Generative Adversarial Networks を用いた半教師あり学習の音響イベント検出への適用

Application of Semi-supervised Learning Using Generative Adversarial Networks to Sound Event Detection

合馬一弥　　山田武志　　牧野昭二
Kazuya Ouma　Takeshi Yamada　Shoji Makino
筑波大学情報学群情報科学類
College of Information Science, University of Tsukuba

1. はじめに

音声アシスタントや自動運転車において，音響イベント検出(SED: Sound Event Detection) は重要な役割を担っている．SED とは，与えられた音響データ内で発生している音響イベントの種類，開始時刻，終了時刻を特定することである．

環境音認識の国際コンペティションである DCASE [1] でも見られるように，最近では SED を行う手法として CNN (Convolutional Neural Networks) や CRNN (Convolutional Recurrent Neural Networks)がよく用いられている．このような NN を用いた SED の学習には大量の強ラベル(音響イベントの種類，開始時刻，終了時刻からなるラベル)付きデータが必要である．しかし，強ラベルの作成には音響データの聴取や波形の確認を行う必要があるため，高い検出精度を得るために必要な数百～数千のデータを用意するのは困難である．従来の解決策として，少量の強ラベルを用いた学習の後，大量の弱ラベル(音響イベントの種類のみからなるラベル)を用いて検出精度を向上させる手法[2]が挙げられる．弱ラベルの活用によって，人的・時間的コストを抑えることができる．

弱ラベルの活用のほかに，ラベルなしデータを活用した手法が期待されている．画像認識分野においては，ラベルなし画像を活用するために GAN (Generative Adversarial Networks)[3]を応用した半教師あり学習手法[4]が提案されている．従来の GAN は，よりリアルなデータを出力する Generator と入力データがリアルなものであるかを判別する Discriminator からなるネットワークであり，学習後の Generator が画像生成等に用いられる．一方，[4]ではこの Discriminator にクラス学習も適用し，学習後の Discriminator をクラス分類に用いることで半教師あり学習を実現している．本稿ではこれを SED に適用し，その有効性を評価する．

2. GAN を用いた半教師あり学習

GAN を用いた半教師あり学習手法の概要を図 1 に示す．Generator は従来の GAN と同様の動作をする．すなわち，ノイズからデータを生成し，Discriminator にどの程度リアルなものだと判定されたかで学習を行う．Discriminator は，入力データがリアルなものかを判別するように学習するほか，入力データが強ラベル付きのリアルデータであった場合は SED についても学習を行う．

3. 実験

上記の手法の有効性を評価するため DCASE 2016 Task 3 [5] を用いて実験を行った．このタスクは，家と住宅街の二つの場所において収録された音響データと強ラベルをもとに SED を行うタスクである．このタスクにおいて，学習データのうち一部の強ラベルを隠し，(i)残った強ラベル付きデータのみで学習する手法，(ii)半教師あり GAN

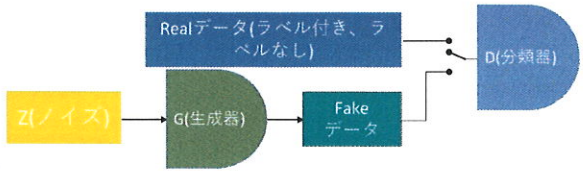

図 1 半教師あり GAN

表 1 実験結果(評価は Error Rate)

強ラベル付きの割合	1	0.9	0.8	0.7
(i)強ラベルのみ	0.65	0.71	0.86	0.90
(ii)強+ラベルなし	0.49	0.61	0.75	0.77

を用いてラベルなしデータも用いて学習する手法を比較する．本実験では音響特徴量としてメルスペクトログラムを用いた．また，(i)の検出器と(ii)の Discriminator には CRNN を，(ii)の Generator には UpSampling＋CNN を用いた．

実験結果は表 1 の通りである．強ラベル付きデータのみを用いた手法(i)に比べてラベルなしデータも用いた手法(ii)は Error Rate が向上していることが分かる．また，画像認識分野において訓練データがすべて正解ラベル付きである場合でも，半教師あり GAN を適用することで精度の向上が得られている[6]が，SED においても同様の向上が見られる．

4. おわりに

GAN を用いた半教師あり学習によって，SED においてもラベルなしデータを用いて検出精度の向上に寄与できることが示された．

謝辞　本研究は JSPS 科研費 19H04131 の助成を受けた．

参考文献

[1] http://dcase.community/.
[2] T. Matsuyoshi, T. Komatsu, R. Kondo, T. Yamada, S. Makino, “Weakly Labeled Learning Using BLSTM-CTC for Sound Event Detection,” Proc. APSIPA 2018, pp. 1918-1923, Nov. 2018.
[3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Networks,” arXiv:1406.2661, 2014.
[4] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, “Improved Techniques for Training GANs,” arXiv:1606.03498, 2016.
[5] http://dcase.community/challenge2016/task-sound-event-detection-in-real-life-audio.
[6] A. Odena, “Semi-Supervised Learning with Generative Adversarial Networks,” arXiv:1606.01583, 2016.