

WorldTSP における MapReduce の適用

山下 尊也[†] 廣安 知之^{††} 三木 光範[‡] 横内 久猛^{††}[†]同志社大学工学部学生 ^{††}同志社大学生命医科学部 [‡]同志社大学工学部

1 はじめに

巡回セールスマン問題 (Traveling Salesman Problem: TSP) は, 都市間を結ぶ最短の巡回路を求める問題である. また, TSP の中でも 1,904,711 都市を巡回する巡回路を最適解に近づける問題として, World Traveling Salesman Program (World TSP)[1] があり, 毎年, 最短の巡回路の更新がされている.

WorldTSP は都市数が多いため, 最適な巡回路を求めるためには, 多くの計算時間がかかる. そのため, PC クラスタを用いた分散処理を用いることで計算時間を軽減する必要がある. そこで大量の Web 情報を処理するために, Google で用いられている分散処理プログラミングモデルである MapReduce を用いる探索手法を提案する.

2 TSP と MapReduce について

2.1 巡回セールスマン問題 (TSP)

TSP とは, あるセールスマンが, 様々な都市を一度ずつ訪問して出発点に戻って来た際の総距離が最短になる経路を求める問題である. TSP は, 都市間の組み合わせを考えるため, 都市数が n 個の場合, 巡回路は $\frac{(n-1)!}{2}$ 通り存在する. そのため, 都市の数が増えるに従い, 組み合わせの数は膨大なものとなる.

2.2 MapReduce

MapReduce は, Google の技術者である Jeffrey Dean と Sanjay Ghemawat により提案された, 大規模 PC クラスタ上での分散プログラミングのモデルである [2].

Google では, 図 1 に示すように Map と Reduce という 2 つのプロセスを組み合わせ, 大量のデータ処理を PC クラスタ上でやっている.

以下では, MapReduce の 2 つのプロセスについて説明する.

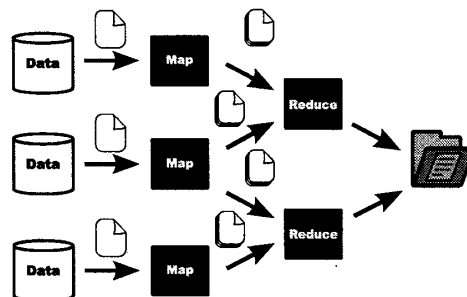


図 1: MapReduce の処理

2.2.1 Map

Map とは, ひとまとまりのデータから新しい有用なデータへと変換するプロセスである. 図 1 に示すように Map では様々なデータを受け取り, データを変更し Reduce に引き渡す.

2.2.2 Reduce

Reduce は, 図 1 に示すように Map から得られた複数のデータを統合処理し, 最終的に得たいデータを出力するプロセスである.

3 TSP における MapReduce の適用

WorldTSP を解くために, 図 2 に示すように, 対象問題の領域を分割し, 分割した巡回路を組み合わせ, 新しい巡回路を生成する手法を MapReduce に適用する. MapReduce の各プロセスは以下のように行なった.

1. 領域分割 (Map)
元の領域から複数の領域に分割を行う.
2. 領域の組み合わせ (Reduce)
分割した領域に対して最適化を行い, 最適化された巡回路を組み合わせ, 全領域を巡回する経路を生成する.

3.1 領域分割 (Map)

探索領域の分割については, 枝組み立て交叉 (Edge Assembly Crossover: EAX)[3] で用いられている中間個体を利用する. 以下に中間個体を用いた領域の分割手順を示し, 図 3 に概念図を示す.

1. 与えられた都市から, 親子体を生成する. この時, 生成された親個体を $parent-A, parent-B$ とする.

Applied MapReduce for WorldTSP

[†] Takaya YAMASHITA (tyamashita@mikilab.doshisha.ac.jp)^{††} Tomoyuki HIROYASU (tomo@is.doshisha.ac.jp)[‡] Mitsunori MIKI (mmiki@mail.doshisha.ac.jp)^{††} Hisatake YOKOUCHI (yoko@mis.doshisha.ac.jp)Department of Life and Medical Sciences, Doshisha University ([†])Department of Knowledge Engineering and Computer Science, Doshisha University (^{††})Graduated School of Knowledge Engineering and Computer Science, Doshisha University ([†])

1-3 Miyakodani, Tatara, Kyotanabe, Kyoto 610-0321, Japan

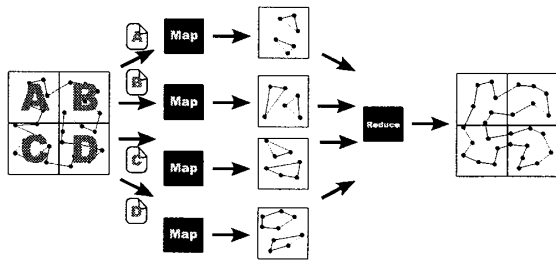


図 2: TSP に MapReduce を適用した際の処理の流れ

2. *parent-A* と *parent-B* を組み合わせ、経路を G' として生成する。
3. G' から部分巡回路である *AB-cycle* を生成する。*AB-cycle* は、*parent-A* と *parent-B* の巡回路の枝を交互に選び、その枝を繋ぐ事で生成される部分巡回路のことである。
4. 生成した複数の *AB-cycle* から、1つの *AB-cycle* を選び、抜き出したものを *E-set* とする。
5. *E-set* を *parent-A* に適用する。*E-set* を構成している *parent-A* の枝は、*parent-A* の巡回路から削除を行い、*parent-B* の枝は、*parent-A* の巡回路に追加することで領域を分割する。

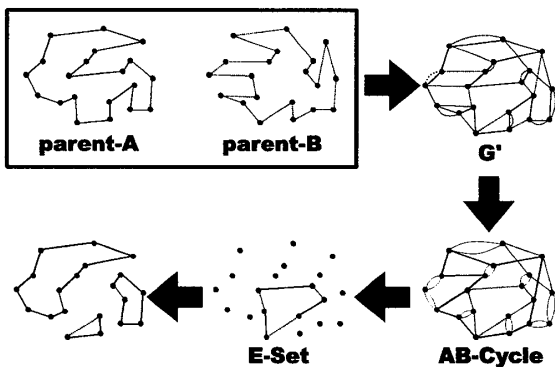


図 3: 中間個体を用いた領域の分割手順

3.2 領域の組み合わせ (Reduce)

Map により分割された領域から全領域を巡回する経路を生成する。以下に分割された領域から全領域を巡回する経路を生成する手法の手順を示し、図 4 に概念図を示す。

1. 分割された領域に 2-Opt を用いて最適化を行い、巡回路を生成する。
2. 図 5 で示すように、分割された領域を *area-A*, *area-B* とし、領域 *area-A*, *area-B* にある都市

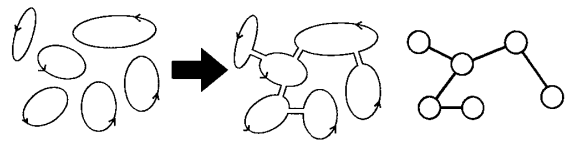


図 4: 最小全域木を用いた領域の組み合わせ

を k, l とする。そして、領域 *area-A*, *area-B* の距離 $d(k, l)$ が最小になるような都市を選び、巡回路として繋げる。

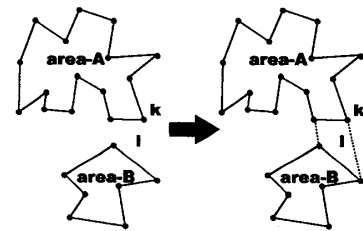


図 5: 分割された領域の組み合わせ

4 まとめと今後の課題

今回、大規模な TSP を解くために分散処理プログラミングモデルの MapReduce を適用する手法を提案した。Map では、元の領域から複数の領域に分割を行い、Reduce では、Map で分割された領域から全領域を巡回する経路を生成した。

今後は、領域の分割手法にクラスタリング手法を用い、領域の分割精度の向上や、各処理に対して効率の良い並列化を行い、巡回路の改善や探索時間の削減を行っていく。

参考文献

- [1] World Traveling Salesman Problem
<http://www.tsp.gatech.edu/world/>
- [2] Jeffrey Dean and Sanjay Ghemawat: MapReduce: Simplified Data Processing on Large Clusters, Google, Inc. (2004)
- [3] 永田裕一, 小林重信: 巡回セールスマン問題に対する交叉枝組み立て交叉の提案と評価, 人工知能学会誌, Vol.14, No.5, pp.848-859(1999)
- [4] 茨木俊秀: 岩波講座 応用数学 [方法 8] 離散最適化法とアルゴリズム, p.89, 岩波書店 (1993)