



深層カプセルネットワークを用いた 高精度アイトラッキングとユーザ情動推定への応用

筑波大学システム情報系

助教 河合 新

1. 研究課題の概説

本研究課題の目的は、1) スマートフォンのインカメラをターゲットデバイスとした深層カプセルネットワークを利用した高精度アイトラッキング技術の開発と 2) そのデータを用いた情動推定手法の開発によって、ゲームとユーザ・社会の関わりを評価するために必要な基礎的な技術を実現すること、の2点である。スマートフォン上のゲームとユーザ・社会との関係を明らかにするためには、ゲームがユーザに与える影響の解析が必要である。本研究課題では、そのような解析手法としてアイトラッキングによる情動推定を提案する。アイトラッキングとは人の視線挙動を計測する技術である。この計測に従来は大掛かりな計測装置を必要としており、そのため、多くのユーザの情報を収集し解析することは困難であった。この計測をスマートフォンのインカメラで実現できれば、あらゆるスマートフォンユーザの情動推定が可能となる。

スマートフォンのインカメラをアイトラッキングのターゲットデバイスとした理由は、スマートフォンを用いた生体情報の計測が容易で身近なものとなってきているためである。近年スマートフォンの保有状況は5年間で約4倍になるなど、その普及が急速に進み、1人が1台を持つ情報端末としてその地位を確固たるものとしてきている。それに付随して、スマートフォンはもっとも身近であり、かつ、その使用に抵抗のないデバイスとなってきている。また、従来のアイトラッキングのデバイスは、いずれも高価かつユーザをなんらかの形で拘束するものであった。一方、それらに比べてスマートフォンは安価かつ拘束がないという特徴があり、生体情報計測デバイスとしても最も理想的なものであると言える。このような状況を鑑み、本研究ではスマートフォンのインカメラをアイトラッキングのためのデバイスとして使用した。

また、対象をスマートフォン上で駆動するゲームとした理由は、これらのゲームの普及にある。スマートフォンの普及に伴って、スマートフォン上で駆動するゲームの利用者も増えている。例えば、そのようなゲームで遊んだ経験者は68.9%という統計報告がなされている。したがって、これらのゲームの利用動向に

も変化が生じていることは想像が難くない。特に、スマートフォンの普及はインターネットへの接続を容易とし、オンラインでのゲームによる人や社会とのつながりを密にしている。そのため、これらのゲームが果たすべき役割を明らかにすることは、ゲームとユーザ・社会との関わりを考えるために重要な項目であると考え、本研究ではこれを対象と想定した。

本研究課題の目的を達成するため、本研究では、i) 深層カプセルネットワークの学習時間短縮手法の提案、ii) AI的アプローチによるアイトラッキング技術精度改善のための可視化手法の提案、の2項目を実施した。スマートフォンのインカメラのアイトラッキング技術の開発は、MIT が先駆けとなっているが、十分な精度が得られていない。その原因として、空間情報を失ってしまう畳み込みNNの使用が考えられる。そこで、本研究では深層カプセルネットワークを用いることで、従来よりも高い精度を実現する。また、カプセルネットワークをアイトラッキングというタスクに用いること自体がこれまでには、新しい独自の枠組みでもある。

2. 深層カプセルネットワークの学習時間短縮手法の提案

畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) は画像認識の分野で一般的な手法として広く認識され使用されているが、幾つかの欠点も指摘されている。CNNは、特徴抽出を行う畳み込み層とシフト不変性のためのプーリング層を繰り返す構成によって線の傾き、太さや色などの低レベルの特徴を認識し、これらの特徴から顔画像などにおける目や口などの高レベルの特徴を認識する。一方、CNNは同じオブジェクトであっても異なる姿勢のときに正しく認識するためには、大規模なデータ拡張と非常に深いネットワークが必要になることが知られている。これは、CNNのプーリング層が抽出した特徴間の位相構造を保存しないため、同じオブジェクトでも姿勢が異なると全く異なる内部表現になるために生じている。そのため、様々な姿勢に対応する認識を実現するためにはCNNでは学習に非常に時間を要する。

CNNのこのような問題の解決として、Google Brainのチームによって提案されているカプセルネットワーク (Capsule

Network: CapsNet) に本研究では着目する。CNN の中間層である特徴マップのピクセルは、対応する場所に特徴が存在しているか否かを出力する。一方、CapsNet の中間層であるカプセルは、オブジェクトの部分の情報をベクトルとして出力する。

これまでのCapsNet は学習に非常に時間を要しているが、これを解決し CNN が大規模なデータ拡張と非常に深いネットワークを必要とする問題を解決できよう。従来のCapsNet が学習に時間を要する原因は、CapsNet が分類クラスごとに特徴空間を生成するため、クラスが増えるほどパラメータが増大することである。しかしながら、CapsNet はカプセルに姿勢を含むオブジェクトの特徴を保持するため、同じオブジェクトであれば異なる姿勢であっても正しく認識できると考えられる。そのため、異なる姿勢のオブジェクトを正しく認識するために CNN では必要であった大規模なデータ拡張と深いネットワークが不要である。

そこで、本研究ではCapsNet のパラメータを削減することで、学習時間の短縮を図る。そのための手段として、CapsNet の一部にL1 正則化を適用し、その後、閾値以下の値を0 とすることによるスパース化をCapsNet の学習手法として提案する。また、副次的な効果として、パラメータの削減によって単純で過学習に強いモデルの獲得が期待できる。

従来手法に比べ提案手法ではパラメータの削減とテストデータの微小変形の有無に関わらず精度の向上が確認できた。これは提案手法によって過学習が抑えられ、CapsNet の汎化性能が向上したためだと考えられる。

3. AI 的アプローチによるアイトラッキング技術精度改善のための可視化手法の提案

視線は人間の情動に関する重要な情報が含まれており個性を容易に反映していることから、その推定が望まれ多くの研究がなされている。視線は表情から読み取れない人間の内部情報が含まれている。そのため、視線を正確に測定できれば、その情報は情報推薦や医学など幅広い分野で活用できる。特に、スマートデバイスYate(Smartpho25)を使用して推定できる場合、視線情報は多くの分野でイノベーションを促進することが期待できる。

本研究では、概観ベースの視線推定手法の1種であるCNN を使用する手法に焦点を当てる。視線推定には、モデルベース（瞳孔の向きを使用して視線を推定）と外観ベース（眼の画像から視線を推定）の2つの方法がある。概観ベースの1種であるCNN を用いた手法は、顔の画像と注視位置で構成される大規模なデータセットを学習して視線を推定する。この手法では、視線を推定するのに人が判断して使用するのが難しい微細な特徴を捉えられる。また、多数のパターンを学習することでロバストな推定が可能である。

ただし、CNN を使用したディープラーニングの推論プロセスはブラックボックスであるため、推定が正しい場合でも、その判断の背後にある推論を決定することはできない。そして、モデルの推論の根拠が明確でないため、ユーザはモデルから適切な特徴を捉えた上で目の特徴を推定しているかどうかを判断できない。したがって、ディープラーニングでは、モデルの可視化はモデルのパフォーマンスを評価する上で不可欠な要素である。

Grad-CAM は、CNN がどこを調べて何を推定したかを可視化する方法である。複数の出力をもつ多クラス分類や確率的出力などの問題で使用できる。Grad-CAM は、特徴マップに関する出力の勾配の符号が正である位置を可視化する。ただし、勾配の符号が負であっても可視化が必要な場合があるため、視線推定の回帰問題には適用できない。

本研究では、CNN を用いた視線推定モデルを可視化する回帰問題に Grad-CAM を利用できる新たな可視化手法を提案する。提案された方法は、勾配の符号を考慮して、推定に使用される特徴を可視化できる。提案する可視化手法は、2次元の値を出力する回帰問題で使用できる。

提案された Grad-CAM の拡張によって、視線推定モデルの可視化を実行できた。その結果、データのクレンジング、画像のカッティングテクニックの改善、および入力画像の削減が、モデルの精度向上と軽量化のために必要であることを明らかにした。今後、入力画像の推定結果への寄与を調べ、手法の有効性を高める。

4. おわりに

本研究課題は、1) スマートフォンのインカメラによる深層カプセルネットワークを利用した高精度アイトラッキング技術の開発と2) そのデータを用いた情動推定手法の開発によって、ゲームとユーザ・社会の関わりを評価するために必要な基礎的な技術を実現すること、の2点を目的として、i) 深層カプセルネットワークの学習時間短縮手法の提案、ii) AI 的アプローチによるアイトラッキング技術精度改善のための可視化手法の提案の2項目を実施した。これらの成果により、視線推定を大掛かりな従来の計測装置ではなく、スマートフォンのインカメラで実現しあらゆるスマートフォンユーザの情動推定を可能とする基盤を開発した。

共同研究者

筑波大学システム情報系 准教授 延原 肇