

令和元年度 情報工学コース卒業研究報告要旨

戸田 研究室	氏 名	犬 塚 雅 也
卒業研究題目	環境音の記号表現獲得に向けた 教師なし学習法に関する検討	

人間の発話は、言語情報に基づいた音声の最小構成要素である音素を利用することで、記号系列として記述することが可能である。一方で、水が流れる音や風が吹く音、人混みの音などのような環境音は、音声における音素のような記号表現を持たないため、記号系列として記述することが困難である。もし、環境音を記号系列として記述することができれば、類似画像検索のように、ある環境音と類似した環境音を検索するシステムや、記号表現を元にした自動タグ付けなど様々な応用が期待される。

環境音が記号表現を持たない理由として、環境音を発生させる要因は多種多様であること、複数の種類の環境音が同時に生じて重畳することなどが挙げられる。これに対し、擬音語を用いた環境音の記号表現も研究されているが、人によって聞こえ方が異なるなど、人手による記号表現の定義自体が難しいという問題がある。このような問題に対して、様々な信号に適用可能な深層学習に基づくデータ駆動方式による記号表現の獲得手法が注目を集めている。中でも、AutoEncoderの一種であるVector Quantization Variational AutoEncoder (VQ-VAE)は、学習データの再構成による教師なし学習に加え、埋め込み空間の学習を行うことで、特徴量の離散的表現を獲得することが可能である。先行研究によると、発話音声を学習することで、埋め込み空間のベクトルが音素に近い表現を獲得することが確認されている。

本研究では、環境音の記号表現獲得に向け、音声と環境音を併用したVQ-VAEの学習方法を提案する。提案手法では、発話音声を用いてVQ-VAEを学習したあと、そのモデルを初期値として、環境音を用いたファインチューニングを行う。発話音声で学習したモデルは、環境音を擬音語化するのに有効な埋め込み空間を備えていると予想されるため、ランダムな初期値から学習を行う場合に比べて学習結果が向上することが期待される。

提案手法の有効性を調査するため、AudioSetを用いた実験的評価を実施する。AudioSetの環境音に対して、A) パワースペクトルの量子化と位相復元処理による再合成、B) 提案手法、C) 発話音声のみで学習したVQ-VAE、D) 環境音のみで学習したVQ-VAEにより、記号化ならびに再構成を行った。元の音と再構成音の類似度、ならびに、各種環境音ラベルに対応した環境音の再現度に関する主観評価実験結果(図2)から、提案法は他の手法と比較し、類似度ならびに再現度の高い再構成音を得ることができ、環境音の種類や時間的特徴などをより適切に捉えた記号表現を獲得できる可能性が示された。

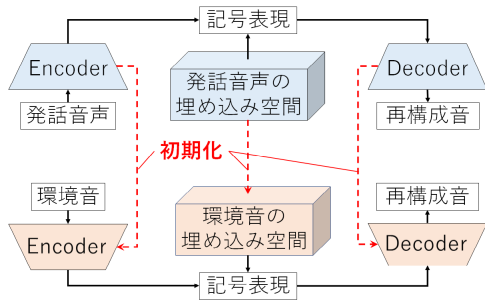


図1 提案手法の概略図

(上：発話音声による事前学習，
下：環境音による学習)

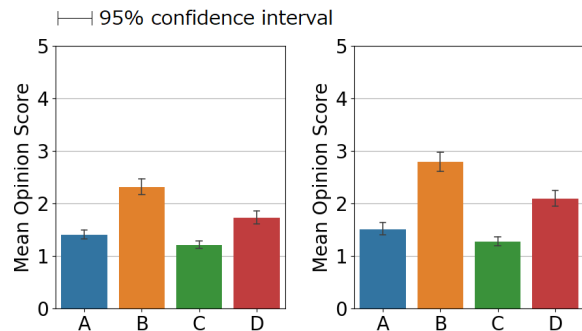


図2 主観評価実験の結果

(左図：類似度，右図：再現度)