

学界 | 细致卷积

2017-11-17 机器海岸线

选自 arXiv

作者: Wenpeng Yin, Hinrich Schütze 等

机器海岸线编译

参与: 方建勇

Attentive Convolution

Wenpeng Yin, Hinrich Schütze
Center for Information and Language Processing
LMU Munich, Germany
wenpeng@cis.lmu.de

论文链接: <https://arxiv.org/pdf/1710.00519>

摘要: 在 NLP 中, 卷积神经网络(CNN)受益于来自关注机制的递归神经网络(RNN)。我们假设这是因为 CNN 中的注意力主要是作为积极的汇集(即应用于汇集)而不是作为注意力的卷积(即它被整合到卷积中)来实现。卷积是 CNN 的区别, 它可以通过考虑输入文本 t_x 中的局部固定大小的上下文来对单词的更高级表示进行功能完全建模。在这项工作中, 我们提出了一个细致的卷积网络 Att-entiveConvNet。它扩展了卷积运算的上下文范围, 不仅从本地上下文中获得一个词的高级特征, 而且还从由 RNN 中常用的注意机制提取出的非本地上下文信息中获得更高级的特征。这个非本地的上下文可以来自 (i) 来自远处的输入文本 t_x 的部分或者 (ii) 来自第二输入文本的上下文文本 t_y 。在关于句子关系分类(文本包含和回答句子选择)和文本分类的评估中, 实验证明 AttentiveConvNet 具有最先进的性能, 并优于 RNN / CNN 变种。所有的代码将被公开发布。

textual entailment example	
premise	three bikers stop in town
hyp ₁	a group of bikers are in the street
hyp ₂	the bikers did n't stop in the town
sentiment analysis example	
With the 2017 NBA All-Star game in the books I think we can all agree that this was definitely one to <i>remember</i> . Not because of the three-point shootout, the dunk contest, or the game itself but because of the <i>ludicrous</i> trade that occurred after the festivities.	

表 1: inter-和 intratext 注意的例子。

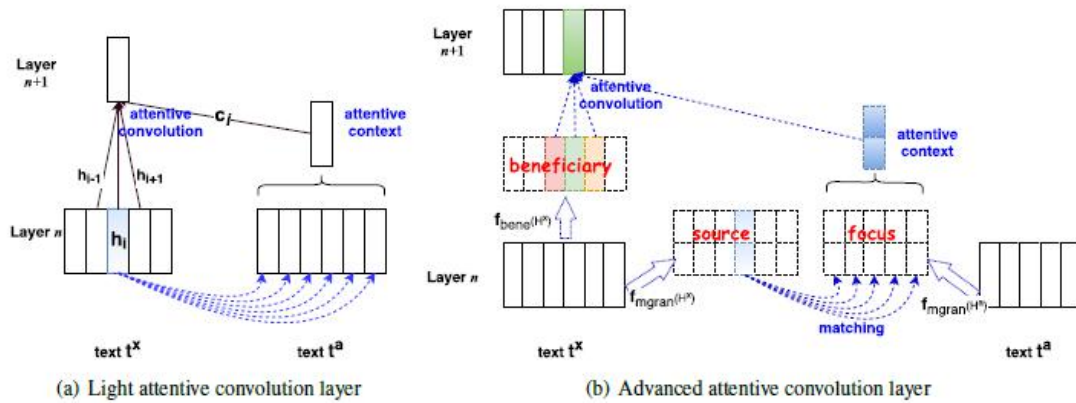


图 1: 细致的卷积层。

role	text
premise	Three firefighter come out of subway station
hypothesis	Three firefighters putting out a fire inside of a subway station

表 2: 文本包含所需的多粒度对齐。

	Systems	acc
w/o attention	bi-CNN	80.3
	bi-LSTM (Bowman et al., 2015)	77.6
	Tree-CNN (Mou et al., 2016)	82.1
	NES (Munkhdalai and Yu, 2017)	84.8
with attention	W-by-W attention (Rocktäschel)	83.5
	Self-Attentive (Lin et al., 2017)	84.4
	Match-LSTM (Wang & Jiang)	86.1
	Decompose Attention (Parikh)	86.8
	<i>LSTMN (Cheng et al., 2016)</i>	89.1
	ABCNN (Yin et al., 2016)	83.7
	APCNN (dos Santos et al., 2016)	83.9
Attentive ConvNet	light	86.3
	advanced	87.8
	advanced&intra-attention	88.4
	ensemble	89.3

表3: SNLI 测试作为隐藏层输入的性能比较1。

Systems	acc
AttentiveConvNet (advanced+intra-attention)	88.7
w/o tri-hidden	88.4
w/o uni-hidden	87.3
w/o gate	88.2

表4: 在 SNLI 开发中的切除测试。

	Systems	p@1
w/o attention	WordC1	68.87
	WordC2	79.59
	CDSSM (Shen et al., 2014)	67.21
	bi-CNN	77.94
	Bi-GRU (Tang et al., 2017)	78.14
with attention	<i>Sentence-Rank</i>	86.00
	ABCNN (Yin et al., 2016)	84.72
	APCNN (dos Santos et al., 2016)	84.14
Attentive ConvNet	light	87.42*
	advanced	88.27*
	advanced&intra-att	88.54*

表 5: SQUAD-AnSS 的系统比较。 现有技术的显著改进标有（等比例检验， $p < 0.05$ ）。

	Systems	acc
w/o attention	Paragraph Vector	58.43
	Lin et al. BiLSTM	61.99
	Lin et al. CNN	62.05
	<i>MultichannelCNN (Kim)</i>	64.62
with attention	CNN+internal attention	61.43
	Lin et al. RNN Self-Att.	64.21
Attentive ConvNet	light	66.75
	advanced	67.36*

表 6: Yelp 的系统比较。 相对于现有技术的显著改进标记为（等比例的检验， $p < 0.05$ ）。

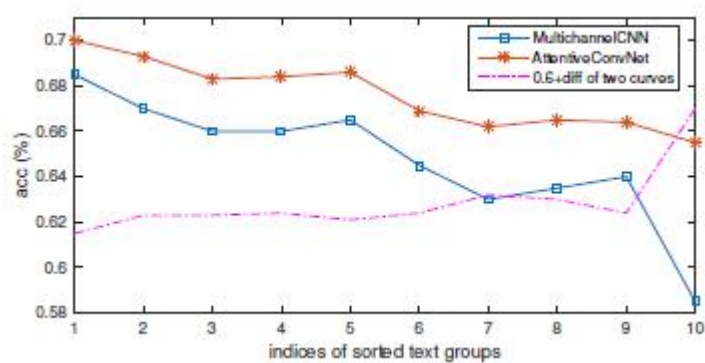


图 2: AttentiveConvNet vs. MultichannelCNN.

本文为机器海岸线编译，转载请联系 fangjianyong@zuu.zju.edu.cn 获得授权。

✂️ -----