

论文阅读笔记

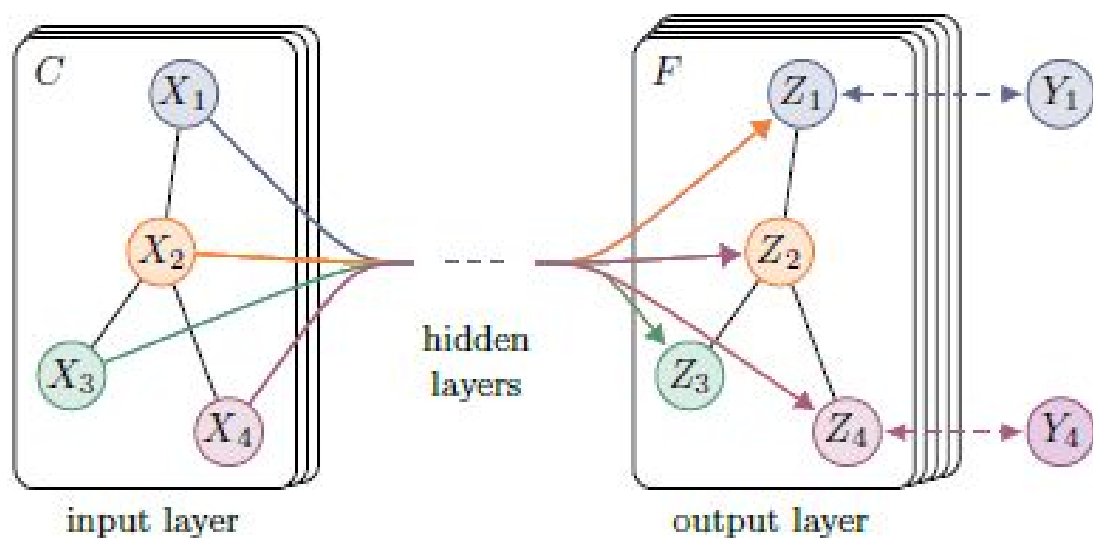
Step8

MF1833063, 史鹏, spwannasing@gmail.com

2019 年 8 月 17 日

1 SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS

本文提出了一种图卷积网络（graph convolutional networks, GCNs），该网络是传统卷积算法在图结构数据上的一个变体，可以直接用于处理图结构数据。从本质上讲，GCN 是谱图卷积（spectral graph convolution）的局部一阶近似（localized first-order approximation）。GCN的另一个特点在于其模型规模会随图中边的数量的增长而线性增长。总的来说，GCN 可以用于对局部图结构与节点特征进行编码。



图卷积神经网络的（单层）最终形式： $H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$

2 Semantic-Unit-Based Dilated Convolution for Multi-Label Text Classification

本文是基于seq2seq的多标签文本分类在attention上的一些改进工作。主要贡献有两点：

- 1.提出了所谓的“语义单元”，因为在多标签文本分类中，word-level的作用没有那么大，而是“semantic units”来决定文本的分类。
- 2.使用了Hybrid Attention将semantic units和word level的attention混合起来。

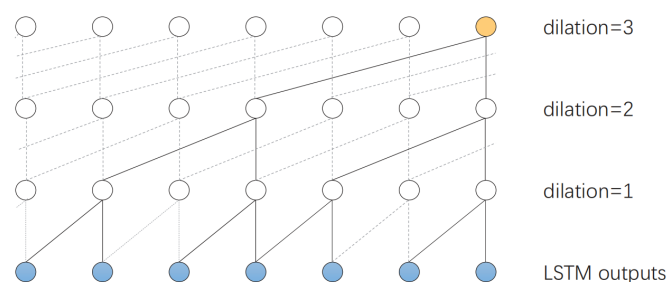


Figure 1: **Structure of Multi-level Dilated Convolution (MDC).** A example of MDC with kernel size $k = 2$ and dilation rates $[1, 2, 3]$. To avoid gridding effects, the dilation rates do not share a common factor other than 1.

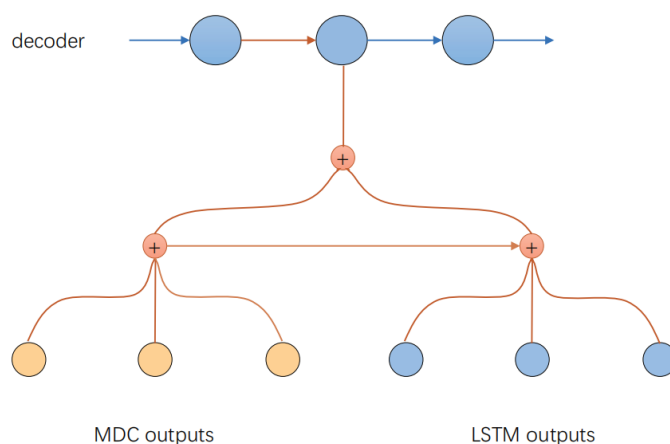


图 1: Structure of Hybrid Attention.

3 A Deep Reinforced Sequence-to-Set Model for Multi-Label Classification

本文的主要Motivation是解决多标签文本分类中Seq2Seq模型的输出序列的顺序问题，因为label之间本来应该是无序的，即交换不变性，这里所做的工作则是提出用强化学习来解决label之间的顺序问题，reward即为预测label序列和答案之间的F1值。

$$\mathcal{L}(\theta) = -\mathbb{E}_{\mathbf{y} \sim p_\theta} [r(\mathbf{y})] \quad (3.1)$$

$$\nabla_\theta \mathcal{L}(\theta) \approx -[r(\mathbf{y}^s) - r(\mathbf{y}^g)] \nabla_\theta \log(p_\theta(\mathbf{y}^s)) \quad (3.2)$$

$$r(\mathbf{y}) = F_1(\mathbf{y}, \mathbf{y}^*) \quad (3.3)$$

其它的结构基本一致。

4 Compositional Questions Do Not Necessitate Multi-hop Reasoning

这篇文章主要是提出了在HotPotQA数据集中存在的一个问题：所谓的multi-hop其实不是必要的，大多数问题可以在只提供single paragraph的情况下就正确的回答出来。然后提出了一个Single-Paragraph QA模型。

分别将paragraph送入Bert

$$S' = \text{BERT}(S) \in \mathbb{R}^{h \times (m+n+1)} \quad (4.1)$$

然后选取 y_{empty} 最小的作为答案输出。

$$[y_{\text{span}} ; y_{\text{yes}} ; y_{\text{no}} ; y_{\text{empty}}] = W_1 \text{ maxpool } (S') \quad (4.2)$$

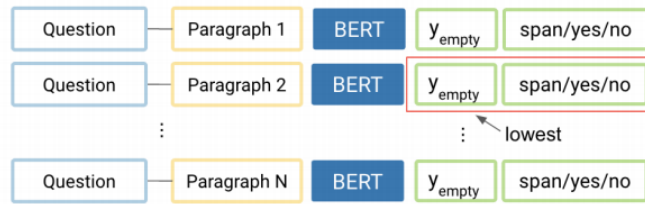


Figure 2: Our model, single-paragraph BERT, reads and scores each paragraph independently. The answer from the paragraph with the lowest y_{empty} score is chosen as the final answer.