

Applying a Fuzzy Approach to Emotion Classification of Song Lyric Summarizations

CS 585, UMass Amherst, Fall 2017

Dani Mednikoff & Jen Zhu

1 Outline

In our project, we focus on performing emotion analysis and classification on a corpus of summarized song lyrics. Emotion classification on songs could potentially make it easier for listeners to find songs they are interested in based on their moods. Despite the large quantity of research conducted on emotion analysis and classification, not nearly as much has been done using song lyrics as the sole input for emotion/mood classification. One past technique, for example, was to analyze the lyrics line by line. Our project goal is to also analyze songs via lyrics but in an alternate method. The idea is twofold; first we aim to take the lyrics of a particular song and create a short summary from those lyrics. We then plan to perform emotion analysis and classification on the previously generated summary. Various techniques we will use to achieve this include the valence-arousal model, fuzzy classification, and psychological feature vectors.

2 Literature Review

Paper [4] Discusses the main reason why we chose to focus on solely using song lyrics to classify songs based on emotion/mood. Since there is not much research in this field we would like to explore and build upon previous research. The paper discusses two important challenges faced when developing a lyric-based approach - feature extraction and schemes used for labelling emotions are not trivial, since we are looking at complexities that need to disambiguate affect from text. Finding ground-truth labels that accurately reflect the emotion conveyed in a group of interconnected words is another challenge. The paper offers insight into several approaches to lyric based classification, such as fuzzy clustering or refining feature vectors by only including sentiment/sentiment related words.

Paper [1] introduces a model of analyzing lyrics of a song line by line. Their approach was to analyze the song lyrics based on a lexicon with emotion units. To deal with multiple sentences in an individual lyric they used fuzzy clustering. An extremely important takeaway from this paper was their use of Russell's model of mood, also known as the circumplex model of emotion. They used the model to classify emotions basing them off of valence and arousal scores, which seems like an excellent way to do so and one we would like to apply to our work in the same manner.

Paper [2] provided insight into multiple previous successful summarization techniques, while also describing their approach into improving and building upon some of those techniques. They look into which might provide better results when comparing the discourse-based approach and the frequency-based approach. In this case their DimSum system attempts to combine and improve both. First, it looks at the frequency of what they call "signature words", which they got from analyzing semantics on words with situations like looking alike but with different meanings. This different from typical techniques like using a single word and looking at their frequencies which happens to introduce a lot of noise. Additionally, the system also uses a large corpus to acquire the domain knowledge where in knowledge based summarization they use conceptual representations. From their new approach we are now aware of subtle issues with summarization techniques and also know of possible ways to avoid or fix them. We can use this knowledge when looking for summarizers to use for our project in order to produce the best possible output to use for classification.

Paper [3] describes the process of music identification from lyrics. The researches introduce us with

a potential dataset - Allmusic.com, where human experts classify songs into 183 moods. Later on, we could evaluate the accuracy of our model by comparing our results to that of the Allmusic.com dataset. Additionally, the paper discusses a novel approach to classification - rather than use a standard bag of words, the researchers chose to focus on psychological features of text, using the General Inquirer. Using a bag of words can be a less effective transformation to reduce high dimensionality of the feature vector.

Paper [5] discusses two detailed approaches to implementing fuzzy clustering for song emotion classification. Once again we see Thayer's model for the description of emotions. An interesting distinction the paper mentions is the model generator, which will construct a model based on the testing set, and then the emotion classifier, which relies on the model to build classify different inputs. Two fuzzy classifier methods mentioned are: Fuzzy k-NN (FKNN) classifier, and the Fuzzy Nearest-Mean Classifier. Additionally, the paper suggests reducing the amount of weak features by applying a feature selection technique such as stepwise backward selection. Overall, the paper outlines the classifiers and methods we will use in our project.

3 Approach

We plan to use a dataset of songs and their lyrics from a predetermined dataset, as well as NLTK to generate our summarization of the song lyrics. From there, we will implement fuzzy clustering. The categories available to us will be from the Thayer's circumplex model of emotion, using valence-arousal. Our main quadrants on this model are: Happy, Sad, Angry and Calm. From there, we plan to build our features potentially using the general inquirer and psychological features. From there, our plan is to use both the Fuzzy k-NN classifier, and the Fuzzy Nearest-Mean classifier to generate which valence-arousal quadrant each song falls under. Afterwards, a potential implementation to reduce the number of weak features could be to implement a stepwise backwards selection.

Since both of our potential datasets contain over 200,000 song lyrics, we plan on reducing our training and testing sets to a percentage of the potential dataset. Our focus will be to use 10% of the dataset and split these into our training and test set. We will have an 80/20 split - 20% of the songs will be in our training set, and 80% will be for the development/testing set. We can adjust these values accordingly depending on expectations of difficulty/ease.

Preliminary Experiment (for Progress Report):

We plan to build a sentiment-like version of our project initially, focusing primarily on two valence-arousal quadrants - happy and sad. Our dataset training dataset will contain 20% of songs from the potential dataset, and our testing dataset will contain the remaining 80%.

After our preliminary experiment, once we are able to determine which quadrant on the circumplex model of emotion a song is in, we can then expand the reach of the project. Through the usage of the Fuzzy k-NN and Fuzzy Nearest-Mean classifiers, we aim to develop a baseline algorithm for classifying two quadrants. By using this initial algorithm, we aim to incorporate the remaining two quadrants (anger and calm) into our song classification.

4 Possible Datasets

When thinking of possible datasets that we can use we have to think of a reasonable amount of data. Obviously every song ever created would be much too large of a dataset to analyze. Our initial thought is to find a dataset that limits the data to a specific year or another attribute that will narrow down the data set but still keep it large enough, rather than having to create our own. It is also necessary that the dataset contains a range of all genres, to eliminate bias.

Here are a list of the current datasets we have encountered, which appear to be pertinent to our goal:

1. MetroLyrics Dataset: <https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics>
 - The data contains information about the song, year released, artist, genre and lyrics.
2. MusixMatch: <https://labrosa.ee.columbia.edu/millionsong/musixmatch>
 - A bag of words approach to lyrics, dataset seems to index each word in associated song lyrics.
 - Also comes with separate training and test datasets.

At the moment, we plan to move forwards with the MetroLyrics Dataset, since it requires less parsing as compared to the MusixMatch dataset. Since the MusixMatch dataset appears more intricate and complex than we expected, it appears we would need extra time to parse through the data. Since this is not the focus of our project, we have decided the MetroLyrics dataset aligns more with our project goals.

5 Scope

The scope of our project is to be able to create a summary from the provided lyrics or dataset and then using that summary classify that song of portraying a certain emotion. It is heavily based on whether or not the summary is able to accurately represent the lyrics themselves. Additionally, it will use psychological features in order to build a valence-arousal category during classification. Aside from the issue of our summary being accurate or not, it is also important to emphasize our focus is to classify the song into one quadrant, although we are well aware it is possible for a song to possess multiple emotions in it.

6 Pre-existing Software Systems that Can Be Used

- Implementation of the code: Python
- Summarization: WordNet/NLTK
- Content Analysis: General Inquirer

Along the way, we may have to change or add existing software to this list.

References

- [1] Yajie H., Xiaoou C., & Deshun Y. (2009), “*Lyric-Based Song Emotion Detection With Affective Lexicon and Fuzzy Clustering Method*,”. 10th International Society for Information Music Retrieval Conference, 123-128.
- [2] Chinatsu A., Mary Ellen O., James G., Bjornar L. “*A Scalable Summarization System Using Robust NLP*”. 66-73
- [3] Yang, Dan, and Won-Sook Lee. “*Music Emotion Identification from Lyrics*.”. 2009 11th IEEE International Symposium on Multimedia, 2009, doi:10.1109/ism.2009.123.
- [4] Kim, Youngmoo E., et al. “*MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW*.” International Society for Music Information Retrieval Conference, vol. 11, 2010, pp. 255–260. ISMIR, www.ismir2010.ismir.net/proceedings/ismir2010-45.pdf.
- [5] Yang, Yi-Hsuan, et al. “*Music Emotion Classification*”. Proceedings of the 14th Annual ACM International Conference on Multimedia - MULTIMEDIA '06, 2006, doi:10.1145/1180639.1180665