NLP4MusA 2021

Proceedings of the 2nd Workshop on NLP for Music and Spoken Audio (NLP4MuSA)

12 November, 2021
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

ISBN 978-1-950737-19-2

Introduction

Welcome to the 2nd Workshop on NLP for Music and Spoken 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 Yunyao Li discusses the challenges posed by the current *Wild West* of NLP research. Invited speakers cover different areas in at the crossroads between Music, Spoken Audi oand NLP, in particular: Longqi Yang focuses on goal-directed music recommendation; Markus Schedl describes different approaches for emotion-aware music exploration; Juham Nam provides a review on music auto-tagging; and finally, Anna Huang discusses a preliminary approach to "tuning" Music Transformer.

In total, we accepted 8 papers (47% 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, Elena Epure, Luis Espinosa-Anke, Rosie Jones, Mohamed Sordo, Massimo Quadrana and Kento Watanabe

Online

November 2021

Organisers:

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

Program Committee:

Andres Ferraro (UPF)
Bruno Massoni (Deezer)
Christos Christodoulopoulos (Amazon)
Elena Cabrio (Université Cote d'Azur)
Ichiro Fujinaga (McGill University)
José Camacho-Collados (Cardiff University)
Kongmeng Liew (Nara Institute of Science and Technology)
Lorenzo Porcaro (UPF)
Manuel Moussalam (Deezer)
Marion Baranes (Deezer)
Mark Levy (Apple)
Masataka Goto (AIST)
Morteza Behrooz (Facebook)
Pasquale Lisena (EURECOM)
Richard Sutcliffe (University of Essex)
Romain Hennequin (Deezer)
Rosa Stern (Sonos)
Scott Waterman (Pandora)
Shuo Zhang (Bose Corporation)
Sravana Reddy (Spotify)

Invited Speakers:

Yunyao Li, IBM
Longqi Yang, Microsoft
Markus Schedl, JKU
Juhan Nam, KAIST
Anna Huang, Google Brain

Invited Talks

Yunyao Li: Taming the Wild West of Natural Language Processing

Natural language processing (NLP) is becoming increasingly adapted in the real-world. To many, NLP is the new resource of growth and wealth. However, the NLP landscape is like the Wild West now: many and growing numbers of players, fast innovations, and limited oversight. In this talk, I will discuss the major challenges in taming the Wild West of NLP. I will present our work in recent years in in addressing these challenges. I will showcase some of the work in concrete domains (e.g. compliance). I will also share thoughts on a general approach towards adapting NLP to solve real-world problems.

Longqi Yang: Towards Goal-directed Content Recommendation

People's content choices (e.g., Podcast, music, etc.) are driven by their short-term intentions and long-term goals, which are often underserved by today's recommendation systems. This is mainly due to the fact that higher-ordered goals are often unobserved, and recommenders are typically trained to promote popular items and to reinforce users' historical behavior. As a result, the utility and user experience of content consumption can be affected undesirably. This talk will cover behavioral experiments that quantify the effects of goal-agnostic recommenders and algorithmic techniques to improve them.

Markus Schedl: Using NLP for emotion-aware music exploration, lyrics and playlist analysis

In this talk, I will showcase the use of NLP techniques for several music-related tasks, which are carried out at the Institute of Computational Perception of the Johannes Kepler University Linz. More precisely, I will briefly introduce our latest research on lyrics analysis, text-based playlist clustering, and emotion-aware music exploration and recommendation.

I will report findings of our studies on genre and temporal differences of song lyrics, and on uncovering the extent to which the sequential ordering of tracks in user-generated playlists matters for different playlist types identified by their title. Furthermore, I will briefly introduce EmoMTB, our emotion-aware music exploration and recommendation interface which adopts emotion recognition techniques from user-generated texts.

Juhan Nam: Music Auto-Tagging: from Audio Classification to Word Embedding

Music auto-tagging is one of the main audio classification tasks in the field of music information retrieval. Leveraging the advances of deep learning, particularly, convolutional neural networks for image classification, researchers have proposed novel neural network architectures for music to improve the annotation and retrieval performances. However, this classification approach has the limitation that the model can handle only a fixed set of labels that describe music and does not consider the semantic correlations between the labels. Recent approaches have addressed the issues by associating audio embedding with word embedding where labels are located in a vector space. This allowed the model to predict unseen labels in the training stage from music or retrieve music from any word query. This talk reviews the advance of music auto-tagging where research interests are moving toward combination with natural language processing techniques.

Anna Huang: Tuning Music Transformer

Music Transformer is an expressive language model for music, offering exciting potential for creative exploration. In the AI Song Contest, we see artists obtain a range of compelling results, by feeding it

different musical fragments to elaborate. However, finding something novel and appropriate could take many iterations. If there's more control, then it could be possible to steer the exploration process. In this talk, I'll discuss preliminary work in taking both ML and HCI approaches to "tuning" Music Transformer towards users' creative goals, and also a common framework for evaluating progress in generative models and interfaces.

Table of Contents