審 査 の 結 果 の 要 旨

氏名　　楊　征路

本論文は「Fast Algorithms for Sequential Pattern Mining　（シーケンシャルパターンマイニングにおける高速化アルゴリズム)」と題し、英文８章から構成されている。一般的なシーケンスデータベースから頻度高く現れるシーケンスパターンを効率よく取り出す方式を提案し、人工的データセットおよび実応用データセットを用いた実験を行い、提案する方式の有効性について論じている。

第１章は、「Introduction (序章)」であり、本論文の背景および目的について概観し、本論文の構成を述べている。

第２章は、「Problem Definition and Related Work（問題定義と関連研究)」と題し、シーケンシャルパターンマイニング処理における問題を明らかにし、関連研究をまとめている。

第３章は、「LAPIN：Efficient Sequential Mining Algorithms (LAPIN：高効率シーケンシャルマイニングアルゴリズム)」と題し、新しいシーケンシャルパターンマイニングアルゴリズム LAPIN を提案している。LAPIN では、パターン探索時に部分シーケンスにおけるアイテム出現最後尾位置に着目することで、アイテムを直接比較せず、出現最後尾位置における比較により探索コストを抑えることで高速化を実現している。提案した方式を詳述すると共に、提案したアルゴリズムの正しさ、および、その効率について理論的に解析し、さらに多様なデータセットを用いた実験を行い、提案した方式の有効性を示している。

第４章は、「Improved Efficient Sequential Mining Algorithms（改良シーケンシャルパターンマイニングアルゴリズム)」と題し、LAPIN アルゴリズムの候補シーケンスの探索において、プレフィックス部分のシーケンスの位置を用いることにより候補シーケンスを生成するコストを削減し、一層の性能の向上が実現されることを示している。二つの改良アルゴリズムを提案し、様々な実験を行い、主記憶容量に制約がある場合も含めて、従来のアルゴリズムと比較し、最も高い性能を得られることを示している。

第５章は、「Applications of Sequential Pattern Mining（シーケンシャルパターンマイニングアルゴリズムのアプリケーションへの適用)」と題し、提案した LAPIN アルゴリズムを実際のアプリケーションに適用し、その結果を示している。即ち、ウェブログ処理および DNA シーケンスデータセットの二つのアプリケーションを用いて詳細な実験を行い、従来のアルゴリズムと比較し、提案するアルゴリズムがより高い性能を示すことを確認している。

第６章は、「Extension of Sequential Pattern Mining（シーケンシャルパターンマイニングアルゴリズムの拡張応用)」と題し、提案したアルゴリズムを拡張しスカイライン問合せに適用することを検討している。スカイライン問合せの実験結果から、提案手法が従来の

スカイライン問合せ手法に比べ高性能であることを確認するとともに、提案方式はわずかに改良を行うことにより、多様なデータセットの解析に適用可能であることを示している。

第７章では、「Discussion（議論）」と題し、提案したシーケンシャルパターンマイニングアルゴリズムが、単純なシーケンシャルパターンマイニングのみならず、制限付きパターンマイニング、推定パターンマイニング、パターン圧縮、データストリームマイニング、Top-K 問合せ等、様々なデータ解析の問題に適用する場合について議論、検討している。

第８章「Conclusions（結論）」では、本論文の成果と今後の課題について総括している。

以上これを要するに、本論文は、処理負荷の高いシーケンシャルパターンマイニングアルゴリズムをデータシーケンスにおける部分パターンの最後尾位置に着目することにより、探索コストを削減、性能向上を図るものであり、理論的解析および実機上での様々なデータセットによる性能解析、さらには、実際のアプリケーションデータを用いた評価により、従来手法より高い性能が得られることを明らかにすると共に、シーケンシャルパターン探索のみならずスカイライン問合せ処理等の様々なデータ解析に広く適用可能なことを明らかにしており、電子情報学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。