用語解説

視線 IEC（Interactive Evolutionary Computation with Gaze）

IEC（Interactive Evolutionary Computation: 対話型進化計算）は、進化計算における解候補の表現型をユーザが評価できる形式（音や画像など）で構成し、評価関数の代わりにユーザの感性評価を用いて解候補を評価し、進化させる手法である。IEC では、解候補はユーザの満足のいくデザインに向かって進化する。IEC の起源は 1986 年に R. Dawkins が提案したバイオモルフであるとされている。その後、IEC という用語は、進化計算の分野で 1990 年代に広まり、ユーザが好む音や衣服デザインなどを生成する IEC システムが数多く提案されている。

視線 IEC は、ユーザの視線情報によって解候補を評価する手法である。「目は口ほどにものを言う」ということわざがあるように、ユーザの視線情報には潜在的な嗜好が隠されていることは様々な分野で指摘されている。このため、ユーザの視線情報を用いた嗜好性調査や行動分析が、マーケティングや高齢者向けシステムの開発、ソーシャルコンテンツなど多くの分野で行われている。視線 IEC では、ユーザの視線情報を IEC の解候補評価に用いて、ユーザの好むものを動的に生成する。視線 IEC に関するシステムは、最近 15 年の間に国内外で提案され、ユーザが満足のいく解候補を生成できることが確認されている。さらに、ユーザは提示された解候補群を自由に眺めるだけでよいという長所から、ユーザの評価負担を軽減できることも確認されている。

以前より、IEC システムでは、ユーザの解候補評価負担が大きいことが問題とされているが、視線 IEC はこの解決にも有効である。通常の IEC では、ユーザは解候補を 5 段階や 10 段階方式などで評価するため、ユーザに明示的に解候補を評価させる作業が発生する。すなわち、ユーザは、何らかの形で解候補に対して抱いた印象を、数値などに変換して表現しなければならない。しかし、ユーザの視線情報を用いれば、ユーザ自身が印象を数値に変換する必要がなくなり、評価負担の軽減につながる。

視線 IEC におけるユーザの視線計測の際には、ユーザが特別な計測装置を着用する必要もなく、ストレスフリーでユーザが自由に解候補評価に参加できる。今後は、複数ユーザの視線情報を用いた合意形成や嗜好性調査などにも応用されることが期待されている。

（福岡工業大学 竹之内 宏）

バイリニアプーリング（Bilinear Pooling）

プーリング（Pooling）は与えられた特徴量ベクトル集合を集約・要約する演算である。畳み込みニューラルネットワーク（CNN）では、畳み込み演算により得られる特徴マップに対して部分的に最大値（または平均値）として要約することで認識対象の位置不変性を得ることができ、Max Pooling（平均の場合、Average Pooling）として知られている。また、ネットワーク全体を通して得られる最終特徴マップに対して、各次元について平均成分を集約する手法が Global Average Pooling（GAP）として提案され、入力画像サイズを制限する必要が無いことや、その汎化能力の高さから、物体認識タスクにおける多くのモデルで採用されている。

画像認識分野において CNN の登場以前より、認識対象の位置不変性や入力画像サイズフリーといった観点からプーリング演算は注目されており、カーネル法等の機械学習手法や統計的手法との関連が深く、より高次元統計量による集約手法が主な研究対象であった。Average Pooling や GAP は一次プーリングのため、高次元統計量の導入によるモデルの表現能力向上が報告されている。

バイリニアプーリング（Bilinear Pooling）は高次元統計量を表現する手法として提案されており、CNN によって得られた最終特徴マップについて共分散行列等の半正定値行列や共起行列として集約し、画像特徴量を表現する。CNN では、得られた画像特徴量に対して全結合層を用いた線形分離により認識・予測するため、特徴量の線形分離性が求められる。半正定値行列として表現された特徴量はリーマン空間上に分布するため、平坦空間への変換方法として、計算効率の良い行列対数による分離性の高い接線空間への写像や、GPU 上で効率的に演算可能である行列平方根の適応により、実用的であり精度面で優れる認識が可能である。

深層学習は学習を通して入力データの特徴量変換を獲得できるため、自然言語・音声・時系列データにおいて特に活用されており、画像からのキャプション生成のような複数の異分野を横断する課題についても発展している。特に、画像と質問文を入力とし回答文を生成する Visual Question Answering 等、入力がマルチモーダルな課題において画像と文章の複数特徴量を組み合わせる際に、バイリニアプーリングによる共起表現は効果的である。実問題の認識・予測ではマルチモーダルな情報を得ることが可能な状況は多く、効果的に複数特徴量を融合するバイリニアプーリングやその派生手法は、様々な実問題解決に向けた有効な一手法として期待できる。

（高知工科大学 四宮 友貴）