人工神经网络 HW4 实验报告

计52 路橙 2015010137

###### Part 1：综述

本次实验中，我用TensorFlow实现了基本的RNNCell，其中包含BasicRNNCell、GRUCell与BasicLSTMCell。并通过初步的调参，比较了三者之间的区别，得到了一个较优性能的结果模型。

最终，使用tensorboard绘制了训练过程中的变化。

###### Part 2：不同RNNCell的训练过程

选择相同超参数的三种RNNCell进行训练，共计训练30个epoch，结果如下。

参数选择：

* **Learning Rate:0.001**
* **Units:512**
* **Embed\_units:300**
* **Layers:1**
* **Batch\_size:16**

**其余超参数为网络默认参数。**

1. **Optimizer为GradientDescentOptimizer：**



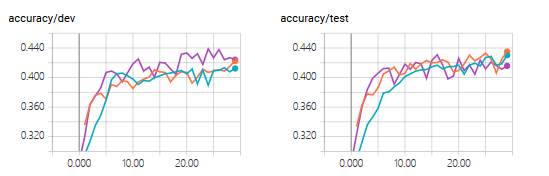


Figure 1.每个epoch网络在dev集与test集的准确率

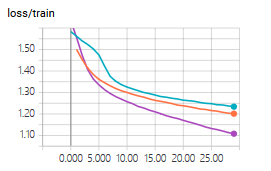
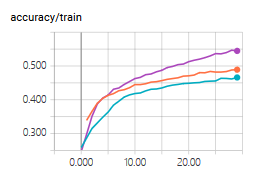


Figure 2.每个epoch网络在训练集上的accuracy与loss

由上图可知，当Optimizer为GradientDescentOptimizer时，三种网络有如下特点：

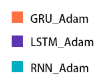
1. BasicRNNCell具有收敛快、性能较好的优点，甚至在dev集上的准确率达到了0.44以上，超过了其余两个网络。这也说明了基础的RNN就可以做到较高的水平，RNN在与序列有关的问题中可以有较优的表现；

但另一方面，RNN的过拟合问题比较严重，当在test集上的准确率已处于徘徊波动状态时，在train集上依然可以观察到loss在稳步快速下降，accuracy在快速上升，说明BasicRNNCell的过拟合问题是三种网络里最明显的。

1. GRUCell的收敛速度最慢，但性能较优，几乎不存在过拟合问题。一方面，由于GRUCell比BasicRNNCell的参数多很多，因而收敛速度慢有一定的道理；而另一方面，在test集最终accuracy的曲线上看，GRUCell依然可以有继续爬升的空间，说明GRUCell的收敛速度慢，但最终可以达到一个较高的准确率水平。
2. BasicLSTMCell收敛速度较快，且过拟合问题不严重，最终性能可以达到很高水平。这也从另一个角度说明，LSTM作为RNN的改进版本，拥有很强的处理序列问题的能力。

一方面，在test集与dev集上，BasicLSTMCell都可以做到收敛速度很快且最终效果较好；另一方面，在train集上，BasicLSTMCell的过拟合程度较轻，loss下降也先快后慢最终趋于平缓。这说明，BasicLSTMCell集成了BasicRNNCell与GRUCell的优点，同时一定程度上规避了两者的缺陷，是处理序列问题中选用RNN模型时的最佳Cell选择。

1. **Optimizer为AdamOptimizer：**



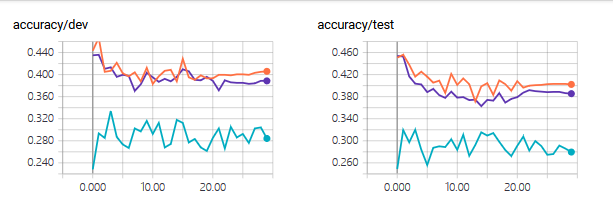


Figure 3. 每个epoch网络在dev集与test集的准确率

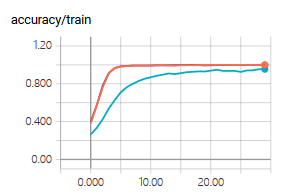
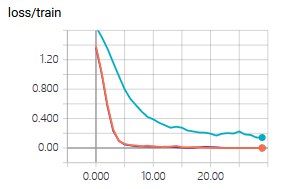
 

Figure 4.每个epoch网络在训练集上的accuracy与loss

由上图可知，当Optimizer为AdamOptimizer时，三种网络都会有严重的过拟合问题（甚至最终可以在train集上达到1.0的准确率，但在test集上的准确率较低）。此外，三种网络还有如下特点：

1. BasicRNNCell性能欠佳，在test集上的准确率处于徘徊状态，难以达到其最优性能，且其在train集上的loss下降也较慢。由于很快产生了过拟合，因此在test集上的性能表现无法再提高，从一开始便决定了性能较差。
2. GRUCell在test集与dev集中都有很高的表现，甚至在test集上可以有超过0.46的准确率。

一方面，AdamOptimizer可以使网络在最初阶段达到很高的test集准确率，从而保存当前网络后便会得到一个极优的性能；另一方面，由于AdamOptimizer的收敛速度较快，使得网络在达到较优结果后会超着向train集收敛的方向迈进，从而使得网络出现严重的过拟合，在test集与dev集上的表现也相应下降，最终稳定的结果并不如GradientDescentOptimizer。

1. BasicLSTMCell也有与GRUCell十分类似的特征，即最初准确率极高，然而随着训练的进行网络会向train集过拟合，导致最终稳定的结果较差。且BasicLSTMCell的结果比GRUCell的结果差一些。
2. **总结**

通过**①**与**②**的比较可知，在RNNCell的训练过程中，若为了追求最高的测试集准确率，则应该选用AdamOptimizer并训练10个以内的epoch即停止；而若为了使训练稳健增长且尽量避免过拟合问题，则应该选用GradientDescentOptimizer。

###### Part 3：BasicLSTMCell的简单调参

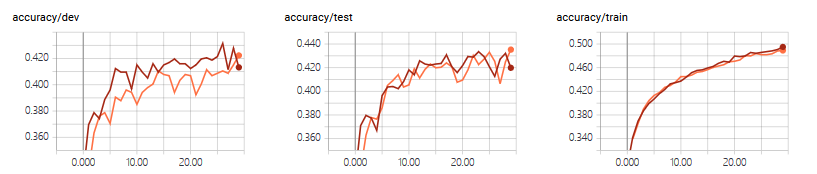
1. **Dropout的选择：**

通过对LSTM网络增加DropoutWrapper来减少过拟合。所用的Optimizer为GradientDescentOptimizer.分别进行keep\_prob为1.0与0.5的对照实验，结果如下图所示。

参数选择：

* **Learning Rate:0.001**
* **Units:512**
* **Embed\_units:300**
* **Layers:1**
* **Batch\_size:16**

**其余超参数为网络默认参数。**



由上图可知，增加dropout后的LSTM网络具有更快的收敛速度、更高的测试集准确率。因此，适当增加dropout的确会使得LSTM网络减小过拟合，使得网络有更稳定、更高效的性能。

1. **Units\_size的选择（讨论在AdamOptimizer下过拟合的主要因素）：**

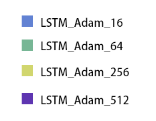
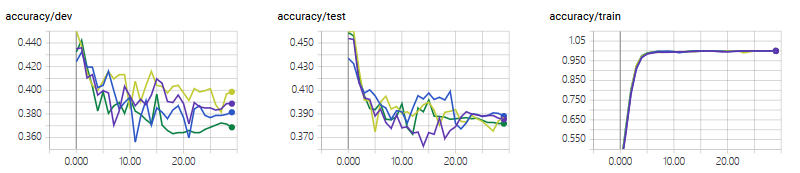
通过对LSTM网络内部gate的units\_size进行调整，分别选取size为512、256、64、16，结果如下图所示。

参数选择：

* **Learning Rate:0.001**
* **Embed\_units:300**
* **Layers:1**
* **Batch\_size:16**

**其余超参数为网络默认参数。**

**选用的Optimizer为**AdamOptimizer.



由上图可知，从宏观而言，units\_size即使为16时也可以得到较好的效果，说明在Optimizer为AdamOptimizer时，网络超参数的数量已经不是过拟合的主要因素，而恰恰是Optimizer的收敛速度导致过拟合的出现。

而当该参数为256时，网络的效果最稳定，甚至可以出现0.46以上的超高测试集准确率。

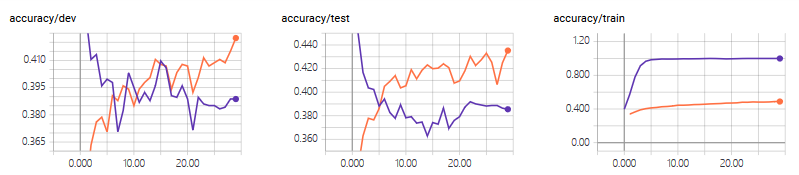
1. **Optimizer的选择：**

分别选取AdamOptimizer与GradientDescentOptimizer，结果如下图所示。

参数选择：

* **Learning Rate:0.001**
* **Units:512**
* **Embed\_units:300**
* **Layers:1**
* **Batch\_size:16**

**其余超参数为网络默认参数。**



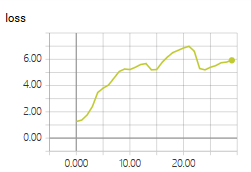
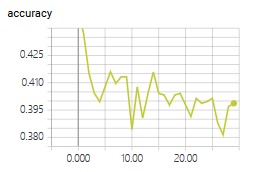
由上图可知，Adam可以让网络在最初时拥有超过的测试集准确率，但随着过拟合的出现而逐渐下降；但GD可以使网络缓慢稳步上升，最终的性能甚至可以达到Adam的性能，且过拟合的程度小很多。因此，可以得出如下结论：

在LSTM的训练过程中，若为了追求最高的测试集准确率，则应该选用AdamOptimizer并训练10个以内的epoch即停止，并选用保存的最佳性能的网络；而若为了使训练稳健增长且尽量避免过拟合问题，则应该选用GradientDescentOptimizer。

###### Part 4：最优网络

通过Part 3的讨论，我最终选取的最优网络有两种，分别对应两种Optimizer.

1. 最优准确率最高：



选用BasicLSTMCell，且Optimizer为AdamOptimizer.由Part 3的讨论知，此时units\_size选择256最佳；而最终网络可以有超过0.46的测试集准确率。

1. 最终准确率最高，且尽可能减少过拟合：



选用BasicLSTMCell，且Optimizer为GradientDescentOptimizer.由Part 3的讨论知，此时需要添加0.5的keep\_prob的dropout；而最终网络可以有超过0.43的测试集准确率。