

Predicting Audio Features with Last.fm Tags

Jaime Ramírez Castillo¹ and M. Julia Flores¹

¹Departamento de Sistemas Informáticos, Universidad de Castilla - La Mancha, Spain

ABSTRACT

In this paper, we discuss a number of experiments to analyze the suitability of music label representations to predict certain audio features, such as danceability, loudness, or acousticness...

Keywords: Music information retrieval, Artificial intelligence

INTRODUCTION

Music information retrieval (MIR) is a field that focuses on the analysis, processing and knowledge discovery of information stored in music pieces Ramirez and Flores (2020).

Many of the challenges in this field use audio features, such as valence and energy, to predict other music-related aspects. For example, some studies have been using these features to predict the probability of a track being a hit [REF].

To make these predictions, researchers have been using machine learning models, specially over the past few years. Previously, researchers commonly used hand-crafted audio computational models to perform these tasks.

But what if we tried to use these audio features as the predicted values? We would need a set of predictor variables to accurately predict the value of audio features. This premise is the core concept of this article.

The idea is, given a set of tags, to predict a set of audio features that a hypothetical track would exhibit. Then, we could build a track selection algorithm that selects actual tracks that are the closest to the predicted audio features. This process could be part of an explainable recommendation pipeline, where users enter a set of tags, and they receive the predicted audio features, the closest tracks to those features, and the distance values between each track and the predicted features.

In the remainder of the article, we explain the data gathering and preparation process, as well as the data input formats and various models. We will explore various models for the same track and provide insights on how accurately the prediction can be, by using only Last.fm tags.

STATE OF THE ART

In recent years, researchers have used Last.fm tags in classification and regression tasks. Several studies have used Last.fm to predict music sentiment or mood.

Additionally, the Spotify audio features...

In general, these studies confirm the possibility of extracting knowledge from Last.fm tags. To the best of our knowledge, no studies have addressed the problem of audio features regression, based solely on Last.fm tags.

To be Continued...

GENERATING A DATASET

Before conducting experiments on predicting audio features from tags, we constructed a dataset, by gathering the data from the Last.fm and Spotify APIs.

A Single-user Dataset

Similar to other intelligent systems, recommender systems must be trained, by using user preference data, to produce adequate recommendations. For our recommendation framework, we have leveraged the knowledge discovery potential of large historical listening logs, gathered from Last.fm.

To characterize the preferences and context of the user, we have chosen to start with a simple scenario, where just data from a single user is available. By training our system with data from a single user, we also want to begin a discussion, given the following question: Is it possible to train recommender systems, and in particular, user-centric systems, by using a single-user dataset?

To the best of our knowledge, research on user-centric recommender systems has concentrated its efforts on explainable AI Wang et al. [2019], and also user-centered evaluation of these systems Knijnenburg et al. [2012]. We also argue that recommender systems that exploit the preferences of a single user, or a reduced number of users, might as well be considered as user-centric models.

Last.fm

Last.fm is a online music service for users to keep track of their music listening habits. Last.fm can also be considered as a community where users tag artists, albums, and tracks, according to the own perception of the user.

For nearly two decades, users have been contributing to Last.fm by tagging tracks with arbitrary text labels. These tags do not necessarily have to be single-worded. Users often use short sentences to define a song, such as 'I like this track', or 'on the beach'.

Community-contributed tags from the Last.fm API. These tags are text labels that Last.fm users assign to artists, albums, or tracks. Users apply these tags to categorize music from their own perspective, which means that tags do not fit into any structured ontology or data model. Tags can refer to aspects such as genre, emotion, or user context.

Last.fm Tags

Last.fm uses the term *scrobble* to refer to a single track playback, in a particular moment. We have queried the Last.fm API to download the user's scrobble logs, reported from 2007 to 2022. For each scrobble, we have gathered the following information:

- Track playback timestamp.
- Track MusicBrainz Identifier (MBID), if exists.
- Track name
- Artist name
- Track tags. If the track does not have any tags assigned, then artist tags have been used.

For each tag assigned to a track, or an artist, Last.fm includes a count property to indicate the popularity of the given tag for the track. Last.fm normalizes this value in the 0-100 range, so the most popular tag for a track can have a count value of 100.

Users normally listen to their favorite tracks many times, so the amount of individual tracks listened is much smaller than the number of track plays. In this case, the amount of individual tracks listened is about 20,000.

The format is as follows:

```
{
  "artist - name": {
    "electronica": 100,
    "rock": 80,
    "pop": 45,
    "jazz": 0,
    "nu-jazz": 0,
    "country": 0,
    "soul": 0,
  }
}
```

Spotify

After gathering Last.fm data and identifying the unique tracks that represent the user music collection, we have collected Spotify audio features. For each of these individual tracks, we have downloaded the Spotify audio features specific to the given track.

The Spotify audio features are numerical values that represent high-level audio information computed from a specific track. These values characterize a track, musically speaking, by measuring relevant musical aspects.

The features provided by the Spotify API are acousticness, danceability, duration_ms, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, and valence. Table 1 describes these features. The reader can find further details about each feature in the Spotify API documentation 4.

A small portion of the tracks do not have features available in Spotify, so they have been filtered out from our experiments.

After filtering songs that miss Last.fm tags or Spotify audio features, our dataset contains 14009 samples.

Track audio features from the Spotify API. These are attributes computed from the audio themselves. They are a way to describe music by using numerical values. For example, a danceability attribute of 0.95 means that a particular song is highly suitable for dancing.

Last.fm Tags Representations for Training

Tabular

Each tag is a column and each cell contains the popularity value of a tag for a track. A cell is 0 if a tag is missing for a track.

The number of columns is limited to the top-K tags.

Tabular Tokens

Tags are converted to text tokens. Columns represent token positions, and cells contain the token at a particular position, for a track. To tokenize tags, we have used the GTP2 tokenizer. Because the tokenizer requires a string as input, we have converted the set of tags for each track into a string. To *stringify* the tags, we have concatenated tags with multiple strategies:

- By including tag popularity: ‘rock 2, pop 1’.
- By repeating tags based on popularity: ‘rock rock, pop’.
- By ordering by popularity: ‘rock, pop’.

Training Data By Track

When generating training data by track, the tabular formats present sparsity problems.

For tabular representations, we need to define a fixed set of columns as tags. For most of tracks, most columns are ‘0’.

The sparsity of a matrix is the number of zero-valued elements divided by the total number of elements (e.g., $m * n$ for an $m * n$ matrix) is called the sparsity of the matrix

EXPERIMENTS

TODO

Models

TODO

CONCLUSIONS

TODO

ACKNOWLEDGMENTS

TODO

REFERENCES

Ramirez, J. and Flores, M. J. (2020). Machine learning for music genre: multifaceted review and experimentation with audioset. *Journal of Intelligent Information Systems*, 55(3):469–499.