



ディープラーニングを用いた大人と子供の手書き文字の識別

今村梨世

文京学院大学女子高等学校

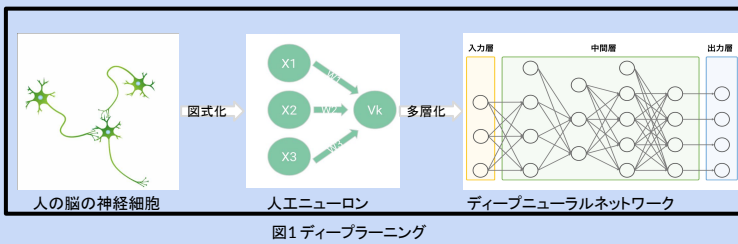


導入

インターネットの普及率が上がったことにより、SNS上で引き起こされる犯罪が増え、子供が犯罪に関わることも増えた。これらの犯罪から子供を守るために、何らかの手段を用いて大人が子供かをネット上でも識別できないかと考え、研究を始めた。本研究では、文字に着目し、ディープラーニングを用いた実験をした。

目的

ディープラーニングで学習する文字データの種類を変えることで、大人と子供が書いた文字の識別率がどのように変化するかを調べる。また、識別の精度向上を目指す。



実験方法

全ての実験に共通する手順

- ①手書き文字データを集める
- ②集めたデータをデータセットにする
- ③ディープニューラルネットワークを作成する
- ④ディープラーニングの学習を実行し、評価を出す
ニューラルネットワークの構築・学習には、NNC(Neural Network Console)を用いた。

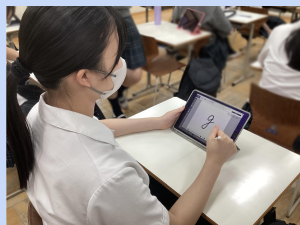


図4 データ集めの様子

実験1 アルファベット全体をデータセットとした識別実験(総データ数4254枚)

実験2 「a」~「z」の文字別にデータセットを作成して行った識別実験(データ数1文字163枚)

実験3 識別率が高い文字に関して、データ量を増やして行った識別実験(データ数1文字520枚)

実験4 文字をいくつか組み合わせた単語での識別実験(データ数1単語520枚)

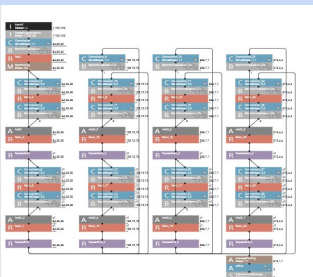


図5 使用したディープニューラルネットワーク

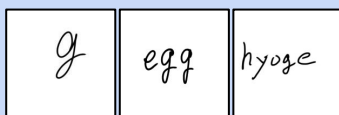


図6 子供の文字のサンプル

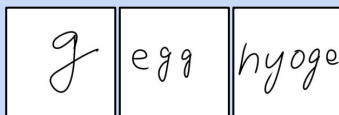


図7 大人の文字のサンプル

結果

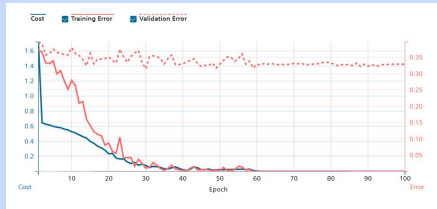


図8 実験1の学習曲線

曲線の下が安定しておりデータ量は十分に足りているが、トレーニングエラーとバリデーションエラーの差が大きい。

表1 実験1の結果

	1回目	2回目	平均値
a-z	72.05%	71.46%	71.76%

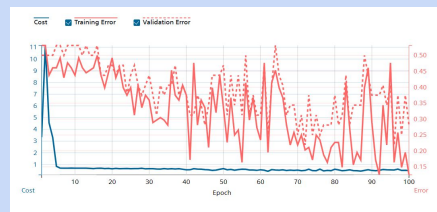


図9 実験2の学習曲線

学習曲線の下が激しいため、データ量不足が推測される。

表2 実験2の結果

	1回目	2回目	平均値
a	43.33%	46.67%	46.50%
b	71.64%	73.25%	72.45%
i	72.35%	72.43%	72.39%
m	70.58%	73.51%	72.05%
e	81.11%	81.67%	81.39%
g	83.33%	87.37%	85.35%
h	84.56%	85.29%	86.43%
o	86.63%	91.39%	89.01%
y	85.44%	88.07%	86.76%

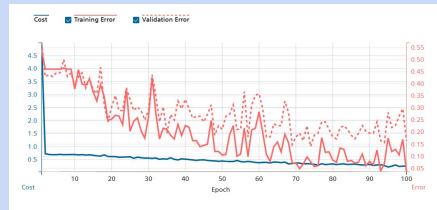


図10 実験3の学習曲線

データ量を増やしたことで、図6より安定した学習曲線になった。

表3 実験3の結果

	1回目	2回目	平均値
h	79.57%	81.62%	80.60%
g	79.48%	82.87%	81.18%
o	77.23%	75.48%	76.36%

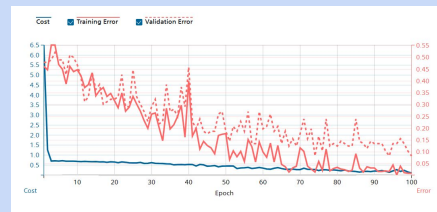


図11 実験4の学習曲線

トレーニングエラーとバリデーションエラーの差が小さく、緩やかに下がっている、良い学習曲線といえる。

表4 実験4の結果

	1回目	2回目	平均値
egg	86.43%	88.85%	87.64%
gyo	91.36%	92.34%	91.85%
hyoge	89.40%	90.48%	89.94%

結論

- データに使用する単語の文字数を増やすことによって識別率が向上する。しかし5文字を超えると識別の精度が落ちたことから大人と子供の文字の識別に最適な文字数が存在する可能性がある。
 - ニューラルネットワークに読み込ませるデータの種類を変えることによって識別精度は変化する。
- 本実験では、実験1の72%から実験4の92%と、識別率を20%も向上させることができた。

参照

- (1)機械学習を用いた筆跡鑑定/大井健太、杉田誠
https://ipsj.ixsq.nii.ac.jp/ei/index.php?action=pages_view_main&active_action=repository_action_common_download&item_id=201450&item_no=1&attribute_id=1&file_no=1&page_id=13&block_id=8
- (2)画像認識のための深層学習の研究動向～畳込みニューラルネットワークとその利用法の発展～/岡谷貴之/東北大学大学院情報科学研究科
https://www.istage.ist.go.jp/article/istai/31/2/31_169/_pdf-char/a
- (3)インターネット利用に係る子供の犯罪被害等の防止について/令和四年統計/警察庁生活安全局人身安全・少年課課長/阿波拓洋
https://www.cfa.go.jp/assets/contents/node/basic_page/field_ref_resources/5a7c1011-f85c-4122-a1bf-aa960545da19/924b655d/20230630_policies_vk_hikouhigai-gakkan-kouenkai2.pdf
- (4)ディープラーニング/岡谷貴之/大学大学院情報科学研究科
https://www.istage.ist.go.jp/article/istai/68/6/68_466/_pdf-char/a