Apprentissage Avancé Project

Xiaoxiao CHEN,Yuxiang WANG,Honglin Li

February 10, 2017

Abstract

We decide to take "Million song Dataset Challenge" on the site kaggle (https://www.kaggle.com/c/msdchallenge/data). Our group consist of three members: Xiaoxiao CHEN/Yuxiang WANG/Honglin LI. The project proposition has four main parts. The first part presents the general description of the project. The second part presents the development tools that are expected to be used. The third part presents the documents that can be helpful to our projects. The last part presents the middle and final objective of the project.

1 Introduction

Project task
The Million Song Dataset Challenge aims at being the best possible offline evaluation of a music recommendation system. We are very interested in this issue for three reasons as follow:
1) We do not have experience about recommendation system yet and it is interesting to try something new.
2) The data for this challenge is completely open, almost everything is possibly available.
3) Any type of algorithm can be implemented and new method are encouraged to explore.
Dataset and evaluation.
The core data is the Taste Profile Subset released by The Echo Nest as part of the Million Song Dataset. It consists of triplets (user ID, song ID, play count). The data is split in two:
1) The train set contains a little over a million users, full history released .
2)The validation and test sets combined contain 110k users, half of their history released (available here on Kaggle).

2 Development tools

To realize this recommendation system, any type of algorithm can be used: collaborative filtering, content-based methods, web crawling,etc.
   Our main idea:
1) Realizing a baseline by some tools or a simple algorithm.
2) Learning some algorithms on the paper and try to realize the algorithms.
3) Trying to get more data from the internet and improve our model.

3 Documents for reference

1) THE MILLION SONG DATASET, THE MILLION SONG DATASET Thierry Bertin-Mahieux, Daniel P.W. Ellis, Columbia University, LabROSA, EE Dept. AND Brian Whitman, Paul Lamere, The Echo Nest,Somerville, MA, USA.
2) A MUSIC RECOMMENDATION SYSTEM BASED ON MUSIC DATA GROUPING AND USER INTERESTS, Hung-Chen Chen and Arbee L.P. Chen, National Tsing Hua University AND Hsinchu, Taiwan 300, R.O.C.
3) Dataset website:http://labrosa.ee.columbia.edu/millionsong/

4 Objective for middle and final report

Objective for middle report For the middle report, we expect to finish the implementation and test of our first version system. With the feedback of the test result, improvement of algorithms or experiments of new method are encouraged.
Objective for final report For the final report, we should have a further understanding of this type of recommendation system. The final solution will be formed with best result and hopefully we can develop our own unique method for this challenge and propose some open questions or advises to further improve this system.