



HOKKAIDO UNIVERSITY

Title	Studies on Depth-First Mining Algorithms for Maximal Flock Patterns from Large Trajectory Data [an abstract of dissertation and a summary of dissertation review]
Author(s)	耿, 晓亮
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第11970号
Issue Date	2015-09-25
Doc URL	https://hdl.handle.net/2115/59948
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Geng_Xiaoliang_review.pdf, 審査の要旨



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 耿 暁亮

審査担当者 主 査 教 授 有村 博紀

副 査 教 授 湊 真一

副 査 教 授 原口 誠

学位論文題名

Studies on Depth-First Mining Algorithms for Maximal Flock Patterns from Large Trajectory Data

(大規模軌跡データからの極大群パターン発見のための深さ優先アルゴリズムに関する研究)

データマイニングの研究は、人工知能・数理統計・データベースの融合領域としての誕生から、分類予測と、クラスタリング、パターンマイニングなどの多様な研究トピックに関して展開され、この20年余りの間に飛躍的な発展をみせている。パターン発見 (またはパターンマイニング) は、データマイニングの一分野であり、さまざまなデータとパターンのクラスを対象に、新しいパターン発見問題の定式化や、効率良いパターン発見アルゴリズムの設計と実装について詳細な研究成果が得られている。また、こうしたパターン発見技術に基づいて、実際のデータと問題の領域における適用研究も盛んに展開されている。しかし、いわゆるビッグデータ処理でいわれるところの超大規模データに対する規模耐性のあるアルゴリズムの設計や、実際の問題領域に出現する非構造データに対する適切なパターン発見問題の定式化、時間に伴って変化するデータ流の適応的処理の問題など、重要な課題が未着手または未解決のまま、数多く残されているのも事実である。

本論文は、これらの問題に関して、ビッグデータ処理で注目されている大規模トラジェクトリデータ (移動体の移動軌跡データ) に焦点をおき、複数の移動体の一定時間内の類似した行動を表す群れパターン (フロックパターン) とその拡張を対象にして、メモリ効率と時間効率の良いパターン発見アルゴリズムの開発を追求し、フロックパターン発見問題における新たなパターンクラスの提案と効率良いパターン発見アルゴリズムの開発を行い、トラジェクトリデータに対するパターン発見技術の新たな展開を行ったものである。このフロックパターンは、複数の軌跡が一定以上の長さの時区間で、互いに近接して移動する状況を表現するパターンであり、それが定める時区間 (開始時間と終了時間) とそれが含む軌跡集合 (移動体の番号の集合) の対で表現される。時区間長をパターンの長さといい、含まれる軌跡数をサイズと呼ぶ。フロックパターンは、データが密集する部分時空間を表すいわゆる密領域パターンの一種と考えられる。

著者は、フロックパターンマイニングにおける多数の類似した冗長パターンの発見の問題を解決するために、長さ極大性とサイズ極大性の二点に着目し、新しくいくつかの極大フロックパターンのクラスを提案している。さらに、これらのクラスに対して、効率良い深さ優先型パターン発見アルゴリズムを開発している。従来から、パターン発見アルゴリズムのメモリと時間の効率を向上させるには、深さ優先探索型アルゴリズムが有効であることが指摘されてきた。しかし、作業メモリを用いた幅優先探索型アルゴリズムと異なり、深さ優先型アルゴリズムでは重複解の回避が必要であり、一般に設計が難しいという問題点をもっていた。これに対して、著者は、導入した極大パターンクラスに対して、その極大パターンがもつ閉包拡張演算を利用することで、効率良い深さ優先型パターン発見

アルゴリズムの実現を行っている。具体的には、著者は以下のような成果を得ている。

第3章では、同じ開始時間と軌跡集合を持つパターンの中で、長さが最も長いようなパターンである「右方向長さ極大ブロックパターン」のクラスを導入している。さらに、このクラスに対して、与えられた軌跡データベースから、定められた最大半径と最小長さをもつような右方向長さ極大ブロックパターンすべてを、入力が多項式のメモリ量と遅延時間（出力されるパターン一個あたりの計算時間）で重複なしに発見する効率良いアルゴリズムを提案している。そのために、与えられた開始時点と軌跡集合から、それらを満たす一意なこのクラスのパターンを与える右方向閉包拡張という演算を考案し、上記のアルゴリズムの重要な部分で用いている。さらに、空間索引データ構造を用いた高速化手法を与えている。

第4章では、第3章の結果を拡張して、同じ軌跡集合を含みながら、左方向にも右方向にもそれ以上時区間を拡張できないようなパターンである「両方向長さ極大ブロックパターン」のクラスを導入している。このクラスに対して、与えられた軌跡データベースから、定められた最大半径と最小長さをもつような右方向長さ極大ブロックパターンすべてを、入力が多項式のメモリ量と遅延時間で重複なしに発見する効率良いアルゴリズムを提案している。そのために、与えられた軌跡集合を満足し、指定した開始時間を含む一意な極大パターンを与える演算を新たに導入し、上記のアルゴリズムで用いている。

第5章では、第3章と第4章とは異なった視点から、大量の冗長パターンを回避する問題を考察している。ブロックパターン発見においては異なる軌跡集合をもつ二つ以上のパターンが同じ区間をもち、さらに空間の同じ領域を占める場合に、これらは時空間に同じ情報を表すので冗長と考えられる。そこで、同じ時区間と領域を占めるブロックパターンの中で、その軌跡集合が集合の包含関係に関して極大なパターンである「サイズ極大ブロックパターン」のクラスを導入している。このクラスに対して、与えられた軌跡データベースから、指定の条件を満たすようなサイズ極大ブロックパターンすべてを、入力が多項式の遅延で重複なしに発見する効率良いアルゴリズムを提案している。ただし、このアルゴリズムでは、発見したパターンをメモリ上の表に格納して重複パターンの回避を行っているため、発見されるパターン数に比例したメモリ量を必要としており、この点の改善が今後の課題となっている。

上記の結果に加えて、3章および、4章、5章では、提案アルゴリズムの実証的な性能評価のために、システムの実装と、生成データおよび実データ上の計算機実験を行っている。実験結果からは、提案した極大パターンのクラスと深さ優先探索型パターン発見アルゴリズムの有効性を確認している。さらに、提案手法の有用性を検証するために、台風軌跡の実データを用いてパターン発見実験を行っている。

これを要するに、著者は、トラジェクトリデータに対するパターン発見において、従来の技術がもつ問題点の解析に基づいて、種々のパターンのクラスと関連するパターン発見問題を提案し、これらの問題に対して効率良いパターン発見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、情報科学におけるデータ工学と知識発見の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。