Assignment Task2

王柯然 2100017727

2023 年 10 月 6 日

1.计算co-occurance的probablity and ratio
设置windowsize=5，采用TreebandWordTokenizer作为分词器
总共出现ice次数：8455次
总共出现steam次数：800次
在ice附近出现的各词次数：
solid：15次，gas：4次，water：118次，fashion：0次
在steam附近出现的各词次数：
solid：1次，gas：4次，water：53次，fashion：0次
可以得到如下表格：

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.77×10^{-3}	4.73×10^{-4}	1.39×10^{-2}	0
$P(k\|steam)$	1.25×10^{-3}	5×10^{-3}	6.63×10^{-3}	0
$P(k\|ice)/P(k\|steam)$	1.42	9.46×10^{-2}	2.10	/

2.使用glove：
a.
最接近physics的词语：chemistry
最接近north的词语：north
最接近queen的词语：king
最接近car的词语：motor
b.
见下表

	France vs. Spain	tree vs. water	water vs. sky	sky vs. bird
cosine similarity	0.274	0.494	0.724	0.677

c.
与text最接近的5个单词：sounds、sound、type、translation、code