Mood Dynamic Playlist: Interpolating a musical path between emotions using a KNN algorithm

Patrick Donnelly [1], Shaurya Gaur [1]

[1] School of Electrical Engineering & Computer Science,

Oregon State University

Corvallis, OR 97331, USA

ABSTRACT

We often to listen to music for its power to change our emotions. Whether selecting music for concentration, tunes for dancing, or lullabies for falling asleep, people often select music based on their desired mood or activity. We propose a method for automatically generating musical playlists that takes the listener on an emotional journey. We represent a playlist as a path of songs through the arousal-valence circumplex space, using existing datasets of songs annotated with affect values. Given a starting and desired affective state, we employ a K-nearest neighbor approach to choose songs that gradually and smoothly step through the affective space. We compare several different distance metrics and we evaluate the smoothness of the resulting playlists using mean squared error. We discuss an example playlist and link to a demonstration of our approach.

Keywords: Arousal-Valence, Machine Learning, Music, Playlist Generation

INTRODUCTION

We listen to music for a variety of reasons, including enjoyment, distraction, joy, comfort, and for focus. We sometimes choose certain music to reinforce our mood, and at other times, we choose music to change our mood altogether. People often curate music playlists based on certain emotions, using these playlists to reflect upon or alter their own moods.

Some music platforms, such as Spotify[1], create automated playlists based on rules and preferences set by the user. Although many people listen to music while they work to help focus, the order and type of music that plays from randomized user libraries are not usually optimized to steer users towards better states of mind. Instead, these playlists are most often designed to keep a listener in a specific mood.

Although previous research studies have attempted to create playlists consisting of songs with a consistent affective state, we propose an automated playlist generation algorithm that takes the user on an emotional journey.

Related Work

Previous studies that generate music playlists often focused on maintaining a consistency throughout the playlist, whether that be staying within a single genre, unified by common emotion, or musical similarity based on acoustic analysis.

Several studies rely only on the similarity of song's metadata and acoustic attributes and use a variety of approaches, including local search (Pauws et al., 2006), graph search (Sakurai et al., 2021), nearest neighbors (Pao et al., 2008), and reinforcement learning (Sakurai et al., 2021), while other studies have also taken into account emotional similarity (Deng and Leung, 2012). In a unique approach, one study attempted to learn coherent playlists solely from other playlists, without using features extracted from the songs themselves (Chen et al., 2012).

Other approaches have focused on optimizing the transition from one song to the next song. One such study recommends songs in a playlist by selecting for acoustic similarity between the end of one song and the start of the next (Ikeda et al., 2017). Another approach creates an ordered playlist based on smooth transitions in audio similarity between adjacent songs (Flexer et al., 2008).

A few approaches to automatic playlist generation attempt to take into account the users' current affective state or activity level. One study queried its users for their mood and energy level (Rumiantcev and Khriyenko, 2020), while others attempted to automatically determine a user's affective state using computer vision (Dureha, 2014) or from physiological data, such as the user's heartrate (Liu et al., 2009) and electrodermal activity (Zhang et al., 2018). In another study, the authors measured body movement, attempting to correlate the tempos of the tracks with accelerometer data in order to recommend the next song based on the activity (Kim et al., 2019).

[1] https://www.spotify.com/

Contributions and Novelty

We present an approach to automatically creating playlists that take the user on an emotional journey. Given a specified starting and a target song, our approach automatically generates a playlist that slowly and steadily transitions from the mood of the starting song to the mood of the target song over the course of the playlist. Such a system would easily enable users to self-identify their current and desired moods and to create a playlist that gently walks through this emotional space. For example, a user might create a playlist that slowly moves from sad to energetic music, perhaps allowing the listener to transition from a lethargic state to a more productive state.

DATASET

We first identified a dataset of songs annotated with their arousal and valence values. The arousal-valence circumplex model is a representation of affect in a two-dimensional continuous space in which the x-axis represents valence and the y-axis represents arousal (Russell, 1980). Valence measures the "positivity" of an emotional response and ranges from negative to positive. Arousal measures the "intensity" of the response and ranges from calm or soothing on one side to exciting or agitating on the other. Together, these coordinates are a common framework used to represent an emotional response to a stimulus. In our approach, we consider a playlist as a path through this emotional space.

Annotating music with arousal and valence values is a difficult and time-consuming task that requires a large-scale human subject study and the navigation of copyright laws. Although a few datasets exist, such as DEAM (Aljanaki et al., 2017) or PMEmo (Zhang et al., 2018), these datasets are often relatively small, given the time and expense of annotating music by multiple raters. Moreover, often these datasets contain relatively obscure songs chosen because they were available in the public domain. Unfortunately, there does not exist a human annotated dataset of sufficient volume to use in our approach.

For this reason, we chose to use the Deezer 2018 dataset, which provides estimated valence and arousal values for 18,644 songs (Delbouys et al., 2018). The authors estimated these values based on text analysis of LastFM[2] tags combined with a word list of 13,915 English words and their embedding in the arousal-valence space (Warriner et al., 2013).

Although the Deezer dataset (see Figure 1) has the advantage of its relatively large size, it is important to note several limitations. These statistically estimated annotations are less reliable than those collected by human annotators. Furthermore, the procedure to generate these estimated annotations resulted in many small sets of songs that have the exact same arousal and valence values, presumably because they each contained the same LastFM tags. This consensus would not occur frequently in a human annotated dataset in which several different listeners rate each song. Despite

[2] http://www.last.fm

these limitations, this dataset has the advantage that most all of the songs are readily available in commercial catalogs.

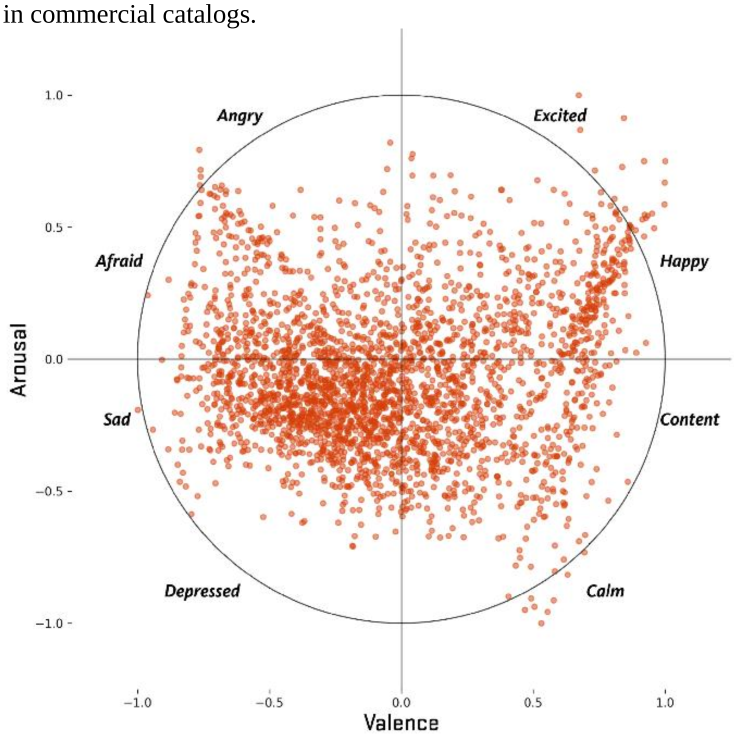

Figure 1. Distribution of the Deezer dataset (n=18,644) along the Valence and Arousal circumplex model

METHODS AND RESULTS

We consider a user's current affective state as a point represented in the arousal-valence space and we desire to create a path to a target affective state, also represented as a point the arousal-valence space. We generate playlists by choosing songs along an emotional path from the starting point to the destination point. This emotional path should be smooth and the songs should be evenly spaced. This ensures the no single song is too drastic a change from a previous song and that the emotionally tone changes gradually over time.

Approach

We use a *k*-nearest neighbors approach to determine songs closest to an idealized smooth trajectory through the arousal-valence space. We first determine this ideal trajectory as a straight line from the starting point to the target point. Next, we determine a hypothetical point along this ideal line. This hypothetical point is spaced from the starting point inversely proportional to number of total songs requested by the user. This represents a single even "step" in the playlist between the starting and target songs. We then apply the *k*-nearest neighbors algorithm to this hypothetical point by calculating the distances from the hypothetical point to all songs in the dataset. Using a distance metric, and *k*=1, we determine which single song is most similar to our hypothetical point on along the smoothed playlist. We break ties at random. The chosen song is added to the playlist. This song becomes the target for the next iteration. This process repeats iteratively as songs are added to the playlist one by one. Each added song steps closer and closer to the target in the arousal-valence space.

Results

We tested several common distance metrics, including Manhattan, Euclidean, and Minkowski (*p*=3) distance, cosine similarity, and Jaccard distance. We compared these metrics against a naive baseline of choosing a neighbor at random.

To evaluate the smoothness of an automatically generated playlist, we calculated the mean-squared error (MSE) of the path of the generated playlist against a perfect line from the origin and destination in the arousal-valence space. We randomly generated playlists of various lengths and evaluated their smoothness. We found that cosine similarity consistently generated the smallest MSE. Over repeated trials, this led to the smoothest line through the arousal-valence space compared to the other distance metrics (see Figure 2). Following cosine similarity, the Minkowski distances (p=2 and p=3) and Jaccard distance metric performed the next best. Surprisingly, the Manhattan distance was often worse than the baseline strategy of choosing a neighbor at random.

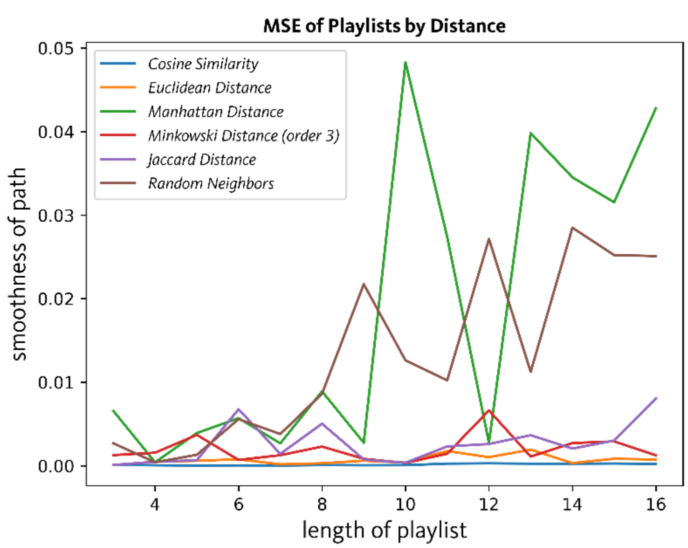

Figure 2. Comparison of the smoothness (MSE) of a playlist path for

different length playlists across different distance metrics.

Demonstration

To demonstrate our approach, we walk through an example. First, we decide we want a playlist of 15 songs. Next, we choose a starting song. We select “How can I help you say good-bye” by Patty Loveless. For this song *valence* = -1.64 and *arousal* = -0.46, which indicts the song is perhaps sad or depressing. Next, we chose a target song: “1963” by Rachael Yamagata. This song has coordinates *valence* = 1.08 and *arousal* = 1.02, which indicates an excited or happy mood.

Using our approach, we generate the playlist shown in Figure 3. This playlist is a very smooth path from the affective state of the starting song to that of the target song. It begins with a few sad songs, transitions through several neutral songs in the middle, and concludes with several happy songs. We can visually observe this playlist’s smoothness, as it closely approximates a straight line through the arousal-valence space (see Figure 3). Although the step size between songs may differ, they remain relatively consistent. This ensures that no individual song differs too dramatically from its predecessor in the playlist.

We provide an interactive demonstration of our automatic playlist generation approach at http://soundbendor.org/playlist/ihiet22/. At this site, users are prompted to choose a starting and ending song from a list of songs. Next, the user chooses the number of songs to include in their playlist. Once the user clicks the “Create” button, our system will generate and display their playlist. This demonstration also automatically creates a link to listen to the newly generated playlist directly on Spotify.

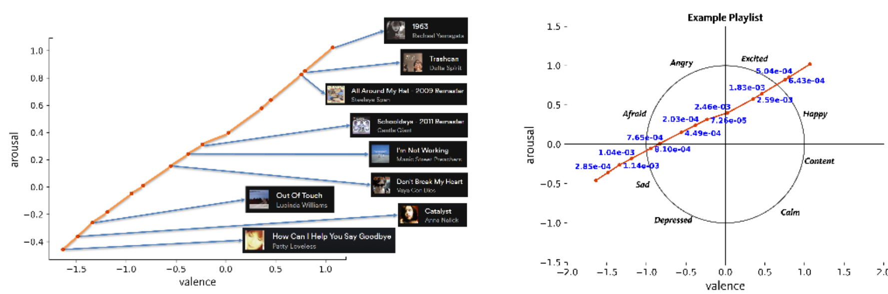

Figure 3. Example playlist with 15 songs that smoothly traverses the affective space (left) and the path of the example playlist shown in the arousal-valence space (right).

DISCUSSION

Although our approach does generate playlists that slowly step through the arousal-valence space, there remain several limitations that we will address with future work.

Sometimes a song in a playlist does not seem inherently musically similar to the preceding song, despite having nearby coordinates in the arousal-valence space. This was not unexpected, since we made no effort to control for the genre or other qualities of the songs at this time. In future work, we will extend our process to include other features, such as acoustics, tempo, lyrics, or even album art.

However, given of the limited number of annotated songs available, the inclusion of these additional features in our similarity metric might limit the number of choices available. Ultimately, this might constrain the diversity of playlists themselves. As a compromise, we intend to use Spotify's genre tags, which would allow creating an emotional journey while at least restricting the playlist to one specified genre.

We recognize that relatively small size of the dataset limits the usefulness of our system, as it cannot extrapolate to the millions of other recordings available in commercial streaming libraries. Using existing dataset of human-annotated arousal-valence values, we are working to build semi-supervised predictive models designed to estimate the arousal-valence values of novel unseen songs, thus expanding the library of annotations and extending the usefulness of this system to real libraries.

Lastly, our current approach relies on the users themselves to specify their current affective state. Unfortunately, a self-assessment may be unreliable or inconsistent. To simplify this process, we plan to add facial recognition to our system that would automatically determine the user's starting emotion using existing datasets of photos annotated with arousal-valence values.

Conclusion

We present a system for dynamically generating playlists that smoothly traverses the arousal-valence space, creating playlists that take the listener on an emotional journey from their current mood to a desired mood.

This system has many potential applications. For example, emotionally dynamic playlists could be useful in music therapy application, allowing listeners to gently guide themselves to happy or peaceful states of mind. Alternatively, an exercise instructor might construct a playlist intended to increase heartrate in conjunction with a specific exercise routine. Likewise, workers and students might use this system to achieve more focused states of mind with a goal of increase productivity.

REFERENCES

Aljanaki, A., Yang, yi-hsuan, Soleymani, M., 2017. Developing a benchmark for emotional analysis of music. PLOS ONE 12, e0173392. https://doi.org/10.1371/journal.pone.0173392

Chen, S., Moore, J.L., Turnbull, D., Joachims, T., 2012. Playlist prediction via metric embedding, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 714–722.

Delbouys, R., Hennequin, R., Piccoli, F., Royo-Letelier, J., Moussallam, M., 2018. Music Mood Detection Based On Audio And Lyrics With Deep Neural Net.

Deng, J.J., Leung, C., 2012. Emotion-based music recommendation using audio features and user playlist, in: 2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012). IEEE, pp. 796–801.

Dureha, A., 2014. An accurate algorithm for generating a music playlist based on facial expressions. International Journal of Computer Applications 100, 33–39.

Flexer, A., Schnitzer, D., Gasser, M., Widmer, G., 2008. Playlist Generation Using Start and End Songs, in: Proceedings of the Symposium on Music Information Retrieval (ISMIR-08). pp. 173–178.

Ikeda, S., Oku, K., Kawagoe, K., 2017. Music playlist recommendation using acoustic-feature transition inside the songs, in: Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia. pp. 216–219.

Kim, H.-G., Kim, G.Y., Kim, J.Y., 2019. Music recommendation system using human activity recognition from accelerometer data. IEEE Transactions on Consumer Electronics 65, 349–358.

Liu, H., Hu, J., Rauterberg, M., 2009. Music playlist recommendation based on user heartbeat and music preference, in: International Conference on Computer Technology and Development. IEEE, pp. 545–549.

Pao, T.-L., Cheng, Y.-M., Yeh, J.-H., Chen, Y.-T., Pai, C.-Y., Tsai, Y.-W., 2008. Comparison between weighted d-knn and other classifiers for music emotion recognition, in: 3rd International Conference on Innovative Computing Information and Control. IEEE, pp. 530–530.

Pauws, S., Verhaegh, W., Vossen, M., 2006. Fast Generation of Optimal Music Playlists using Local Search. pp. 138–143.

Rumiantcev, M., Khriyenko, O., 2020. Emotion Based Music Recommendation System, in: Proceedings of Conference of Open Innovations Association (FRUCT). pp. 639–645.

Russell, J.A., 1980. A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178. https://doi.org/10.1037/h0077714

Sakurai, K., Togo, R., Ogawa, T., Haseyama, M., 2021. Music Playlist Generation Based on Graph Exploration Using Reinforcement Learning, in: 3rd Global Conference on Life Sciences and Technologies (LifeTech). IEEE, pp. 53–54.

Warriner, A.B., Kuperman, V., Brysbaert, M., 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods 45, 1191–1207.

Zhang, K., Zhang, H., Li, S., Yang, C., Sun, L., 2018. The PMEmo Dataset for Music Emotion Recognition, in: Proceedings of the 2018 ACM on

International Conference on Multimedia Retrieval. ACM, Yokohama Japan, pp. 135–142.