汉语词向量

模型：Skip-Gram with Negative Sampling (SGNS)

超参数：

前后2窗口

维数：100

训练语料：

https://dumps.wikimedia.org/backup-index.html

汉语数据

可能的话，自己尝试写代码，不行的话可以用库。

获得词向量后，利用余弦距离来计算pku\_sim\_test.txt文件中每行两个词之间的相似度，最终是要输出每行两个词之间的余弦距离值。

结果输出要求(因为是机器判定，请一定按如下格式输出)：

1)输出文件的编码： utf-8

2)输出格式：词之间以及词和相似度之间使用一个tab符分开，如下例：

没戏 没辙 4.3

3)不要打乱pku\_sim\_test.txt中原来的行序，输出文件以自己的学号命名。

4)当pku\_sim\_test.txt中的某个词没有获得词向量时，对应的该行的词间相似度标识为OOV，即，如果没有“没辙”这个词的词向量，则输出：

没戏 没辙 OOV