F-026

非構造化データに基づく感情分析の検討

A Study of Emotion Analysis Based on Unstructured Data

福家 穂乃佳† 浦野 昌一†
Honoka Fuke Shoichi Urano

1. はじめに

スポーツの試合のハイライト動画は SNS 上で多く見受けられるようになったが、編集にかかるコストは現在も課題である。これまでに、野球の放送動画から選手の背番号のテロップを基に分析し、打席シーンを特定することでハイライト動画の作成を自動で行う研究が報告されている[1]。しかし、野球に特化したテロップ分析を行っているため、野球以外のスポーツには適応できず、根本的な課題解決は計れない。そこで、スポーツの種類に捉われないハイライト動画自動作成システムの構築に向け、観客の表情から感情分析を行い、ハイライトシーンを抽出することを目指す。今回は、ハイライトシーンの際の感情として考えられる喜びに着目する。

また、新型コロナウイルスなどの影響により近年、マスクを着用する人が増加している。そのため、観客の表情から喜びを分析するには、目元からの情報が重要となることも想定される。一方、喜びは口元で笑顔を表現することにより分析可能な感情と推測する。そこで、本稿では、深層学習を用いた笑顔判別では顔全体、目元、口元のいずれが効果的か検証を行う。

2. 笑顔の判断

人間が笑顔を判断する際にどの部分が重要な判断材料となるかを調査した研究がある[2]。その研究では、真顔、弱い笑顔(口角を引く動作)、優しい笑顔(頬を引き上げる動作を追加)、強い笑顔(口を開ける動作を追加)の 4 種類の表情のうち、顔全体、目元、口元のどの部分から人間は笑顔をどれだけ強く感じるかを検証している。

研究結果より、顔全体、目元、口元のいずれにおいても笑顔が強くなるにつれ笑顔をより強く認識できるということが分かっている。一方で、いずれの笑顔の強さにおいても顔全体と口元に比べ、目元の場合は実際の笑顔よりも弱く認識された。そのことから、人間は笑顔が弱い場合も強い場合も口元の微妙な変化からどのような笑顔なのかを判断していると考えられ、笑顔を認識する際、口元の情報は重要な手がかりである可能性が高いと結論付けられた。

3. 分析手法

3.1 畳み込みニューラルネットワーク(CNN)

畳み込みニューラルネットワークは画像認識の分野で広く使用される手法の一つである。すべてのピクセルから特徴を抽出しているため、画像内のどの部分に認識目標物があっても認識可能な移動不変性を持つ。畳み込みニューラルネットワークの構造例を図 1 に示す。

†明治大学大学院先端数理科学研究科 Meiji University
Graduate School of Advanced Mathematical Sciences

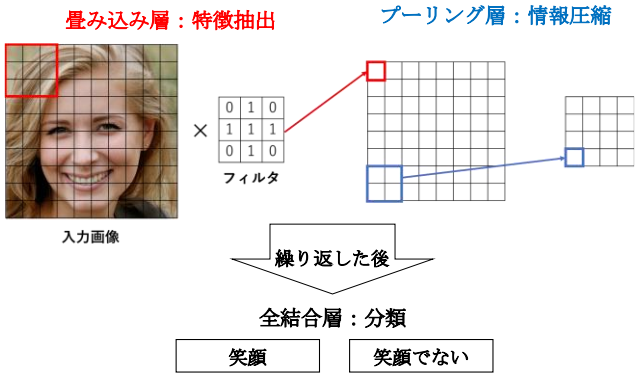

図 1 CNN の構造例

3.2 顔のランドマーク検出

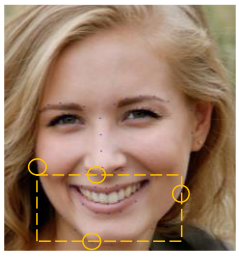

図 2 口元画像の抽出例

本稿では、顔のランドマーク検出機能を持つ dlib を使用し、任意の 4 点を指定することで口元と目元の画像を抽出する。その例を図 2 に示す。

3.3 ヒートマップによる可視化

本稿では、モデルが画像内のどの部分を笑顔の判断材料としているかをヒートマップにより確認する。ヒートマップ表示には Grad-CAM[3]を使用する。Grad-CAM は特徴マップの勾配から重みを算出し、モデルが注視している部分を可視化している。

4. シミュレーション

本稿では、外国人の顔画像で構成される CelebA データセットを使用する。顔全体、口元のみ、目元のみの画像を用いて 3 つのモデルを作成し、検証を行う。図 3 にシミュレーションの概要、図 4 にモデルの構造を示す。

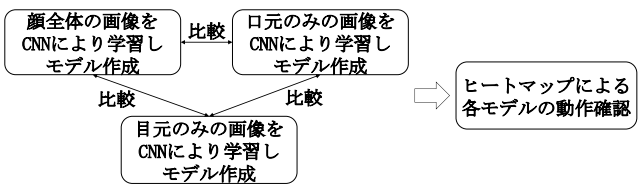

図 3 シミュレーションの概要

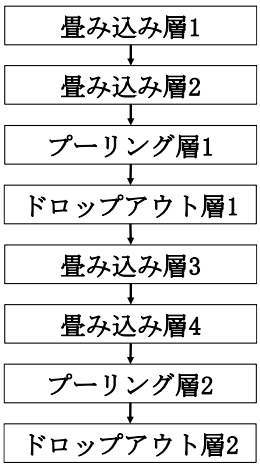

図 4　モデルの構造

シミュレーションの手順を以下に示す。

1　顔の全体画像として 178×178 ピクセルの画像を使用
2　CNN モデルの作成
　Case1：顔全体の画像を用いてモデル作成
　Case2：口元のみの画像を用いてモデル作成
　Case3：目元のみの画像を用いてモデル作成
3　ヒートマップを用いてモデルの動作確認

4.1 シミュレーション条件

データセットの内訳は以下で設定する。
＜学習用データ＞
笑顔画像：28,517 枚
笑顔でない画像：28,517 枚

＜検証用データ＞
笑顔画像：9,053 枚
笑顔でない画像：9,053 枚

＜評価用データ＞
笑顔画像：837 枚
笑顔でない画像：837 枚

評価指標は正解率、適合率、再現率、F 値を使用する。適合率と再現率はトレードオフの関係であるため、F 値でモデルのバランスを確認した上で、正確に笑顔を識別できるモデルに向け、本稿では、適合率の高さに注目する。

5. シミュレーション結果

表 1 に各モデルの評価指標の結果を示す。また、図 5 と図 6 には各モデルにおいて笑顔と正解した際の画像を用いて、ヒートマップによる判別根拠の可視化を行った例を示す。

表 1　モデルの精度比較

	正解率	適合率	再現率	F 値
Case1	64.5%	62.1%	74.3%	0.676
Case2	76.3%	86.6%	62.4%	0.725
Case3	57.6%	83.7%	18.1%	0.297

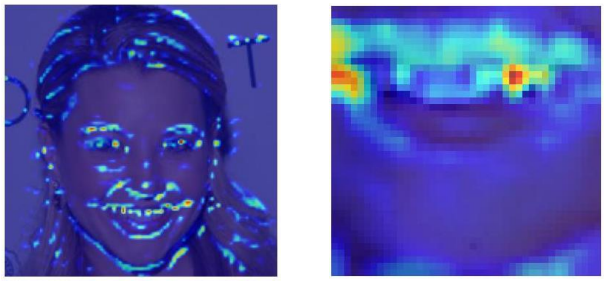

図 5　判別根拠の可視化例（左：Case1，右：Case2）

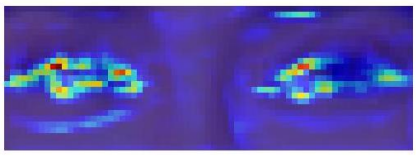

図 6　判別根拠の可視化例（Case3）

6. まとめ

本稿では、喜びの際の表情である笑顔を判別する際に顔全体、口元、目元のいずれが有効かを検証するため、それぞれの画像を学習に使用し、3 つの CNN モデルを作成した。

図 5、6 より、Case1 では顔のパーツ一つ一つを注視し、顔全体から笑顔を判別していることを確認した。Case2 では口が開いている部分や口角から笑顔を判別していることを確認した。Case3 では目の輪郭から判別を行っていることを確認した。このことから、いずれのモデルにおいても笑顔を判別する際に重要と考えられる特徴を捉えることが可能と考える。

表 1 より、最も高い適合率を示したのは口元のみの画像を使用した Case2 であり、次いで Case3、Case1 となった。しかし、正解率と F 値においては Case2 の次に Case1 が高い値となった。このことから、バランス良く総合的に高い精度を示しているのは Case2 であり、深層学習を用いた場合においても口元は笑顔の判別に有効な顔のパーツであると考える。また、Case3 の正解率が低かったのは、今回のデータセットに使用した顔画像の目元部分だけでは、人間が見ても笑顔か笑顔でないかの判別が容易ではなかったことが原因と考えられる。このことから、人間の判断と同様に深層学習を用いた場合においても目元のみでは笑顔かどうかの判別は困難であることが示唆され、目元による笑顔判別には更なる検討が必要であることが分かった。

今後は、目元からの感情分析が可能な表情の調査に取り組み、感情分析の汎用性を高めていく。

参考文献

[1] 岡田健司, 山田航平, 平川豊, 大関和夫, “野球放送動画のテロップ情報を用いたハイライトシーンの作成”, 情報処理学会第 75 回全国大会, pp.445-446, (2013).
[2] 渡辺有香, 松本秀彦, 諸冨隆, “目と口の部分提示に対する笑顔の強さの評価について”, 作新学院大学人間文化学部紀要第 7 号, pp.61-79, (2009).
[3] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, IEEE, International Conference on Computer Vision (ICCV), pp. 618-626, (2017).