

Ensemble CNN による文字認識を用いた訪問介護システムの実証実験 Verification of Home-visit Nursing Care System Using Character Recognition by Ensemble Learning CNN

高橋 大成[†]
Taisei Takahashi

大枝 真一[‡]
Shinichi Oeda

1. 研究背景

現在、日本では少子高齢化などの理由により、訪問介護の需要が高まっている。訪問介護とはホームヘルパーが要介護者の自宅に直接訪問し、食事や入浴などの身体介助をはじめとした要介護者のサポートを行うサービスである。訪問介護では、業務終了後にデイ日誌と呼ばれるその日の介護記録を指定の用紙に記入する作業がある。記入したデイ日誌は事務所に保管される。デイ日誌は主に他のホームヘルパーと情報共有されるために使われるが、保管されていくデイ日誌は日を追うごとに増え続け、事務作業を圧迫していく一方である。そこで本研究では訪問介護のデイ日誌の電子化および、訪問介護システムの構築を目指す。

2. 先行研究

先行研究 [1] では電子化の手法として、携帯情報端末を用いた要介護者の情報検索および電子データの入力可能なシステムの構築がなされた。しかし、実際の介護現場では、携帯情報端末を用いて電子データを入力する作業は手間となってしまうのが現状であり、記録媒体としては紙とペンが適しているという知見を得た。そこで本研究では紙媒体でのデータを電子化する方針をとる。また先行研究 [2] では手書きの苗字部分を CNN によって文字認識を行うシステムを実際に提案している。この研究では文字認識の正答率が 90% となったが、この研究では文字認識の正答率が 90% であり、10% の誤認識が発生していた。この誤認識がどのようなものかということ、システムは入力された手書きの苗字画像を、分類対象となっている苗字 20 種類から予想したが、それが間違っていたことを意味する。介護医療の現場で、A さんを B さんと誤認識することは絶対にあってはならない。そこで、システムの予想に確信が持てない場合、その画像は除外して人間が判断することにする。その代わりに、システムが認識した画像に関しては絶対に間違いが発生しないシステムを構築したい。そこで本研究では Convolutional Neural Network(CNN) と Ensemble Learning と呼ばれる手法を組み合わせ、適合率を最大化するのを目的とする。ここでの適合率とは Precision のことであり、システムが出した結果に対して本当に正しかったものの割合を指す。

[†]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, Information and Computer Engineering, National Institute of Technology, Kisarazu College

3. 使用手法

3.1. Convolutional Neural Network

最近では計算機の演算の向上から Neural Network を用いた画像認識が数多く研究されている。特に画像の識別に特化した Convolutional Neural Network(CNN) が用いられる。CNN は通常の Neural Network に畳み込み演算を行う畳み込み層と情報を圧縮するプーリング層を新たに追加したモデルである。畳み込み層では空間的な情報を抽出する演算を行うため、画像の識別に適したモデルと言える。

3.2. Ensemble Learning

CNN の認識精度を高めても数パーセントの誤差は発生してしまう。識別した結果に誤差が多い、つまり適合率が低いシステムは実際に運用する際に良いとは言えない。そこで適合率を高められる手法である Ensemble Learning を用いる。適合率 (Precision) を式 (1) に示す。TP とはシステムが識別した結果が正答だった総数を示し、FP はシステムが識別した結果が誤答だったことを示す。つまり Ensemble Learning では FP の最小化を目指す。

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

Ensemble Learning は複数の弱学習器を生成し、弱学習器のそれぞれの推論の結果を統合して一組の分類器を生成する手法である。Ensemble Learning では弱学習器のエラーが無相間であれば、汎化性能が上がるということが知られている。また単一の弱学習器よりも正答率が高くなることも知られている。先行研究 [3] では、木更津高専の学生に作成を依頼したサンプル苗字データを用いて、Ensemble Learning による適合率の最大化を行った。本研究では、訪問介護で使われている実際の苗字データを入手したため、それらのデータを用いて実証実験を行う (以降画像データを“サンプル苗字データ”、“実際の苗字データ”と呼び分ける)。

4. 計算機実験

本研究では弱学習器を 10 個作成し、10 個の多数決を取ることで Ensemble Learning を行う。弱学習器は汎化性能を上げるために、CNN の畳み込み層・プーリング層の数やフィルタの大きさを変更しそれぞれ作成した。実際の苗字データは計 2047 枚あり、1000 枚を学習用に 500 枚を検証用に 547 枚をテスト用に割り当てた。クラスは 14 種類である。弱学習器を作成したあと、実際にテスト用データに対してそれぞれ推論を行い、結果を統合した。強学習器の推論は弱学習器の推論の結果の多数決によって決めた。ここで多数決の閾

値を決め、弱学習器の推論結果がどれだけ一致するかで識別を行うかどうかを判定し、識別を行なった場合の Precision と識別を行なった枚数の割合を調査した。

5. 実験結果

サンプル苗字データと実際の苗字データに対して Ensemble Learning を行なった。結果から得られた Precision と識別を行なった割合を表 1 に示す。サンプル苗字データに対する結果は先行研究 [3] を参考にした。この結果から多数決の閾値を高くすると識別を行う枚数は減少するが、Precision が高くなることがわかった。本研究では Precision を最大化することともに、識別枚数を増やすことも目的であるため、Precision が 1.00 の状態で何枚の画像を識別できたのかも評価の対象となる。サンプル苗字データに対しては Precision が 1.00 の状態で 93.9% の割合で画像を識別することができた。一方で実際の苗字データに対しては Precision が 1.00 の時に 71.7% の割合で画像を識別することができた。やはり、サンプル苗字データよりも複雑な実際の苗字データの方が正答率や Precision や採用された画像の割合も低いという結果になった。

弱学習の作成時に畳み込み層・プーリング層の数やフィルタの大きさを変更しそれぞれ作成したが、これの有効性を調べるために次の実験を行う。

実験 B: 前の実験で作成した弱学習器の中でテスト用データに対する正答率が一番高かったモデルの初期値を変えて弱学習器として 10 個用意する。用意した弱学習器を用いて Ensemble Learning を行い、前の実験と同様に多数決の閾値を決め、弱学習器の推論結果がどれだけ一致するかで識別を行うかどうかを判定し、識別を行なった場合の Precision と識別を行なった枚数の割合を調査した。

実験 C: 前の実験で作成した弱学習器の中でテスト用データに対する正答率が一番低かったモデルの初期値を変えて弱学習器として 10 個用意する。用意した弱学習器を用いて Ensemble Learning を行い、前の実験と同様に多数決の閾値を決め、弱学習器の推論結果がどれだけ一致するかで識別を行うかどうかを判定し、識別を行なった場合の Precision と識別を行なった枚数の割合を調査した。

実験 B と実験 C の結果を表 2 に示す。この表から単一のモデルのみを組み合わせると Precision が 1.00 にならないことがわかった。このことから前の実験で畳み込み層・プーリング層の数やフィルタの大きさを変更し弱学習器を作成したことは汎化性能の向上に繋がったと言える。

6. 今後の展望

今回の実験で畳み込み層・プーリング層の数やフィルタの大きさを変更することで Precision を最大化することができた。一方で実際に識別された割合は 71.7% であったため、残りの 28.3% は人の手で分類することになってしまう。そのため次の目標は Precision を最大化しつつ、識別する画像の枚数を増やすことである。具体的な手法としては Ensemble Learning の手法の一つ

表 1 サンプル苗字データと実際の苗字データの実際に識別された画像の割合と Precision

CNN の数	サンプル苗字データ		実際の苗字データ	
	識別された画像の割合	Precision	識別された画像の割合	precision
1	100.0%	0.953	100.0%	0.894
2	96.5%	0.985	100.0%	0.894
3	93.9%	1.00	100.0%	0.894
4	93.2%	1.00	99.5%	0.897
5	92.0%	1.00	96.0%	0.916
6	90.8%	1.00	94.7%	0.967
7	90.5%	1.00	85.7%	0.980
8	89.4%	1.00	79.5%	0.995
9	88.2%	1.00	71.7%	1.00
10	87.7%	1.00	58.1%	1.00

表 2 実験 B と実験 C の実際に識別された画像の割合と Precision

CNN の数	実験 B		実験 C	
	識別された画像の割合	Precision	識別された画像の割合	precision
1	100.0%	0.963	100.0%	0.832
2	100.0%	0.963	100.0%	0.832
3	100.0%	0.963	100.0%	0.832
4	99.8%	0.965	96.7%	0.860
5	98.9%	0.972	93.4%	0.875
6	95.6%	0.973	89.0%	0.901
7	94.3%	0.977	82.0%	0.930
8	92.3%	0.986	74.4%	0.948
9	90.3%	0.988	65.9%	0.972
10	85.6%	0.994	53.1%	0.979

である boosting を使用したい。また引き続き訪問介護システムの開発を行う。

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, SOLA および木更津高専技術振興交流会の支援を受けているものです。

参考文献

- [1] 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.
- [2] 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.
- [3] 秦 優哉, 高橋 大成, 大枝 真一, “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”, FIT2018(第 17 回情報科学技術フォーラム), CJ-005, 2018.
- [4] Thomas G. Dietterich, “Ensemble Methods in Machine Learning”, MCS2000, PP1-15, 2000.