

300行で理解する 拡散モデル

手塚 峻典

近年の画像生成タスクではやっている深層学習手法として拡散モデルがあります。これは2020年の論文⁽¹⁾で提案された手法です。現在はOpenAIによって開発されたDALL・E 2やStability AIによって開発されたStable Diffusionにこれが用いられています。さらに、画像生成だけでなくさまざまな生成手法にもこれが使われています。

本稿はコードを通してこの拡散モデルの基本を理解することを目的としています。

拡散モデルの基本

拡散モデルは拡散過程と逆拡散過程という2つの過程で成り立っています(図1)。拡散過程 q は画像に対してノイズをかける操作です。各時間ステップごとにわずかにかけることで少しずつ画像の分布から標準正規分布へ近づけていきます。この過程は単純な確率的なアルゴリズムで表せます。

逆拡散過程 p_θ はノイズをかけられた画像に対してノイズを取り除く操作です。拡散過程の逆操作です。時間ステップごとにノイズ除去を少しずつ行うことで標準正規分布から画像の分布へ近づけることができます。

拡散モデルでは逆拡散過程の一部にニューラル・ネットワーク(NN)を用いています。このモデルはこの2つの過程を用いて拡散モデル内部のニューラル・ネットワークの学習と推論を行います。拡散モデルに限らず生成モデルでは特定の条件を与えてその条件に沿ったデータ生成を行うことが求められます。例えば、特定の条件としてクラス・ラベルやテキスト、画像などのどれかが条件として与えられる場合もあります。このようなタスクを条件付き生成と呼びます。

また本稿では条件付き生成と対比することになる都合上、何も条件を付けずに生成するタスクを条件なし生成と呼ぶことにします。拡散モデルの条件付き生成タスクに頻繁に使われる手法としてClassifier Free Guidance(CFG)があります。

この手法は条件付けへの忠実度を制御する引数が与えられ、それを操作して生成することで条件に強く

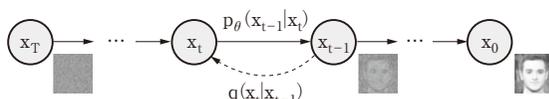


図1(1) 単純な確率的なアルゴリズムで表せる拡散過程

沿ったデータの生成や、逆に条件に弱く沿ったデータの生成が行える特徴を持ちます。由来としてもともとClassifier Guidance(CG)という手法があり、その改良手法がCFGになります。

CGは別途その拡散モデル専用の学習済みクラス分類モデルを用意する必要があります。その分類モデルの誘導で条件付き生成を行う特徴があります。CFGはこのような分類モデルが必要ないという意味合いで分類器フリーな手法です。本記事では拡散モデルの基本としてこれらのことを説明していきます。

実際のオープンソース・ソフトウェア を使って説明する

これらについてConditional Diffusion MNISTというオープンソース・ソフトウェアを用いて説明していきます。これは手書き数字の画像データセットであるMNISTデータベースを生成タスクの学習データとして利用し、拡散モデルで条件付き生成タスクを行うソフトウェアです。

MITライセンスでスクリプト(script.py)全体で400行程度しかない上に、その中で現在の拡散モデルの基本が約100行でまとめられているという教育目的であればありがたい仕様になっています。

```
https://github.com/TeaPearce/Conditional\_Diffusion\_MNIST/blob/main/script.py
```

拡散モデルの学習と推論は論文⁽¹⁾の手法を用いています。条件付き生成タスクは0~9の数字をクラス・ラベルとし、指定された数字の画像を生成するタスクになっています。

図2はその拡散モデルが生成した画像で左から各列ごとに0~9の数字を条件付けています。補足として

◆参考・引用*文献◆

(1) *Ho et al : Denoising Diffusion Probabilistic Models. arxiv.org.
<https://arxiv.org/abs/2006.11239>