

顔認証システムの人種バイアスに影響を与える潜在的要因の調査

Research on Potential Factors Affecting Racial Bias in Face Recognition

佐藤 佑哉* 土屋 純† 成田 惇† 西垣 正勝† 大木 哲史†
Yuya Sato Jun Tsuchiya Jun Narita Masakatsu Nishigaki Tetsushi Ohki

キーワード 顔認証, 公平性, 生体認証, 人種バイアス

あらまし

顔認証システムは非接触で対象の動作を必要とせず、カメラ等の安価な装置で実現できるなどの利点があり、使い勝手の良い生体認証であることから市場が広がっている。しかし、近年では人工知能の公平性について人種や性別などの情報に関して差別的な振る舞いを行うことが問題視されており、顔認証システムにおいても学習に用いる顔画像データセットの人種割合が偏っていることを原因として、人種間の認証精度に偏りが認められている。さらに、顔認証システムの誤認識を原因とした有色人種の誤認逮捕や、米国の白人男性警察官による黒人男性殺害事件をきっかけとした人種差別抗議運動により、IBM, Amazon, Microsoft といった大手テック企業が顔認証技術の出荷停止の措置を行う事態へと発展し、顔認証システムの人種差別に関する問題は大きく取り上げられている。顔認証システムの公平性に関して十分な検討を行うことは、このような社会問題の解決に取り組む上で重要であると考えられる。

顔認証モデルを学習させる際に一般的に使用される CASIA-WebFace, VGGFace2, MSCeleb-1M などの大規模な顔画像データセットは、Web サイトのスクレイピングによって無造作に収集されて構築されているが、アジア人とインド人が世界人口の 44% を占めているにも関わらずデータセットには 8% ほどしか含まれていない。このように、データセットの人種割合が実際の人口割合とは大きく異なっており、先行研究によって、これらの

人種割合が偏っているデータセットを用いて学習を行った顔認証モデルは白人と比較して黒人の認証精度が低くなり、BUPT-Balancedface などの人種割合を統一した顔画像データセットで顔認証モデルの学習を行うことで人種間の認証精度の偏りが軽減されることから、学習に用いる顔画像データセットの人種割合の偏りが顔認証システムの公平性に影響を与えることが示されている。

一方で、データセットの人種割合の偏り以外の顔画像の明度やコントラストなどが持つ要素が人種間の認証精度の偏りに与える影響について十分な検討がされていない。データセットの人種割合の偏り以外に人種バイアスに影響を与える要因が存在する場合、人種に対する公平性を評価するためのデータセット等で評価を行い、公平性が認められるような顔認証システムにおいても、利用する状況に応じて人種バイアスが顕在化する可能性が考えられる。

そこで本研究では、表現学習に基づく顔認証システムを対象として、潜在的な要因の調査を行う。解像度、明度、彩度等など、環境要因の変動を対象とし、これらの変動を加えたテストセットを用いて、人種間の認証精度の差を検証することで、実利用時の環境要因の変動が人種間の認証精度の偏りに与える影響の調査を行なう。

さらに、人種間のスコア正規化アプローチや、モデル生成時のデータ均等化など、既存の公平性向上アプローチによる影響を同時に評価する。

* 静岡大学情報学部
静岡県浜松市中区城北 3 丁目 5-1, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka, Japan

† 静岡大学大学院総合科学技術研究科
静岡県浜松市中区城北 3 丁目 5-1, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka, Japan