NLP4MusA 2020

Proceedings of the 1st Workshop on NLP for Music and Audio

16–17 October, 2020
Online

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

Welcome to the 1st Workshop on NLP for Music and Audio. The aim of NLP4MusA is to bring together researchers from various disciplines related to music and audio content, on one hand, and NLP on the other. It embraces the following topics.

- NLP architectures applied to music analysis and generation
- Lyrics analysis and generation
- Exploiting music related texts in music recommendation
- Taxonomy learning
- Podcasts recommendations
- Music captioning
- Multimodal representations

The workshop spans one day split into two days to accommodate an online format while preserving a timezone friendly schedule, which features both live and asynchronous presentations and Q/A sessions. The main topics covered in the accepted papers

The talks of our keynote speakers highlight topics of high relevance in the intersection between music, audio and NLP. The presentation by Colin Raffel discusses what Music Information Retrieval (MIR) can learn from recent transfer learning advances in NLP. Sam Mehr focuses in his talk on the notion of universality in music. NLP4MusA also features an impressive number of industry-led talks by Tao Ye, Fabien Gouyon, Elena Epure, Marion Baranes, Romain Henequin, Sravana Reddy, Rosa Stern, Alice Coucke, Isaac Julien and Shuo Zhang. We include the abstract of their talks in this volume.

In total, we accepted 16 long papers (53% of submissions), following the recommendations of our peer reviewers. Each paper was reviewed by three experts. We are extremely grateful to the Programme Committee members for their detailed and helpful reviews.

Sergio Oramas, Luis Espinosa-Anke, Elena Epure, Rosie Jones, Mohamed Sordo, Massimo Quadrana and Kento Watanabe

October 2020

Organisers:

Sergio Oramas (Pandora)
Luis Espinosa-Anke (Cardiff University)
Elena Epure (Deezer)
Rosie Jones (Spotify)
Mohamed Sordo (Pandora)
Massimo Quadrana (Pandora)
Kento Watanabe (AIST)

Program Committee:

José Camacho-Collados (Cardiff University)
Francesco Barbieri (Snap)
Andrés Ferraro (UPF)
Minz Won (UPF)
Lorenzo Porcaro (UPF)
Albin Correya (UPF)
Pablo Accuosto (UPF)
Morteza Behrooz (Facebook)
Christos Christodoulopoulos (Amazon)
Mark Levy (Apple)
Fabien Gouyon (Pandora)
Scott Waterman (Pandora)
Gregory Finley (Pandora)
Zahra Rahimi (Pandora)
Ann Clifton (Spotify)
Sravana Reddy (Spotify)
Aasish Pappu (Spotify)
Manuel Moussallam (Deezer)
Marion Baranes (Deezer)
Romain Hennequin (Deezer)
Horacio Saggion (UPF)
Iñigo Casanueva (Poly AI)
Giuseppe Rizzo (LINKS Foundation)
Pasquale Lisena (EURECOM)
Masataka Goto (AIST)
Peter Knees (TU Wien)
Shuo Zhang (Bose Corporation)
Markus Schedl (TU Linz)
Ichiro Fujinaga (McGill University)
Elena Cabrio (Université Cote d'Azur)
Michael Fell (Université Cote d'Azur)

Richard Sutcliffe (University of Essex)

Invited Speakers:

Colin Raffel, University of Carolina, Chapel Hill & Google Brain
Sam Mehr, Harvard Music Lab
Tao Ye, Amazon
Fabien Gouyon, Pandora/SiriusXM
Elena Epure, Deezer
Marion Baranes, Deezer
Romain Henequin, Deezer
Sravana Reddy, Spotify
Rosa Stern, Sonos
Alice Coucke, Sonos
Isaac Julien, Bose
Shuo Zhang, Bose

Invited Talks

Colin Raffel: What can MIR learn from transfer learning in NLP?

Transfer learning has become the de facto pipeline for natural language processing (NLP) tasks. The typical transfer learning recipe trains a model on a large corpus of unstructured, unlabeled text data using a self-supervised objective and then fine-tunes the model on a downstream task of interest. This recipe dramatically mitigates the need for labeled data and has led to incredible progress on many benchmarks that had previously been far out of reach. In this talk, I'll first give an overview of transfer learning for NLP from the lens of our recent empirical survey. Then, I will argue that transfer learning is massively underutilized in the field of music information retrieval (MIR), particularly in light of the scarcity of labeled music data. To prompt future research, I'll highlight some successful applications of transfer learning in MIR and discuss my own work on creating a large, weakly-labeled music dataset.

Sam Mehr: Universality and diversity in human song

What is universal about music, and what varies? In this talk I will present some highlights from analysis of the Natural History of Song Discography, which includes audio recordings from 86 human societies, to uncover what makes music sound the way it does around the world. Using data from music information retrieval, amateur and expert listener ratings, and manual transcriptions, we find that acoustic features of songs predict their primary behavioral context; that tonality is widespread, perhaps universal; that music varies in rhythmic and melodic complexity; and that elements of melodies and rhythms found worldwide follow power laws. The findings demonstrate basic facts of the human psychology of music that may inform our understanding of aesthetics and our preferences for music.

Tao Ye: Inside a real world conversational music recommender

When a user asks Alexa "Help me find music", there are in fact a multitude of interesting problems to be solved, in the cross-section of Natural Language Understanding, recommendation systems, and advanced natural language generation. In Natural Language Understanding, we encounter intent identification, slots filling, and particular challenges of spoken language understanding (SLU) in the music domain. This is also different from a one-shot command SLU, where users tend to give a clear "play XYZ" intent. In a dialog, users increase variation in their speech and often answer a question casually such as "whatever, I don't care". We rely on both grammatically rules and statistical models to set intent, triggers and fill slots. Machine learning is also applies directly to construct an interactive recommender that makes recommendations more relevant. With real time user critique and feedback, we need to integrate long term user preferences and immediate user requests. Finally, how Alexa speaks to the users also makes a difference in the experience. We tackle the tough problem of making the entire conversation sound natural rather than robotic. Particularly, emotional and empathic tagged speech are used. The challenge is to know when to use these tags to vary speech.

Fabien Gouyon: Lean-back or Lean-in?

In this talk I will go over some of Pandora's latest research and product developments in the realm of voice interactions. I will address how NLU powers unique music listening experiences in the Pandora app, and highlight exciting opportunities for further research and development.

Elena Epure, Marion Baranes & Romain Henequin: "Je ne parle pas anglais"""", dealing with multilingualism in MIR

Deezer is a local player, on a global scale. Our goal is to serve a very diverse audience providing a seamless experience worldwide. Consequently, dealing with multilingualism, and more generally with multiculturalism is essential to us. In this talk, we address two topics for which the generalisation to multilingual data and users is particularly important: the user-app interaction through the search engine and the catalogue annotation with multilingual metadata. We conclude by contemplating the state of multilingualism in the music information retrieval (MIR) community.

Sravana Reddy: The Spotify Podcasts Dataset

We present the Spotify Podcasts Dataset, a set of approximately 100K podcast episodes comprised of raw audio files along with accompanying ASR transcripts, that we released for the TREC 2020 Challenge. We will talk about some of the characteristics of this dataset, and our experiments running baseline models for information retrieval and summarization.

Rosa Stern & Alice Coucke: Music data processing for voice control

The focus of the Voice Experience team at Sonos is to bring together the profuse world of music and the slick user experience of a voice assistant within the Sonos home sound system. Supporting music related voice commands and a music catalog in our SLU (Spoken Language Understanding) system carries challenges at the various stages of our pipeline, which we'll present and discuss in this talk. We'll bring our focus on the main issues we encounter in our data processing pipeline, especially related to speech and voice recognition.

Isaac Julien & Shuo Zhang: Building a Personalized Voice Assistant for Music

The Bose Music Assistant was a former year-long research project that focused on building a personalized, conversational voice interface for music, with the goal of helping our customers find the content they enjoy. We will discuss the creation of a hybrid Grammar- and ML-based NLU engine that supported the Assistant and allowed us to quickly prototype and expand the experiences that it offered. We will also describe some of the NLP challenges we encountered in the music domain, and the opportunity that these challenges provided for personalization.

Table of Contents