

表情認識を用いた関心度推定方式の研究 —オンラインミーティング参加者を対象に—

2017SE017 本田将己
指導教員: 野呂昌満

1 はじめに

現在, 表情認識に関する研究が数多く存在し, 様々な分野に活かされている. 映像コンテンツに対する関心度推定システム [1] が提案されており, オンラインミーティングを対象とした関心度推定の実現が期待できる.

オンラインミーティングを対象とする場合, 既存研究では以下の問題が挙げられる. 既存の関心度推定システム [1] では, 推定対象ごとの表情特徴を事前登録する必要があるが, 推定対象の人物の変更や増加が想定される場面に適応されていない. 表情認識を用いた感情認識の研究 [2] では, 顔特徴の動きから予測される感情をあらかじめ定義し, 推定対象の表情から感情を推定を行うので, 関心度に結びつく顔表情の特徴についての知識が必要である.

本研究の目的は, 表情認識を用いた関心度推定手法の設計と, その妥当性の検証である. 表情特徴の手動定義を行わずに, 表情認識を用いて特徴量の検出を行い, 参加者の変更や増加が想定されるオンラインミーティングの場面でも適応させるために, 不特定多数の人物を推定対象とする関心度推定器の設計を行う.

本研究の技術課題は以下のとおりである.

1. 関心度推定処理の設計
2. 提案手法の妥当性検証

これらの解決によって, 表情の手動定義が求められる表情認識において, 特徴の手動定義を行わない場合でも有効であることが明らかになり, 不特定多数の人物を対象とした関心度推定が有効であると明らかにできると考える.

2 技術課題の解決

2.1 提案手法の概要

本研究では, 推定識別器に CNN を用いた関心度推定手法を提案する. Web カメラで撮影した顔画像から表情の特徴を抽出し, 抽出した情報をもとに関心度推定処理を行う上で, CNN が適していると考えた. 入力した顔画像から特徴を抽出し, 関心度を表す「関心なし」と「関心あり」の2種類に分類する.

提案する手法の概要を図 1 に示す. 本研究では, まず学

習用データセットとテスト用データセットの作成を行う. 次にニューラルネットワークの設計を行い, 学習用データセットで学習させる. 学習したニューラルネットワークにテスト用データセットを入力し, 出力値から入力した顔画像の関心度を推定する.

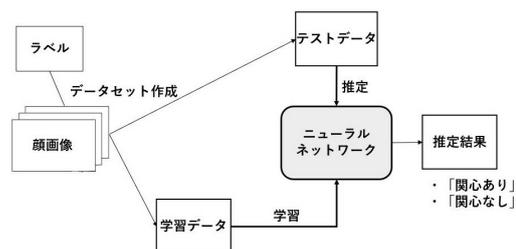


図 1 提案手法の概要

2.2 CNN

CNN の構造は, Fine-tuning [3][4] と呼ばれる学習手法に基づいて設計を行った. CNN の特徴として, 入力に近い層では汎用的な特徴を抽出し, 出力に近い層ほど学習に特化した特徴を抽出する傾向がある. 大量の学習データの重みを入力に近い層に再利用することで, 特徴量の抽出と画像分類の精度向上に期待できる.

本研究では, 学習済みニューラルネットワークとして高い精度結果を得ている VGG-16 [3] の構造とハイパーパラメータをベースに, 過学習の発生を緩和させるために, ドロップアウト層とバッチ正規化層を追加した設計を行った. 出力に近い畳み込み層と全結合層のみ学習データによって重みを更新し, 入力に近いそれ以外の層は VGG-16 の持つ重みを固定する.

2.3 データセットの作成

本研究で用いる学習データとテストデータには, オンラインミーティングに参加している 6 名 (男性 3 名, 女性 3 名) の RGB カラーの顔画像を用いる. 事前処理として, 画像の顔領域のみをトリミングし, CNN の入力層に合わせ, 画像サイズを 224×224 にリサイズした. 参加者自身が, 撮影時点で関心がなければ「関心なし」, 関心があれば「関心あり」と撮影した顔画像にそれぞれラベル付けをした. データ数を増やすために, それぞれ反転と平滑化を行い, 学習データ 9246 枚, テストデータ 2312 枚のデータセットを作成した.

3 実験

3.1 ニューラルネットワークの学習と精度検証

設計したニューラルネットワークの学習と精度検証を行った。各データの RGB 値を 0~1 に正規化したベクトルを入力し、各ラベルは「関心なし」を 0、「関心あり」を 1 とそれぞれ One-hot ベクトルに変換した値を入力した。

バッチサイズ 64, エポック数 200 での学習時とテスト時、それぞれの損失と正確率の推移を図 2, 図 3 に示す。学習時とテスト時のそれぞれの損失の値は、学習ごとに減少し続け、テスト時では収束していることから、過学習が抑えられたと考える。テスト時の正確率は約 93 % となり、関心度推定器として妥当であると考えられる。

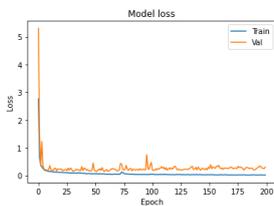


図 2 損失とエポック数

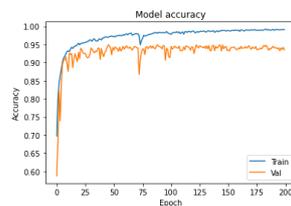


図 3 正確率とエポック数

3.2 関心度推定の汎化性能検証

学習していない 2 名の顔画像を用いて、学習したニューラルネットワークの汎化性能検証を行った。

検証結果を表 1 に示す。「関心あり」の画像には高い精度で推定を行っているのに対して、「関心なし」には 2 名の画像とも誤った推定を行っている。このことから、不特定人物の関心度推定は可能であるが、汎化性能が優れているとは言い切れないと考える。

表 1 汎化性能検証結果

人物	顔画像	関心なし推定精度	関心あり推定精度	出力値
男性	関心なし 1	0.14%	99.86%	1
	関心なし 2	100.00%	0.00%	0
	関心あり 1	13.42%	86.58%	1
	関心あり 2	9.32%	90.68%	1
女性	関心なし 1	13.29%	86.71%	1
	関心なし 2	99.11%	0.89%	0
	関心あり 1	15.64%	84.36%	1
	関心あり 2	47.47%	52.53%	1

4 考察

関心度推定手法の実現には、更に多くの表情を学習したニューラルネットワークの作成が必要であると考えられる。本研究では、6 名の顔画像を学習に用いたが、「関心あり」に対して「関心なし」のデータ数が少なく、クラスごとのデータ数が偏っている。推定対象が不特定多数の場合、クラスごとのデータ数が均等であり、多くの人物を顔画像を含ん

だデータセットを用いた学習を行う必要がある。

CNN の推定の判断根拠となった特徴量を分析する必要があると考える。表情特徴を手動定義せずに関心度推定を行うことができたが、表情特徴のどの部分に着目して判断したか理解できていない。表情以外に着目していた場合、ニューラルネットワークの構造やハイパーパラメータを変更しても精度向上が期待できない。CNN の判断根拠となった特徴量の可視化手法 (Grad-CAM[5] など) を用いて、設計したニューラルネットワークの妥当性を検証し、その後調整を行う必要があると考える。

5 おわりに

本研究では、オンラインミーティングを対象とした関心度推定方式を、表情認識を用いて研究を行った。作成したデータセットで学習を行った結果、参加者の関心度を高い精度で推定することができた。CNN を用いることで、表情特徴を手動定義せず不特定多数の人物を対象とした関心度推定の実現が可能であると示された。

学習データを改良した学習済みニューラルネットワークの作成が今後の課題であり、関心度推定の判断根拠となる顔表情の特徴を分析し、ニューラルネットワークの妥当性を検証することで、関心度推定の精度をさらに向上できると考える。

参考文献

- [1] 宮原正典, 青木政樹, 滝口哲也, 有木康雄: “顔表情からの関心度推定に基づく映像コンテンツへのタギング”, 情報処理学会論文誌, Vol. 49, No. 10, pp. 3694-3702, 2008.
- [2] トーンカム チャイヤブルック, 内田理, “Kinect を利用した感情推定手法”, 映像情報メディア学会技術報告, Vol. 38, No. 33, 2014.
- [3] 中山秀樹: “深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習”, 信学技報, Vol. 115, No. 146, pp. 55-59, 2015.
- [4] 岡谷貴之: “画像認識のための深層学習の研究動向 畳み込みニューラルネットワークとその利用法の発展”, 人工知能, Vol. 31, No. 2, pp. 169-179, 2016.
- [5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra: “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618-626, 2017.