

平成28年度 情報工学コース卒業研究報告要旨

関 研究室	氏 名	武 田 健 志
卒業研究題目	木文法に基づくグラフ圧縮法および圧縮グラフに対する頂点選択問合せ評価法の提案と実装	
<p>グラフはその汎用性の高さから様々な関係を表現することに用いられている。一方で、実用データを表現したグラフは非常に大規模になりやすいため、メモリ量を節減した記憶方式の開発が必要となっている。このような問題を解決するため、グラフ圧縮の研究が行われている。特に、gRePairは後述する木文法を用いた木構造圧縮をグラフに拡張した手法であり、RDF等の辺ラベル付きグラフに対しても高い圧縮性能を示している。</p> <p>これらの圧縮データに対して検索や更新などの問合せ処理を行うことを考えたとき、圧縮データをいったん解凍し、解凍されたデータに対して問合せを行い、更新がなされた場合はさらに再圧縮を行うという方法は効率が悪い。一方、文法を用いた圧縮法では圧縮データ中に構造情報が一部保存されていることから、圧縮データに対して直接問合せ評価が行える場合がある。しかし、圧縮グラフに対する直接問合せ評価を実装し実験的評価を行ったという研究はまだ報告されていない。</p> <p>圧縮データに対する直接問合せ評価法の従来研究として、圧縮されたXML文書に対する研究があげられる。TreeRePairでは、入力として与えられた木について、隣接するノードのラベルの対(木ダイグラムとよぶ)の中で最頻出するものを非終端記号に置き換えていくことで、直線的文脈自由木文法(Straight Line Context-Free Tree Grammar: SLCFTG)に変換する。こうして得られた圧縮データ、すなわちSLCFTGは圧縮率(元のデータと圧縮データのサイズ比)および、圧縮に要する時間効率の双方が非常に優れていることが知られている。また、TreeRePairで圧縮されたXML文書に対して、それらを解凍することなく直接更新する手法についても研究が行われている。</p> <p>本研究では、グラフをSLCFTGに圧縮する手法と、SLCFTGに圧縮されたグラフに対して、それを解凍することなく直接頂点選択を行う問合せ評価法を提案し、実データを用いて提案手法を評価する。圧縮法の基本的な方針としては、与えられたグラフの指定された頂点を根頂点とした深さ優先探索木tを構成し、tに対してTreeRePairアルゴリズムを適用することでグラフからSLCFTG G_tを得る。G_tには深さ優先探索木tに含まれる辺の情報しか含まれていないため、tを構成する時に、tに含まれない辺(逆辺)の情報Rを別に作成する。次に、グラフに対する問合せモデルとして決定性選択グラフオートマトン(Deterministic Selecting Graph Automaton: DSGA)を導入した。これは頂点に状態を割り当てていき、頂点のラベルと状態の組を指定することで頂点選択を行うモデルである。そして、圧縮されたグラフ(G_t, R)とDSGA Aが与えられたとき、G_tに対してAを直接適用することで頂点選択を行う手法を提案した。提案手法を実データに対して適用した結果、圧縮率に関してはgRePairで圧縮されたものと比較するとわずかに劣っていたが、多くのデータに対して、圧縮データへの直接問合せ評価法に要する時間は、解凍してから問合せを行う方法と比べて15%~40%、要したメモリ量については5%未満であった。</p>		