

CSDN新首页上线啦，邀请你来立即体验！(http://blog.csdn.net/)

立即体

验



博客 (//blog.csdn.net?ref=toolbar) 学院 (//edu.csdn.net?ref=toolbar)

下载 (//download.csdn.net?ref=toolbar) GitChat (//gitbook.cn/?ref=csdn)

更多 ▾



登录 (https://passport.csdn.net/account/login?ref=toolbar) 注册 (http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister)

# 深度学习的Attention模型

原创 2017年10月12日 09:16:13

标签：深度学习 (http://so.csdn.net/so/search/s.do?q=深度学习&t=blog) /

attention (http://so.csdn.net/so/search/s.do?q=attention&t=blog) /

注意力 (http://so.csdn.net/so/search/s.do?q=注意力&t=blog) /

seq2seq (http://so.csdn.net/so/search/s.do?q=seq2seq&t=blog) /

机器学习 (http://so.csdn.net/so/search/s.do?q=机器学习&t=blog)

2393

## 前言

前面看到谷歌发表的运用在机器翻译上的论文《Attention is all you need》，很是让人惊讶，这是一种全新的模型，与之前的经典的seq2seq模型改动较大，它完全摒弃了RNN或CNN神经网络，大大简化了模型的复杂度，而且效果还相当好。当然Attention模型可以单独使用，但这篇文章我们来看看Attention的机制及怎么与经典的seq2seq结合。

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！



汪洋之舟---seaboat (http://...)

+ 关注

(http://blog.csdn.net/wangyangzhizhou)

码云

未开通

(https://gitee.com/wangyangzhizhou)

原创	粉丝	喜欢
305	1178	0

### 他的最新文章

更多文章

(http://blog.csdn.net/wangyangzhizhou)

从JDK源码看字节流和字符流 (http://blog.csdn.net/wangyangzhizhou/article/details/78839294)

从JDK源码看Reader (http://blog.csdn.net/wangyangzhizhou/article/details/78825744)

中文文本相似度的研究有什么用 (http://blog.csdn.net/wangyangzhizhou/article/details/78814777)

从JDK源码看OutputStream (http://blog.csdn.net/wangyangzhizhou/article/details/78810679)



内容举报



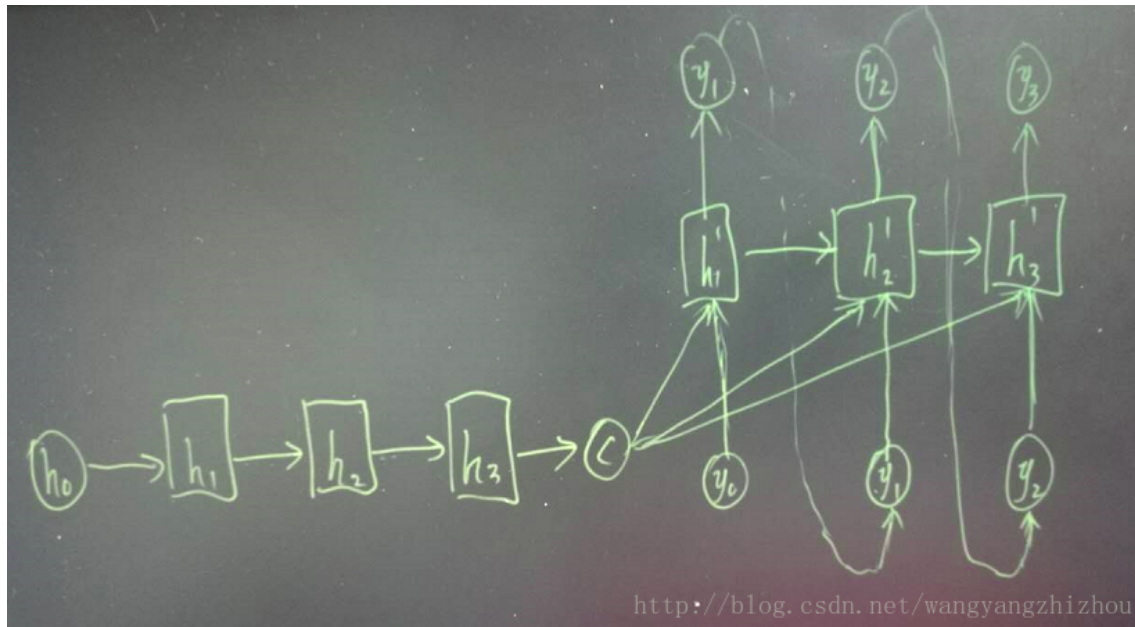
返回顶部



## seq2seq

前面我们有详细的文章介绍了seq2seq的机制以及如何用TensorFlow来实现seq2seq。可到《深度学习的seq2seq模型》(<http://blog.csdn.net/wangyangzhizhou/article/details/77883152>)、《TensorFlow实现seq2seq》(<http://blog.csdn.net/wangyangzhizhou/article/details/77977655>)阅读。

seq2seq模型结构基本都大同小异，下面用一种常见结构进行说明，我们可以看到decoder将不同时刻的输入最后用一个状态C来表示，encoder部分的不同时刻的输出则为 $y_t = g(y_{t-1}, h'_t, C)$ ，且有 $h'_t = f(h'_{t-1}, y_{t-1}, C)$ ，可以看到，不同时刻的输入被编码成C，而不同时刻的输出与上一时刻的输出、当前时刻的隐含状态、编码状态C都相关，而当前时刻隐含状态由于上一时刻的隐含状态、上个时刻输出、编码C相关。



加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

seq2seq缺点

tails/78793202)

(<https://passport.csdn.net/account/mol>

JDK不同操作系统的FileSystem (Windows&Unix-like) (<http://blog.csdn.net/wangyangzhizhou/article/details/78781872>)

### 相关推荐

深度学习中的Attention模型介绍及其进展 (<http://blog.csdn.net/tansuo17/article/details/78728522>)

深度学习笔记——Attention Model (注意力模型) 学习总结 ([http://blog.csdn.net/mpk\\_no1/article/details/72862348](http://blog.csdn.net/mpk_no1/article/details/72862348))

深度学习笔记(六): Encoder-Decoder模型和Attention模型 (<http://blog.csdn.net/u014595019/article/details/52826423>)

用深度学习解决大规模文本分类问题93理论和实践 (转)



博主专栏

登录

注册

×

seq2seq模型是处理序列问题的大杀器，由它实现的encoder-decoder模型让其在很多场景都发挥着强大的作用，得到了很不错的效果。但它也有自己的局限性，被认为最大限制了seq2seq能力的就在于编码和解码之间的唯一联系就是一个固定长度的语义向量C。

我们可以看到上面，encoder的不同时刻的输入都被编码成状态C，这个便是语义向量。从这个角度看，它将整个序列的信息压缩到一个固定长度的向量中去了。对于短序列来说，这种做法可能问题不大，但如果是较长较复杂的队列，则该语义向量可能无法完全表示整个序列的信息。而且先输入的序列的信息会被后输入的序列信息稀释掉。输入序列越长，这个现象就越严重。如果我们得到一个有损的语义编码，那对后面的解码也很可能会产生影响。

另外seq2seq模型一般会使用RNN，而RNN因为与上一时刻状态相关，所以不能并行运算，效率低。但这是所有使用RNN的局限性，除非不使用它。

## Attention模型

为了解决语义损失和信息稀释的问题，提出了Attention模型，Attention即注意力，它是模拟了人类的视觉注意机制而来，比如当观察某个画面时，注意力聚焦到其中某一部分，其余部分则变得模糊。

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！



java开源研究 (https://passport.csdn.net/account/mol)  
(http://blog.csdn.net/column/details/s  
Java并发  
(http://blog.csdn.net/column/details/s  
concurrent.html)  
(http://blog.csdn.net/column/details/jc  
concurrent.html)  
tomcat内核

展开



### 他的热门文章

75篇关于Tomcat源码和机制的文章 (http://blog.csdn.net/wangyangzhizhou/article/details/53944121)

17714

分布式系统调用链监控 (http://blog.csdn.net/wangyangzhizhou/article/details/53844668)

14676

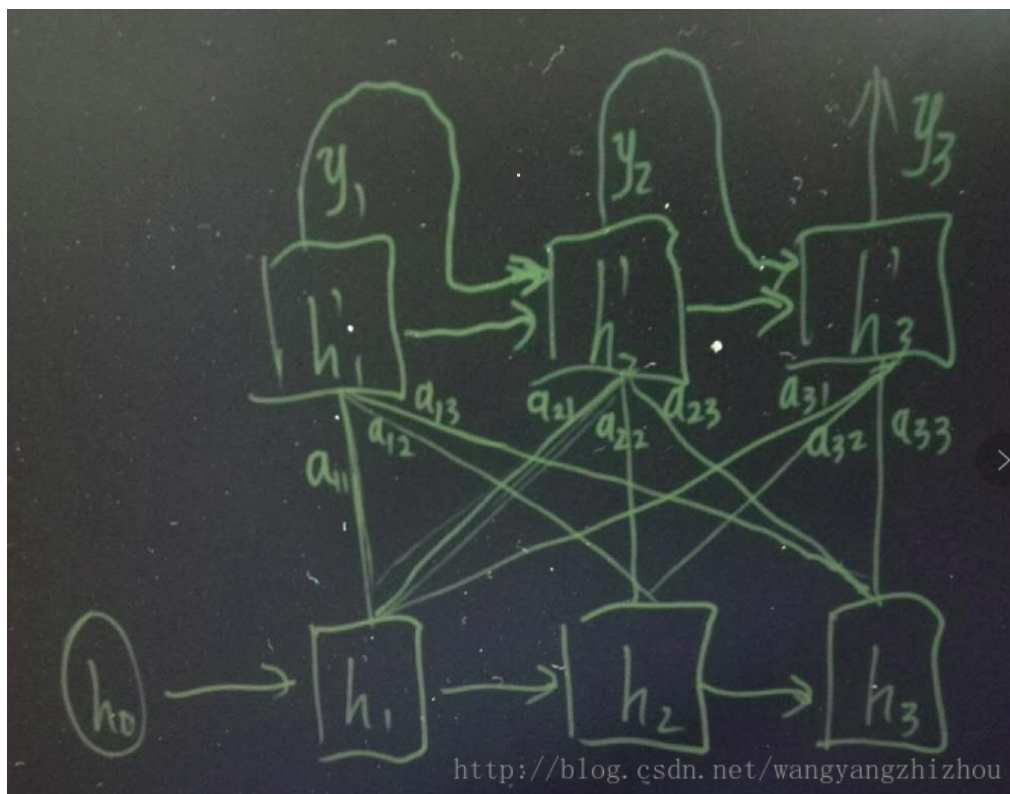
登录

注册



内容举报

返回顶部



web安全认证机制知多少 (<http://passport.csdn.net/account/moln.net/wangyangzhizhou/article/details/51336038>)

13601

注解机制及其原理 (<http://blog.csdn.net/wangyangzhizhou/article/details/51698638>)

13570

如何设计一个web容器 (<http://blog.csdn.net/wangyangzhizhou/article/details/50662125>)

12278

按照图，咱们往下详细看Attention模型怎么与seq2seq结合。对于decoder，输出为

$$y_t = g(y_{t-1}, h'_t, C)$$

这里的C已经与前面说到的seq2seq的语义向量C不同了，已经不再是将所有输入进行编码，下面会说到C怎么计算。再看隐含状态，

$$h'_t = f(h'_{t-1}, y_{t-1}, C)$$

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

登录

注册

内容举报

返回顶部

×

C向量计算公式为，

(<https://passport.csdn.net/account/mol>

$$C_t = \sum_{j=1}^{T_x} a_{tj} h_j$$

我们可以将其中的a看成是各个时刻的输出的不同时刻的输入对应的权重，其实就可以使用softmax来计算，它的计算公式为，

$$a_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^{T_x} \exp(e_{tk})}$$

其中，

$$e_{tj} = w(h'_{t-1}, h_j)$$



0



内容举报



返回顶部

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

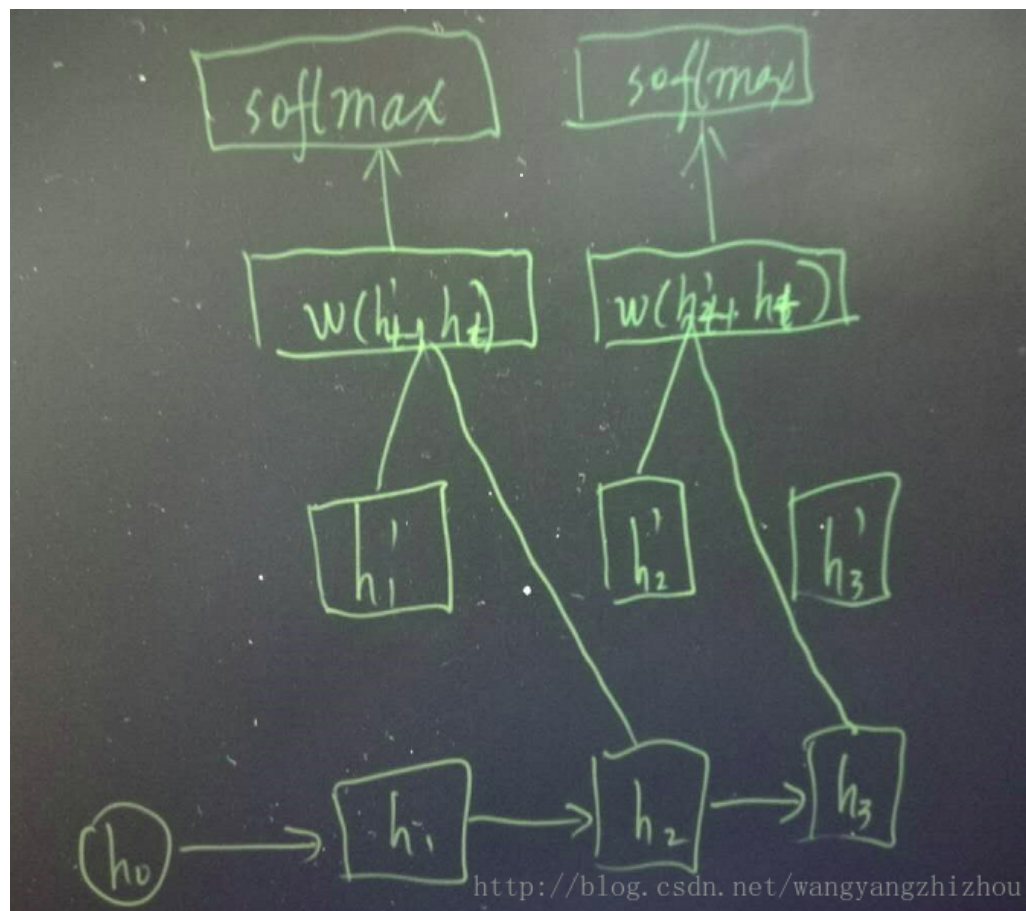
登录

注册





(https://passport.csdn.net/account/mol



通过上面几个公式就基本描述了seq2seq与Attention的结合，其中每个输出多了注意力向量参与，注意力的权重由上一时刻隐含状态与encoder的各个输入的隐含状态共同决定。

在seq2seq模型上加上Attention模型，克服了语义损失和信息稀释的问题，但是它也引入了额外的成本，对于m个输入，n个输出的结构中，Attention参数也达到了m\*n的数量级。

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！  
以下是**广告**和**相关阅读**

[登录](#)
[注册](#)


内容举报

返回顶部

=====广告时间=====

(<https://passport.csdn.net/account/mol>)

公众号的菜单已分为“分布式”、“机器学习”、“深度学习”、“NLP”、“Java深度”、“Java并发核心”、“JDK源码”、“Tomcat内核”等，可能有一款适合你的胃口。

鄙人的新书《Tomcat内核设计剖析》已经在京东销售了，有需要的朋友可以购买。感谢各位朋友。

为什么写《Tomcat内核设计剖析》(<http://blog.csdn.net/wangyangzhizhou/article/details/74080321>)

相关阅读：

《深度学习的seq2seq模型》(<http://blog.csdn.net/wangyangzhizhou/article/details/77883152>)

《TensorFlow实现seq2seq》(<http://blog.csdn.net/wangyangzhizhou/article/details/77977655>)

欢迎关注：



<http://blog.csdn.net/wangyangzhizhou>



加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

登录

注册



内容举报

返回顶部

版权声明：本文为博主原创文章，未经博主允许不得转载。

本文已收录于以下专栏：机器学习&深度学习 (<http://blog.csdn.net/column/details/14969.html>)



0



myy629464 (/myy629464) 2017-10-14 21:51

1楼

(/myy629464), 学习一下！

回复

## 相关文章推荐

### 深度学习中的Attention模型介绍及其进展 (<http://blog.csdn.net/tansuo17/article/details/787...>

近期对深度学习中的Attention模型进行了深入研究，该模型在图像识别、语音识别和自然语言处理三大深度学习的热门领域均有广泛的使用，是2014和2015年深度学习领域的重要进展。现对其原理、主要应用...



tansuo17 (<http://blog.csdn.net/tansuo17>) 2017年12月06日 11:09 45

### 深度学习笔记——Attention Model（注意力模型）学习总结 ([http://blog.csdn.net/mpk\\_no1/a...](http://blog.csdn.net/mpk_no1/a...)

Attention Model（注意力模型）学习总结，包括soft Attention Model，Global Attention Model和Local Attention Model，静态AM，...



mpk\_no1 ([http://blog.csdn.net/mpk\\_no1](http://blog.csdn.net/mpk_no1)) 2017年08月06日 21:49 7067

### AI程序员年薪百万起，男友非要学，我该支持他吗？

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

吴恩达说未来是人工智能的天下，另外无论是薪酬还是就业前景，我该支持男友学吗？他这么跟我说的...

登录

注册



内容举报



返回顶部



(https://passport.csdn.net/account/mol



(http://www.baidu.com/cb.php?c=lgF\_pyfqHmknjnvPjn0IZ0qnfK9ujYzP1f4PjDs0Aw-

5Hc3rHnYnHb0TAq15HfLPWRznjb0T1YYnHmkrHckPW0vuHnmvY0AwY5HDdnHn3njm4PHn0lgF\_5y9YIZ0IQzq-

uZR8mLPbUB48ugfEIAqspynEmybz5LNYUNq1ULNzmvRqmhkEu1Ds0ZFb5HD0mhYqn0KsTWYs0ZNGujYkPHTYn1mk0AqGujYknWb3rjDY0APGujYLnWm4n1c0ULI85H00TZbqnV



0



## 深度学习笔记(六)：Encoder-Decoder模型和Attention模型 (http://blog.csdn.net/u01459501...

这两天在看attention模型，看了下知乎上的几个回答，很多人都推荐了一篇文章Neural Machine Translation by Jointly Learning to Align and ...



u014595019 (http://blog.csdn.net/u014595019) 2016年10月15日 23:09 19200



## 用深度学习（CNN RNN Attention）解决大规模文本分类问题 - 综述和实践（转载）(http://blo...

近来在同时做一个应用深度学习解决淘宝商品的类目预测问题的项目，恰好硕士毕业论文题目便是文本分类问题，趁此机会总结下文本分类领域特别是应用深度学习解决文本分类的相关的思路、做法和部分实践的经验。业...



u013818406 (http://blog.csdn.net/u013818406) 2017年04月05日 21:42 2521

## 用深度学习（CNN RNN Attention）解决大规模文本分类问题 - 综述和实践（转载）(http://blo...

本文转载自:http://blog.csdn.net/u013818406/article/details/69359816 近来在同时做一个应用深度学习解决淘宝商品的类目预...



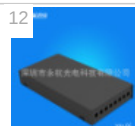
chivalrousli (http://blog.csdn.net/chivalrousli) 2017年08月07日 11:29 4206



**30.00/件**  
12芯 24芯 ST SC FC  
机架式光纤盒，光缆终



**18.00/包**  
【批发】塑料吸盘定位  
片25\*25 不干胶自粘式



**4.80/个**  
厂家直销 YH-F8挂墙  
式光纤终端盒8口

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

## 深度学习笔记(1)：caffe 添加新层 attention LSTM layer和LSTM layer代码精读 (http://blo...

登录

注册



内容举报

返回顶部

深度学习笔记（1）：caffe 添加新层 attention LSTM layer和LSTM layer代码精读

(https://passport.csdn.net/account/mol



u013110060 (http://blog.csdn.net/u013110060) 2017年03月08日 14:33 1165

## 深度学习笔记（2）：caffe 加新层 Attention LSTM layer (http://blog.csdn.net/u013110060/...

深度学习笔记（2）：caffe 加新层 Attention LSTM layer



u013110060 (http://blog.csdn.net/u013110060) 2017年03月09日 16:15 2864

## 用深度学习（CNN RNN Attention）解决大规模文本分类问题 - 综述和实践 (http://blog.csdn....

转处：https://zhuanlan.zhihu.com/p/25928551 用深度学习（CNN RNN Attention）解决大规模文本分类问题 - 综述和实践 清  
淞 ...



xiewenbo (http://blog.csdn.net/xiewenbo) 2017年05月31日 20:29 1540

## 用深度学习（CNN RNN Attention）解决大规模文本分类问题 (http://blog.csdn.net/ch120949...

一、传统文本分类方法 文本分类问题算是自然语言处理领域中一个非常经典的问题了，相关研究最早可以追溯到上世纪50年  
代，当时是通过专家规则（Pattern）进行分类，甚至在80年代初一度发展到利用知识工...



ch1209498273 (http://blog.csdn.net/ch1209498273) 2017年11月05日 21:09 141

## [深度学习论文笔记][Attention] Spatial Transformer Networks (http://blog.csdn.net/Hao\_Zh...

Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in Ne...



Hao\_Zhang\_Vision (http://blog.csdn.net/Hao\_Zhang\_Vision) 2016年11月15日 22:02 1072

## 【深度学习】聚焦机制DRAM(Deep Recurrent Attention Model)算法详解 (http://blog.csdn.n...

Visual Attention基础，Multiple object recognition with visual attention算法解读。

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！


登录

注册



内容举报


返回顶部

 shenxiaolu1984 (<http://blog.csdn.net/shenxiaolu1984>) 2016年06月28日 22:14 6434

(<https://passport.csdn.net/account/mol>)


## 深度学习Attention机制在短文本分类上的应用——qjzcy的博客 (<http://blog.csdn.net/qjzcy/ar...>)

平常我们对分类的判断也是基于标题中的某些字，或者某些词性。比如《姚明篮球打的怎样》应该判别为体育，这时候“姚明”，“篮球”应该算对我们比较重要的词汇。词性我们关注点在“人名”和“名词”上面，深度学习的...

 qjzcy (<http://blog.csdn.net/qjzcy>) 2017年04月17日 20:48 2032


## Attention and Memory in Deep Learning and NLP(深度学习和NLP中的注意和记忆机制) 阅读..

什么是Attention Mechanisms(注意机制) 神经网络中的注意机制类似人类的视觉注意机制，人类视觉本质上看图片会聚焦在“高清晰度”的区域，同时也会感知周围“低清晰度”的区域，然...

 aPYXa (<http://blog.csdn.net/aPYXa>) 2016年07月26日 20:19 1664


## 深度学习笔记——基于双向RNN (LSTM、GRU) 和Attention Model的句子对匹配方法 (<http://...>)

本文主要是结合RNN和Attention Model做一些关于句子对匹配的模型总结。

 mpk\_no1 ([http://blog.csdn.net/mpk\\_no1](http://blog.csdn.net/mpk_no1)) 2017年08月06日 23:06 1275

## 浅谈深度学习尝常识：泛化、过拟合、Dropout、Attention ([http://blog.csdn.net/Errors\\_In\\_L...](http://blog.csdn.net/Errors_In_L...))

刚开始接触深度学习，难免有很多名词概念的陌生、不清晰；当然，如果要快速考核一个人的深度学习“深度”，尤其是在面试过程中，同行会问到，有没有出现过拟合呀？怎么解决呀？这些问题看起来很高深，或者需要很深厚...

 Errors\_In\_Life ([http://blog.csdn.net/Errors\\_In\\_Life](http://blog.csdn.net/Errors_In_Life)) 2017年04月19日 23:46 1086

## 用深度学习 (CNN RNN Attention) 解决大规模文本分类问题 - 综述和实践 (<http://blog.csdn....>)

转载自<https://zhuanlan.zhihu.com/p/25928551> 近来在同时做一个应用深度学习解决淘宝商品的类目预测问题的项目，恰好硕士毕业论文题目便是大规模文本分类问题，趁此机会特将一些心得分享给大家，与5000万程序员共同成长！


登录

注册

×


  
内容举报

  
返回顶部

 hlang8160 (<http://blog.csdn.net/hlang8160>) 2017年11月20日 22:54 175


## 浅谈深度学习常识：泛化、过拟合、Dropout、Attention ([http://blog.csdn.net/qq\\_284134...](http://blog.csdn.net/qq_28413479)

转载链接：[http://blog.csdn.net/errors\\_in\\_life/article/details/70249256](http://blog.csdn.net/errors_in_life/article/details/70249256) 刚开始接触深度学习，难免有很多名词概念的陌生、不清晰；当然，如果要快...

 qq\_28413479 ([http://blog.csdn.net/qq\\_28413479](http://blog.csdn.net/qq_28413479)) 2017年08月09日 16:27 243


## Attention and Memory in Deep Learning and NLP(深度学习和NLP中的注意和记忆机制) (htt...

A recent trend in Deep Learning are Attention Mechanisms. In an interview, Ilya Sutskever, now the ...

 aPYXa (<http://blog.csdn.net/aPYXa>) 2016年07月21日 19:03 882

## 深度学习tracking学习笔记(2)：图像/视觉显著性检测技术发展情况梳理(Saliency Detection、...

reference: <http://blog.csdn.net/anshan1984/article/details/8657176> [http://blog.csdn.net/huangbo10/...](http://blog.csdn.net/huangbo10/)

 u011534057 (<http://blog.csdn.net/u011534057>) 2016年06月23日 10:44 3731



## compare-aggregate模型 基于文档的问答系统（基于深度attention\_LSTM）...

([http://download.csdn.net/detail/qq\\_28413479/1011534057](http://download.csdn.net/detail/qq_28413479/1011534057)) 2017年12月11日 09:50 111KB [下载](#)

  
内容举报

  
返回顶部

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

登录

注册

×