人工知能学会第二種研究会資料
AIチャレンジ研究会
SIG-Challenge-064-04

Self-Supervised Learning for Bird Vocalization Embedding Extraction

Runwu Shi [1*]　Itoyama Katsutoshi [2]　Nakadai Kazuhiro [1]

[1] 東京工業大学
[1] Tokyo Institute of Technology
[2] (株) ホンダ・リサーチ・インスティチュート・ジャパン
[2] Honda Research Institute Japan, Co. Ltd.

Abstract: Efficient bioacoustics analysis requires automatic processing for vast collections of bird vocalizations, for which low-dimensional vocalization embedding is commonly used. In this study, the bird vocalization embedding from the whole song level is extracted using self-supervised representation learning. Existing methods such as Variational Autoencoder (VAE) based methods have shown the performance in generating these compact embeddings from smaller vocalization units, such as notes or syllables. However, some bird species such as the Great Tits have structured songs consisting of different repeated syllable compositions. In order to directly extract embeddings from the entire song level, this study regards each vocalization as comprising both generalized and discriminative components and employs dual encoders to capture these aspects. The effectiveness of the proposed approach is demonstrated through its superior clustering performance on the Great Tits dataset when compared to both pre-trained models and the standard VAE. Additionally, this study delves into the most informative aspects of the embedding and elucidates the disentangled representation's efficacy in capturing bird vocalizations.

* Contact:
〒 152-8552 東京都目黒区大岡山 2-12-1
E-mail: shirunwu@ra.sc.e.titech.ac.jp