WARSAW UNIVERSITY OF TECHNOLOGY

Project proposal

Music recommendation system

EINIS 2016L

Author: Michał Godek

Supervisor: prof. dr hab. inż. Mieczysław Muraszkiewicz

April 22, 2016

Contents

1 Task summary

The subject of the project is to create a recommendation system which will aid finding music tracks of ones liking. To achieve this, Collaborative Filtering together with Matrix Factorization technique will be used. Such an approach makes the solution domain agnostic and does not require to know feature details of ranked items. This makes it highly reusable for other content if only user preferences are available.

1.1 Goals and objectives

Objectives have been divided into two groups. The obligatory part, which is the key of the project, will be complete in whole. The optional piece will be left to decide if it should be executed at all, depending on the time required and result obtained from the first stage.

1.2 Obligatory

Main goals of the project are to:

- create a music recommender program, which outputs tracks based on an input playlist;
- evaluate the resulting output.

1.3 Optional

- A few track features will be incorporated into the Collaborative Filtering model to aid the latent feature matrices creation.
- A possible extension is to use a much larger music data set and improve the computation time using either:
  - the Apache Spark library and deploying the solution to a cloud cluster;
  - the TensorFlow[1] library and utilizing the GPU of a local machine or deploying the program to a cloud for computing.

2 Task analysis

This section lists the elements constituting the solution of the problem at hand.

2.1 Data set

The essence part of the project is the data set on which the prediction model will be created and evaluated. The archive named 'The Million Song Dataset'[2] provides a listing of a million tracks along with their features and a real user preference set. In this project the features of the tracks are ignored, instead the focus is put on the users' taste in music. The data set used in the obligatory part is limited only to a randomly chosen 1% of the whole 280 GB repository.

2.2 Prediction model

Collaborative Filtering is an approach to finding new content for recommendation to a user, which depends mostly on 'the wisdom of the crowd', where suggested music comes from other users' playlists. This project will utilize the idea presented in the paper "Matrix Factorization Techniques

for Recommender Systems"[3] written by the winners of the 'Netflix Prize'. The concept behind it is that given a user - track preference matrix:

	Track 1	Track 2	Track 3	Track 4
User 1	7	?	1	?
User 2	?	?	?	4
User 3	9	0	?	2
User 4	1	8	?	0

Table 2.1: User - Track preference matrix

the algorithm searches for latent feature matrices:

	Feature 1	Feature 2
User 1	?	?
User 2	?	?
User 3	?	?
User 4	?	?

Table 2.2: Users - Latent features matrix

	Track 1	Track 2	Track 3	Track 4
Feature 1	?	?	?	?
Feature 2	?	?	?	?

Table 2.3: Latent features - tracks matrix

which when multiplied would create the input preference matrix. The latent matrices are calculate by iteratively improving them to find their optimal values and predict the missing values in the original user - item matrix.

2.3   Evaluation method

Assessment of the results of a recommender system is not straight forward, since for true grading of the solution, real users would need to evaluate it and share their opinion on the usefulness of the program. However, to achieve the goal of this project, two ways of testing are proposed:

- prediction will be checked against a test set, which will be manually prepared by removing ratings for some of the tracks known by a user. Than a check will be performed of the position of those manipulated tracks in the extended output ranking suggested to the user.

- the output playlist will be evaluated depending only on the subjective taste of the author.

2.4   Assumptions and constraints

User input playlist is given with a play count for each of the tracks. Play count is treated as a measure of liking a user has to each track.

3   References

[1] Katherine Bailey. Matrix factorization with tensorflow, 2016.

[2] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

[3] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, August 2009.