
Tweet 情感分类

语义计算与知识检索

韩 云

yunhan@pku.edu.cn

April 22, 2018

1 情感分析

文本情感分析是一个重要的研究领域，像 Twitter 这样的社交媒体平台吸引了研究者在语言处理以及政治和社会科学方面的兴趣。任务通常包括检测一段文字是否表达 positive, negative 或 neural 情绪; 情绪可以是一般的，也可以是关于特定主题的。

在本次实验中，我们针对 SemEval-2017 的 Task 4 中的 Subtask A，也就是给定一篇推文，我们需要将其划分为 positive, negative 和 neural 三类。

2 实验方法

情感分析属于 NLP 中一类比较重要的问题，而对于此类问题，简单的神经网络已经不能满足我们的需求，常常需要使用 CNN, RNN, LSTM 等等来处理。

1. LSTM & Full-connected Layer

针对此次试验，使用一层 LSTM 和一个 softmax 全连接层来对推特数据进行分类，具体的网络结构如下：

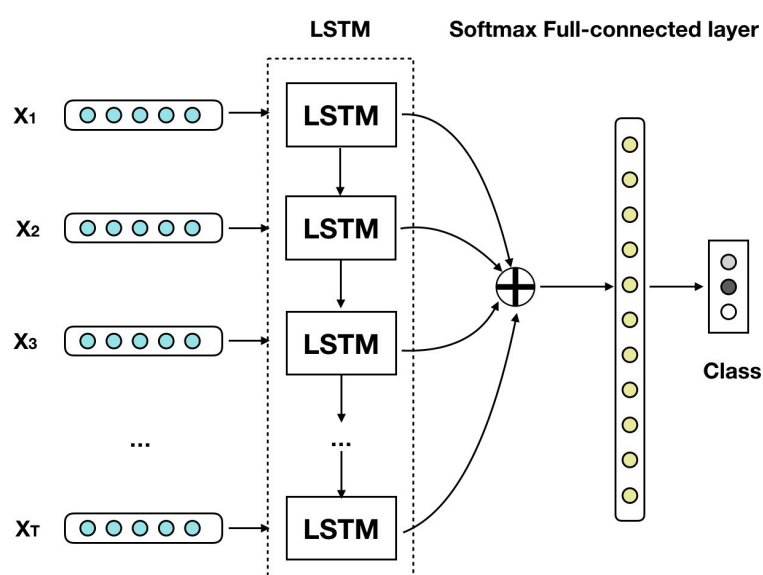


图 1: LSTM & Full-connected Layer

2. Bidirectional LSTM & Full-connected Layer

为了提高实验的性能，将一层 LSTM 改进为一层双向的 LSTM，然后连接一个 softmax 全连接层来对推特数据进行分类，具体的网络结构如下：

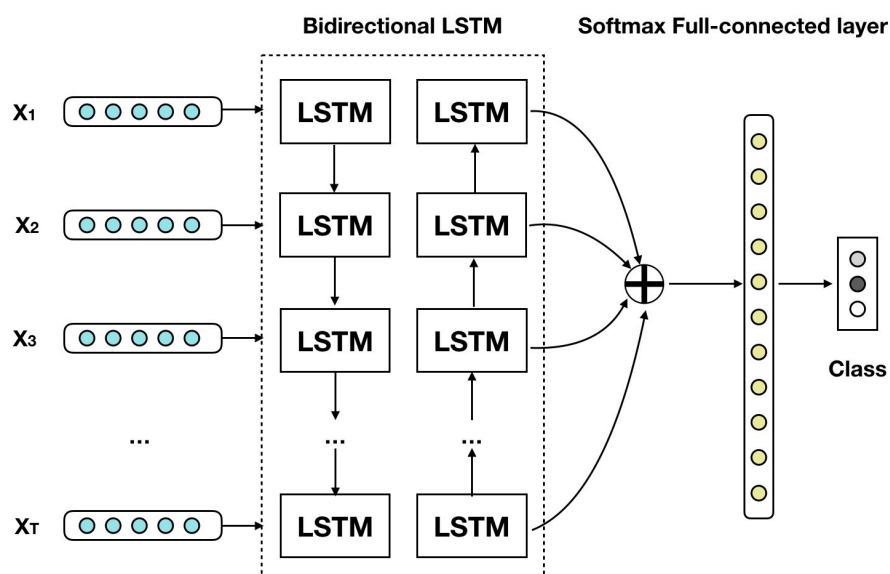


图 2: Bidirectional LSTM & Full-connected Layer

3 实验数据和环境

- 实验数据

SemEval-2017/Task 4/Subtask A/2016 年的推特数据。这些推文是基于国际上流行的流行时事，包括一系列命名实体（例如唐纳德特朗普，iPhone），地缘政治实体（例如阿勒颇，巴勒斯坦）和其他实体（例如叙利亚难民，达科他通道管道，西方媒体，枪支管制和素食主义）。

- 实验环境 (Python3.6)

- GloVe 预训练的 300 维的词向量。
- Nltk 工具包中的 word_tokenize。
- keras 中的数据预处理 to_categorical，模型构建 Sequential, LSTM, Embedding, Dense, Dropout, Bidirectional 等等。

- 评价方法

准确率。

4 实验结果和分析

4.1 实验结果

使用 LSTM 和一个全连接层搭建的网络，执行 50 个 epoch，每个 epoch 的训练集的 loss 和 accuracy，验证集的 loss 和 accuracy 具体的实验结果如下：

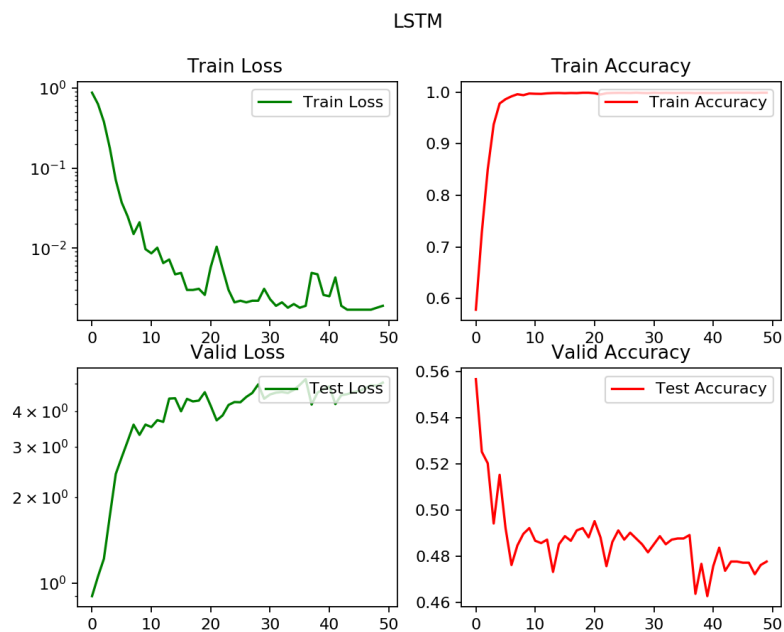


图 3: LSTM 运行结果

使用 Bidirectional LSTM 和一个全连接层搭建的网络，执行 50 个 epoch，每个 epoch 的训练集的 loss 和 accuracy，验证集的 loss 和 accuracy 具体的实验结果如下：

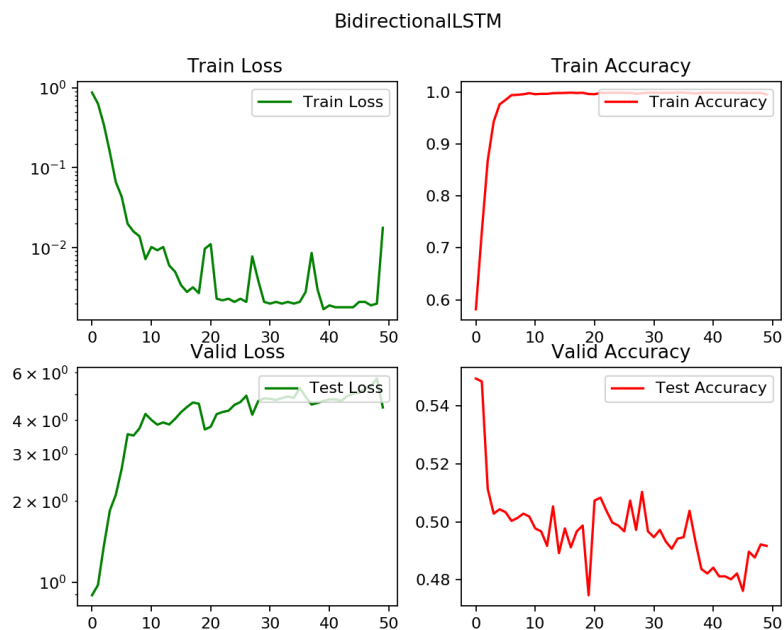


图 4: 使用 Bidirectional LSTM 运行结果

4.2 结果分析

根据 learning curve 我们可以看出，最终模型在验证集上的准确度最优可以是 0.55。

虽然两个模型在训练集的准确度随着 epoch 的增加而逐渐增高，但是在验证集上的准确率都是逐渐降低的，说明对于两个模型都存在 overfitting 的问题。但是本次实验中没有考虑使用正则化等问题对此进行修正，而是直接选取两个模型中对于验证集的最高点来预测测试集上的结果，最终 LSTM 和 Bidirectional LSTM 在测试集的准确度分别为 0.5001 和 0.5106。

5 参考文献

[1] <http://alt.qcri.org/semeval2017/task4/> .

[2] <https://nlp.stanford.edu/projects/glove/> .