本周进度

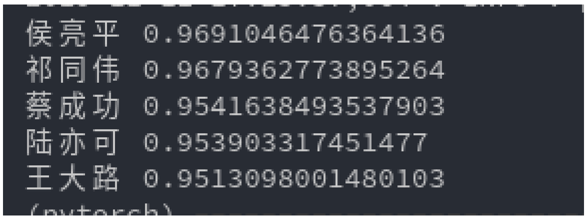
- **word2vec Demo**
- **中文Demo** -- 未完成

word2vec Demo

- **源代码**

```python
import logging
from gensim.models import word2vec

logging.basicConfig(
    format='%(asctime)s : %(levelname)s : %(message)s',
level=logging.INFO)

sentences =
word2vec.LineSentence('./in_the_name_of_people_segment.txt')

model = word2vec.Word2Vec(sentences, hs=1, min_count=1, window=3,
size=100)

# 找出某一个词向量最相近的词集合
req_count = 5
for key in model.wv.similar_by_word('李达康', topn=100):
    if len(key[0]) == 3:
        req_count -= 1
        print(key[0], key[1])
        if req_count == 0:
            break
```

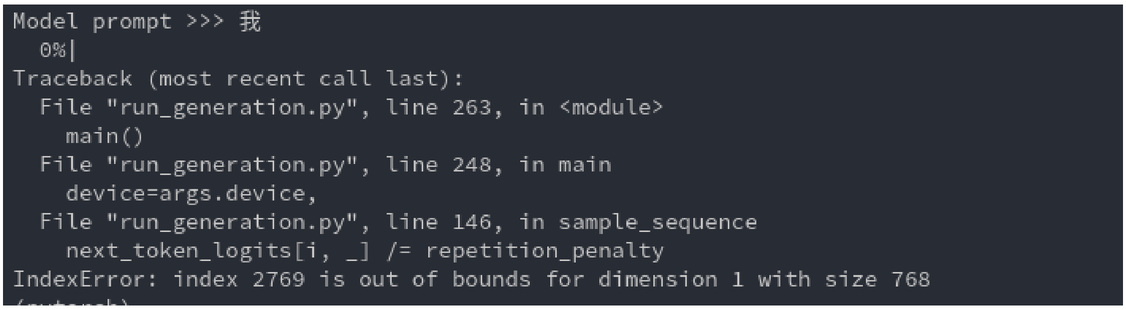
- **运行结果**



```
侯亮平 0.9691046476364136
祁同伟 0.9679362773895264
蔡成功 0.9541638493537903
陆亦可 0.953903317451477
王大路 0.9513098001480103
(pytorch)
```

中文 Demo -- 未完成

- 使用 transformers 官方 example，使用BERT模型运行时报错，需要对代码进行修改
- **错误信息**



```
Model prompt >>> 我
  0%|
Traceback (most recent call last):
  File "run_generation.py", line 263, in <module>
    main()
  File "run_generation.py", line 248, in main
    device=args.device,
  File "run_generation.py", line 146, in sample_sequence
    next_token_logits[i, _] /= repetition_penalty
IndexError: index 2769 is out of bounds for dimension 1 with size 768
```

下周任务

- CS224n 课程学习
- 读论文
- 中文Demo