

先端人工知能論最終課題レポート

下村拓

2018/08/03

1 設定した問題の概要

日本人、韓国人、中国人などはとても似た顔をしている。しかし、我々日本人は微妙な違いを見分けかなりの精度でこれらの国を見分けることができる。このタスクを機械が同様の精度で行うことができれば様々な場所で活躍できると考えられる。例えば、液晶に映る日本語と外国語の表示切り替えを前に立っている外国人の数等で、時間配分を変えることが可能であると考えている。よって、今回は日本人か外国人かの判定を CNN に行わせることとする。

2 利用した手法の概要と工夫点

モデルのアーキテクチャとしては基本的に AlexNet のものを用いるところから始めた。ただし、 50×50 の画像を入力するため、次元数は異なる。

まずは画像データについて、正規化するために 255 で割ることで 0 から 1 の浮動小数点とした。また、全ての画素値の平均と分散を求め、

$$\frac{data - mean}{std} \quad (1)$$

とすることで標準化を行った。

次にデータオーギュメンテーションを行った。画像を水平に移動させるものと、画像を水平に反転させるものの二種類のデータ拡張の手法を用いた。

モデルについては、AlexNet をベースにしたが、次元が減ることが不利になるのではないかと考え、MaxPooling 層を削った。Max Pooling 層の代わりに Convolution 層を増やした。また、学習をしていく上で、常にヴァリデーションデータとトレインデータのロスの差が大きかったことから、過学習を抑制できていないと考えられたため、Dropout の層を増やした。最後の全結合については、データを単に平滑化すると重みの数が増えるため、Global Average Polling 層を用いて 1 次元データにした。

今回のデータセットは正解ラベルに偏りが大きいものであったため、その課題を修正するために工夫が必要だった。ダウンサンプリングはデータ数が少なかったため適切ではなく、オーバーサンプリングは顔の特徴量を損なうことなくデータ拡張する手法が見つからなかった。そのため損失関数に重みをつけることとした。交差エントロピー誤差を損失関数としていたが、外国人のラベルで誤った際の損失を 6 倍とした。6 倍の値はデータ数の偏りに起因する。

3 利用したデータ

今回は J リーグ選手の顔を用いた。理由は、外国人選手が含まれていることと一括で大量のラベル付き顔写真が得られるからである。J リーグ公式の選手名鑑から 50×50 の画像と身長や体重、出身地の情報を取得した。出身地の情報から日本の都道府県であるかないかで日本人と外国人のラベルをつけた。

データは全部で 601 件あり、日本人と外国人はそれぞれ 516, 85 件であった。日本人の数は外国人のおよそ 6 倍であり、この値から損失のハイパーパラメータを決定した。

4 得られた結果と評価

全データの 7 割を訓練用に、1 割を検証用、2 割をテスト用として扱った。下表は最終的な訓練、検証、テストの精度、ロスを並べたものである。

表 1 学習結果

train accuracy	0.9228
train loss	5.3523
validation accuracy	0.8871
validation loss	8.0523
test accuracy	0.9024
test loss	12.66

テスト画像は 123 件あり、日本人が 103 件、外国人が 20 件存在した。その中で誤ったテストは 12 件で日本人が 11 人、外国人が 1 人となっている。上表の環境の下、テストで誤った画像を下に示す。

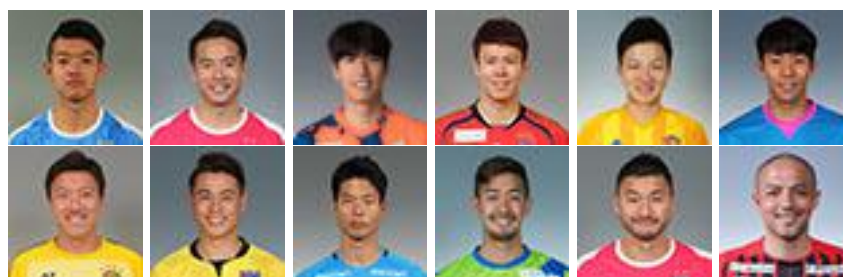


図 1 テストで誤った画像

また、学習の進捗状況を数に示す。

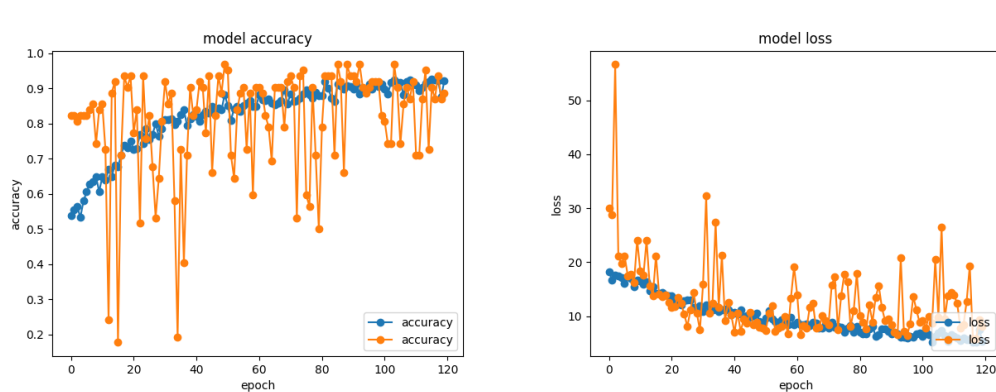


図2 学習の進捗

図からは学習がそれほど過学習せずに進んだことがわかる。これはデータ拡張や Dropout 層を積んだことの成果であると考えられる。

精度はおよそ 9 割となっており、外国人の誤りは 1 件だけであったことから損失関数のペナルティに関する重み付けも効果があったと考えられる。

本モデルを向上させるためにさらにできる工夫は

- データを増やす
- 外国人のデータを拡張する

などが考えられる。

気になる点として、validation の学習曲線がジグザグになってしまっていることがあげられる。この現象は頑健性が乏しかったことに起因するのかと予想したが、改善する手法が思いつかなかった。