

適合率最大化を目的とした CNN アンサンブル学習による文字認識

Character Recognition to Maximize True Positive Rate
by CNN with Ensemble Learning秦 優哉[†]
Yuya Hata高橋 大成[†]
Taisei Takahashi石井 大成[‡]
Taisei Ishi大枝 真一[‡]
Shinichi Oeda

1. はじめに

日本では年々高齢化が進んでおり、介護の必要性が高まっている。特に、訪問介護の需要は高く、今後も利用者の増加が予想されるが、解決しなければならない問題が残されている。その一つとして介護記録の管理方法が挙げられる。訪問介護の業務の一つとして、介護者は実施した介護の内容や要介護者の健康状態などを紙の記録用紙に記入し、他の介護士と情報共有を行っている。しかし、記録用紙が綴じられたファイルを持ち出すことは困難であるため、介護者は施設外の訪問先では記録用紙を参照することができない。また、閲覧の際にはファイリングされた大量の記録用紙から必要な情報を探し出さなければならない。この問題を解決するため、CNN(Convolutional Neural Network) [1,2] を用いた手書き文字の認識による訪問介護支援システムを構築する。本研究では、Ensemble Learning を用いて、複数の CNN の結果からシステム全体の出力を判定及び評価する。

2. 先行研究

先行研究 [3] では、携帯情報端末を用いて介護の記録を入力・参照するシステムが構築された。しかし、携帯情報端末によるデータ入力は現場の業務に適しておらず、紙を用いたデータ入力が最善であることがわかった。これをふまえて、先行研究 [4] では、マークシート方式の記録用紙に手書きでデータ入力を行い、携帯情報端末で閲覧を行うシステムが構築された。しかし、2つの問題点により介護士の負担を軽減させることができなかった。1点はマークシート方式でのデータ記入は、要介護者ごとに設定した識別 ID を全て覚える必要がある点。もう1点はマークシート方式に対応した紙を扱う必要があるため、従来の記録用紙とは別のフォーマットを使わなければならない点である。

そこで我々はこれらを改善するために、データ入力は現状の業務形態を変えずに従来の記録用紙によってペンで行い、データ閲覧は携帯情報端末から記録用紙の閲覧が行えるシステムを提案してきた [5-8]。これらの先行研究では、手書きの記録用紙の苗字部分を CNN によって文字認識を行うシステムを提案した。文字認識とはいえ認識する必要のある苗字は対象となる訪問介護先の介護者の苗字だけでよい。平均的には 10 人から 20 人の介護者の苗字を認識できればよいため、数多

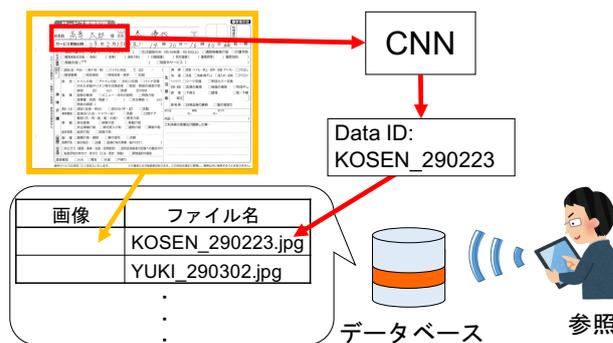


図 1: 構築する訪問介護支援システム。

くの手書き文字の認識を目標とする文字認識とは大きく異なる。

我々の先行研究の特徴は、介護士の従来の業務を変更せずに記録用紙の電子化を図ることである。電子化の際には手書きの日付と名前を CNN により自動的に認識してデータ ID を作成することでデータベースに格納し、データ参照を容易にする。特に、本研究では、Ensemble Learning を用いて誤認識を防ぎつつ認識精度を向上させる手法を提案し、またシステム構成について説明を行う。

3. 訪問介護支援システムの処理フロー

構築する訪問介護支援システムの概要を以下に示す。また、概略図を図 1 に示す。

1. 記録用紙をスキャンし、画像化する。
2. 画像が Web サーバに送られ、テンプレートマッチングによる特定領域の抽出が行われる。
3. 抽出された画像を入力とする CNN で文字認識が行われる。
4. 文字認識の結果からデータ ID を作成、画像に名前が付けられる。
5. 画像がデータベースに登録される。
6. 携帯情報端末からデータベースにアクセスすることで、訪問先で記録用紙が閲覧できるようになる。

訪問介護の現場でのデータ参照を可能とするため、画像化された記録用紙をデータベースに登録し、携帯情

[†]木更津工業高等専門学校 制御・情報システム工学専攻, Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

[‡]木更津工業高等専門学校 情報工学科, Department of Information and Computer Engineering, National Institute of Technology, Kisarazu College

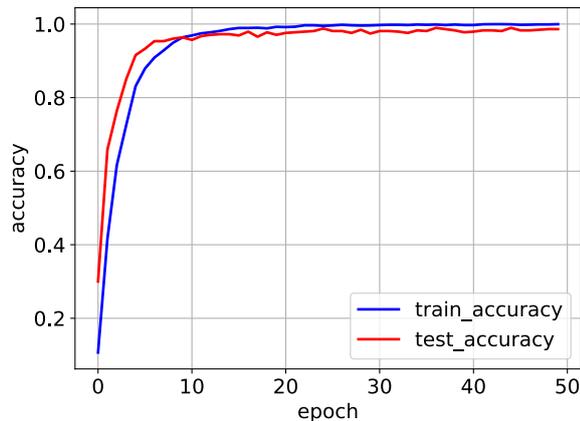


図 2: 単一の CNN の予測精度.

表 1: 2 値分類の指標.

		正解	
		真	偽
予測	真	TP (False Negative)	FP (True Negative)
	偽	FN (False Negative)	TN (True Negative)

報端末を用いて閲覧する。記録用紙をデータベースに登録するために CNN による文字認識の結果を用いてデータ ID を作成し、データ ID を画像に対応付けてデータベースに登録する。このデータベースを検索・参照することで記録用紙の閲覧を行う。

4. Ensemble Learning による文字認識

4.1. 単一 CNN による苗字認識

システムを構築する際に、CNN による文字認識の精度を調べる必要がある。そのために、手書き苗字の画像を対象として実験を行った。実験結果を図 2 に示す。CNN による文字認識の精度は約 98.9% だった。

4.2. Ensemble Learning による文字認識

単一の CNN での苗字の識別は 100% ではないため誤識別は必ず発生する。誤識別が起こることで、違う要介護者のデータが出力される恐れがある。そこで Ensemble Learning を用いて文字認識の結果の信頼性を調べる。Ensemble Learning は個々に学習した複数の弱学習器を用いて汎化性能を向上させ、1 つの強学習器を作成する手法である。

次に、評価方法について説明する。

2 値分類の評価では表 1 のような指標を用いることが多い。表 1 の FP (偽陽性) は '正しい' と予測して、正解が '正しくない' ことを表している。本研究では FP を 0 にすることが目的であるため Precision (適合率) に着

表 2: 多クラス分類の指標.

		正解	
		田中さん	田中さんでない
予測	田中さん	TP (False Negative)	FP (True Negative)
	田中さんでない	FN (False Negative)	TN (True Negative)

目する。Precision は以下のように表す。

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

また、多クラス分類では、表 2 がそれぞれのクラスにおいて作成される。N クラス分類の場合の Precision を導く式は以下のように表される。ここで c はクラスを表す。

$$Precision = \frac{\sum_{c=0}^{N-1} TP_c}{\sum_{c=0}^{N-1} (TP_c + FP_c)} \quad (2)$$

介護で用いられる支援システムであるため間違いがあってはならない。そのための、本研究では誤認識を 0 とした上で、全体の認識精度を向上させることを目的としている。したがって、式 (2) の値が 1 となるようなシステムを構築する必要がある。

5. 計算機実験

本研究では、誤識別が起こることを防ぐために確実に分類できるデータのみを扱い、それ以外は 'Unknown' としてどのクラスにも属さないデータとする。これにより識別できるデータ数は減少するが、識別したデータの識別精度は保証されるシステムの構築が可能となる。

実験の内容は以下ようになる。CNN を弱学習器として、Ensemble Learning を適用することで強学習器を作成する。学習が完了した強学習器にテストデータを入力することで Precision を測る。テストの際には弱学習器の出力によってテストデータは採用・不採用の判定がされる。採用の場合を図 3 に、不採用の場合を図 4 に示す。採用された画像のみを教師信号と比較することで Precision を測る。一つでも異なる出力があった場合は不採用となり、比較の対象から除く。学習時の CNN の個数を変化させ、採用した画像数と Precision の変化を調査する。

5.1. 手書き苗字の認識

Ensemble Learning を用いて CNN の個数を変化させ、Precision の変化を調べる。苗字の種類は 20 種類であるため、各 CNN の出力層は 20 となる。トレーニングデータ 2900 枚、テストデータ 580 枚で実験を行う。

20 種類の苗字を入力として識別を行った結果を表 3 に示す。

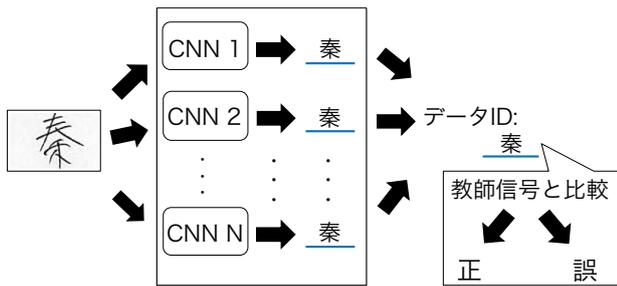


図 3: 全ての推定クラスが一致した場合.

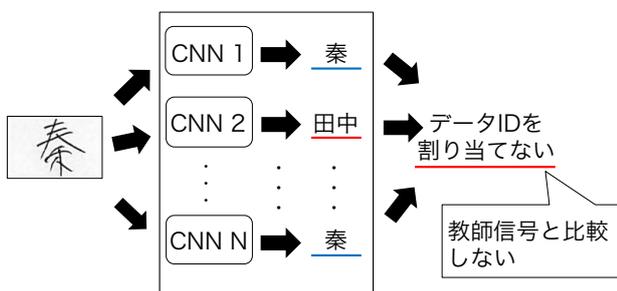


図 4: 推定クラスが一致しなかった場合.

CNNの個数を増加させると Precision を 1.00 にすることができた。ここで、CNNの個数が3個の場合と、10個の場合を比較する。どちらも Precision は 1.00 だが、採用された画像の割合は CNNの個数が3個の場合のほうが多い。そのため、CNNの個数が10個の場合よりも3個の場合の方が、多くの画像を間違いなく認識できていることがわかる。本実験では、採用した画像の枚数を高く維持したまま Precision が 1.00 である、CNNの個数が3個の場合が最善であることがわかる。

5.2. 手書き数字の認識

Ensemble Learning を用いて MNIST と呼ばれる手書き数字のデータセットを入力として Precision を測る。数字は 0 から 9 までの 10 種類であるため、各 CNN の出力層は 10 となる。トレーニングデータ 50000 枚、テストデータ 10000 枚で実験を行う。

10 種類の数字を入力として識別を行った結果を表 4 に示す。

CNNの個数を増加させても Precision が 1.00 にはならなかったことから、この実験では確実な推定を行うことができなかった。この原因として考えられるのは、Ensemble Learning の適用方法である。Ensemble Learning は多角的に物事を捉える手法である。しかし、弱学習器としていた CNN の構造に大きな差異がなかったため、真価を発揮できなかったと推測する。

5.3. 考察

先述の実験結果をふまえ、弱学習器の構造や画像の採用方法を検討することで、更に多くの画像の識別を

表 3: 苗字の分類結果.

CNN の数	採用された画像の枚数	採用された画像の割合	Precision
1	580	100.0%	0.953
2	560	96.5%	0.985
3	545	93.9%	1.00
4	541	93.2%	1.00
5	534	92.0%	1.00
6	527	90.8%	1.00
7	525	90.5%	1.00
8	519	89.4%	1.00
9	512	88.2%	1.00
10	509	87.7%	1.00

表 4: 数字の分類結果.

CNN の数	採用された画像の枚数	採用された画像の割合	Precision
1	10000	100.0%	0.992
2	9924	99.2%	0.996
3	9883	98.8%	0.997
4	9856	98.5%	0.998
5	9842	98.4%	0.998
6	9829	98.2%	0.998
7	9807	98.0%	0.998
8	9795	97.9%	0.998
9	9783	97.8%	0.998
10	9758	97.5%	0.999

目指す。弱学習器は畳み込み層・プーリング層の数やフィルタ、受容野の大きさを変更することで改善を試みる。また、先述の実験では全ての弱学習器が満場一致で同じ出力をした場合のみデータを採用していたが、弱学習器ごとに重みを設定、多数決によって出力を求める方式に変更する。この際、最も得票数が高くても一定の値に達していない場合は対象のデータは不採用となる。

5 種類の CNN を 5 個ずつ計 25 個の CNN で学習を行い、改善した採用方法でテストを行う。扱うデータは先述の実験で用いた苗字の画像である。結果は Precision が 1.00、不採用画像が 16 枚となった。先述の実験で、最適なモデルの不採用画像の枚数が 35 枚であったため、モデルの改善が成功していることがわかる。

6. データベース管理

提案システムのデータベース管理の概要図を図 5 に示す。

6.1. データの登録

本研究で提案するシステムは、介護士が記入した記録用紙をスキャナにセットし、シェルスクリプトを実行すると自動でデータベース登録が行われるものである。データベース登録を行うシェルスクリプトの概要を以下に示す。また、使用する機器は「ScanSnap S1500M(富



図 5: データベース管理の概要.

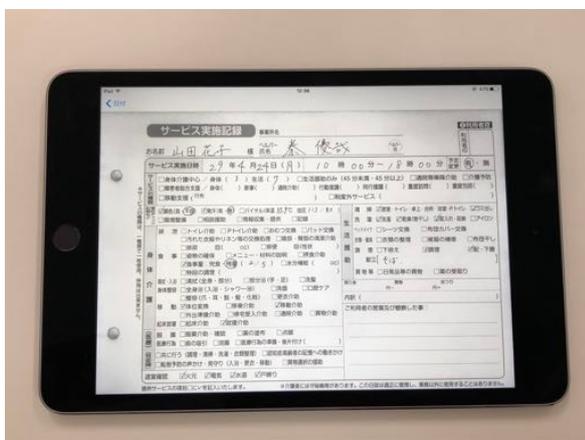


図 6: 携帯情報端末からの参照.

士通株式会社)」、「Raspberry Pi 3 Model B(Raspberry Pi Foundation)」である。Raspberry Pi は ScanSnap を管理するためのサーバとして使用する。

1. Raspberry Pi から SANE(Scanner Access Now Easy) を用いて ScanSnap のスキャンを実行する
2. テンプレートマッチングなどの画像処理によりスキャンした画像から必要な領域を抽出する
3. Ensemble Learning と CNN による文字識別を行い、データ ID を作成する
4. 画像とデータ ID を対応付け、データベースに登録する。

6.2. データの参照

データベースに登録された画像を訪問先で閲覧するために、外部からデータベースにアクセスする必要がある。そこで、本研究では、携帯情報端末上で画像の閲覧を可能とする iOS アプリを作成する。作成する iOS アプリは日付と名前を入力することで、データベースから対応する画像を表示するものである。携帯情報端末からのデータ参照を図 6 に示す。

7. まとめ

Ensemble Learning を適用することで苗字画像での実験では Precision を 1.00 にすることができた。よって、本研究で構築したシステムを用いることで、誤りが全く無い記録用紙の電子化が可能となった。今後の課題としては、CNN のモデルや Ensemble Learning の構造の検討が挙げられる。また、一貫したシステムの作成や実際の現場のデータを用いての実験、プロトタイプを現場で運用することによる評価などを行う必要がある。

謝辞

本研究は、木更津商工会議所 医療・福祉・教育部会および、株式会社 R.O.F., 富沢産業株式会社, TDC ソフト株式会社との共同研究です。また、株式会社宙 SOLA の支援を受けているものです。

参考文献

- [1] Le Cun, Y., L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86, 2278-2324, 1998.
- [2] Alex Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 1097-1105, (2012).
- [3] 川崎直輝, 大枝真一, "携帯情報端末を用いた特別養護老人ホームの利用者データの電子化とその有効性の検証", 情報処理学会第 70 回全国大会, 2ZF-7, 2008.
- [4] 加藤雄大, 大枝真一, "介護現場で用いる記録用紙の電子データ化システムの構築", 情報処理学会第 74 回全国大会, 1ZJ-2, 2012.
- [5] 小森一誠, 秦優哉, 高橋大成, 大枝真一, "データ統合による CNN を用いた訪問介護支援システムの構築", FIT2017(第 16 回情報科学技術フォーラム), CJ-009, 2017 年.
- [6] Yuya Hata, Shinichi Oeda, "Home-visit Nursing Care Support System by CNN with Ensemble Learning", International Workshop on Effective Engineering Education (IWEEE2017), 2017.
- [7] 秦優哉, 小森一誠, 川名晴也, 大枝真一, "CNN のアンサンブル学習による文字認識の正誤判定評価", 情報処理学会第 80 回全国大会, 5ZD-09, 2018.
- [8] 小森一誠, 秦優哉, 川名晴也, 大枝真一, "CNN の文字認識を用いた訪問介護支援システム構築のための訓練データ加算", 情報処理学会第 80 回全国大会, 5ZD-08, 2018.