Master Thesis

Evaluation of synergistic effects of tensor decomposition methods within (deep) neural network applications

Project Plan (Final version)

Tobias Engelhardt Rasmussen, s153057

February 12, 2021

Project Description

This M.Sc. project aims at assessing possible synergistic effects between tensor decompositions, such as TUCKER and/or PARAFAC and CNN applications. In this context, the tensor decomposition shall be understood as a data compression method. It shall be applied either prior to training the neural network to reduce the input dimensionality, or within the neural network to "compress the fitted parameters" and to potentially replace convolutional layers. The synergistic effects of the tensorized neural networks (TNNs) are meant to lead to increased computational efficiency and simpler network architectures, while at the same time performing at maximal prediction accuracy.

The project will implement tensor decompositions, such as TUCKER and/or PARAFAC, prior/within a chosen CNN architecture. Initially, a simple (image) data set will be simulated to test the methodology under well controlled conditions. In a second step, a suitable data set (from state-of-the-art papers) will be selected to assess and benchmark the developed TNNs against state-of-the-art performance(s).

This master thesis is to be based on research and trial and error, which is why there are no clear objectives. Also a part of the data set that is going to be used is not determined prior to the start of the project. Having this in mind, an approximation of the objectives of the project is given below:

1. Investigate different tensor decomposition methods and neural network architectures.

2. Investigate how others have attempted to solve the problem
3. Create a simple algorithm that works for a simple architecture for MNIST
4. Find a new data set to develop the algorithm even further
5. Develop the algorithm for the new data set
6. Evaluate the synergy by assessing different results using different performance measures

Initial Project Plan

Week	Activity	Risk
1	Write project plan, initial models on MNIST, literature study	1
2 - 4	Literature study, initial models on MNIST, and studying different methods	1
5 - 6	Create and test improved model for MNIST	3
7	Writing about MNIST methods and results	3
8	Explore new data set and establish baseline / state-of-the-art	3
9 - 15	Develop and test model for the new data set	5
16 - 20	Writing about methods and results for new data set and discussion	2
21 - 22	Correcting, writing conclusion and abstract and making everything come together	1

In the plan the risk is classified using a scale from 1 (no risk) to 5 (high risk).The risk is described as the chance of the activity being delayed.

Risk Analysis

Before week 5 In the first weeks it is all about getting to know different tensor decomposition methods, and how different architectures for convolutional neural networks work. It is also to study the problem itself and what have previously been done to solve it. This has a very low risk as literature is easily accessible and reading does not usually pose problems.

Weeks 5 - 8 In this period I plan to create a simple algorithm for MNIST using some decomposition method to help a simple neural network architecture perform better. In the end of this period the new data set should also be chosen and ready to work with. This has a slightly higher risk since developing an algorithm can cause problems, but not a high risk since the MNIST data set is relatively simple and well-studied.

Weeks 9 - 15 Since this period will be used to develop an algorithm for a more complex data set, the risk is high. This can be approached in many different ways, hence problems are likely to occur. Structure and a clear working order is needed during this time.

Weeks 16 - end The last weeks will be all about writing the missing sections, correcting and making everything come together. It should be noted that some of the writing will already have taken place, since some sections are better to write having the methods and thoughts clear in mind. This means that this period should have a relatively low risk of being delayed. Also allowing a couple of weeks of corrections means a potential buffer if something goes wrong.

Revised Project Plan

Below is the realisation of the project with the approximate activity and what sections of the thesis was written in the different weeks. The first week is the one starting on Monday 31st of August. The total period was extended for 2 weeks due to sick leave after a tonsil removal operation in the end of week 4.

Week	**Activity**	**Writing**
1 - 2	Write project plan, initial models on MNIST, literature study	
3	Literature study	Introduction
4	Literature study and results for MNIST 3s and 4s	
5-6	– Sick leave after tonsil removal operation	
7	Literature study	Previous Work
8	All digits for MNIST and implementation of decomposition for convolutional layer	
9 - 12	Investigating and preprocessing the THETIS data set and brainstorming methods	Data Methodology
13-14	Implementation and getting to know the DTU Computing Center (HPC)	
15-17	Theoretical speed-up (FLOPs), timing results, investigating the 1x1 convolution	Methodology
18-21	Timing results	Theory Results
22-24	Corrections	Discussion Conclusion Abstract

Self-evaluation

I have been trying to stick to the project plan all along, however some problems with getting useful timing results made me revise the entire timing procedure why the results were still obtained relatively late (up to week 21). Like planned I have been writing all along due to the concepts being fresh in the memory.

In the beginning, I started out getting to know all the tools I was going to use while playing around with them. I was also reading all the literature that I was going to base my methods upon. I combined reading and playing around in order to get a better understanding of what was done. This resulted in the literature study taking one week more than planned.

After having obtained some results using the different methods on the MNIST data set, I was presented with the more complicated THETIS data set that took some investigation and playing around to become confident with. After pre-processing the data set, the methods were developed and implemented which was pretty straight forward due to the knowledge from working with the MNIST data set, but still took time.

As mentioned above I used a lot of time investigating the best way to time the layers of a neural network, and ended up changing the whole procedure, resulting in a delay in obtaining the results. Fortunately I managed to get the results in time to get everything finished.