

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

立即体



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/mobileregister?ref=toolbar&action=mobileRegister)

Attention注意力机制--原理与应用



2017年04月24日 21:51:10

9268



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，或者在解码端引入A模型，对目标数据进行加权变化，可以有效提高序列对序列的自然方式下的系统表现。

什么是Attention？

Attention模型的基本表述可以这样理解成（我个人理解）：当我们人在看一样东西的时候，我们当前时刻关注的一定是我们当前正在看的这样东西的某一地方，换句话说，当我们目光移到加入别处时，注意力会随着目光的移动而转移，这意味着，当人们注意到某个目标或某个场景时，



joshuaxx316 (http://blog...

+ 关注

(http://blog.csdn.net/joshuaxx316)

码云

原创

粉丝

喜欢

未开通

34

14

0

(https://gite
utm_sourc

他的最新文章

更多文章 (http://blog.csdn.net/joshuaxx316)

shell遍历文件夹内文件分割字符串-工
程小札记 (http://blog.csdn.net/joshuaxx
316/article/details/72831246)

shell脚本中变量的赋值与循环--工程小
札记 (http://blog.csdn.net/joshuaxx316/
article/details/71434325)

Python 脚本调试如何C代码--工程小札
记 (http://blog.csdn.net/joshuaxx316/ar
ticle/details/70757628)



内容举报



返回顶部



该目标内部以及该场景内每一处空间位置上的注意力分布是不一样的。这一点在如下情形下同样成立：当我们试图描述一件事情，我们当前时刻说到的单词和句子和正在描述的该事情的对应某个片段最先关，而其他部分随着描述的进行，相关性也在不断地改变。从上述两种情形，读者可以看出，对于Attention的作用角度出发，我们就可以从两个角度来分类Attention种类：空间注意力和时间注意力，即Spatial Attention 和Temporal Attention。这种分类更多的是从应用层面上，而从Attention的作用方法上，可以将其分为Soft Attention和Hard Attention，这既我们所说的，Attention输出的向量分布是一种one-hot的独热分布还是soft的软分布，这直接影响对于上下文信息的选择作用。

为什么要加入Attention？

再解释了Attention做了一件什么事之后，我们讨论一下为什么需要Attention模型,即Attention加入的动机：

1. 序列输入时，随着序列的不断增长，原始根据时间步的方式的表现越来越差，这是由于原始的这种时间步模型设计的结构有缺陷，即所有的上下文输入信息都被限制到固定长度，整个模型的能力都同样收到限制，我们暂且把这种原始的模型称为简单的编解码器模型。
2. 编解码器的结构无法解释，也就导致了其无法设计。

Attention到底是什么原理？

下面我们来看一下Attention的具体原理：

首先让编码器输出结构化的表示，假设这些表示，可以用下述集合表示，（Hold不住了，我要截图了，太麻烦了！！！！）

Beam Search Algorithm (Draft by Andrew Jungwirth) (<https://passport.csdn.net/article/details/69953841>)

【论文阅读笔记】CVPR2015-Long-term Recurrent Convolutional Networks for Visual Recognition and Description (<http://blog.csdn.net/joshuaxx316/article/details/69650952>)

相关推荐

广告

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>)

attention 机制 (http://blog.csdn.net/qj_26609915/article/details/52086772)

注意力机制（Attention Mechanism）在自然语言处理中的应用 (<http://blog.csdn.net/jdbc/article/details/52948351>)

Attention机制 (<http://blog.csdn.net/u011531010/article/details/59123388>)



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

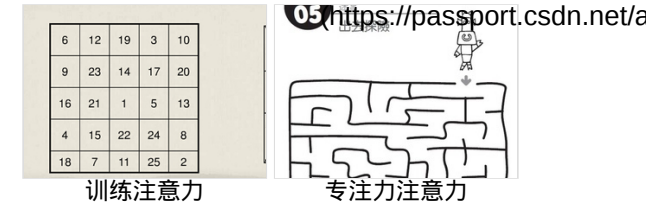




5



加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！



训练注意力

专注力注意力



注意力怎么训练



种头发原理

广告

广告



训练孩子注意力



他的热门文章



java中转义字符(回车\r\n)的替换 反转义 (h 内容举报
tp://blog.csdn.net/joshuaxx316/article/details/47760785)

15885



返回顶部

梯度下降算法中的Adagrad和Adadelta (ht
tp://blog.csdn.net/joshuaxx316/article/details/52062291)

登录

注册



首先让编码器输出结构化的表示，假设这些表示，可以用下述集合表示，

$$C = \{c_1, c_2, \dots, c_M\}$$

每个集合中的元素代表输入信息中某个时间片或者空间位置上的输入信息，输出 e 就是当前时刻 t 下，某个上下文分别表示对应的注意力，在经过 softmax 得到最后的得分。



5

$$e_i^t = f_{ATT}(z_{t-1}, c_i). \quad (16)$$

这里介绍 soft attention 机制，其实 attention 就是另外一个模型，这个模型负责给输入的上下文信息进行打分，好像一个解码器前面的预处理过程，其作用就是告诉解码器的输入，哪些编码器输出的内容重要的，这也就是注意力的意思，这个对应关系很好的反应了编码器对输入注意位置和权重的与解码器对输出注意位置



置和权重之间的相关性。



$$e_i^t = f_{ATT}(z_{t-1}, c_i, \{\alpha_j^{t-1}\}_{j=1}^M), \quad (10)$$

Soft 的意思是，这个模型所打的分数是通过 softmax 进行归一化的，最后每一个输入的上下文表示上对应的权重（即得分）和为 1，

$$\alpha_i^t = \frac{\exp(e_i^t)}{\sum_{j=1}^M \exp(e_j^t)}, \quad (11)$$

解码器对集合 C 中不同的上下文信息表示注意的程度（即概率）可以通过这个得分来反应。因此经过 Attention 模型处理后的上下文信息可以改写成下式子

$$c^t = \varphi \left(\{c_i\}_{i=1}^M, \{\alpha_i^t\}_{i=1}^M \right) = \sum_{i=1}^M \alpha_i^t c_i. \quad (13)$$

这种我们可以认为是一种方式的 Attention 机制，而另外一种通过 Attention 模型通过控制解码器去从 C 中挑

加入CSDN，享受更精准的内容推荐与5000万程序员共同成长的作用，上述两种方式即 Attention Model 的两

种类别：content-based 和 location-based

9457

(https://passport.csdn.net/a

Attention注意力机制--原理与应用 (http://blog.csdn.net/joshuaxx316/article/details/70665388)

9115

cannot open shared object file: No such file or directory如何解决 (http://blog.csdn.net/joshuaxx316/article/details/50553185)

6255

VMware Workstation中客户机与主机之间无法复制文件和剪切板问题解决 (http://blog.csdn.net/joshuaxx316/article/details/44901477)

4654



内容举报



返回顶部

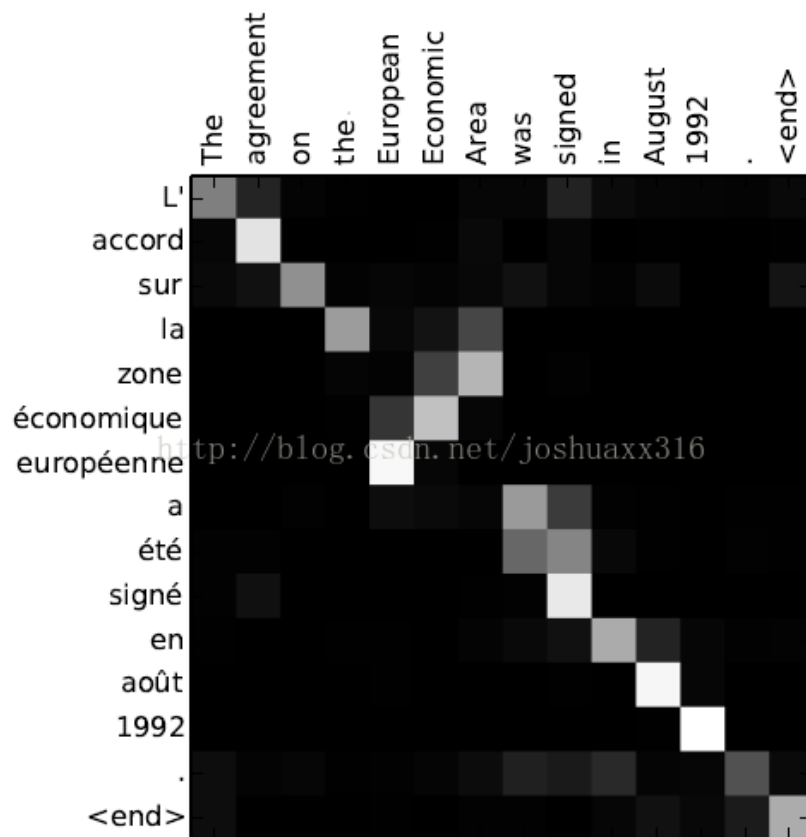
登录

注册



竹大空: content-based 和 location-based。

(<https://passport.csdn.net/a>



广告

由于定长上下文特征表示带来的信息损失，同时也是一种缺陷，由于不同的时间片或者空间位置的信息量明显有差别，利用定常表示无法很好的解决损失的问题，而Attention则恰好解决了这个问题。

我们甚至可以进一步解释，编解码器内部大致是如何工作的，当然从我看来这有点马后炮的意思，毕竟Attention是根据人的先验设计出来的，因此导致最后训练的结果朝着人指定的目标在走。可以说，就是输入第j维的上下文表示信息与第t时间片输出的关系，第j维可以是空间维度上，也可以是时序上。由于加入Attention会对输入的上下文表示进行一次基于权重的筛选,然而加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！
这种显示的筛选模式并不是人工制定这种机制所要看到的结果,而是通过此种加权的方式,让网络

⚠
内容举报

⬆
TOP
返回顶部

登录

注册

×

能学到空间上或者时序上的结构关系，前提当然是假设这里有一种原先不可解释的相互关系。以上图1可以很清晰地看出机器翻译问题中，加入的attention模型输出权重的分布与输入与输出信息的关系。

所以说Attention的作用是？



Attention的出现就是为了两个目的：1. 减小处理高维输入数据的计算负担，通过结构化的选取输入的子集，降低数据维度。2. “去伪存真”，让任务处理系统更专注于找到输入数据中显著的与当前输出相关的有用信息，从而提高输出的质量。Attention模型的最终目的是帮助类似编解码器这样的框架，更好的学到多种内容模态之间的相互关系，从而更好的表示这些信息，克服其无法解释从而很难设计的缺陷。从上述的研究问题可以发现，Attention机制非常适合于推理多种不同模态数据之间的相互映射关系，这种关系很难解释，很隐蔽也很复杂，这正是Attention的优势——不需要监督信号，对于上述这种认知先验极少的问题，显得极为有效。



让我们来看一个具体的例子！

这里直接上一幅图，举个具体的例子，然后咱们慢慢来解释：

让我们来看一下论文里其他研究者都是如何利用AttentionModel的：

最新一篇CVPR2017年accepted paper关于VQA问题的文章中，作者使用到了基于Spatial和基于Temporal两个层面的Attention模型，效果肯定提升了不用说，该问题更是极好的利用了这两点，其实这两种应用方式早在MT中得到了利用。

广告



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/a

广告

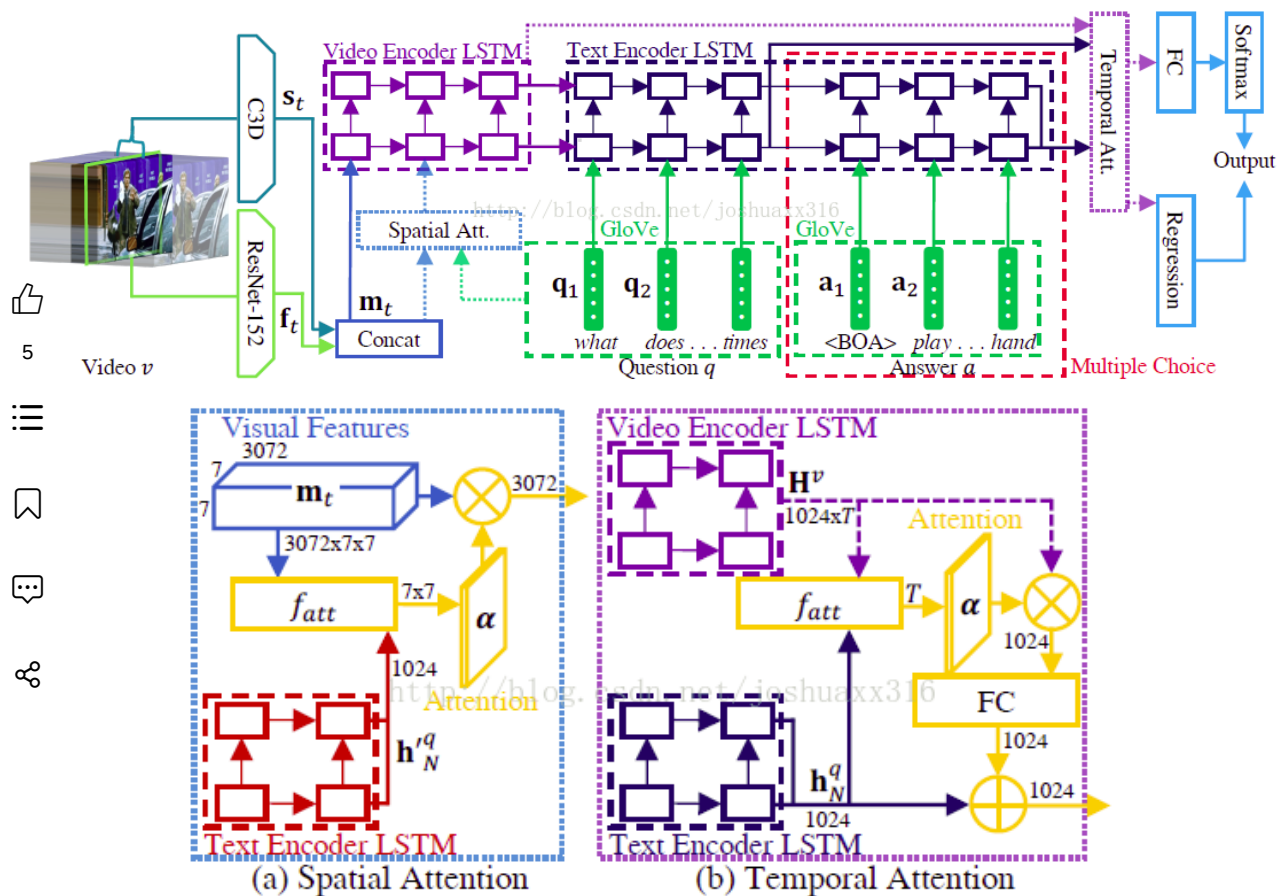


Figure 4. Our spatial and temporal attention mechanisms.

**今天实在太晚了，我要回家睡觉，先写到这，不然又不知道几点睡，不能总熬夜😴，周五前更新完

**上次说周五前更新完，结果拖了一周了，不过想想也没说哪个周五，啊哈哈。这里赶紧补上

上图的两种attention用法都属于soft attention，即通过确定性的得分计算来得到attended之后的编码状态，享受来自内容推荐，与5000万程序员共同成长！

内容举报

返回顶部

登录

注册

×

<https://passport.csdn.net/a>

Jang Y, Song Y, Yu Y, et al. TGIF-QA: Toward Spatio-Temporal Reasoning in Visual Question Answering[J]. arXiv preprint arXiv:1704.04497, 2017. (https://scholar.google.com/scholar_url?url=https://arxiv.org/pdf/1704.04497&hl=zh-CN&sa=T&oi=gsb-gga&ct=res&cd=0&ei=f6MKWZejGNa8jAGQmL6oCQ&scisig=AAGBfm3ylEBM4JKcJ1I7cgd418oZ-vGiHQ)

我们来讨论一下图示中的左边和右边两种attention是如何实现的。

(a) 空间注意力 Spatial Attention

对与视频转文字，图像转文字，VQA问题啊，这种task都需要视觉信息，那么和text信息一样，对visual content进行编码了以后，将编码后的特征表示直接接入解码器，可以得到我们想要的输出，例如text sentences。这就是基础的结构，但是现在加入空间Attention之后，如何改变结构呢。

假设前面视觉编码后的特征表示为 $7 \times 7 \times 3072$ 维featuremaps（论文里的参数） m_t ，然后来自text encoder的隐状态为 h_q 1×1024 ，生成一个 7×7 的attend mask， $a_t = f(h_q, m_t)$ ，attended featuremaps 为 $a_t m_t$ ，所以现在的维度不就变成了 1×3072 了吗？

广告



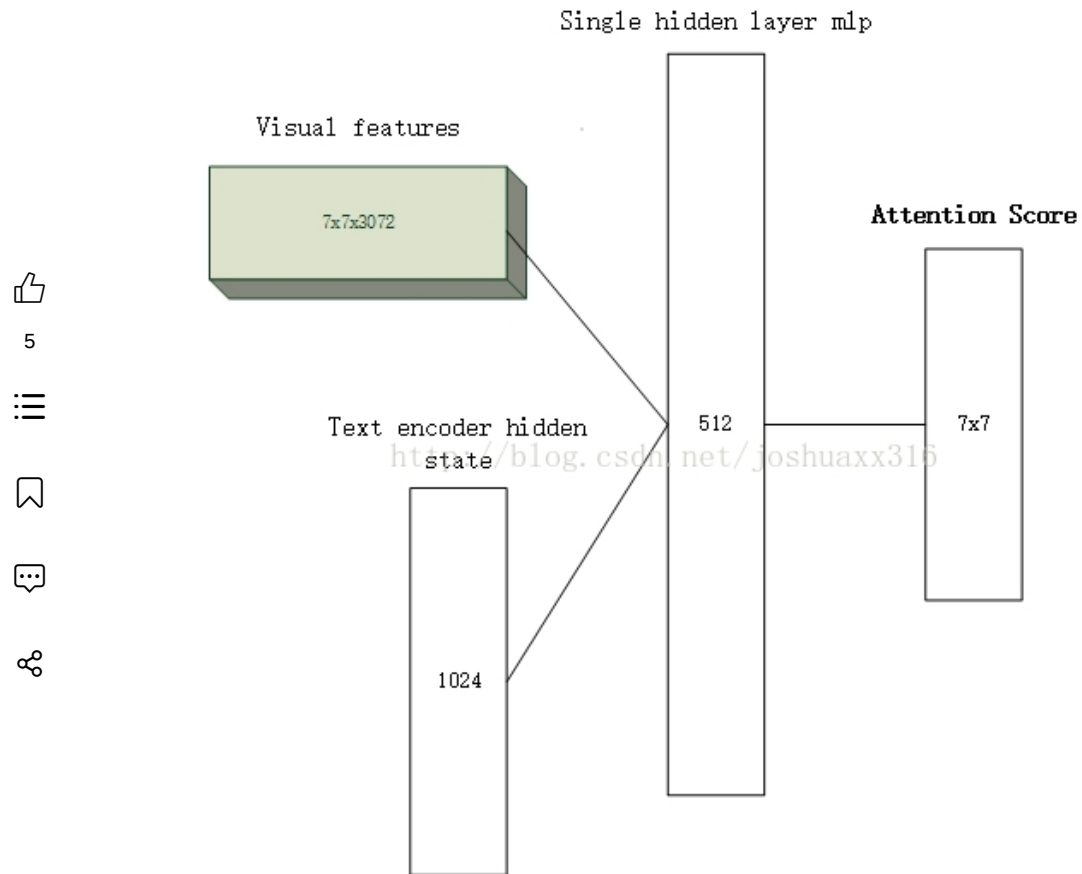
内容举报



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](#)



广告

(b) 时序注意力 Temporal Attention


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





wjf1997 (/wjf1997) 2017-11-30 10:17

1楼

(/wjf1997)啊，向博主学习

5 回复



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 57417

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 5210



【前端逆袭记】我是怎么从月薪4k到40k的！

谨以此篇文章献给我奋斗过的程序人生！我第一次编码是在我大一的时候....

加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！



内容举报



返回顶部

登录

注册



<https://passport.csdn.net/a>

(http://www.baidu.com/cb.php?c=lgF_pyfqHmknj0dP1f0IZ0qnfK9ujYzP1ndPWb10Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1dBnhmLnHlhnyDdPHmYPWRk0AwY5HDdnHn3njTdnWT0lgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEIAqspynElvNBnHqdlAdxTvqdThP-5yF_UvTkn0KzujYk0AFV5H00TZcq0KdpyfqHRLPjnvnfKEpyfqHc4rj6kP0KWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqPjn4rjn)

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



关于attention的发展综述文章



jdbc (<http://blog.csdn.net/jdbc>) 2016年10月27日 18:17 7397

广告

Attention机制 (<http://blog.csdn.net/u011531010/article/details/59123388>)

转载自：<http://blog.csdn.net/malefactor/article/details/50550211> Attention机制的基本思想是相当直观简洁的，是从人脑注意力模型引入的概念...



u011531010 (<http://blog.csdn.net/u011531010>) 2017年03月01日 20:09 766

序列模型中的注意力机制 (http://blog.csdn.net/shijing_0214/article/details/75194103)

现在很多研究的NLP问题都可以转换成一个Sequence to Sequence模型来解决，比如说机器翻译，智能问答，语音识别等。Sequence to Sequence模型由一个encoder和...



shijing_0214 (http://blog.csdn.net/shijing_0214) 2017年07月15日 23:43 1475

名称	不足表现
不注意	上课常走神、注意力不集中
分不清	影响记忆、理解、听错听懂
不记忆	听过的知识易忘记,英语学
如何训练注意力	

1	10	20	30	40
15	2	6	14	17
9	12	20	8	24
21	5	3	25	1
13	23	16	19	11
专注力注意力				



注意力怎么训练



注意力的训练



短信验证码接口



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



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...>)

<https://passport.csdn.net/a>

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 6405

常见的两种注意力机制 (<http://blog.csdn.net/AMDS123/article/details/65938986>)

为什么需要注意力机制 seq2seq模型已经在机器翻译、自然语言处理领域取得了很大的进步，基于贝叶斯条件概率公式，seq2seq由一个编码器和一个解码器构成，编码器将源序列A编码成一个固定长度的特征...



AMDS123 (<http://blog.csdn.net/AMDS123>) 2017年03月25日 18:45 7765

广告

人类注意力机制讨论 (<http://blog.csdn.net/u010165147/article/details/50964108>)

注意力是每个人几乎生下来就会的一项技能，但是我们对其工作机制却还知道的太少。随着认知科学，生物神经学，以及人工智能的发展我们对人类注意力机制的研究也在加速。尤其在人工智能领域，实现人类水平的AI，急需...



u010165147 (<http://blog.csdn.net/u010165147>) 2016年03月23日 18:15 1388

神经网络机器翻译Neural Machine Translation(2): Attention Mechanism (<http://blog.csdn.n...>)

端到端的神经网络机器翻译（End-to-End Neural Machine Translation）是近几年兴起的一种全新的机器翻译方法。前篇NMT介绍的基本RNN Encoder-Decoder结...



u011414416 (<http://blog.csdn.net/u011414416>) 2016年04月04日 18:48 11817

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

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


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...

(https://passport.csdn.net/a

 malefactor (<http://blog.csdn.net/malefactor>) 2016年01月26日 00:18 21284


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AM⁵，...

 mpk_no1 (http://blog.csdn.net/mpk_no1) 2017年08月06日 21:49 7067

广告

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 19201

attention model (<http://blog.csdn.net/apsvfb/article/details/59110169>)

How did I select papers?First, I tried to search for "attention" in CVPR2014-2016, ICCV2009-2015 an...

 apsvfb (<http://blog.csdn.net/apsvfb>) 2017年03月01日 16:16 506

王小草【深度学习】笔记第七弹--RNN与应用案例：注意力模型与机器翻译 (<http://blog.csdn.n...>

标签（空格分隔）：王小草深度学习笔记1. 注意力模型1.2 注意力模型概述注意力模型（attention model）是一种用于做图像描述的模式。在笔记6中讲过RNN去做图像描述，但是精准度可能差强...

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长 2017年12月08日 17:12 3012

 内容举报

 返回顶部

登录

注册

×

视觉注意力的循环神经网络模型 (http://blog.csdn.net/Leo_Xu06/article/details/53491400)

视觉注意力的循环神经网络模型



Leo_Xu06 (http://blog.csdn.net/Leo_Xu06) 2016年12月07日 20:43 3926



深度学习中的注意力机制 (<http://blog.csdn.net/TG229dvt5I93mxaQ5A6U/article/details/784...>)

作者 | 张俊林 责编 | 何永灿 最近两年，注意力模型（Attention Model）被广泛使用在自然语言处理、图像识别及语音识别等各种不同类型的深度学习任务中，是深度...



TG229dvt5I93mxaQ5A6U (<http://blog.csdn.net/TG229dvt5I93mxaQ5A6U>) 2017年11月02日 00:00 416



论文阅读: 图像分类中的注意力机制(attention) (<http://blog.csdn.net/Wayne2019/article/detai...>)



本文简要总结一下attention机制在图像分类任务中的应用。attention作为一种机制，有其认知神经或者生物学原理: 注意力的认知神经机制是什么？如何从生物学的角度来定义注意力？在计算...



Wayne2019 (<http://blog.csdn.net/Wayne2019>) 2017年11月09日 12:10 348

神经网络中注意力机制（Attention Mechanism）(http://blog.csdn.net/program_developer/...)

这篇博客讲解注意力机制很好，保存下来，忘记的时候，在看看！强烈推荐：<http://www.cnblogs.com/robert-dlut/p/5952032.html>...



program_developer (http://blog.csdn.net/program_developer) 2017年09月08日 16:17 745

CVPR 2017论文集锦（论文分类）——附录部分翻译 (<http://blog.csdn.net/dengheCSDN/art...>)
加入CSDN，享受更精准的内容推荐，与5000万程序员共同成长！

广告



内容举报



返回顶部

登录

注册



作为计算机视觉领域的三大顶级会议之一，CVPR 2017 又收录了很多优秀的文章。具体可参见 CVPR 的论文官网：<http://www.cvpapers.com/cvpr2017.html> ...



dengheCSDN (<http://blog.csdn.net/dengheCSDN>) 2017年08月23日 09:51 6452

(<https://passport.csdn.net/a>

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/liuchongge>...



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...



liuchongge (<http://blog.csdn.net/liuchongge>) 2017年06月06日 09:24 2531

广告



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

