

# 令和元年度 情報工学コース卒業研究報告要旨

関 研究室	氏 名	小 島 和 之
卒業研究題目	木文法に基づくグラフに対する木分解結果の圧縮法および圧縮データへの直接問合せ処理法の提案と実装	
<p>グラフ構造は実世界の大規模データを表現するために用いられている。グラフに対する問題で、効率の良いアルゴリズムが知られていないものが多数ある。こうした問題の計算量を削減するために、グラフを木分解と呼ばれる手法で木構造に分解し、得られた木(分解木とよぶ)に対して動的計画法等の技法を用いて元のグラフに対する問題を解くというアプローチが知られている。ここで分解木の頂点は元のグラフの部分グラフであり、その部分グラフの頂点集合をバッグとよぶ。</p> <p>木分解の計算にはグラフ全体の構造情報を必要とし計算コストが高い。そのため木分解結果を複数回利用し計算を行うために結果を保存することが求められる。しかしながら分解木に含まれる部分グラフには重複が多いためデータサイズが元グラフに比べ大きく、データの圧縮を行う必要がある。グラフに対して単純な圧縮アルゴリズムを適用した場合には構造情報が損なわれてしまう。そのためグラフに対する問合せや更新時には一旦解凍してから処理を行わなければならない非効率的である。従来研究では分解木についてバッグに含まれる頂点の出力を子バッグの情報を用いて符号化することで使用メモリ量を削減したものがあげられる。この方法は親と子バッグの共通部分が多い場合においては効率よく削減可能である。しかし、分解木自体を圧縮し、圧縮データに対する問合せ、更新を行う方法は報告されていない。</p> <p>本研究ではグラフに対する分解木を圧縮可能な形式に変換し出力する方法と、圧縮された分解木に対して解凍せず直接的に問合せアルゴリズムを適用する方法を提案し実データを用いた評価を行った。</p> <p>分解木の出力時には各頂点の親頂点を複製し、更にバッグ内の辺を子頂点として追加する。圧縮方法として TreeRePair を使用した。TreeRePair は入力として与えられた木において隣接する頂点のラベル対のうち最頻出するものを非終端記号に置換することで直線的文脈自由木文法 (Straight Line Context-Free Tree Grammar: SLCFG) に変換する。TreeRePair により得られる SLCFG は圧縮率や圧縮にかかる時間の面で優れている。更に SLCFG に対して解凍することなく直接問合せや更新を行う方法について研究が行われている。</p> <p>問合せ例としてグラフにおけるクリーク探索問題を使用した。木分解の定義からグラフに含まれるクリークは分解後においても必ず一つ以上のバッグに含まれる。具体的なアルゴリズムはまずクリークを構成するために必要な数の辺をバッグから選択する。頂点数 <math>m</math> のクリークには辺が <math>m(m-1)/2</math> 個含まれる。選択した辺の端点の個数を数え上げることでクリークであるか判定を行う。</p> <p>ベンチマークは頂点数 432, 辺の数 750 であるグラフデータ ht-wiki-talk について以下の結果が得られた。木分解の結果, 頂点数 816, バッグに含まれる辺の総計 902 である分解木が得られた。分解木の圧縮率は 32.3% であり圧縮データに対する直接クリーク探索法に要した時間は, 解凍し探索する方法に比べ 28.7% であり使用された最大メモリ量は 89.5% であった。</p>		