605.744: Information Retrieval
Emotion Extraction From Lyrics

Sabbir Ahmed

December 11, 2022

Contents

Abstract

Mood classification in music has become more prevalent with the growing streaming industry. Categorizing songs based on the perceived emotions allow music streaming services to improve their recommendation systems and automatic playlist generation. Services such as Spotify use audio features such as duration, energy, tempo, etc. and other combinations of audio features such as danceability and instrumentalness to group similar tracks. These classifications focus solely on the audio production of the songs while ignoring the lyrical content. Emotion extraction from text can be a difficult task, due to the subjectivity in quantifying or discretely categorizing emotions. In this project, 2 of the popular models of emotions, the Plutchik's Wheel and the VAD model, have been used to attempt emotion extraction.

1 Introduction

This paper discusses the methods used to

2 Background

2.1 Natural Language Processing on Lyrics

Natural language processing on English song lyrics assume additional restrictions. There are no standards set for preprocessing lyrics, but the following lists the exceptions address for this project:

- Songs may be entirely composed of stopwords.
- Repetition is considered significant.
- Lyrics may contain songwriting directions, such as "[gang vocals]", "[instrumental]", "[hook]", "[Speaker A]", etc.
- Lyrics may contain adlibs. For this project, adlibs are not considered part of the dictionary.

2.2 Sentiment Analysis

Sentiment scores are evaluated as the ratios of weights of the relatively positive and negative terms in a document. Context is typically emphasized to increase the accuracy of the perceived sentiment. This is achieved by using long n-grams. This project tokenizes the documents as 1-grams, and therefore loses contexts. The sentiment scores are instead evaluated as the ratio of the positive terms (p) and negative terms (n) over the total number of terms. Terms are labeled positive or negative if they are found in their corresponding datasets.

$$S = \frac{|p| - |n|}{N} \tag{1}$$

If a document contains more negative terms, then the sentiment score will naturally be negative as well, and vice versa for a positive score.

2.3 Emotion Classification

The field of emotion classification is highly subjective. It is often difficult to quantify with absolute values or discretely classify emotions. There have been several psychological models developed over the decades. This project utilizes 2 of the popular models, the Plutchik's Wheel of Emotions and the VAD Emotional State Model.

2.3.1 Plutchik's Wheel of Emotions Model

Psychologist Robert Plutchik proposed a model composed of eight primary emotions: anger, fear, sadness, disgust, surprise, anticipation, trust, and joy [1]. The model, pictured in Figure 1, consists of several concentric circles, with the outer circles being combinations of emotions from the inner circles.

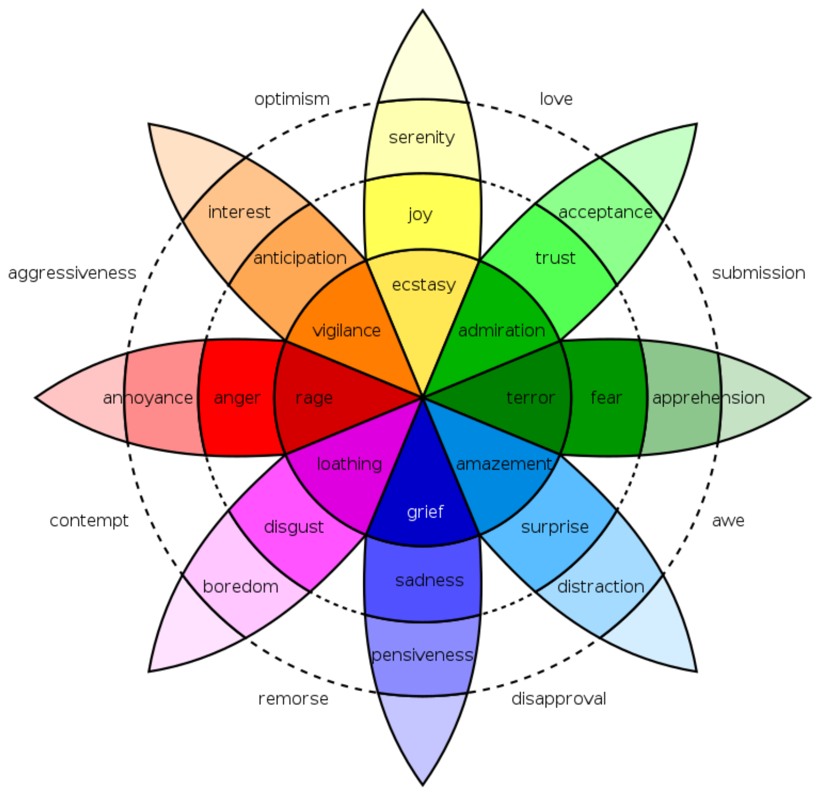

Figure 1: Plutchik's Wheel of Emotions Model

Several categories described by the model do not translate well over text. For example, a human reviewer may find it difficult to extract emotions of *trust* or *anticipation* from a document without explicit usages of synonyms of such emotions.

2.3.2 VAD Emotional State Model

The VAD (Valence-Arousal-Dominance) Emotional State Model was proposed by psychologist Albert Mehrabian. The model plots emotional states across these 3 dimensions of emotion. Valence measures how pleasant or unpleasant an emotion is, arousal determines the energy of the emotion, and dominance refers to the sense of control over the particular emotion. The model, pictured in 2, implies a more granular approach to categorizing emotions.

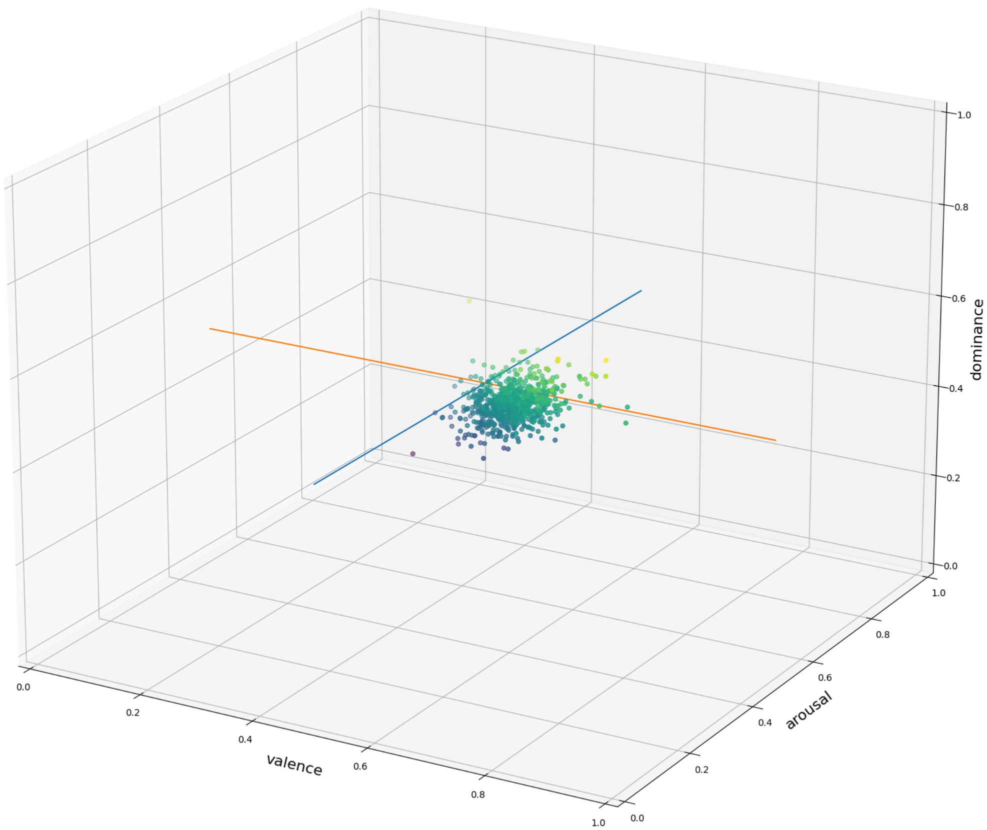

Figure 2: VAD Emotional State Model

The third dimension of the model can be disregarded to describe the more popular Valence-Arousal Emotional State Model (also known as the Circumplex Model), developed by psychologist James A. Russell. The two dimensions of this model allows for emotions to be categorized into quadrants which are sufficient in determining the general sentiment of the emotion. The four quadrants can be labeled as:

- **Quadrant I**: High-arousal/positive-valence, "joy"
- **Quadrant II**: High-arousal/negative-valence, "anger"
- **Quadrant III**: Low-arousal/negative-valence, "sadness"
- **Quadrant IV**: Low-arousal/positive-valence, "calm"

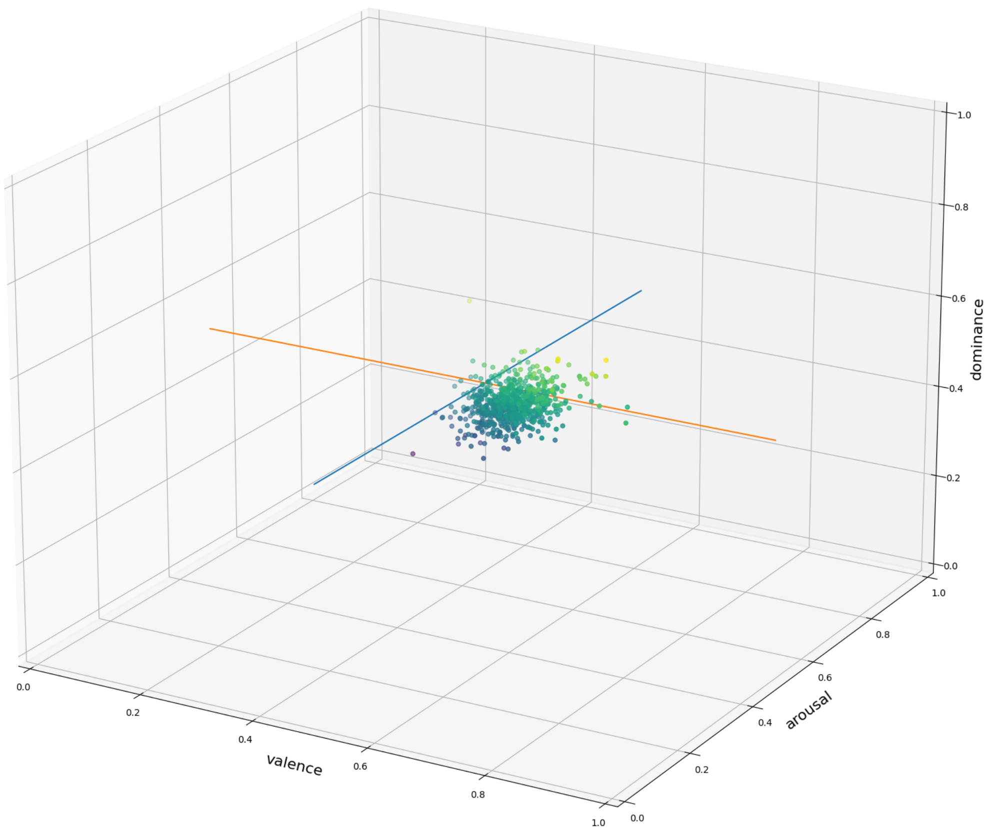

Figure 3: VAD Emotional State Model

3 Source Datasets

The datasets for this project were obtained from various sources.

3.1 Playlist Datasets

Playlist datasets were obtained from Spotify via their web application programming interface (API) [2]. Top playlists curated by the Spotify staff were randomly selected, and the API was used to gather playlist names, lists of tracks, and the apparent targeted moods.

3.2 Lyrics

The lyrics of all of the tracks from the playlist datasets were obtained via scraping from various web sources. In total, 748 documents were obtained for analysis.

3.3 NRC Emotion Lexicons

The word associations, emotion intensities and the VAD lexicons for the emotion models were obtained from the National Research Council Canada [3, 4, 5]. The dictionaries for the positive and negative terms required to compute the sentiment scores were also extracted from these datasets.

4 Scores

The scores generated are based on sums of the emotion intensities and frequencies of word associations.

4.1 Emotion Intensity

Emotion intensity scores are generated for the 8 emotions identified by the Plutchik's Wheel Model. The scores are simply the sums of the corresponding emotion intensity values present in the NRC datasets. To normalize the intensity scores, each of the 8 emotions also have their corresponding intensity ratio scores, which are calculated by computing the ratio of the emotion intensities over the total sum of intensities. Equation 2 defines the ratio for all emotions e in the 8 emotions identified by the Plutchik's Wheel Model, E.

$$e_{ratio} = \frac{\sum_{i=0} e_i}{\sum_{e \in E} \sum_{i=0} e_i} \tag{2}$$

4.2 VAD Scores

Similar to the emotion intensity scores, the valence, arousal and dominance scores were evaluated by summing their intensities and averaging them over the number of terms identified in the word associations. Since every term in the dataset are associated with a 3-tuple of valence, arousal and dominance values, their sums are averaged by the same value. Equation 3 defines the ratios.

$$v = \frac{\sum_{i=0}^{N} v_i}{N}, a = \frac{\sum_{i=0}^{N} a_i}{N}, d = \frac{\sum_{i=0}^{N} d_i}{N} \tag{3}$$

5 Analysis

Table 1: No transform

Emotion	**mean**	**median**	**mean**	**median**	**mean**	**median**
anticipation	1.730709	1.5940	3.719128	2.8350	1.730709	1.5940
disgust	1.069293	0.7200	2.601124	1.1640	1.069293	0.7200
anger	1.648021	1.2190	3.986169	1.8750	1.648021	1.2190
trust	2.476561	2.1790	6.776188	4.9530	2.476284	2.1790
sadness	1.790973	1.5090	3.842015	2.6960	1.701007	1.4210
joy	2.392433	2.2070	6.319867	4.4240	2.392433	2.2070
fear	1.888261	1.5315	3.869922	2.4785	1.888261	1.5315
surprise	0.851107	0.7420	1.840037	1.0710	0.851107	0.7420

Table 2: ratio

Emotion	**mean**	**median**	**max**
joy_ratio	0.216300	0.218978	0.668980
trust_ratio	0.186233	0.187097	0.638021
sadness_ratio	0.121942	0.114320	0.457091
surprise_ratio	0.057528	0.052510	0.407974
anticipation_ratio	0.138459	0.138032	0.456406
anger_ratio	0.101198	0.090794	0.694792
disgust_ratio	0.065279	0.057565	0.316643
fear_ratio	0.113061	0.106388	0.409779

5.1 Emotion Playlist

$$acc = \frac{\sum_{p \in E^+} \frac{\sum_i [p_i > 0]}{|p|} + \sum_{n \in E^-} \frac{\sum_i [n_i < 0]}{|n|}}{|E|} \tag{4}$$

Table 3: ratio

Metric	**Tracks Categorized**	**Accuracy**
25%	747	0.544
50%	747	0.643
75%	730	0.747
mean	747	0.673

Table 4: ratio

Metric	**Tracks Categorized**	**Accuracy**
25%	626	0.572
50%	626	0.719
75%	603	0.820
mean	626	0.713

Table 5: ratio

wheel_playlist	**count**
fear	490
disgust	474
anger	470
sadness	470
anticipation	469
trust	461
surprise	458
joy	444

References

[1] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[2] "Web api reference: Spotify for developers."

[3] S. M. Mohammad, "Word affect intensities," in *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, (Miyazaki, Japan), 2018.

[4] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[5] S. M. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words," in *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, (Melbourne, Australia), 2018.