科学研究費助成事業（特別推進研究）公表用資料〔追跡評価用〕

「非構造マイニングの超高速アルゴリズムの研究」

（平成 17～19 年度 特別推進研究「知識基盤形成のための大規模半構造データからの超高速パターン発見」）

所属・氏名：北海道大学・大学院情報科学研究科・教授・有村　博紀

１．研究期間中の研究成果

・背景（事象の初歩的な説明）

『データマイニング』は，大量のデータから人間にとって有用な規則性やパターンを発見するための効率よい計算手法の研究であり，1990 年代の登場以後，急速に発展している．2000 年代に入って，大規模で非均質な非構造データ（いわゆる『ビッグデータ』）を対象としたマイニングが世界的に注目を集めているが，これらの大規模非構造データに対する理論的に性能保証をもつ高速なアルゴリズムは，ほとんど知られていなかったのが現状であった．

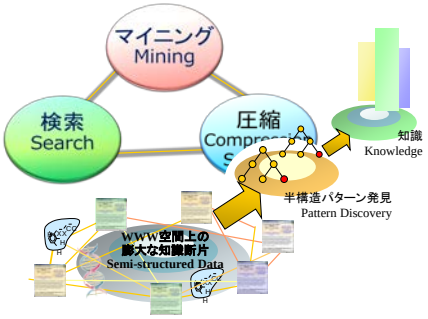

図 1．超高速非構造マイニング

・研究内容及び成果の概要

そこで本研究では，このような大規模非構造データを系列，木，グラフ等の離散構造データとしてモデル化し，大規模離散構造データを対象に，超高速かつ頑健なパターン発見手法を究明する．そのため離散構造アルゴリズムと計算量理論を援用し，次の課題を研究した．

① 中核技術として，集合と，系列，木，グラフ等の離散構造データに対する理論的性能保証（多項式遅延・領域量）をもつ超高速マイニング手法を研究開発した．
② 周辺技術として，高速な照合技術と圧縮技術にもとづく大規模非構造データ連携技術を研究開発した．
③ ゼロサプレス二部決定グラフに基づく非構造データ向け知識索引技術を研究開発した．

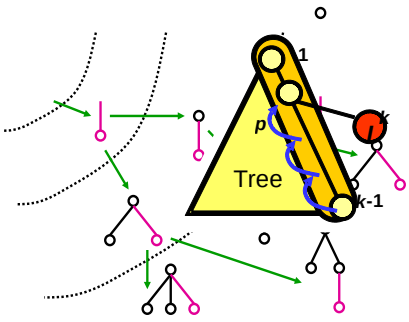

図 2．高速非構造パターン発見手法

２．研究期間終了後の効果・効用

・研究期間終了後の取組及び現状

① **高速イベントストリームデータからエピソードと呼ばれる時系列パターンを発見する新種のマイニング問題をとく効率よいアルゴリズムを開発した．**
② **誤差やノイズを許したマイニングや，連続的時空間を扱う大規模時空間データからのマイニングの理論的性能保証をもつ極大パターン発見アルゴリズムを開発した．**
④ 上記の解一個あたりの理論的性能保証をもつ極大パターン発見アルゴリズムの設計法の一般理論を構築した．

・波及効果

① 開発した超高速パターン発見手法は，非構造データに対する国際的標準手法になっている．
② 深さ優先極大マイニングや大規模知識索引等のビッグデータ時代の先験的成果が生まれた．
③ 本研究の着想から，大型研究教育プロジェクト（湊 ERATO，北大情報 GCOE）が育った．