

# Shape Expression Schema の下での パターン問合せに対する Reverse Engineering

坂本 康介

近年、グラフデータが社会に浸透し、様々な場面で利用されるようになってきているが、グラフデータは大規模で複雑なため、一般の人には問合せを行うのが難しいという問題がある。問合せをより容易に行う手法として **Reverse Engineering** がある。**Reverse Engineering** とは、利用者が求めている解や省きたい解の例を複数提示することで、システムがそれらを満たすような問合せ式を出力する手法である。関連研究として、与えられた各グラフの対応を網羅的に調べることで問合せ式を作成するアルゴリズムの提案がある。しかし、一般に **Reverse Engineering** における解の探索空間は指数関数的に増加するため、より効率的に問合せ式を求める手法が必要になると考えられる。

本研究では、グラフデータに対する問合せの中でもパターン問合せに着目する。パターン問合せとは、グラフの断片であるパターングラフを指定し、指定したパターングラフに対応する部分グラフをデータグラフの中から検索する問合せである。この場合、問合せ式がグラフデータの形式になるため、提示された解の最大共通部分グラフを求めることで問合せ式を生成することができる。最大共通部分グラフを求めるには、各グラフの隣接行列を作成し比較することが必要になる。しかし、各グラフのノードの対応が事前には分からないため、ノードの対応をすべての場合で試す必要があり、その処理は非効率的である。そこで、効率化を図るために **Shape Expression Schema**(以下 **ShEx**)に着目する。

本研究では、**ShEx** の下でパターン問合せの **Reverse Engineering** を効率的に行う手法を提案する。**ShEx** は、グラフデータのスキーマ言語の 1 つであり、**Regular Bag Expression** という規則に基づいて型を定義し、各ノードが必ず 1 つ以上の型を持つという特徴がある。**ShEx** スキーマに基づくグラフの最大共通部分グラフを求める際には、各グラフのノードは必ず同じ型を持つもの同士が対応するため、作成する隣接行列の数を減らし、処理の効率化を図ることができる。

事前に **ShEx** スキーマの各型のエッジの入力と出力をまとめたテーブルを作成し、それを参照することで与えられたグラフの各ノードの持ちうる型を絞り込み、その情報に基づいて隣接行列を作成し比較することで最大共通部分グラフを求める。得られた最大共通部分グラフから問合せ式を構成・出力する。

提案アルゴリズムを **Ruby** で実装し、**ShEx** スキーマとそれを満たすグラフを用意し評価実験を行った。その結果、ノード数が一定以上の場合に、効率化を行わない手法よりも処理時間が短くなり、より効率よく **Reverse Engineering** を行うことができることを確認した。

(指導教員 鈴木伸崇)