

論文の内容の要旨

論文題目 深層生成モデルによる環境の構成的な認識と生成

氏名 小林 由弥

深層学習の登場により、人工知能で実現可能な情報処理の種類や精度は飛躍的に向上した。しかし現状ではこれは適切に設計された狭い範囲の課題に限られており、人間のように広範な課題に対して柔軟に対応する人工知能の実現は未だ遠いものとなっている。より柔軟で汎用的な情報処理を行うために考慮すべき要素の一つとして、構成性があると考えている。構成性とは、ある系や意味表現がそれ自体よりも単純な構成要素の組み合わせによって規定されるような性質のことで、言語や我々の住む実世界に見られる特徴である。通常、人間は全体を一度に認識するのではなく、個々の要素に分解して認識を行っている。ここでの分解の単位となる要素は主に物体であるが、抽象的な概念である場合もある。現実世界はこうした単純な構成要素の組み合わせによって無数の異なる状態が実現され得るが、組み合わせを考慮しない場合には無数のパターンを個々に認識・理解する必要がある、汎化や転移能力の面で非効率的であると考えられる。また、物体は系の発展の単位となることが多く、個々の物体を認識することは因果推論 (causal inference) や物体間の関係性の推論 (relational inference) などに関して有利になることが期待される。

このような背景から本研究では、系を複数の物体の組み合わせとして捉えるような構成的な認識に関する研究を行うことを考え、特に画像を対象として教師なしで物体認識と各物体の表現学習を行う深層生成モデルを扱う。このような手法は物体中心表現学習 (object-centric representation learning) と呼ばれるが、現状では深層生成モデルを基にした物体中心表現学習の手法は学習の不安定性や処理の精度、適用範囲が簡単なデータに限定されていることなどの課題を抱えている。本研究でこうした問題が根本的には教師なし学習によって物体に関する前提知識 (帰納バイアス) が十分に得られていないことによると考え、適切な帰納バイアスを導入する方法について考えるものである。

本研究は上記のテーマに関する3つの研究によって構成されている。研究1は背景に関する補助情報を導入することで、物体を認識するために必要な帰納バイアスを与えるものである。また、研究1では背景の情報を与えるために新たなデータを用いているが、研究2では新たなデータなしに、学習やネットワーク構造の工夫によって帰納バイアスを導入する方法を提案する。研究1と2では静止画一枚のみを与え、物体を認識するという二次元空間での課題となっていたが、研究3では多視点の入力に拡張した問題設定に取り組んでいる。これは、複数の物体を含む、ある3次元的な空間に関して複数視点からの画像を与え、観測していない任意の視点からの画像の予測と物体認識を行うという課題である。また、研究1と研究2ではデータの与え方や学習の工夫による帰納バイアスの導入を試みたが、研究3ではモデル構造に前提知識を組み込む形の導入を行う。また、8章ではこれらの研究をまとめ、本研究の発展や今後の研究の方向について考察を行う。