# RNN情感分析实验报告

牛远卓

2019302973

2022/04/22

## 任务描述

我的任务目标是采用torch构建一个RNN，用于情感分析。选用的数据集是NLPCC2012情感分析任务，链接为：<http://tcci.ccf.org.cn/conference/2012/pages/page10_dl.html>，使用的是ipad.xml这份数据。采用的词向量为Word2Vec 300维的词向量。

## 数据集介绍

基于word2vec 模型的文本分类任务:文本共包含七个类别,每个类别包含200个文本。

本实验主要基于网上的一个demo（<https://blog.csdn.net/qq_35357274/article/details/120854976>）。我主要是做到看懂这个代码的来龙去脉，以及做了实验结果的可视化，从而进一步分析。

## K-Fold

由于数据量过少（ipad数据集中带有情感倾向的句子仅101条），为了使得训练有效果，这里采用k``折交叉验证的方法来进行训练。

<https://blog.csdn.net/qq_36535820/article/details/119762665>

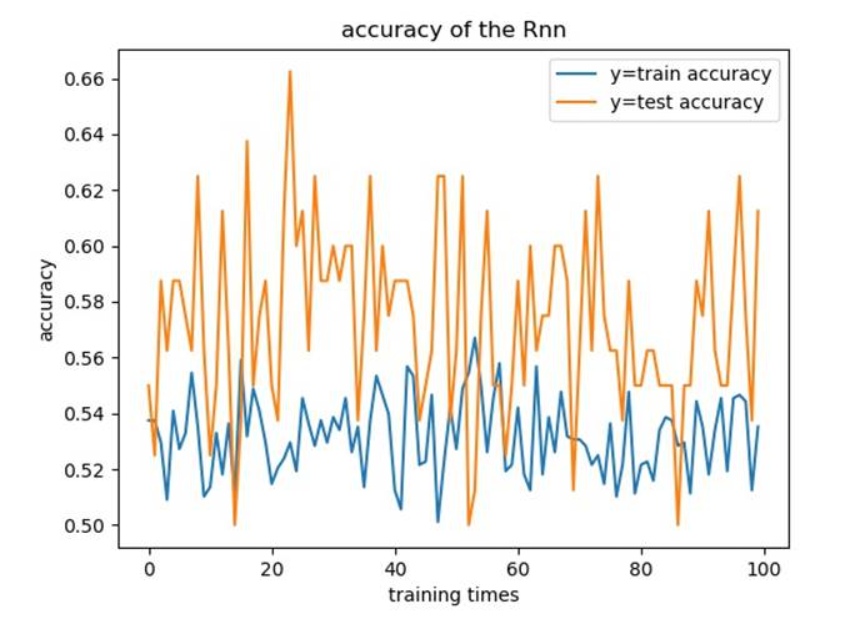
交叉验证既可以解决[数据集](https://so.csdn.net/so/search?q=数据集&spm=1001.2101.3001.7020)的**数据量不够大**问题，也可以解决**参数调优**的问题。

1、首先，将全部样本划分成k个大小相等的样本子集；

2、依次遍历这k个子集，每次把当前子集作为验证集，其余所有样本作为训练集，进行模型的训练和评估；

3、最后把k次评估指标的平均值作为最终的评估指标。在实际实验中，k通常取10.

## 实验结果：



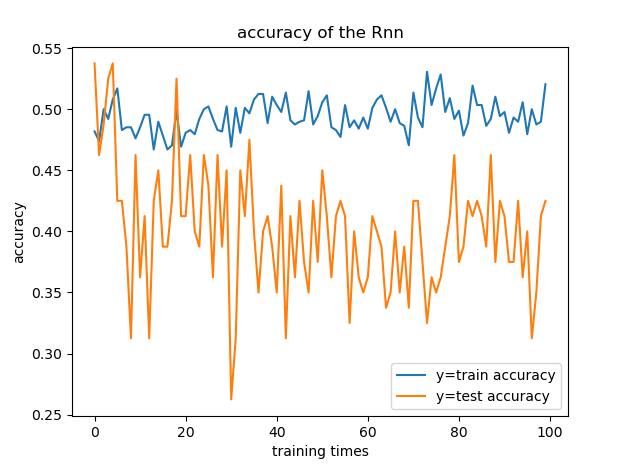
图片中的accuracy是由k折取平均，每个epoch为x轴取得的。

由图可知，训练正确率比测试正确率大部分时候高，但是两者并没有随着epoch深入而增高。

这是原作者的源代码。是单层单向rnn，k取10，epoch取100，max\_seq\_length取30的结果。k是指k\_fold中分几份，max\_seq\_length会在之后讲。这里的实验结果是本次报告的最好了，之后我的更改都是失败的。

## 更改代码

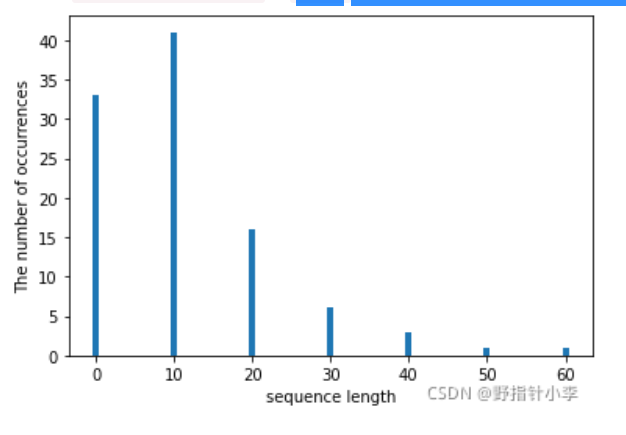
1）会看了代码后，发现有存在没有backward成功的问题。loss = nn.CrossEntropyLoss()中的backward在vscode中本应该显示黄色，但是却显示白色，而且点不进去。于是我成了现在torch版本内有的包torch.functional.cross\_entropy



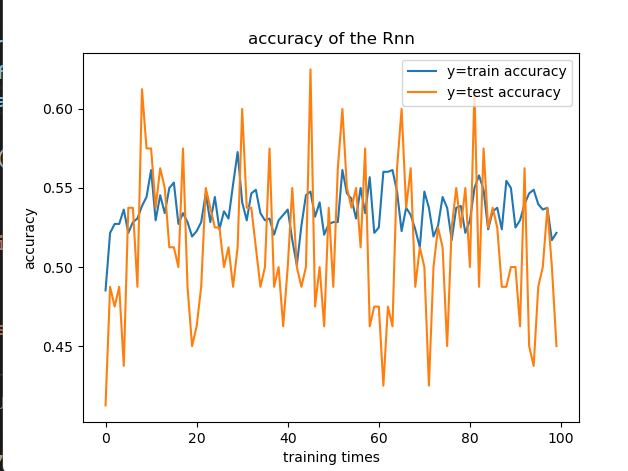
结果是训练情况和测试表现都更加不好，尤其是测试正确率。也就是说还不如不反向传播，这也太奇怪了。哈哈

2）作者说max\_seq\_length的设置可能会影响结果，于是我怕将它调小至27

由于文本长度是不固定的（比如曾经微博要求140字以内），如果我们不确定文本长度，就会导致神经网络无法进行训练（比如RNN中，如果不能确定文本长度，就无法确定sequence length）。同样，我们也不能将最大长度的句子作为文本长度。想象一下，假设最大长度为1000，但是其余句子的长度都在10。如果我们以1000作为文本长度，那么对于其余的句子，剩下的990维都要以0填充。这样不仅增大了内存的消耗，减慢了训练的速度，而且几乎整个语料库中有用的信息都集中在前10个向量中，会导致神经网络不一定学得到东西，或者过拟合。 于是我们选用以上的代码，统计整个语料库中词语数目的区间，用于确定sequence length。ipad语料库中输出的结果如下：



结果



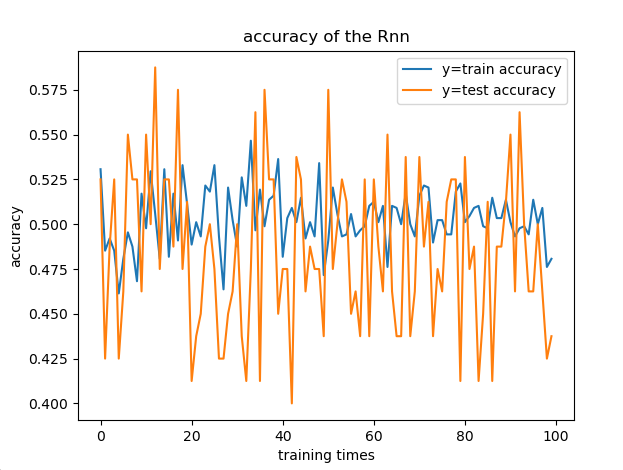
情况有所好转，相比上一个情况尤其是测试正确率回暖了一些。但是依旧没有随着训练次数增多实现正确率上升。原因我想了一下，可能是训练数据太少：ipad数据集中带有情感倾向的句子仅101`条，即使运用了k fold还是不尽如人意。

3）结果还可能与rnn层数和是否双向有关。

以上都是单层单向的

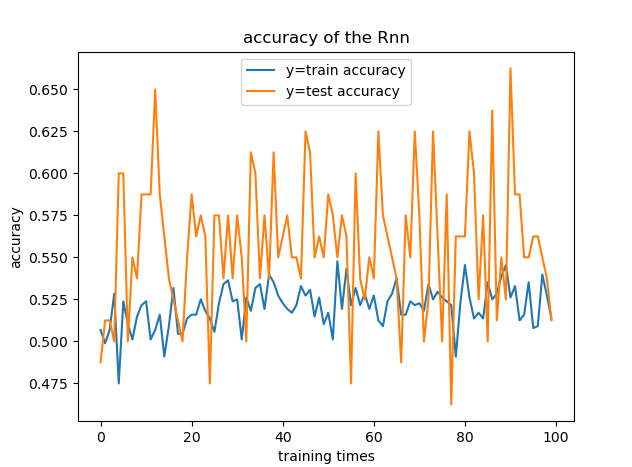
还是以max\_seq\_length=27来算

单层双向：



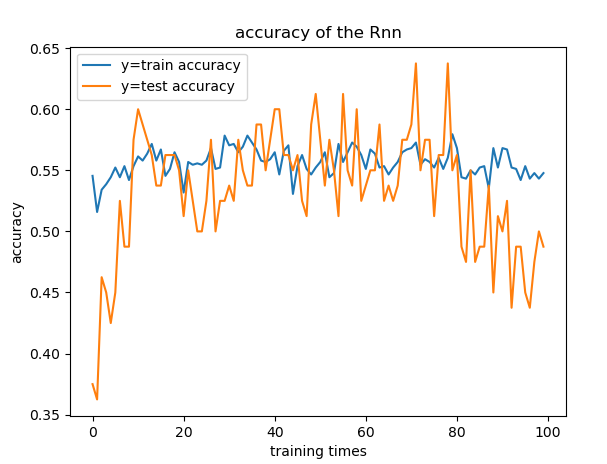
单层双向中训练准确率基本维持在50%，测试正确率位置在47.5%。

双层单向



双层单向中训练准确率基本维持在51%，测试正确率位置在55%。

双层双向



双层双向中训练准确率基本维持在55%，测试正确率位置在53%。

仅由上3图可知，训练效果最好的情况是双向双层，测试效果最好的是双层单向的。

## 实验结论

本次是基于网上一个采用torch构建一个RNN，用于情感分析。选用的数据集是NLPCC2012情感分析任务，使用的是ipad.xml这份数据。采用的词向量为Word2Vec 300维的词向量。我主要是按照每次训练的k个fold的训练结果的平均来对实验结果可视化。实验结果发现正确率大多都不像Alexnet在cifer10上会随着实验次数显著增加，而是在一个区间内起伏。我想大概主要是因为ipad数据集太小，训练完了没啥好的泛化能力。在试着换了一下训练层数、方向以及最大单句词汇数量后，发现还是原作者的模型参数大概是较为合理的。