

衛星画像データ解析システムにおける 高速化の研究

森 駿[†] 中村 彰宏[†] 朴 鍾杰^{††} 布広 永示^{††}

東京情報大学では、戦略的研究基盤形成支援採択事業として、NASA によって開発された地球観測衛星 Terra/Aqua の MODIS センサーで受信している衛星画像データ (MODIS データ) を利用し、「環境変動」に関する研究を進めている。この戦略的研究事業の研究を支援する情報基盤として、Web システム、PC クラスタシステム、データベースシステムなどで構成される衛星画像データ解析システムを開発し、システムの利便性やアプリケーションプログラムの実行環境の改善を行っている。本報告では、衛星画像データの利便性を向上するためのピース化処理・コンポジット処理、解析処理性能の向上するための MODIS データの仮想化、PC クラスタなどの計算資源のスケジューリング処理の概要、性能評価結果について報告する。

Research on acceleration of satellite image data analysis system

Hayao Mori[†] Akihiro Nakamura[†]
Jong Geol Park^{††} Eiji Nunohiro^{††}

Tokyo University of Information Sciences receives MODIS (Moderate Resolution Imaging Spectroradiometer) data, one of the sensors equipped by NASA's Terra and Aqua satellites, and researches of the analysis on change of environment as part of the academic frontier project. For the information infrastructure of this frontier research, we are developing a satellite image data analysis system to support of web system, a parallel distributed system configuration using multiple PC clusters, database for MODIS data to open the research results and MODIS data for public use. This paper presents the overview of satellite data analysis system, new feature which are peace processing of satellite data and composite of multiple satellite data for same location to improve convenience of data, virtualization of storage system for satellite data, scheduling for multiple cluster nodes and performance evaluation.

1. はじめに

東京情報大学では、戦略的研究基盤形成支援採択事業として NASA の衛星 Terra/Aqua に搭載された MODIS (Moderate Resolution Imaging Spectroradiometer) センサーで受信される衛星画像データ (以下、MODIS データ) を利用した「環境変動」の解析に関する研究を進めている。本報告では、気象変化や環境変動などに関する研究を支援する衛星画像データ解析システム (以下、解析システム) の利便性や高性能化を実現する上で必要なデータベース化、計算資源の仮想化やスケジューリング機能などについて記述する。2章では、本解析システムの概要や MODIS データについて説明し、3章で MODIS データのデータベース化の概要、4章でアプリケーションの実行制御とその基盤技術であるデータの仮想化、スケジューリング機能、5章で性能評価について報告する。

2. 衛星データ解析システム

2.1 システム概要

解析システムは、戦略的研究基盤形成に関する研究の一環として進められている MODIS データの解析処理を支援するデータ解析システムである。この解析システムを利用するユーザは Web 環境を利用し、解析処理を行うアプリケーションプログラム (以下、AP) の実行や解析結果の確認を行う事が出来る。解析処理としては、林野火災探索、類似画像探索、土地被覆変化解析や気象変動予測などがあり、現在のそれらの AP を開発中である。MODIS データを使用する解析処理は、システムのリソースを大量に利用するため、ターンアラウンドタイムやスループットに大きく影響する。そこで、システム資源の有効利用や実行性能を向上させるため、データの仮想化やスケジューリング機能を開発した。また、解析処理で利用する衛星画像データのデータベース化においては、データのピース化処理やコンポジット処理を行うことによって、解析処理に応じたデータの利便性を向上した。

2.2 衛星画像データ (MODIS データ)

MODIS センサーは、NASA によって開発された地球観測衛星 Terra/Aqua に搭載されている光学センサーの一つである中分解能撮像分光放射計である。MODIS の観測幅は 2330km で 36 のバンド (可視域から熱赤外域まで: 0.405~14.385 μ m) と 3つの空間分解能 (250m, 500m, 1000m) を持っている。東京情報大学では MODIS データ

[†] 東京情報大学大学院総合情報学研究科
Graduate School of Tokyo University of Information Sciences

^{††} 東京情報大学総合情報学部
Tokyo University of Information Sciences

を東経 105 度から 180 度まで、北緯 15 度から 65 度の範囲で受信している。解析システムは、受信した MODIS データの中から MODIS 標準プロダクトである MOD02 (大気上段反射率データ), MOD03 (衛星姿勢データ), MOD09 (地表面反射率データ), MOD11 (地表面温度), MOD14 (異常温度), MOD28 (海面温度) を管理する。MODIS データのバンド情報を表 1 に示す。

表 1 バンド情報

バンド	波長 (nm)	空間解像度	利用分野
1	620~670	250m	土地被覆変化 植物葉緑素
2	841~876	250m	雲量 土地被覆変化
3	459~479	500m	土 / 植物識別 雲判断
4	545~565	500m	緑の植物
5	1230~1250	500m	葉 / 天蓋差
6	1628~1652	500m	スノー / 雲識別
7	2105~2155	500m	雲プロパティ 土地プロパティ
21	3929~3989	1000m	森林火災 / 火山
22	3929~3989	1000m	雲の温度 / 森林火災
32	11770~ 12270	1000m	雲の温度 / 森林火災 / 火山 / 表面温度

MOD02 は、異常温度探知のため、36 バンドの中から熱赤外域であるバンド 21, 22 とバンド 32 を用いる。MOD03 は、衛星の姿勢情報から、各画素の緯度経度情報とセンサーの角度や太陽の位置などが記録されている。この情報は MODIS データの地図化やコンジョイント処理時に有効なデータである。MOD09 は MOD02 データから大気のエアロゾルの散乱や吸収の補正を行った高次処理データである。MOD09 は 7 つのバンドで構成され、3 つの可視域バンド (バンド 1, 4, 3) を用いた衛星データの可視化、

近赤外バンド (バンド 2) を用いた植生指数 (光合成活動に相関が高い)、3 つの短波長赤外バンド (バンド 5, 6, 7) で雪と雲の識別や異常高温地域検出に有効なデータである。主に陸域変化抽出や土地利用解析に多く用いられる。

2.3 MODIS データの利用例

解析システムの目的の一つとして、MODIS データを利用して、環境の変化や地表面の変化などに関する研究を支援することがある。これらの研究の解析手法としては、自己回帰分析などの統計解析やニューラルネットワークなどの知識情報処理などを適用し、土地被覆変化の抽出、NDVI の時系列予測などがある。これらの研究の一部を下記に示す。

①東アジアの土地被覆変化抽出の傾向分析

2001 年から現在まで東アジアの土地被覆変化を抽出し、その原因を調べる。

②東アジアの災害図 (火災, 洪水) の作成

洪水や火災など自然災害の抽出の自動抽出について研究する。

③自己回帰分析を用いた NDVI の時系列予測

ある地点の気象変化は遠く離れた海水域や大陸の温度によって決まる。そこで、統計的手法を用いて広域観測が可能な衛星データを用いて水表面温度と地表面温度から正規化植生指数 (NDVI : Normalized Difference Vegetation Index) の変化を調べる。

④知能情報処理を用いた土地被覆領域の境界抽出

ニューラルネットワーク, 自己組織化マップなどの知能情報処理手法を用いた土地被覆分類の研究を行い、土地被覆分類の精度向上を目指す。

2.4 システム構成

解析システムの構成を図 1 に示す。解析システムは、Web サーバ、PC クラスタ (32 台) と BladeCenter で構成される計算ノード、データベースで構築される。

ユーザは、Web サーバにアクセスし、AP の実行や MODIS データを利用する。Web サーバは、ユーザからの処理を受け付けて、計算ノードで実行し、その結果を表示する。計算ノードは、AP の特性に応じて PC クラスタや BladeCenter 上で AP を実行する。データベースは、AP が利用し易いようにピース化処理やコンジョイント処理された MODIS データを管理する。

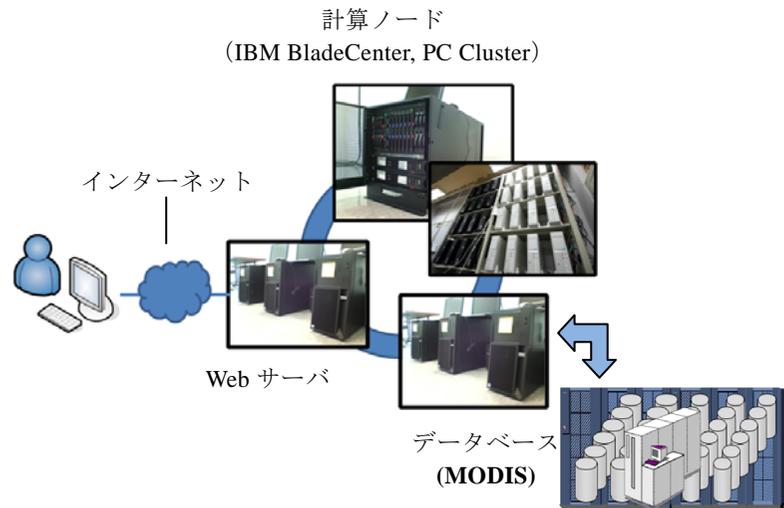


図1 システム構成図

3. MODIS データのデータベース化

MODIS データのデータベース化においては、MODIS データの公開や各種解析処理が利用し易いように MODIS データを加工（ピース化処理・コンポジット処理）して管理する。

3.1 ピース化処理

異なる衛星軌道から観測したデータを時系列的に解析するためには、各データを決まった地図投影法で表示させる必要がある。地図投影はセンサー観測データ（MOD02 又は MOD09）とデータの地図座標情報（MOD03）を用いて再サンプリングする。このため、あるピクセルが撮影範囲外の経緯度の場合には、撮影で記録する事がほとんどないパラメータの値（MOD02 の時 65000、MOD09 の時 -10000、図2における黒い範囲）を入れてデータを作成する。ピース化処理は、このような不要な部分を削除してデータを管理するために、MODIS データを経緯度を元に格子状に区切る処理である。また、ピース化処理によって、AP は小さな単位で MODIS データを利用することが可能となる。ピース化処理のイメージを図2に示す。現在、ピース化処理の範囲は緯度

10~67 度、経度 100~167 度で 2 度単位に分割して作成中である。

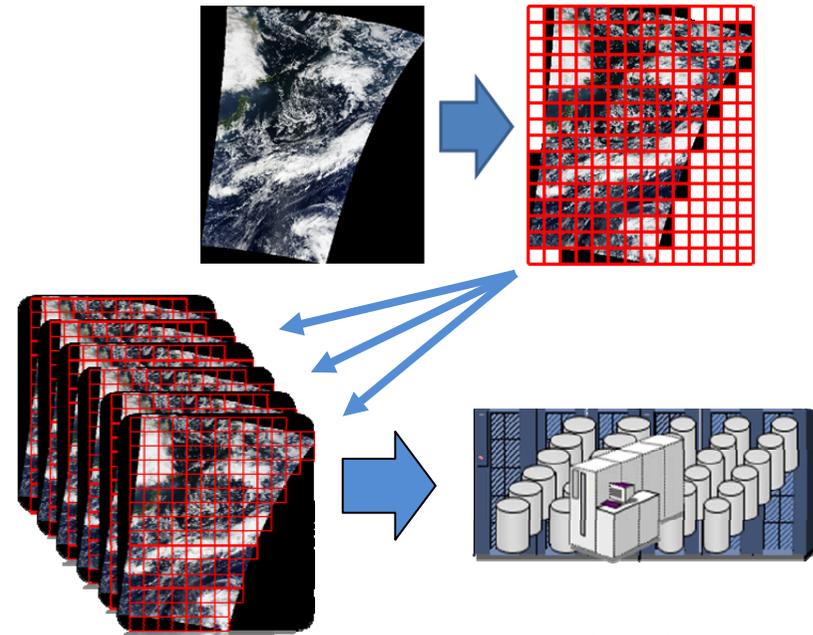


図2 ピース化処理の例

3.2 コンポジット処理

MODIS データのように広範囲を観測するデータを用いて地表面解析を行う場合、衛星軌道が異なるなどの観測条件の違いや雲などの自然現象の影響が問題となる。このような影響を排除して、有効なデータを選択するには、データのコンポジット処理が必要になる。コンポジット処理には、次の2つの場合がある。

- ① 同じ日に複数枚受信したデータを一つのデータに合成する
- ② 同じ場所で一定期間のデータの中から雲などの影響が少ない日時のデータを選択し、その期間の代表値としてデータを合成する

コンポジット処理のイメージを図3に示す。図3は、①の場合であり、この合成によって、衛星画像データの利便性が向上する。また、②の場合のコンポジット期間は、

研究目的や地域によって異なる。例えば、植生の季節別変化特徴を用いた自然変化などを解析する場合は、短いコンポジット期間が必要になる。

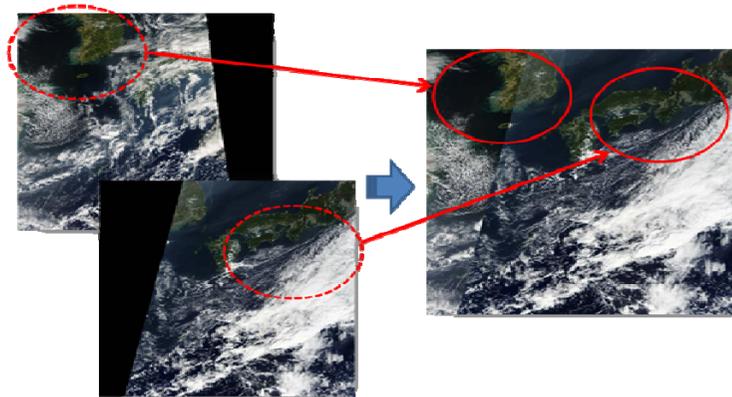


図3 コンポジット処理の例

4. アプリケーションの実行制御

4.1 実行制御の流れ

MODIS データを使用する解析処理は、データを大量に利用する。そこで、システム資源の有効利用や実行性能を向上させるため、LSM (Logical Storage Manager) と呼ぶ管理プログラムが PC クラスターの構成やローカルディスク内に格納されているデータを管理する。LSM は、衛星画像データの仮想化機能や PC クラスターのスケジューリング機能を実装しており、AP の実行で利用する PC やデータを仮想的に構成する。

4.2 データの仮想化

AP の実行イメージを図 4 に示す。AP が実行する PC は、AP が利用するクラスター台数やデータの日時・場所、及び各 PC のローカルディスクにキャッシングされているデータなどの管理情報から LSM によって動的に決定される (①)。一方、MODIS データは、LAN で接続されたグローバルディスクに格納されており、AP が実行する PC が決定した後に、ローカルディスク上のデータ、あるいはローカルディスク上にデー

タが無い場合にはローカルディスクにデータを転送して AP とデータを動的に結合する (②)。このように、AP で利用する MODIS データを仮想化することによって AP が実行する PC や AP が利用するデータの動的資源調整が可能となり、AP の実行におけるシステム資源の利便性が向上する。

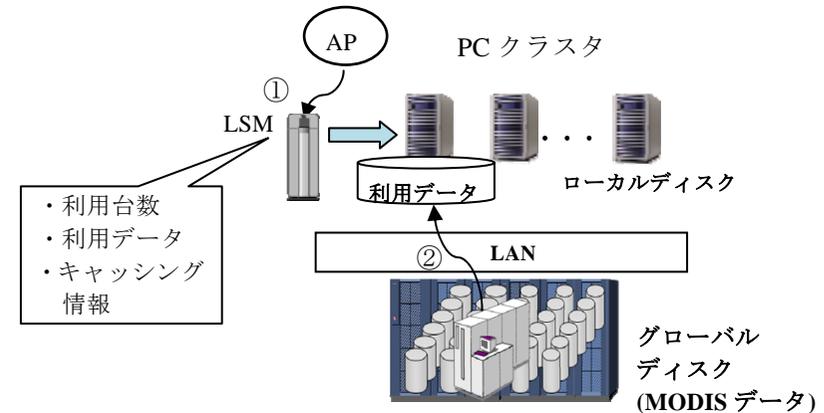


図4 データの仮想化

4.3 PC クラスターのスケジューリング機能

PC クラスターのスケジューリングは、PC クラスターの有効利用、及び PC 内のローカルディスクとグローバルディスク間のデータ転送の削減が目的である。スケジューリングの例を図 5 に示す。LSM は、PC クラスター全体を一元管理し、動作中の PC、各 PC のローカルディスク上のキャッシング状況などの情報を持っている。そして、それらの情報と AP が利用するデータの期間や場所を元にして、実行する PC の数や配置を決定する。この処理手順を次に示す。

- ① AP から利用する PC の台数や使用データが要求される。
- ② ①の内容から、AP が実行する計算ノード数を決定する。
- ③ LSM は、過去の実行状況からローカルディスク上に配置されているデータと AP が利用するデータを調べて、AP が実行する PC の構成を決定する。
- ④ AP と利用するデータを結合する。

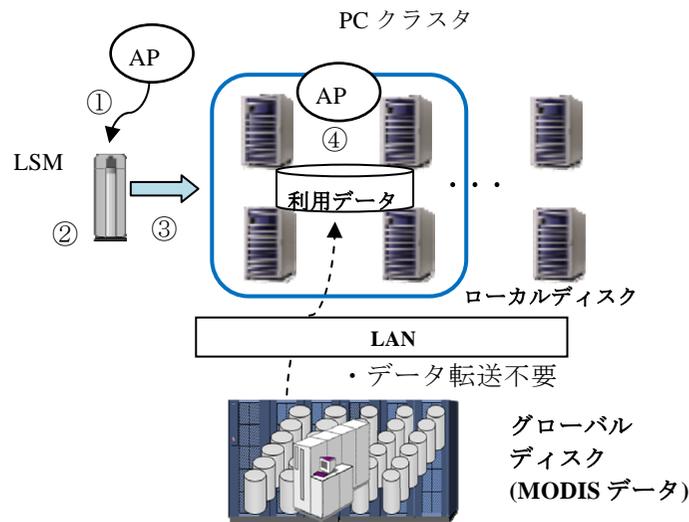


図5 PC クラスタのスケジューリング

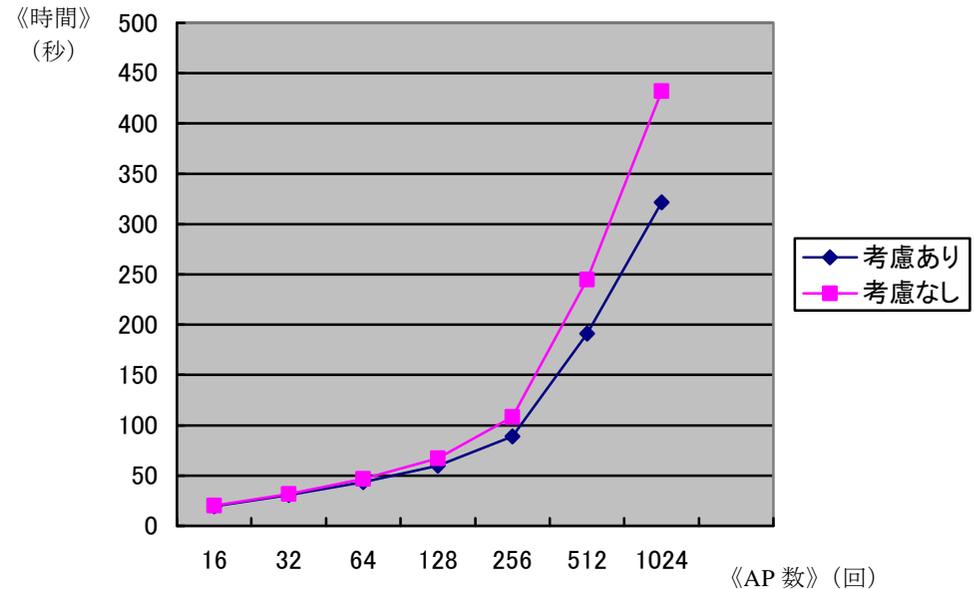


図6 性能評価

5. 性能評価

テスト用の AP を作成し、その AP の実行において、ローカルディスク上のデータの再利用するようにスケジューリングした場合と再利用を考慮しない場合の2つのパターンで性能を比較した。性能評価方法を次に示す。

- 同時に発行する AP を変化させ、発行した全ての AP が終了する時間を計測する。
- AP の使用するデータの日時、場所、期間は乱数で決める。
- 1 つの AP が使用する PC は 1 台であり、その AP が利用するデータは動作する PC 上のローカルディスクに保存して利用する。
- AP の処理内容は、データの入出力のみである。

6. まとめ

性能評価の結果を図6に示す。今回の評価では、ローカルディスク上のデータの再利用性を考慮することで、AP 起動数が 1024 で処理時間が約 72% 短縮したことが確認出来た。また、ローカルディスク上のデータの再利用性を考慮した場合、ローカルディスクのデータ再利用率（ヒット率）は AP の起動回数が増すほど上昇し、再利用率の平均は 31% であった。これにより、PC クラスタのスケジューリングは、AP 実行性能の向上に効果があると考えられる。実際に使用される衛生データは今回の評価で用いたデータよりも大きいので、更に有効に働くものと思われる。

今後の課題として、スケジューリングアルゴリズムの改良や計算ノードの追加、解析処理の実装を行い、より高速なシステム構築を行う研究を進める。

謝辞

本研究は、私立大学戦略的研究基盤形成支援採択事業東京情報大学 国際環境研究プロジェクト「アジア東岸域の環境圏と経済・社会圏の持続的発展のための総合研究」の補助を受けた。

参考文献

- 1) Eiji Nunohiro, Kei Katayama, Kenneth J. Mackin, Jong Geol Park: Development of Searching system for Fire Areas in Forests and Fields utilizing MODIS data, SCIS&ISIS, Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems Proceeding Index FR-H3-4, pp.1200~1203 (2006)
- 2) Eiji Nunohiro, Kei Katayama, Kenneth J. Mackin, Masanori Ohshiro, and Kazuko Yamasaki: Image Match Search System using Distributed Genetic Algorithm, SCIS&ISIS, Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems Proceeding Index FR-A5-1, pp.1536~1541 (2006)
- 3) Kei Katayama, Kenneth J. Mackin, Kotaro Matsushita, Eiji Nunohiro: Applying Brightness Information in Satellite Image Data Search using Distributed Genetic Algorithm, International Conference on Hybrid Information Technology(ICHIT06) IEEE Computer Society, Volume II, pp.84-89 (2006),
- 4) Eiji Nunohiro, Kei Katayama, Kenneth J. Mackin, Jong Geol Park: Forest and Field Fire Search System using MODIS Data, Journal of Advanced Computational Intelligence and Intelligent Informatics(JACIII), vol.11 No.8, pp.1043~1048 (2007)
- 5) 伊藤淳也, 中村彰宏, 朴鍾杰, 松下孝太郎, マッキン・ケネス・ジュームス, 大城正典, 山崎和子, 布広永示: 衛星画像データ解析システムの開発, 画像電子学会第 237 回研究会 in 広島 講演予稿, pp.85-88 (2008)
- 6) 伊藤淳也, 中村彰宏, 朴鍾杰, 松下孝太郎, マッキンケネスジュームス, 布広永示: Web に対応した衛星画像データ解析システムの開発, 画像電子学会 第 243 回研究会 in 鹿児島 講演予稿, pp.27-30 (2009)
- 7) Eiji Nunohiro, Jong Geol Park, Kenneth J. Mackin, Takashi Yamaguchi, Zeyu Zheng, Kotaro Matsushita, Yukio Yanagisawa, Masao Igarashi: Development of satellite data analysis system, The 5th International Conference on Information, ISBN 4-901329-06-5, Proceeding,143-146 (2009)
- 8) 中村彰宏, 朴鍾杰, 松下孝太郎, 布広永示: 衛星画像データ解析基盤の開発, 画像電子学会 第 250 回研究会講演予稿集, pp.15-18 (2010)
- 9) NASA MODIS
<http://modis.gsfc.nasa.gov/about/specifications.php>
- 10) JAXA
<http://www.eorc.jaxa.jp/hatoyama/experience/dataproc/p03.html>