



Ian J. Goodfellow et al. : Generative Adversarial Nets

Advances in Neural Information Processing Systems (NIPS), pp.2672-2680 (2014)

創造的活動とコンピュータ

絵画や音楽、小説等を創作する力は、人間を特徴づける最も重要な能力の1つと古くから考えられてきた。これを計算論的に実現する技術はコンテンツ生成と呼ばれており、1950年代から始まる長い研究の歴史を有する。しかしながら、コンテンツ生成は情報科学的な観点からは非常に難しい問題である。たとえば、リンゴの画像を生成する作業は、「リンゴ」という概念を表す符号のみから無数に存在し得る高次元のパターン情報(画像)を推定するプロセスであり、本質的に不良設定な問題である。このため、深層学習が認識問題において大成功を収めたあとも、コンテンツ生成の実現は相当先であろうと考えられてきた。

ところが近年、コンテンツ生成技術は大方の予想を裏切り驚くべき速さで進歩を続けている。特に、画像生成の技術は圧倒的に進化しており、人間が見ても本物と見分けがつかないレベルのリアルな画像をゼロから生成することが可能になっている。2018年には、人工知能が描いた絵画が史上初めてオークションにかけられ、43万ドルという高価格で落札されたことが大きな話題となった。また、画像生成を応用した技術である Deepfake (実際の映像を改変し精巧な偽映像を作成する技術)の脅威はテレビ番組等でも盛んに取り上げられるようになり、コンテンツ生成は良くも悪くも社会に大きな影響をもたらし得る技術として、研究者のみならず広く一般に注目を集めている。今回取り上げる論文は、コンテンツ生成、特に画像生成において飛躍的なブレイクスルーをもたらした機械学習手法である敵対的生成ネットワー

ク (Generative Adversarial Networks ; GAN) を提案したものである。図-1に、GANによる顔画像生成の進歩の様子を示す。このように、GANの登場以降大きく技術革新が進み、わずか3年ほどできわめてリアルな画像生成が実現できるようになった。

敵対的生成ネットワーク

統計的機械学習に基づく画像生成では、画像の生成モデルの学習、すなわち本物の多数の(理想的にはすべての)画像が為す確率分布を表現する確率モデルを構築することが目的となる。仮にそのような良いモデルが得られれば、モデルからのサンプリングによりリアルな画像を生成することが実現できる。

GANはディープニューラルネットを利用した生成モデルの1つであり、敵対的学習と呼ばれる新しい学習の枠組みを導入した手法である。図-2に示すように、生成ネットワーク(Generator Network)と識別ネットワーク(Discriminator Network)の2つのモジュールから構成される。生成ネットワークは、ランダムに発生させるノイズ(生成する画像に対応した潜在変数に相当する)を種として入力し、これを変換してできるだけ本物に近いリアルな画像を生成することを目的とする。一方、識別ネットワークは、生成ネットワークが作り出した偽



図-1 GANによる顔画像生成の進歩。画像は左からそれぞれ [Goodfellow, et al., NIPS'14], [Radford, et al., ICLR'16], [Liu and Tuzel, NIPS'16], [Karras, et al., ICLR'18] より引用

の画像と、本物の画像のいずれかを入力とし、入力画像が本物か・偽物かを見分ける2クラス識別問題を学習する。つまり、偽物を見破る能力をできるだけ高めるように訓練されるのである。数理的には、GANの目的関数は識別ネットワークと生成ネットワークが為すミニマックス問題として定式化され、学習が成功すれば局所的なナッシュ均衡点へ収束することが示されている。このように、騙すモデル(生成ネットワーク)と見破るモデル(識別ネットワーク)を敵対的に学習させることで、相互に相手を鍛える効果が生まれ、結果として性能の良い画像生成モデルを得ることができる。

さて、GANがその他の生成モデルと比べてなぜ高精細にリアルな画像を生成できるかはよく分かっていないことも多いが、説明としてしばしば為されるのは、GANは多峰の分布のモデル化に優れているという考え方である。1つの概念に対応する画像はさまざまなバリエーションがあるように、実際の画像の分布は正規分布のような単純な形ではなく、多峰の複雑な形になっている。このような分布へモデルを正確にあてはめることは一般に非常に難しく、全体を包んだような大雑把なモデルになりやすい。これを避けるため、GANでは明示的に「本物と見分けにくい」という基準を学習に加えることにより、分布全体を最大公約数的にモデル化するのではなく、実際の画像の分布からはみ出さないように局所領域をモデル化することに成功していると考えられる。

このようにGANが提案した学習アルゴリズムは非常にユニークなものであったが、これは識別ネットワークの学習が収束したという条件のもとで、Jensen-Shannonダイバージェンスと呼ばれる尺度に基づき、実画像の分布とモデルが為す分布を近づける働きになっていることが分かっている。これを

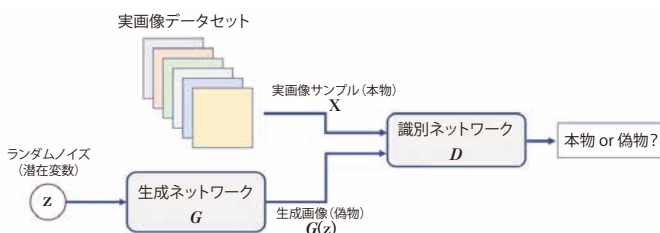


図-2 GANの構造図

KL ダイバージェンスや Wasserstein 距離に置き換えた改良手法など、現在に至るまで数多くの後続研究が出ている。また、ランダムノイズからの生成ではなく、別の入力に条件つけた生成を行うことにより、キーワードや文章からの画像生成や、画像変換(白黒画像着色、超解像度化等)など、実用的にも有用なアプリケーションが次々に実現されている。

AIによる創造の限界

このようにGANは画像生成に驚くべき進化をもたらし、ともすれば何でも作り出せる魔法の技術であるかのように喧伝されがちであるが、もちろんそのような摩訶不思議なものではなく、やっていることはあくまで与えられたデータの補間にすぎないことに注意が必要である。つまり、最初に与えたデータセットの分布に含まれないような新しい画像を作り出すことは原理的に不可能であり、しよせんは今までに見たことのあるデータを模倣しているにすぎないのである。

このように聞くと、やはりコンピュータはまだただと安心する方も多いだろう。しかしながら、ここからは筆者の主観であるが、近年の機械学習によるコンテンツ生成技術の進化は、創造活動とは何かということをあらためて人間に問いかけているように感じられない。実際のところ、我々が住む世の中ではどこかで見たことのあるような絵、小説、映画などの創作物をしばしば目にするであろう。ごく一部の歴史に残るような天才を除き、大半の人間は記憶の中にある何らかの事例の模倣や組合せを(意識的にせよそうでないにせよ)出力しているだけなのかもしれない。我々研究者も、機械学習アルゴリズムに後れをとらないように、真に無から有を生み出す創造的な研究活動ができるよう努力したいものである。

(2019年10月1日受付)

中山英樹 (正会員) nakayama@ci.i.u-tokyo.ac.jp

2011年東京大学大学院情報理工学系研究科博士課程修了。博士(情報理工学)。2012年同大学院講師、2018年同准教授。画像認識、自然言語処理、深層学習等に関する研究に従事。