

07

ネットワーク構造から深層学習のしくみを知る

～ニューラルネットの理解に向けたコミュニティ抽出技術～



どんな研究

深層ニューラルネット (DNN) は様々な実データに対し高精度な予測を実現していますが、その複雑な学習結果を人間が理解することは困難です。本研究では、**ネットワーク解析によりDNNの中に潜むコミュニティ構造を自動的に発見し、可視化する方法を提案します。**

どこが凄い

データから学習されたDNNに対してコミュニティ抽出と呼ばれる手法を適用することで、**DNNの大局的な構造を捉える**ことを世界で初めて可能としました。これは、DNN内部の各部分が予測や認識において果たす役割を解明するための、新たな糸口となります。

めざす未来

DNNの内部の仕組みを人間が解釈しやすい形で表現することで、自動運転や医療など、**機械学習技術が導き出した予測結果に対して根拠を説明**できることが必要となる応用分野においても、安心してDNNが使えるような未来をめざしています。

① 入出力データの組からニューラルネットを学習

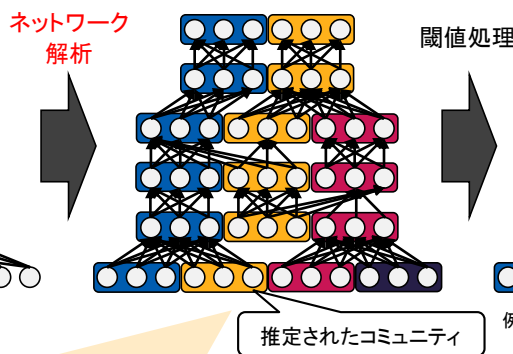
画像に写った野菜を識別する問題の場合

例) トマト、にんじん、なす、...



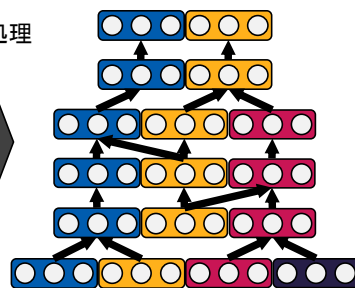
② 隣接する層のユニットに対し、似た結合パターンを持つユニットのコミュニティを推定

コミュニティごとにユニットを「並び替え」



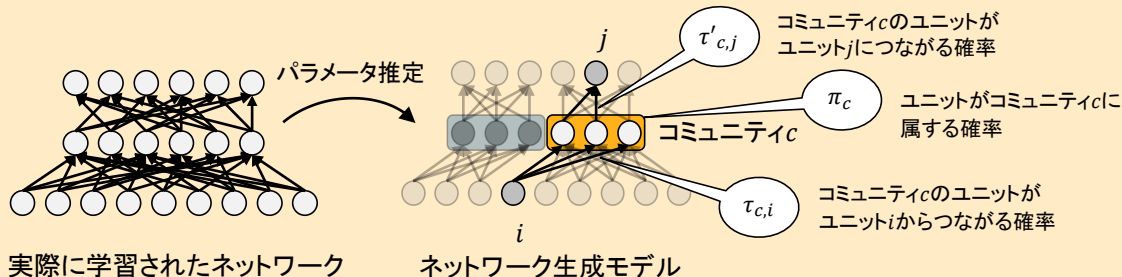
③ 各コミュニティ間に存在する複数の結合をまとめて表示

例) 類似した推論情報を受け取る野菜の組み合わせを表すコミュニティ



例) 類似した推論情報を受け渡す入力画素の組み合わせを表すコミュニティ

- ネットワーク構造の生成モデルに基づくコミュニティ抽出法を、学習済ニューラルネットに適用
- 従来技術では捉えられなかった、**ニューラルネットの全体構造をひとめで理解できる!**



関連文献

- [1] C. Watanabe, K. Hiratsatsu, K. Kashino, "Modular representation of layered neural networks," *Neural Networks*, Vol. 97, pp. 62-73, 2018.
- [2] 渡邊千紘, 平松薫, 柏野邦夫, "深層ニューラルネットにおける正負の結合重みに基づく大局構造抽出," *情報科学技術フォーラム*, 2017.
- [3] C. Watanabe, K. Hiratsatsu, K. Kashino, "Recursive extraction of modular structure from layered neural networks using variational Bayes method," in *Proc. Discovery Science*, Lecture Notes in Computer Science, Vol. 10558, pp. 207-222, 2017.

担当者

渡邊 千紘 (Chihiro Watanabe) メディア情報研究部 メディア認識研究グループ