**Digitale Bildverarbeitung**

**Lectures**

1. Backpropagation and an Introduction to Tensorflow

In this lecture, the basic structure of a neural network was developed. The principle and math behind backpropagation was explained. In the later phase we tried different optimizations and regulations as activation functions, loss functions or different types of layers.

1. Transfer Learning with Tensorflow for Object Classification

In this lecture, the VGG16 model has been trained for a similar task, but with a different dataset. By freezing the layers and adding a new layer, right before the flatten layer, generality can be maintained.

1. Segmentation with U-Net

In the last lecture, segmentation with the U-Net was done, an application when objects need to be located and classified. An image is divided into different areas and analyzed, processed and then an output is generated again. A dice loss was implemented, the data prepared and a generator for training was written.

The project is about an appropriate handling of the MovieLens 100K Dataset, a recommender system.

The Data Set includes 943 users, with age, gender, occupation, and zip code. These users had the ability to watch 1682 movies and gave 100000 ratings. Each user had to rate at least 20 movies.

The goal of a recommender system is to provide the user a recommendation of movies he/she could enjoy, based on his/her previous ratings. Therefore collaborative-filtering was used, which is based on matrix factorization, the reason therefore, it can scale significantly better to massive datasets. But make better recommendations based on user's tastes.

Data Analysis:

In the beginning of the project data analysis was performed. The main goal is to experience the user data since their impact on the output is the most powerful.

Here students made by far the biggest number of users, out of 21 possible occupations. With nearly 200 students they are a fifth. This could result into a huge bias in the direction of students. That has a bad performance in the direction of all other users.

The next finding is that 71% of the users are men. This again results in a bias to men and a bad performance for female users.

The last finding is that most of the users are in the age between 25 and 34, and again a bias in their direction will be the outcome and the performance for older or younger users will decrease.

Unfortunately, the data frame is not big enough to train the model on these criteria.

Data Pre-Processing:

To use the information we gathered about users for collaborative-filtering, their properties and the movies they rated, it is necessary to encode that information, so that in the end it is possible to describe the users with vectors. These vectors are going to be updated by training the model, so that a user described more specific and recommendation fit better.

Work on in the future:

About the model, it would be interesting how the output would be if instead of only focusing on the user rating, also other parameters such as gender, age or occupation had an impact.

Unfortunately, I had problems implementing my ideas, so it isn’t possible for me to demonstrate some examples.

Also a thing to try could be another strategy to split the data set by using the timestamp. To create a real-world performance with the leave-one-out strategy is used. That means every user recent viewed movies is kept as train data, to prevent the model of a data leakage and look-ahead bias.