HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | Studies on Enumeration of Acyclic Substructures in Graphs and Hypergraphs [an abstract of dissertation and a summary of dissertation review] |
| Author(s) | 和佐, 州洋 |
| Degree Grantor | 北海道大学 |
| Degree Name | 博士(情報科学) |
| Dissertation Number | 甲第12184号 |
| Issue Date | 2016-03-24 |
| Doc URL | https://hdl.handle.net/2115/61622 |
| Rights(URL) | https://creativecommons.org/licenses/by-nc-sa/2.1/jp/ |
| Type | doctoral thesis |
| File Information | Kunihiro_Wasa_abstract.pdf, 論文内容の要旨 |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

# 学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　和佐 州洋

## 学 位 論 文 題 名

Studies on Enumeration of Acyclic Substructures in Graphs and Hypergraphs
（グラフや超グラフに含まれる非巡回部分構造の列挙に関する研究）

近年, 計算機の性能向上やネットワーク環境の発展に伴い, グラフやその一般化である超グラフで定義された情報を容易に, かつ大量に得られるようになった. しかし, その情報すべてを, 自らの目で確かめるのは困難である. さらに, その中から, 隠れた規則性といった有用な知識を発見するのは, 人力では到底不可能な状況である. そこで, 我々は, 蓄積されたグラフや超グラフに対して, データマイニングや機械学習を用いることで, 先に述べた困難な状況を克服しようとしている. しかし, これらの技術を用いる際に, 入力データに含まれる, 指数的な数の部分構造を取り扱わなければいけない. そこで, このような問題を列挙問題と呼び, 次のように定義する: グラフ, または超グラフで定義される入力データと制約条件が与えられた時に, 入力データ含まれ, かつ条件を満たすすべての部分構造を漏れ無く重複なく求めよ.

1950 年代から, 計算機を用いた部分構造列挙に関する研究が本格化してきた. 1975 年に Read と Tarjan は, 電力ネットワーク網の性能評価やプログラムフローの研究を行うために, 全域木や, パス, サイクルを列挙するアルゴリズムを提案した. また, クリークや部分木の列挙アルゴリズムは, データマイニングや機械学習へ応用されている. このように, 現在までに, 実用面の要請から, 全域木や, パス, サイクル, クリーク, マッチング, カットなどの部分構造列挙問題が研究されてきた. さらに, 列挙自体に対する理論的な興味からも積極的に研究が行われている. 列挙アルゴリズムは, 他のアルゴリズムに比べて, 独特な点を持つ. 入力を離散データとして, そのサイズを $n$ とすると, 入力の部分構造を解とするような列挙問題の場合, $O(2^n)$ 個の解が存在する. したがって, そのような問題に対して, 列挙アルゴリズムの時間計算量は, 少なくとも $O(2^n)$ 時間である. しかし, 一般に, 解の個数は $O(2^n)$ よりも小さく, 入力サイズのみで時間計算量を測るのは妥当ではない. よって, 列挙アルゴリズムの計算量は, 入力と出力の両方に関して評価する. 本研究では, 列挙アルゴリズムの評価を遅延に, つまり, 解を出力したあと, 次の解を出力するまでに係る時間, に基づいて行う. これまでに, 遅延が定数時間であるようなアルゴリズムが与えられている対象は, 非常に少ない. それどころか, 多項式遅延で, つまり, 遅延が多項式で抑えられるかどうか未解明な列挙問題も残されている.

本研究では, 近年, 生命情報科学や, 機械学習, データベースのスキーマ設計で注目されている非巡回性を, 列挙問題の制約として与え, 次の 4 つの問題について考察し, 遅延の観点で, 高速な列挙アルゴリズムを構築することを目標とする. まず, $k$-部分木列挙問題について考察する. Ferreira と, Grossi, Rizzi によって与えられた本問題は, 入力として, グラフ $G$ と整数 $k$ が与えられた時に, $G$ に含まれる $k$ 頂点からなる連結かつ非巡回な部分グラフ, つまり $k$-部分木を列挙せよ, という問題である. 本問題に対して, Ferreira らは, $O(k)$ ならし遅延で列挙するアルゴリズムを与えた. しかし, 定数の遅延で列挙できるかは未解決問題であった. そこで, 本研究では, 入力グラフを木に限定した際に, Avis と Fukuda が提案した逆探索法を基に, 列挙アルゴリズムを構築する. 逆探索法は, 家系木と呼ばれる, 解全体空間全体を覆う樹状の探索木を定義し, 家系木上を深さ優先探索することで列挙

する. 提案の列挙アルゴリズムは, 同じ根を持つ $k$-部分木間の親子関係と, 異なる根を持つ $k$-部分木間の親子関係の, 2 種類の親子関係を定義し, これに基づいて家系木を構築し探索することで, 木に含まれるすべての $k$-部分木を遅延を, 最悪時に定数遅延で列挙するアルゴリズムである. さらに, $k$-部分木列挙の応用として, 木の類似度計算についても考察する. 類似度計算のひとつとして, 与えられた 2 つの木に対して, それらに出現する部分木を数え上げ, 頻度を比較することで類似度を計算する手法があげられる. これまで, 部分木のサイズを制限しない場合や, 部分木をパスに制限するなど非常に厳しい制約を与えた場合, 木の類似度計算を効率よく行うアルゴリズムが与えられてきた. しかし, サイズのみを制限する, つまり, $k$-部分木を基にした類似度の計算アルゴリズムは知られていなかった. この問題に対して, 本研究では, $k$-部分木の列挙にかかる時間と等しい時間計算量で, この類似度を計算するアルゴリズムを与える.

これまで, 部分グラフの列挙問題については広く研究されてきた一方, 誘導グラフの列挙問題については, クリークや, 誘導連結グラフ, 誘導パス, 誘導サイクルの列挙を除いて, ほとんど研究されてこなかった. そこで, 本研究では, グラフに含まれるグラフ中の連結かつ非巡回な誘導グラフ, つまり, 誘導木を列挙する問題について考察する. この問題に対して, 逆探索法や分割法といった枠組みを用いて, 素朴なアルゴリズムを設計した場合, $d$ を入力グラフの最大次数とすると, $O(d)$ 遅延の壁を破るのは非常に難しい. そこで本研究では, グラフの縮退数に着目する. グラフ $G$ の縮退数が $k$ であるとは, $G$ 中のすべての誘導グラフ $S$ に対して, $S$ が高々次数 $k$ の頂点を持つときをいい, 縮退数が定数であるようなグラフとして, 木 ($k = 1$) や, 格子グラフ ($k = 2$), 平面グラフ ($k = 5$) などが知られている. また, グラフ $G$ の縮退数が $k$ である時, 次を満たす頂点の全順序が存在する: $G$ 中の任意の頂点 $v$ に対して, (A) 列中で $v$ より大きく, (B) $v$ に隣接する, の 2 つを満たす頂点 $u$ は, 高々 $k$ 個しかない. この列を, 縮退列と呼ぶ. 本研究で提案する誘導木列挙アルゴリズムは, 空集合から, つまり, 空な誘導木から始まり, 縮退列上で最小の頂点から順に, 誘導木に追加しながら列挙するアルゴリズムである. このように縮退列の順で頂点を追加していくことで, 追加に係る時間計算量を, 将来出力する誘導木に分配することができる. この計算量の解析によって, 提案アルゴリズムが, グラフ中のすべての誘導木を $O(k)$ ならし遅延で列挙することを示す. さらに, 提案のアルゴリズムは, 入力グラフの縮退数が定数の場合, 遅延が定数となる最適な列挙アルゴリズムである.

次に, グラフの一般化である超グラフ中の非巡回部分超グラフの列挙問題について考察する. 超グラフの非巡回性は, データベースのスキーマ設計において重要な概念である. 非巡回性の定義は, 階層的に与えられており, 制約の少ない順に, Berge-, $\alpha$-, $\beta$-, $\gamma$-非巡回性と定義されている. これまでに, Daigo と Hirata により, 極大な $\alpha$-部分超グラフの列挙アルゴリズムが示されている. しかし, 他の非巡回性に対しては, 生成検査法による列挙アルゴリズム以外は知られていない. そこで, 本研究では, 与えられた超グラフに含まれる連結かつ Berge-非巡回な部分超グラフを列挙する問題を考察する. 我々は, 連結な Berge-非巡回な部分超グラフに対する家系木を, 除いても連結性を維持する超辺に着目し, そのような超辺を追加削除することで親子関係を定義する. 提案のアルゴリズムは, この家系木を用いることで, 解 1 つあたり, $O(\|S\| + \|N(S)\|)$ 時間で列挙する. ただし, $S$ と $N(S)$ を, それぞれ, 出力する解の超辺集合と $S$ に接続する超辺の集合とし, $\|S\| = \sum_{s \in S} |s|$ とする. これは, 連結かつ Berge-非巡回な部分超グラフを多項式遅延で列挙する初めてのアルゴリズムである. 最後に, いくつかの未解決問題を示し, 本研究の今後の方向を与える.