cases	doc_1	doc_2	decision	id

doc_1

authors	• Virtanen, Tuomas • Xie, Huang
title	Zero-Shot Audio Classification via Semantic Embeddings
publication_date	2021-01-01 00:00:00
source	SupportedSources.CORE
journal	IEEE/ACM Transactions on Audio Speech and Language Processing
volume	
doi	10.1109/taslp.2021.3065234
urls	• https://core.ac.uk/download/542974366.pdf
id	id8571280307586439386
abstract	In this paper, we study zero-shot learning in audio classification via semantic embeddings extracted from textual labels and sentence descriptions of sound classes. Our goal is to obtain a classifier that is capable of recognizing audio instances of sound classes that have no available training samples, but only semantic side information. We employ a bilinear compatibility framework to learn an acoustic-semantic projection between intermediate-level representations of audio instances and sound classes, i.e., acoustic embeddings and semantic embeddings. We use VGGish to extract deep acoustic embeddings from audio clips, and pre-trained language models (Word2Vec, GloVe, BERT) to generate either label embeddings from textual labels or sentence embeddings from sentence descriptions of sound classes. Audio classification is performed by a linear compatibility function that measures how compatible an acoustic embedding and a semantic embedding are. We evaluate the proposed method on a small balanced dataset ESC-50 and a large-scale unbalanced audio subset of AudioSet. The experimental results show that classification performance is significantly improved by involving sound classes that are semantically close to the test classes in training. Meanwhile, we demonstrate that both label embeddings and sentence embeddings are useful for zero-shot learning. Classification performance is improved by concatenating label/sentence embeddings generated with different language models. With their hybrid concatenations, the results are improved further.Comment: Submitted to Transactions on Audio, Speech and Language Processin
versions	

doc_2

authors	• Huang Xie • Tuomas Virtanen
title	Zero-Shot Audio Classification via Semantic Embeddings
publication_date	2020-11-24 14:42:22+00:00
source	SupportedSources.ARXIV
journal	None
volume	
doi	
urls	• http://arxiv.org/pdf/2011.12133v2 • http://arxiv.org/abs/2011.12133v2 • http://arxiv.org/pdf/2011.12133v2
id	id-8641644356207759223
abstract	In this paper, we study zero-shot learning in audio classification via semantic embeddings extracted from textual labels and sentence descriptions of sound classes. Our goal is to obtain a classifier that is capable of recognizing audio instances of sound classes that have no available training samples, but only semantic side information. We employ a bilinear compatibility framework to learn an acoustic-semantic projection between intermediate-level representations of audio instances and sound classes, i.e., acoustic embeddings and semantic embeddings. We use VGGish to extract deep acoustic embeddings from audio clips, and pre-trained language models (Word2Vec, GloVe, BERT) to generate either label embeddings from textual labels or sentence embeddings from sentence descriptions of sound classes. Audio classification is performed by a linear compatibility function that measures how compatible an acoustic embedding and a semantic embedding are. We evaluate the proposed method on a small balanced dataset ESC-50 and a large-scale unbalanced audio subset of AudioSet. The experimental results show that classification performance is significantly improved by involving sound classes that are semantically close to the test classes in training. Meanwhile, we demonstrate that both label embeddings and sentence embeddings are useful for zero-shot learning. Classification performance is improved by concatenating label/sentence embeddings generated with different language models. With their hybrid concatenations, the results are improved further.
versions	

decision	id
DUPLICATES	233