CS7IS4 - Group 10 – Midterm paper

Exploration of the similarity of the music and corresponding lyrics

Yuanpei Teng	Peichen Sun	Harsha Vardhan Gajendra Kumar
yteng@tcd.ie	sunpe@tcd.ie	gajendrh@tcd.ie
Haokun Zhang	**Jiawei Shen**	**Barath Nithish Lingasamy**
zhangh9@tcd.ie	shenj3@tcd.ie	lingasab@tcd.ie

Abstract

This paper was written to examine the similarities between the music and lyrics of the same artist: the music-based similarity analysis will be based on the ABC notation format and combined with statistical analysis tools and machine learning methods, the lyrics-based similarity analysis will be based on the text analysis tools. The goal of this paper is trying to prove whether the same musician would maintain a similar style across his different music and lyrics.

Keywords: Music, text analysis, machine learning, ABC format notation, similarity analysis

I. Introduction

Music is an essential part of people's daily lives and there is a wide variety of different genres of music, these different genres of music combine to create a colorful world, and the sentimental conveyed by the music indirectly connect people from different countries and languages. There are times when despite the differences in the types of music, the people who listen to it feel the similar emotions. The lyrics in the different songs may also express the similar meanings. Therefore, multiple studies focused on analyzing the similarities between different genres of music, and several studies focused on analyzing the similarities between the lyrics by using several statistical analysis tools or machine learning methods. However, there are very few studies research on whether there are similarities between songs written by the same composer, for example, same composer may use of similarly rhythmic instruments or similar melodic. And whether there are some similarities between lyrics and music written by the same musician. With the advancement of technology, there are several methods allowed researchers to use to analysis the similarity of music and the lyrics such as ABC notation. ABC notation encoding the music into a text-based format that can be readable by computer and allowed researchers to analysis the similarity of the music through the patterns, structures, and keys by using text-analysis tools. And ABC notation can also convert MIDI files and audio files such as MP3s into a standardized format which provides opportunities for similarity analysis between different genres of music, as well as between music and corresponding lyrics. Therefore, in this study, we aim to investigate the potential relationship between music-based similarity and lyrics-based similarity for the same musician using ABC notation as a framework.

II. Articulation of Research Question

The research question for this study is to understand for the same musician, to what degree of correspondence between musical compositions and corresponding lyrics by using ABC notation as a framework and combined with several machine learning methods to apply similarity analysis. To ensure the persuasiveness of the study, several musicians, along with their songs and lyrics, will be selected for this study, and it will be determined whether there are certain similarities between each musician's music and lyrics.

III. Related Works

1. ABC notation introduction

The structure of the ABC notation includes two parts, tune header and tune body. Tune header contains descriptive meta-data that can be used to distinguish each music, such as the title, the meter of this song and the key signature of this song. Tune body contains the rhythms and note information, possible score instructions, like key signatures, instruments info, time signatures. And tune body may also indicate the repetitive structures of a music from a music theory perspective (Walshaw, 2014). Therefore, convert the music to ABC notation can be better expressed in the form of text, which makes it easier to perform similarity analysis.

2. Lyric-based similarity analysis methods

The application of lyrics related to music information retrieval has been a hot research topic in recent years. (Logan, 2004)The author explored the use of lyrics for automatic music indexing and found that lyrics could be used to discover natural genre clusters. However, they also found that using lyrics to determine artist similarity was better than randomization, but not as good as state-of-the-art acoustic similarity techniques. This suggests that combining different methods may lead to better results.

To analysis the similarity of the lyrics between different songs, Markus Schedl and Peter Knees in their paper (Markus Schedl, Peter Knees, 2009) introduced the way that which first implement PLSA model to the song lyrics to extract the topics of these lyrics. And the next step, for each artist, collected all his lyrics, and created a N-dimensional vectors, and each dimension of vectors indicates the chance that the artist's tracks pertain to the relevant topic. Once got N-dimensional vectors of an artist, we can calculate the L1 distance, which is also called 'Manhattan distance' to compare the similarity between two songs or two artists. And in this section, the author also introduced other ways to compare, for example, using Term Frequency-Inverse Document Frequency to compute the semantic content of lyrics, and using Cosine distance to analysis the similarity.

The use of Word2Vec model for lyrics similarity has been explored in various research works. In the paper (Chandra, 2020),the authors encoded the lyrics as well as trained them on a dataset containing 50,000 rap music lyrics by using the Word2Vec framework and the Continuous Bag Of Words algorithm. An average accuracy of 85% was achieved in three rounds with different dataset sizes.

A paper from (Harish, 2019) also propose an algorithm. The clustering algorithm can combine nodes that possess similar properties (i.e., shingles). The algorithm is stable and can detect outliers. The algorithm has a linear storage requirement for the node representation. The n nodes are randomly sampled from the original data, and they are further partitioned if the algorithm allows parallel execution of data.

Another paper (Sasaki, October) set up a new modelling algorithm LDA which can find out the similarity between different lyrics. Since LDA assigns each word which constitutes lyrics to a different topic independently, the lyrics include a variety of topics according to the variation of words in the lyrics. K typical topics which constitute many lyrics in the database are estimated and a ratio to each topic is calculated for lyrics with unsupervised learning. As a result, the appearance probability of each word in every topic can be calculated. The typical representative word for each topic can be decided simultaneously.

Another study from (Ranganathan, 2011) gives us a scoring method which can get the popularity of each word. The popularity score of a word has been identified from the lyrics. In lyrics, the words are mainly attached to the suffix. So, the root words are taken into consideration for determining their frequency count. The root words are identified using the morphological analyzer.

There has also been some research in recent years into analyzing similarities between lyrics from a sentiment analysis perspective. In the paper (Bao, 2022), the author mainly used the IR emotion ontology library to obtain the emotion vector of lyrics and calculate the similarity between lyrics using the cosine similarity formula, to prepare for the subsequent music recommendation system.

With the rise of the BERT model, variant models based on BERT are also being used in sentiment analysis. In the paper (Kim, T., Vossen, P., 2021),the authors train the model on two popular emotion recognition datasets. The authors emphasize that the inclusion of the speaker's name and contextual information in the dialogue enables the model to learn intra- and inter-speaker states and thus predict the current speaker's emotions. The model can understand and categorize sentiment details in the text better than traditional BERT models. This enables it to identify finer-grained emotion categories.

A study from (Afif Hijra Ferdinan, Andrew Brian Osmond, Casi Setianingsih, 2018) introduced a way to analysis the similarity of the lyrics from the sentimental perspective. In their approach, TF-IDF, which is called 'Term Frequency – Inverse Document Frequency', is applied as a weighting process to help identify the most important keywords or phrases in lyrics, as it will measure the frequency of a word in lyrics and its prevalence across the entire lyrics. By using this method, important words with less frequently will be assigned a larger weight, and common words with most frequently will be assigned a lower weight. KNN method will be used as calculation to classify the lyrics into one of the closest pre-defined sentimental classes based on the Euclidean Distance or Cosine Distance, and the weight of each word that generated by the TF-IDF will form a vector space model and used as a variable in the calculation of the Distance.

3. Music-based similarity analysis methods

Although there hasn't been much research in Music Similarity through ABC notation, there has been some research in finding music similarity through Mel spectrogram and other means related to frequency. (Tzanetakis, G., Cook, P., 2002) delved into the extraction and representation of features for music signals, focusing on pitch, timbre, rhythm, and harmony. (Downie, J. S., Ehmann, A. F., Tcheng, D., 2004) addressed the assessment of the similarity of musical instrument sounds, focusing on feature extraction and representation. Their method employs Mel-frequency cepstral coefficients (MFCCs), derived from Mel spectrograms, to capture and compare the acoustic characteristics of musical instruments.

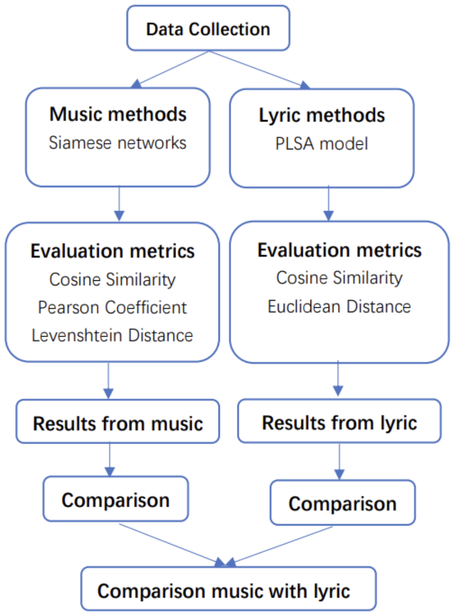
(Simonetta, F., Carnovalini, F., Orio, N., et al, 2018) explored symbolic music similarity using a graph-based representation. In their work, they introduced a novel approach that leverages graphs to model the structural relationships within symbolic music data. Music is represented as a sequence of segments, allowing for the highlighting of contrapuntal relationships and the reduction of each part by progressively deleting the least relevant notes. This approach is designed to capture both the harmonic and contrapuntal textures of polyphonic works. The method is considered feasible because it provides a detailed representation of music that includes both contextual and harmonic information, which is essential for capturing the complexity of musical similarity beyond mere melodic comparison.

(Berenzweig, A., Logan, B., Ellis, D. P. W., et al, 2004) conducted a comprehensive study, wherein they evaluate acoustic and subjective music similarity measures by comparing their performance on a database of 400 popular artists. It uses acoustic techniques based on Mel-frequency cepstral coefficients (MFCCs) and subjective techniques that include data from The All-Music Guide, surveys from playlists, personal collections, and web-text mining.

(Kasif, G., Thondilege, G., 2023) in their work offers a fresh approach to finding creative analogies in music, assisting artists in avoiding unintentional copyright infringement. A Siamese Convolutional Neural Network (CNN) model trained on the Who Sampled dataset—which includes audio samples, remixes, and cover songs—is used in the study. The model's capability to distinguish between similar and dissimilar data points. is shown by a weighted similarity score that considers Pearson correlation, cosine similarity, and Euclidean distance. Likewise, (Uzan, L., Wolf, L., 2015) presented a fresh technique for telling voice actors apart, even when they purposefully change their voices to sound like different people. A dataset of 29733 utterances from 19 hours of speech captured by 114 performers who portrayed 647 characters was used in the study. Spectrophotometers were utilized to identify the speaker using a Convolutional Neural Network (CNN).

IV. Methodology Proposed

Methodology flow chart



Data Collection

In our approach, the lyrics dataset for this paper will be taken from the API provided by the Genius website, with a selection of 2 genres (Country music and Hip hop music) with 2 popular artists and the lyrics of 20 of their respective songs. The music dataset will be collected from various music collection websites, downloaded in MIDI version, and converted to ABC notation format using a converter. For data pre-processing, data cleansing, word splitting, and stop word removal operations will be performed to ensure the quality of the dataset.

Lyrics Similarity Analysis Method

It is true that there are many models and evaluation matrices that can be used, but the PLSA model and the cosine distance will be one of the most preferred ways to analyze the lyrics, because the cosine distance is also applied in the similarity analysis of the music, and the same computation facilitates the analysis of the similarity between the music and the lyrics that will be carried out later.

Music Similarity Analysis Method

We are planning to use CNN as our model with Siamese networks, which is then given to various layers and the output is calculated. The base CNN model involves designing a Convolutional Neural Network (CNN) to extract features from an input (.wav) file. This network is then duplicated to form the two legs of the Siamese network. The architecture includes convolutional layers with increasing numbers of filters and a small kernel size. The activation function uses ReLU for non-linearity after each layer. Pooling layers are applied after some convolutional layers to reduce spatial dimensions. Fully connected layers are added towards the end of the network to learn non-linear combinations of high-level features. Batch normalization layers are optional to stabilize and accelerate training. The final layer outputs a feature vector, typically between 128 to 512 elements and then similarity score is calculated by using three different ways, cosine distance, Pearson coefficient and Levenshtein distance.

References

Afif Hijra Ferdinan, Andrew Brian Osmond, Casi Setianingsih. (2018). *Emotion Classification in Song Lyrics.* Bandung: 2018 International Conference on Control.

Bao, N. (2022). *Analysis of Music Retrieval Based on Emotional Tags Environment.* Journal of Environmental and Public Health.

Berenzweig, A., Logan, B., Ellis, D. P. W., et al. (2004). *A large-scale evaluation of acoustic and subjective music-similarity measures.* Computer Music Journal.

Chandra, J. S. (2020). *Artificial Intelligence based Semantic Text Similarity for RAP Lyrics.* International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 1-5.

Downie, J. S., Ehmann, A. F., Tcheng, D. (2004). *Evaluating rhythmic similarity measures.* International Conference on Music Information Retrieval (ISMIR), 251-258.

Harish, P. R. (2019). *K-Means and Locality Sensitive Hashing for Song Clustering.*

Kasif, G., Thondilege, G. (2023). *Exploring Music Similarity through Siamese CNNs using Triplet Loss on Music Samples.* International Research Conference on Smart Computing and Systems Engineering (SCSE).

Kim, T., Vossen, P. (2021). *EmoBERTa: Speaker-Aware Emotion Recognition in Conversion with RoBERTa.*

Logan, B. K. (2004). *Semantic analysis of song lyrics.* 2004 IEEE International Conference on Multimedia and Exp(ICME) (IEEE Cat. No.04TH8763). https://doi.org/10.1109/ICME.2004.1394328.

Markus Schedl, Peter Knees. (2009). *Context-based Music Similarity Estimation.* Linz.

Ranganathan, K. G. (2011). *Lyric mining: word, rhyme & concept co-occurrence analysis.*

Sasaki, S. Y. (October). LyricsRadar: A Lyrics Retrieval System Based on Latent Topic of Lyrics. *In Ismir*, 585-590.

Simonetta, F., Carnovalini, F., Orio, N., et al. (2018). *Symbolic music similarity through a graph-based representation.* Audio Mostly 2018 on Sound in Immersion and Emotion, 1-7.

Tzanetakis, G., Cook, P. (2002). *Musical genre classification of audio signals.* IEEE Transactions on Speech and Audio Processing, 10(5), 293-302.

Uzan, L., Wolf, L. (2015). *I know that voice: Identifying the voice actor behind the voice.* International Conference on Biometrics (ICB).

Walshaw, C. (2014). *A STATISTICAL ANALYSIS OF THE ABC MUSIC NOTATION CORPUS: EXPLORING DUPLICATION.* London.

Declaration for work contribution

Name	Student ID	Role	Contribution
Yuanpei Teng	17343768	Chair	1. Reviewed methods for the lyric similarity analysis 2. Found three literatures that referenced in the related work section in the midterm paper. 3. Wrote the Introduction, Articulation of Research Question, part of the related work and part of the methodology proposed sections in the midterm paper. 4. Arranged the weekly group meetings. 5. Organized the final format and content of group midterm paper. 6. Researched multiple literature papers relate to ABC notation format and lyrics analysis. 7. Tried to collect datasets of lyrics and music. 8. Implement PLSA model and Cosine distance for lyric analysis.
Peichen Sun	23333198	Ambassador	1. Reviewed methods for music similarity comparison, focusing on both acoustic techniques and subjective measures. 2. Analyzed the feasibility and detailed processes of different music similarity methodologies. 3. Conducted research on ABC notation, understanding its application in digital music processing. 4. Identified evaluation metrics for music similarity methods, such as Precision at 20 (P-20) and Mean Reciprocal Rank (MRR).
Haokun Zhang	23330168	Monitor	1. Reviewed the methods of the lyrics analysis part. 2. Searched papers on lyrics similarity and the lyrics dataset. 3. Tried to write some code about analyzing the lyrics.
Jiawei Shen	23332898	Recorder	1. Search for literature on lyrics and music 2. Seek methods for calculating lyrical similarity. 3. Try to find datasets of lyrics and music. 4. Finish the lyrics-related parts of the first draft.

Harsha Vardhan Gajendra Kumar	23331187	Accountant	1. Research on existing Methods for Music Similarity found methods related to mel spectrogram and other methods related to frequency. 2. Research on ABC notation, conversion from music to ABC notation, MIDI to ABC notation and music to MIDI 3. Explored evaluation techniques for existing similarities such as Cosine Similarity, Levenshtein distance, and Sequence Alignment. 4. Worked on Siamese Neural Network, and feasibility in our domain to provide ABC notation input. 5. Wav2vec implementation for converting wav/mp3 to ABC notation.
Barath Nithish Lingasamy	23339268	Verifier	1. Took some time to understand the expectation of research and reflecting on research question. 2. Explored the question of music similarity and found out that mel spectrography is an efficient way to evaluate the same. 3. The papers were found based on the usage of ML techniques such as CNN's for finding music similarity. I identified a paper that specifically converts music from sound domain to frequency domain (Siamese CNN) which we will be implementing in the coming weeks. 4. Also found a way to convert music into ABC by first converting .wav to MIDI and then from MIDI to ABC. 5. Established various methods of evaluating the similarity of music including cosine similarity, Chord Progression, Similarity matrices and Patterns of note duration and resets. 6. Implemented an algorithm to calculate cosine similarity. 7. Summarized the papers related to CNNs in the first draft along with the methodology derived from the main paper (Siamese CNN).

Group member signatures

1. **Yuanpei Teng** Yuanpei Teng ____________________ **Date: 27/02/2024**
2. **Peichen Sun** Peichen Sun ____________________ **Date: 27/02/2024**
3. **Jiawei Shen** Jiawei Shen ____________________ **Date: 27/02/2024**
4. **Haokun Zhang** Haokun Zhang ____________________ **Date: 27/02/2024**
5. **Barath Nithish Lingasamy** B@rath ____________________ **Date: 27/02/2024**
6. **Harsha Vardhan Gajendra Kumar** G. HarshaVardhan ____________________ **Date: 27/02/2024**