

机器爱学习

- 专注机器学习、深度学习及其应用

博客园
新随笔
订阅

首页
联系
管理

随笔 - 66 文章 - 0 评论 - 8

昵称：AI-ML-DL
园龄：10个月
粉丝：15
关注：0
+加关注

<	2017年10月						>
日	一	二	三	四	五	六	
24	25	26	27	28	29	30	
1	2	3	4	5	6	7	
8	9	10	11	12	13	14	
15	16	17	18	19	20	21	
22	23	24	25	26	27	28	
29	30	31	1	2	3	4	

搜索

找找看

CV : image caption(Review Networks for Caption Generation)

这里介绍一篇比较新颖的框架，Review Network，听说最近获得了NIPS 2016的关注。

1、引言

本文首先指出传统的（attention）encoder-decoder模型具有两个不足之处：1、传统的attention机制，在每一个时间步长的输出受当时的attention结果，未跟之后的attention机制联系起来，这样每一个生成就不是基于全局来看的，不是最优的；2、之前都工作都证明discriminative supervision是有效的，但不清楚是如何将其应用到端到端的学习机制中，不清楚其工作机制。本文就是基于这两个想法，研究出来Review Network，其在编码器的隐藏层采用带有attention机制的review step，输出一些thought vector，其能通过解码器的attention机制来获取全局属性。总的来说，Review Network的灵感就是review编码器所得的所有信息产生一个更加抽象，全局、紧密的向量，相比之前方法所得到隐藏层状态。

2、模型

以下是传统的encoder-decoder模型和本文所提的review network：

谷歌搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

随笔分类

[CV\(35\)](#)
[DL\(10\)](#)
[ML\(20\)](#)
[NLP\(1\)](#)

随笔档案

[2017年3月 \(5\)](#)
[2017年2月 \(19\)](#)
[2017年1月 \(8\)](#)
[2016年12月 \(23\)](#)
[2016年11月 \(11\)](#)

最新评论

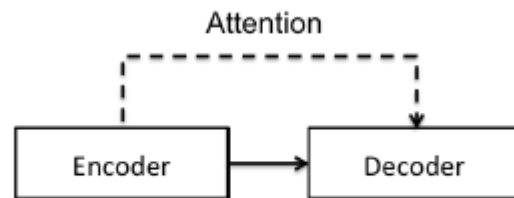
1. Re:生成对抗式网络
详细

--StudyAI_com

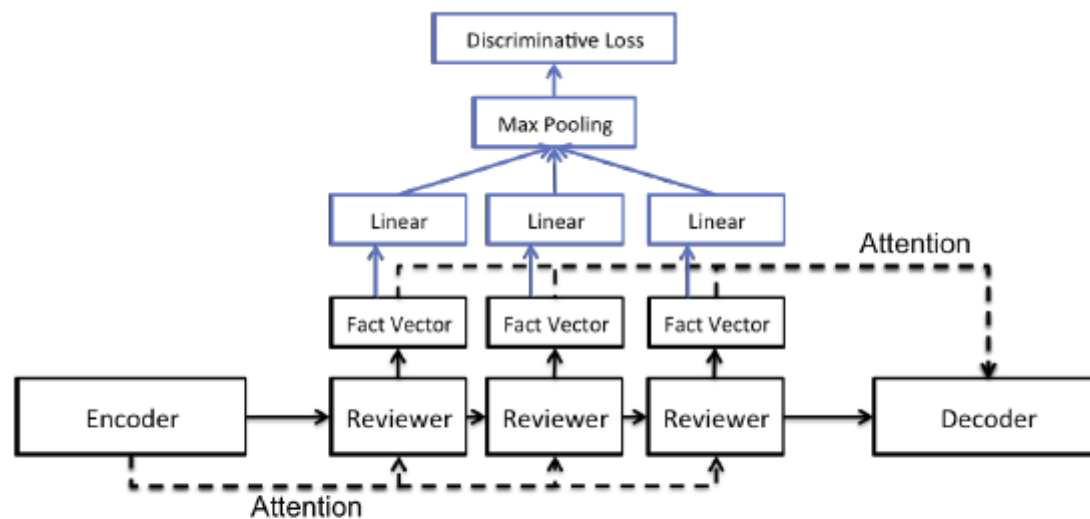
2. Re:CV : object detection(YOLO)
@马春杰杰 可以一起交流...
--flysnow_88

3. Re:CV : object detection(YOLO)
@flysnow_88还没有呢，现在在看SSD了...
--马春杰杰

4. Re:CV : object detection(YOLO)



(a) Attentive Encoder-Decoder Model.



(b) Review Network. Blue components denote optional discriminative supervision. T_r is set to 3 in this example.

可以看到相关的不同点，这里我们详细说一下每一部分的内容。

三个方面说明本文的结构，分别是encoder，reviewer，decoder。其相应的图如下：

@马春杰杰你好：想问下，你更改了源码没？可以输出每一类的recall,AP,以及mAP了吗？我也在做这一步。...

--flysnow_88

5. Re:CV : object detection(YOLO)

@马春杰杰recall和mAP都是分类任务的指标，只是需要针对多标签任务进行一些修改，具体的，百度即可知道...

--AI-ML-DL

阅读排行榜

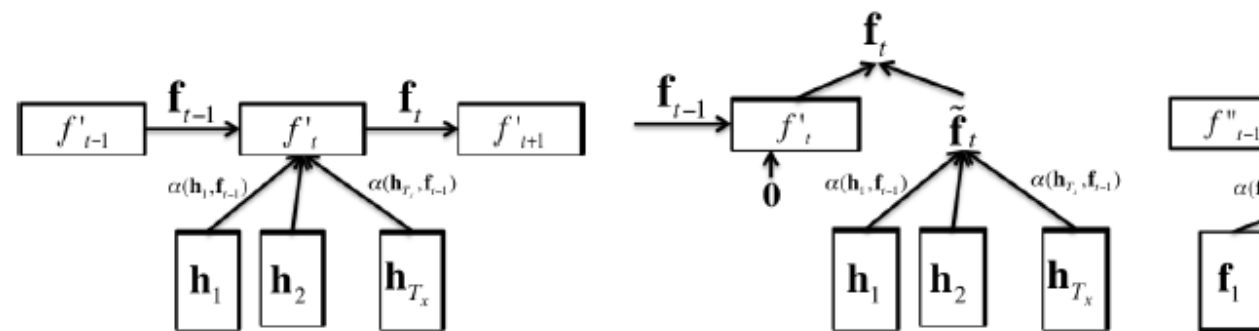
1. LSTM与GRU结构(8605)
2. 聚类算法 (clustering) (3629)
3. CV : object recognition(ZFNet)(3615)
4. 生成对抗式网络(2707)
5. CV : image caption(Show, Attend and Tell: Neural Image Caption Generation with Visual Attention)(1700)

评论排行榜

1. CV : object detection(YOLO)(5)
2. 时间序列分析(1)
3. 聚类算法 (clustering) (1)
4. 生成对抗式网络(1)

推荐排行榜

1. 时间序列分析(2)
2. CV : object recognition(ZFNet)(1)
3. LSTM与GRU结构(1)
4. 聚类算法 (clustering) (1)



(a) Attentive Input Reviewer.

(b) Attentive Output Reviewer.

encoder部分：

讲解了RNN和CNN两种形式，如图所示，该部分并没有什么不同，不做讲解。

reviewer部分：

首先设置review step的数量，用T表示。其实encoder隐藏层状态序列和前一时刻thought vector的函数：

$$\mathbf{f}_t = g_t(H, \mathbf{f}_{t-1})$$

其中，g函数是带有attention的LSTM函数。

attentive input reviewer：首先在隐藏层状态序列上应用attention，然后作为LSTM的输入，总的输入为：

$$\tilde{\mathbf{f}}_t = \text{att}(H, \mathbf{f}_{t-1}) = \sum_{i=1}^{|H|} \frac{\alpha(\mathbf{h}_i, \mathbf{f}_{t-1})}{\sum_{i'=1}^{|H|} \alpha(\mathbf{h}_{i'}, \mathbf{f}_{t-1})} \mathbf{h}_i, \quad g_t(H, \mathbf{f}_{t-1}) = f'_t(\tilde{\mathbf{f}}_t,$$

α 函数决定每个隐藏层的权重，既可以是点乘也可以是MLP。

attentive output reviewer：用0向量作为LSTM的输入，thought vector由attention加权结果和LSTM的输出得来，总结如下：

$$\tilde{\mathbf{f}}_t = \text{att}(H, \mathbf{f}_{t-1}), \quad g_t(H, \mathbf{f}_{t-1}) = f'_t(0, \mathbf{f}_{t-1}) + \mathbf{W}\tilde{\mathbf{f}}_t$$

attention机制用在RNN的最上面一层。

weight tying : 学习两种weight tying , wt表示LSTM的参数 , 第一种是传统的RNN , 第二种则不一样。

decoder部分 :

用F表示thought vector序列 , St是LSTM的第t个隐藏层状态 , decoder的公式如下 :

$$\tilde{s}_t = \text{att}(F, s_{t-1}), \quad s_t = f''([\tilde{s}_t; y_{t-1}], s_{t-1}), \quad y_t = \arg \max_y \text{softmax}(s_t)$$

Discriminative Supervision :

其就是模型的蓝色部分 , 可知其首先利用线性过程计算每个thought vector的得分 , 然后应用一个最大池化 , 提取大部分信息 , 此处增加一个多标签边缘损失函数作为Discriminative Supervision , Discriminative loss如下 :

$$\mathcal{L}_d = \frac{1}{Z} \sum_{j \in W} \sum_{i \neq j} \max(0, 1 - (s_j - s_i))$$

在训练时 , 总的损失函数为 :

$$\mathcal{L}(x, y) = \frac{1}{T_y} \sum_{t=1}^{T_y} -\log \text{softmax}_{y_t}(s_t) + \lambda \mathcal{L}_d$$

3、实验

本文做了两个实验 , 一个是image caption , 一个是code caption。相关结果如下 :

Table 1: Comparison of model variants on MSCOCO dataset. Results are obtained using VGGNet. Scores in the brackets are without beam search. We use RNN-like tied weights unless otherwise indicated. “Disc Sup” means discriminative supervision.

Model	BLEU-4	METEOR
Attentive Encoder-Decoder	0.278 (0.255)	0.229 (0.215)
Review Net	0.282 (0.259)	0.233 (0.218)
Review Net + Disc Sup	0.287 (0.264)	0.238 (0.223)
Review Net + Disc Sup + Untied Weights	0.290 (0.268)	0.237 (0.222)

Table 2: Comparison with state-of-the-art systems on the MSCOCO evaluation models. *Feat.* means using task-specific features or attributes. *Fine.* means using fine-tuned models.

Model	BLEU-4	METEOR	ROUGE-L
Attention [20]	0.537	0.322	0.654
MS Research [5]	0.567	0.331	0.662
Google NIC [18] [†]	0.587	0.346	0.682
Semantic Attention [21] [†]	0.599	0.335	0.682
Review Net (this paper) [†]	0.597	0.347	0.686

Table 3: Comparison of model variants on HabeasCorpus code captioning dataset. “LLH” refers to log-likelihood, “CS- k ” refers to top- k code snippet accuracy.

Model	LLH	CS-1	CS-2	CS
Language Model	-5.34	0.2340	0.2763	0.3
Encoder-Decoder	-5.25	0.2535	0.2976	0.3
Encoder-Decoder (Bidir)	-5.19	0.2632	0.3068	0.3
Attentive Encoder-Decoder (Bidir)	-5.14	0.2716	0.3152	0.3
Review Net	-5.06	0.2889	0.3361	0.3

4、总结

看出review network是一种有效的结构，不仅在本文所提出的两个任务上，其在encoder-decoder框架中，都能起到不同程度的效果，所以值得研究和扩展。

分类: [CV](#)

好文要顶

关注我

收藏该文



AI-ML-DL

关注 - 0

粉丝 - 15

+加关注

0

0

« 上一篇: [Aprior算法、FP Growth算法](#)

» 下一篇: [CV : object recognition\(LeNet\)](#)

posted @ 2016-12-12 16:15 AI-ML-DL 阅读(175) 评论(0) 编辑 收藏

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库

【推荐】报表开发有捷径：快速设计轻松集成，数据可视化和交互



最新IT新闻:

- Win10市场份额继续攀升：微软要把Win7赶出局
 - iPhone X砍掉128GB版 苹果每周多赚39亿
 - 《守望先锋》总监长文扎心：求玩家对我们温柔一点
 - 苹果公开感谢腾讯发现iOS漏洞：曝出不为人知的秘密
 - 盖茨切换到Android暗示Surface Phone可能无见光之日
- » 更多新闻...



最新知识库文章:

- 实用VPC虚拟私有云设计原则
 - 如何阅读计算机科学类的书
 - Google 及其云智慧
 - 做到这一点，你也可以成为优秀的程序员
 - 写给立志做码农的大学生
- » 更多知识库文章...