

1 任务背景

我近期在做一个句子分类的任务，做法是使用了各类循环神经网络（RNN）对句子的词进行特征抽取然后进行分类。如果用集合描述这个过程即句子词向量集合 $E = \{x_i, (0 \leq i \leq n)\}$ 通过RNN进行特征编码产生高级特征集合 $H = \{h_i, (0 \leq i \leq n)\}$ ，之后通过池化产生句子的特征表示 V_{sent} ，最后交给分类器判定句子属于某一类。

2 关于RNN

首先来看一个简单的RNN对特征进行编码的公式：

$$h_i = \tanh(\mathbf{W} \cdot h_{i-1} + \mathbf{U} \cdot x_i + \mathbf{b})$$

从公式上看，如果RNN对句子的特征编码方向不同，所产生的高级特征H集合也不同。我们可以从句子开头到句子结尾进行特征编码，也可逆着来。若我们只用一个方向上的RNN特征，那没有什么问题，直接就可以进行池化，进一步编码成为一个句子向量表示。但是在使用双向的RNN特征的情况下，其特征就需要结合后才能池化。结合方式就有下文的两种可供选择。

3 两种特征结合方式

对于双向RNN编码出来的两个高级特征集合 $H_L = \{h_{li}, (0 \leq i \leq n)\}$ 和 $H_R = \{h_{ri}, (0 \leq i \leq n)\}$ 有两种结合方式产生新的特征集合 $H' = \{h'_i, (0 \leq i \leq n)\}$ 。

其一：我们可以采用特征直接拼接的方式，这里没有模型参数。

$$h'_i = h_{li} \oplus h_{ri}$$

其二：两种特征集合经过一个隐层，产生一个新的特征的集合。 W_1, W_2, b 是模型参数， $\tanh()$ 是双曲正切激活函数。

$$h'_i = \tanh(\mathbf{W}_1 \cdot h_{li} + \mathbf{W}_2 \cdot h_{ri} + \mathbf{b})$$

4 特征结合方式的分析

两种特征的结合结果从本质上来讲，还是有很大的差别。那么哪种结合方式更为合适呢？首先说第一种结合方式，实际上它彻底保留了双向RNN的特征，这个网络如果和使用单向RNN的神经网络比较，我们可以发现在两者在句子向量 V_{sent} 的表示上，使用双向RNN比单向RNN的多出了一个方向上的特征，另一个方向上的特征表示是一模一样的。那么使用双向RNN的最坏情况就是有一个方向的特征完全是噪声，这样在模型训练的时候，全是噪声的模型参数 W 可以通过学习全置为0，抹去这部分特征。所以从理论上来说，这样的结合方式的双向RNN最坏情况的性能和单向RNN完全一致。

另一种结合方式，从本质上来讲它通过两类特征产生了新的特征。因为不但有新的模型参数作为特征权重，而且还使用了一个非线性变化函数，这个过

程很复杂，我们目前不能从直观理解角度来看这个过程到底发生了什么。我们还是和原先的单向RNN神经网络进行比较，这样就可以发现两者的句子的特征 V_{sent} 完全不同。至于这种方式结合出来的句子特征 V_{sent} 性能上没有保证，是有可能低于单向RNN的。

5 实验实例和结论

在我语料测试集上的实验结果具体见Table 1，在实验结果上能看出我使用第二种结合方式的模型性能要低于只用单向特征的模型，使用第一种结合方式的模型性能要高于只用单向RNN特征的模型。所以在使用双向RNN 的时候要慎重考虑它特征的结合方式，建议采用第一种。

	单向F-macro	第一种方式F1-macro	第二种方式F1-macro
RNN	82.2	82.8	81.0
GRNN	83.1	84.2	83.1
LSTM	83.6	84.1	82.7

Table 1: 实验结果

我的语料不公开，诸位可以通过其他实验来观察这种现象。我把代码发布在github上：

<https://github.com/3cQScbrOnly1/NN4REOS>