

敵対的生成ネットワークを用いた画像生成と音声変換

金子 卓弘^{1,a)}

概要：画像や音声は、私たち人間がコミュニケーションを行う上で重要かつ不可欠なものであるが、必ずしも理想的なものが得られるとは限らない。この理想とのギャップを解消する手段として、近年、深層学習に基づく生成モデル（深層生成モデル）を用いた自動生成・変換が注目を集めている。特に、近年最も脚光を浴びている深層生成モデルの一つとして、敵対的生成ネットワーク（GAN: Generative Adversarial Networks）がある。GANは生成器と識別器の二つのニューラルネットワークから構成されており、これら二つのネットワークを互いに競合する条件下で学習させることによって、実際のデータと見分けがつかないようなデータを生成可能な生成器の学習を可能にしている。このような学習機構を用いることで、GANでは、学習対象のデータ分布を陽に仮定することなく学習することができ、この柔軟性によって様々なタスクで解決の糸口を与えてきている。本講演では、まず、GANの黎明期から主要な研究対象である「画像生成」を題材に、GANの近年の発展および拡張について、講演者らの取り組みを紹介しながら解説する。また、深層学習に基づくモデルの強みの一つは、表現学習をも内包することによって、分野をまたいだ技術応用が容易になった点である。GANに関しても同様のことが成り立っており、画像分野における知見を活かした他分野への技術応用が盛んに行われている。その一例として、本講演では、GANの「音声変換」への応用について、講演者らの研究事例に触れながら説明する。最後に、GANの音学分野での応用可能性などについて議論しながら、GAN研究の今後を展望する。

¹ 日本電信電話株式会社
NTT, 3-1, Morinosato Wakamiya Atsugi-shi, Kanagawa
Pref., 243-0198, Japan

^{a)} takuhiro.kaneko.tb@hco.ntt.co.jp