Abdur Rafay Saleem
University of Central Florida
ab464825@ucf.edu

Abstract—**This paper review is based on the research paper about a new solution for music interpretation that is based on the natural language processing (NLP) techniques. It is structured in three major parts: a discussion and background of the research problem and the paper's technical contributions, a detailed explanation of the suggested methodology, and a personalized critical analysis of the success of these efforts.**

I. RESEARCH PROBLEM

Music interpretation is an important concept for a lot of technologies to work around music. However, due to its lack of clear interpretation and variety of structures it is challenging to interpret musical instruments. Since it is created by humans, just like language, the meaning can vary from item to item and there is no standard agreed by all for determining its meaning. But, the introduction of MIDI files presents potential for new approaches to understand music. MIDI files contain symbolic representation of music content and the instruments and devices used in its composition which provide a better overview of the music's meaning.

The research problem exists in exploring the potential of NLP techniques to learn music's composition and understand its meaning. NLP techniques can provide alternate representations for musical data and interpretations related to gain and performance. The paper proposes a new NLP based technique for music data representation that can perform musical composer classification. This technique learns from the information available in MIDI files to understand what the original artist wanted to convey and compose music that preserves its original message and tone. The paper points out why this technique is different and essential for effective musical processing.

The technical contributions include development of a method for feature vector extraction that is associated with each piece of music to improve classification performance. It also explores the application of NLP techniques for music interpretation especially word/subword segmentation and validates the concept of music data representation. Then it presents comprehensive details of the new methodology and explains its workings. It also demonstrates the utilization of various machine learning algorithms while conducting experiments. The experiments are conducted on a variety of parameters to mimic real world possibilities.

Finally, the work of this paper helps music be measured using scientific measures and encourages future research in the field of music comprehension.

II. TECHNIQUES

The methodology employed in this research has several stages. Each contributes to the overall goal of musical classification. Based on the composers, it is classified into one of five groups. The techniques that are used include feature extraction for musical methods, word/subword segmentation, and the application of multiple different classification models.

MAESTRO Dataset

The MAESTRO Dataset was used to obtain the MIDI files that are utilized in this study. It comprises 200 hours of piano performances collected over a decade from international competitions and all are concert level quality. It contains precise alignments of musical notes and also additional parameters from piano performances such as key striking velocities, pedal position and note duration. Throughout competitions, the MIDI capture and playback system was embedded in all pianos to capture this information in MIDI files with high precision and exceptional quality. The dataset also contains metadata related to the song such as the composer name, year of performance, duration of every music piece, and composition title. Different versions of maestro were partitioned and merged using segment-wise prediction and other datasets to produce large-scale corpus of music. This helps the model during pre-training on different tasks of music understanding such as genre classification, melody completion, style classification, and accompaniment suggestion. The selected dataset contains a huge number of piano arrangements by multiple different music composers which helps in covering a significant number of musical varieties.

Musical Feature Extraction Methods

The first step in the process involves the extraction of musical features from the MIDI files. This is achieved using the pretty_midi library. It is a Python library that extracts musical features from different symbolic representations of music and nicely supports MIDI files. The MIDI files are processed to extract relevant musical features such as start time and end time, the pitch of the note, and velocities of piano-key striking. Another feature is computed called note duration using the extracted start time and end time. This produces each note as a coordinate pair of MIDI integer encoded pitch and duration of the note in seconds. The velocity is ignored from the encoding due to two reasons.

1. The velocity has a high variance and including this will result in a huge number of tuples which can hinder the pattern recognition ability of the model.
2. Ignoring velocity of key striking may produce a bad sounding version of the original but it can still be recognized as the original. Classification and pattern recognition of the music can be based on its notes and their durations.

The concurrent occurrence of musical characters is common since the technique considers the whole polyphonic music piece and not a single channel. Applying NLP here is a practical approach since in some languages this is common. However, the correct ordering of these simultaneous notes was still a problem which needed a tie-breaking technique. The following sequence of actions take place to obtain the relevant representation for the next stage:

1. The MIDI pitch value of each note was used as the classifier for concurrent notes. The tuple pairs are arranged in descending order based on this value.
2. Then the tuples are directly mapped into Unicode characters and put through a SentencePiece algorithm. This algorithm groups the characters of common occurrence into words or subwords. This helps in comparison of commonly occurring notes.
3. Then the Word2Vec approach transforms these previously extracted words and subwords into vectors.
4. Then the vectors of each musical piece are averaged and concatenated with their SD which is the standard deviation vector. This step helps derive a complete vector representation of every whole musical piece.

Hence, the features are encoded into a symbolic NLP-based representation that can be processed by the machine learning models during composer classification tasks.

Classification Models

The research employs a variety of machine learning models to classify music based on its composer. These models include support vector machines (SVM), logistic regression (LR), random forest classifier (RFC), k-nearest neighbors (kNN), and lastly multilayer perceptron (MLP). Each of these models offers unique strengths in handling the classification task, and their combined use provides a comprehensive approach to the problem.

The dataset was divided into training and testing subsets to facilitate the development and evaluation of the models. The labels of the composers were preserved and used in both the training and testing phases. The data is split in a way that 80% of it is used for training the models, and the remaining 20% is reserved for testing their performance.

For the training phase, a technique known as fivefold cross-validation is applied to all the traditional machine learning models, except for the MLP. This technique involves partitioning the training dataset into five equal subsets. In each iteration of the training process, it uses a different subset as the validation set and the remaining subsets are used for training the model.

In the case of the MLP model, a different approach is taken. The performance of the MLP model is evaluated based on the averaged F1-score. The score is a measure of a model's accuracy that considers both its precision and recall and is chosen as the performance metric for the classification task. This metric is particularly suitable for a task that suffers from the problem of imbalanced classification. This problem means that some classes may have fewer instances than others. By giving significant weight to these low instances/values the models' performance becomes uninfluenced by low values. This ensures that the models are evaluated based on their ability to correctly classify music pieces into their respective composers despite having low numbers of pieces for some of them. It was calculated using 200 epochs from the validation set during the model training step. This approach helps to ensure that the parameters of the MLP model converge to an optimal solution.

In summary, the techniques used in this research provide a robust approach to the composer classification of music. It combines a variety of machine learning models, a rigorous training and evaluation process, and a suitable performance metric. This ensures that the classification task is handled in a way that is both effective and fair.

III. CRITICAL ANALYSIS

The area of music retrieval and composition classification is a new field in the era of technology. It faces various challenges and limitations due to its highly variant nature. The research presented in the paper offers a novel approach to musical composer classification using an NLP-based data representation technique. The conducted experiments were quite detailed and compared the result metrics against existing methods such as the extended Wolkowicz method. The results of the experiments demonstrate the effectiveness of this approach.

The proposed method takes into account the standard deviation vectors and outperforms other methods even without the need for fine-tuning. This suggests that the proposed data representation provides impactful features for classification and results in solid and well-generalized machine learning models. It clearly explains the technical details of the methodologies and its different stages such as feature extraction, word vectorization and classification. The validation F1-scores for the traditional machine learning models align well with the scores obtained from the testing dataset. Thus, indicating that the models are not overfitted. The research in the paper also extends beyond the 5-composer classification by demonstrating the ability to classify up to 14 composers with a high F1-Score. This confirms and validates the composer classification pipeline that was followed in this approach and highlights its potential.

However, the paper also opens up avenues for further exploration. One area of interest is the connection between machine and human intelligence. The application of machine learning transformer techniques could potentially contribute to a more concrete interpretation of musical data. For example Dis-Cover AI Minds to Preserve Human Knowledge. Such enhancements could advance music generation and recommendation technologies.

In conclusion, the authors have done a decent job of incorporating knowledge from the field of NLP into music information retrieval and interpretation. It focuses on the most common limitations and challenges at each sub stage and applies statistical solutions to solve and optimize them. Overall, the paper shows outstanding results for composer classification and also serves as a stepping stone toward a thorough understanding of music.