

# GPGPU を用いた頻出部分グラフマイニングの実装と プロファイリング

土岐 達哉

## 要旨

複数のグラフからなるデータベースから特徴的かつ有意性のあるパターングラフを発見したり、グラフの分類やラベル付けなどを行う手法をグラフマイニングという。例えば実世界では化学式、WWW や文法構造など非常に多くのデータ構造がグラフで表現でき、以前から研究が盛んなデータマイニング分野の一つである。指定した頻度以上に出現するパターンを列挙する手法は頻出部分グラフマイニングと呼ばれ、グラフマイニングの中でも代表的な手法である。グラフマイニング分野の発展とともに、頻出部分グラフマイニングも探索に幅優先や深さ優先を用いた手法などが数多く提案されてきた。しかし一般に、頻出部分グラフマイニングは、パターン数の多さや部分グラフ同型問題を解く必要があるなど、計算コストが非常に大きいことが知られている。この問題に対する解決策として、近年、GPGPU による部分グラフマイニングの並列化が提案されている。画像処理ユニットである GPU は、CPU と異なり膨大な並列演算を得意とする。GPGPU は、この並列性を画像処理以外の一般目的に応用する技術一般をさす。頻出部分グラフマイニングの分野において提案された GPGPU による高速化は、GPU メモリの制約に加え、基となる既存アルゴリズムそのものの配列化の困難さから、必ずしも数多くの研究があるわけではないが、近年、深さ優先探索に基づく基本的なアルゴリズム gSpan に対する GPU 実装が提案され、改めて GPGPU による頻出部分グラフマイニングの高速化が注目を集めることとなった。

本研究では、既存研究における疑似コードを中心に、実際に GPGPU を用いて頻出部分グラフマイニングアルゴリズムの実装を行った。加えて、実装したシステムに対してプロファイル技術を用いた各処理における処理時間の分析を行い、処理メソッドやデータ構造の観点から並列化におけるボトルネックの検証を行うとともに、いくつかの解決策を提案した。これにより、プロファイル前と比較し 30~40%程度の高速化を達成することに成功した。