

流风，飘然的风

在关键时刻听从自己内心的声音！

博客园 首页 新闻 新随笔 联系 管理 订阅

随笔- 504 文章- 0 评论- 150

昵称：流风，飘然的风

园龄：8年11个月

粉丝：250

关注：3

+加关注

< 2017年12月 >						
日	一	二	三	四	五	六
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

搜索

<input type="text"/>	找找看
<input type="text"/>	谷歌搜索

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签
更多链接

用TensorFlow实现文本分析模型，做个聊天机器人

用TensorFlow实现文本分析模型，做个聊天机器人

聊天机器人的架构简图

用 TensorFlow 实现 Chatbot 的模型

如何准备 chatbot 的训练数据

Chatbot 源码解读

1. 聊天机器人的架构简图

聊天机器人是可行的

我：chatbot，你好！

chatbot：你也好！

我：聊天机器人可行吗？

chatbot：你不要怀疑这是天方夜谭，我不就在这里吗？世界上还有很多跟我一样聪明的机器人呢，你听过IBM公司在2010年就研发出来了的Watson问答系统吗？它可比我要聪明100倍呢

我：噢，想起来了，据说Watson在智力竞赛中竟然战胜了人类选手。但是我了解到它有一些缺陷：因为它还只是对信息检索技术的综合运用，并没有进行各种语义关系的深刻计算，所以它能回答的问题也仅限于事实类的问题，所以它能赢得也就是知识类的智力竞赛，如果你给它出个脑筋急转弯，它就不行了

chatbot：是的呢，所以你任重道远啊

聊天机器人工作原理是什么

最新随笔

1. java实现 HTTP/HTTPS请求绕过证书检测代码实现
2. pyspider源码解读--调度器scheduler.py
3. pyspider操作千万级库，pyspider在对接量级较大库的策略
4. 尼克《人工智能简史》谈人工智能的历史、现实与未来
5. CentOS7使用yum命令安装Java1.8
6. php ci nginx 伪静态rewrite配置方法
7. vim编辑文件警告Swap file already exists，如何删除vim编辑产生的.swp文件？查看隐藏文件命令
8. php Allocator Jemalloc TCMalloc那个内存分配器比较好？
9. AlphaGo设计师黄士杰：“最强的学习技能在人类的脑袋里”
10. php 获取ip地址的5种方法，插入用户登录日志实例

我的标签

php(50)
python(43)
java(34)
python3(32)
android(31)
linux(30)
mysql(24)
深度学习(21)
人工智能(20)
Js(17)
更多

随笔分类(893)

android(57)
bootsharp(15)
c++(3)
CentOS服务器(60)
hadoop(9)

我：chatbot，我问的每一句话，你都是怎么处理并回答我的呢？

chatbot：我身体里有三个重要模块：提问处理模块、检索模块、答案抽取模块。三个模块一起工作，就能回答你的问题啦

我：是嘛，那么这个提问处理模块是怎么工作的呢？

chatbot：提问处理模块要做三项重要工作：查询关键词生成、答案类型确定、句法和语义分析。

我：那么这个查询关。。。

chatbot：别急别急，听我一个一个讲给你听。查询关键词生成，就是从你的提问中提取出关键的几个关键词，因为我本身是一个空壳子，需要去网上查找资料才能回答你，而但网上资料那么多，我该查哪些呢？所以你的提问就有用啦，我找几个中心词，再关联出几个扩展词，上网一搜，一大批资料就来啦，当然这些都是原始资料，我后面要继续处理。再说答案类型确定，这项工作是为了确定你的提问属于哪一类的，如果你问的是时间、地点，和你问的是技术方案，那我后面要做的处理是不一样的。最后再说这个句法和语义分析，这是对你问题的深层含义做一个剖析，比如你的问题是：聊天机器人怎么做？那么我要知道你要问的是聊天机器人的研发方法

我：原来是这样，提问处理模块这三项工作我了解了，那么检索模块是怎么工作的呢？

chatbot：检索模块跟搜索引擎比较像，就是根据查询关键词所信息检索，返回句子或段落，这部分就是下一步要处理的原料

我：那么答案抽取模块呢？

chatbot：答案抽取模块可以说是计算量最大的部分了，它要通过分析和推理从检索出的句子或段落里抽取和提问一致的实体，再根据概率最大对候选答案排序，注意这里是“候选答案”噢，也就是很难给出一个完全正确的结果，很有可能给出多个结果，最后还在再选出一个来

我：那么我只要实现这三个模块，就能做成一个你喽？

chatbot：是的

聊天机器人的关键技术

我：chatbot，小弟我知识匮乏，能不能告诉我都需要学哪些关键技术才能完成我的梦想

chatbot：小弟。。。我还没满月。说到关键技术，那我可要列一列了，你的任务艰巨了：

1) 海量文本知识表示：网络文本资源获取、机器学习方法、大规模语义计算和推理、知识表示体系、知识库构建；

html5(38)
iphone-ios(11)
java(49)
javascript(55)
linux(62)
lua(3)
NLP(5)
nodejs(13)
php(74)
python(44)
SEO(17)
UI设计(4)
web前端(8)
蓝牙开发(2)
人工智能(40)
数据库(42)
思维导图(14)
微信公众号(25)
系统架构(9)
杂文随笔(53)
综合(181)

随笔档案(504)

2017年12月 (4)
2017年11月 (16)
2017年10月 (9)
2017年9月 (22)
2017年8月 (13)
2017年7月 (32)
2017年6月 (27)
2017年5月 (7)
2017年4月 (23)
2017年3月 (16)
2017年2月 (9)
2017年1月 (13)
2016年12月 (22)
2016年11月 (6)
2016年10月 (23)
2016年9月 (17)

2) 问句解析：中文分词、词性标注、实体标注、概念类别标注、句法分析、语义分析、逻辑结构标注、指代消解、关联关系标注、问句分类（简单问句还是复杂问句、实体型还是段落型还是篇章级问题）、答案类别确定；

3) 答案生成与过滤：候选答案抽取、关系推演（并列关系还是递进关系还是因果关系）、吻合程度判断、噪声过滤

聊天机器人的技术方法

我：chatbot，我对聊天机器人的相关技术总算有所了解了，但是我具体要用什么方法呢？

chatbot：看你这么好学，那我就多给你讲一讲。聊天机器人的技术可以分成四种类型：1) 基于检索的技术；2) 基于模式匹配的技术；3) 基于自然语言理解的技术；4) 基于统计翻译模型的技术。这几种技术并不是都要实现，而是选其一，听我给你说说他们的优缺点，你就知道该选哪一种了。基于检索的技术就是信息检索技术，它简单易实现，但无法从句法关系和语义关系给出答案，也就是搞不定推理问题，所以直接pass掉。基于模式匹配的技术就是把问题往已经梳理好的几种模式上去靠，这种做推理简单，但是模式我们涵盖不全，所以也pass掉。基于自然语言理解就是把浅层分析加上句法分析、语义分析都融入进来做的补充和改进。基于统计翻译就是把问句里的疑问词留出来，然后和候选答案资料做配对，能对齐了就是答案，对不齐就对不起了，所以pass掉。选哪个知道了吗？

我：知道了！基于自然语言理解的技术！so easy！妈妈再也不用担心我的学习！o(′□′)o

chatbot：好，那我的任务结束啦，我该走了，再见！

我：不要走——不要走——ka ji ma——

在呼唤chatbot的声音中，我从梦中醒来，这时我才意识到这是一个梦，揉揉双眼，突然发现床边有一张纸，上面写着：

你会成功的！读了这篇文章的人也会成功的！加油！

聊天机器人的工作流程大体为：提问 - 检索 - 答案抽取。

提问：就是要分析主人的问句中关键词，提问类型，还有真正想知道的东西。

检索：根据前一步的分析，去找答案。

答案抽取：找到的答案，并不能直接应用，还要整理成真正有用的，可以作为答案的回答。

2016年8月 (2)
2016年7月 (3)
2016年6月 (25)
2016年5月 (13)
2016年4月 (9)
2016年3月 (6)
2016年2月 (3)
2016年1月 (10)
2015年12月 (11)
2015年11月 (5)
2015年10月 (9)
2015年9月 (7)
2015年8月 (1)
2015年7月 (18)
2015年6月 (7)
2014年6月 (15)
2014年5月 (8)
2014年4月 (1)
2014年2月 (1)
2014年1月 (7)
2013年11月 (3)
2013年10月 (1)
2013年9月 (1)
2013年8月 (3)
2013年7月 (1)
2013年6月 (5)
2013年3月 (5)
2013年2月 (2)
2012年12月 (5)
2012年11月 (7)
2012年10月 (3)
2012年9月 (6)
2012年8月 (7)
2012年7月 (2)
2012年6月 (2)
2012年5月 (3)
2012年4月 (5)
2012年2月 (2)
2012年1月 (10)

涉及到的关键技术如图中所示。



看不清图的话，就是酱紫：

问句解析：

中文分词、词性标注、实体标注、概念类别标注、句法分析、语义分析、逻辑结构标注、指代消解、关联关系标注、问句分类、答案类别确定；

海量文本知识表示：

网络文本资源获取、机器学习方法、大规模语义计算和推理、知识表示体系、知识库构建

答案生成与过滤：

候选答案抽取、关系推演、吻合程度判断、噪声过滤

2. 用 TensorFlow 实现 Chatbot 的模型

之前有根据 Siraj 的视频写过一篇《自己动手写个聊天机器人吧》，

<http://www.jianshu.com/p/d0f4a751012b>

文章里只写了主函数的简单过程：Data - Model - Training，是用 Lua 实现的。

下面这篇文章是用 TensorFlow + tflearn 库 实现，在 建模，训练 和 预测 等环节可以学到更多细节：

学习资源：自己动手做聊天机器人 三十八-原来聊天机器人是这么做出来的

<http://www.shareeditor.com/blogshow/?blogId=121>

两篇的共同点是都用了 Seq2Seq 来实现。

2011年12月 (1)

2009年1月 (10)

积分与排名

积分 - 584880

排名 - 182

最新评论

1. Re:linux复制指定目录下的全部文件到另一个目录中，linux cp 文件夹

哈哈，一个CP里面别别乾坤

--rayinnight

2. Re:在eclipse中安装properties插件PropertiesEditor及设置(附图)，ASCII码转换成中文

设置了，可以了

--apollo_chen

3. Re:tomcat输出servlet-api.jar - jar not loaded 解决办法

@Koori_Cc非常高兴能帮到你...

--流风，飘然的风

4. Re:tomcat输出servlet-api.jar - jar not loaded 解决办法

问题解决 ~ 感谢

--Koori_Cc

5. Re:MySQL字符串函数substring：字符串截取

@sunny_an在公司的mapper中看到的，从1开始取的。本来还以为是错误的，上网搜了一下，才发现mysql的substring是从1开始的。看来这些细节还是要注意啊。...

--相健欢

6. Re:升级adt插件后，eclipse突然出现Unable to build: the file dx.jar was not loaded from the SDK folder 错误

@东风吹又吹不错，解决问题就ok...

--流风，飘然的风

7. Re:升级adt插件后，eclipse突然出现Unable to build: the file dx.jar was not loaded from the SDK folder 错误

LSTM的模型结构为：



此图片来自微信公众平台

未经允许不可引用

细节的话可以直接去看上面那篇原文，这里 po 出建立模型阶段简要的流程图和过程描述：



此图片来自微信公众平台

未经允许不可引用

- 先将原始数据 300w chat 做一下预处理，即 切词，分为 问答对。
- 然后用 word2vec 训练出词向量，生成二进制的词向量文件。



此图片来自微信公众平台

未经允许不可引用

作为 Input data X 传入下面流程：

- question 进入 LSTM 的 encoder 环节，answer 进入 decoder 环节，
- 分别生成 output tensor。
- 其中 decoder 是一个词一个词的生成结果，将所有结果加入到一个 list 中。

我发现是Android SDK Build-tools版本问题
我的ADT是24.0.2,我把Android SDK Build-tools升级到最新26.X版本,就出现这个问题了,我把26.X全部卸载.....

--东风吹又吹

8. Re:中国用户通过rchange用银联充值到PerfectMoney再给BTC-E充值进行搬砖的方法

@流风，飘然的风因为比特币分叉吗？...

--WY1116

9. Re:中国用户通过rchange用银联充值到PerfectMoney再给BTC-E充值进行搬砖的方法

@WY1116我很久没买比特币了，现在是高点风险比较大，暂时不适合进入...

--流风，飘然的风

10. Re:中国用户通过rchange用银联充值到PerfectMoney再给BTC-E充值进行搬砖的方法

你好，偶然看到你的这篇分享，了解到你也是比特币和山寨币投资者，我是刚入门，能否偶尔请假你一下？我有一些关于电子货币的渠道也可以分享给你，

--WY1116

11. Re:安心之法-斯多葛哲学学派，近年来在美国非常流行

@牛顿的小脑引用不错！！有勇气去改变那些可以改变的事；有度量去容忍那些不能改变的事；有智慧区别以上两类事。终于有活人评论了，博客园会评论的人太少了...

--流风，飘然的风

12. Re:安心之法-斯多葛哲学学派，近年来在美国非常流行

不错！！

有勇气去改变那些可以改变的事；有度量去容忍那些不能改变的事；有智慧区别以上两类事。

--牛顿的小脑

- 最后和 encoder 的输出，一起做为下一环节 Regression 的输入，并传入 DNN 网络。



此图片来自微信公众平台
未经允许不可引用

3. 如何准备 chatbot 的训练数据

学习资源：

自己动手做聊天机器人 三十八-原来聊天机器人是这么做出来的

<http://www.shareeditor.com/blogshow/?blogId=121>

训练数据的生成过程如下：

- 首先在 input file 里读取每一行，并根据 ' | ' 拆分成 question 和 answer 句子。
- 每个句子，都将 word 通过 word2vec 转化成词向量。
- 每一句的向量序列都转化成相同维度的形式：`self.word_vec_dim * self.max_seq_len`
- 最后 answer 构成了 y 数据，question + answer 构成了 xy 数据，再被投入到 model 中去训练：

```
model.fit(trainXY, trainY, n_epoch=1000, snapshot_epoch=False, batch_size=1)
```


13. Re:JavaScript调试技巧之console.log()
 详解
 引用if(!window.console){window.console =
 {log : function(){};}}Opera还是不能用 console.log 加上下面两句代码兼容： wind.....
 --xiashengwang

14. Re:tomcat输出servlet-api.jar - jar not loaded 解决办法
 谢谢楼主指教，同样的错误，遇到你真幸运
 --踏雪飞虹

15. Re:MySQL字符串函数substring：字符串截取
 @sunny_an谢谢你的反馈，已经修改了...
 --流风，飘然的风

16. Re:MySQL字符串函数substring：字符串截取
 您好，文章中关于substring的结果有点问题
 比如select substring('example.com', 4);字符是从1开始的，而不是类似于java从0开始的，所以正确的结果应该为mple.....
 --sunny_an

17. Re:python3.4学习笔记(十六) windows
 下面安装easy_install和pip教程
 @东北三环TypeError: chown() missing 1 required positional argument: 'numeric_owner'看这个提示应该是使用的安装帐号权限不够，用.....
 --流风，飘然的风

18. Re:python3.4学习笔记(十六) windows
 下面安装easy_install和pip教程
 我的出错信息是这个，这怎么解决呢？=====
 ===== RESTART: E:\Projects\ez_setup-0.9\ez_setup.py =====
 Extrac.....
 --东北三环

19. Re:python3.4学习笔记(十六) windows
 下面安装easy_install和pip教程
 @Fenjs是的，删掉再运行会重新根据代码生成。-----pyc



此图片来自微信公众平台

未经允许不可引用

代码如下：

```
def init_seq(input_file):
    """读取切好词的文本文件，加载全部词序列
    """
    file_object = open(input_file, 'r')
    vocab_dict = {}
    while True:
        question_seq = []
        answer_seq = []
        line = file_object.readline()
        if line:
            line_pair = line.split('|')
            line_question = line_pair[0]
            line_answer = line_pair[1]
            for word in line_question.decode('utf-8').split(' '):
                if word_vector_dict.has_key(word):
                    question_seq.append(word_vector_dict[word])
```

文件介绍pyc文件，是python编译后的字节码（bytecod.....

--流风，飘然的风

20. Re:python3.4学习笔记(十六) windows

下面安装easy_install和pip教程

老师你好，我找到这个文件了，第6行是从pip.commands.install import InstallCommand看不出个所以然。。解决问题的方法：1.可以排除，命名没有和Python.....

--Fenjs

阅读排行榜

1. Linux下的压缩zip,解压缩unzip命令详解及实例(249357)
2. MySQL字符串函数substring：字符串截取(174734)
3. 谈谈几个月以来开发android蓝牙4.0 BLE低功耗应用的感受(129465)
4. php 5.4中php-fpm 的重启、终止操作命令(128625)
5. linux 中更改用户权限和用户组的命令chmod,chmod实例(127300)
6. linux复制指定目录下的全部文件到另一个目录中，linux cp 文件夹(99696)
7. js中没有contains()方法可以用indexOf("str") != -1代替,一个改变自定义checkbox状态的代码(90192)
8. linux查看日志文件内容命令tail、cat、tac、head、echo(81499)
9. android 让一个控件按钮居于底部的几种方法(80949)
10. 谈谈android反编译和防止反编译的方法(64620)

评论排行榜

1. 谈谈几个月以来开发android蓝牙4.0 BLE低功耗应用的感受(43)
2. Google Maps API v2 android版本开发国内手机不支持google play Service相关问题解决--图文教程(20)

用TensorFlow实现文本分析模型，做个聊天机器人 - 流风，飘然的风 - 博客园

```
for word in line_answer.decode('utf-8').split(' '):
```

```
    if word_vector_dict.has_key(word):
```

```
        answer_seq.append(word_vector_dict[word])
```

```
    else:
```

```
        break
```

```
question_seqs.append(question_seq)
```

```
answer_seqs.append(answer_seq)
```

```
file_object.close()
```

```
def generate_trainig_data(self):
```

```
    xy_data = []
```

```
    y_data = []
```

```
    for i in range(len(question_seqs)):
```

```
        question_seq = question_seqs[i]
```

```
        answer_seq = answer_seqs[i]
```

```
        if len(question_seq) < self.max_seq_len and len(answer_seq) < self.max_seq_len:
```

```
            sequence_xy = [np.zeros(self.word_vec_dim)] * (self.max_seq_len-len(question_seq)) +
```

```
list(reversed(question_seq))
```

```
            sequence_y = answer_seq + [np.zeros(self.word_vec_dim)] * (self.max_seq_len-len(answer_seq))
```

```
            sequence_xy = sequence_xy + sequence_y
```

```
            sequence_y = [np.ones(self.word_vec_dim)] + sequence_y
```

```
            xy_data.append(sequence_xy)
```

```
            y_data.append(sequence_y)
```


3. android 最新 NDK r8 在window下开发环境搭建 安装配置与使用 详细图文讲解，完整实际配置过程记录(原创)(20)
4. 谈谈android反编译和防止反编译的方法(9)
5. python3.4学习笔记(十六) windows下面安装easy_install和pip教程(6)
6. MySQL字符串函数substring：字符串截取(5)
7. python3.4学习笔记(十三) 网络爬虫实例代码，使用pyspider抓取多牛投资吧里面的文章信息，抓取政府网新闻内容(4)
8. tomcat输出servlet-api.jar - jar not loaded 解决办法(3)
9. 中国用户通过rchange用银联充值到PerfectMoney再给BTC-E充值进行搬砖的方法(3)
10. 升级adt插件后，eclipse突然出现Unable to build: the file dx.jar was not loaded from the SDK folder 错误(3)

推荐排行榜

1. 谈谈几个月以来开发android蓝牙4.0 BLE低功耗应用的感受(17)
2. 谈谈android反编译和防止反编译的方法(9)
3. MySQL字符串函数substring：字符串截取(5)
4. linux查看日志文件内容命令tail、cat、tac、head、echo(5)
5. 用shell脚本监控进程是否存在 不存在则启动的实例(4)
6. Git和GitHub使用总结(3)
7. eclipse下编写android程序突然不会自动生成R.java文件和包的解决办法(3)
8. PHP获取当前类名、函数名、方法名(2)
9. Java继承时构造函数的关系(2)
10. Android Market google play store帐号注册方法流程 及发布应用注意事项(2)

```
return np.array(xy_data), np.array(y_data)
```

4. Chatbot 源码解读

提炼出步骤如下：

其中 2. 准备数据， 3. 建立模型 就是上文着重说的部分。

- 引入包
- 准备数据
- 建立模型
- 训练
- 预测

1. 引入包

```
import sys

import math

import tflearn

import tensorflow as tf

from tensorflow.python.ops import rnn_cell

from tensorflow.python.ops import rnn

import chardet

import numpy as np

import struct
```

2. 准备数据

```
def load_word_set()
将 3000 万语料，分成 Question 和 Answer 部分，提取出 word。
```

```
def load_word_set():  
    file_object = open('./segment_result_lined.3000000.pair.less', 'r')  
    while True:  
        line = file_object.readline()  
        if line:  
            line_pair = line.split('|')  
            line_question = line_pair[0]  
            line_answer = line_pair[1]  
            for word in line_question.decode('utf-8').split(' '):  
                word_set[word] = 1  
            for word in line_answer.decode('utf-8').split(' '):  
                word_set[word] = 1  
        else:  
            break  
    file_object.close()  
  
def load_vectors(input)  
从 vectors.bin 加载词向量，返回一个 word_vector_dict 的词典，key 是词，value 是200维的向量。  
  
def init_seq(input_file)  
将 Question 和 Answer 中单词对应的词向量放在词向量序列中 question_seqs， answer_seqs。  
  
def init_seq(input_file):  
    """读取切好词的文本文件，加载全部词序列  
    """  
  
    file_object = open(input_file, 'r')
```

```
vocab_dict = {}

while True:

    question_seq = []
    answer_seq = []

    line = file_object.readline()

    if line:

        line_pair = line.split('|')

        line_question = line_pair[0]

        line_answer = line_pair[1]

        for word in line_question.decode('utf-8').split(' '):

            if word_vector_dict.has_key(word):

                question_seq.append(word_vector_dict[word])

        for word in line_answer.decode('utf-8').split(' '):

            if word_vector_dict.has_key(word):

                answer_seq.append(word_vector_dict[word])

    else:

        break

    question_seqs.append(question_seq)

    answer_seqs.append(answer_seq)

file_object.close()

def vector_sqrtlen(vector)
用来求向量的长度。

def vector_sqrtlen(vector):
```

```
len = 0
```

```
for item in vector:
```

```
    len += item * item
```

```
len = math.sqrt(len)
```

```
return len
```

```
def vector_cosine(v1, v2)
```

用来求两个向量间的距离。

```
def vector_cosine(v1, v2):
```

```
    if len(v1) != len(v2):
```

```
        sys.exit(1)
```

```
    sqrtlen1 = vector_sqrtlen(v1)
```

```
    sqrtlen2 = vector_sqrtlen(v2)
```

```
    value = 0
```

```
    for item1, item2 in zip(v1, v2):
```

```
        value += item1 * item2
```

```
    return value / (sqrtlen1*sqrtlen2)
```

```
def vector2word(vector)
```

给定一个词向量，去 word - vector 字典中查找与此向量距离最近的向量，并记忆相应的单词，返回单词和 cosine 值。

```
def vector2word(vector):
```

```
    max_cos = -10000
```

```
    match_word = "
```

```
    for word in word_vector_dict:
```

```
v = word_vector_dict[word]

cosine = vector_cosine(vector, v)

if cosine > max_cos:

    max_cos = cosine

    match_word = word

return (match_word, max_cos)
```

3. 建立模型

```
class MySeq2Seq(object)
```

```
def generate_trainig_data(self)
```

由 question_seqs , answer_seqs 得到 xy_data 和 y_data 的形式。

```
def model(self, feed_previous=False)
```

用 input data 生成 encoder_inputs 和带GO头的 decoder_inputs。

将 encoder_inputs 传递给编码器，返回一个输出(预测序列的第一个值)和一个状态(传给解码器)。

在解码器中，用编码器的最后一个输出作为第一个输入，预测过程用前一个时间序的输出作为下一个时间序的输入。

4. 训练

```
def train(self)
```

用 generate_trainig_data() 生成 X y 数据，传递给 上面定义的 model，并训练 model.fit，再保存。

5. 预测

用 generate_trainig_data() 生成数据，用 model.predict 进行预测，predict 结果的每一个 sample 相当于一句话的词向量序列，每个 sample 中的每个 vector 在 word - vector 字典中找到与其最近的向量，并返回对应的 word，及二者间的 cosine。

```
if __name__ == '__main__':
```

```
phrase = sys.argv[1]

if 3 == len(sys.argv):

    my_seq2seq = MySeq2Seq(word_vec_dim=word_vec_dim, max_seq_len=max_seq_len,
input_file=sys.argv[2])

else:

    my_seq2seq = MySeq2Seq(word_vec_dim=word_vec_dim, max_seq_len=max_seq_len)

if phrase == 'train':

    my_seq2seq.train()

else:

    model = my_seq2seq.load()

    trainXY, trainY = my_seq2seq.generate_trainig_data()

    predict = model.predict(trainXY)

    for sample in predict:

        print "predict answer"

        for w in sample[1:]:

            (match_word, max_cos) = vector2word(w)

            #if vector_sqrtlen(w) < 1:

            #    break

            print match_word, max_cos, vector_sqrtlen(w)
```

本人微信公众帐号：心禅道（xinchandao）

心禅道 追求心灵安宁之道的思考 将持续探索更新给你总结的文章



长按，识别二维码，关注我们

天天搞笑开心吧，每天给你带来开心搞笑内容



长按，识别二维码，关注我们



长按二维码加关注

双色球预测合买(ssqyuace)

微信号: ssqyuace

本人微信公众帐号：双色球预测合买（ssqyuace）

流风，飘然的风，本人博客网站：<http://www.ssqhm.com>

QQ群: python技术交流群：453879716，人工智能深度学习群：251088643, 有兴趣的可以加入

微信公众帐号：双色球预测合买（ssqyuace），心禅道（xinchandao），天天搞笑开心吧（ttgxkxb）

分类: [人工智能](#)

标签: [DeepLearning](#), [tensorflow](#), [chatbot](#), [聊天机器人](#), [人工智能](#), [深度学习](#)



[流风，飘然的风](#)

[关注 - 3](#)

[粉丝 - 250](#)

[+加关注](#)

0

0

« 上一篇：[Windows10 64位下安装TensorFlow谷歌人工智能系统已官方原生支持](#)

» 下一篇：[牛逼的文案，个个都是金句，爆款的文章，一定会有几个金句。好看的文案，一定有一个金句。](#)

posted @ 2017-06-09 19:34 流风，飘然的风 阅读(2267) 评论(0) 编辑 收藏

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库

【促销】腾讯云技术升级10大核心产品年终让利

【推荐】高性能云服务器2折起，0.73元/日节省80%运维成本

【新闻】H3 BPM体验平台全面上线



最新IT新闻:

- 郭台铭详解鸿海工业互联网战略 拟分拆在上海上市
 - 扎克伯格休假照片曝光 配娃娃吃喝玩乐
 - 金刚狼死侍回归漫威，迪士尼收购福克斯让好莱坞「变天」
 - 乐视网发布公告：聘任刘淑青为公司总经理
 - 面试软件工程师，这些准备工作你做了吗？
- » 更多新闻...



最新知识库文章:

- 以操作系统的角度述说线程与进程
 - 软件测试转型之路
 - 门内门外看招聘
 - 大道至简，职场上做人做事做管理
 - 关于编程，你的练习是不是有效的？
- » 更多知识库文章...

Copyright ©2017 流风，飘然的风