グラフ構造データを対象とした
解釈可能決定集合の拡張

松山　航太

要旨

深層学習を中心とした機械学習技術の発展に伴い，近年，非常に高精度な機械学習モデルが構築されている．しかし，モデルの多くは，利用者がその内容を理解・解釈することが困難な大規模かつ複雑なブラックボックスであり，モデルのデバッグ・更新が容易ではなく，また安全面や倫理面からもその応用範囲が限定されてしまうことも考えられる．これらの問題を軽減するため，モデル理解の支援や解釈可能（容易）なモデルの構築など，機械学習の解釈性が着目され，多数の研究が行われている．本研究では，分類規則抽出に関する代表的な解釈性研究の一つである解釈可能決定集合（Interpretable Decision Sets; IDS）に着目する．IDS は，一般的な表形式データを入力とする手法であり，データから導出されるクラス相関ルール集合から，解釈性を考慮した評価関数を最大化する部分集合を抽出することで，精度を維持したまま解釈性の高い少数ルール群を特定する．本研究では，IDS のアイディアを拡張し，グラフ構造データを対象とした解釈可能決定集合を導出するアルゴリズム GIDS を提案する．具体的には，頻出部分グラフを用いてグラフデータベースを表形式へと変換し，クラス相関ルールを導出する．加えて，クラス相関ルールの構成要素がグラフであることに着目し，グラフ形状を考慮した解釈性に関する新たな評価尺度をルール集合選択基準に組み込む．これらにより，グラフ構造データからの解釈容易なルール集合の導出を目指す．GIDS を定量的に評価するため，間取り図に関する実データを対象とした実験を行った．実験では，様々なルール集合選択基準を用い，得られるルール集合を解釈性の観点から定量的に評価した．その結果，グラフを考慮しない評価関数を用いた場合と比較し，GIDS がより解釈性の高いルールを導出できることが確認された．