Universal Music Context Representation with Multi-modal Analysis for efficient Retrieval using Parallel Processing Paradigm

MAKARAND VELANKAR, MKSSS'S Cummins College of Engineering,SPPU, India
VAIBHAV KHATAVKAR, College of Engineering Pune, India
DR. PARAG KULKARNI, Kvinna Limited, India

Musical context with metadata tags (data about data such as song title, artist, etc.) and musical contents (such as melody, rhythm, etc.) is a superior combination for efficient music information retrieval (MIR). Representation learning applied to extract and learn relevant features from music representations such as lyrics, symbolic notations, and audio tracks have produced good results individually. A Multi-modal approach with combining features is a better alternative for various applications, as the impact of music is a combination of all musical aspects. Our initial work focuses on lyrics, historical background, composition and audio rendition for nursery rhymes. Analysis of multi-modal information is a step towards identifying important themes that are used to build context. Context is represented using a novel representation of key-value pairs and it is extendable with adding relevant features as per genre or task and useful for various applications in MIR.

ACM Reference Format:
Makarand Velankar, Vaibhav Khatavkar, and Dr. Parag Kulkarni. 2019. Universal Music Context Representation with Multi-modal Analysis for efficient Retrieval using Parallel Processing Paradigm. 1, 1 (November 2019), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Music in different forms of communication such as audio, lyrics, notations, videos, etc. provides useful cues for the music consumers to enjoy and build specific memory and impression. Consumers refer to these cues which are recalled later on, as search terms for searching music. Present music search is mainly meta-data based and does not cover overall contextual information. This paper presents context [1,2] building for nursery rhymes from different sources as a sample case and the process is applicable to any music genre and sources of info. As a representative case, this paper covers multi-modal analysis from music notations, lyrics, audio and historical information available for nursery rhymes [3] to build the universal context. The historical background may or may not be available for every song. Lyrics available play a vital role in meaning to be conveyed and familiarity of language is necessary to understand the context from the lyricist viewpoint. Notation analysis is done to find a key notation sequence which generally represents repeated notation sequence. This repetitive sequence referred to as theme generally represents song title. Audio track analysis attempts to extract different content based musical features such as tempo, energy, amplitude, timbre, music ornamentation. Different content-based features can be extracted using signal processing tools available. The sample features are extracted using librosa which is a python open source library for audio analysis. Feature summaries of audio tracks are stored and are used to build the required context. Meta-data available with a track such as an album name, artists, and the year is also useful to build the musical context along with extracted patterns and feature values.

The universal musical context is built as a combination of the relevant feature values extracted from different available sources. This representation needs to be simple, extendable and useful for fast access and retrieval. Considering the huge musical data across the globe with songs from various genres and the growth rate of new song generation, the need is to process this data using a parallel programming paradigm. From this huge music data, users should be able to extract the required information quickly for different tasks. The tasks can be like a search for songs with specific parameters or features. Representation of universal context is a challenging problem considering the knowledge is extracted from multiple modes of information of music.

This paper proposes a solution to the problem of knowledge representation. The solution proposed is the use of key-value pair which is one of the simplest generic representations. It can be implemented using any programming language and it allows potential expansions without the requirement of modifying the code. Key-value pairs can be further effectively used by the map-reduce algorithm for parallel processing of data. Although the work presented here is for nursery songs as a proof of concept, it is extendable to millions of songs and can be useful for effective music information retrieval. The proposed representation is extensible to accommodate contextual features for evolving music with new modal info and is independent of any music genre.

2 MULTI-MODAL ANALYSIS

A multi-modal analysis is extracting useful information from different sources or representations of the data [4, 5]. For songs or music, different sources of information such as notations, historical background, lyrics, audio tracks, videos, comments by listeners are available as shown in figure 1. Context building can be done using information from all possible sources and extracting required knowledge from it. Music contents and human perception plays a major role in musical context [6]. Different musical patterns can be extracted from audio or notations [7]. Music meta-data available

Authors' addresses: Makarand Velankar, MKSSS'S Cummins College of Engineering,SPPU, Karvenagar, Pune, Maharashtra, India, makarand.velankar@cumminscollege.in; Vaibhav Khatavkar, College of Engineering Pune, shivajinagar, Pune, Maharashtra, India; Dr. Parag Kulkarni, Kvinna Limited, Shivajinagar, Pune, Maharashtra, India.

XXXX-XXXX/2019/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

provides information about the singers, solo/duet/chorus song, composers, lyricists, albums/films, year of release. This information may or may not be available always. This information is used effectively for meta-data based music information retrieval at present. The content-based multi-modal analysis will add power to the present music search and recommendation [8] to make it more effective and user-friendly.

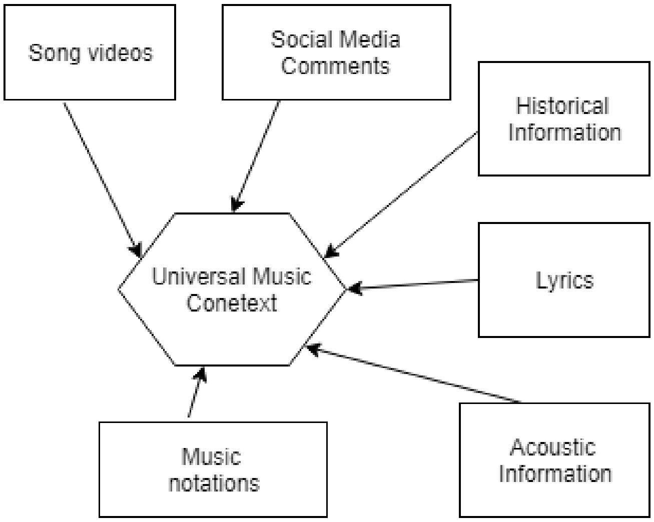

Fig. 1. Multi-modal Music Analysis for Universal Context Representation

The conflation algorithm is used for keyword identification from the text available of the historical information and lyrics data. The algorithm uses 3 major logical steps as removal of fluff words, suffix stripping and identification of equivalent stems. Musical notations are represented in different forms such as sheet music, MIDI, western notations, etc. One can directly use tools available to extract notations from sheet music or use MIDI notations if directly available. Available western notations are used as a source to extract melodic information. Melody is represented as a sequence of melodic phrases. Melodic phrases consist of different notes rendered in a specific order. The auto association for all phrases helped to derive dyads (2 note sequence), triads (3 note sequence) mainly based on the n-gram concept. The majority of the phrases found were involving dyads and triads with occasional use of 4 or more notes in the phrase. The auto association revealed the phrases and sub-phrases within the song useful for building context based on melodic patterns or motifs. Application-specific context can identify a song from the notation patterns and associate it to the specific scale used. The python tool Librosa is used for feature extraction of audio for sample experiments. The features extracted are chroma, spectral and rhythm.

3 MUSIC CONTEXT REPRESENTATION

Music audios are growing at a very fast rate which can be noticed with uploads every day on a website like YouTube. Considering the context for millions of songs to be stored and retrieval/classification on the basis of context, it becomes a typical big data challenge. Possible data storage options explored were SQL databases, semi-structured databases, and No-SQL databases. SQL databases are not considered as they don't provide the required flexibility. Another choice is semi-structured databases such as JSON or XML but considering the enormous volume, these models don't suit the required application. No-SQL databases are a better choice in such requirements. Different variants of No-SQL are explored such as MongoDB, Cassandra, CouchDB. MongoDB which is a document-oriented database based on the key-value model was found suitable for the storage of musical context. MongoDB is a document-based database system that stores data for the particular record in the document form as a set of similar documents in the collection with enhanced performance [9]. Each song's context here is represented by a single document. The advantage of No-SQL representation is that it is flexible and can add any number of attributes/ fields/ keys and they can be different for different documents. Considering songs in different genres, the requirements are different such as for Indian classical music raga and related information or western music chords, mode and related data needs to be stored. The flexibility of No-SQL systems is best suited for such requirements to have a universal context representation of music. Another advantage of MongoDB is easy to interface with map-reduce functionality which is a parallel programming paradigm for fast retrieval [10]. We have stored the combined multi-modal music context as a document in the collection as shown with the following example for one song.
/ " id" : ObjectId (59835448741331ads45dc8), "title": "A Tisket A Tasket" , "type": "nursery rhyme", "year" : 1900, "lyricsK1": "I dropped it", "HistoryK1": "child", "HistoryK2": "play", "Notes1": "GE", "Notes2": "EA", "tempo": 112, "tonnetz": 0.014487, "rmse": 0.12909, "spectral-centroid": 2619.95, "spectralbandwidth": 2529.93, "spectralrolloff": 5522.64/
It shows key-value pairs in the form of "key": "value". We can use the power of MongoDB queries to search for specific key values or filter data as per the requirements. K1, K2 denotes keywords and Notes1, Notes2 denotes prominent note sequences associated with the song. One can add further key-value pairs representing rhythms, singer, album, etc. As per our findings, no one has proposed this solution so far for music feature representation.

4 CONCLUSION AND FUTURE DIRECTIONS

A multi-modal framework for universal context building is essential to incorporate context from different modes of the information source. The exercise presented here to build context for nursery rhymes is a proof of concept which can be extended to any musical form. The flexible context representation using a NoSQL database is a proposed universal representation for various applications. This universal context representation is useful for efficient music retrieval and recommendation. We propose to build a framework by integrating the different sources to build a universal music context. The musical video, comments of the listeners can be added further as a source of information to build a universal context using multi-modal analysis. Further, the system can be tested with building the data-set of large music data for multi-modal analysis and perform experimentation to explore the true power of map-reduce functionality to search efficiently within this big data.

5 REFERENCES

[1] Shen, Jialie, Dacheng Tao, and Xuelong Li. "QUC-tree: Integrating query context information for efficient music retrieval." IEEE Transactions on Multimedia 11, no. 2, 313-323, 2009.

[2] Yang, Yi-Hsuan, and Jen-Yu Liu. "Quantitative study of music listening behavior in a social and affective context." IEEE Transactions on Multimedia 15, no. 6, 1304-1315, 2013.

[3] http://abckidsinc.com/top-nursery-rhymes-time-lyrics-origins accessed on 25-3-2019.

[4] Maestre, et.al. "Enriched multimodal representations of music performances: Online access and visualization." Ieee Multimedia 24, no. 1, 2017.

[5] Müller, et.al. "A multimodal way of experiencing and exploring music." Interdisciplinary Science Reviews 35, no. 2, 138-153, 2010.

[6] Makarand Velankar and Hari V. Sahasrabuddhe. “Novel Approach for Music Search Using Music Contents and Human Perception.” International Conference on Electronic Systems, Signal Processing and Computing Technologies 1-6, 2014.

[7] Velankar, M., and Kulkarni, P. A. "Pattern recognition approaches in music analytics" i-manager’s Journal on Pattern Recognition, 5(2), 37-46, 2018.

[8] Su, et.al. "Music recommendation using content and context information mining." IEEE Intelligent Systems 25, no. 1, 16-26, 2010.

[9] Klein, John, et al. "Performance evaluation of NoSQL databases: a case study." Proceedings of the 1st Workshop on Performance Analysis of Big Data Systems. ACM, 2015.

[10] Ajdari, Jaumin, and Brilant Kasami. "MapReduce Performance in MongoDB Sharded Collections." INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS 9.6 (2018): 115-120.