## Basic Information

Expressive power of recurrent nerual networks.
Author: Valentin Khrulkov(Skoltech (Russia)), Alexander Novikov, Ivan Oseledets
ICLR 2019, Cited by 21

## Summary

This paper mainly proposed a theoretical analysis of the expressive power of a class of recurrent neural networks. By corresponding RNN to Tensor Train decomposition, the authors prove that RNN is exponentially more efficient than a shallow convolutional network with one hidden layer. They then also compare recurrent(TT) with convolutional (Hierarchical Tucker) and shallow(Canonical Decomposition) networks with each other.

In detail, the authors first introduce several tensor decomposition models to us. Then they built connection between tensor decomposition and network training. Then they focus on the tensor train decomposition which is analog with RNN. Finally the authors show that almost all tensor train networks require exponentially large width to represent in CP networks. It is proved that the space of TT-type networks with rank O(r) can be as large as the space of CP-type networks with rank $poly(r)$.

## Strong points

1. **Originality and contribution**

It is quite novel to introduce the tensor decomposition idea into analyzing the network expressiveness. Also it is also interesting idea to connect train tensor decomposition with RNN. This is a brand new point of view to understand the networks. By analyzing networks using tensors, it enables us to introduce and apply mathematical properties to help. Later works may also try to apply this to have a deeper understanding of networks.

2. **Writing**

I really enjoy reading the first several sections, namely the introduction, the deep learning and tensor networks and the tensor format reminder parts. They gave a clear and sound overview of the past studies and the background knowledge. These parts are easy to follow and help readers to understand the following parts.

## Weak points

1. **Clearness**

The authors did not clearly explain how to connect RNN with TT decomposition. First of all, RNN reuse the same parameters against all the input from $x_1$ to $x_d$. This means the decomposition tensors(TT-cores) $G^i$ will be all the same. Will everything still be the same under this condition? Also the analogy between the multilinear unit and an RNN unit is also hard to understand. RNN does not take such inputs but take linear combination of the previous hidden state and current input. I think the authors could have explain more using a simple RNN as an example. And as a result, I cannot understand the ”bad” example’s (low TT-rank but exponentially large CP-rank) translation into a RNN. Even the analogy is true, how precise it the corresponding? Could we recover the networks if we only have the decomposed tensors? I also don’t know how to understand the finding with respect to related neural networks, namely, RNNs and shallow MLPs.

2. **Technical details**

The comparison between CP networks and TT networks may not be fair due to the shallowness restricted CP networks’ expressivity.

3. **Experiment**

1) I am very curious how the authors implement the experiment on the MNIST and CIFAR=1- datasets. Namely, how to construct a CP or TT networks. They could have provide more details, such as the detailed structure.

2) Since RNN is good at dealing with sequence data, and we could also apply CNN to sequence data, should a sequence example will be more convincible and also give us a better understand of the idea of the paper?

3) I believe compare the TT and CP decomposition of the same rank will be more convincible if the purpose was to compare the expressiveness. However, is it possible to conduct numerical experiments for comparing the ranks?