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_review.pdf, 審査の要旨 |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　和佐 州洋

審査担当者　主　査　教　授　有村 博紀
副　査　教　授　原口 誠
副　査　教　授　湊 真一

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）

本論文では, グラフや超グラフといった大規模な離散構造データから, その中に含まれる与えられた制約をみたす部分構造を網羅的に見つけ出すための列挙アルゴリズムについて研究する. このような離散構造に対する列挙アルゴリズムは, 1970 年代から研究されており, これまでに全域木や, パス, サイクル, クリーク, マッチング, カット, 部分木などの部分構造列挙アルゴリズムが提案されている. またこれらのアルゴリズムは, データマイニングや機械学習へも応用されている.

列挙アルゴリズムは, 他のアルゴリズムに比べて, 次のような特徴を持つ. 一般に, アルゴリズム理論ではアルゴリズムの効率を入力サイズ $n$ に対する総計算時間で測り, それが $n$ の多項式で抑えられるときに「効率良い」と考える. ここで サイズ $n$ の入力のすべての部分構造を解とするような列挙問題の場合, 最悪時に $n$ の指数個の解が存在するので, その場合にはどのような列挙アルゴリズムでもすべての解を出力するには $n$ の指数計算時間をもつので,「効率が悪い」ことになる. 一方で, 一般に解の個数は $O(2^n)$ よりも小さい場合も多い. 実際の状況では解の数が少ないときには計算が早く終了して欲しいことが多く, 解の数が多いときにはそれに比例して計算時間がかけても良いと考えるのは自然である. そこで, 上記の 2 点を考慮して, 列挙アルゴリズムの計算量解析では, 入力と出力の両方の関数として計算時間を評価する. とくに本研究では, 列挙アルゴリズムの評価を「遅延」(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の存在が示されている列挙問題の数は非常に少ない. それどころか, 多項式遅延を持つような問題, すなわち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多く残されている. 本論文では近年, 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種々の列挙問題の制約として与える. 具体的には, 次の 4 つの問題について考察し, 遅延の観点から高速な列挙アルゴリズムを構築することを目標とする.

3 章では, $k$-部分木列挙問題について考察している. 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 $k$-部分木を列挙する問題である. 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している. 基本的なアイディアとして, 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 種類の親子関係に基づいて探索空間を網羅

する探索木 (「家系木」とよぶ) を構築し探索することで, 木に含まれるすべての $k$-部分木を最悪時に定数遅延で列挙するアルゴリズム を与えている.

4 章では, 誘導グラフの列挙問題, とくに, グラフに含まれる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している. この問題に対して, 従来手法を用いて $d$ 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, 縮退列という特徴的な頂点列をもつことが知られている. 定数であるグラフ族として, 平面グラフなどの興味深いクラスが知られている. 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している. 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に遅延が定数となる最適な列挙アルゴリズムである.

5 章では, 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している. 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し. 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ことで, この探索木を用いることで, 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.

さらに 6 章では, $k$-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している. 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ている.

これを要するに, 著者は, 離散アルゴリズムにおける種々の非巡回部分構造の列挙問題において, 問題の理論的解析に基づいて効率良い部分構造の列挙アルゴリズムの設計と実装が可能であるという新知見を得たものであり, 情報科学における離散アルゴリズムと大規模データ処理分野において貢献するところ大なるものがある. よって著者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める.