Progress Report

1. Adjustment

Our project is to build a music recommendation system and we have been working on our project so far. After receiving some suggestions from the TAs, we have planned to apply more course related technology or instrument on our project, such as hadoop. So, we are going to implement mapper and reducer programs for our project to run on hadoop. Apart from this, considered of the limitation of time and the content , we would like not to use CNN and RNN in our project and we would apply some other algorithms related to deep learning in our project and compare their performance with collaborative filtering.

2. Current work

Up till to now, we find the database called ***million songs,*** but we notice that there are only some characteristics through ***million songs***. So we crawler the data from the platform of NetEase Cloud Music. You can see our SQL file in the github repository in the database folder. In terms of data, we still have challenges to be faced, because the limit of resource in NetEase Cloud Music. We only get around 30,000 lines of data. We realize that it is not a big data. We will explain clearly the rest of related problems in Part Three.

Another module we have already done is collaborative filtering. For the collaborative part, we use the python package surprise. We use the package mainly to find the similarity between the users for our user-based recommended system. We calculate their similarity using KNN algorithm. Basing on the similarity between the users, we can recommend the items to the user.

3. Challenges

In the next stage, we will try to figure out how to apply Hadoop to solve the collaborative filtering algorithm. The whole algorithm should be broken down into several simple steps with several mappers and reducers. We are going to design the framework of the mapreduce system, only by this way, the scaled data processing is feasible.

Besides, the source of data is another problem. Now we only acquired about 30000 users and music data, while the audio data has not been obtained. In the next stage, we are going to obtain more music and user ratings data from million song data set. For audio file in deep learning, we plan to crawl data from **netease music**.

4. Future work

For the following steps, we are going to analyse the ***voice frequency.*** Firstly, we'll transform the voice frequency into the image. And then select some apparent characteristcs from the image by Fourier transform. So that we can do the machine learning and find the key parameter in the model.

What’s more, currently we build a user-based recommended system. We plan to build an item-based recommended system and compare its performance with the user-based recommended system.