The Tutorial Report

Runpeng Li
rli3@oxy.edu
Occidental College

0.1 Introduction

This project aims to build a highly accurate music recommender system using Python. Music recommender systems are vital for the music industry as they help suggest new music to listeners, promote upcoming artists, and enhance music consumption. To assist me in achieving my objective of building a highly accurate music recommender system using Python, I used the tutorial by Velardo (2022). This tutorial is highly relevant to my objectives as it explores how music recommender systems such as Spotify are developed using collaborative filtering. It highlights the collaborative filtering techniques known as the user-item matrix that can be used to develop a music recommender system based on users' opinions. By following this tutorial, a successful outcome for the project would be creating a highly accurate music recommender system using a matrix factorization algorithm known as alternating least squares.

0.2 Methods

The music recommender system will be built using a collaborative filtering technique, which is considered highly effective in pairing users with similar tastes together into groups. The collaborative filtering music recommender system that will be built will be based on the collective preferences of the users within the different groups. The main algorithm that will be used for the system is alternating least squares (ALS). As Velardo (2022) states, the main reason for using the ALS algorithm is to utilize its ability to increase prediction accuracy by solving overlapping issues. There are other alternative algorithms that can be used, such as the K-Nearest Neighbors, User-Item Linear Regression, and Autoencoders, among many others. However, a major strength of the ALS algorithm over other alternatives is its high scalability, as it is capable of handling large datasets. A major deviation made from the tutorial is the use of the user-music matrix instead of the user-artist matrix. Velardo (2022) recommends the use of a user-artist matrix as the system he is looking to build aims to connect users to artists. Instead of recommending certain artists to the users, this project's music recommender looks to connect users to certain music types. The goal behind this deviation from the tutorial is to enhance the user experience and limit the potential of bias favoring employees from certain regions. The implementation of the ALS algorithm will rely on data from the last.fm dataset, which contains a large collection of music. It is an ideal dataset to use as it tracks the listening habits of individuals from all over the world. Hence, it can be highly reliable due to the generalizable findings. Another valuable tool during the implementation of the ALS algorithm is the implicit library. This library contains a lot of functionalities that can be used in collaborative functioning.

0.3 Metrics and Results

The main performance metric that will be used is root mean square error (RMSE). RMSE is an effective evaluation metric as it measures the average difference between a predicted model and the actual values. These performance metrics can help achieve the project's objective as they outline the level of accuracy of the music recommender system. The lower values of RMSE are considered more desirable as a hypothetical model that would recommend the exact value has an RMSE value of 0. In cases where a high RMSE is recorded, it suggests that the music recommender's accuracy is low and needs to be improved. A high RMSE can be improved by introducing more influencer variables in the trained dataset.

0.4 Reflection

Following the tutorial was extremely helpful as it provided me with much-needed guidance for the successful development of a music recommender system using Python. The successful creation of a music recommender system can be vital for supporting the music industry as it exposes artists to a larger target audience. Despite these positive aspects associated with the topic, there are certainly some ethical concerns that may arise from the project. One of the major ethical concerns is the likelihood of the project perpetuating bias. Based on the training data, certain regions and their artists are likely to be prioritized ahead of others. Another major ethical concern that may emerge is ensuring that the system is inclusive. As diversity and inclusivity have become increasingly sensitive topics, ensuring that there is a balance in the communities can be a challenge. In order to address these concerns associated with the system, the project should look to ensure that diverse and representative data is implemented. Using diverse and representative data enhances generalizability, thus reducing bias. In addition, various fairness metrics should be employed to ensure that there is enhanced diversity and inclusivity. Examples of fairness metrics that can be utilized to enhance the music recommender's efficiency include demographic parity (DP) and equal opportunity (EO). Implementing these fairness matrices ensures that discrimination is minimized, reducing the likelihood of legal battles or reputational damages.