

基于Word2Vec的问题空间收束方案

1. 问题空间收束

1.1 问题空间

问题空间是指所有可能的问题的集合，问题空间收束是指在问题空间中，对于一个给定的问题，找到与之最相似的问题。

1.2 问题相似度

问题相似度是指两个问题之间的相似程度，相似度越高，两个问题越相似。问题相似度的计算方法有很多，这里采用的是基于Word2Vec的方法。

1.3 Word2Vec

Word2Vec是一种将词语映射到向量空间的方法，它的基本思想是：对于一个给定的词语，它的上下文词语越相似，它们的向量越接近。

这里采用python中gensim的Word2Vec模块，具体的使用方法可以参考[这里](#)。

2. 问题空间收束方案

2.1 问题空间收束方案

问题空间收束方案是指在问题空间中，对于一个给定的问题，找到与之最相似的问题的方法。这里采用的方法是：对于一个给定的问题，首先将其转换为一个向量，然后计算该向量与问题空间中所有问题向量的相似度，最后取相似度最高的前k个问题作为与之最相似的问题。

2.2 问题向量的计算

问题向量的计算方法有很多，这里采用的是：将问题中的每个词语转换为一个向量，然后将所有词语向量相加，最后取平均值作为问题向量。这里的平均值是加权平均池化方法。

3. 问题空间收束方案的实现

3.1 问题空间收束方案的实现

问题空间收束方案的实现是指将问题空间收束方案实现为一个python模块，在这里是demo/QASNet.py，集成了句子向量计算以及最邻近搜索最相似前k个问题返回的功能，集成在ann_search(user_q, top_n)函数中，其中user_q是用户输入的问题，top_n是最相似问题的个数。

具体的实现可以参考demo/QASNet.py。

这里放上目前的方法在一个规模极小的数据集上的测试结果：

```
(virtualguide) D:\Proj\VirtualGuide\demo>python tools.py --mode=test --input-file-
path=./test_data/data.json --model-path=./model/zhwiki.model
launching test mode...
Please input your question: 西湖周边的住宿和酒店有什么推荐的?
-----
top 3 questions and answers:
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\HUAWEI\AppData\Local\Temp\jieba.cache
Loading model cost 0.400 seconds.
Prefix dict has been built successfully.

question: 在西湖边有哪些好的住宿选择?
answer: 西湖边有多个好的住宿选择, 如凯悦酒店、西湖世家度假村、西湖银泰阁等。
similariry: 0.857102392861801

question: 西湖边的民宿如何? 有哪些推荐?
answer: 西湖边的民宿备受游客欢迎, 较受欢迎的有半山墅院、绿野仙踪、烟雨楼等。
similariry: 0.8564139706365645

question: 西湖周边有哪些购物场所?
answer: 西湖周边有很多购物场所, 如西湖银泰城、河坊街、南宋御街等。
similariry: 0.8261205140200423
-----
Please input your question: 夏季西湖热不热
-----
top 3 questions and answers:

question: 西湖边的天气如何?
answer: 西湖边的天气四季分明, 春秋宜人, 夏季略有些炎热, 冬季寒冷干燥。
similariry: 0.7040561573154875

question: 西湖门票价格如何?
answer: 西湖门票价格因景点不同而异, 一般在30元到150元之间。
similariry: 0.6870606311419293

question: 西湖的水质如何?
answer: 西湖的水质非常优良, 水清鱼美。
similariry: 0.6803660202016528
-----
Please input your question: 西湖历史文化故事
-----
top 3 questions and answers:

question: 西湖边有哪些好的文化活动?
answer: 西湖边有很多好的文化活动, 如杭州西湖国际文化艺术交流中心举办的各种艺术展览和表演、西湖博物馆的讲座和展览等。
similariry: 0.6910760358978711

question: 西湖的夜景如何?
answer: 西湖的夜景非常美丽, 游客可以观赏湖岸灯光秀、水上舞台秀等夜间表演。
similariry: 0.5669785532717239

question: 西湖边有哪些著名的美食?
```

```
answer:  西湖边有很多著名的美食，如西湖醋鱼、龙井虾仁、宋嫂鱼羹、锅盔等。
similariry:  0.5493400511785823
-----
Please input your question: 西湖推荐的旅游路线有拿些？
-----
top 3 questions and answers:

question:  西湖边有哪些好的旅游路线？
answer:  可以选择游览西湖周边的著名景点，如孤山、花港观鱼、曲院风荷等，也可以选择参观西湖周边的历史文化遗址，如雷峰塔、岳王庙等。
similariry:  0.870610812975461

question:  西湖边的民宿如何？有哪些推荐？
answer:  西湖边的民宿备受游客欢迎，较受欢迎的有半山墅院、绿野仙踪、烟雨楼等。
similariry:  0.7724447893426437

question:  在西湖边有哪些好的住宿选择？
answer:  西湖边有多个好的住宿选择，如凯悦酒店、西湖世家度假村、西湖银泰阁等。
similariry:  0.7416500471944119
-----
```