**Word2Vec\_SkipGram**

1. Sources:

* Training\_text.txt
* stopwords.txt

1. Flow:

* Remove all the stopwords and the punctuation marks
* Choose embedding\_size (the num of dimensions) = 5 (because of too many topics, 2 axes are not enough to express the closeness of the words)
* Generate random matrices of values for vectors for each word in center\_words and context\_words => make the length of each vector = 1 (to avoid dot\_prod of any word is too big for similar words)
* No update thru 25 epochs:
* Set the window\_size (the num of neighbors on each side) = 3, the num of random unrelated words in the text = 3 (to label 0 to train together with 1)
* Loop to collect each center word and its neighbors and the random unrelated words
* Cal the score, errors (using sigmoid: 1 is very similar, 0 is very different)
* Update by dif.errors.learning\_rate (make what’s supposed to be closer closer, what’s supposed to be farer farer)

\***Note**: afte update here, the length changes

* Normalize the vector length