

Untangling indices of emotion in music using neural networks

Dorien Herremans

Singapore University of Technology and Design, Singapore, Singapore

Kin Wai Cheuk

Singapore University of Technology and Design, Singapore, Singapore

Yin-Jyun Luo

Singapore University of Technology and Design, Singapore, Singapore

Kat Agres

Institute of High Performance Computing, A*STAR, Singapore, – Select State/Province –, Singapore

Abstract

Emotion and music are intrinsically connected, and researchers have had limited success in employing computational models to predict perceived emotion in music. Here, we use computational dimension reduction techniques to discover meaningful representations of music. For static emotion prediction, i.e., predicting one valence/arousal value for each 45s musical excerpt, we explore the use of triplet neural networks for discovering a representation that differentiates emotions more effectively. This reduced representation is then used in a classification model, which outperforms the original model trained on raw audio. For dynamic emotion prediction, i.e., predicting one valence/arousal value every 500ms, we examine how meaningful representations can be learned through a variational autoencoder (a state-of-the-art architecture effective in untangling information-rich structures in noisy signals). Although vastly reduced in dimensionality, our model achieves state-of-the-art performance for emotion prediction accuracy. This approach enables us to identify which features underlie emotion content in music.