Assignment3.part2. QuestionAnswer

MF1733071，严德美，1312480794@qq.com

2017 年 12 月 21 日

1 任务描述

使用给定的课本知识（以句子为单位）和训练数据，训练数据主要是标好的地理选择题（正确和错误答案已经标出），设计高考地理选择题答题系统，对于测试数据中每一个问题选出正确答案。

2 简述分析与设计

将所给的课本知识和训练数据（背景知识，问题和正确选项拼接在一起）整合在一起作为测试数据的知识库，使用测试数据中给定的背景材料和问题去知识库中寻找最相似的top_k个知识点，对于测试数据中问题对应的四个选项，计算每个选项和top_k个知识点的最大的相似度，最后以四个选项中相似度最大的那个选项作为正确答案，主要流程如下图。计算相似度，主要是使用中文jieba分词工具先进行分词，然后将句子表示成关于tf-idf的文档向量，最后应用lsi模型计算余弦相似度。

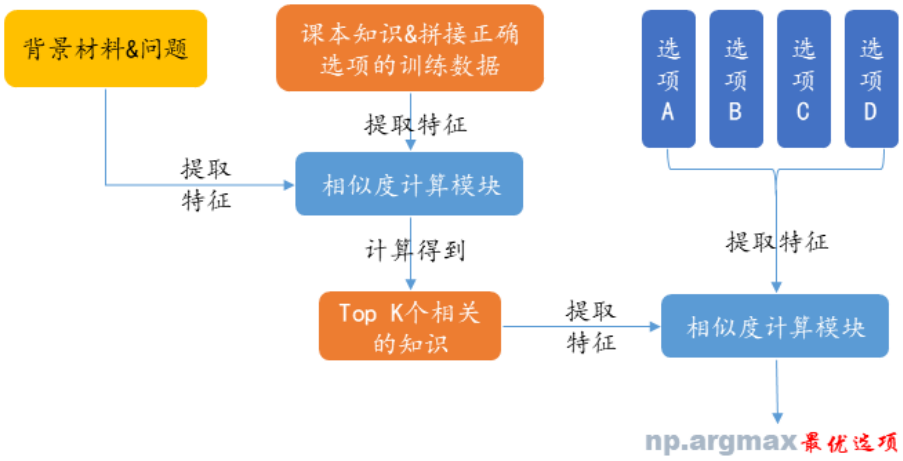

图 1: 主要方法

3 遇到的问题与解决方案

从网上找了相关论文，现在主流的都是神经网络，做的好的也是使用神经网络，但是由于知识水平和时间的局限性，没有对其中的神经网络进行实现，而是最后选择了最为简单的余弦相似度。

4 运行方式和依赖的包

直接执行questionanswer.py文件，输出答题的准确率，外部依赖：jieba,gensim，安装方式：pip install jieba,pip install gensim

5 运行结果

最终答题的准确率为37.84%，运行结果如下图。

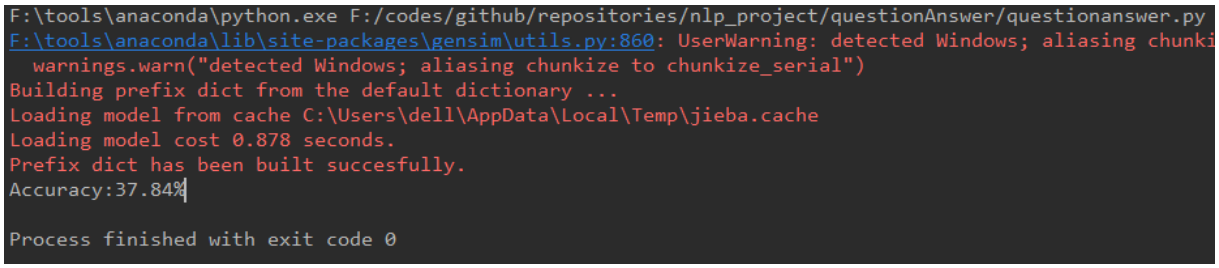

图 2: 运行结果