

文本分类问题不需要ResNet? 小夕解析DPCNN设计原理 (上)

原创 夕小瑶 夕小瑶的卖萌屋 2018-04-04

来自专辑

卖萌屋@自然语言处理

>

历史回顾

回顾一下图像和文本的发展史，似乎这就是一场你追我赶的游戏。在上一阶段的斗争中，朴素贝叶斯、最大熵、条件随机场这些理论完备的统计机器学习模型使得文本分类、中文分词、NER等诸多自然语言处理问题取得了差强人意（释义：基本使人满意。顺带嘲讽一波误用该词的媒体们(¬_¬)）的性能，而这些理论完备的模型却在图像分类这种基本的计算机视觉问题上都严重碰壁。

如今深度学习，或者具体点说卷积神经网络（CNN），为图像分类、目标检测等计算机视觉问题带来了蜕变式发展。从LeNet、AlexNet等，到Resnet、DenseNet，深层CNN一次次的刷新图像分类的极限，直至超越人类。但是在文本分类这个简单的问题上却迟迟未能等到深层CNN带来的惊喜。甚至都已经2017年了，依然有paper在讨论文本分类问题有没有必要用深度CNN网络的问题，并且最后的主要结论是没啥必要，一层就够了[1][2]。

猜想

出现这个现象的原因有三种，一种是文本分类问题真的已经做的足够好了（喂喂，真的要这样欺骗记几吗），一种是深度CNN网络确实不适合文本分类问题（emmmm这就比较忧伤了），还有一种是NLP领域还没有出现何恺明男神研究出一种适合学习自然语言文本的深度CNN网络。

其实第二种可能性确实让人比较沮丧，一旦有人说出来，可能就会出现恶性循环，直到有人做出有效的模型才能结束这个sad story。而冷静看的话，第三种则是很有可能的，最优化理论里有个经典的no-free-lunch (NFL) 定理，暗示我们不要试图找一个解决所有AI问题的通用模型，而要根据具体问题来设计具体的算法和模型。而现有的很多paper的研究却貌似在忽略这个问题，从而得出一些不太靠谱的经验性结论，甚至包括一些顶会paper。

no-free-lunch延伸出的公平性比较问题

比如[1]试图说明文本分类问题里，word-level的模型（注：word-level模型是指以词为单位，character-level是指以字母/字符为单位）没必要做深，作者却拿TextCNN这个专门为文本分类问题设计的浅卷积模型与DenseNet这个为图像分类问题设计的深卷积模型做比较，发现DenseNet不如TextCNN于是结论说文本分类问题没什么必要用深层模型的呢，费力不讨好的呢。可是亲爱的，你可知道no-free-lunch？你可知道一个模型不针对数据集精调超参可有多大性能损失？

同样的，[2]也是做了一个没有什么意义的尝试，其作者精心为文本分类问题设计了一个character-level的深度卷积网，包括为其精调参数，却直接将设计完成的网络用在word-level的问题上，然后结果发现word-level的模型性能很差，显得好像word-level的模型是真的不能用深度CNN的哦，如果要用，请切换为char-level哦。看似让word-level和char-level保持同样的网络结构、层数和超参数是为了单一变量，实则恰好无视了no-free-lunch，私以为这样的比较结果是没有什么意义的。当然，这篇文章带来的数据资源以及其他贡献还是值得肯定的。

在no-free-lunch衍伸出的公平比较问题上，[3]是一篇做的不错的文章。其通过多维度比较word-level的精心设计的TextCNN模型与char-level的精心设计的VDCNN模型，从而相对谨慎的得出char-level的模型确实没有word-level模型高效的结论。表1是测试集性能比较。

(b) Error rates (%)

Models	depth	AG	Sogou	Dbpedia	Yelp.p	Yelp.f	Yahoo	Ama.f	Ama.p
Linear model <i>best</i> [9]	0	7.64	2.81	1.31	4.36	40.14	28.96	44.74	7.98
char-CNN <i>best</i> [1]	9+2	9.17	3.58	1.35	4.88	36.73	27.60	37.95	4.70
	29+2	8.67	3.18	1.29	4.28	35.28	26.57	37.00	4.28
word-CNN w/o tv-embed.	1	6.95	2.21	1.12	3.44	34.21	26.06	37.51	4.27
word-CNN w/ tv (300-dim)	2	6.57	1.89	0.84	2.90	32.39	24.85	36.24	3.79

表1

表2是两种模型的空间复杂度与时间复杂度的比较。

		Depth	Dimensionality of layer outputs (#layers)	#param	Time	Error rate(%)
char-CNN [1]		9+2	16(1), 64(3), 128(2), 256(2), 512(2), 2048(2)	2.2M	†215	36.73
		29+2	16(1), 64(11), 128(10), 256(4), 512(4), 2048(2)	4.6M	‡700	35.28
word-CNN	w/o tv-embed.	1	500(1)	45M	6	34.21
	w/ 2 tv (100-dim)	2	100(2), 500(1)	68M	21	32.77
	w/ 4 tv (100-dim)	2	100(4), 500(1)	91M	36	32.55
	w/ 4 tv (300-dim)	2	300(4), 500(1)	184M	72	32.39

表2

由此各自的优劣就非常明显了。

那话说回来，**文本分类问题中到底存不存在有效的深层word-level模型呢？**

最起码小夕在前面指出了，前人的贬低深层word-level模型的依据都是片面、不可靠的，那么说明文首提到的第二种可能性依然没有成立，这时第三种可能性的机会来了。

2017年的ACL上有这么一篇文章，**Deep Pyramid Convolutional Neural Networks for Text Categorization [5]**，论文中提出的DPCNN是严格意义上第一篇word-level的对文本分类有效的深层卷积网络。这篇文章貌似目前并没有引起广泛关注，然而其性能提升却撩起了小夕的浓厚兴趣。

小夕曾经做过相关研究，当时做实验时首先用的是1层CNN1层全连接的word-level模型，也就是2014Kim[4]的这个俗称的TextCNN模型，如图1

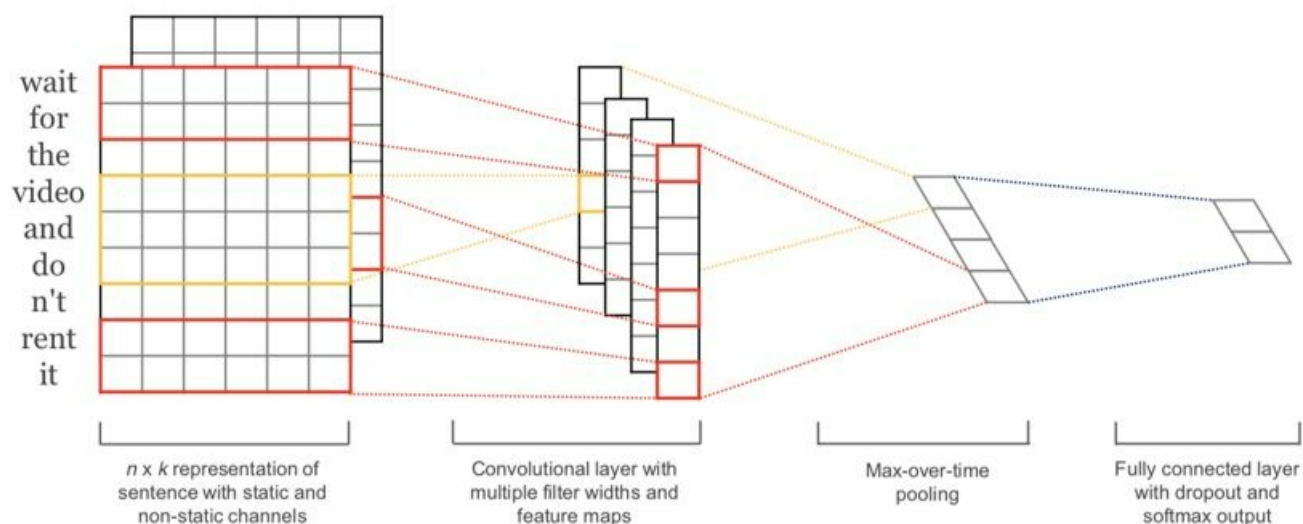


图1

小夕发现TextCNN果然在几个大规模文本分类数据集上取得了不错的效果。然而，小夕多次尝试加深CNN却发现模型性能总是收效甚微甚至不升反降。当时小夕似乎都要信了[1]的结论，感觉word-level的模型在文本分类问题上太难做深了。虽然觉得不能像[1]那样直接把DenseNet拿来用，然而并没有探索出太有效的深层结构。所以认认真真研究了一下这篇文章，并且做了一小部分复现，发现确实是有必要好好体会和总结一下这个网络的设计思想的。

DPCNN一方面在冷启动和梯度消失问题上借鉴了ResNet中的Highway/Shortcut结构，另一方面让DPCNN在冷启动阶段去等效为TextCNN，使得其理论上不至于比TextCNN更差，另一方面就是一个巧妙的设计trick将各个block的embedding约束到相似（甚至可能相同）的空间（毕竟文本里的embedding是比较“扁平”的，很可能三四个词描述的语义完全可以用一个词替代）。下面详细介绍这三点。

首先丢上(a)DPCNN与(b)TextCNN（文中称为ShallowCNN）、(c)ResNet的对比图，如图2。

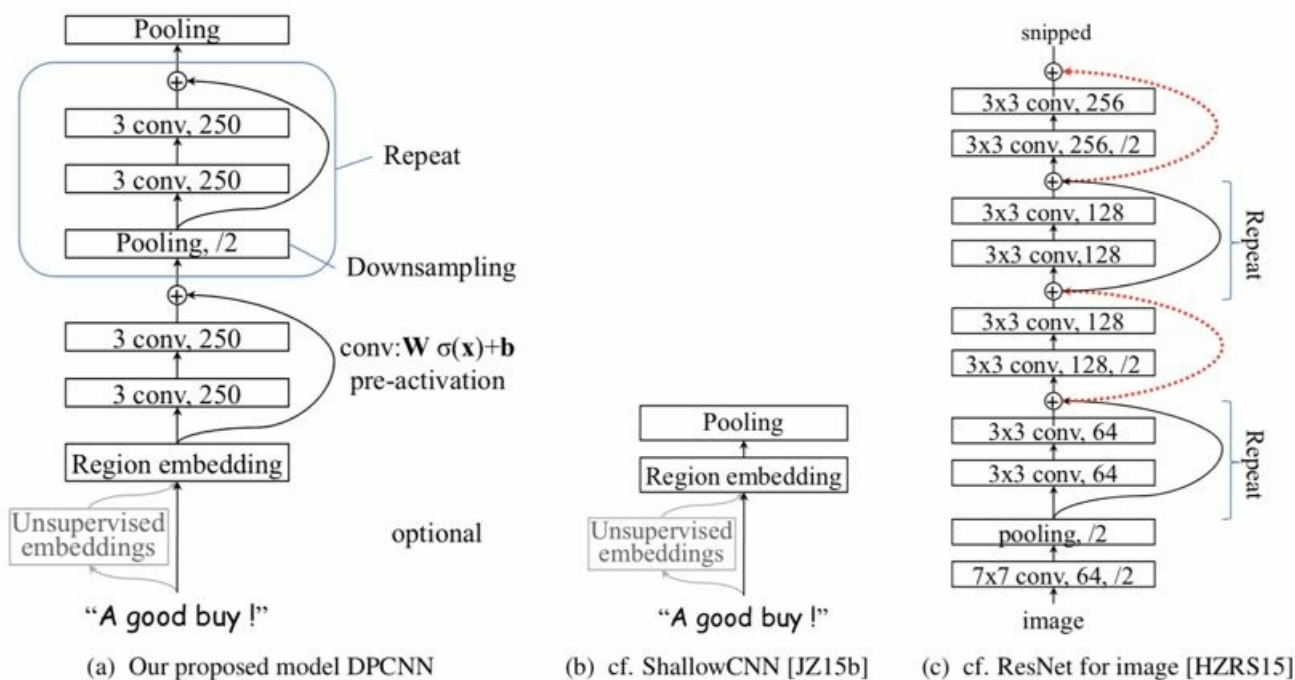


Figure 1: (a) Our proposed model DPCNN. (b,c) Previous models for comparison. \oplus indicates addition. The dotted red shortcuts in (c) perform dimension matching. DPCNN is dimension-matching free.

图2

呜，好困哇，坚持了一下还是没写完。迷迷糊糊的发现后半部分写的有点乱，还是从中间截断，先发引人入胜又戛然而止的前半部分吧~明天补完尾巴再把后半部分重新整理一下再推送哦，啾咪啾咪~

参考文献

- [1] Le H T, Cerisara C, Denis A. Do Convolutional Networks need to be Deep for Text Classification?[J]. arXiv preprint arXiv:1707.04108, 2017. [2] Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification[C]//Advances in neural information processing systems. 2015: 649-657.[3] Rie Johnson and Tong Zhang. 2016. Convolutional neural networks for text categorization: Shallow word-level vs. deep character-level.arXiv:1609.00718.[4] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.[5] Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017, 1: 562-570.

蟹蟹你o(≧v≦)o



微信支付



Transfer to 夕小瑶

声明：pdf仅供学习使用，一切版权归原创公众号所有；建议持续关注原创公众号获取最新文章，学习愉快！