Music Recommender Systems: Progress Report 15-780

Deepika Bablani, Raghav Gupta, Tejas Gokhale

20 April 2017

1 Dataset Pruning and Identifying Metadata

We are using the following files provided by the Million Song Dataset:

1. Million Song Subset: A subset of 10,000 songs (1.8 GB) selected at random along with their metadata.
2. Taste Profile Subset: A subset which contains triplets of the form (user id, song id, # of plays) for 1 million users.
3. musicXmatch : A collection of lyrics for 210,519 tracks.

We pruned the taste profile subset to contain only those triplets for which we have song entries in the million song subset. Our pruned triplet dataset has 418,252 entries. We selected metedata attributes for content and collaborative recommendation based on whether they had valid entries atleast for most songs. From the metadata we identified the following attributes:

1. Artist id
2. Song Duration
3. Key Signature
4. Song Tempo
5. Time Signature

2 Collaborative Filtering

We implemented a collaborative filtering mechanism based on a nearest neighbor approach. Given a user id, we find the n most similar users using Pearson Correlation. Then using songs heard by the n similar users and not by the user, we pick a recommend a song based on a prediction function:

$$pred(user) = max_{songs}\ (\sum_{n_i}(n_i[s] - \overline{n_i}) * sim[n_i, user])$$

where $n_i[s]$ is the rating given to song s by neighbor i, $\overline{n_i}$ is the average rating given by neighbor i and $sim[n_i, user]$ is the similarity between neighbor i and the user.

We tested this recommendation model on a test set of 10,000 samples. We looked at the deviation of the recommended song from the average user profile based on the attributes mentioned above (see Figure 1). We see that some attributes are better reflected in our recommendation method. For instance in case of time signature, more than 70% of the samples have a deviation of less than 10%. We also see that there is not a huge amount of overlap in artists between the user and the most similar neighbor. This is primarily because of sparsity in the dataset where not all users have heard a lot of songs.

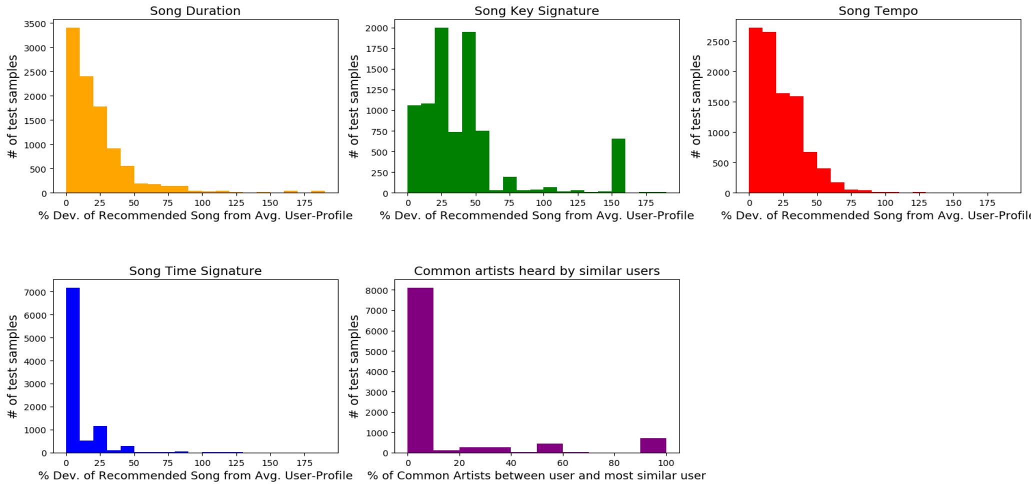

Figure 1:

3 Content Based Recommendation

We also implemented a content based mechanism for predicting songs. The approach makes an average user profile for every user based on the attributes of the songs played and the no. of plays for each song. Then the algorithm tries to find songs with similar attribute values as the average user profile. This is expected because in content based recommendation the song is recommended based on minimizing the deviation whereas collaborative filtering makes recommendations just based on songs liked by similar users.

We tested this model on a test set of 10,000 samples. We looked at the deviation of the recommended song from the average user profile (see Figure 2). We see that in this model the deviation decreases in general for all attributes.

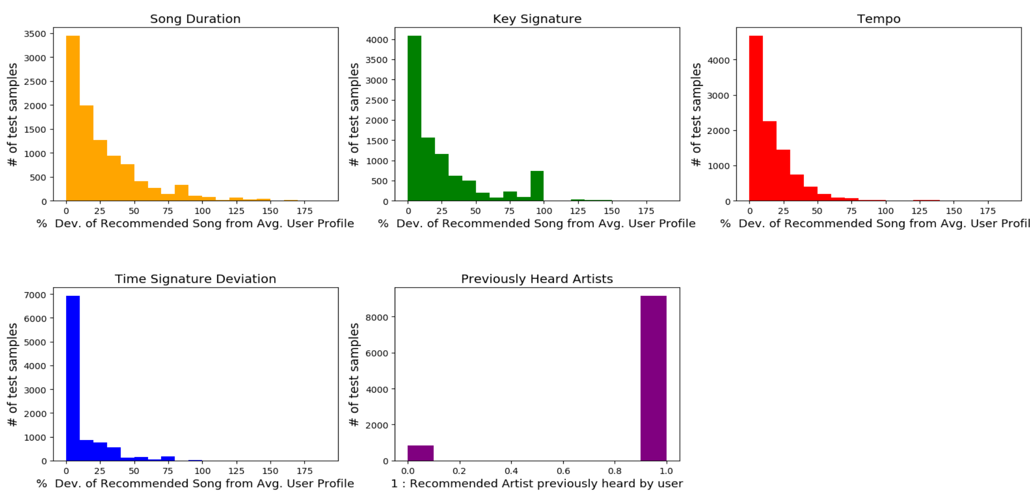

Figure 2:

4 Deep Neural Network

The motivation behind using a deep neural network for music recommendation is to overcome the cold start problem associated with collaborative filtering. Due to the cold start problem songs which have not been rated in the past cannot be recommended. Using a deep network, we propose to learn latent factor representation of a song using a bag-of-words approach from the song lyrics. Recent literature shows that this has worked better than content or collaborative based recommendation [4]. Ideally we would have liked to use log-compressed-mel-spectrograms as input to the network but due to unavailability of raw audio we were unable to follow this approach. Instead, we have extracted feature vectors containing word counts for song lyrics.

Future Work: In our implementation, we plan to concatenate the bag-of-words of a song and the user-id as an input. We will train the network to predict the rating of a song given a user-id. Then to predict a song for a user, we will go through the songs listened to by the *n* nearest neighbours and find the rating of the songs for the given user. The song with the highest rating will be predicted. Since we have a latent representation of every user's song preference, we don't need collaborative data to predict the rating of a new song for that user. Hence we claim that this approach will solve the cold start problem.

References

[1] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman and Paul Lamere *The Million Song Dataset* Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), 2011

[2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl *Grouplens: An open architecture for collaborative filtering of netnews* Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work, pages 175-186, New York. ACM

[3] Dietmar Jannach and Gerhard Freidrich *Tutorial: Recommender Systems* International Joint Conference on Artificial Intelligence, 2013

[4] Aaron van den Oord, Sander Dieleman and Benjamin Schrauwen Deep content-based music recommendation, NIPS 2013