

[CSDN首页 \(http://www.csdn.net?ref=toolbar\)](http://www.csdn.net?ref=toolbar)[学院 \(http://edu.csdn.net?ref=toolbar\)](http://edu.csdn.net?ref=toolbar)[下载 \(http://download.csdn.net?ref=toolbar\)](http://download.csdn.net?ref=toolbar)

更多 ▾

[下载 CSDN APP \(http://www.csdn.net/app?ref=toolbar\)](http://www.csdn.net/app?ref=toolbar)[写博客 \(http://write.blog.csdn.net/postedit?ref=toolbar\)](http://write.blog.csdn.net/postedit?ref=toolbar)[首页 \(http://blog.csdn.net/\)](http://blog.csdn.net/)[移动开发 \(http://blog.csdn.net/mobile/newarticle.html\)](http://blog.csdn.net/mobile/newarticle.html)

全部 ▾

[登录 \(https://passport.csdn.net/account/login?ref=toolbar\)](https://passport.csdn.net/account/login?ref=toolbar) | [注册 \(http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister\)](http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister)**CSDN** (<http://www.csdn.net>)

【图像理解】之Show, attend and tell算法详解

原创 2016年08月13日 22:10:16

6138

8

2

Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." arXiv preprint arXiv:1502.03044 (2015).

聚焦机制 (Attention Mechanism) 是当下深度学习前沿热点之一, 能够逐个关注输入的不同部分, 给出一系列理解。这篇论文是聚焦机制代表作, 完成了图像理解中颇具难度的“看图说话”任务。

作者提供了基于Theano的源码(戳这里) (<https://github.com/kelvinxu/arctic-captions>), 另外有热心群众在Tensorflow上给出了实现(戳这里) (https://github.com/jazzsaxmafia/show_attend_and_tell.tensorflow)。

本文对照Tensorflow版本源码, 详解论文算法。

数据结构

从输入到输出经历编码和解码两个部分。

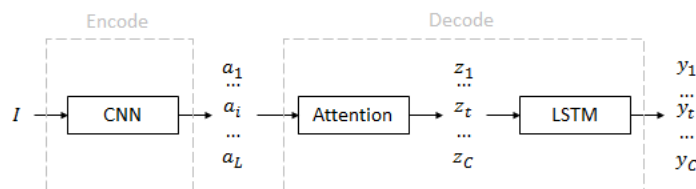
类比: 在机器翻译中, 编码部分把源语言变成基本语义特征, 解码部分把基本语义特征变成目标语言。

输入: 图像 I

特征(annotation): $\{a_1 \dots a_i \dots a_L\}$

上下文(context): $\{z_1 \dots z_t \dots z_C\}$

输出(caption): $\{y_1 \dots y_t \dots y_C\}$



I 是输入的彩色图像。

有顺序的 y_t 组成一句“说明”(caption)。句子长度 C 不定。每个单词 y_t 是一个 K 维概率, K 是词典的大小。

a_i 是一个 D 维特征, 共有 L 个, 描述图像的不同区域。

z_t 也是一个 D 维特征, 共有 C 个, 表示每个单词对应的上下文。

释义 a_i 是一次生成的, 但单词 z_t 是逐个生成的, 所以使用下标 t 来强调每一次估计。

网络结构

shenxiaolu1984 ([http://bl...](http://blog.csdn.net/shenxiaolu1984))

+关注

<http://blog.csdn.net/shenxiaolu1984>

原创

粉丝

喜欢

码云

未开通

63

0

0

[https://gite](https://github.com/shenxiaolu1984)

他的最新文章

[更多文章 \(http://blog.csdn.net/shenxiaolu1984\)](http://blog.csdn.net/shenxiaolu1984)

【平价数据】SimGAN: 活用合成数据和无监督数据 (/shenxiaolu1984/article/details/76160306)

【平价数据】GAN用于半监督学习 (/shenxiaolu1984/article/details/75736407)

【深度学习】One Model to Learn Them All详解 (/shenxiaolu1984/article/details/73736259)

【深度学习】DeepMind关系推理网络 (/shenxiaolu1984/article/details/73522056)



Unable to Conn

The Proxy was unable to connect to the remote site. responding to requests. If you feel you have reached please submit a ticket via the link provided below.

URL: <http://pos.baidu.com/s?hei=250&wid=300&di=u%2Fblog.csdn.net%2Fshenxiaolu1984%2Farticle%2F>

编辑推荐

最热专栏

[深度学习论文笔记][Attention]Show, ...
showand tell笔记 (/lijianqing/article/de...
Show, Attend and Tell: Neural Image ...
Show and Tell: Lessons learned from t...
全局变量报错: UnboundLocalError: I...

在线课程

编码($I \rightarrow a$)

输入图像 I 归一化到 224×224 。特征 a 直接使用现成的VGG网络1中conv5_3层的 $14 \times 14 \times 512$ 维特征。区域数量 $L = 14 \times 14 = 196$ ，维度 $D = 512$ 。
为了更好地描述局部内容，所以使用了较低层级的特征。

编码只进行一次，解码是逐个单词进行的，所有以下网络变量均带有步骤下标 t 。

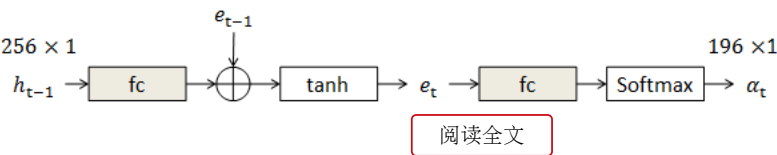
上下文生成($a_t \rightarrow z$)

当前步骤的上下文 z_t 是原有释义 a 的加权和，权重为 α_t 。和 a_i 类似， z_t 也是一个 D 维向量。：

$$z_t = \alpha_t^T \cdot a$$

α_t 维度为 $L = 196$ ，记录释义 a 每个像素位置获得的关注。

权重 α_t 可以由前一步系统隐变量 h_t 经过若干全连接层获得。编码 e_t 用于存储前一步的信息。灰色表示模块中有需要优化的参数。



版权声明：本文为博主原创文章，未经博主允许不得转载。 ⚠ 举报

标签：深度学习 (<http://so.csdn.net/so/search/s.do?q=深度学习&t=blog>) /
CV (<http://so.csdn.net/so/search/s.do?q=CV&t=blog>) /
图像理解 (<http://so.csdn.net/so/search/s.do?q=图像理解&t=blog>) /
机器学习 (<http://so.csdn.net/so/search/s.do?q=机器学习&t=blog>)

自然语言处理在“天猫精灵”中的应用
实验应用
(http://edu.csdn.net/huiyiCourse/series_detail/64?utm_source=blog9)

大数据平台架构及Druid引擎实践
(http://edu.csdn.net/huiyiCourse/series_detail/64?utm_source=blog9)

xaoshuan (/xaoshuan) 2017-02-28 11:27 6楼

作者写的非常好，我想请教一下，该论文实现 隐变量生成 该部分的时候用到了x_t，请问这里的x_t什么意思？文章只提到：输入i、输出o和遗忘f是三个“门变量”，用来控制其他状态的强度，都可以通过上一步骤的隐状态h，以及当前上下文z决定

回复

xaoshuan (/xaoshuan) 2017-02-28 11:25 5楼

作者写的非常好，我想请教一下，该论文实现 隐变量生成 该部分的时候用到了x_t，请问这里的x_t什么意思？文章只提到：输入i、输出o和遗忘f是三个“门变量”，用来控制其他状态的强度，都可以通过上一步骤的隐状态h，以及当前上下文z决定

回复

喜欢

xaoshuan (/xaoshuan) 2017-02-28 11:23 4楼

作者写的非常好，我想请教一下，该论文实现 隐变量生成 该部分的时候用到了x_t，请问这里的x_t什么意思？文章只提到：输入i、输出o和遗忘f是三个“门变量”，用来控制其他状态的强度，都可以通过上一步骤的隐状态h，以及当前上下文z决定

回复

评论

收藏

分享

查看 8 条热评

相关文章推荐

[深度学习论文笔记][Attention]Show, Attend, and Tell: Neural Image Caption Generation with Visual Attention (/hao_zhang_vision/article/details/53175979)

Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." a...

Hao_Zhang_Vision (http://blog.csdn.net/Hao_Zhang_Vision) 2016-11-15 19:53 2083

showand tell笔记 (/lijianqing/article/details/70738879)

lijianqing (http://blog.csdn.net/lijianqing) 2017-04-25 14:17 188

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention (/cuixiaoxue/article/details/70138784)

动机: 以前的工作都是只关注图像一次，然后使用最后一层的全连接层得到图像最有用的信息。这样的缺点就是在描述图像的时候丢失了很多有用的信息。而且，产生每个词用到的图像信息都是相同的。因此，在这篇论文中考...

cuixiaoxue (http://blog.csdn.net/cuixiaoxue) 2017-04-12 08:55 336



Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge代码 (/liyaohhh/article/details/53955715)

Show and Tell: Lessons learned from the 2015 MSCOCO ...

liyaohhh (http://blog.csdn.net/liyaohhh) 2016-12-31 15:19 1038



全局变量报错：UnboundLocalError: local variable 'l' referenced before assignment (/my2010sam/article/details/17735159)

总结：内部函数，不修改全局变量可以访问全局变量内部函数，修改同名全局变量，则python会认为它是一个局部变量在内部函数修改同名全局变量之前调用变量名称（如print sum），则引发Unbou...

 my2010Sam (<http://blog.csdn.net/my2010Sam>) 2014-01-01 16:35  40362



Centos出现-bash: unzip: command not found的解决办法 (/youcai35/article/details/51435662)

利用unzip命令解压缩的时候，出现-bash: unzip: command not found的错误。unzip——命令没有找到，其原因肯定是没有安装unzip。利用...

 youcai35 (<http://blog.csdn.net/youcai35>) 2016-05-17 14:11  6399

weblogic学习笔记：域创建+应用部署 (/lh87522/article/details/44353685)

参考地址：http://docs.oracle.com/cd/E13222_01/wls/docs92/index.html 文档基于版本：weblogic92 域（domain）是weblogic...

 lh87522 (<http://blog.csdn.net/lh87522>) 2015-03-17 22:04  3581



自然语言处理中的Attention Model: 是什么及为什么 (/malefactor/article/details/50550211)

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>) 2016-01-20 18:26  44805



attention model (/u014568921/article/details/52634977)

自然语言处理中的Attention Model: 是什么及为什么深度学习和自然语言处理中的attention和memory机制

 u014568921 (<http://blog.csdn.net/u014568921>) 2016-09-23 10:38  702



【深度学习】聚焦机制DRAM(Deep Recurrent Attention Model)算法详解 (/shenxiaolu1984/article/details/51518578)

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

 shenxiaolu1984 (<http://blog.csdn.net/shenxiaolu1984>) 2016-06-28 22:14  4513























【论文笔记】Show and Tell: Lesson learned from the 2015 MSCOCO Image Captioning Challenge (/qq_35647180/article/details/61921782)

AK在其开源的neuraltalk和neuraltalk2项目中和此模型进行比较其相似之处，并承认说“but the Google release should work significantly ...

 qq_35647180 (http://blog.csdn.net/qq_35647180) 2017-03-13 17:05  399



TensorFlow下运行Google的Im2txt: show and tell inception v3 (/gbbb1234/article/details/70543584)

我的设备：ubuntu14.04+GPU TensorFlow1.0.1 相关论文《Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge》

<div><div> g 34 (http://blog.csdn.net/gbbb1234)</div><div>2017-04-23 19:06</div><div> 1904</div></div> <div>(http://download.csdn.net/detail/jian_5030624/1144075)</div> <div>一句话搞定面向对象 Keep it DRY, Keep It shy, and tell the other guy (http://download.csdn.net/detail/jian_5030624/1144075)</div>	
<div><div> 2009-03-25 19:29</div><div>5KB</div><div> 下载</div></div> <div><div> 眼</div><div>show and tell,让机器看图说话 (show attention and tell) (/luoyang224/article/details/76599736)</div><div> 喜欢</div><div>对show and tell 模型做的一个总结, 包括其中应用的attention机制。</div><div> luoyang224 (http://blog.csdn.net/luoyang224)</div><div>2017-08-02 23:02</div><div> 102</div></div> <div><div> 收藏</div><div> 评论</div><div> 分享</div></div>	
<div><div> What Microsoft, Oracle, IBM, And SAP Don't Tell Customers (/samurais/article/details/7000941)</div><div>Share The four big software vendors -- Microsoft, Oracle, IBM, and SAP -- have hidden...</div><div> watson243671 (http://blog.csdn.net/watson243671)</div><div>2011-11-22 16:16</div><div> 663</div></div>	
<div><div> ahao214 (http://blog.csdn.net/ahao214)</div><div>2013-06-15 20:01</div><div> 890</div></div> <div>Tales of Mystery and Imagination——5、The Tell-Tale Heart (/codinglab/article/details/9102871)</div> <div>It is true that I had been-and I ma-very nervous,but do you really think that I am mad?I could see a...</div>	
<div><div> qibofang (http://blog.csdn.net/qibofang)</div><div>2016-05-23 19:09</div><div> 4217</div></div> <div>基于《Combining Sketch and Tone for Pencil Drawing Production》的图像铅笔画算法的实现 (/qibofang/article/details/51482431)</div> <div>一, 借鉴: 本文借鉴了CSDN博主风吹夏天对此论文算法的理解: 风吹夏天的图像铅笔画算法, 以及香港中文大学Cewu Lu等人写的该论文的主页。原文作者和博主风吹夏天都给过代码, 但是代码不全。我仔细看了原...</div>	
<div><div> leixiaohua1020 (http://blog.csdn.net/leixiaohua1020)</div><div>2013-11-23 00:11</div><div> 8642</div></div> <div>图像检索: CEDD (Color and Edge Directivity Descriptor) 算法 (/leixiaohua1020/article/details/16883379)</div> <div>颜色和边缘的方向性描述符 (Color and Edge Directivity Descriptor, CEDD) 本文节选自论文《Android手机上图像分类技术的研究》。 CEDD具有抽取...</div>	
<div><div> rocky69 (http://blog.csdn.net/rocky69)</div><div>2012-08-02 10:22</div><div> 2844</div></div> <div>《Computer Vision: Algorithms and Applications》學習筆記 (一) ——圖像旋轉算法與實現 (/rocky_shared_image/article/details/7821665)</div> <div>http://www.cnblogs.com/mlv5/archive/2012/02/02/2336321.html 昨天和今天學習了《Computer Vision: Algor...</div>	

图像重建的迭代算法 **Iterative procedure for image reconstruction and OSEM** (/colddie/article/details/6185621)

除了简单的back projection 和 改进的filtered back projection,用的更多的图像重建方法是迭代法，其中包括EM。下面是M edical imaging 里面用的较多的...

 colddie (<http://blog.csdn.net/colddie>) 2011-02-15 11:47  1423



目录



喜欢



收藏



评论



分享