

衛星画像データ解析システムの効率的利用に関する研究

森駿^{†1} 岩崎信也^{†2} 山口崇志^{†2} 花田真樹^{†2} 朴鍾杰^{†2} 布広永示^{†2}

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、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。

Study on Efficient Use of Satellite Image Data Analysis System

HAYAO MORI^{†1} SHINYA IWASAKI^{†2} TAKASHI YAMAGUCHI^{†2}
MASAKI HANADA^{†2} JONG GEOL PARK^{†2} EIJI NUNOHIRO^{†2}

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, 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 (SIDAS : Satellite Image Data Analysis System) 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、データ再利用機能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章で衛星データの概要、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再利用機能、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. 衛星画像データ解析システム

2.1 システム概要

SIDAS 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、それらの AP を研究開発中である。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。このため、AP の処理時間を短縮されるために、システム資源の有効利用や実行性能を向上を目的としたデータ再利用機能やスケジューリング機能を開発した。

2.2 衛星画像データ (MODIS データ)

MODIS は、NASA によって打ち上げられた地球観測衛星 Terra (午前に観測) と Aqua (午後観測) 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 μm) と 3 つの空間分解能 (250m, 500m, 1000m) を持っている。MODIS データのバンド情報を表 1 に示す。

東京情報大学では北海道、東京、沖縄の 3 つの場所で受信をしており、各受信所からは 1 日昼 3~4 回、夜 3~4 回受信する。3 か所から受信する範囲は東経 80 度から 180 度まで、北緯 0 度から 80 度である。

†1 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
†2 東京情報大学総合情報学部
Tokyo University of Information Sciences

表1 バンド情報

バンド	波長 (nm)	空間分解能	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

- MOD02 は、衛星データの可視化と植生指数の計算をするためのデータであり、7つのバンドで構成される。3つの可視域バンド（バンド1, 4, 3）は衛星データの可視化、近赤外バンド（バンド2）は植生指数（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）は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。
- MOD03（空間分解能 1000m）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。MOD03 は MOD02, MOD09, MOD11, MOD28 と MOD35 を地図投影法に合わせて変換する際の地図情報データである。
- MOD11（空間分解能 1000m）は土地被覆の放射率を用いて地表面温度を推定したデータである。都市域のヒートアイランド現象解析や土壌水分の状態解析や、ヒートフラックス解析に有効なデータである。
- MOD13（空間分解能 250m）は大気補正を行った植生指数データである。
- MOD28（空間分解能 1000m）は海面温度データである。海面温度の変化や海流の流れの解析に多く用いるデータである。

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自動抽出について研究する。
- (3) 自己回帰分析を用いた SST 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。衛星データを用いて水表面温度の分析をし地表面温度との
 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、AP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が利用しやすいようにコンポジット処理されたMODISデータを管理する。

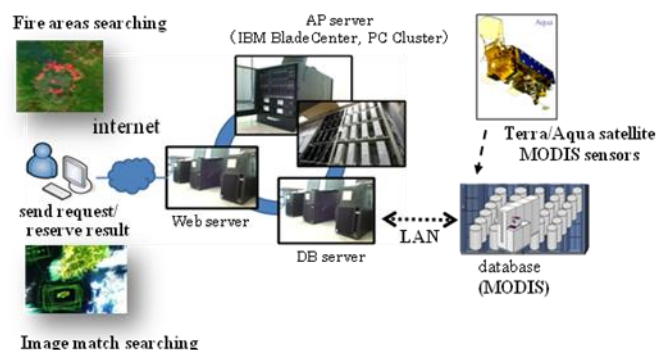


図1 システム構成

3. 衛星画像データ

3.1 データフォーマット

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バンド毎に Raw データに変換して保存・提供を行っている。Raw データは一般的な画像フォーマットに近いデータ形式であるがデータ部しか存在せず、データ部内は撮影データが等間隔・直列に並んでいる。位置情報や時間等の負荷情報は別ファイル(ヘッダファイル)に保存しておく必要があるが、受信した直後よりもデータ量は少ない。

3.2 コンポジット処理

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 データを加工して管理する。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。

3.3 モザイク処理

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、データのモザイク処理が必要になる。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といった処理を行う。

図2にモザイク処理の例を示す。

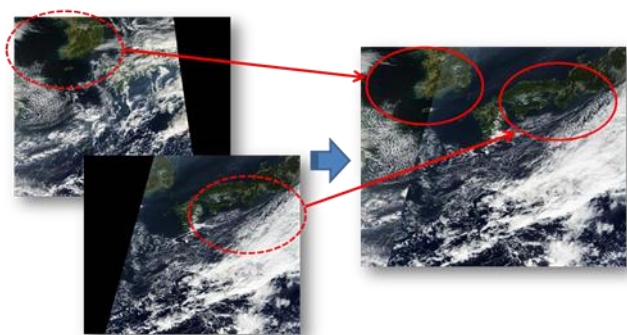


図 2 モザイク処理の例

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、PC クラスターのスケジューリング機能や衛星画像データの再利用機能を実装しており、AP の実行で利用する PC やデータの構成を決定する。

4.2 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 クラスター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4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。

AP とデータの結合を動的に行うことにより、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。この処理手順を次に示す。

- ① AP から利用する PC クラスターの台数や使用データが要求される。
- ② 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。

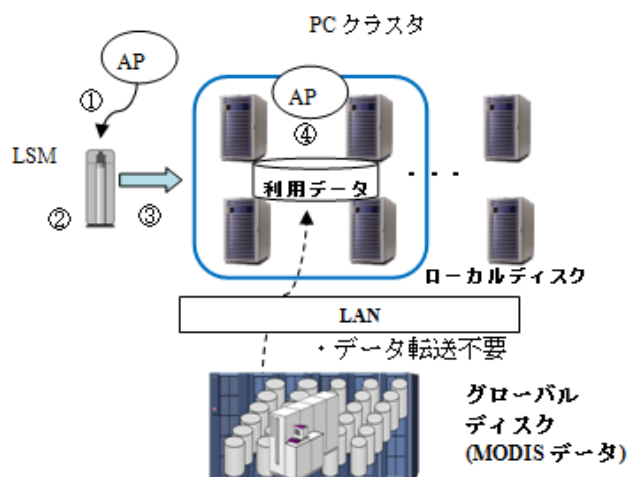


図 3 PC クラスターのスケジューリング

4.3 データの再利用機能

AP の実行イメージを図3に示す。AP が実行する PC クラスターは、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 クラスタ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 クラスタが決定した後に、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他の PC クラスタのローカルディスクにキャッシングされているデータ (②)、あるいはグローバルディスクからデータを転送して (③) AP とデータを結合する。このように、キャッシングされているデータを再利用することで 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が削減される。

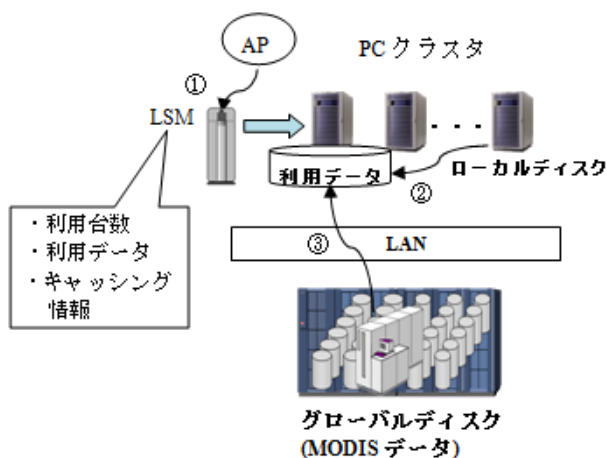


図 4 データの再利用

5. 性能評価

5.1 評価方法

データの再利用機能の効果を評価するための条件を次に示す。

- PC クラスタ台数：32 台
- PC クラスタで実行するタスク数：200 個
- タスクを LSM 内のキューに登録しておき、順次キューから取り出して実行する。
- スケジューリング機能ではクラスタ内の PC に保持されているデータの種類や配置状況、PC クラスタの空き状況などを調べ、タスクを割り当てる PC の場所と台数を決定する。
- 決定した複数の PC にタスクを割り振って実行する。

5.2 評価結果

図 5 に評価結果を示す。今回の評価では LSM からタスクを 200 個発行し、従来のタスクの実行が完了するごとのヒット率を評価した。処理を開始してからヒット率はタスク 30 までは急上昇し、その後ゆるやかに上昇することが

確認された。

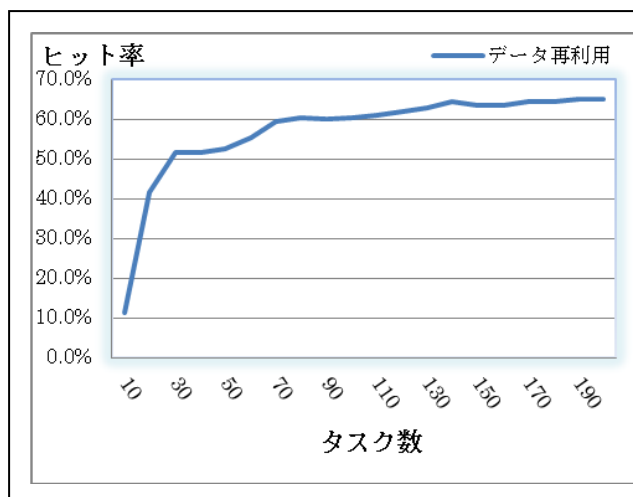


図 5 タスクの実行数とヒット率との関係

6. まとめ

今回は LSM のデータ再利用機能・スケジューリング機能の評価として、PC クラスタのローカルディスクにキャッシングされているデータのヒット率を計測した。

今後はスケジューリング機能のパターンの改善、再利用方法の改善を行い、更にヒット率を向上を実現することにより、システムの効率的利用を目指す。

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) Akihiro Nakamura, Jong Geol Park, Kenneth J. Mackin, Eiji Nunohiro, et al, "Development and Evaluation of Satellite Image Data Analysis Infrastructure", The Sixteenth International Symposium on Artificial Life and Robotics, Proceeding Index OS18-3, (AROB 16th' 2011)
- 2) Eiji Nunohiro, Jong Geol Park, Kenneth J. Mackin, et al, "Development of satellite data analysis system", Proceedings of the 5th International Conference on Information (2009), pp.143-146
- 3) 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 (JAC 3, 2007), vol.11 No.8, pp.1043~1048
- 4) Kotaro Matsushita, Kei Katayama, Kenneth J. Mackin, Eiji Nunohiro, "Development of Satellite Image Searching using Distributed Genetic Algorithm with Normalized Correlation", The twelfth International Symposium on Artificial Life and Robotics, Proceeding Index GS12-7, (AROB 12th' 2007)
- 5) National Aeronautics and Space Administration (NASA), MODIS Web, <http://modis.gsfc.nasa.gov/>