Subsampling 적용시

```
(base) krsnik@krsnik-MS-7A37:~$ python fastText.py 20 part
loading...
preprocessing...
build training set...
Vocabulary size
1815

# of training samples
2896200

Progress: |#----------------------------------------------| 1.7% Complete Loss : 13.837511
Progress: |##---------------------------------------------| 3.5% Complete Loss : 7.799623
Progress: |###--------------------------------------------| 5.2% Complete Loss : 5.065654
Progress: |###--------------------------------------------| 6.9% Complete Loss : 4.235257
Progress: |####-------------------------------------------| 8.6% Complete Loss : 3.378349
Progress: |#####------------------------------------------| 10.4% Complete Loss : 3.819187
Progress: |######                                          12.1% Complete Loss : 3.803500
```

```
=============================================
The most similar words to "narrowmindedness"
killed   : tensor(0.9512)
carry    : tensor(0.9495)
dramatic         : tensor(0.9463)
biological       : tensor(0.9459)
shall    : tensor(0.9453)
=============================================


=============================================
The most similar words to "department"
treatment        : tensor(0.9706)
development       : tensor(0.9695)
railroad         : tensor(0.9685)
carthage         : tensor(0.9675)
processing       : tensor(0.9672)
=============================================


=============================================
The most similar words to "campfires"
significant      : tensor(0.9569)
industry         : tensor(0.9560)
cultural         : tensor(0.9558)
zone     : tensor(0.9552)
separate         : tensor(0.9551)
=============================================


=============================================
The most similar words to "knowing"
allowing         : tensor(0.9727)
thinking         : tensor(0.9681)
growing : tensor(0.9667)
following        : tensor(0.9663)
string   : tensor(0.9643)
=============================================


=============================================
The most similar words to "urbanize"
university       : tensor(0.9593)
surrounding      : tensor(0.9583)
creek    : tensor(0.9580)
industrial       : tensor(0.9566)
branden : tensor(0.9563)
=============================================


=============================================
The most similar words to "imperfection"
collection       : tensor(0.9742)
election         : tensor(0.9729)
protection       : tensor(0.9723)
fiction : tensor(0.9711)
elections        : tensor(0.9693)
=============================================
```

```
=============================================
The most similar words to "principality"
december         : tensor(0.9768)
transcontinental         : tensor(0.9762)
desert   : tensor(0.9757)
military         : tensor(0.9748)
characteristics : tensor(0.9747)
=============================================


=============================================
The most similar words to "abnormal"
normal   : tensor(0.9775)
formal   : tensor(0.9654)
potential        : tensor(0.9623)
formally         : tensor(0.9606)
regarded         : tensor(0.9593)
=============================================


=============================================
The most similar words to "secondary"
second   : tensor(0.9790)
secretary        : tensor(0.9738)
continues        : tensor(0.9718)
ecological       : tensor(0.9712)
continue         : tensor(0.9711)
=============================================


=============================================
The most similar words to "ungraceful"
ever     : tensor(0.9073)
movies   : tensor(0.9005)
feeling : tensor(0.8999)
anarchist        : tensor(0.8990)
about    : tensor(0.8981)
=============================================
```

Subsampling 미적용시


```
error_log.txt  Memo.txt
(base) krsnik@krsnik-MS-7A37:~$ python fastText.py 20 part
loading...
preprocessing...
build training set...
Vocabulary size
3971

# of training samples
20004690

Progress: |--------------------------------------------| 0.2% Complete Loss : 12.106852
Progress: |--------------------------------------------| 0.5% Complete Loss : 7.645260
Progress: |--------------------------------------------| 0.7% Complete Loss : 5.531726
Progress: |--------------------------------------------| 1.0% Complete Loss : 4.545908
Progress: |#-------------------------------------------| 1.2% Complete Loss : 4.129728
Progress: |#-------------------------------------------| 1.5% Complete Loss : 3.821685
Progress: |#-------------------------------------------| 1.7% Complete Loss : 3.425719
Progress: |#-------------------------------------------| 2.0% Complete Loss : 3.400018
Progress: |#-------------------------------------------| 2.2% Complete Loss : 3.521554
Progress: |#-------------------------------------------| 2.5% Complete Loss : 3.422898
Progress: |#-------------------------------------------| 2.7% Complete Loss : 3.182032
```

```
============================================
The most similar words to "narrowmindedness"
intelligent    : tensor(0.9817)
aggressive     : tensor(0.9809)
corruption     : tensor(0.9803)
stereotyped    : tensor(0.9800)
intelligence   : tensor(0.9797)
============================================

============================================
The most similar words to "department"
department     : tensor(1.)
treatment      : tensor(0.9790)
investment     : tensor(0.9788)
category       : tensor(0.9765)
scheduled      : tensor(0.9754)
============================================

============================================
The most similar words to "campfires"
expanded       : tensor(0.9641)
majority       : tensor(0.9638)
manifesto      : tensor(0.9633)
melting : tensor(0.9632)
helping : tensor(0.9630)
============================================

============================================
The most similar words to "knowing"
knowing : tensor(1.0000)
nevertheless   : tensor(0.9806)
programming    : tensor(0.9797)
studying       : tensor(0.9782)
describing     : tensor(0.9778)
============================================

============================================
The most similar words to "urbanize"
urban   : tensor(0.9790)
restored         : tensor(0.9744)
density : tensor(0.9740)
heavens : tensor(0.9735)
melbourne        : tensor(0.9733)
============================================
```

```
============================================
The most similar words to "imperfection"
protection     : tensor(0.9801)
connection     : tensor(0.9793)
direction      : tensor(0.9778)
collection     : tensor(0.9778)
selection      : tensor(0.9768)
============================================

============================================
The most similar words to "principality"
principal      : tensor(0.9912)
principle      : tensor(0.9891)
principles     : tensor(0.9884)
incompetent    : tensor(0.9874)
autonomous     : tensor(0.9852)
============================================

============================================
The most similar words to "abnormal"
normal  : tensor(0.9771)
treatises      : tensor(0.9687)
mixture : tensor(0.9681)
functioning    : tensor(0.9680)
friedrich      : tensor(0.9680)
============================================

============================================
The most similar words to "secondary"
second  : tensor(0.9802)
seconds : tensor(0.9794)
census  : tensor(0.9648)
struggle       : tensor(0.9639)
interior       : tensor(0.9639)
============================================

============================================
The most similar words to "ungraceful"
highest : tensor(0.9565)
forcing : tensor(0.9551)
atmosphere     : tensor(0.9542)
ancestry       : tensor(0.9540)
stadler : tensor(0.9535)
============================================
```

컴퓨터 사양이 좋지가 않아서 part로만 실험을 할 수 있었는데, part로 테스트하면 학습량이 많지가 않아서 오히려 subsampling을 하지 않았을 때가 정확도 및 타당성이 더 높았습니다. (knowing, department 단어가 일치하는 단어가 그대로 나온 점, secondary 단어의 결과 차이 등등) 요구사항에는 subsampling을 하라고 되어있지만 실험 결과 유의미한 요소라고 생각해서 두가지 버전 모두 제출해봅니다. 제출한 코드는 일단 subsampling이 되도록 만든 버전이며, python fastText.py 20 part(or full) 로 실행하시면 됩니다.