

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

立即体

验

CSDN

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 

0

浅谈Attention-based Model【原理篇】

原创

2017年07月22日 19:24:02

1662



转载请标明出处：http://blog.csdn.net/wuzqchom/article/details/75792501

(http://blog.csdn.net/wuzqchom/article/details/75792501)

计划分为三个部分：

浅谈Attention-based Model【原理篇】（你在这里）

浅谈Attention-based Model【源码篇】(http://blog.csdn.net/wuzqChom/article/details/77918780)

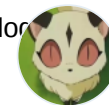
浅谈Attention-based Model【实践篇】

0. 前言



weixin_3506...

(//my.csdn.net?ref=toolbar)
(//write.blog.csdn.net/postedit?ref=toolbar)source=csdnblor



wuzqchom (http://blog.cs...

+ 关注

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

码云

原创

粉丝

喜欢

未开通

56

6

5

(https://gite
utm_sourc

他的最新文章

更多文章 (http://blog.csdn.net/wuzqchom)

使用BibTex格式时缩小参考文献的字体
(http://blog.csdn.net/wuzqChom/article/
details/78592038)

【转】计算机类学术论文 28个常见出
版社一般写法（参考文献用）(http://bl
og.csdn.net/wuzqChom/article/details/
78196847)

Batch Normalization (http://blog.csdn.n
et/wuzqChom/article/details/78024653)

看了台大的李宏毅老师关于Attention部分的内容，这一部分讲的不错（其实李宏毅老师其它部分的内容也不错，比较幽默，安利一下），记录一下，本博客的大部分内容据来自李宏毅老师的授课资料：Attention-based Model (http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLSD15_2.html)。如发现有误，望不吝赐教。

1. 为什么需要Attention

最基本的seq2seq模型包含一个encoder和一个decoder，通常的做法是将一个输入的句子编码成一个固定大小的state，然后作为decoder的初始状态（当然也可以作为每一时刻的输入），但这样的状态对于decoder中的所有时刻都是一样的。

attention即为注意力，人脑在对于的不同部分的注意力是不同的。需要attention的原因是非常直观的，比如，我们期末考试的时候，我们需要老师划重点，划重点的目的就是为了尽量将我们的attention放在这部分的内容上，以期用最少的付出获取尽可能高的分数；再比如我们到一个新的班级，吸引我们attention的是不是颜值比较高的人？普通的模型可以看成所有部分的attention都是一样的，而这里的attention-based model对于不同的部分，重要的程度则不同。

2. Attention-based Model是什么

Attention-based Model其实就是一个相似性的度量，当前的输入与目标状态越相似，那么在当前的输入的权重就会越大，说明当前的输出越依赖于当前的输入。严格来说，Attention并算不上是一种新的model，而仅仅是在以往的模型中加入attention的思想，所以Attention-based Model或者Attention Mechanism是比较合理的叫法，而非Attention Model。

浅谈Attention-based Model【源码篇】
(<http://blog.csdn.net/wuzqChom/article/details/77918780>)

Python切片赋值操作 (<http://blog.csdn.net/wuzqChom/article/details/77917825>)

相关推荐

自然语言处理中的Attention Model：是什么及为什么 (<http://blog.csdn.net/malefactor/article/details/50550211>)

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)

以Attention Model为例谈谈两种研究创新模式 (<http://blog.csdn.net/malefactor/article/details/50583474>)

深度学习方法（九）：自然语言处理中的Attention Model注意力模型 (<http://blog.csdn.net/xbinworld/article/details/54607525>)

没有attention机制的encoder-decoder结构通常把encoder的最后一个状态作为decoder的输入（可能作为初始化，也可能作为每一时刻的输入），但是encoder的state毕竟是有限的，存储不了太多的信息，对于decoder过程，每一个步骤都和之前的输入都没有关系了，只与这个传入的state有关。attention机制的引入之后，decoder根据时刻的不同，让每一时刻的输入都有所不同。

再引用tensorflow源码attention_decoder()函数关于attention的注释：

0

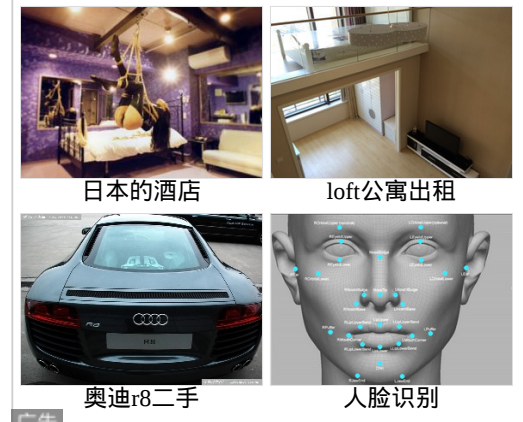
“In this context ‘attention’ means that, during decoding, the RNN can look up information in the additional tensor attention_states, and it does this by focusing on a few entries from the tensor.”



3. Attention



对于机器翻译来说，比如我们翻译“机器学习”，在翻译“machine”的时候，我们希望模型更加关注的是“机器”而不是“学习”。那么，就从这个例子开始说吧（以下图片均来自上述课程链接的slides）



日本的酒店

loft公寓出租

奥迪r8二手

人脸识别

广告



老年智能机

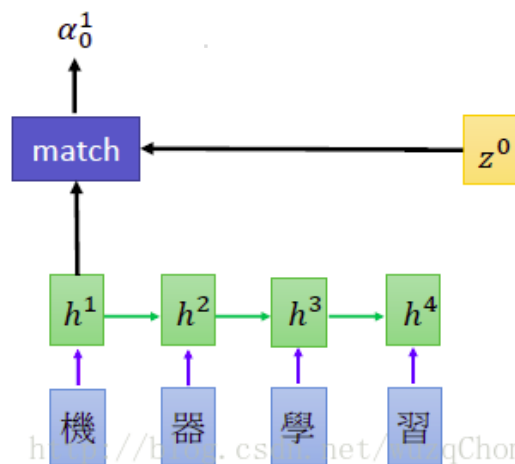


他的热门文章

TensorFlow中CNN的两种padding方式“SAME”和“VALID” (<http://blog.csdn.net/wuzqchom/article/details/74785643>)

4769

tensorflow.nn.bidirectional_dynamic_rnn()函数的用法 (<http://blog.csdn.net/wuzqchom/article/details/75453327>)



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

3036

python中scrapy点击按钮 (<http://blog.csdn.net/wuzqchom/article/details/52677579>)

2305

浅谈Attention-based Model【原理篇】 (<http://blog.csdn.net/wuzqchom/article/details/75792501>)

1649

Seq2Seq模型 (<http://blog.csdn.net/wuzqchom/article/details/76651479>)

1515



0



刚才说了，attention其实就是一个当前的输入与输出的匹配度。在上图中，即为 h^1 和 z^0 的匹配度（ h^1 为当前时刻RNN的隐层输出向量，而不是原始输入的词向量， z^0 初始化向量，如rnn中的initial memory），其中的match为计算这两个向量的匹配度的模块，出来的 α_0^1 即为由match算出来的相似度。好了，基本上这个就是attention-based model 的attention部分了。那么，match什么呢？

对于“match”，理论上任何可以计算两个向量的相似度都可以，比如：

- 余弦相似度
- 一个简单的神经网络，输入为 h 和 w ，输出为 α

- 或者矩阵变换 $\alpha = h^T W z$ (Multiplicative attention , Luong et al., 2015)

现在我们已经由match模块算出了当前输入输出的匹配度，然后我们需要计算当前的输出（实际为decoder端的隐状态）和每一个输入做一次match计算，分别可以得到当前的输出和所有输入的匹配度，由于计算出来并没有归一化，所以我们使用softmax，使其输出时所有权重之和为1。那么和每一个输入的权重都有了（由于下一个输出为“machine”，我们希望第一个权重和第二个权重越大越好），那么我们可以计算出其加权向量和，作为下一次的输入。



0

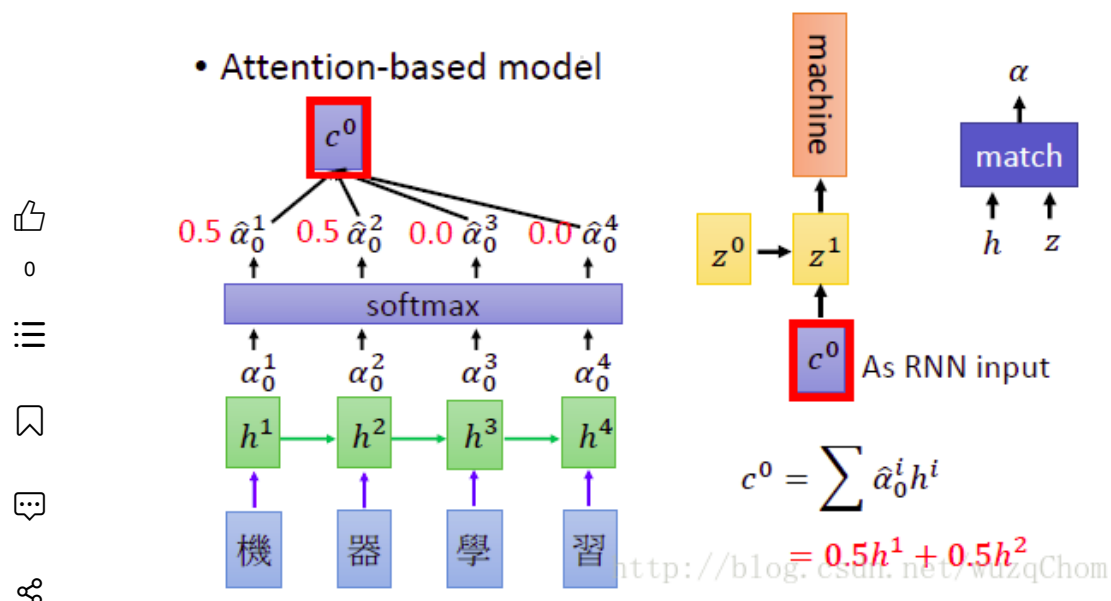


这里有一个问题：就是如果match用后面的两种，那么参数应该怎么学呢？



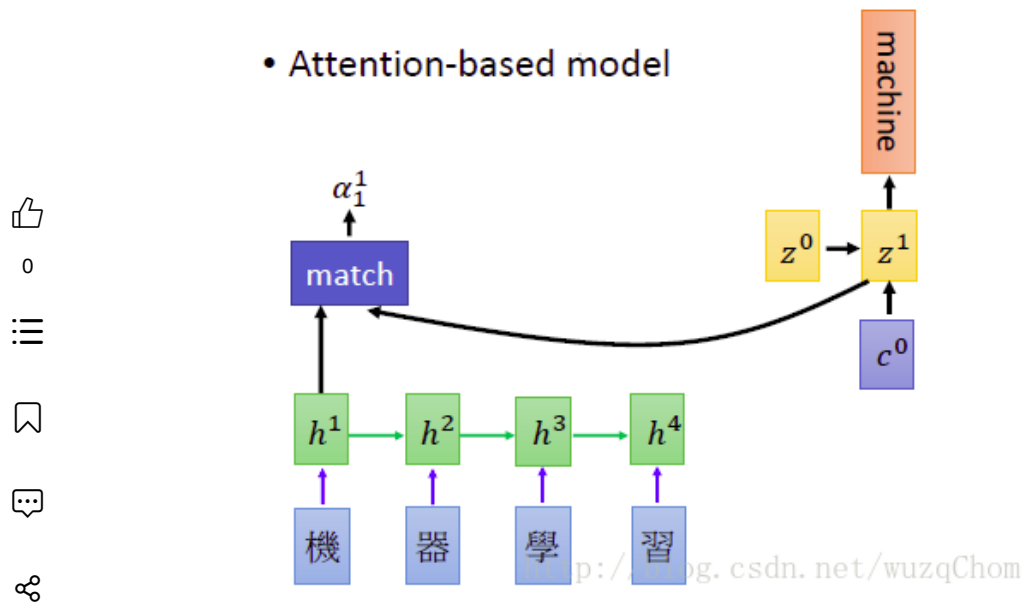
如下图所示：





那么再算出了 c^0 之后，我们就把这个向量作为rnn的输入（如果我们decoder用的是RNN的话），然后d第一个时间点的输出的编码 z^1 由 c^0 和初始状态 z^0 共同决定。我们计算得到 z^1 之后，替换之前的 z^0 再和每一个输入的encoder的vector计算匹配度，然后softmax，计算向量加权，作为第二时刻的输入.....如此循环直至结束。

• Attention-based model



再看看Grammar as a Foreign Language (<https://arxiv.org/abs/1412.7449>)一文当中的公式：

$$\begin{aligned}
 u_i^t &= v^T \tanh(W_1' h_i + W_2' d_t) \\
 a_i^t &= \text{softmax}(u_i^t) \\
 d_t' &= \sum_{i=1}^{T_A} a_i^t h_i
 \end{aligned}$$

上面的符号表示和前面描述的不太一样，经统一符号的公式如下：

$$u_t^i = v^T \tanh(W_1 h^i + W_2 Z^t)$$

$$\alpha_t^i = \text{softmax}(u_t^i)$$



$$c^t = \sum_i \alpha_t^i h^i$$



得到 c^t 之后，就可以作为第t时刻RNN的input，而 Z^t 可以作为t时刻RNN的隐状态的输入，这样就能够得到新的隐状态 Z^{t+1} ，如此循环，直到遇到停止符为止。



上式 v^T 为一个向量， u_t^i 即为t时刻的decoder隐状态向量和第i个输入的相似度（未归一化）。



在第一个式子当中的 \tanh 里面的内容，tensorflow对于第一部分的实现是采用卷积的方式实现，而第二部分使用的是线性映射函数linear实现。

v^T 向量的长度即为attention_vec_size，tensorflow中直接把输入向量的长度设置为attention向量的长度，你要设置其他的值也可以，这里比较奇怪，tensorflow没有将其作为函数的参数，难道设成和输入的向量相同的效果更好？（有知情者也烦请告知）

4. 遗留问题

本来我们已经结束了，但是仔细想想，其实还有一个地方有所疑问。就是加入match是一个简单地神经网络或者一个矩阵，神经网络的权值和矩阵里面的值怎么来？

其实这些都是可以BP的时候就可以自动学到的。比如我们明确输出是什么，在第一时刻的时候，那就会调整 z^0 和 c^0 的值，进而就会调整所有的 α 值，之前说过 α 是match的输出，如果match是后两种的话，就会进而调整match中的参数。

5. to be continued

最近在用双向LSTM加上attention做知乎的比赛，但目前效果不是很好，具体等待比赛结束之后再详细记录。



0

版权声明：本文为博主原创文章，未经博主允许不得转载。



发表你的评论

(http://my.csdn.net/weixin_35068028)



相关文章推荐

自然语言处理中的Attention Model：是什么及为什么 (<http://blog.csdn.net/malefactor/article...>)

要是关注深度学习在自然语言处理方面的研究进展，我相信你一定听说过Attention Model（后文有时会简称AM模型）这个词。AM模型应该说是过去一年来NLP领域中的重要进展之一，在很多场景被证明有...



malefactor (<http://blog.csdn.net/malefactor>) 2016年01月20日 18:26 58071

深度学习笔记——Attention Model（注意力模型）学习总结 (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) 2017年08月06日 21:49 7778



惊呆了！微博和阿里背后的数据库有多厉害？

想不到！数据库作为最关键的基础设施，渗透技术领域的方方面面，我阿里和微博的师哥们是这么分享的...

(http://www.baidu.com/cb.php?c=lgF_pyfqHmknjTzrjb0lZ0qnfK9ujYzP1nsrjD10Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1YvrycknynsPyN-P1lbnyDL0AwY5HDdnHfznWRYPHR0IgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEpZNGXy-jULNzTvRETVnzpyN1gVw-IA7GUatLPjqdAdxTvqdThP-5yF_UvTkn0KzujYk0AFV5H00TZcqn0KdpyfqHRLPjnvnfKEpyfqHc4rj6kP0KWpyfqP1cVrHnz0AqLUWYs0ZK45HcsP6KWThnqPW0zrjm)



以Attention Model为例谈谈两种研究创新模式 (<http://blog.csdn.net/malefactor/article/detail...>)



各位观众朋友好，也许此刻您刚打开电梯.....上了年纪的读者估计能看出上面一句是引用了韩乔生老先生的名言，我写东西就喜欢用名人名言开头，这好习惯这么多年怎么也改不了。您问韩乔生是谁？恭喜您，作为90后您以...



malefactor (<http://blog.csdn.net/malefactor>) 2016年01月26日 00:18 21393

深度学习方法（九）：自然语言处理中的Attention Model注意力模型 (<http://blog.csdn.net/xb...>)

上一篇博文深度学习方法（八）：Encoder-Decoder模型，基本Sequence to Sequence模型描述了基本的Encoder-Decoder模型，在作为翻译模型的时候，这种基本的Enc...



xbinworld (<http://blog.csdn.net/xbinworld>) 2017年02月04日 00:27 6505

浅谈Attention-based Model【源码篇】 (<http://blog.csdn.net/wuzqChom/article/details/779...>)

转载请标明出处：<http://blog.csdn.net/wuzqchom/article/details/77918780>此为本人阅读tensorflow源码的记录，主要在一些步骤上加了一些注释和少...



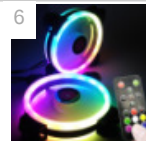
wuzqChom (<http://blog.csdn.net/wuzqChom>) 2017年09月09日 22:54 694



24.00/个
双铜管CPU散热器超
静音1150AMD1155



560.00/台
1U服务器工业工控机
箱多USB高档铝面板



17.00/个
RGB机箱风扇 12cm台
式电脑散热风扇 七

广告



自然语言处理中的Attention Model：是什么及为什么 (<http://blog.csdn.net/jdbc/article/detail...>)

要是关注深度学习在自然语言处理方面的研究进展，我相信你一定听说过Attention Model（后文有时会简称AM模型）这个词。AM模型应该说是过去一年来NLP领域中的重要进展之一，在很多场景被证明有...



jdbc (<http://blog.csdn.net/jdbc>) 2016年01月26日 11:16 2336



【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。



shenxiaolu1984 (<http://blog.csdn.net/shenxiaolu1984>) 2016年06月28日 22:14 6483

attention 机制 (http://blog.csdn.net/qq_26609915/article/details/52086772)


attention 机制什么是attentionattention机制是（非常）松散地基于人类的视觉注意机制。就是按照“高分辨率”聚焦在图片的某个特定区域并以“低分辨率”感知图像的周边区域的模式，然后...



qq_26609915 (http://blog.csdn.net/qq_26609915) 2016年08月01日 16:24 5285


深度学习中的Attention模型介绍及其进展 (<http://blog.csdn.net/jteng/article/details/52864401>)

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...


 jteng (<http://blog.csdn.net/jteng>) 2016年10月20日 15:29 7022


阅读理解任务中的Attention-over-Attention神经网络模型原理及实现 (<http://blog.csdn.net/liuchonge>...

本文是“Attention-over-Attention Neural Networks for Reading Comprehension”的阅读笔记。这篇论文所处理的任务是阅读理解里面的完形填空问...

 liuchonge (<http://blog.csdn.net/liuchonge>) 2017年06月06日 09:24 2569

注意力机制 (Attention Mechanism) 在自然语言处理中的应用 (<http://blog.csdn.net/jjdbc/art...>

 关于attention的发展综述文章

 jjdbc (<http://blog.csdn.net/jjdbc>) 2016年10月27日 18:17 7539

Attention注意力机制--原理与应用 (<http://blog.csdn.net/joshuaxx316/article/details/706653...>

注意力机制即Attention mechanism在序列学习任务上具有巨大的提升作用，在编解码器框架内，通过在编码段加入A模型，对源数据序列进行数据加权变换，或者在解码端引入A模型，对目标数据进行加权...

 joshuaxx316 (<http://blog.csdn.net/joshuaxx316>) 2017年04月24日 21:51 9574



纯干货13 2017年-李宏毅-最新深度学习/机器学习中文视频教程分享-前篇 (<http://blog.csdn.net...>

上一次分享了台湾大学李宏毅老师讲的GAN视频课程，今天继续与大家分享李老师的在2017年初陆续发放出来的关于深度学习/机器学习的视频课程，所有的课程都是用中文讲的，课程讲得详细、简单易懂，PPT做的非...

 lqfarmer (<http://blog.csdn.net/lqfarmer>) 2017年07月24日 09:45 3598



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

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  19550



debug tensorflow的seq2seq的attention_decoder方法 (<http://blog.csdn.net/guotong1988/>...

写这个attention_decoder的testcase来用debug的方式看看注意力机制的实现import tensorflow as tf from tensorflow.python.ops ...

 guotong1988 (<http://blog.csdn.net/guotong1988>) 2017年09月04日 20:08  269



【增强学习】Recurrent Visual Attention源码解读 (<http://blog.csdn.net/shenxiaolu1984/art...>

Mnih, Volodymyr, Nicolas Heess, and Alex Graves. "Recurrent models of visual attention." Advances in...

 shenxiaolu1984 (<http://blog.csdn.net/shenxiaolu1984>) 2016年06月17日 21:18  8261

attention机制 (<http://blog.csdn.net/chcyj/article/details/78645220>)

Attention注意力机制--原理与应用 注意力机制即Attention mechanism在序列学习任务上具有巨大的提升作用，在编解码器框架内，通过在编码段加入A模型，对源数据序列进行数据加权变换...

 chcyj (<http://blog.csdn.net/chcyj>) 2017年11月27日 15:07  43

Seq2Seq源码解析（基于Theano框架）(<http://blog.csdn.net/yezhenxu1992/article/details/...>

这是一篇关于seq2seq源码解析的文章。seq2seq是一个通用型 端到端（end-to-end）的学习框架，包括2个部分，一个是Encoder（编码器，负责编码源句子的输入），另外一个Decod...