel or SupervisedModel any more, but a `FastText` object which is very similar.
```

```
In [28]: model.get_nearest_neighbors('飲料', k=10)
```

```
Out[28]: [(0.8788970708847046, '輝劍'),
          (0.803864061832428, '名松'),
          (0.80056232213974, '飲料類'),
          (0.7759467959403992, '飲料機'),
          (0.7476385831832886, '飲品'),
          (0.7389094829559326, '飲料罐'),
          (0.7211691737174988, '軟飲料'),
          (0.6965533494949341, '茶飲料'),
          (0.6946642398834229, '經米濱'),
          (0.6877132058143616, '乳酸飲料')]
```

```
In [29]: word = "這肯定沒見過 "
```

```
# 若強行取值會報錯
try:
    vec = model[word]
except KeyError as e:
    print(e)
```

```
In [ ]:
```