Tutorial 4

Few-Shot and Zero-Shot Learning for Musical Audio

Yu Wang, Hugo Flores García, Jeong Choi

Abstract

While deep neural networks achieved promising results in many MIR tasks, they typically require a large amount of labeled data for training. Rare, fine-grained, or newly emerged classes (e.g. a rare musical instrument, a new music genre) where large-scale data collection is hard or simply impossible are often considered out-of-vocabulary and unsupported by MIR systems. To address this, few-shot learning (FSL) and zero-shot learning (ZSL) are learning paradigms that aim to train a model that can learn a new concept based on just a handful of labeled examples (few-shot) or some auxiliary information (zero-shot), mimicking human ability. By doing so, the trained model is no longer limited to a pre-defined and fixed set of classes but ideally can generalize to any class of interest with the cost of little human intervention. In addition, few-shot and zero-shot models naturally incorporate human input without asking for significant effort, making them useful tools when developing MIR systems that can be customized by individual users.

In this tutorial, we will go over

- FSL/ZSL foundations Task definition and existing approaches.
- Recent advances of FSL/ZSL in MIR Techniques and contributions in recent studies. We will also discuss the remaining challenges and future directions.
- Coding examples Showcasing the training and evaluation pipeline of FSL and ZSL models on specific MIR tasks. Code and references to the tools and datasets will be provided.

We aim for this tutorial to be useful to researchers and practitioners in the ISMIR community who are facing labeled data scarcity issues, looking for new interaction paradigms between users and MIR systems, or generally interested in the techniques and applications of FSL and ZSL. We assume the audience is familiar with the basic machine learning concepts.

Biographies of Presenters

Yu Wang Yu is a Ph.D. candidate in Music Technology at the Music and Audio Research Laboratory at New York University, working under Prof. Juan Pablo Bello. Her research interests focus on machine learning and signal processing for music and general audio. Specifically, she is interested in adaptive and interactive machine listening with minimal supervision. She has interned with Adobe Research and Spotify. Before joining MARL in 2017, she was in the Music Recording and Production program at the Institute of Audio Research. She holds two M.S. degrees in Materials Science & Engineering from Massachusetts Institute of Technology (2015) and National Taiwan University (NTU) (2012), and a B.S. in Physics from NTU (2010). Yu is a guitar player and also enjoys sound engineering. Japanese math rock is her current favorite music genre.

Hugo Flores García is a Ph.D. student in Computer Science at Northwestern University, working under Prof. Bryan Pardo in the Interactive Audio Lab. Hugo's research interests lie at the intersection of machine learning, signal processing, and human computer interaction for music and audio. Hugo has previously worked on a deep learning framework for Audacity, an open source audio editor, and will be a research intern at Spotify and Descript during the latter half of 2022. Hugo holds an B.S. in Electrical Engineering from Georgia Southern University (2020). He is a jazz guitarist, and can be seen playing with various groups local to the Chicago area. Hugo enjoys augmenting musical instruments with technology, as well as making interactive music and art in SuperCollider and Max/MSP.

Jeong Choi is a machine learning researcher at Naver, where he leads NOW AI team that's working on a multi-modal recommendation system for a video streaming service, Naver NOW. Before joining Naver, he was a researcher at

NCSOFT, working on a recommendation system in a music game FUSER. He also interned at Deezer Research. He received a M.S. in Culture Technology at Korea Advance Institute of Science and Technology, under the supervision of Prof. Juhan Nam. His research interest is on representational learning of various signals that can further contribute to diverse music recommendation strategies. Previously, he pursued a long music career as a composer and a bassist. His passion for music research originates from the experience. He also received a M.S. and a B.E. in Digital Media at Ajou University, and majored in French at Daewon Foreign Language High School.