Project Proposal: Playlist Generation Using Twitter Account Post History and Sentiment Analysis

Emily Parrish emparrish93@gwu.edu
Shiva Omrani somrani@gwu.edu
Vedant Shrivastava vedant_s@gwu.edu
Marcelino Velasquez mav20@gwu.edu

1 Introduction

In our project, we propose to develop a playlist generation capability inspired by a user's or entity's social media posting history. In this proposal, we present the data sources, projected models that could be used, in addition to metrics and validation methodologies.

2 Goals

1. Collect sample data using web scrapers and Twitter/Spotify's Application Programming Interfaces (API).

2. Perform binary/multi-class sentiment analysis on Twitter account data (tweets) and music lyrics from songs and independently validate these labels.

3. Use the distributions of the polarities in the respective corpora (account for Twitter, songs/-playlists) to connect the input account to a playlist of appropriate songs according to sentiment.

4. Potentially integrate this capability into a user interface or applet.

3 Proposed Project Plan

The process of achieving these goals consist of five sections, starting with data acquisition. Then our plan is to pre-process all of the data for sentiment analysis, which potentially will involve different techniques for the song lyrics and Twitter data. For example, the Twitter data might have supplementary features like symbols, numbering, punctuation and URLs while the music data (depending on the file format) might contain new line characters.

Figure 1 breaks down the proposed project plan, including datasets, model implementation, and outputs. The subsequent sections will describe some of the specific methods and sources in more detail.

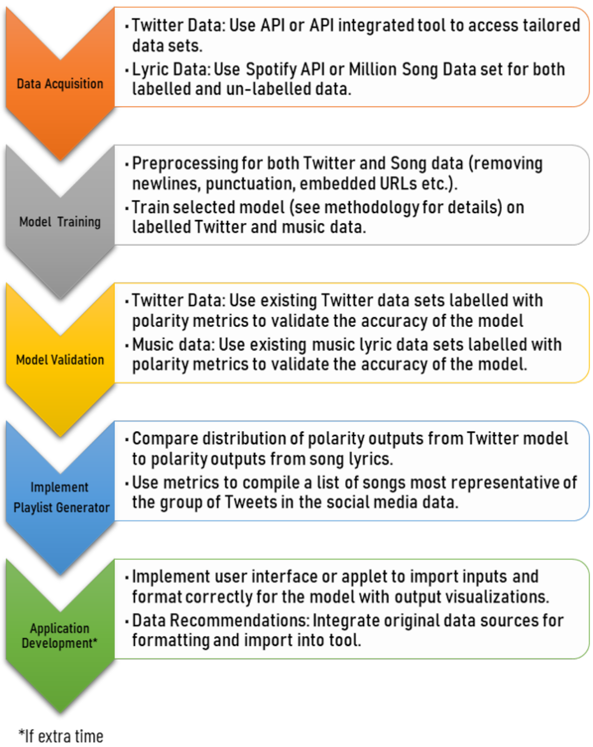

Figure 1: Proposed Project Plan

4 Methodology

As already mentioned, the goal of our project is to recommend music to users based on their tweets. To do so, we need to perform sentiment analysis on a user's Tweeter feed to gain knowledge about their mood. Also, we need to perform sentiment analysis on a number of songs to match them with a user's general emotional state.

4.1 Sentiment Analysis of Twitter Data

We will explore applying deep convolutional neural networks and recurrent neural networks, specifically LSTMs, to Twitter sentiment analysis. As Kim suggested, a simple CNN with a single convolutional layer can produce excellent results for sentence classification (Kim, 2014). Furthermore, bidirectional LSTMs have been popular for sentence classification for their ability to not only learn long-term dependencies but also read the sentences in two directions. Since sentiment analysis of tweets can be a special type of classification at the sentence level, we will focus on the literature's two popular techniques: CNNs and LSTMs.

For sentiment analysis on Twitter data, we intend to closely follow the framework proposed by Cliche who ranked first in SemEval-2017 Twitter competition (Cliche, 2017). By combining CNNs with RNNs (LSTMs), he leverages the advantages of both models and achieves state of the art results for Twitter sentiment classification. His framework consists of following steps:

1. Training initial word embeddings using a neural language model

2. Further tuning the embeddings using distant supervision

3. Training CNNs and RNNs on the labeled SemEval-2017 Twitter corpus using the embeddings and parameters obtained in the previous step.

We will be using word embeddings to represent the underlying text of each tweet. In other words, each tweet will be represented as the set of the word embeddings of all of its constituent words. We will obtain these embeddings by training a neural language model on a large, unlabelled corpus of Twitter data. Then, we will feed this data to a convolutional neural network to further tune the embeddings as in this state they do not encode much useful sentiment data.

As Tang et al. mention, traditional word embedding algorithms "only use the contexts of words but ignore the sentiment of texts" (Tang et al., 2016). Therefore, to make the embeddings more sentiment-aware, we need to further tune them. To do so, we will initialize a CNN with parameters obtained from the previous step. We will train this CNN using a distant supervision approach; we will train it on a dataset of tweets that are labeled positive if they contain a happy emoticon and are labeled negative if they contain a sad emoticon.

Finally, we will use the embeddings and parameters obtained from the distant supervision training step and feed them to initialize CNNs and LSTMs. We will "ensemble 10 CNNs and 10 LSTMs through soft voting" (Cliche, 2017). These models will be trained on the training portion of the labeled SemEval-2017 Twitter corpus. The resulting ensemble model will be evaluated by its ability to classify sentiment of the tweets in the test set. After validation, the model is ready to be used for classifying sentiment of tweets so that they can be matched with songs in the next step.

It is important to note that although we aim to explore both CNNs and LSTMs for tweet classification, we may reduce our framework to only taining CNN's without ensembling them with LSTMs if we find the task to be too challenging. Additionally, we will likely be focusing on classifying sentiment of tweets based on whether they generally contain a positive or negative sentiment. Therefore, our classification task will be on a 2-point scale.

4.2 Sentiment Analysis of Songs

"Sentiment analysis of songs is about utilizing data mining or AI techniques in combination with different features to correctly classify songs in mood categories based on the most typical emotion types they express" (Çano, 2018). According to Çano, there are several plausible features that go into song emotion classification, such as genre and epoch, but the most popular feature types in the field are tags, lyrics, audio, and a combination of these (Çano, 2018). Lyrics are a commonly utilized feature because, unlike audio, they are exempt of copyright laws and are found free on the internet. Furthermore, they naturally contain crucial information for accurately classifying a song. Researchers have approached using lyric features with two methods: lexicon-based and corpus-based.

In short, tags are loan words like "mellow", "sweet", "80s", etc. that are useful for conveying the mood, instrument, or genre of a song. They are favored because they are easy to work with yet full of information. They are usually seen as a consensus by listeners on the traits of a song, making them somewhat reliable.

Lexicon-based lyric features attempt "to predict emotionality by mapping lyrics words with their synonyms in affect lexicons." (Çano, 2018) An affect lexicon is a collection of emotion-related words classified as positive, negative, or neutral/ambiguous and is further categorized into subgroups of emotions such as joy, fear, pain, and others. Çano provides references to researchers who used ANEW, an affect lexicon, to categorize the emotions of the refrains and intros of song lyrics. This is an unsupervised learning approach to song emotion classification unlike its counterpart.

Corpus-based lyric features use collections of mood-annotated lyrics to train models (Çano, 2018). Word oriented metrics like tf-idf vectorization and term frequency have been shown to process the lyrical feature well enough to train traditional classifiers to be able to predict the emotion of unlabeled songs. Being a supervised learning solution, this requires the lyrics to already be mood-annotated by an independent source, typically other humans, making this somewhat difficult to replicate on a large-scale.

Çano goes into detail about the difficulties of identify emotions in song, an example being that song emotion is inherently subjective and based on listener perception. Each feature mentioned before also has their respective drawback. Tags lack a common vocabulary, which can lead to ambiguity, and require careful preprocessing since they can also contain noise, or irrelevant feedback (Çano, 2018). Lyrics are not universal since songs that are purely instrumental or classical do not contain them. Thus, researchers typically combine these features, including audio, to achieve acceptable results.

For this project, experimentation with both lyric methods, audio, and tags will be necessary to achieve the right type of model for our usage.

5 Datasets

In this project, we will need to access two types of data to accomplish the playlist generation goals: social media (Twitter) data and music (Spotify) lyric data. In the following section, we have outlined potential data sources for acquiring this data.

5.1 Twitter Data

As explained in section 5.1, our training framework requires both labeled and unlabeled datasets. The unlabaled dataset will be constructed by using data mining from Twitter API. For the labeled dataset that is required for step 3 of our framework, we will be using the training portion of the human-labeled dataset provided by SemEval-2017 Twitter competition. The dataset can be found here: `http://alt.qcri.org/semeval2017/task4/index.php?id=results`

Another avenue to extract Twitter data that we as the developers can tailor easily, is utilizing George Washington University's Social Feed Manager (`https://sfm.readthedocs.io/en/latest`). This tool not only uses Twitter's API to access tweet data, but also can get post information from Tumblr, Flickr, and Weibo. Using this functionality would allow us to tailor certain accounts'b tweets to our dataset, specific searches, and even sample subsets of these tailored datasets. The data is also pulled from Twitter in real time, so we could even do some temporal studies and look at the changing sentiment based on the addition of data for a particular account.

5.2 Song Lyric Data

To acquire the song lyric data, we will make use of the musiXmatch dataset in order to obtain the collection of song lyrics in bag-of-words format. Each track is described as the word-counts for a dictionary of the top 5,000 words across the set. Although copyright issues prevent this dataset from distributing the full, original lyrics, we hope and believe that this format is sufficient for our academic purposes. The dataset comes in two text files, describing training and test sets. It can be found at `http://millionsongdataset.com/musixmatch/`.

6 Baseline

Çano's paper (Çano, 2018) on text-based sentiment analysis for music emotion recognition provides potential, as well as tested, base architectures for accomplishing our goals. Most of the suggested base architectures are inspired by the record-breaking CNNs seen at the yearly ImageNet challenge, as well by networks using RNN features such as LSTM in studies regarding text analysis. The following are possible neural architecture designs that we will experiment with for our purposes.

6.1 NgramCNN

Extending from the designs of image recognition CNNs, this network mainly consists of a selected number of parallel convolutions of growing kernel lengths and max-pooling layers that select the maximum value from each feature map region. These layers are repeated several times depending on the training data set, and end with the concatenation of the output features of the last max pooling layer for feed-forward classification (Çano, 2018). There are also variations of this model that are further discussed in this paper.

6.2 BLSTM-2DCNN

A bidirectional LSTM extended by a two-dimensional CNN with pooling layers, this architecture is known to score high accuracy in sentiment polarity analysis. It should be noted that these datasets are usually small with short documents, which is favorable to RNNs.

6.3 Optimized LR or SVM

Machine Learning Classifiers like Logistic Regression and Support Vector Machines enhanced with grid-search regularization parameters (Çano, 2018) have also been shown to perform efficiently in sentiment polarity analysis, even outdoing CNNs. Being computationally quicker to train, I believe these make good options for the Twitter sentiment polarity aspect of our project.

Despite their differences, most of these architectures utilize the same fundamental concept of "using deep and complex neural network constructs for extracting and selecting features in combination with one or few simple feed-forward layers for classification" (Çano, 2018). Following this formula will be crucial in developing a model that will serve our purposes well.

7 Evaluation Metrics

While Twitter sentiment analysis and song lyric sentiment analysis implementations can be similar, the evaluations of the two are completely different problems. Thus, different manners of validation must be used to verify our project's accuracy and ability. The following is our current plan of evaluation.

7.1 Evaluation of Twitter Sentiment Analysis

As mentioned in section 5.1, we will be evaluating our Twitter sentiment classifier on the evaluation dataset provided by SemEval-2017 Twitter competition. This dataset consists of human labeled tweets for sentiment on a 2-point, 3-point, and 5-point scales. For our project, we will be only focusing on the dataset that classifies tweets as either positive or negative. The datasets can be found in this link: `http://alt.qcri.org/semeval2017/task4/index.php?id=results`.

7.2 Evaluation of Song Lyric Sentiment Analysis

Our research has led us to existing song emotion datasets that can possibly serve for song emotion validation, as well as criteria for finding the most quality data. One potential set is MoodyLyrics4Q, which contains 2000 songs labeled with 4 categories based on a model that can be found in this article (Çano and Morisio, 2017). MoodyLyricsPN, a dataset with 5000 songs labeled as positive or negative, can be found in the same paper. A link to both these datasets is provided here: `http://softeng.polito.it/erion/`.

References

Erion Çano and Maurizio Morisio. 2017. Music mood dataset creation based on last.fm tags. pages 15–26.

Mathieu Cliche. 2017. BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 573–580, Vancouver, Canada. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou. 2016. Sentiment embeddings with applications to sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(2):496–509.

Erion Çano. 2018. Text-based sentiment analysis and music emotion recognition.