

画像認識の学習用画像生成

玉川学園高等部 2年 原田 悠世

動機

文字認識モデルを学習させるための画像(データセット)を画像の**収集以外の方法で増やす**ことはできないかと考えた。

目的

画像生成モデル(GAN)を用いて画像を生成し、**生成した画像を用いてデータ拡張**が行えるかを検証する。

GANとは

・機械学習の手法の一つ
・イタチごっこに例えられる学習工程で学習する。
→生成器と識別器が交互に学習を行い、生成器は識別器を騙す画像を生成することを目標に学習し、識別器は生成器が作り出した画像と本物の画像を識別することを目標として学習を行う。
GANには様々な派生形があり、今回はその中でも**CGAN**(Conditional GAN)というGANを使用した。CGANは基本的なGANと同じような仕組みだが、**アルファベットの種類ごとに画像を生成**することを可能にしている。

実験方法

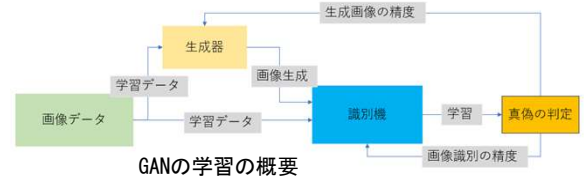
使用したデータセット: EMNIST
内容: A~Zの画像

EMNIST画像の例→



1文字あたり約2400~2600枚

- ・学習に使用しない**テスト用の画像**を1文字あたり**400枚**ランダムに抽出した。このテストセットを用いてモデルの精度を評価した。
- ・文字ごとに残りの約2000枚から指定した枚数をランダムに抽出してGANや文字認識モデルの学習を行った。
- ・文字認識モデルの学習ではランダムなシード値(初期値)から5回学習を行い、その**平均の正答率を実験結果の正答率**として学習を行った。



GANの学習の概要

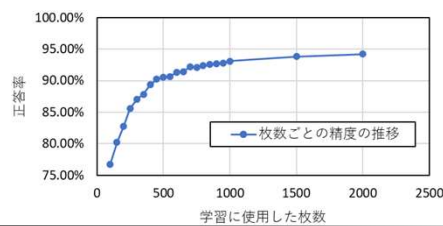
前提実験

(目的) 文字認識モデルの学習に使用した**画像枚数の変化とモデルの精度の関係**を調べる。

(方法) EMNISTの大文字のアルファベットA~Zまでの画像データを用意し、それぞれ1文字あたりの画像の枚数を決めた枚数にそろえて文字認識の学習を行った。

(結果) 文字認識の学習は**画像枚数が多くなればなるほど精度が上昇**し続けることが分かった。しかし、使用する**枚数が多くなればなるほど精度の上昇がほぼ横ばい**になっていくことがわかった。

学習に使用した枚数ごとの精度の推移



←画像の枚数と精度の推移のグラフ。
100~500枚までは50枚ずつ、500~1000枚までは100枚ずつ、1000~2000枚は500枚ずつ増やしている。

実験1

(目的) 生成した画像が**文字認識に有効**かを調べる。

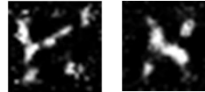
(方法) EMNISTの大文字のアルファベットA~Zの画像をそれぞれ1文字あたり300枚ずつ使用して学習したGANを用いて、アルファベットA~Zの画像を1文字あたり300枚生成した。これを用いて文字認識モデルを作成し、EMNISTデータ300枚の画像を使った文字認識モデルと精度を比較し、生成画像の評価を行った。

(結果) **A以外のアルファベットの精度がとても低くなってしまっている**ことがわかる。また、すべての文字において正答率が前提実験の時より下回っていることがわかる。

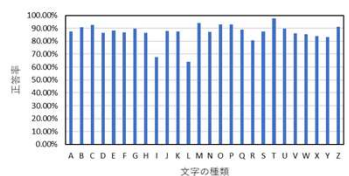
(考察) 画像の**生成時の精度が偏ってしまっている**ことが精度の低迷の原因だと考えられる。つまり、A以外の画像がうまく生成できていないといえる。生成された画像を確認したところ、**ノイズのような意味のない模様の画像**が多くみられたことから生成に問題があることがわかる。

(結論) 生成に問題があり、**精度が低くなっている**。

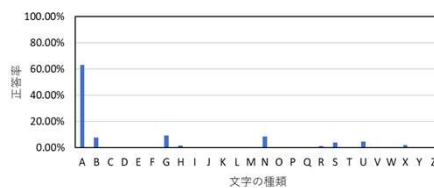
ノイズのような画像→



EMNISTデータ300枚での文字ごとの正答率



生成画像のみのモデルの文字ごとの正答率



↑前提実験での1文字あたり300枚の文字ごとの精度

←生成画像1文字あたり300枚の文字ごとの精度

実験2

(目的) 生成した画像の**フィルタリング**が有効かを調べる。

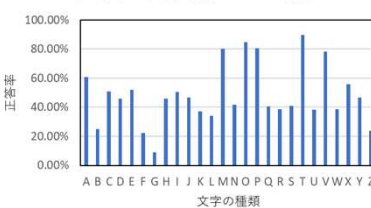
(方法) 実験1で作成したモデルから画像を生成する際に、GANで使った識別器を用いて精度の低い画像を弾くようにする。また、フィルタリングの強度を変えて実験を行う。

(結果) **A以外のアルファベットの精度が改善**された。しかし、しだいに前提実験の結果と比較すると精度が低いことがわかる。また、**フィルタリングの強度によって精度の変化**が起こった。

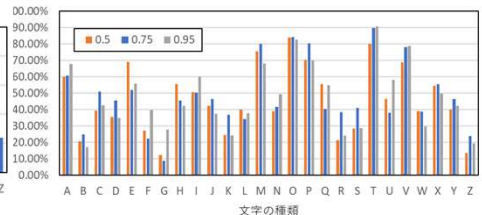
(考察) フィルタリングにより生成画像の内容が改善され、**精度が向上**したことがわかる。

(結果) **フィルタリングにより精度が改善**された。

フィルタリング後の文字ごとの正答率



生成画像による文字ごとの正答率



実験3

(目的) 実験1、2より、**生成された画像だけでは精度が高くない**こと分かったので、今度は生成された画像を元の画像データと合わせて使用し、**拡張用のデータとして使用した場合**はどのようにモデルに影響するのかを調べる。

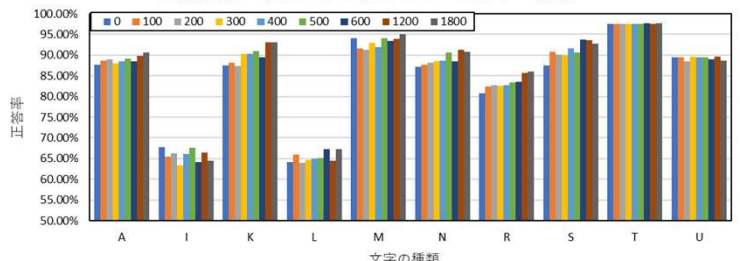
(方法) EMNISTのアルファベット1文字あたり300枚で学習したGANで生成した画像の枚数を変えてGANを学習するために用いた画像と合わせて文字認識モデルの学習を行い、精度がどのように変動するかを調べる。

(結果) 生成した枚数を増やすと**精度はわずかに向上**したが、前提実験での結果ほど正答率は上昇しなかった。また、生成枚数によって文字ごとに精度が変化していていることが分かった。

(考察) 図よりKのように**精度が大幅に上昇**している文字もあれば、Uのように**精度がわずかに下が**っている文字もある。しかし、大体のアルファベットは**AやNのように精度の上がり幅が少なく**、これが全体の**精度があまり上昇しなかった原因**だと考えられる。また、**Tはほぼ全く精度が変わらな**かった。これは、**ももとの精度が高い**からだと考えられる。I、Lの精度が低いのは**I、Lの画像が似通っていた**ことにより、**識別が困難**だったことが考えられる。

(結論) 精度が生成画像によって**上昇**したことから、生成画像によるデータ拡張は成功したといえる。

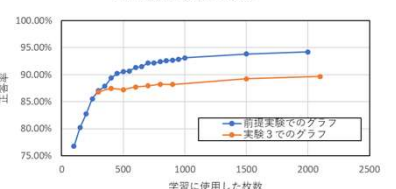
生成枚数の変化による文字ごとの正答率の変化



I、Lの画像→



前提実験と実験3の比較



参考文献

- 1) EMNIST: an extension of MNIST to handwritten letters (EMNISTの論文) G. Cohen, S. Afshar, J. Tapson and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017, pp. 2921-2926, doi:10.1109/IJCNN.2017.7966217.
- 2) Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks Sandfort, V., Yan, K., Pichard, P.J. et al. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 9, 16884 (2019). <https://doi.org/10.1038/s41598-019-52737-x>
- 3) PyTorch <https://pytorch.org/>
- 4) CGAN (条件付きGAN) のサンプルを理解と実装 #Python - Qiita https://qiita.com/miya_ppp/items/f444fc00e4c6e708c047
- 5) PyTorchでCNNを徹底解説 #Python - Qiita <https://qiita.com/mathlive/items/8e1f9a8467ff8df03c>
- 6) Generative Adversarial Networks (GANの論文) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. "Generative Adversarial Networks". Advances in Neural Information Processing Systems 27 (NIPS 2014).
- 7) [Pytorch] MNISTのCGAN (Conditional GAN) を実装する <https://qiita.com/gensal/items/527e350ca7cc4c9851e9>