

# 要 旨

## 題 目

### 多重集約演算のための 並列パイプライン処理方式

前田庸亮

複数の集約演算を並列に処理するアルゴリズムとして、**m2P**、**mRep**、**mDC**などのアルゴリズムが提案されている。これらのアルゴリズムはすべてのノードが直接結ばれているモデルを想定すると、より大規模な並列計算機ではこの仮定が成り立たない。そこで本稿では、より拡張性に優れたリング構造向きのノード間通信アルゴリズムを提案し、並列パイプライン化を適用して通信時間の隠蔽による高速化を図る。提案するアルゴリズムは各ノードが、演算とノード間の通信を交互に繰り返し、集約処理を進める。通信は隣接した特定のノード間で行われ、衝突が起こることは無い。

既存のアルゴリズム評価に用いられているコストモデルに最小限の変更を加えて、**mDC**、**mRep**、並列パイプライン化適用前及び、適用後の本アルゴリズムを評価したところ、パイプライン化は全ての場合に効果があり、特にノード数が少ない場合に有力であった。次に、通信速度を変化させた場合、ノード間の通信速度が比較的遅い場合に改善の効果が大きかった。さらに、目標としていたノード数が大きい場合はいずれのアルゴリズムよりも高速であるという良好な見通しを得た。

**キーワード** データマイニング, 集約演算, 並列パイプライン, ノード間通信

# Abstract

## Parallel Pipelining Method for Multiple Aggregate Queries

YOUSUKE MAEDA

Existing parallel algorithms to do multiple aggregations—for instance m2P, mRep, and mBC—are based on a model in which every node is directly connected with each other. But, the model is inadequate for up-to-date super parallel computers. A novel inter-node communication algorithm adopting the ring structure and its parallel pipelined improving trial is described. In this paper using that every node processes aggregations repeating a set of computations and communications. Inter-node communications occur between the next node respectively, so that contentions would not be happen. The cost model used in evaluations of existing algorithms is modified to support the ring structure. The modified cost model shows.

- (1) Pipelining is always effective, especially in case of the number of nodes is small.
- (2) A significant improvements of performance is obtained at relatively low band width in inter-node communication.
- (3) As the number of nodes maximized, the new algorithm exceeds the other.

*key words* Data Mining, Aggregation, Parallel Pipelining, Inter-node Communication