Tutorial 3

Metric Learning for Music Information Retrieval
Brian McFee, Jongpil Lee and Juhan Nam

Abstract

Metric learning is a paradigm of representation learning, in which proximity between the representations of items is optimized to correspond to a notion of similarity. Compared to the classification, metric learning can leverage more flexible forms of supervision, for example, two audio clips belong to the same artist or not, or have the same tempo or not. This enables the learning model to take a large or indefinite number of classes. Moreover, metric learning handles the embedding space directly by measuring the distance between the transformations of different examples. This facilitates usage of different domains or modalities of inputs in the same framework (e.g., audio embedding in one input and word embedding in another input). Such flexible and adaptable characteristics of metric learning have been enjoyed in many of machine learning tasks, particularly, similarity-based content retrieval. In recent years, interest in metric learning from the MIR community has also increased. Considering the multi-faceted and hierarchical-level of notions in similarity (e.g., semantic-level, score-level or audio-level) and diverse forms of data (e.g., audio, MIDI, text labels, lyrics, album covers, and user data), we see a great potential of metric learning in music. Therefore, introducing the method in an educational manner and surveying recent progress will be timely and helpful to relevant researchers. In this tutorial, we plan to present three lectures as follows:

1. **Metric learning foundations:** This lecture introduces mathematical foundations of metric learning.
2. **Deep metric learning and applications to MIR (1): core tasks -** This lecture introduces recent deep metric learning methods and their applications to music classification and similarity-based retrieval tasks.
3. **Deep metric learning and applications to MIR (2): variations -** This lecture introduces various applications of deep metric learning in MIR, showing how researchers have bridged diverse domains and modalities of input in metric learning.

Biographies

Brian McFee is Assistant Professor of Music Technology and Data Science New York University. He received the B.S. degree (2003) in Computer Science from the University of California, Santa Cruz, and M.S. (2008) and Ph.D. (2012) degrees in Computer Science and Engineering from the University of California, San Diego. His work lies at the intersection of machine learning and audio analysis. He is an active open source software developer, and the principal maintainer of the librosa package for audio analysis.

Jongpil Lee received the B.S. degree in electrical engineering from Hanyang University, Seoul, South Korea, in 2015, the M.S. degree, in 2017, from the Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, where he is currently working toward the Ph.D. degree. He interned at Naver Clova Artificial Intelligence Research in the summer of 2017 and at Adobe Audio Research Group in the summer of 2019. His current research interests include machine learning and signal processing applied to audio and music applications.

Juhan Nam is an Associate Professor of the Graduate School of Culture Technology at the Korea Advanced Institute of Science and Technology (KAIST), South Korea. Before joining KAIST, he was a staff research engineer at Qualcomm. Before his research career, he was a software/DSP engineer at Young Chang (Kurzweil). He received the Ph.D. degree (2013) in Music from Stanford University, studying at the Center for Computer Research in Music and Acoustics (CCRMA). He is interested in various research topics at the intersection of music, signal processing, and machine learning.