

論文内容の要旨

博士論文題目

時系列パターンマイニングアルゴリズムの高速化手法に関する研究

氏名

松原 裕貴

(論文内容の要旨)

時系列パターンマイニングの分野では大規模なデータベースを処理した際に、多くの処理時間が必要とされることが問題視されており、過去の研究において様々なアルゴリズムの提案が行われてきた。

本論文では、既存の時系列パターンマイニングアルゴリズム PAID のアルゴリズムのコアを変更することなく CPU キャッシュの利用効率の向上と計算量及びキャッシュミス削減と並列化による高速化を目的としたアルゴリズムの提案を行った。

一つ目は、PAID で利用される特定のデータ構造の時間的局所性を処理対象へのアクセスパターンを変更することにより向上させ、CPU キャッシュに取り込んだデータを有効的に利用することを目的とした CC-PAID アルゴリズムを提案した。複数の時系列パターンで CPU キャッシュに取り込まれたデータの再利用を行うことにより、CPU キャッシュの利用効率が向上し、CPU のキャッシュミスが削減されることによりキャッシュミスにより生じるレイテンシを少なくし、処理時間を短縮する。

二つ目は、CC-PAID のデータ構造のアクセス効率に着目し、特定のデータ構造を動的に再構築し、それを複数の処理で有効的に利用することにより、特定のデータ構造に対するアクセスと処理の省略を行うことを目的とした CCDR-PAID の提案を行った。動的に再構築されたデータを用いることにより、省略可能なアクセスを省略すると共に特定の処理も省略し、更に不要なデータが取り除かれた状態で処理が行われるため、CPU キャッシュに読み込まれるデータ量と計算量を削減することができる。

また、時系列パターンマイニングは並列処理に適しているため、並列処理による更なる速度向上のために PAID アルゴリズムと上記の二つの提案手法の並列化も行った。

更に PAID, CC-PAID, CCDR-PAID の処理時間、キャッシュミス、実行命令数、メモリ使用量、並列処理時のスケーラビリティ等の評価を行い、有効性の確認を行った。

(論文審査結果の要旨)

本論文は、既存の時系列パターンマイニングアルゴリズムである PAID の中核を変更することなく処理を高速化する方法として、CPU のキャッシュミスと計算量の削減に着目したアルゴリズムに関して述べている。

本論文の主眼は、時系列パターンマイニングという基本的にワーキングセットの大きいアプリケーションに対して、データ構造を改良しワーキングセットを縮小することでキャッシュミスを削減し、また、アルゴリズムの特徴を活かして不要な処理を判定し、省略することで計算量を減少させ、処理を高速化することにある。そのために、(1) 処理対象となる時系列パターンのアクセスパターンに着目し、処理可能な時系列パターンを幅優先で一括処理することにより CPU のキャッシュの時間局所性を向上させキャッシュミスを減らす CC-PAID アルゴリズム、(2) CC-PAID アルゴリズムを改良し、データアクセスやある処理を省略可能とするために特定のデータ構造を動的に再構築し、異なる処理間で共有することによりキャッシュミスと計算量の削減を行う CCDR-PAID アルゴリズムの二つを段階的に提案している。また、これら二つの提案アルゴリズムの並列化も同時に行っている。提案手法の評価では、様々な特徴のデータセットを用いて、処理時間だけでなくキャッシュミス回数や命令実行回数などの詳細なデータを示して提案手法の性質を明らかにしつつ、提案アルゴリズムが最大約 2.8 倍の高速化が可能であると共に、並列化により 6 コア使用時に更に最大約 4 倍高速化されることを実証している。

これらの提案手法は、キャッシュメモリの大容量化ならびにマルチコア化という近年のプロセッサ技術の特徴を巧みに活用し、データマイニングという極めて重要なアプリケーションにうまく適合させたという点で、新規性、有効性に不足はない。また、実証実験により提案手法の性質を網羅的かつ詳細に調査している点で信頼性も十分である。

論文の主要部分に相当する内容は、既に学術論文誌に掲載されると共に、国際会議においても公表が決定している。以上のように意義深い成果が得られていることから、本論文を博士(工学)の学位論文として価値あるものと認める。