

部分グラフ同型問題と模倣関係の融合

菅原 知倫

グラフデータは情報のつながりを表現することに適したデータ構造であり、人間関係や交通ネットワークの管理など幅広く利用されている。グラフデータの効率的なパターンマッチは、そうした実際の利用に応用できる重要な課題の一つである。

グラフデータにおける主なパターンマッチの定義として、部分グラフ同型問題 (subgraph isomorphism) と模倣関係 (graph simulation) の二つがある。部分グラフ同型問題は、発見したいパターンを定義したグラフ (以下、パターングラフ) と探索対象のグラフ (以下、データグラフ) が与えられたときに、データグラフに含まれるパターングラフと同型の部分グラフを全て見つけるというもので、SNSの分析や化学式の類似構造の発見などに利用できる。

部分グラフ同型問題に関する先行研究として、Ullmann によって提案された初の実用的なアルゴリズムがある。また、近年では Ullmann のアルゴリズムを拡張する形で VF2, QuickSI, GraphQL, GADDI, SPath といった多くのアルゴリズムが提案された。しかし、部分グラフ同型問題は NP 完全であり、上述のアルゴリズムを用いたとしても、この問題を解くには最悪の場合指数時間を要する。

一方、模倣関係は、ウェブサイトの分類やソーシャルネットワーク上の役割の発見などに利用される。この概念によるパターンマッチは Henzinger らの研究により多項式時間で計算できることが示されている。しかし、模倣関係はパターングラフとデータグラフ間の対応を関係として求めるため、解の構造がパターングラフと一致しないという問題がある。

そこで本論文では、まず部分グラフ同型問題と模倣関係を融合し、両者の利点を活かした新たな概念を定義する。次に、定義した新概念に基づくパターンマッチングアルゴリズムを提案する。具体的には、パターングラフのノードに対して、その中から部分グラフ同型問題の定義を適用するノード (以下、キーノード) を任意に設定する。そして、キーノードに対しては部分グラフ同型問題の定義、それ以外のノードに対しては模倣関係の定義を適用したマッチングを行う。これにより本論文が提案するアルゴリズムは、部分グラフ同型問題と比べて計算コストが低く、模倣関係と比べて解の構造がパターングラフと一致するマッチングを可能とする。ただし、よりパターングラフの構造と一致するノードとマッチさせるために、本研究では一般的な模倣関係の定義を拡張し、制約を強めた模倣関係を扱う。

提案アルゴリズムを実装し、ラベル付き有向グラフを対象とした評価実験を行った。その結果、提案アルゴリズムが所望の動作効率でグラフデータにおけるパターンマッチを行えることを確認した。

(指導教員 鈴木伸崇)