Music Recommendation Platform

Rob Verbeek 1003226
Sakshi Udeshi 1003197

In this project we propose to use the data from this Kaggle competition to build a music recommender engine. Input data would be user data and a song. Output would be how likely the user is it that the would like the song. It is trivial to see that this can be extended to function as a recommendation system based on a corpus of data.

Milestones

1. The first step would be to do a survey of the technology that exists in the recommender systems. Another parallel goal would be to try and source alternate data that we may be able to use to supplement and extend the project. This should be done by Nov 1, 2017.
2. The next step would be trying to experiment with the data and cleaning up the data, if necessary. This will also include writing code to transform the data into vectors and data modelling. We target to finish this by Nov 7, 2017.
3. We aim to code up up to Collaborative filtering, Content-based filtering and Hybrid recommender systems and test their performance against the training data. This should be complete by Nov 21, 2017. We also aim to use a standard implementation of K Nearest Neighbors (KNN) as a baseline.
4. A stretch goal would be to build a website to supplement the recommender engine, using real data and real artists. This would be done by Dec 1, 2017
5. Another stretch goal is to explore the feasibility of dynamic recommender systems, where the system is continuously supplied with new data and learns on that new data as well. This deadline would be Dec 9, 2017
6. The final step would be to write a report of what we have accomplished over the course of this project. This

We will reflect after each step to verify if our future steps make sense and modify our strategy according to what we have learnt so far.

Potential Problems

1. Low accuracy. This is a problem that we will most likely face. A good way to mitigate this would be to talk to our peers at SUTD to guide us with respect to recommender systems and having periodic code and logic review with the TA, Balu.

2. Anonymized data. Our stretch goal relies on the assumption that we will be able to find high quality anonymized data outside this Kaggle project.
3. Lack of time. We may run into unforeseen problems with this project. We have given ample slack in our deadlines. And our intermediate steps can be presented as a complete project. Our reflection after each step will also help us to recalibrate our goals.
4. We will not face the cold start problem as we have sufficient data.

Responsibilities

- Rob will be responsible for researching the appropriate recommender systems. He will also be responsible for coding up Collaborative filtering project.
- Sakshi will be responsible for data clean up and transforming them into useful vectors. In addition to this, he will code Content-based filtering and Hybrid recommender systems
- Both members will be equally responsible for the website implementation and stretch goals.