高精度な文字認識を実現したAIによる学習手法

Learning Methods for Realization of Optical Character Recognition with High Accuracy Using AI

古畑　彰夫 FURUHATA Akio　　田中　遼平 TANAKA Ryohei　　長田　邦男 OSADA Kunio

近年，定型作業の一層の省力化が叫ばれる中，ソフトウェアロボットによる業務自動化（RPA：Robotic Process Automation）の需要が高まっている。我が国では，いまだに紙に依存する業務が多く残っており，多様な帳票の手書き文字などを高精度に認識できる，AIを活用した光学的文字認識（AI-OCR）への期待が大きい。認識精度を高めるには，多種多様な文字データを学習させる必要があるが，データの収集・教示作業に多大なコストが掛かるため，文字データを大量に集めることは難しい。

そこで，東芝デジタルソリューションズ(株)は，文字データの収集・教示コストを抑えながら文字認識の精度を向上させるため，AIを活用した文字データ生成手法や，未教示データを学習に活用する半教師あり学習手法の開発を進めている。これらの手法を評価した結果，コストを抑えながら，AI-OCRの文字認識精度を向上させることができた。

Attention has been focused in recent years on robotic process automation (RPA) using software robots amidst the ongoing improvement of routine, repetitive business operations in Japanese companies. As part of these efforts, the introduction of optical character recognition using artificial intelligence (AI-OCR), which makes it possible to recognize various handwritten characters on a wide variety of business forms with a high degree of accuracy, is expected to improve the efficiency of paper-dependent business processes. However, the higher data collection and teaching costs of AI-OCR, due to the need for a large number of different character samples including a vast number of handwritten kanji characters in order to improve recognition accuracy, are a serious issue.

To rectify this situation, Toshiba Digital Solutions Corporation is engaged in the development of the following methods for AI-OCR to improve character recognition accuracy while suppressing increases in costs: (1) a data augmentation method that can generate a variety of character data based on a small number of actual handwritten data and (2) a semi-supervised learning method using virtual adversarial training to recognize character strings. We have confirmed the effectiveness of these AI-based methods through verification tests.

1．まえがき

昨今，働き方改革が強く推進される中で，定型作業の省力化を目的としたRPAの活用が進んでいる。我が国のビジネス環境では，紙文書を取り扱う事務作業が現在でも多く残っており，紙文書処理を含むワークフローのRPA化を実現するため，OCRの需要も高まっている。**図1**に，RPAでのOCR利用例を示す。

RPAで利用されるOCRには，非定型帳票や，乱雑に書かれたものを含む手書き文字などの認識が求められることが多いが，従来技術では認識精度の向上が難しかった。そこで，ディープラーニングに代表される新しいAI技術でこれらを実現するAI-OCRが広がりつつある。

東芝デジタルソリューションズ（株）は，50年以上の文字認識研究開発の歴史を持ち，広くOCR関連のソリューションを提供してきた。ここでは，AI-OCRの実用化に向けた当社の取り組みについて述べる。

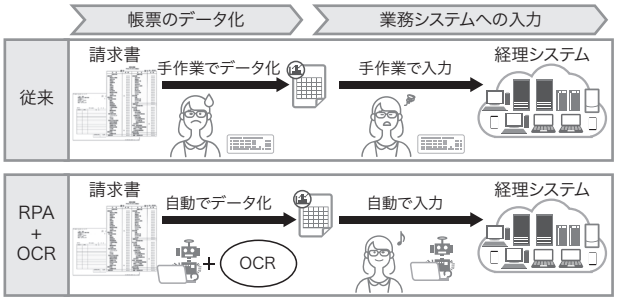

図1．RPAでのOCR利用例

OCRで読み取ったデータをRPAによって自動でデータ化してシステムに入力し，作業コストを削減する。

Example of collaboration between RPA and OCR

2．AI-OCRとその認識モデルの学習

OCRとAI-OCRの違いの明確な定義は難しいが，一般的には次のような特徴を持つものがAI-OCRと呼ばれることが多い。

(1)　ディープラーニングなどの新しい機械学習を活用しており，手書き文字も高精度に読み取れる。
(2)　運用中に認識したデータを学習して，認識精度を更に向上させていく。
(3)　細かな読み取り設定をしなくても，非定型帳票が認識できる。

ディープラーニングなどの機械学習手法では，多種多様なデータを認識モデルに学習させることで，認識精度を向上させることができる。文字認識の場合，多種多様な文字データ（文字パターン）を収集するには多くの人に実際に文字を記入してもらう必要があるが，日本語は文字の種類が多く，記入・収集コストが大きくなるという問題がある。また，ディープラーニングでは，カテゴリーが教示されたデータ（教示データ）を用いて学習（教師あり学習）させることが一般的であり，収集した文字データにカテゴリーを教示するコストも必要になる。

このような文字データの収集・教示コストなどを抑えながら認識精度を向上させる手法として，3章では，多種多様な手書き文字データを生成する学習用文字データ生成手法について，4章では，教示データだけでなくカテゴリーが教示されていないデータ（未教示データ）も学習に活用できる半教師あり学習手法について，それぞれ述べる。

3. 学習用文字データ生成手法

手書き文字は，人ごとの書き癖や記入の丁寧さなどによるばらつきが大きいため，高精度な認識モデルを得るには，多くの人が書いた文字データを学習させる必要がある。しかし，日本語は第2水準漢字などの余り使われない文字も多く，全てのカテゴリーについて十分な数の文字データを収集することは難しい。

これを解決する手法の一つが学習用データ生成（データオーグメンテーション）手法である。実際に記入された少数の文字データを基に，実際には記入されなかった多数の文字データを生成する。

3.1　画像生成手法

学習用データを基に，その特徴を捉えた新たなデータを生成する手法として，VAE（Variational Auto-Encoder）[1]やGAN（敵対的生成ネットワーク）[2]などがよく知られている。VAEは，入力データをより扱いやすい潜在変数と対応付ける手法である。データ生成時に潜在変数を変更することで，実際に観測された変形を模擬して入力データを変形させたデータが得られる。GANは，生成モデル（データを生成）と，識別モデル（入力されたデータが生成データであるか実データであるかを識別）とを敵対的に学習させ，実データと見分けの付かないデータを生成する生成モデルを得る手法である。これらの手法に，指定したカテゴリーのデータを生成できる機能を追加したものが，それぞれConditional VAE（CVAE），Conditional GAN（CGAN）である。

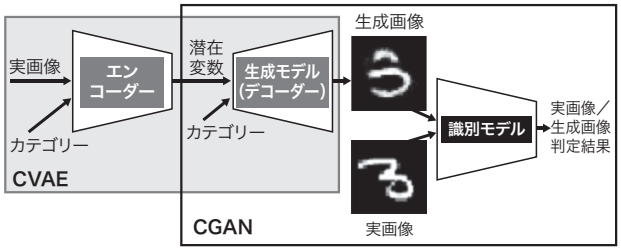

図2.　CVAE-GANの構成
前半部分がCVAE，後半部分がCGANで構成される。二つを組み合わせることで，高品質な文字データが生成できる。
Outline of conditional variational autoencoder and generative adversarial network (CVAE-GAN)

CVAEは，生成データの変形は制御しやすいが，必ずしも実データに近いデータが生成できるとは限らない。一方，CGANは実データに近いデータを生成できるが，VAEと違って生成データの変形を制御できないという問題がある。

そこで，CVAEとCGANを組み合わせることで，高品質なデータの生成と生成データの制御性を両立する手法が，CVAE-GAN[3]である（**図2**）。前半部分がCVAE，後半部分がCGANで構成されている。

3.2　文字データの生成例

CVAE-GANを用いて文字データを生成した。比較のために，CVAE，CGAN，及びCVAE-GANを用いて，様々な書き癖の「あ」の文字データを生成した例を**図3**に示す。

CVAEで生成したものには，右端列上段の文字データなど，不自然なものが含まれている。また，CGANで生成した文字データには，左から5列目中段のドットノイズや，5列目下段の線の途切れのような，実際の筆記では発生しにくい劣化が見られる。これらに比べ，CVAE-GANで生成した文字データには，壊れたパターンや不自然な劣化が少なく，より品質の良い文字データが生成されている。このように，生成した文字データを認識モデルに学習させることで，AI-OCRの認識精度を高めることができる。

4. 未教示データを学習に活用する半教師あり学習

AI-OCRの多くはクラウドサービスとして提供されており，クラウドシステム上に蓄積される文字データを学習用文字データとして利用できる。ただし，これらは，未教示データであることが多く，教示にはコストが掛かるため，未教示データも学習に活用できることが望ましい。

図3. 3種類の手法による文字データの生成例

CVAE-GANで生成した文字データは，壊れたパターンや不自然な劣化（線の途切れ）などが少なく，最も良い品質が得られた。

Images generated by CVAE, CGAN, and CVAE-GAN

4.1 VATによる半教師あり学習

未教示データを学習に活用する手法に，一部は教示データ，残りは未教示データを用いる半教師あり学習がある。

半教師あり学習にも適用できるディープラーニングの手法では，VAT（Virtual Adversarial Training）[4]が知られている。VATは，ある学習用データとその周辺で，認識結果（事後確率分布）が大きく変動しない（学習用データとその周辺のデータの確率分布間距離が小さい）という制約を付けて，学習させる手法である。直観的には，カテゴリー間の識別境界を決定する際に，学習用データ周辺も学習用データと同じカテゴリーになるような制約を与えていると解釈できる。VATには，過学習を抑制して汎化性能を向上させる効果があることが知られている。VATでは，学習用データ周辺での事後確率分布の計算に，学習用データのカテゴリー情報は不要であるため，未教示データも学習用データとして用いることができる。

VATを用いた半教師あり学習の特長を**図4**に示す。VATを用いた半教師あり学習では，半教師あり学習によって学習用データの密度が上がり，更に各学習用データの近傍にあるデータは学習用データとほぼ同じ結果であるという制約があるため，認識結果が文字データの変動に対して安定する。これにより汎化性能を向上させるとともに，認識精度も向上できる。

4.2 文字列認識へのVAT適用

AI技術の進展により，文字単位ではなく文字列の単位で認識する文字列認識手法が実用化されている。文字列認識

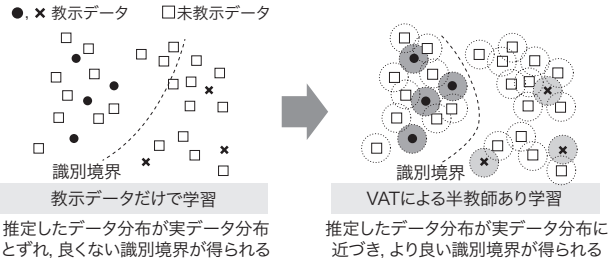

図4. VATを用いた半教師あり学習による識別境界の適正化

半教師あり学習で学習用データの数を増やし，更に，VATの活用で周辺データとの確率分布間距離を考慮することで，より正確な識別境界が得られる。

Changes in decision boundary by means of semi-supervised learning

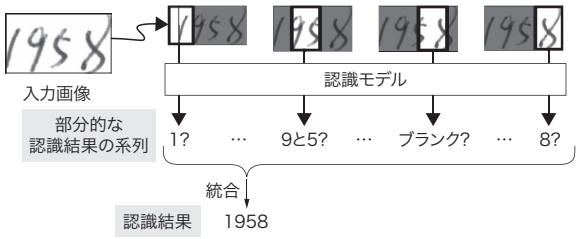

図5. 文字列認識結果

ウインドウをずらしながら部分的な認識結果の系列を求め，最終的にそれらを統合して文字列全体を認識する。文字の切れ目が分かりにくい手書き文字などに有効である。

Results of character string recognition

では，**図5**のように，部分的な認識結果の系列を求め，それを統合して文字列全体としての認識結果を求める。代表的な手法としては，CTC（Connectionist Temporal Classification）[5]が知られている。

文字列認識にVATを適用する場合，文字列の事後確率分布，つまり認識結果となる可能性のある全ての文字列の事後確率を計算するコストが膨大であるという問題があった。例えば，数字列認識の場合，認識対象カテゴリー数N_cを11（0～9の数字とブランク文字），部分的な認識結果の系列長Lを25と仮定すると，演算回数は11^{25}回にもなり，事実上計算することは不可能である。日本語認識の場合，N_cは数千に達するため，更に演算回数が多くなる。

そこで，より簡単に確率分布間距離の計算を行うFDS（Fast Distributional Smoothing）という手法を開発した[6]。FDSでは，式(1)の関係を利用した近似計算を行うことで演算回数を大幅に削減し，文字列認識でもVATと同様の学習が行えるようにした。

$$\text{文字列の確率分布間距離} \leq \sum^{L} \text{系列各部の確率分布間距離} \quad (1)$$

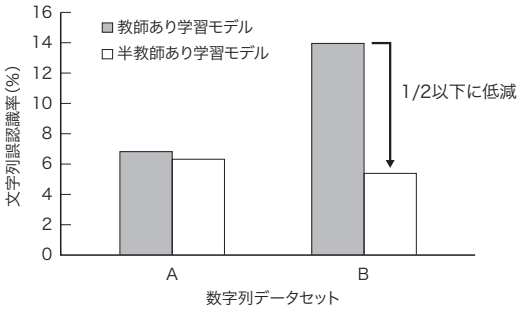

図6. 半教師あり学習による文字列誤認識率の低減

FDSを用いた半教師あり学習により，認識精度を大幅に向上できることが確認できた。

Improvement of character recognition accuracy by means of semi-supervised learning

同様の数字列認識の場合，演算回数は11×25回に削減され，実行可能になる。

4.3　未教示の数字列データを用いた学習の効果

FDSによる，数字列データの半教師あり学習を実施し，認識精度向上の効果を確認した。ここでは，教示データ及び教示データとは性質が異なる未教示データを使って，半教師あり学習を実施する例を示す。

性質が異なる二つの筆記者グループから収集した，手書きの数字列データセットA及びBを用いる。数字列データセットAは教示済み，Bは未教示である。実験では，以下二つのモデルによる各データセットの文字列誤認識率を評価した。

(1)　数字列データセットAだけを用いて教師あり学習をさせた教師あり学習モデル

(2)　(1)に加え，数字列データセットBを用いて半教師あり学習をさせた半教師あり学習モデル

数字列データセットA，Bそれぞれに対する各モデルの文字列誤認識率を**図6**に示す。数字列データセットBでは，教師あり学習モデルに比べて半教師あり学習モデルの誤認識率が，1/2以下に低減されており，半教師あり学習の効果が確認できた。一方，数字列データセットAでは，半教師あり学習の前後で誤認識率の目立った変化はなく，半教師あり学習による悪影響は見られなかった。

この結果から，未教示のままであっても，新たに与えられた数字列データセットを学習させることで，認識精度を向上できることが確認された。このような学習手法を組み込むことにより，運用中に未教示データを学習して認識精度を向上させていくAI-OCRが実現できる。

5. あとがき

AI-OCRの実用化のために開発した，認識モデルの学習手法へのAI技術適用事例について述べた。学習用文字データ生成手法や未教示データを用いた半教師あり学習などの手法を適用し，文字認識精度を高められることを確認した。

今後も，認識モデルの継続的な改良で認識精度を向上させるとともに，OCRサービスの使い勝手の向上などにもAI技術を活用していく。

文　献

(1) Kingma, D. P.; Welling, M. "Auto-Encoding Variational Bayes". International Conference on Learning Representations (ICLR) 2014. Banff, Canada, 2014-04, ICLR. arXiv.org e-Print archive, 2014, arXiv:1312.6114v10. <https://arxiv.org/pdf/1312.6114.pdf>, (accessed 2019-05-27).

(2) Goodfellow, I. J. et al. "Generative Adversarial Nets". Proceedings of Neural Information Processing Systems (NIPS 2014). Montreal, Canada, 2014-12, NIPS. arXiv.org e-Print archive, 2014, arXiv:1406.2661v1. <https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>, (accessed 2019-05-27).

(3) Bao, J. et al. "CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training". 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy, 2017-10, IEEE. 2017, p.2764–2773. <https://arxiv.org/pdf/1703.10155.pdf>, (accessed 2019-05-27).

(4) Miyato, T. et al. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access). 2018, arXiv:1704.03976v2. <https://arxiv.org/pdf/1704.03976.pdf>, (accessed 2019-05-27).

(5) Graves, A. et al. "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks". Proceedings of the 23rd International Conference on Machine Learning (ICML). Pittsburgh, PA, 2006-06, International Machine Learning Society. Association for Computing Machinery, 2006, p.369–376. <https://www.cs.toronto.edu/~graves/icml_2006.pdf>, (accessed 2019-05-27).

(6) 田中遼平，ほか．CTC-VATのための高速事後分布平滑化手法及びその文字列認識への応用．信学技報．2018，**118**，362，p.29–34.

古畑　彰夫　FURUHATA Akio
東芝デジタルソリューションズ（株）
ソフトウェア& AI テクノロジーセンター 知識・メディア処理技術開発部
電子情報通信学会会員
Toshiba Digital Solutions Corp.

田中　遼平　TANAKA Ryohei
東芝デジタルソリューションズ（株）
ソフトウェア& AI テクノロジーセンター 知識・メディア処理技術開発部
Toshiba Digital Solutions Corp.

長田　邦男　OSADA Kunio
東芝デジタルソリューションズ（株）
ソフトウェア& AI テクノロジーセンター 知識・メディア処理技術開発部
Toshiba Digital Solutions Corp.