

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

15

以Attention Model为例谈谈两种研究创新模式

原创

2016年01月26日 00:18:00

标签：自然语言处理 (http://so.csdn.net/so/search/s.do?q=自然语言处理&t=blog) /

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

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

21394

/* 版权声明：可以任意转载，转载时请标明文章原始出处和作者信息。*/

author: 张俊林

各位观众朋友好，也许此刻您刚打开电梯.....读这一篇之前，请您最好先拜读一下本篇的前传：文本处理中的Attention Model：是什么及为什么 (http://blog.csdn.net/malefactor/article/details/50550211)。因为那里有些背景知识需要交代。

话接上回书，在研读AttentionModel相关文献过程中，我再次深切感受到了科研中的两种创新模式：模型创新与应用创新。若干年前，也就是在我年轻不懂事的花样年华里，具体而言，就是在科学院读博士的后期，这种感受就已经比较明显，所以曾经在2006年写过一篇博客：自然语

weixin_3506... (//my.csdn.net?ref=toolbar)
张俊林博客 (http://blog.cs...)



+ 关注

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

码云

| | | | |
|-----|------|----|---------------|
| 原创 | 粉丝 | 喜欢 | 未开通 |
| 111 | 1936 | 2 | (https://gite |
| | | | utm_sourc |

他的最新文章

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

通过不断重置学习率来逃离局部极值点
(http://blog.csdn.net/malefactor/article/details/78768210)

深度学习中的注意力机制(2017版) (http://blog.csdn.net/malefactor/article/details/78767781)

机器码农：深度学习自动编程 (http://blog.csdn.net/malefactor/article/details/72853720)

言处理领域的两种创新观念 (<http://blog.csdn.net/malefactor/article/details/1415328>) (<http://blog.csdn.net/malefactor/article/details/1415328>)。当时谈的相对务虚一些，而且由于年富力
强，少不经事，更强调重大计算模型的创新，曾经对当时横扫NLP大多数领域的CRF式应用创
新深感失望（当然我承认我连应用创新都做不好，所以主动撤出了科研界去工业界卖苦力，不
加上这一句估计作者本人会被喷得狗血淋头^@@^）。而深度学习其实就是重大的模型创新，
在DL汹涌澎湃的大潮下，今天再来谈一谈这两种创新模式，而且目标更聚焦一些，我们就以
AM模型的研究过程来看看，我觉得AM的研究还是非常典型能明显说明这个问题的，当然这是
个普遍现象，从AM来谈只是作为典型例子而已。

15

首先把文本处理领域里面采用AM模型的论文尽量找全，这个倒是不难，因为AM被关注也就是
最近1年里的事情；然后分析下每个论文的创新点，再梳理下相互之间的关系，其实很容易看出
隐藏在后面的这两种创新模式。掌握这种模式对于做研究是很有帮助的，因为这代表了创新里
的一种固定的研究或者说是思维模式。就是说如果有这种意识，其实你很容易指导自己如何在
现有工作基础上去想创新思路，而不至于一说创新就感到很茫然。

应用创新

应用创新相对模型创新来说简单一些，核心思想就是：如果有一个新模型，那么我就拿来试试
不同的领域问题，看看能不能解决，如果能解决，那么这就是一种典型的应用创新。说得不好
听点就是说手里有把锤子，那就把很多问题看作钉子，然后到处敲敲，看看是不是能把钉子砸
到地里面。其实也没什么好听不好听的，因为非理论学科比如NLP，你看到的相当多数论文，
哪怕是顶级会议的论文，其实都是这种类型。这种创新的价值主要是能够证明：某个现成的模
型应用在某个领域是有效的。其实我们这些在公司食堂里吃饭的人更偏爱这种类型的论文，因
为简单粗暴对口，对我们来说，能简单方便可行地解决手上问题的论文就是好论文。

下面我们拿AM模型研究来具体说明下这种创新模式。

上一篇介绍AM模型的文章已经讲了AM的基本思路，那里提到的模型一般被称作Soft Attention
Model，是很通用的，目前大多数使用AM模型的场景其实都是这个模型的应用或者模型变体。

极深网络（ResNet/DenseNet）：Skip
Connection为何有效及其它 (<http://blog.csdn.net/malefactor/article/details/67637785>)

NeuralFinder:集成人工生命和遗传算法
自动发现神经网络最优结构 (<http://blog.csdn.net/malefactor/article/details/63250929>)

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)

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

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>)

论文[1]“Neural machine translation by jointly learning to align and translate”是首先把Soft Attention Model使用到机器翻译的论文，后面的NLP领域使用AM模型的文章一般都会引用这篇论文，但是这篇论文的思路也是从图像处理领域里的AM模型借鉴到MT领域的，所以也是典型的应用创新。就是说看到AM模型在图像处理领域有用，那么拿来做下机器翻译看一看，果然有用，就形成了一篇应用创新型的论文。图1是该论文中讲解AM模型的示意图。这算影响力比较大的论文了，但是你仔细一看，其实也是应用创新。这种创新相对简单，这里说的简单不是否认论文算法简单，而是说这种创新模式是相对简单的，但是不能否认它的价值，因为各个领域科学研究的大多数进展就是这样逐步前行的。

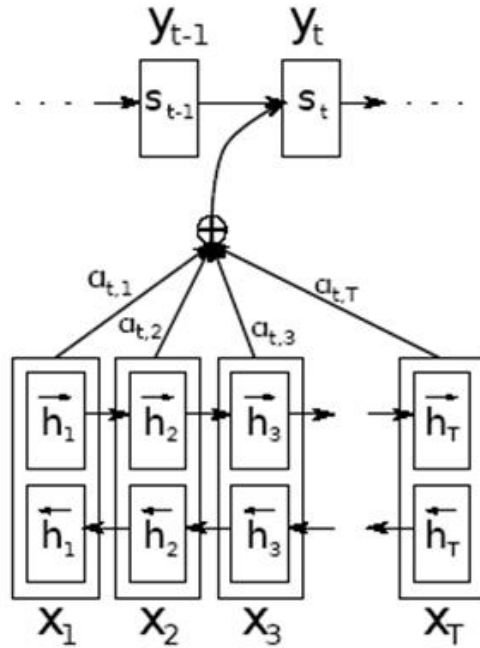
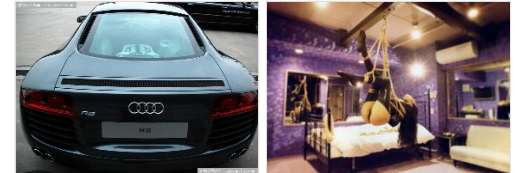


图1 论文1中的AM示意图

论文[2]“Neural Responding Machine for Short-Text Conversation”是华为李航老师研究组提出来的，论文中提出的Neural Responding Machine也是很典型的应用创新（图2），将带AM模型的Encoder-Decoder框架应用到对话机器人应用中，输入Encoder-Decoder的是一句对话，而Encoder-Decoder输出的则是对话机器人的应答，其训练数据采用了微博评论里的对话数据。

广告



奥迪r8二手

日本的酒店

香港住宿多少钱 人工智能网站

魔方公寓 二手玛莎拉蒂 行星减速机

4100万像素手机 透明手机价格

8寸大屏手机 学信网可查文凭

他的热门文章

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>)

57976

机器码农：深度学习自动编程 (<http://blog.csdn.net/malefactor/article/details/72853720>)

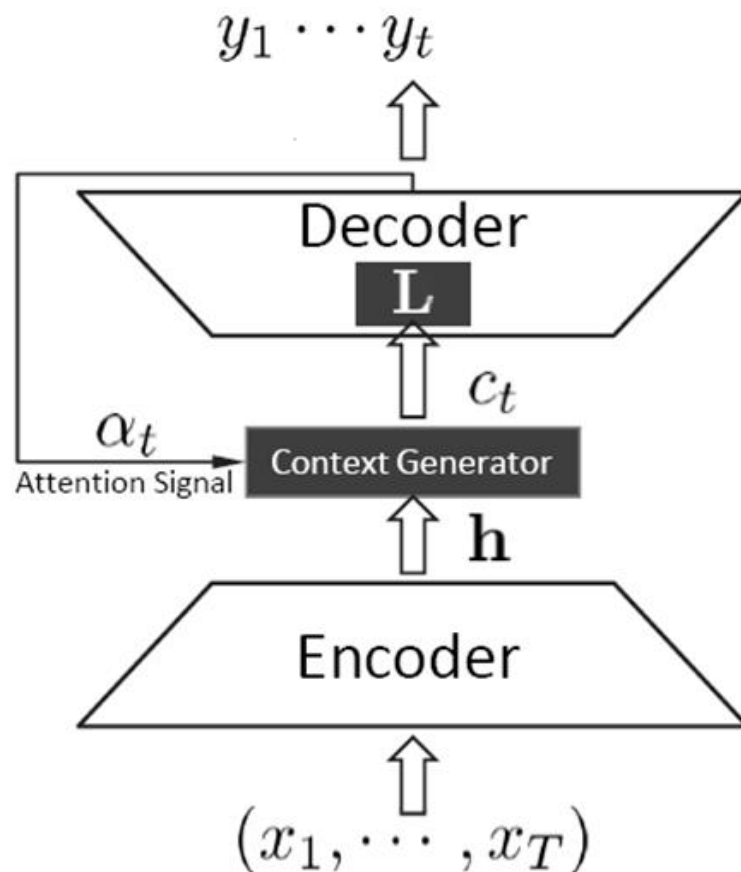


图2. Neural Responding Machine

论文中提到的Global Scheme就是不带AM的Encoder-Decoder框架，Local Scheme其实就是Soft Attention Model，Local and Global Model则是把Global和Local混合起来用。这篇论文证明了AM-Encoder-Decoder应用在对话系统上是有效的。说实话，这篇论文理论创新不算突出，但是相对传统的方法来说，这个思路还是非常简洁实用的。我个人也比较喜欢，因为很可能会照搬拿来做。

46745

深度学习与自然语言处理之五：从RNN到LSTM (<http://blog.csdn.net/malefactor/article/details/50436735>)

39988

Batch Normalization导读 (<http://blog.csdn.net/malefactor/article/details/51476961>)

38888

深度学习在搜索和推荐领域的应用 (<http://blog.csdn.net/malefactor/article/details/52040228>)

38773

Google的论文[3]:“Grammar as a Foreign Language”把带AM模型的Encoder-Decoder框架用在句法分析上，也是用的Soft AM模型。这也是一种应用创新，但是我个人觉得这个论文的创新性还是比较高，主要在于思考问题的思路，一般做句法分析就是去把句子Parse成一颗树，而他们则把句法的树形结构拍平搞成线性序列，然后把<句子，句法关系>当做一个不同语言的翻译问题，这个给人的感觉还是非常新颖的。当然也许我这方面文献看得少，已经有人这么做过，但是我不清楚。



论文[4]则是把带AM的Encoder-Decoder应用在生成式文本摘要中，本质上也是Soft AM模型，并证明模型有效。



从上面所述可以看出，目前NLP里面应用AM模型的论文很多属于这种应用创新类型的。这种创新模式相对简单些，你可以想个新的领域，并试试带AM的Encoder-Decoder框架是否有效，如果有效而且没人做过，那么发一篇好会议的论文应该不成问题。



模型创新

如果把应用创新比作拿着锤子找钉子敲，那么模型创新也可以做类似的比喻：如果你手里有把锤子，但是死活不能把你手里的钉子敲到地里去，那么你就去改造一下锤子，比如加大重量，或者把锤子柄改得长一点，做出敲钉子敲得更利落的锤子2.0版（此处非为罗永浩做广告，特此声明），那么这就是模型创新。意思是改造一下从别人手里直接拿来的模型，让它更适合解决手头的问题。

模型创新的难度要比应用创新难些，因为它要求你对锤子的构造要透彻了解，你得清楚锤子到底是哪里有问题，才不能敲打手头的钉子，怎么改造才能让它升级。相对去找个新的钉子，这种方式要求会更高一些。

但是其实改造锤子也是有一些模式在里面的，下面我们还是以AM研究作为例子来谈谈AM领域里的模型创新，并抽象出三种模型创新模式：混搭风；坦白从宽型和抗拒从严型（^@@^）。

Soft AM vs Hard AM

上一篇讲AM模型的科普文介绍了Soft Attention Model，所谓Soft，意思是在求注意力分配概率分布的时候，对于输入句子X中任意一个单词都给出个概率，是个概率分布。那么相对Soft，就有相应的Hard Attention Model，提出Hard版本就是一种模型创新。既然Soft是给每个单词都赋予一个单词对齐概率，那么如果不这样做，直接从输入句子里面找到某个特定的单词，然后把目标句子单词和这个单词对齐，而其它输入句子中的单词硬性认为对齐概率为0，这就是Hard Attention Model的思想。

15

所以这就是从Soft AM衍生出Hard AM的典型思维过程。Hard AM在图像里证明有用，但是估计在文本里面用处不大，因为这种单词——对齐明显要求太高，如果对齐不齐对后续处理负面影响很大，所以你在NLP的文献里看不到用Hard AM的，估计大家都试过了，效果不好。

既然有Soft AM了，也有Hard AM了，这个思路好像没法再创新了吧？其实不然，你可以试试两者的中间地带：不软不硬AM。就是介于Soft 和Hard之间：Soft AM要求输入句子每个单词都要赋予单词对齐概率，Hard AM要求从输入句子中精确地找到一个单词来和输出单词对齐，那么可以放松Hard的条件，先找到Hard AM在输入句子中单词对齐的那个单词大致位置，然后以这个单词作为轴心，向左向右拓展出一个大小为D的窗口，在这个 $2D+1$ 窗口内的单词内进行类似Soft AM的对齐概率计算即可。你看，这就是又一种基于AM的模型创新方法：两头取中间混搭方法。孔子他老人家其实是这种创新方式的最早提出者，所谓“中庸之道”么。

斯坦福大学Manning研究组（题外话，Manning是NLP领域大神级人物，我就是从小看他写的那本《统计自然语言处理》长大的^@@^）在ACL2015发表的论文“Effective Approaches to Attention-based Neural Machine Translation”其实就是这个混搭AM模型的思路。在论文里面，提出了两种模型：Global Attention Model（图3）和Local Attention Model（图4），Global Attention Model其实就是Soft Attention Model，Local Attention Model本质上是Soft AM和Hard AM的一个混合或折衷。一般首先预估一个对齐位置 P_t ，然后在 P_t 左右大小为D的窗口范围来取类似于Soft AM的概率分布。

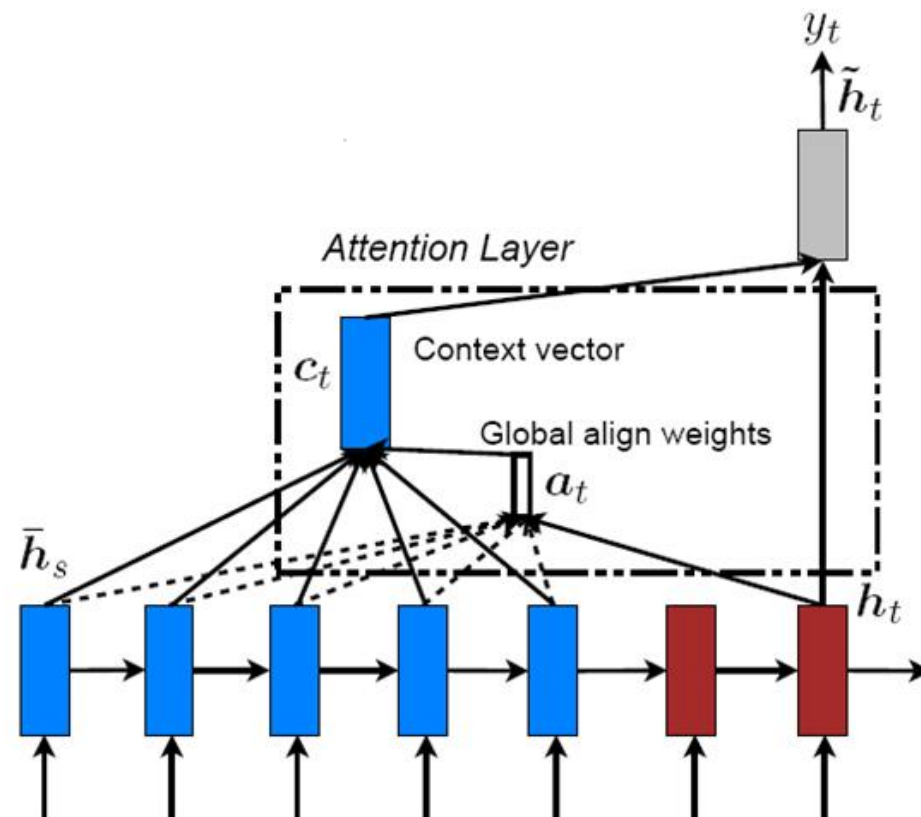


图3. Global Attention Model

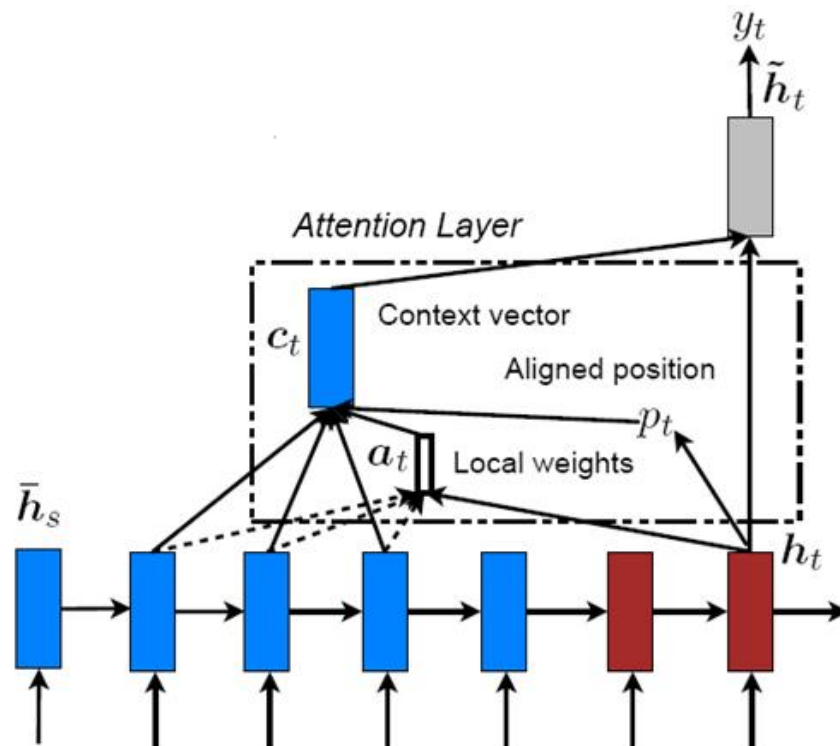


图4. LocalAttention Model

因为论文[1]在斯坦福之前提出将AM用于机器翻译，所以斯坦福的论文中提的那个Global AM模型作为一个论文创新性是不足的，不能算应用创新。斯坦福的这篇论文的主要创新就是提出了这个混搭AM模型，并在机器翻译里面证明有效。这是一种典型的模型创新。模型创新给人的感觉往往是：如果你看了会觉得怎么思路这么简单，但是在没提出来之前你还真不一定能想到。这也是为何说模型创新难度要大于应用创新的原因之一。

动态AM vs 静态AM

Soft AM和Hard AM是在求输入句子中注意力分配概率所覆盖单词范围大小上做文章。那么还有没其它可改进的角度？其实还有。

Soft AM有一个要求：就是对于输出句子中的每个单词，都需要计算针对这个单词来说，输入句子每个单词的单词对齐概率。那么能不能要求松一点，宽大处理呢？就是说对于整个输出句子来说，整体对输入句子求出一个注意力分配概率分布就行，而不要求每个单词都需要算呢？这是可以的，这也算一种模型创新。一般把这种模型叫做静态AM，而Soft AM叫做动态AM。之所以说是动态，意思是每个单词都需要动态去算的意思，而静态是指整个输出句子算一个就够了。



Google的论文“Teaching Machines to Read and Comprehend”是用AM模型来解决问答系统问题的，就是给出一个问题，需要在文章里面找到对应的答案。这篇论文提出了两个AM模型：“Attentive Reader”（图5）和“Impatient Reader”（图6），其实Impatient Reader就是Soft Attention Model，即动态AM，而Attentive Reader就是静态AM。因为在问答场景下，对于一个问题，它的焦点是文章中的答案，所以问句整体有一个注意力分配概率就行。

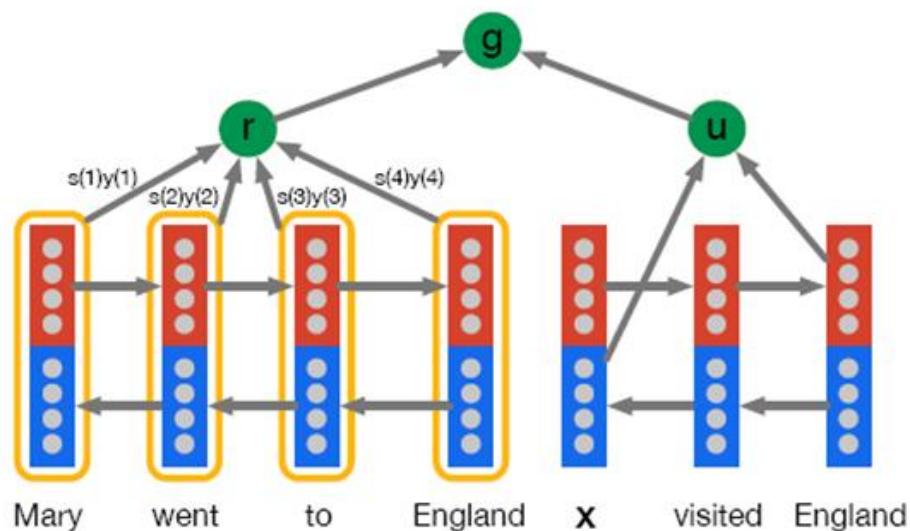


图5. Attentive Reader

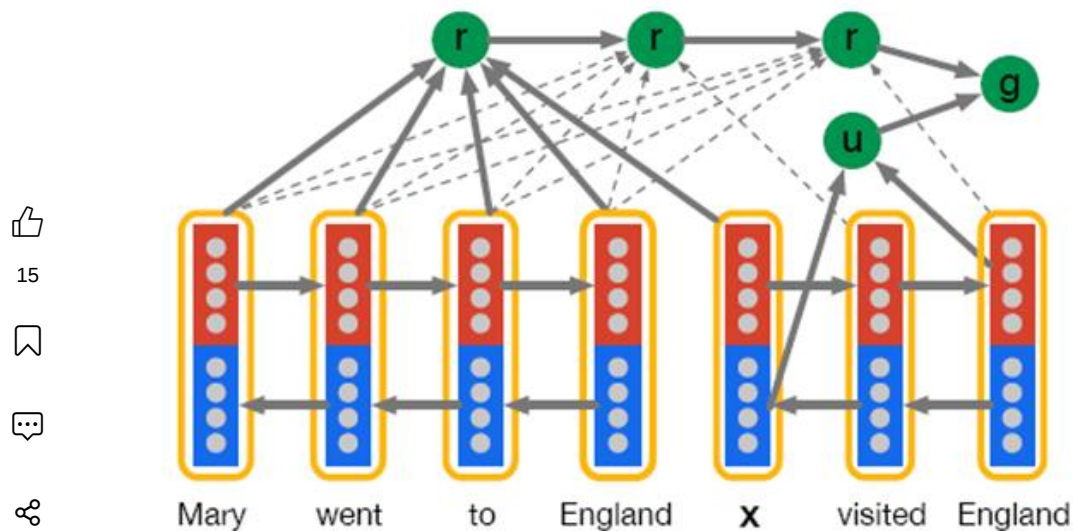


图6. impatient Reader

从这里也能归纳出模型创新的一种方法：就是放松约束，坦白从宽的方式；

Soft AM vs 强制前向AM

对于AM来说，能改的地方差不多了吧？还能有其他改进AM的思路吗？你自己可以想想。

答案是还可以继续改。

Soft AM在逐步生成目标句子单词的时候，是由前向后逐步生成的，但是每个单词在求输入句子单词对齐模型时，并没有什么特殊要求。强制前向AM则增加了约束条件：要求在生成目标句子单词时，如果某个输入句子单词已经和输出单词对齐了，那么后面基本不太考虑再用它了，因为输入和输出都是逐步往前走的，所以看上去类似于强制对齐规则在往前走。

论文[7]使用带AM的Encoder-Decoder框架来做语音识别问题，在论文里就是采取了强制前向AM的思路。所以，这也是一种典型的模型创新方式，就是说，增加约束，抗拒从严的思路。在目前NLP中采用AM模型的论文中，能看到的较明显的模型创新基本都在上边罗列出来了。看上去好像对AM进行改进已经穷途末路，走投无路，问天天不应了，但这肯定是假象，后面一定会陆续冒出新的改进AM模型，只是一时半会你想不出来而已。就比如，我拍下脑袋瞎咧咧几句，也许还可以从这几个角度来改造AM这个锤子：



角度1：目前AM指的只是对于某个输出单词来说，对输入句子单词求注意力分配概率分布，那么能不能把输出句子当前要生成的单词 y_i 之前已经输出的部分单词序列(y_1 到 y_{i-1})也求一个注意力分配概率分布呢？当然，乍一想这不算太好的主意，因为Decoder如果是RNN或者LSTM的话，其实已经在隐层节点状态隐含了部分这种信息了，但是你别忘了，不带AM的Encoder-Decoder框架里面，如果Encoder用的是RNN，理论上也已经部分考虑进来位置的影响力信息了，但不是还是提出了AM吗？那为啥要歧视Decoder部分再单独加上一个AM模型呢？你说是不是这个道理？如果显示地把它融入到AM模型里面，也许对于输出句子很长的场景也是有帮助的。

角度2：目前AM模型在输出 y_i 的时候，是对每一个单词求单词对齐概率的，那么能否对于输出序列单词以及输入序列单词进行组团呢？比如不只一个单词而是若干个单词组成的短语级别来计算这种对齐概率呢，这类似于把短语对齐的概念引入到AM模型中来；

当然，如果你要是像我一样喜欢做空中的梦想家，这么天马行空地瞎想，可能还有其它各种拿AM来折腾的思路，也许哪天你就鸿运当头，看美女看得目瞪口呆正好天上掉馅饼砸到你嘴里，瞎猫碰见个死耗子什么的，是吧？在A股市场走势飞流直下，还不如被关停之际，我们要勇于逆流而上去拥抱“中国梦”，是吧。

在实际做创新的时候，并非是应用创新和模型创新泾渭分明的，有时候会混用两种创新模式，就比如：你可以把(混合AM+强制前向AM)再混合一下，然后找个新的应用领域，那么肯定算是一种崭新的创新思路。也就是可以用目前的已有的模型创新和应用创新再排列组合一下，当然这种创新一般人听起来觉得意思不大，但是您肯定属于那种每天早上对着镜子冲自己举起小拳

头说自己不是一般人那种，对不，值不值得去做在于你自己怎么看待这个问题。立志于做大创新的同志肯定不屑于做这种组合式创新，但是对于憋论文憋得都快流产的同志们来说，这也不失为一种创新思路和生路，毕竟即使这么做，你也得去踩很多雷，才能趟出一条路，是吧。

对了，话都说到这个份上了，天台上排队的同学们，你们这回总能安心下来了把，还有很多雷等着你去踩呢。



NLP领域AM相关参考文献：

[1]Dzmitry Bahdanau, Kyungh yun Cho, and Yoshua Bengio.2015. Neural machine translation by jointly learning to align and translate. In ICLR 2015.

[2]LifengShang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short-Text Conversation.ACL 2015.



[3]Oriol Vinyals.etc. Grammar as a Foreign Language.arXiv.2015.google.

[4]Alexander M. Rush. etc.ANeural Attention Model for Sentence Summarization. EMNLP 2015.

[5]Minh-Thang Luong Hieu PhamChristopher D. Manning. Effective Approaches to Attention-based Neural Machine Translation.ACL 2015.

[6]Karl Moritz Hermann.etc.Teaching Machines to Read and Comprehend.arXiv.2015.google.

[7]Jan Chorowski.etc. End-to-end Continuous Speech Recognition using Attention-based Recurrent NN: First Results.arXiv.2015.



扫一扫关注微信号：“布洛卡区”，深度学习在自然语言处理等智能应用的

技术研讨与科普公众号。

15



发表你的评论

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



mao_feng (/mao_feng) 2017-08-31 00:06

16楼

(/mao_feng)

回复



Alan_Corleone (/Alan_Corleone) 2017-03-08 17:24

15楼

(/Alan_Corleone) 非常感谢您的分享

回复



Alan_Corleone (/Alan_Corleone) 2017-03-08 17:19

14楼

(/Alan_Corleone) 分享内容很赞

回复

查看 19 条热评

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

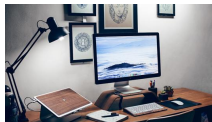


深度学习笔记——Attention Model（注意力模型）学习总结 (http://blog.csdn.net/mpk_no1/a...)

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



广告

票选结果：Python再上天，微软要求全员学Python？

宇宙语言Python荣登年度排行榜，吴恩达，微软纷纷为它站台，Python这么牛逼的原因是....

(http://www.baidu.com/cb.php?c=lgF_pyfqHmknjnvPjc0IZ0qnfK9ujYzP1nYPH0k0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1YkuAD3uAwBnHu9ujw9mW6v0AwY5HDdnHfznWRYPHT0lgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEIAqspynETZ-YpAq8nWqdlAdxTvqdThP-5yF_UvTkn0KzujYk0AFV5H00TZcqn0KdpyfqHRLPjnvnfKEpyfqHc4rj6kP0KWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqnH6Yrj6)

浅谈Attention-based Model【原理篇】 (<http://blog.csdn.net/wuzqChom/article/details/757...>)

转载请标明出处：<http://blog.csdn.net/wuzqchom/article/details/75792501> 计划分为三个部分：浅谈Attention-based Model

【...



wuzqChom (<http://blog.csdn.net/wuzqChom>) 2017年07月22日 19:24 1662

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



10.80/条
HDMI高清线 HDMI高
清线 高清投影仪专用



2.35/条
HDMI线厂家 HDM高
清线 3D 带网带环1.4



0.20/条
数据延长线usb 2.0延
长线1米公对母背包电

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.net/shenxiaolu1984>)


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



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

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  7539

15

以Attention Model为例谈谈两种研究创新模式 (<http://blog.csdn.net/jdbc/article/details/5058...>)



本文主要介绍以Attention Model为例谈谈两种研究创新模式。

 jdbc (<http://blog.csdn.net/jdbc>) 2016年01月26日 11:19  1338





基于attention的video描述 (http://blog.csdn.net/qq_21190081/article/details/53069237)

引言这篇博文内容主要基于2015年Y Li的论文Video Description Generation Incorporating Spatio-Temporal Features and a S O...

 qq_21190081 (http://blog.csdn.net/qq_21190081) 2016年11月07日 19:26  1796

神经网络、深度学习创新点的思考 (<http://blog.csdn.net/lanchunhui/article/details/53006238>)

从特征学习 (Feature Learning) 的观点来看，神经网络尤其是深度神经网络 (也即是深度学习) 是十分强大的特征学习方法。例如就可将 Autoencode 视为一种特征降维的方法。也正因如此，在...

 lanchunhui (<http://blog.csdn.net/lanchunhui>) 2016年11月02日 11:21  935

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 20:09 781

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 12:10 380



ICLR 2017 | Attention 和 Memory Networks (<http://blog.csdn.net/omnispace/article/detail...>)



ICLR 2017 | Attention 和 Memory Networks 原创2016-11-09小 S程序媛的日常程序媛的日常 今天分享 ICLR 2017，主题是 Attention ...



omnispace (<http://blog.csdn.net/omnispace>) 2017年02月09日 06:40 1587

使用深度学习打造智能聊天机器人 (http://blog.csdn.net/qq_34281100/article/details/518333...)

深度学习入门



qq_34281100 (http://blog.csdn.net/qq_34281100) 2016年07月05日 17:41 1511

CNN在句子相似性建模的应用续--基于attention的多角度CNN模型 (<http://blog.csdn.net/liuch...>)

今天我们来介绍一下“UMD-TTIC-UW at SemEval-2016 Task 1: Attention-Based Multi-Perspective Convolutional Neura...



liuchonge (<http://blog.csdn.net/liuchonge>) 2017年03月24日 14:17 1934