

Algorithmic Concept-based Explainable Reasoning [?]

Georgiev, DobrikBarbiero, PietroKazhdan, DmitryVeličković, PetarLiò, Pietro

15 July 2021

概要

グラフニューラルネットワーク（GNN）モデルに関する最近の研究では、古典的なグラフアルゴリズムや組合せ最適化問題への GNN の適用に成功した。これには、前提条件が満たされない場合のアルゴリズムの適用や、十分な学習データが得られない、あるいは生成できない場合の学習済みモデルの再利用など、数多くの利点があります。しかし、GNN はブラックボックスモデルであり、直接解釈することができないため、これらのアプローチの主な障害は、説明可能性に欠けることである。本研究では、概念に基づく説明に関する既存の研究を GNN モデルに適用することで、この制限を解決する。GNN の読み出し機構を修正した概念ボトルネック GNN を導入する。3つのケーススタディを用いて、我々は以下のことを実証する。(i) 提案モデルが、各ターゲットクラスに対して、正確に概念を学習し、学習した概念に基づく命題式を抽出できること、(ii) 提案概念ベース GNN モデルが、最先端モデルとの比較性能を達成できること、(iii) グラフレベルの概念に対して明示的にスーパービジョンを与えずに、グローバルグラフ概念を導出できること、である。

concept: 概念 [?]

1 どういう論文？

- グラフアルゴリズムにおける概念に基づく推論機構を持つ Concept Bottleneck Graph Neural Networks を提案
- これを通して、性能に影響を与えることなく、ノードレベルの概念を正確に学習できることを示した
- 学習データとモデルの重みを調べることで、定義された概念に基づき、各ノードレベルの出力クラスを数式で説明することが可能である
- 概念によって、特定のグラフレベルのタスク（終了時期の決定など）の教師なしルール抽出を行うことができる
- 抽出されたルールは解釈可能であり、ルールを適用しても精度に大きな影響を与えない

2 先行研究と比べてどこがすごい？

3 技術や方法のポイントはどこ？

4 どうやって有効と検証した？

5 議論はある？

6 次に読むべき論文は？