刘扬 171850524

论文通过提出具有最少门数量的门控单元MGU, 简化RNN复杂的结构, 使得对于RNN的理解和训练更加容易.

1. 复杂结构RNN存在的问题

经典的RNN隐藏门控单元, 如LSTM与GRU, 虽然性能优越, 但是其中包含较多的门和隐状态, 结构较为复杂, 参数较多, 这导致对于RNN的理论分析和训练成为一件困难的事情. 与之相比, 简化的模型具有简单的结构, 便于理解RNN的结构和研究RNN的性质, 同时也具有较少的参数, 对RNN的训练开销也会减小. 因此, 为支持之后对于RNN的研究, 在保持准确率的基础上简化RNN的结构, 成为了一个重要的问题.

2. 想法: 基于现有门控单元进行简化

实验评估证明, 门控单元中门的数量与模型准确率之间不存在正相关. 例如, GRU作为LSTM的变种, 其隐状态和门的数量与LSTM相比, 都少了一个, 但是通常情况下, GRU的准确率要高于LSTM. 另有评估证明, 拥有一个门的单元, 其性能明显优于不包含任何门的双曲正切单元. 综合上述两个评估, 可以得到结论: 门控单元对于RNN而言是不可缺少的, 但并非门的数量越多越好. 于是, 简化RNN结构一个很自然的想法便是: 在保留门控单元的同时, 尽可能减少单元中门的数量, 便可以尽可能简化模型结构.

GRU基于LSTM模型, 减少了门的数量和隐状态的数量, 简化了结构, 并保持了准确率. 因此, 为得到更加简化的模型, 可以基于GRU模型进行进一步的更改, 有望得到更简化同时保持准确率的模型, 这正是作者的目标. 所以一个看似可行的想法便产生了: 基于现有的GRU模型, 进一步减少门的数量, 并保持性能不下降, 则寻找到了更简化的模型.

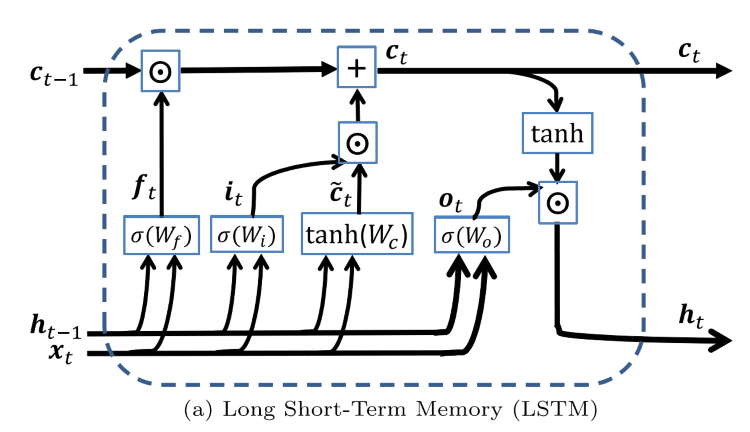
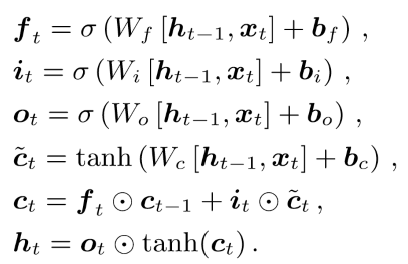
3. 想法的形式化

为了在GRU基础上进行修改, 首先需要了解LSTM与GRU的结构.

(1)LSTM模型

LSTM模型包括: 遗忘门, 输入门, 输出门; 隐状态, 单元状态; 单元 状态更新值.

上述数据之间的关系为:



,,三个向量中元素的值介于0到1之间, 可以将他们视为权重.

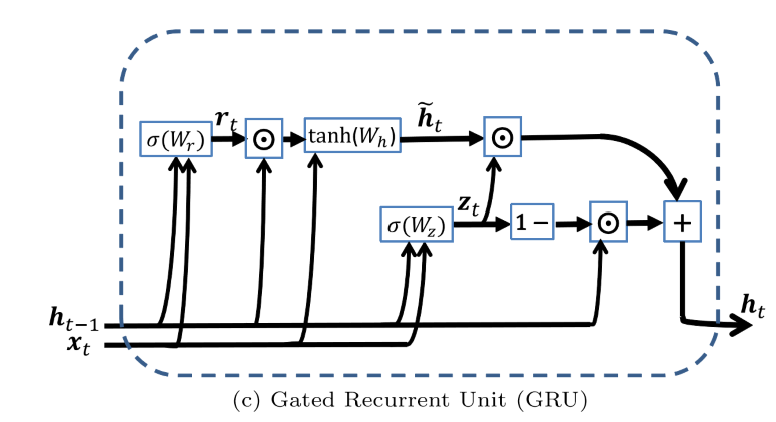
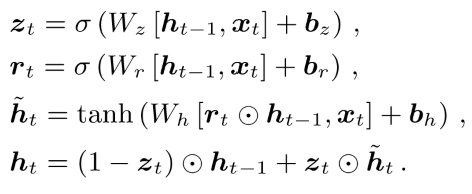
视为与的加权和, 权重分别由和规定.

LSTM模型共包括三个门, 两个隐状态, 四组参数.

(2)GRU模型

GRU模型包括: 更新门(作用与遗忘门类似), 重置门; 隐状态.

上述数据之间的关系为:



GRU是LSTM的 变种, 它对LSTM进行了较大的修改.

GRU模型共包括两个门, 一个隐状态, 三组参数.

(3)基于GRU设计MGU

LSTM有3个门与2个隐状态, GRU对其进行修改, 保留2个门与1个隐状态. 由 于隐状态和门都是不能缺少的, 所以基于GRU进一步的修改便是减少一个门.

汇总现有评估结果, 并提取其中一致的结论, 得到以下三条结论:

①门控单元对于RNN的高性能是有利的;

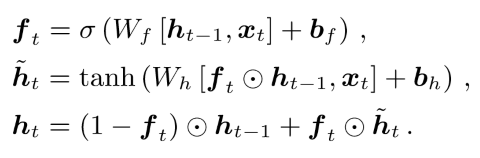
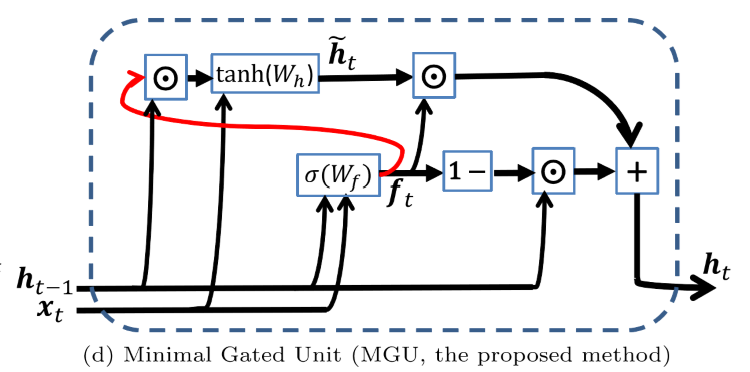
②所有评估都肯定了遗忘门的重要性;

③简化的模型能降低复杂度并且保持准确率.

基于上述结论, 将GRU模型的更新门与重置门都替换为遗忘门(更新门的作用与 遗忘门类似), 保持GRU其他结构不变, 得到了新的模型MGU.

MGU只有一个门, 是所 有门控单元中门个数最少的单元, MGU因此得名. MGU并 非旨在提出更好的RNN模型, MGU提出的目的, 一是利用更简化的模型, 在保持准确 率的同时降低开销; 二是简化对RNN的分析和理解.

MGU结构如下:

MGU与GRU结构非常相似, 上图中的红线是MGU对GRU的修改.

4. MGU的性能测试

MGU虽然结构简单, 但是如果其性能表现不佳, 那么MGU将不具有价值: 既不适合应用在实际问题中解决问题, 又无法为理论分析提供帮助. 因此, 需要对MGU性能进行测试.

作者选择了四个数据集用于测试MGU的性能: The adding problem, IMDB, MNIST, PTB. 四次实验覆盖不同长度, 不同取值的有序数据, 这样结果具有代表性.

作者通过对比LSTM模型与其多个变种模型的性能, 得出结论: GRU是综合性能最好的模型. 因此选择GRU作为基线方法, 用于和MGU进行性能对比.

通过在四个数据集上测试的结果可知, MGU在准确率上与GRU相当, 但它只有GRU三分之二的参数. 虽然MGU在某些问题, 如PTB数据集上收敛地比GRU慢, 但是综合考虑所有实验结果, MGU是一个可应用的模型.

5.结论

MGU只使用一个门, 它只有LSTM二分之一的参数和GRU三分之二的参数, 但是它拥有和GRU相当的准确率, 并且能更快的训练. 基于作者们的上述评估, MGU可以作为RNN中的隐藏门控单元, 更重要的是, 它最小化的设计将会促进RNN的理论研究与分析.

6. 进一步的讨论

MGU仅有一个遗忘门, 达到了门控单元的最简结构. 沿着这种最简的思想, 很容易猜测: 是否存在比MGU更简化的模型. 由于MGU仅存的一个门是至关重要的, 我们不可能删除这个门来寻找具有更简单结构的门控单元. 一个可能的方式是: 修改MGU中,,的定义, 或者删除,通过简化函数来降低参数数量, 寻求参数数量上的最简形式.

通过网络查找, 论文Simplified Minimal Gated Unit Variations for Recurrent Neural Networks(<https://arxiv.org/abs/1701.03452>)中提出了MGU的三种变种来降低参数数量, 三个变种都是对遗忘门的定义进行了修改.

MGU:

MGU1:

MGU2:

MGU3:

通过在MNIST和RNT数据集上测试, MGU2在这两个数据集上的测试性能优于MGU, 它在保持与MGU大致相同准确率的同时, 进一步降低了训练开销.

7. 参考资料

<https://zhuanlan.zhihu.com/p/34203833>

<https://zh.d2l.ai/chapter_recurrent-neural-networks/gru.html>

<https://zh.d2l.ai/chapter_recurrent-neural-networks/lstm.html>