



HOKKAIDO UNIVERSITY

Title	知識情報処理を用いた衛星データの自動判読システムに関する研究
Author(s)	橋本, 秀太郎
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第11316号
Issue Date	2014-03-25
DOI	https://doi.org/10.14943/doctoral.k11316
Doc URL	https://hdl.handle.net/2115/55672
Type	doctoral thesis
File Information	Hashimoto_Shutaro.pdf



SSI-DT79115041

博士論文

知識情報処理を用いた衛星データの自動判読システムに関する研究

Development of an Automatic Satellite Data Interpretation System
using Knowledge Information Processing

橋本 秀太郎

2014 年 2 月

北海道大学 大学院情報科学研究科
システム情報科学専攻

本論文は北海道大学大学院情報科学研究科に
博士 (情報科学) 授与の要件として提出した博士論文である。

橋本 秀太郎

審査委員： 主査 田殿 武雄 客員教授

副査 小野里雅彦 教授

金子 俊一 教授

知識情報処理を用いた衛星データの自動判読システムに関する研究*

橋本 秀太郎

概要

衛星リモートセンシングによる観測は広域同時性、定期反復性、多波長性という特徴を持ち、その特性を活かして幅広い分野に応用されてきた。衛星リモートセンシングにおける近年の主要な動向には、取得情報量の増大、新たな解析手法・アプリケーションの増加、そして計算機・ネットワーク性能の向上が挙げられる。しかし衛星データの解析時には解析の目的や解析対象地域、解析日時を踏まえた衛星データの選定、補正処理、解析処理の選定、パラメータチューニングや教師データの取得などリモートセンシングの専門家による知識経験を活かした作業が不可欠となっており、前述の技術革新の恩恵が限定的となっているほか、衛星データの活用場面や解析規模を制限している。衛星データや解析手法等を効率的に管理・運用可能な自動判読システムへの期待はますます高まるものと予想される。

本論文ではこうした課題に資する提案として、光学衛星データ解析を対象とし、判読に用いられるデータやプロセスの蓄積共有、及びそれらを用いたクエリ駆動型の高精度自動判読システムの枠組みを構築し、将来の実利用への適用可能性を検討することを目的とする。提案する自動判読の枠組みはデータの選択から高次解析までの一連の解析ワークフローの自動化を目指すものであり、そのような研究はこれまでに無く、高い新規性を持つ。簡易な判読クエリの入力から解析の自動実行ができるようになることで、リモートセンシング技術に関する高度な知識経験を持っていない一般利用者が衛星データ解析を手軽に実行できるようになりデータ利用の拡大に繋がる事が期待されるほか、災害時被害状況把握における時間的ロスを削減でき、解析結果の迅速な提供にも寄与すると考えられる。

キーワード: リモートセンシング, 光学センサ, ALOS, AVNIR-2, 変化検出, 土砂災害, 洪水, 土地被覆分類, 知識情報処理, オントロジー, セマンティクス

*北海道大学 大学院情報科学研究科 システム情報科学専攻 博士論文, SSI-DT79115041, 2014 年 2 月 19 日.

Development of an Automatic Satellite Data Interpretation System using Knowledge Information Processing[†]

Shutaro Hashimoto

Abstract

Satellite remote sensing has been used in various applications because of its advantages in large area coverage, periodic and repetitive observation with multispectral bands. In recent years, there are three major trends of the remote sensing field: increase of the amount of available information, the growth of the number of the proposed methods and applications, and improvement of computing and networking environments. However, current practice analysis requires manual intervention by remote sensing experts. For example, technical knowledge and experiences are required in the selection of data and processes, parameter tuning, and acquisition of training data. Such manual intervention limits the use of satellite data and the scale of satellite data analysis. From the trends and the limitations mentioned above, there is a great demand for an automatic satellite data interpretation system.

The objective of this paper is to develop a framework of automatic satellite data interpretation system that can implement query-driven interpretation of optical satellite data using various data and processes. The proposed framework aims to implement the series of analysis workflow from the selection of data and processes to the execution of higher-order analysis. This leads to the expansion of the use of satellite data by users who do not have a technical knowledge and experience in remote sensing, and contributes to the rapid provision of the results of urgent analysis, such as disaster damage detection.

Keywords: remote sensing, optical sensor, ALOS, AVNIR-2, change detection, mud-slide, flood, LULC classification, knowledge information processing, ontology, semantics

[†]Doctoral Thesis, Division of Systems Science and Informatics, Graduate School of Information Science and Technology, Hokkaido University, SSI-DT79115041, February 19, 2014.

目次

第1章 序論	1
1.1 衛星リモートセンシングの意義と自動判読の必要性	1
1.2 光学衛星データ解析の概要	3
1.3 本研究の目的	6
1.4 ALOS および AVNIR-2 の概要	7
1.5 既往の関連研究に対する本研究の位置付け	8
1.6 本論文の構成	13
第2章 不確実性を考慮した情報表現・推論手法とそれを応用した変化検出システム	15
2.1 はじめに	15
2.2 不確実性を考慮した情報表現及び推論手法	16
2.3 変化検出システム	17
2.3.1 変化検出システムの概要	17
2.3.2 地物・地物変化のオントロジー	19
2.3.3 知識の表現	22
2.3.4 変化検出の流れ	23
2.4 本章のまとめ	26
第3章 光学衛星データの土地被覆推論の自動化・高精度化	27
3.1 はじめに	27
3.2 土地被覆分類の現状と課題	28
3.3 生成モデルを用いた確率的推論による土地被覆分類	30
3.4 カーネル密度推定を用いた高精度かつロバストなデータ密度分布推定 . . .	31

3.5	土地被覆の季節性を考慮した土地被覆分類	33
3.6	多時期土地被覆分類	35
3.7	本章のまとめ	36
第 4 章	データ及びプロセスの蓄積と解析ワークフローの自動構築	39
4.1	はじめに	39
4.2	自動判読システムの構成	41
4.3	自動判読システムのためのオントロジーの拡張	44
4.3.1	自動判読システムに必要なオントロジーの要件	44
4.3.2	時間オントロジー	45
4.3.3	幾何オントロジー	46
4.3.4	観測センサオントロジー	46
4.3.5	関係オントロジー	47
4.3.6	データオントロジー	49
4.3.7	プロセスオントロジー	52
4.4	データストアおよびナレッジストア	52
4.5	解析ワークフロー自動構築および自動実行のアルゴリズム	53
4.6	本章のまとめ	56
第 5 章	AVNIR-2 データを用いた災害被害域の推論	59
5.1	実験の概要	59
5.2	実験で利用した知識モジュール	59
5.3	事例 1：山口県における集中豪雨にともなう土石流跡の検出	61
5.3.1	解析事例	61
5.3.2	変化検出システムによる解析手順および解析結果	64
5.3.3	評価及び考察	65
5.4	事例 2：パキスタンにおける豪雨による洪水冠水域の検出	69
5.4.1	解析事例	69
5.4.2	変化検出システムによる解析手順および解析結果	70

5.4.3	評価および考察	72
5.5	本章のまとめ	76
第 6 章	AVNIR-2 データを用いた日本周域における土地被覆分類図作成と精度評価	79
6.1	実験の概要	79
6.2	使用データおよび解析エリア	79
6.3	日本周域における土地被覆分類図作成	80
6.4	精度評価を行う手法および精度評価指標	85
6.5	日本広域に分布した検証点を用いた精度検証結果および考察	87
6.6	本章のまとめ	91
第 7 章	多時期データと土地被覆解析に基づく変化抽出の自動実行	93
7.1	実験の概要	93
7.2	データストアおよびナレッジストアの構成	93
7.3	森林から裸地への変化検出の実行	102
7.4	本章のまとめ	112
第 8 章	結論	115
8.1	本論文の成果	115
	謝辞	117
	参考文献	119
	研究業績	127

目 次

1.1	本論文の全体像	7
2.1	変化検出システムの流れを示す IDEF0 図	18
2.2	地物・地物変化オントロジー	20
2.3	ベイジアンネットワークの構築	25
3.1	畑地の AVNIR-2 データの密度分布推定結果	34
4.1	提案するシステムアーキテクチャ	41
4.2	時間オントロジー	46
4.3	幾何オントロジー	47
4.4	観測センサオントロジー	48
4.5	関係オントロジー	49
4.6	データオントロジー	50
4.7	データ形式オントロジー	51
4.8	変量オントロジー	51
4.9	プロセスオントロジー	52
4.10	プログラムモジュールの有する機能	53
4.11	解析ワークフロー構築処理および自動実行のアルゴリズムを表す擬似コード	55
5.1	山口県における土石流発生前後の AVNIR-2 データ	62
5.2	山口県庁による山地災害発生箇所の調査結果	63
5.3	JAXA EORC による山口県の土石流跡の手動解析結果	63
5.4	構築された mudslide 推論のためのベイジアンネットワーク	64
5.5	山口県の土石流跡の推論結果 1	66

5.6	山口県の土石流跡の推論結果 2	67
5.7	山口県の土石流跡の推論結果 3 および土石流発生前後の AVNIR-2 データ (再掲)	68
5.8	土石流跡検出結果精度の事後確率のしきい値変化に対する変化	69
5.9	土石流跡検出結果と山口県庁による山地災害発生箇所の差分を表すカラー 合成画像	70
5.10	パキスタンにおける洪水発生前後の AVNIR-2 データ	71
5.11	JAXA EORC によるパキスタンの洪水冠水域の手動解析結果	71
5.12	構築された water layer appearance 推論のためのベイジアンネットワーク	72
5.13	パキスタンの洪水冠水域の推論結果	73
5.14	パキスタンの洪水冠水域の推論結果 2 および洪水発生前後の AVNIR - 2 データ (再掲)	74
5.15	洪水冠水域検出結果精度の事後確率のしきい値変化に対する変化	75
5.16	洪水冠水域検出結果と JAXA EORC による洪水冠水域抽出結果の差分を 表すカラー合成画像	76
6.1	日本周域における AVNIR-2 画像の通算取得数	80
6.2	月毎の日本周域における AVNIR-2 画像の取得数	81
6.3	土地被覆検証点データの空間分布	82
6.4	KPC-PROD による日本周域土地被覆分類図	84
6.5	KPC-PROD による日本周域土地被覆分類図における札幌市周辺の拡大図	84
6.6	つくば市近郊における土地被覆分類図の比較	86
6.7	多時期データ数の増加に対する KPC-PROD の分類精度変化	90
6.8	教師データ数の増加に対する各手法の分類精度変化	91
7.1	データストア内のディレクトリ／ファイルツリー	95
7.2	AVNIR-2 データのメタデータを記述した RDF グラフ	96
7.3	ナレッジストア内のディレクトリ／ファイルツリー	99
7.4	土地被覆変化検出プロセスのメタデータを記述した RDF グラフ	100

7.5	土地被覆分類プロセスのメタデータを記述した RDF グラフ	101
7.6	入力した解析クエリ (RDF グラフ)	103
7.7	構築された解析ワークフロー (RDF グラフ)	104
7.8	変化前の森林の土地被覆分類図を探索するクエリ (RDF グラフ)	107
7.9	変化後の裸地の土地被覆分類図を探索するクエリ (RDF グラフ)	107
7.10	サンプルデータを探索するクエリ (RDF グラフ)	108
7.11	マルチスペクトルセンサによる地表面反射率のリモートセンシング画像を 探索するクエリ (RDF グラフ)	108
7.12	マルチスペクトルセンサによる放射輝度のリモートセンシング画像を探索 するクエリ (RDF グラフ)	109
7.13	構築された解析ワークフローとその制御構造 (RDF グラフ)	110
7.14	自動判読の実行結果	113

表 目 次

1.1	ALOS 主要諸元	8
1.2	AVNIR-2 主要諸元	9
1.3	衛星データの汎用自動判読の関連研究と本研究の比較	12
6.1	土地被覆検証点データと観測値サンプルの要約	82
6.2	土地被覆分類の検証実験における実行環境	83
6.3	検証点データを用いた 10 分割交差検定による精度評価結果	88
6.4	検証点データを用いた 10 分割交差検定による精度評価における KPC-PROD の分類精度行列	89
7.1	自動判読の検証実験における実行環境	112

略語一覽

ALOS	Advanced Land Observing Satellite
API	Application Programming Interface
AVNIR-2	Advanced Visible and Near Infrared Radiometer type 2
CUDA	Compute Unified Device Architecture
DEM	Digital Elevation Model
DN	Digital Number
DSM	Digital Surface Model
DTM	Digital Terrain Model
EORC	Earth Observation Research Center
GCP	Ground Control Point
GIS	Geographic Information System
GPU	Graphics Processing Unit
GPGPU	General-Purpose computing on Graphics Processing Units
JAXA	Japan Aerospace Exploration Agency
KDE	Kernel Density Estimation
KPC	Kernel-based Probabilistic Classification
KPC-FREQ	Kernel-based Probabilistic Classification with Frequency Score
KPC-PROD	Kernel-based Probabilistic Classification with Product Score
KPC-SUM	Kernel-based Probabilistic Classification with Summation Score
LULC	Land Use and Land Cover
MLC	Maximum Likelihood Classification
MLE	Maximum Likelihood Estimation

MODIS	Moderate Resolution Imaging Spectroradiometer
NDSI	Normalized Difference Snow Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
OA	Overall Accuracy
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
PA	Producer's Accuracy
PCA	Principal Component Analysis
PolSAR	Polarimetric Synthetic Aperture Radar
PRISM	Panchromatic Remote-sensing Instrument for Stereo Mapping
RBF	Radial Basis Function
RESTEC	Remote Sensing Technology Center of Japan
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RMS	Root Mean Square
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
SVMs	Support Vector Machines
UA	User's Accuracy
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WKT	Well-Known Text
WWW	World Wide Web
XML	Extensible Markup Language

第1章 序論

1.1 衛星リモートセンシングの意義と自動判読の必要性

衛星リモートセンシングによる観測は広域同時性、定期反復性、多波長性という特徴を持つ。その特性を活かし、静止軌道上からの気象観測をはじめとして、環境変動の監視といった全球規模の陸域・海域・大気情報の収集から土地利用調査、農林水産業利用、災害時被害観測といった地域規模の詳細な情報収集まで、これまでも幅広い分野に応用されてきた。特に地震や豪雨等の大規模な自然災害発災直後においては現地での情報収集・伝達が極めて困難で、どこにどの程度の被害が発生しているか分からない状況となる。そのため現地の状況に依らず広域観測可能な衛星データの解析によって広範囲の被害状況を迅速に調査することが極めて有効であり、救援活動計画の立案やより高分解能のセンサによる観測計画立案の重要な役割を担っている。宇宙航空研究開発機構（JAXA）地球観測研究センター（EORC）では2006年の陸域観測技術衛星「だいち」（ALOS）の打ち上げ以来、国内外で280件以上の災害時緊急観測および解析実績があり、解析によって得られた情報は防災関係機関に提供され有効に役立てられてきた [1]。2011年3月11日に発生した東北地方太平洋沖地震に伴う東日本大震災直後においても発災翌日の3月12日から4月20日までにALOSにて総計643シーンの観測を実施し、これを用いて津波による被害状況把握および冠水域検出、土砂崩れ発生箇所の探索、海上流出がれきの把握、地殻変動検出等への利用が行われ、大きく貢献した [2]。

衛星リモートセンシングを取り巻く環境の近年の主要な動向には、取得情報量の増大、新たな解析手法・アプリケーションの増加、そして業務用と民生用の両方における計算機・ネットワーク性能の向上が挙げられる。近年の衛星リモートセンシングのハードウェア面、インフラ面での発展は著しく、衛星数の増加、センサの性能向上による分解能・観測範囲の向上、センサ種別の多様化が進み、取得した衛星データの蓄積もますます増え、解析に

利用可能な衛星データの情報量は爆発的に増加している [3]. また、衛星データの多種多様な解析手法が提案され続けており [4], 複数の種類の衛星データや地理空間情報を複合的に利用するアプリケーションも増加傾向にある [5]. 加えて近年は計算機技術やネットワーク技術の進歩も著しく、一般の利用者であっても巨大なサイズの衛星データを公開元からダウンロードし、民生用のコンピュータを用いて解析を行うことが容易に実現できるようになってきている.

しかし衛星データの解析時には解析の目的や解析対象地域、解析日時を踏まえた衛星データの選定、補正処理、解析処理の選定、パラメータチューニングや教師データの取得などリモートセンシングの専門家による知識経験を活かした作業が必要不可欠となっており、上記のような近年の技術革新の恩恵が限定的となっているほか、衛星データの活用場面や解析規模を大きく制限している. 即時性が求められる災害時被害状況解析においても人の判断が介在することによる時間的ロスのほか、土地鑑を有する人員と衛星データ解析が行える人員のギャップによる問題が度々発生している. 加えて前述した衛星データや解析手法、アプリケーションの多様化・蓄積が今後も進むであろうことも踏まえると、これらを効率的に管理し、運用していくことができるような自動判読システムへの期待はますます高まるものと予想される.

本論文ではこうした課題に資する提案として、光学衛星データ解析を対象とし、判読に用いられるデータやプロセスの蓄積共有、及びそれらを用いたクエリ駆動型の高精度自動判読システムの枠組みを構築し、将来の実利用への適用可能性を検討することを目的とする. 提案する自動判読の枠組みはデータの選択から高次解析までの一連の解析ワークフローの自動化を目指すものであり、そのような研究はこれまでに無く、高い新規性を持つものである. 簡易な判読クエリの入力のみを与えることでデータの選択から高次解析までの解析が自動実行できるようになり、その結果、リモートセンシング技術に関する高度な知識経験を持っていない一般利用者が衛星データ解析を手軽に実行できるようになり、データ利用の拡大に繋がる事が期待されるほか、例えば災害時被害状況把握において発生していた人の判断が介在することによる時間的ロスを削減でき、解析結果の迅速な提供にも寄与すると考えられる.

次節より、一般的な光学衛星データ解析の概要について紹介した後、光学衛星データの

自動判読を実現する上での課題について整理し、本研究の目的と本論文で取り扱う範囲について述べる。また本章の最後では本論文の2章以降の構成を概説する。

1.2 光学衛星データ解析の概要

リモートセンシングとは、“離れた所から対象物に直接手を触れないで、対象物を同定あるいは計測し、またその性質を分析する技術”と定義される [6]。その手段は対象物から反射または放射される電磁波を観測する方法が主に用いられる。このとき電磁波を観測する装置のことをセンサと呼び、センサを搭載する移動体をプラットフォームと呼ぶ。衛星リモートセンシングはプラットフォームとして人工衛星を用いるものである。今後、本論文中で単にリモートセンシングというときには、この衛星リモートセンシングを指すこととする。

リモートセンシングは用いられる電磁波の観測波長帯（バンド）によって、性質が大きく変わる。本論文で用いる光学衛星データは可視光線と赤外線を光学センサによって観測したものであり、個々の観測値を画素値として記録したデジタル画像形式で得られる。光学衛星データの特徴は、波長が短い光を用いるために高空間分解能化がしやすいこと、雲の粒子により反射するため被雲によって地表面の様子を観測できなくなることなどが挙げられる。

主に波長 $0.4 \sim 1.3 \mu\text{m}$ の可視域・近赤外域バンドによる観測は可視・反射赤外リモートセンシングと呼ばれ、太陽を放射源とし、地表面で反射された電磁波を観測する。太陽は波長約 $0.5 \mu\text{m}$ をピークとする電磁波を放射しており、地球に到達する電磁波の性質は季節等によりわずかな変動はあるが、常に同質と考えることができる。その電磁波が地表にて反射され、衛星に搭載されたセンサに入射される。その電磁波の放射輝度（ある単位面積の面から放射される、あるいはその面を通過する、単位立体角、単位時間あたりの放射エネルギー）を観測する。従って、得られる観測値は地表の反射率に大きく依存し、観測値から地表の反射率を知ること、地表の情報を得ることができる。反射率は、ある面への入射光束に対する反射光束の比率と定義される。この反射率は定義上、電磁波の波長帯や入射・反射方向が明確ではなく、使用される場面によって変化する。微小波長帯ごと

の反射率は分光反射率、または反射スペクトルと呼ばれ、波長に対する関数として表される。分光反射率特性は物体によって異なり、マルチスペクトルセンサ（バンドを複数持つセンサ）が物体を識別できる原理は分光反射率特性の差異によるものであるほか、われわれが物体の色情報を認識することができるのも、これが理由である。

また、主に波長 $8 \sim 14 \mu\text{m}$ の熱赤外域バンドによる観測は熱赤外リモートセンシングと呼ばれる。可視・反射赤外リモートセンシングで観測する電磁波の放射源が太陽であったのに対し、熱赤外リモートセンシングで観測する電磁波の放射源は地表面である。常温の地表面は波長約 $10 \mu\text{m}$ をピークとする電磁波を放射しており、熱赤外域においては太陽放射の反射光より地表面放射が支配的である。すなわちこの波長域における観測では、地表面の熱放射を観測していることになる。

リモートセンシングではセンサの種類によって得られるデータの性質が決まるため、衛星データを解析する場合にはセンサやそのセンサが搭載されているプラットフォームについて熟知することが必要となる。センサの特性は主に下記の3つがある。

1. 保持するバンドの性質である分光特性（波長域、応答関数など）
2. 観測される放射輝度の特徴を決定する放射量特性（信号とノイズの比である S/N 比、観測可能な最大・最小放射輝度の幅であるダイナミックレンジ、そのレンジを何段階の離散値で表現するかを決める量子化レベルなど）
3. 画像の幾何的特徴を決定する幾何学特性（得られるデータの観測幅を決める視野角、地上分解能を決める瞬間視野角など）

また、プラットフォームの特性は主に下記の3つがある。

1. 地上分解能や観測幅を決める衛星高度
2. 幾何的な歪みに影響を与える位置姿勢精度
3. データの取得頻度や観測地域を決める軌道要素

これらの特性を考慮し、判読目的の条件に合致したセンサによるデータを選択することが重要である。

衛星データから得られる情報には、空間情報（対象物の空間的な広がりや分布に関する情報）、スペクトル情報（観測された電磁波の波長方向の強度に関する情報）、時間情報（対象物の変化に関する情報）の3つの基本要素が含まれており、これらの情報を最大限に活用して利用者の目的に応じた役立つ情報に加工することが、衛星データ解析の目的となる [7]。こうした情報を組み合わせて評価することにより、特定の時期における植生分布、土地被覆状況、森林・農作物等の生育状況、岩石・鉱物資源分布が把握できる [7]。また、植生状況の変動、氷河・雪氷の変動、土地被覆変化など長期間にわたる地表面の変化、災害時の被災地の抽出、復興状況などを把握することが可能となる [7]。

最後に、衛星データの一般的な判読の流れについて説明する。衛星データはデータ提供側で施された処理に応じて様々なレベルのプロダクトとして得られる。プロダクトの種類によっては衛星データの解析を行う前処理として、放射量補正、バルク幾何補正、精密幾何補正、大気補正、オルソ補正といった補正処理を施す必要がある。このうち放射量補正及びバルク幾何補正は系統補正と呼ばれ、補正のためにはセンサに関する技術的なデータが必要となるため、通常は系統補正が施されたプロダクトが提供される。それぞれの補正処理についての簡単な説明を下記に示す。

放射量補正

センサに入射した電磁波の強度を量子化して生成された DN 値を、物理量である放射輝度に変換する処理。

バルク幾何補正

バルク幾何補正とは衛星の位置-姿勢-地球の幾何学的モデル、及びセンサの幾何学的特性のみの情報を用いてデータに含まれる幾何学的な歪みを補正し、地図座標系へと簡易に変換する処理。

精密幾何補正

地上基準点（GCP）の緯度経度で表される位置情報を用いて地図座標系へと精密に変換する処理。

大気補正

大気の状態を推定することで大気中の分子やエアロゾルによる散乱・吸収・放射の

影響を除去し、地表面反射率へと変換する処理。

オルソ補正

センサ位置からの中心投影により得られた画像に対し、地表面の高低差を考慮しつつ正射投影の画像へと変換する処理。

補正処理の後、判読目的に応じて様々な高次解析処理が行われる。

1.3 本研究の目的

本論文の目的は、光学衛星データ解析を対象とし、判読に用いられるデータやプロセスの蓄積共有、及びそれらを用いたクエリ駆動型の高精度自動判読システムの枠組みを構築し、将来の実利用への適用可能性を検討することである。Fig. 1.1 は本論文で提案する自動判読システムの枠組みの全体像と、取り組むべき研究課題を表している。提案システムでは、利用者は解析結果として得たいデータを表す解析クエリを入力し、特に高度な知識経験を要しない簡単な操作を行うだけで目的の解析結果を得ることが可能となる。システムは衛星データや地理データ等のデータが置かれたデータストア、補正処理や解析処理等のプロセスモジュールが置かれたナレッジストアを持ち、それらを用いて利用者が入力した解析クエリを出力するための解析ワークフローを自動構築し、解析を行う。

本論文ではまず陸域観測技術衛星「だいち」(ALOS) 搭載の可視・近赤外域の観測波長帯を持つ光学センサである高性能可視近赤外放射計 2 型 (AVNIR-2) によるデータを対象とし、高次解析処理における主要なプロセスである土地被覆推論及び地物・地物変化推論のそれぞれについて自動化を行うための検討を行う。

次に様々なデータ及びプロセスの記述・蓄積方法と、それらを選択し組み合わせることによりクエリとして入力された目的の解析結果を生成するための解析ワークフローの自動構築処理に関する検討を行う。ここでは、対象とする衛星データを光学衛星データに限らず、様々な種類の衛星データを扱うことのできる枠組みの構築を行う。

最後に、以上の研究内容を組み合わせて光学衛星データの自動判読システムを構築して実証を行う。

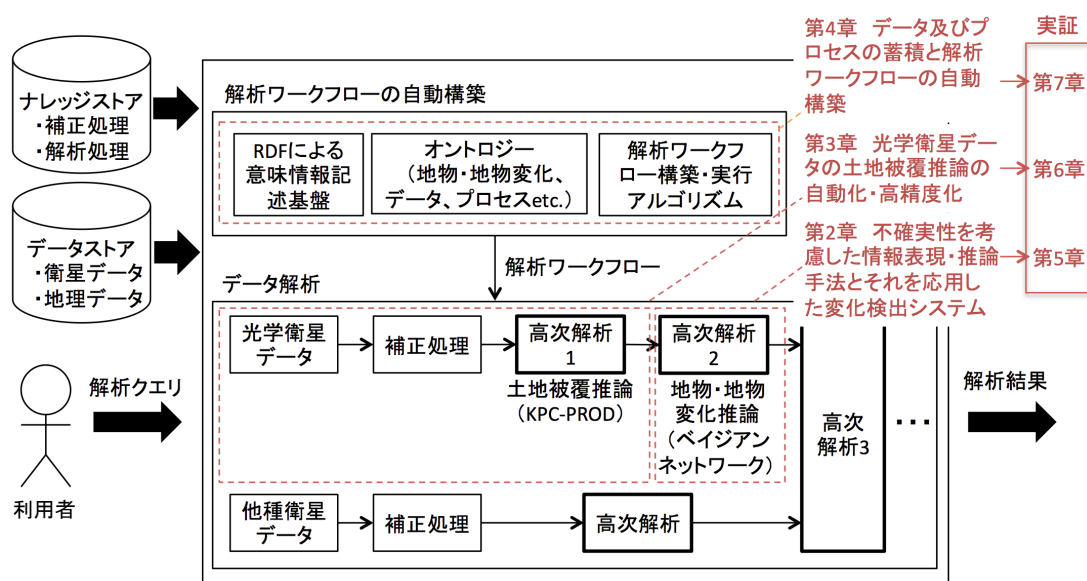


Fig. 1.1: 本論文の全体像

1.4 ALOS および AVNIR-2 の概要

本節では、本論文で主に利用される衛星データに関する情報として、観測センサが搭載された衛星である陸域観測技術衛星「だいち」(ALOS)と、観測センサである高性能可視近赤外放射計2型(AVNIR-2)についての概要を述べる。Table 1.1 は ALOS の主要諸元を、Table 1.2 は AVNIR-2 の主要諸元をそれぞれまとめたものである。

ALOS は地図作成、地域観測、災害状況把握、資源探査等への貢献を図ることを目的として JAXA が開発・運用した地球観測衛星である [8]。ALOS は、高精度で標高抽出を行うためのパンクロマチック立体視センサ (PRISM)、土地被覆の観測を高精度に行うための AVNIR-2、および昼夜や天候によらず陸域の観測が可能なフェーズドアレイ方式 L バンド合成開口レーダ (PALSAR) の3つの観測センサを搭載し、高分解能かつ高観測幅の陸域観測を得意とする。ALOS は 2006 年 1 月 24 日に打ち上げられ、2011 年 4 月 22 日の電力低下による機能停止までの 5 年 3 ヶ月の間継続的に運用され、その間に全球規模の膨大な量の観測データの取得および数々の技術検証が行われ、プロジェクトの成功基準として設定されたエクストラサクセス基準を達成した [9]。

AVNIR-2 は可視域 3 バンド、近赤外域 1 バンドの計 4 バンドの観測波長帯を持つマルチスペクトルセンサで、直下視にて地上分解能 10 m、観測幅 70 km での広域高分解能観測を

Table 1.1: ALOS 主要諸元 [8][9]

運用情報	打ち上げ日	2006 年 1 月 24 日
	打ち上げ機	H-IIA 8 号機
	打ち上げ場所	種子島宇宙センター
	設計寿命	3 年以上 (5 年目標)
	機能停止日	2011 年 4 月 22 日
	運用終了日	2011 年 5 月 12 日
軌道	軌道種類	太陽同期準回帰軌道
	降交点通過地方時	午前 10 時 30 分 ± 15 分 (赤道上)
	軌道高度	691.65 km (赤道上)
	軌道傾斜角	98.16 度
	軌道周期	98.7 分
	周回数	14+27/46 周/日
	回帰日数	46 日 (サブサイクル 2 日)
	軌道回帰精度	± 2.5 km (赤道上)
	隣接軌道間距離	59.7 km (赤道上)
搭載センサ	パナクロマチック立体視センサ (PRISM)	
	高性能可視近赤外放射計 2 型 (AVNIR-2)	
	フェーズドアレイ方式 L バンド合成開口レーダ (PALSAR)	

行なうことができる。ALOS は回帰日数 46 日の太陽同期準回帰軌道で周回しており、同一観測パスを年間最大約 8 回の頻度で観測している。また、ALOS の赤道上での隣接軌道間距離は 59.7 km であり、AVNIR-2 の観測範囲のオーバーラップが北緯 25 度で約 16 km、北緯 45 度で約 28 km の幅で生じており、これらの領域は年間最大約 16 回の頻度で観測が行われる。

1.5 既往の関連研究に対する本研究の位置付け

本論文で扱う光学衛星データの自動判読の要件は、様々なデータ、様々な解析手法、様々なアプリケーションに対応可能であるという汎用性、および日々行われる追加・更新にも対応していくことのできる柔軟性を有することである。そのような汎用性、柔軟性を実現可能な手段として、知識情報処理システムがある。

知識情報処理システムは機械可読な形式で表現された問題解決のための知識を知識ベースに蓄積しておき、知識を用いた推論を行うことで問題解決を行うシステムであり、特に

Table 1.2: AVNIR-2 主要諸元 [8]

観測波長帯	バンド 1: 0.42–0.50 μm
	バンド 2: 0.52–0.60 μm
	バンド 3: 0.61–0.69 μm
	バンド 4: 0.76–0.89 μm
地上分解能（瞬時視野角）	10 m（約 14.28 μrad ）
観測幅	70 km 以上（約 5.8 deg）
ディテクタ数	7,100 個/バンド
走査周期	1.48 msec
ポインティング角	± 44 deg
量子化ビット数	8 ビット/画素

専門家の知識を知識ベースに蓄積する知識情報処理システムはエキスパートシステムと呼ばれる。エキスパートシステムをはじめとした知識情報処理システムの大きな特徴は知識ベースと推論機構が分離された構造となっている点にある [10]。知識情報処理システムでは知識がプログラム中に埋め込まれず知識ベースとして外部に記述されることでシステム開発後の修正が容易となり、汎用的なアルゴリズム化が困難でシステム開発後に多くの修正が予想される悪構造問題（ill-structured problem）に対して適している。光学衛星データや航空写真の自動判読は悪構造問題であり、これまでに提案されている汎用性を持った自動判読を行うシステムの多くは知識情報処理システムである。

知識情報処理システムを利用した先駆的な研究として、松山による航空写真判読システムがある [11]。この研究ではエキスパートシステムを採用し、空間的情報に関する知識を IF-THEN ルール形式で表現・蓄積し、知識を用いた推論を行うことで航空機や家といった地物の認識を行なっている。

その後、典型的なエキスパートシステムは知識の意味的な前提や知識記述の指針が暗黙的であるため、知識の共有・再利用が阻害されてしまうことや既存の知識ベースに新たに知識を追加していくことが難しく、知識ベースを常にゼロから構築しなければならないという問題点が Crevier ら [12] や溝口 [13] によって指摘されており、これを解決するために合意の得られた概念体系であるオントロジーを知識記述の基盤とすることが提唱されている。そのため、近年提案されている汎用自動判読を目指したシステムの多くはオントロジーを利用したものとなっている。オントロジーを採用する利点には主に次の 4 点が挙げ

られる。

1. 定義の一貫性が保証された概念を知識のビルディング・ブロックとして用いることができる
2. 概念定義が分離されるため知識記述が容易になる
3. 記述された知識の意味的な解釈を形式的に行えるようになる
4. どのような観点から概念を捉えるかが規定されるため知識記述の指針となる

衛星データの汎用自動解析を試みた研究には、Datcu らによる人間中心のインタラクティブな解析システム [14][15][16], Durbha らによる意味ベースでの画像検索システム [17][18][19], Moodley らによるエージェント間でやり取りされるデータの意味を規定するオントロジーを用いたマルチエージェントシステム開発のためのフレームワーク [20][21][22][23][24] がある。

Datcu らは専門家が行ったデータ融合 (data fusion) のプロセスを学習して利用する衛星データ自動判読システムを提案している [16]。このシステムではまず、専門家がシステムを用いてインタラクティブに解析を行う。専門家は複数種のセンサの衛星データから画像特徴 (e.g. PolSAR L-バンドの偏波シグネチャやハイパースペクトルシグネチャの第一主成分) を算出し、それらを用いて教師付き分類を行い、目的の情報を抽出する。この間にシステムは、解析結果に対して貢献度の高いセンサおよび画像特徴の組を学習し、知識として蓄える。この知識を用いることで、別の事例に対して自動判読を行う。さらに Datcu らは解析結果の信頼度を高めるために、異なるドメインにて得られた知識をオントロジーを利用して統合する手法を提案している。その例として地雷原検出のアプリケーションを紹介しており、地雷が埋まっていやすい地形に関するドメイン知識と、地雷を設置する場所に関するドメイン知識を統合して危険度マップを生成している。

Durbha らは衛星データアーカイブからの意味ベースでの検索を目的として I³KR (Intelligent Interactive Image Knowledge Retrieval) という枠組みを提案している [17]。この枠組みではまず、衛星データを教師なしセグメンテーションによって同質の画像小領域に分割し、各領域に対して色やテクスチャ、形状に基づく特徴量を計算する。その後、特

微量空間でのカーネル主成分分析 (Kernel PCA) とサポートベクターマシン (SVM) を組み合わせた分類手法により, オントロジーで定義された地物概念との対応付けを行う. 提案手法の例として, 提案した枠組みをもとに MODIS データや Landsat データ中から沿岸部における地物を意味ベースで検索するシステムを構築している.

Moodley らは知識駆動型のセンサーウェブ (Sensor Web) の実現を目的とし, SWAP (Sensor Web Agent Platform) という地球観測アプリケーションのためのアプリケーションインフラストラクチャを提案している [20]. SWAP はそれぞれが固有の役割を持つ複数のエージェントプロセスから構成され, エージェントを組み合わせた様々なアプリケーションを構築することができる. 観測に関する概念やエージェントに関する概念, データに関する概念等のシステム内で登場する概念を定義した SWAP オントロジーを構築し, システム内での意味情報共有基盤として利用している.

Table 1.3 はそれぞれの研究の特徴をまとめたもので, システムの入力には衛星データ以外に必要な入力情報や出力可能な情報に加え, 蓄積されるデータやプロセスの拡張性, そして一連の解析ワークフロー構築の自動化が行えるかという観点で整理した. また同時に, 本研究で提案する枠組みについても同じ観点からの特徴を掲載する.

まずデータの拡張性に関してみると, Datcu らや Durbha らのシステムは衛星データをデータベースに蓄積しておき, それらの観測日時や観測範囲を見ながら自動で選択できるようにはなっているものの, MODIS データや Landsat データというように予め指定されたセンサの衛星データのみしか利用することができない. それに対して Moodley らのシステム, および本研究による提案システムでは MODIS データや Landsat データというような個々の衛星データから光学衛星データのようなセンサの種類を読み取り, 光学衛星データを入力とするアプリケーションで同様に利用できる仕組みを備えたアプローチを取り, 共有性や再利用性を持たせたデータ蓄積が可能である.

次にプロセスの拡張性に関してみると, Datcu らや Durbha らのシステムはそれぞれ固有の解析プロセスのみを行うものであり, 様々な解析プロセスを蓄積し, 選択して解析に利用することは想定されていない. それに対して Moodley らのシステムは入出力データの種類の具体的な指定されたプロセスを蓄積し, 解析に利用することが可能となっている. さらに本研究による提案システムでは入出力データの種類の抽象的に指定されたプロセス

Table 1.3: 衛星データの汎用自動判読の関連研究と本研究の比較

研究	システムの入力	システムの出力	データの拡張性	プロセスの拡張性	解析ワークフロー構築の自動化
Datcu 2002, 2003, 2005	・アプリケーションの指定	・分類によって得られる解析結果	× ・予め指定されたセンサの衛星データのみ ・集中管理のみ	× ・各種画像特徴をもとに分類を行うプロセスのみ	× ・アプリケーション毎、地域毎に教師付き学習が必要
Durbha 2005, 2008, 2009	・探索したい地物	・指定した地物が含まれる画像（解析地域は限定）	× ・予め指定されたセンサの衛星データのみ ・分散管理可能	× ・画像のセグメンテーション後、小領域に対して各種領域特徴をもとに地物分類を行うプロセスのみ	× ・地域毎に教師付き学習が必要
Moodley 2006, 2011	・解析対象の衛星データの観測時間と場所 ・アプリケーションの指定	・高次解析結果	○ ・置かれた衛星データからセンサの種類を読み取ることが可能 ・分散管理可能	△ ・具体的に入出力が指定されたプロセスを利用可能 ・分散管理可能	× ・手動でプロセスを組み合わせた解析ワークフローの構築が必要
本研究	・解析結果を表すクエリ	・高次解析結果	○ ・置かれた衛星データからセンサの種類を読み取ることが可能 ・分散管理可能	○ ・抽象的に入出力が指定されたプロセスを利用可能 ・分散管理可能	○ ・自動でプロセスを組み合わせた解析ワークフローを構築可能 ・用意された教師データから自動で教師付き学習が可能

を蓄積・利用することができる。その結果、特定の問題に対して適用できるプロセスだけではなく、幅広い問題に対して汎用的に適用することができるような共有性や再利用性の高いプロセスを記述することができる。

最後に解析ワークフロー構築の自動化に関してみると、Datcu らや Durbha らのシステムは先に述べたとおりそれぞれ固有の解析プロセスのみからなる解析ワークフローを実行するものであり、また事前に専門家が解析地域において教師付き学習を行っておく必要がある。また Moodley らのシステムは判読を行うためのアプリケーションを事例ごとに手動で構築して与える必要がある。それに対して本研究による提案システムは、複数の汎用性を持ったプロセスモジュールを機械可読な形で用意しておき、クエリで指定した解析結果を得るための解析ワークフローをプロセスモジュールを組み合わせることで自動構築するというアプローチを取っている。加えて教師データを用意しておくことで、そこから解析の条件にあった教師データのみを自動で選択した教師付き学習を行うことができ、人の手を介すること無く解析ワークフローを実行することが可能となっている。

以上のことをまとめると、本研究で提案するシステムは機械可読に蓄積されたデータやプロセスを利用することでデータの選択から高次解析までの一連の解析ワークフローの自動化を行うものであり、蓄積されたデータやプロセスの共有性や再利用性の面で既存の関連研究に対する優位性を持つものと考えられる。

1.6 本論文の構成

本論文の構成は以下に記す通りである。

本論文の2章及び3章では、光学衛星データである AVNIR-2 データを対象とし、高次解析処理における主要なプロセスの自動化について検討する。高次解析処理として土地被覆推論及び地物・地物変化推論のそれぞれについて自動化を行うための検討を行う。

2章では光学衛星データを用いた地物・地物変化推論手法を提案し、それを応用することで衛星データの災害状況把握解析における専門家の判断処理／作業を軽減するための自動変化検出システムを提案する。このシステムは、利用者が推論したい地物・地物変化の名称を入力すると、その地物・地物変化の意味を調べ、知識ベースに蓄積されている単純な情報抽出や推論を行うための知識モジュールを組み合わせ自動的にその地物・地物変化の推論を行う。地物・地物変化の意味を調べるための辞書として、本研究では地物・地物変化のオントロジーを構築し、システムで利用する。また、複数の知識モジュールによって得られた情報を統一的に扱い推論を行う仕組みとして、ベイズ確率による個々の情報の表現と、オントロジーから自動構築したベイジアンネットワークによる推論手法を採用する。

3章では光学衛星データを用いた土地被覆推論の自動化・高精度化を行う手法をそれぞれ提案する。土地被覆推論の自動化・高精度化のために、データベースに蓄えられた大量の土地被覆教師データを、特定の種類の土地被覆に依存せず、効果的に活用できる土地被覆推論手法を新たに提案する。提案手法はデータ分布に対する仮定を設けないノンパラメトリックな密度推定手法であるカーネル密度推定を採用し、大量の教師データがつくる分布を利用した高精度な確率的推論を行う。また、観測された地表面反射率に加えて年周期性を連続的に表現できる観測日変量を入力データとすることにより土地被覆クラスの季節性を考慮した土地被覆推論を行う手法、及び多時期の観測データを用いて高精度な土地被覆推論を行うための拡張について提案する。

4章では様々なデータ及びプロセスの記述・蓄積方法と、それらを選択し組み合わせることにより、目的の解析結果を生成するための解析ワークフローの自動構築処理に関する検討を行う。ここでの検討では、対象とする衛星データを光学衛星データに限らず、様々な種類の衛星データを扱うことのできる枠組みの構築を行う。データやプロセスを様々な

解析事例に対して汎用的に利用できるよう問題非依存に記述・蓄積するため、データやプロセスの意味をメタデータとして明示的かつ機械可読に記述することを目的とし、RDFによるメタデータ記述の形式的枠組みと、システム内でのデータやプロセスの定義や持つべき属性を規定した独自のオントロジーを採用する。また、メタデータが付記されたデータやプロセスが蓄積されたデータストア及びナレッジストアを利用し、目的の解析結果を出力できる解析ワークフローを、データストアにあるデータ及びナレッジストアにあるプロセスの出力データを意味を考慮しながら探索・マッチングしていくことで構築する処理を提案する。

5章、6章、および7章ではそれぞれ2章、3章、4章で提案した手法についての検証を行う。まず2章および3章における提案手法について個別に精度検証を行い、高精度の結果が得られる事を示す。その後、2章および3章における提案手法を要素技術とし、4章で提案した枠組みに基づき構築したシステムによる自動判読の実行例を示し、目的の判読結果が自動的に生成される事を示す。

最後に、8章にて本論文の結論を述べる。

第2章 不確実性を考慮した情報表現・推論手法とそれを応用した変化検出システム

2.1 はじめに

高次解析において、衛星データのシーン中における地物の有無や地物変化の発生の有無といった情報を得るには、観測データの値そのもの、もしくは土地被覆情報といった比較的プリミティブな情報を総合的に考慮し、推論する必要がある。現実的には、推論結果は証拠となる情報が与えられると一意に決定できるのではなく、経験的あるいは統計的に作成された推論ルールを用いることによる不確実性や、証拠となる情報そのものが持つ不確実性を踏まえる必要がある。

そこで本章では、情報の不確実性を考慮しつつ、複数のプリミティブな情報からより複雑な情報である地物・地物変化情報を推論する手法を提案する。不確実性を持つ情報を統一的に扱い、それらをもとにした推論を行う仕組みとして、ベイズ確率による個々の情報の表現方法、並びにベイジアンネットワークによる推論手法を採用する。

加えて本章では、衛星データの災害状況把握解析における専門家の判断処理／作業を軽減するための自動変化検出システムの枠組みも提案し、後の章における検証実験にて提案システムを用いた精度評価を行う。本章で提案するシステムでは、利用者が推論したい地物・地物変化の名称を入力すると、システムはその地物・地物変化の意味を調べ、知識ベースに蓄積されている単純な情報抽出や推論を行うための知識モジュールを組み合わせることで自動的にその地物・地物変化の推論を行う。地物・地物変化の意味を調べるための辞書として地物・地物変化のオントロジーを構築し、システムで利用する。

本章の構成は以下の通りである。

2.2節では、不確実性を持つ複数の情報をもとにした推論を行う仕組みとして採用した、ベイズ確率による個々の情報の表現方法、並びにベイジアンネットワークによる推論手法

について説明する.

2.3 節では, 提案手法を応用して構築した変化検出システムの構成や動作を示す. 目的の変化を衛星データから検出するために使用される知識モジュールをどのように選択し, 組み合わせて情報抽出・推論を行い, 変化検出結果を得るかを説明する.

本章で構築した変化検出システムの検証実験結果は5章で言及する.

本章の提案内容及び5章で示す検証実験結果は著者らによる投稿論文[25]にて報告したものである.

2.2 不確実性を考慮した情報表現及び推論手法

不確実性を伴った情報を表現するためにベイズ確率を採用し, 個々の情報の信頼度を区間 $[0, 1]$ の確率値 (1 に近いほど信頼できることを表す) で記述する. ベイズ確率は頻度の表現である客観確率に加えて信頼度の表現である主観確率も含めた広義の確率解釈である[26]. ベイズ確率には古典的確率の理論体系をそのまま適用することができるため, 知識や情報の不確実性表現とその演算に適している.

また, ベイズ確率で記述された複数の情報をもとに別の情報を推論する仕組みとして, ベイジアンネットワークを用いた推論を採用する. ベイジアンネットワークは確率的な因果関係を有向グラフによって表現したものであり, Pearl によって提案され, 理論体系がまとめられた[27][28]. ベイジアンネットワークの定義を以下に示す.

1. 非循環有向グラフである.
2. グラフを構成するノードはそれぞれある確率変数を表す.
3. グラフを構成するアークは両端のノードで表現された確率変数の間にアークの方向に沿った因果関係があることを表し, アークの根本である原因事象となるノードを親, アークの先端である結果事象となるノードを子と呼ぶ.
4. 親を持たないノードで表された確率変数は独立した確率分布 (事前確率分布) が定義され, 1 つ以上の親を持つノードで表された確率変数は親の状態に依存した確率分布 (条件付き確率分布) が定義される.

ベイジアンネットワーク上での確率的推論では、いくつかの確率変数の事前確率分布が与えられたとき、もしくはいくつかの確率変数の値が観測されたときの、目的の確率変数の事後確率分布を求める。ここで、推論のもととなる情報が持つ確率分布が事前確率分布として利用される。目的の確率変数 X_k についての事後確率分布は次のように計算できる。

$$p(X_k|X_{\text{obs}}) = \alpha \sum_{X_{\text{hid}}} \cdots \sum_{X_j \in X_{\text{all}}} \prod p(X_j|\text{Parents}(X_j)) \quad (2.1)$$

ここで、 $\text{Parents}(X)$ は確率変数 X へのアークを持つノード（親）の集合、 X_{all} はベイジアンネットワーク上の全確率変数集合、 X_{obs} は観測済み確率変数集合、 X_{hid} は隠れノード集合である。 X_{hid} は X_{all} から X_k および X_{obs} を除いた集合として次のように定義される。

$$X_{\text{hid}} = X_{\text{all}} \setminus (X_k \cup X_{\text{obs}}) \quad (2.2)$$

また、 α は正規化定数であり、 $\sum_{X_k} p(X_k|X_{\text{obs}}) = 1$ を満たすように決定される。

Eq. 2.1 における $p(X_j|\text{Parents}(X_j))$ は、 $\text{Parents}(X_j)$ が空集合の場合には事前確率分布であり、1つ以上の要素（親）を持つ場合には親の状態のもとでの条件付き確率分布となる。条件付き確率分布は推論を行うためのルールとなるものであり、予め知識として与えられる。

2.3 変化検出システム

2.3.1 変化検出システムの概要

本節では、2.2節で提案した推論手法をもとに提案した変化検出システムの枠組みを説明する。本システムを用いた検証実験結果は5章にて言及し、その際に検証実験で利用したシステムの具体的な構成を示す。

本システムはオントロジーを用いて検出対象の概念の意味を解釈し、それに応じて知識モジュールでの情報抽出や推論のアプローチを選択することで、検出対象に適した解析を行う。本システムは主に次の3つのプロセスから構成される。

- 検出対象の分析および知識の統合

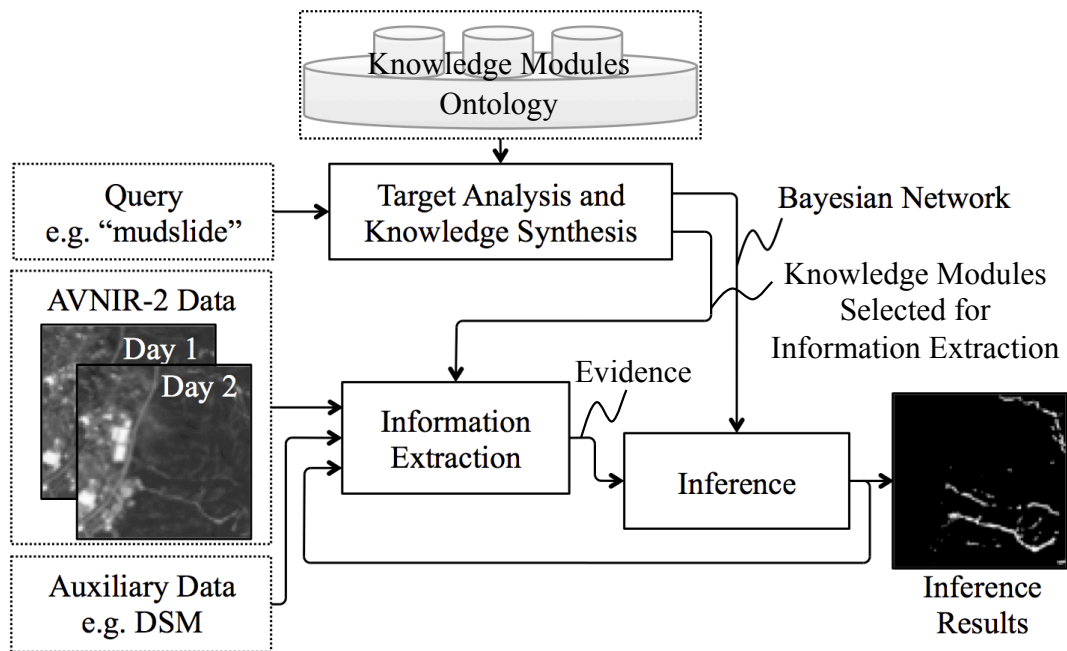


Fig. 2.1: 変化検出システムの流れを示す IDEF0 図

- 情報抽出
- 地物・地物変化の推論

Fig. 2.1 は本システムを構成するプロセスと、各プロセスの入力、出力、制約、および機構となる概念を IDEF0 図により記述したものである。

2.3.2 項では、本システムにおいて知識記述の基盤として用いられるオントロジーについて説明する。ここでは、オントロジーの構成や概念を定義する際に用いられる公理的な関係について述べる。また、地物や地物変化の概念を衛星データ解析という観点から定義した地物・地物変化オントロジーについても説明する。

2.3.3 項では、概念間の定義以外の関係や各種情報抽出処理といった知識を容易に追加・変更可能なモジュールとして表現した知識モジュールに関する説明を述べる。

2.3.4 項では、本システムによる変化検出の流れを解説する。

2.3.2 地物・地物変化のオントロジー

本研究で構築するオントロジーはドメインオントロジーと呼ばれる種類のオントロジーであり、ある問題領域内で共有されるべき概念の定義を記述したオントロジーである。オントロジーの定義にはいくつかの流儀が存在するが、本研究では溝口 [29] によるオントロジーの定義を採用する。溝口はオントロジーを「人間が対象世界をどのように見ているかという根本的な問題意識をもって物事をその成り立ちから解きあかし、それをコンピュータと人間が理解を共有できるように書き記したものである」と定義し、オントロジーの構成物として次の4つを挙げている。

1. 対象世界から基本概念を切り出した結果としての「概念」の集合
2. 概念の is-a 関係（上位・下位関係）による階層化
3. is-a 関係以外で必要となる概念間の関係
4. 抽出した概念と関係の定義、あるいは意味制約の公理化

本研究ではリモートセンシングの分野に登場する地物や地物変化の概念を衛星データ解析という観点から定義したドメインオントロジーを構築する。リモートセンシングに関するオントロジーの先行研究としては長井ら [30] のオントロジーがあるが、これは辞典に掲載されている用語を体系的に整理したものであり、オントロジーの構成要素や哲学的考察の深さという分類観点ではライトウェイトオントロジーと呼ばれるオントロジーに属する。それに対して本研究で構築する地物・地物変化オントロジーは同じ分類観点ではヘヴィウェイトオントロジーに属し、その性質や用途は異なる。地物・地物変化オントロジーは、リモートセンシングの分野に登場する概念を「その概念にとって何が本質的性質であるか」という基準で分析・定義することにより、コンテキスト依存となる曖昧な性質が取り除かれた概念定義の記述を目指している。例えば、あらゆる「川」という概念が共通に持つ性質は何か、どのような性質を持つならば「川」とであると一意に決まるのか、という観点から分析することにより「川」という概念が持つ本質的性質を定義する。

本研究で構築するオントロジーでは、概念間の意味的關係に形式表現を与えるための公理を次のように定める。ここで、 S_X を概念 X の実体の集合とし、 $has(X)$ を、概念 X を部

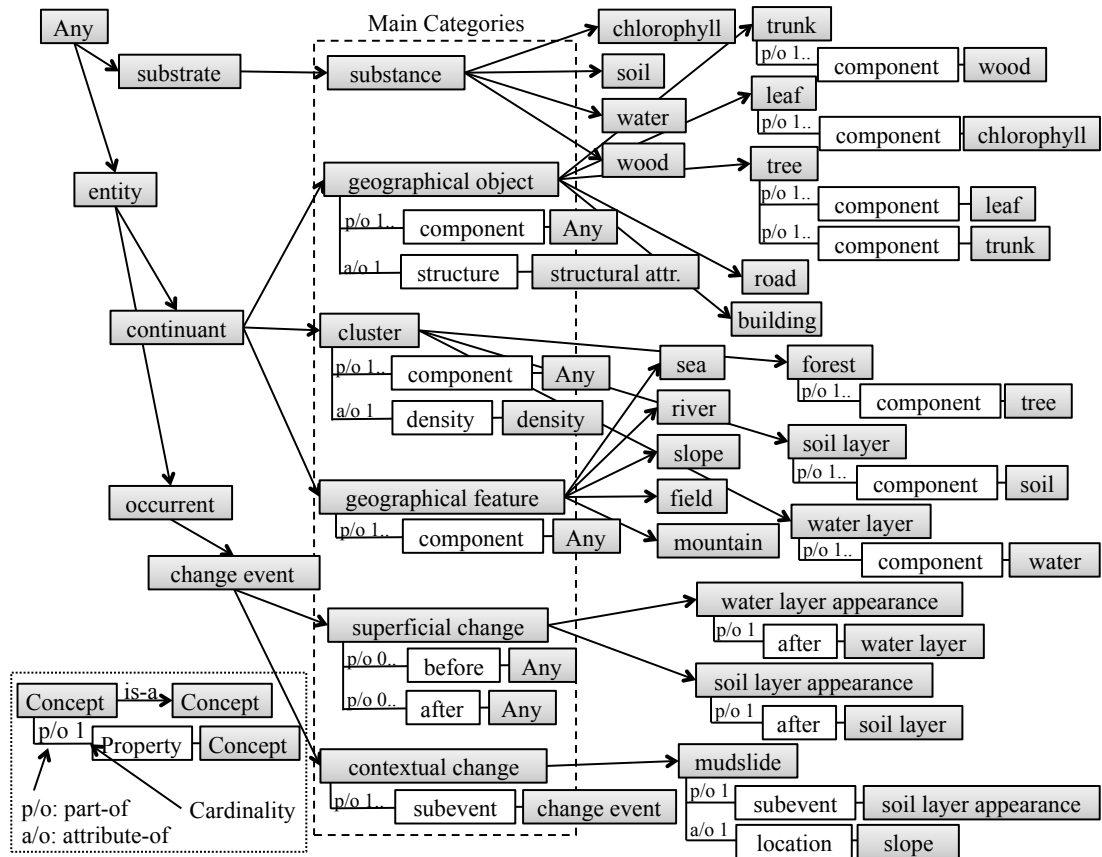


Fig. 2.2: 地物・地物変化オントロジー

分あるいは属性として持つ実体の集合とする。

1. 概念 A と概念 B が is-a 関係によって B is an A と結ばれているとき, S_B は S_A の部分集合である。(これは, S_B は S_A の持つ全ての part-of 関係や attribute-of 関係を継承していることを意味する)
2. 概念 A と概念 B が part-of 関係で B is a part of A と結ばれているとき, S_A は has(B) の部分集合である。
3. 概念 A と概念 B が attribute-of 関係で B is an attribute of A と結ばれているとき, S_A は has(B) の部分集合である。

オントロジーの記述には, 古崎ら [31] によるオントロジー記述環境『法造』を用いた。

Fig. 2.2 は地物・地物変化オントロジーの一部を抜粋したものである。

地物・地物変化オントロジーでは地物の概念をその性質に従って substance, cluster, geographical object, そして geographical feature という4つの地物カテゴリーに分けて定義している。このうち, substance は地物の根本的概念であり, ほかのカテゴリーに属するあらゆる地物概念は, その構成を表現する part-of 関係を掘り下げていくと substance から派生的に定義されている。この定義方法は Aristotle の形而上学 [32] における実在物の表現である「質料」と「形相」の考え方に基づいており, 「形相」をもたない「質料」(素材, 材料のことで, 物体の性質を決定する最小単位) が, 「形相」(物体の形状や機能) を与えられることではじめて実在する物体として成り立つと考える。

それぞれの地物カテゴリーについて説明する。

substance

リモートセンシングにおける「質料」として扱われる, リモートセンシングにて捉えることが可能な最小の素材・材料である概念のカテゴリー (例えば, soil, water)

cluster

具体的な形状や構造は不定であるが, 何かが局所的に集積していることを表す実在物概念のカテゴリー (例えば, soil layer, forest)。

geographical object

ある固有の形状・構造を持ち, それがアイデンティティとして考えられる実在物概念のカテゴリー (例えば, tree, building)。

geographical feature

単純に形状や構造だけでは定義することができず, 成因や機能, 周囲の地物との関係といったコンテキストがアイデンティティになっている実在物概念のカテゴリー (例えば, field, river)。

また本研究では地物変化の概念を, 変化の意味に従って superficial change と contextual change という2つの地物変化カテゴリーに分けて定義している。

superficial change

「地物 A が地物 B に変化した」「地物 A が出現した／消滅した」「地物 A の色が変わ

化した」というように、地物の状態の変化から定義される地物変化概念カテゴリー（例えば, leafing, soil layer appearance）.

contextual change

地物状態の表面的な変化ではなく、自然現象や土地開発のように発生原因等のコンテキストから定義される地物変化概念カテゴリー（例えば, mudslide, land development）.

以上の地物・地物変化カテゴリーはそれぞれ意味が大きく異なり、解析対象概念が属するカテゴリーによってその解析アプローチも異なるものになる．そのため 2.3.4 項で提案する変化検出システムの推論の流れでは地物・地物変化カテゴリーごとの解析処理を行うというアプローチを取っている．

2.3.3 知識の表現

ある地物や地物変化の概念、データ、データから求められる特徴量の間の関係に関する知識は知識モジュールとして記述され、知識ベースに蓄積される．概念間の関係はオントロジーにも記述されるが、オントロジーには概念間の本質的・体系的で定義として考えることが可能な関係が記述されるのに対し、知識ベースにはいくつかの概念間の不確実性を持つ関係が分散的に蓄積される点で異なる．

それぞれの知識モジュールは既知の推論結果やデータ等の情報を入力することで独立に情報抽出や推論を実行し、その結果を出力する．知識モジュールの入出力となる情報はオントロジーで定義された概念の生起についての推論結果か AVNIR-2 によって観測された反射率というような特定可能なデータであり、暗黙的な仮定を含まず一貫性が保証される．また、知識モジュールの入出力となる概念の生起についての推論結果の表現にはベイズ確率を用い、推論結果の信頼度を表す区間 $[0, 1]$ の確率値で記述する．ベイズ確率は頻度の表現である客観確率に加えて信頼度の表現である主観確率も含めた広義の確率解釈である [26]．ベイズ確率には古典的確率の理論体系をそのまま適用することができるため、知識や情報の不確実性表現とその演算に適している．推論結果の記述にベイズ確率を用いることで、不確実性を伴った推論結果を知識モジュール間で共有・再利用しつつ推論を進めることが可能になる．

知識モジュールとして実装される知識の例として次の3種類の知識を想定する。

複数の概念の確率的な共起関係

あるピクセルにおける複数の概念の共起関係を、それぞれの概念の生起事象の間の条件付き確率分布によって記述する。例えば、*mountain* と *forest* との間の共起関係は、2概念の生起事象間の条件付き確率分布 $p(\text{forest}|\text{mountain})$ によって記述する。確率的関係のグラフィカルモデルであるベイジアンネットワークを用いることで、複数の事象間の共起関係であっても簡素に表現することができる。

概念と証拠となる観測値の関係

ある概念の生起事象と反射率や数値標高といった証拠となる観測値の関係を条件付き確率分布によって記述する。概念の生起事象は2値確率変数で表されるが、証拠となる値は主に連続値をとる確率変数で表される。例えば、*substance* カテゴリーに属する概念である *chlorophyll* と *chlorophyll* の証拠となる観測値である正規化植生指数 (NDVI) の関係は、*chlorophyll* の存在で条件付けられた NDVI の確率分布 $p(\text{NDVI}|\text{chlorophyll})$ によって記述する。

手続き的知識

地物の認識手法のような手続き的知識はプログラムモジュールとして実装される。例えば、*cluster* の解析を行うための知識として、入力された画像のセグメンテーションを行ったのち、各セグメントの特徴量を求めるようなプログラムモジュールが考えられる。

2.3.4 変化検出の流れ

本システムを用いた変化検出を行うにはまず、検出対象の概念を表すクエリ（例えば、土石流を検出対象とする場合は “*mudslide*” という文字列）を入力する。検出対象クエリが入力されると、本システムは検出対象クエリが表す概念の意味をオントロジーによって分析し、その概念を推論するためのベイジアンネットワークを自動構築する。Ding ら [33] はオントロジーにおける概念間の論理関係から条件付き確率分布表を自動生成し、ベ

イジアンネットワークを構築するための手法を提案しており、本研究では条件付き確率分布表の生成のためにこの手法を採用する。またベイジアンネットワークを自動構築する際に、概念を推論するための証拠情報の抽出に適用可能な知識モジュールを知識ベースから探索する。

Fig. 2.3 にベイジアンネットワークを構築する過程を例示する。この処理の実行手順を以下に説明する。

Step 1:

検出対象概念をオントロジーから検索し、その概念の定義に用いられる概念を再帰的に展開することで、検出対象概念の定義構造に現れる全ての概念を探索する。オントロジーに登場する概念は必ず根本的概念から派生して定義され、かつ循環参照が無いように構築されているため、この処理は根本的概念に帰着することで終了する。

Step 2:

得られた定義構造からベイジアンネットワークを構築する。ベイジアンネットワークでは、定義構造に登場する各概念はその概念の存在を表す2値確率変数ノード（概念ノード）として表現される。また、定義構造に登場する概念間の関係はノード間のアークとして表現され、オントロジーの公理に従う論理関係を持つ条件付き確率分布が割り当てられる。例えば、概念 A が概念 B と概念 C から定義されているとき、ベイジアンネットワーク上では概念 B から概念 A へのアークと概念 C から概念 A へのアークで表現され、概念 B と概念 C の両方が存在しているときのみ概念 A が存在するという論理積関係を表す条件付き確率分布が割り当てられる。

Step 3:

それぞれの概念ノードについて、その概念の証拠情報の抽出に適用可能な知識モジュールを知識ベースから探索する。探索された知識モジュールの実行に必要な情報として新たな概念の情報が出現した場合には、その概念について Step 2 から Step 3 を適用する。

本システムでは変化検出のプロセスが推論対象の概念のカテゴリによって substance

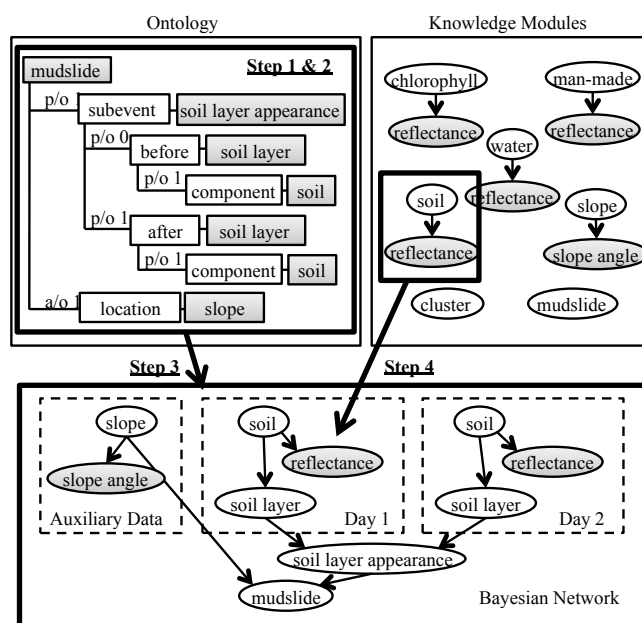


Fig. 2.3: ベイジアンネットワークの構築

推論, object 推論, change 推論という3つの段階に分割され, それぞれの段階ごとに情報抽出と推論が実行される。

substance 推論

ピクセルごとに substance カテゴリーに属する各概念の事後確率が推論される。この推論の入力となる情報は, ピクセルベース画像解析で得られた情報のみとなる。

object 推論

ピクセルごとに cluster や geographical object に属する概念の事後確率が推論される。このとき, substance 推論によって得られた各 substance の事後確率が, 事前確率として推論に用いられる。object 推論の証拠となる値には, ピクセルベース画像解析によって抽出された情報の他に, オブジェクトベース画像解析によって抽出された情報も使われる。

change 推論

ピクセルごとに superficial change または contextual change の事後確率が推論される。このとき, 2 時期の衛星データにおいて行われた substance 推論と object 推論

によって得られたそれぞれの地物の事後確率が、事前確率として推論に用いられる。

これら 3 つの段階の推論が順に実行され、最終的に対象の地物変化の事後確率が得られる。

このように本システムでは推論の順序が固定的であり、解析処理の自由度が低いという制約がある。4 章ではより自由度の高い解析処理を実現するための手法について言及する。

2.4 本章のまとめ

本章では情報の不確実性を考慮しつつ複数のプリミティブな情報からより複雑な情報を推論するため、ベイズ確率による情報表現及びベイジアンネットワークによる確率的推論を採用した枠組みを提案した。推論では予め知識として与えたベイジアンネットワークを利用し、推論の証拠としてベイジアンネットワークを構成するいくつかの確率変数が持つ確率分布、もしくはそれらの確率変数の値が与えられたときの、目的の確率変数の事後確率分布を求める。

提案手法の検証のため、衛星データの災害状況把握解析のための自動変化検出システムの枠組みを提案した。本システムは本研究にて構築した地物・地物変化のオントロジーを地物・地物変化の定義を調べるための辞書として利用し、利用者が入力した検出対象の地物・地物変化名からその定義を調べ、知識ベースに蓄積されている単純な情報抽出や推論を行うための知識モジュールを組み合わせその地物・地物変化を推論するためのベイジアンネットワークを自動構築し、推論を行う。ただし本システムは解析処理の自由度が低いという制約があり、4 章ではより自由度の高い解析処理を実現するための手法について言及する。また、本章で構築した変化検出システムの検証実験結果は 5 章で言及する。

第3章 光学衛星データの土地被覆推論の自動化・高精度化

3.1 はじめに

本章では光学衛星データの土地被覆推論の自動化と高精度化を目的とし、新たな土地被覆推論手法を提案する。提案手法は次の4つの特徴的な点を有し、これにより高精度な分類を実現する。

- 確率的推論を導入することで、分類クラス間の重なりや教師データの密度といった不確実性を定量的に評価可能な事後確率として分類結果を得る。
- ノンパラメトリック推定手法であるカーネル密度推定により、大量の教師データがつくる分布を利用した高精度な推論を行う。
- 観測された地表面反射率に加えて年周期性を連続的に表現できる観測日変量を入力データとすることで、土地被覆クラスの季節性を考慮した土地被覆分類を行う。
- 多時期データのそれぞれに対して得られた分類結果の事後確率を統合し、分類スコアを求めることで、土地被覆クラスを良好に識別可能な時期の分類結果を強く反映した最終結果を得る多時期分類が可能である。

本章の構成は以下の通りである。

3.2 節では、土地被覆分類の現状と課題、並びに一般的な土地被覆分類手法の類別毎の特徴を紹介する。

3.3 節では、分類問題に適用される確率的モデルである生成モデル上での確率的推論を土地被覆分類問題に適用するための検討を行う。

3.4 節では、データ分布に対する仮定を設けないノンパラメトリックな密度推定手法であるカーネル密度推定を採用し、大量の教師データがつくる分布を利用した高精度な確率的推論手法を提案する。

3.5 項では、観測された地表面反射率に加えて年周期性を連続的に表現できる観測日変量を入力データとすることにより土地被覆クラスの季節性を考慮した土地被覆推論を行うための拡張を提案する。

3.6 項では、多時期の観測データを用いて高精度な土地被覆推論を行うための拡張を提案する。

本章で構築した土地被覆推論手法の検証実験結果は 6 章で言及する。

本章の提案内容及び 6 章で示す検証実験結果は著者らによる投稿論文 [34] にて報告したものである。

3.2 土地被覆分類の現状と課題

土地被覆分類図は地図作成や土地利用計画、植生調査、森林・農地管理、災害被害把握などの様々な分野のアプリケーションに活かすことができる基盤情報として利用されている。広域を同時に、かつ定期的に反復観測可能な衛星データから土地被覆分類図を作成・更新していくことは非常に有効であり、これまでも多くの研究機関が広域の土地被覆分類図を作成し、広く利用できるよう公開している。著名な土地被覆分類図プロダクトには、Boston University による MODIS Land Cover (MOD12) [35][36] や European Commission による Global Land Cover 2000 (GLC2000) [37], United States Geological Survey (USGS) による Global Land Cover Characterization (GLCC) [38] がある。宇宙航空研究開発機構 (JAXA) 地球観測研究センター (EORC) においても、AVNIR-2 が取得した高分解能マルチスペクトルデータを主に用いて作成した日本周域の高解像度土地利用土地被覆分類図を 2010 年から公開している [39][25]。JAXA EORC では継続的に土地被覆分類図の高精度化を行っており、2013 年 12 月現在での最新版は 2013 年 2 月に公開されたバージョン 13.02 となっている。

日々継続的に広域を観測した膨大な衛星データを有効活用するためには、人の手を必要

としない自動処理によって衛星データから土地被覆分類図を作成できることが望ましい。しかし広域の土地被覆分類図を自動作成する上で、分類精度は既存手法の大きな課題となっている。土地被覆分類図の実用的な分類精度については事例によって様々な基準が提唱されており、例えば Thomlinson らは総合精度である OA で最低 85%、かつクラス毎の精度である PA で最低 70% という基準を設けている [40] が、従来の広域土地被覆分類図プロダクトはその精度まで到達できていない。前述の JAXA EORC による土地利用土地被覆分類図バージョン 13.02 の OA は 66.23% であったという結果を示している [39]。他にも、岩男らはユーラシア大陸上の 749 点の検証点情報を用いて 4 種類の土地被覆分類図プロダクトの精度評価を行い、OA の最も良いもので MOD12 global land cover product の 58% であり、他のプロダクトでも大差は無かったという結果を示している [41]。同様に Ran らは、中国科学院が作成した中国の 10,000 分の 1 土地利用図を利用し、21 箇所に設置した各 100 km 四方の領域において 4 種類の土地被覆分類図プロダクトの面的精度評価を行い、OA は最も良くて GLC2000 の 59% であった事を述べている [42]。

既存の自動分類手法には教師付き分類、教師無し分類、半教師付き分類という 3 つのカテゴリがある。

教師付き分類とは事前に教師データと呼ばれる分類クラスでラベル付けを行った観測値データを用意して、それをもとに分類器を構築しておき、観測値を分類する手法の総称であり、それに対して教師無し分類とは事前にラベル付けを行った教師データを用意せず、観測値同士の類似度をもとにクラスタリングを行う手法の総称である。教師付き分類は教師無し分類に比べて分類精度が高いことが知られており、本節の冒頭で紹介したような多くの実用例では教師付き分類に属する手法が利用されている。しかし教師付き分類で精度の良い分類結果を得るためには、分類対象の季節性や地域性、そして観測値の観測条件に適合した教師データが必要となり、広域かつ全ての時期について網羅するには大量の教師データが必要となる。異なる地域や異なる時期、異なる観測条件の教師データが混在してしまうと、分類クラス間で観測値分布の混合が発生してしまい、分類精度が悪化してしまう。

もうひとつの分類手法のカテゴリである半教師付き分類とは、少量のラベル付けを行った教師データに加えて大量のラベル無しデータを用意し、その混在データをもとに分

類器を構築することで、同量の教師データのみを用いた教師付き分類よりも分類精度を向上させることを目的とした手法の総称である。半教師付き分類であれば少量の代表的な観測値を示す教師データさえ用意できれば、それらと季節性や地域性、観測条件に合ったラベル無しデータを用いて分類することで精度良く分類できる可能性がある。しかし、半教師付き分類では分類クラス毎にクラスター分割可能であるという仮定をもとにした手法であるため、分類クラス毎にクラスター分割が困難な問題に対しては精度が悪化する可能性がある。しかし光学衛星データを用いた土地被覆分類ではクラスター分割が難しい土地被覆クラス同士が分類クラスに設定されることが多く、高精度化には限界があると推測する。

そのため本研究では、データベースに蓄えられた大量の土地被覆教師データを特定の種類の土地被覆に依存せず、効果的に活用できる教師付き分類手法を新たに提案する。

3.3 生成モデルを用いた確率的推論による土地被覆分類

土地被覆分類は、ある地点について得られた観測値を手がかりに、その地点の土地被覆状態を予め設定しておいた複数の土地被覆クラス中の 1 クラスに決定する問題を解く。ここでは入力データとなる D 次元の観測値ベクトルを $\mathbf{x} \in \mathbb{R}^D$ 、 J 個の分類クラスを $C_j (j = 1, 2, \dots, J)$ とする。

提案手法では、カテゴリカルな隠れパラメータである土地被覆クラスによって入力データが従う確率分布が決定されるという確率的モデルである生成モデル (generative model) を仮定し、観測された入力データの値から土地被覆クラスを確率的推論によって求める。生成モデルは次式のように表される。

$$p(\mathbf{x}, C_j) = p(C_j)p(\mathbf{x}|C_j) \quad (3.1)$$

この生成モデルは 2 種類の確率分布から構成されており、確率分布 $p(C_j)$ は土地被覆クラスに対する事前分布、もう一つの確率分布 $p(\mathbf{x}|C_j)$ はクラスで条件付けられた入力データの密度分布である。

入力データの観測後におけるクラス事後確率分布は、ベイズの定理を用いることで次のように求められる。

$$p(C_j|\mathbf{x}) = \frac{p(C_j)p(\mathbf{x}|C_j)}{p(\mathbf{x})} = \frac{p(C_j)p(\mathbf{x}|C_j)}{\sum_{j'=1}^J p(C_{j'})p(\mathbf{x}|C_{j'})} \quad (3.2)$$

ここで、求めたい土地被覆クラスの集合は排他的かつ網羅的である（すなわち、土地被覆クラス間におけるクラス事後確率の総和は1である）という分類問題の仮定を利用している．Eq. 3.2 を用いて入力データからクラス事後確率を求めるためには事前分布とクラス条件付き密度分布を知る必要がある．

提案手法では、事前分布には無情報分布である一様分布を仮定することとし、クラス条件付き密度分布は教師データの密度推定によって求める．その結果、得られるクラス事後確率分布の値は、観測された入力データの値においてクラス毎の教師データがなす密度分布の重なり具合を表す．すなわち、クラス毎の密度分布の重なりが少なく、ほぼ確信的にあるひとつのクラスに識別できるような入力データの場合、そのクラスにおけるクラス事後確率の値は1に近くなり、他のクラスは0に近くなる．また逆にクラス毎の密度分布の重なりが大きく、複数のクラスが分類候補として尤もらしくなり識別が困難であるような入力データの場合、複数のクラスにおけるクラス事後確率の値が高くなり、事後確率の最大値が低くなる．この性質を利用することで分類クラス間の重なりによる不確実性を定量的に評価することができる．また、Eq. 3.2 における $p(\mathbf{x})$ はクラス毎の密度分布を総合したデータ密度分布であり、観測された入力データ値付近において存在する教師データの密度を表す．入力データ値付近において存在する教師データが多いほど、それをもとにした推論は信頼できると考えられるため、データ密度分布の値も不確実性の評価に利用できると思われる．

3.4 カーネル密度推定を用いた高精度かつロバストなデータ密度分布推定

既存手法では、データ密度分布の推定にはデータがある密度分布モデルに従うことを仮定し、そのモデルパラメータを教師データから推定するというパラメトリックなモデル推定が用いられることが多い．その代表的な手法には最尤分類法（MLC）があり、これは分

類クラスごとに多次元ガウス分布の平均ベクトルと共分散行列を教師データにもとづく最尤推定によって予め求めておき、分類したい入力データに対して多次元ガウス分布が最も高い値を取る分類クラスに分類する手法である。こうしたパラメトリックな密度分布推定は少数のパラメータから密度分布の大域的な形状が決定できるため、教師データのサイズが小さい場合であっても密度分布を容易に推定することができる。しかしその反面、分布の自由度には制約があるため複雑な密度分布の表現は難しい。仮定した密度分布モデルが実データの密度分布に近い形状であれば良好な密度推定が可能と考えられるが、後に示すように、土地被覆分類の入力データとなる観測値はガウス分布等の単純な形状の分布をしていない。そのため本研究では、高い自由度の分布を表現可能なノンパラメトリック密度推定手法を採用し、大量の教師データを用いた密度推定を行うことで高精度化な推論を行う手法を提案する。

本研究では、クラス条件付き密度を教師データを用いたカーネル密度推定 (KDE)、別名 Parzen 窓法 [43] によって求め、3.3 節で示した確率的推論手法により土地被覆分類を行う手法である Kernel-based Probabilistic Classification (KPC) を提案する。カーネル密度推定は連続値の確率密度分布をノンパラメトリックに推定する手法である。ある点における確率密度を、各サンプルデータを中心に置いたカーネル関数の値の総和を取ることで推定する。カーネル密度推定を用いることで、観測値データの分布モデルに対する先見の仮定や制約を置くことなく実データに忠実な密度推定が可能となる。

ある土地被覆クラスの教師データ集合を

$$\{ \mathbf{x}_n = [x_{n,1}, x_{n,2}, \dots, x_{n,D}]^T \mid 1 \leq n \leq N \} \quad (3.3)$$

とすると、観測値 $\mathbf{x}$ においてカーネル密度推定によって推定される確率密度は次式で得られる。

$$p(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \left\{ \prod_{d=1}^D \frac{1}{h_d} K \left(\frac{x_d - x_{n,d}}{h_d} \right) \right\} \quad (3.4)$$

ここで $K(u)$ はカーネル関数であり、次のガウシアンカーネルを用いる。

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \quad (3.5)$$

また、 $h = [h_1, h_2, \dots, h_D]^T$ はそれぞれのカーネル関数の広がり決定するバンド幅パラメータであり、Scott の経験則 [44] を利用し、次式で求める。

$$h_d = N^{-1/(D+4)} \cdot \sigma_d \quad (3.6)$$

ここで σ_d は教師データにおける d 次元値の標準偏差を表す。

Fig. 3.1 は畑地の地表面反射率サンプルデータを用いてカーネル密度推定を行い、得られた確率密度分布をバンド 3 ($0.61 \sim 0.69 \mu m$, 赤色域), バンド 4 ($0.76 \sim 0.89 \mu m$, 近赤外域) について図示したもので、比較対象としてサンプルデータのヒストグラムおよび多次元ガウス分布の最尤推定にて得られた確率密度分布も同時に図示している。多次元ガウス分布の最尤推定では楕円体状の幅広い分布を推定しているが実データの分布には沿っておらず、反面、カーネル密度推定が実データの分布にほぼ忠実な密度分布を推定できていることが確認できる。

3.5 土地被覆の季節性を考慮した土地被覆分類

土地被覆によっては、得られる観測値の季節性を調べることで分類種目の細分化や分類精度向上が期待できる。例えば、水田の地表面反射率は多くの時期で水田以外の農地や草地と似たものになり判別が難しいが、水田の水入れ時期には他と異なる地表面反射率となり容易に識別可能となる。すなわち、同じ土地被覆であっても観測時期によって観測値の分布が変化することを考慮に入れることで、土地被覆クラス間の分布の重なりを低減することができる。

そこで提案手法では、マルチスペクトルバンドの地表面反射率と年間通算日での観測日を結合したデータを入力データとして土地被覆分類を行う。観測日の軸を追加することで季節性を持つ土地被覆クラスの密度分布が分離しやすくなり、分類精度を向上することができる。

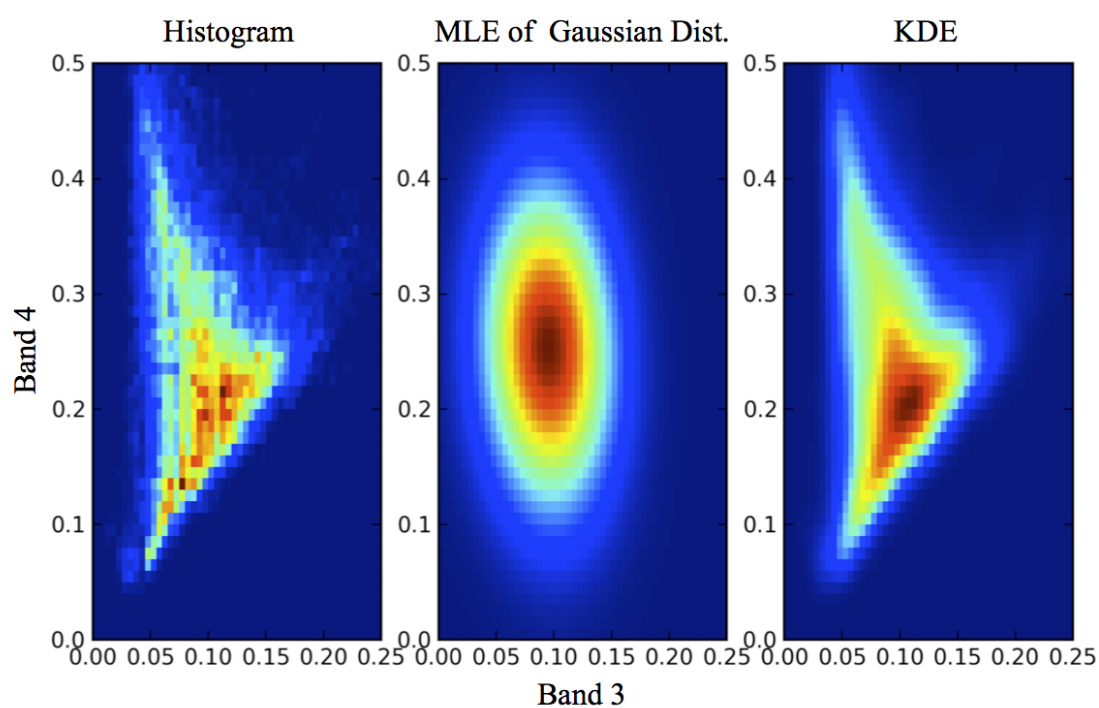


Fig. 3.1: 畑地の AVNIR-2 データの密度分布推定結果

左図は畑地における AVNIR-2 データのバンド 3 およびバンド 4 の地表面反射率のヒストグラムを表す。中央および右図は畑地における AVNIR-2 データを教師とした多次元ガウス分布の最尤推定およびカーネル密度推定により得られた密度分布を表す。

年間通算日での観測日をそのまま導入した場合、年末年始において数値的に不連続な点が生じてしまう。季節の年周期性を連続的に表現するため、観測日は次式によって2次元の単位円の円周上を動く変数 t_1, t_2 へと変換した上で分類に用いる。

$$[t_1, t_2] = [\cos(2\pi\text{DOY}), \sin(2\pi\text{DOY})] \quad (3.7)$$

ここで、DOY は年間通算日での観測日を区間 $[0, 1]$ に正規化した変数である。 t_1, t_2 は2次元変数であるが、これらは線形従属であり単位円の円周上に拘束されているため、実質的には1次元変数である。

3.6 多時期土地被覆分類

既存の土地被覆分類手法では、数日から数ヶ月おきに同一地点を観測した時系列データを利用し、土地被覆毎の地表面反射率の季節変化を調べて分類することがよく行われる。本章の冒頭で紹介した MOD12[35][36] や GLC2000[37], GLCC[38] はどれも比較的低分解能ではあるが高頻度で観測された衛星データから時系列データを構築し、それを用いた分類によって作成された土地被覆分類図プロダクトである。

しかし AVNIR-2 のような高分解能・低頻度観測の光学衛星データにおいてはシーン間での観測条件の差異や被雲による欠測の影響が大きく、同一地点における地表面反射率の季節変化を追うことのできる良好な時系列データを得ることが難しい場合が多い。高分解能・低頻度観測の光学衛星データを用いた高精度な広域土地被覆分類図作成のためには、そうした場合であっても土地被覆の季節変化を踏まえた分類が可能な手法が必要とされる。

そこで、多時期データを用いて土地被覆を判別するための事例を増やし、土地被覆を容易に判別可能な事例における分類結果を最終的な分類結果に強く反映させる手法を提案する。これは、地表面反射率の季節変化を調べるために時系列データを構築する従来手法とは多時期データに対する考え方が根本的に異なる。

提案手法では、まず単時期（単シーン）毎の解析を行い、その後、同一地点における多時期の解析結果を総合して最終的な土地被覆分類結果とする。多時期の解析では複数時期の解析によって得られたクラス事後確率を総合し、最終的な土地被覆クラスを決定する。

ここでは、同一地点における多時期データ中での土地被覆変化は考えず、ひとつの土地被覆クラスに分類されるという仮定のもとで分類を行う。

複数時期のクラス事後確率を総合する最もシンプルな方法は、クラス事後確率が最も高くなったクラスの頻度を調べて最も高頻度となったクラスを多時期の分類結果とする多数決方法である。しかしこの方法では、確率的推論によって得られたクラス毎の確信度の情報が使われずに失われてしまう。そこで、Eq. 3.8 に示す事後確率の総乗をスコアとした分類手法である Kernel-based Probabilistic Classification with Product Score (KPC-PROD) と、Eq. 3.9 に示す事後確率の総和をスコアとした分類手法である Kernel-based Probabilistic Classification with Summation Score (KPC-SUM) の2手法を検討する。

$$S_{\text{prod}}(C_j) \propto \prod_{\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots\}} p(\mathbf{x}|C_j) \quad (3.8)$$

$$S_{\text{sum}}(C_j) \propto \sum_{\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots\}} p(\mathbf{x}|C_j) \quad (3.9)$$

これにより得られたスコアが最も高かった土地被覆クラスを最終的な分類結果とする。総乗をスコアとする手法である KPC-PROD は全ての時期の観測値データがひとつの土地被覆のものであるという同時事象を考えることに等しい。すなわち、各土地被覆クラスを識別するために重要な時期での分類結果が最終的な解析結果を求める際に強調されるが、その反面、ノイズの影響を受けやすいというデメリットがあると考えられる。一方、総和をスコアとする手法である KPC-SUM は事後確率の平均値を考える事に等しく、ノイズの影響が低減されるが、識別しやすい時期での観測値が多くなればクラス間の差異が出にくく、観測時期の偏りに影響されやすいというデメリットがある。

3.7 本章のまとめ

本章では光学衛星データを用いた広域土地被覆分類のための新たな分類手法である KPC を提案した。KPC は教師データを用いたカーネル密度推定によって地表面反射率とその観測日を統合した観測値の密度分布を推定し、推定された密度分布を用いた確率的推論を

行なうことでシーンごとの分類結果の信頼度を表す事後確率を得る。また、最後に多時期の事後確率を総合したスコアを求めることで精度の向上を行う多時期分類手法として、事後確率の総乗をスコアとした手法である KPC-PROD と事後確率の総和をスコアとした手法である KPC-SUM の2種類を提案した。

提案手法の特徴として、次の4つの点が挙げられる。

- 分類クラス間の重なりや教師データの密度といった不確実性を定量的に評価可能な事後確率として分類結果が得られる。
- 分布に対する先見的な仮定や制約を置くことなく、実データに忠実な分布を推定でき、かつ外れ値には影響されにくい高精度かつロバストな推論ができる。
- 検証点データおよび衛星データセットを用意するだけで自動処理を行うことができるスケーラブルなアルゴリズムである。
- 低頻度観測かつ欠測が多いため季節変動を追いつらいデータセットであっても、得られているデータのみから季節性を考慮した土地被覆推論が可能である。

本章で提案した土地被覆推論手法の検証実験結果は6章で言及する。

第4章 データ及びプロセスの蓄積と解析ワークフローの自動構築

4.1 はじめに

衛星データ解析には衛星データや地理データ、検証済みデータ、補正処理や解析処理といった様々なデータやプロセスが登場し、これらを組み合わせることで解析が実現される。一般的な衛星データ解析では主に利用者がそのデータやプロセスの組合せを考えるという作業を担っている。典型的な作業の流れは、目的の解析結果を得るため、様々なデータやプロセスがある中から必要なものを選定し、それらを組み合わせた解析ワークフローを考え、解析用のソフトウェアを介して実行する、というものである。様々な特性を備えたデータやプロセスがある中でどのデータに対してどのようなプロセスを適用させていけばよいかを判断するには固有の知識や経験を要するため、この作業は特殊な訓練を積んだ専門家でなければ困難な作業である。それ故、衛星データ解析を行うことのできる人は限られてしまい、衛星データの利用拡大の阻害要因となっているほか、災害発生時の被害状況解析のような大規模な判読を行う際の高速化のボトルネックとなっている。加えてデータの種類の増加や提案されている解析処理の増加などにより、専門家であってもそれらを網羅的に把握することは難しくなっているという問題もある。

そこで本章では、衛星データ解析に登場するデータやプロセスを機械可読な形で表現・蓄積しておき、それらを組み合わせた解析ワークフローを自動構築および実行するための枠組みを提案し、それに基づく自動判読システムを構築する。将来的に実際に運用される自動判読システムでは、扱うデータやプロセスが多種多様かつ膨大となることが想定される。そのため、提案する枠組みはそのようなデータやプロセスを扱うことを想定し、下記のような特徴を備えたデータやプロセスの蓄積方法を実現するものである。

機械可読性

データやプロセスの意味が、人間だけでなく機械が読み取ることができるよう記述される。

共有性

データやプロセスの意味が、その整合性が保たれ、複数の人やシステム間で共有できるよう記述される。

再利用性

データやプロセスがある特定の問題に対してのみ利用できるよう記述されるのではなく、様々な問題に対して繰り返し利用できるよう汎用的に記述される。

モジュール性

データやプロセスの追加・削除・編集が容易となるように、個々のデータやプロセスを外部からのアクセス手段が明示化されたひとまとまりのモジュールとして記述することができる。

分散管理性

データやプロセスを集中管理せず、ネットワーク上で分散的に管理・運用することができる。

柔軟性

将来的にデータやプロセスの意味を記述するためのデータモデルに変更が必要となったとしても、既に記述されたデータやプロセスを無駄にすることなく利用することができる。

本章の構成は以下の通りである。

4.2 節では、提案する自動判読システムの構成および動作概要について述べる。また、提案システムにおけるデータやプロセスの意味情報記述の基盤として採用する RDF についても説明する。

4.3 節では、提案システムにおける意味情報記述に対する重要な役割を担うオントロジーについて説明する。提案システムでは 2.3.2 項で説明した地物オントロジーおよび地物変

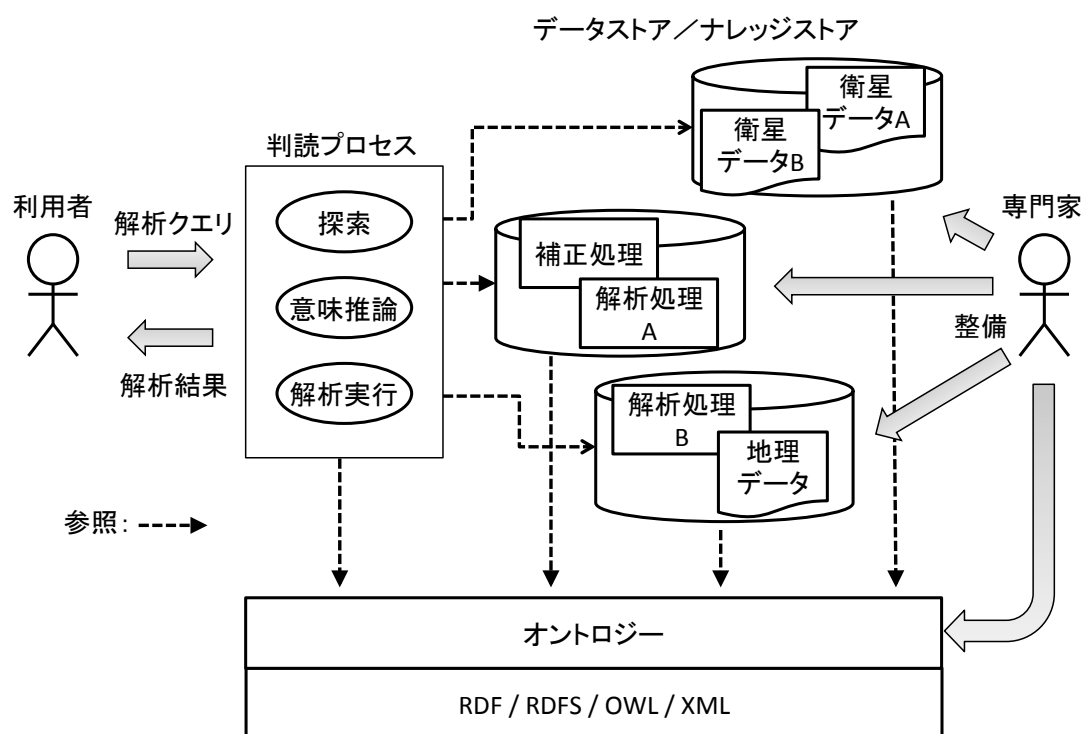


Fig. 4.1: 提案するシステムアーキテクチャ

化オントロジーの他に、時間オントロジー、幾何オントロジー、観測センサオントロジー、関係オントロジー、データオントロジー、そしてプロセスオントロジーという6種類のオントロジーを利用する。この6種類のオントロジーについてそれぞれ説明する。

4.4節では、提案システムにおいて蓄積するデータやプロセスへの要件、およびそれらが蓄積されるデータストアやナレッジストアの構成について説明する。

4.5節では、提案システムによる自動判読の主要なプロセスである解析ワークフローの自動構築の流れについて説明する。

本章で構築した自動判読システムの検証実験結果は7章で言及する。

4.2 自動判読システムの構成

Fig. 4.1は提案するシステムアーキテクチャを示した図である。提案システムでは、利用者は目的の解析結果を表すクエリをシステムに入力し、解析途中で発生する簡単な選択を行うという操作のみを行うことで、その解析結果が得られる仕組みとなっている。

提案するシステムアーキテクチャは RDF というメタデータ記述の枠組みを基盤としている。RDF は WWW 上に存在するリソースについての情報を記述するための言語であり、Semantic Web 実現のための構成要素として開発され、W3C によって規格化されている [45]。RDF は単に人間が解釈するための情報記述ではなく、機械読解可能な形式での情報記述の枠組みを提供している。提案システムでは RDF をメタデータ記述の基盤とすることで、蓄積されるデータやプロセスの意味を機械読解可能な形式で記述する。

RDF の主な特徴として、トリプル (triple) 形式と URI によるリソース参照が挙げられる。トリプル形式は主語 (subject)、述語 (predicate)、目的語 (object) の 3 種類のデータの組を 1 つのエントリとしたデータ構造であり、あるデータについてのメタデータと、それらの間の関係を 1 つのエントリに含めて記述できる。テーブルデータのように、スキーマおよび記述されたデータ間の配置の両方を用いてそれぞれのデータが表す意味を読み取らなければならないデータモデルに対し、トリプル形式によるデータモデルではデータが表す意味を 1 つのエントリのみから陽に独立に読み取ることが可能である。これにより、スキーマと記述されたデータとの依存関係を考える必要性がなくなり、スキーマが更新されたとしても記述されたデータの永続性を保つことができるという柔軟性を持つ。また、1 つのトリプルエントリは主語と目的語を表す 2 つのノードを、述語を表す有向辺で結んだ有向グラフとして考えることができ、複数のトリプルエントリを合成した大きな意味ネットワークも記述可能である。しかしその際には、異なるトリプルエントリで記述されているリソースが同一のリソースであることを保証する仕組みが必要となる。RDF ではトリプルエントリとして記述されるリソースへの参照は URI によって記述されるため、主語や述語、目的語を一意に特定することができる。この 2 つの特徴により、あるリソースについての意味ネットワークを複数の RDF データとして分散的に記述・管理・利用することが容易に実現できる。

RDF ではデータとメタデータ、およびそれらの間の関係を記述できるが、データとメタデータの間の関係もデータとして扱われるため、さらにその関係についてのメタデータも同様に記述することができる。このことから、RDF は自身が記述するデータについてのスキーマを同じ RDF データとして記述できる自己記述型データであるという特徴を持つ。すなわち、RDF はリソースについての情報であるメタデータを記述するという役割

の他に、そのリソースについて記述できる、あるいは記述されるべきメタデータの規定を記述するという役割も同時に担う。ある対象が本来持つべき情報の体系的定義であるオントロジーも RDF データとして記述することができる。RDF でスキーマを記述するための枠組みは RDFS として、同様に RDF でオントロジーを記述するための枠組みは OWL として定義され、どちらも W3C によって規格化されている [46][47]。

提案システムではメタデータ記述の形式的なプロトコルとして RDF を用いるのに対し、メタデータ記述の意味的なプロトコルとして独自に構築したオントロジーを利用する。オントロジーはシステム内において、解析対象やデータ、プロセスの意味情報を記述する際の指針としての役割、記述された意味情報の整合性を担保する役割、記述された意味情報を解釈するための辞書としての役割を担い、共有性や再利用性を実現するための重要なメカニズムである。ここで用いられるオントロジー中には次に列挙するようなカテゴリーがある。

- 地物オントロジー
- 地物変化オントロジー
- 時間オントロジー
- 幾何オントロジー
- 観測センサオントロジー
- 関係オントロジー
- データオントロジー
- プロセスオントロジー

このうち、地物オントロジーおよび地物変化オントロジーについては 2.3.2 項で既に説明している。それ以外のオントロジーについては次の 4.3 節で詳しく説明する。本研究では、オントロジーは 2.3.2 項と同様にオントロジー構築環境『法造』を用いて構築する。その後、RDF 処理系で扱うことができるように OWL 形式でエクスポートして利用する。そ

のため、ロール概念等のような法造独自のオントロジー実装は利用せずにオントロジーを構築する。

システムで利用されるデータやプロセスはそれらの意味を機械可読に記述した RDF データとともに、分散的に管理されているデータストアやナレッジストアに蓄積される。衛星データや地理データといった各種データはデータストアに、補正処理、解析処理といった各種プロセスはナレッジストアにそれぞれ蓄積される。RDF は URI 参照を用いており、データストアやナレッジストアが置かれている場所によらずそれぞれのデータやプロセスを一意に参照することができるため、これらのデータストアやナレッジストアは分散的に構築可能であり、様々な専門家や機関が個別に管理・整備できる。そうした分散管理のシステムにおいても記述されたデータやプロセスの意味を一意に定めるため、すべての RDF データは唯一のオントロジーを参照し、それに基づき記述される。

システムの利用者が目的の解析結果を表すクエリをシステムに入力すると、システムはそれを用いてデータストアに蓄積されているデータおよびナレッジストアに蓄積されているプロセスを探索し、意味推論を行いながらそれらを組み合わせることで目的の解析結果を得るための解析ワークフローを構築して実行する。システムに入力される解析クエリはデータストアに格納されているデータと同様の RDF データであり、これにより最終的に得られる解析結果のデータを指定する。RDF データは記述の自由度が高く、これを直接作成・編集するためには習熟が必要であるため、利用者や専門家らがデータやプロセスの意味を記述した RDF データを作成する際には、オントロジーを参照しながら必要な情報のみを入力していくだけで作成できるようにする平易なインターフェースのエディタを用いることを想定する。

4.3 自動判読システムのためのオントロジーの拡張

4.3.1 自動判読システムに必要なオントロジーの要件

本研究で構築するオントロジーの目的は、自動判読システム内における解析対象やデータ、プロセスの意味情報記述の基盤としての利用である。解析対象である地物や地物変化の概念についてまとめたオントロジーは 2.3.2 項にて説明している。本節では、残りのデー

タやプロセスの意味情報記述の基盤として利用されるオントロジーについて述べる。

データやプロセスの意味情報の記述のためは、時間、幾何、観測センサといった基本的な概念が必要となる。例えば衛星データの意味情報を記述する際には、いつ、どこを、どのセンサで観測して得られたデータなのかを記述することになる。時間、幾何、観測センサに関する情報の記述の基盤として利用するため、これらの概念に関するオントロジーを用意する。4.3.2 項では時間オントロジーについて、4.3.3 項では幾何オントロジーについて、そして4.3.4 項では観測センサオントロジーについてそれぞれ説明する。

柔軟な意味情報記述のために、概念同士の間、概念とインスタンスの間、またはインスタンス同士の間における関係を記述するための関係概念の定義も重要である。関係概念には、例えば2個のインスタンスの同一性や非同源性といった関係、時間的關係や幾何的關係が含まれる。4.3.5 項では関係オントロジーについて説明する。

本研究で構築するオントロジーにおけるデータやプロセスの定義は、広く一般的に用いることのできる定義ではなく、提案する自動判読システムの枠組みの中で利用される定義を想定する。目的の自動判読を実現する上で必要な情報がデータやプロセスの属性として定義される。4.3.6 項ではデータオントロジー、4.3.7 項ではプロセスオントロジーについてそれぞれ説明する。

4.3.2 時間オントロジー

Fig. 4.2は構築した時間オントロジーである。時間オントロジーでは、時刻 (time point)、期間 (time interval)、および時間間隔 (time duration) という3種類の概念を定義している。

時刻概念は時間軸上のある一点の座標を表す。時刻の記述のため、年、月、日、時、分、秒を属性として持つ。

期間概念は始点と終点を持つ時間軸上のある範囲を表す。始点と終点は時刻概念のインスタンスとして記述する。

時間間隔概念は時間の長さを表す。時間間隔の記述には年、月、日、時、分、秒を指定する。時間間隔の各属性間には繰り上げの関係があり、同じ時間間隔に対して複数通りの記

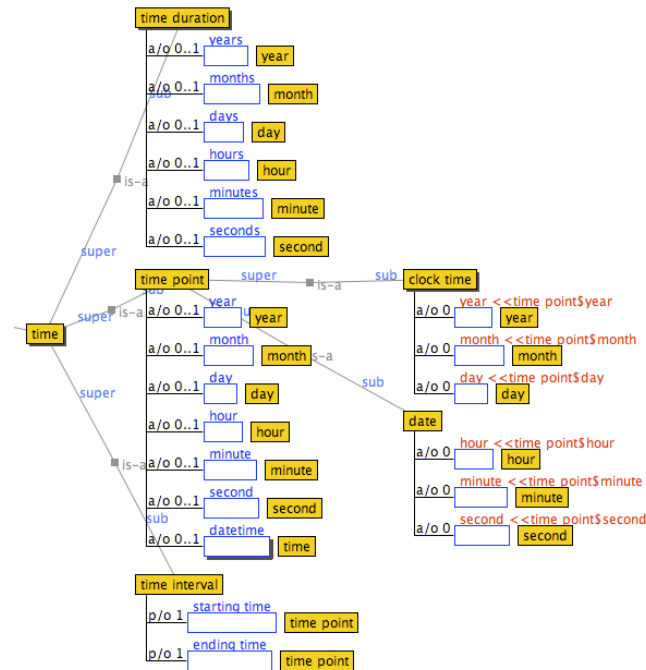


Fig. 4.2: 時間オントロジー

述が存在する。

4.3.3 幾何オントロジー

Fig. 4.3 は構築した幾何オントロジーである。幾何オントロジーでは、主に 2 次元形状概念を定義しており、点 (point)、線 (line)、多角形 (polygon)、円 (circle) などが含まれる。これらの概念における幾何形状は属性 "as WKT" にて、Well-known text (WKT) 形式 [48] のテキストとして記述される。

4.3.4 観測センサオントロジー

Fig. 4.4 は構築した観測センサオントロジーである。観測センサオントロジーは天体 (astronomical object)、プラットフォーム (platform)、およびセンサ (sensor) という 3 種類の概念が含まれる。

天体概念は観測センサが観測対象とする天体の概念を定義したものであり、地球観測の場合には惑星概念 (planet) のインスタンスである地球 (earth) が観測対象となる。

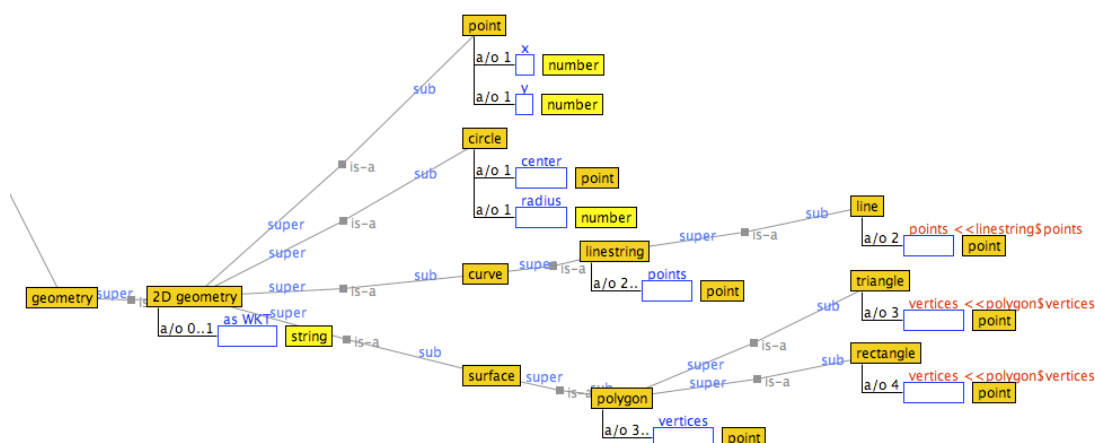


Fig. 4.3: 幾何オントロジー

プラットフォーム概念は観測センサが搭載される地球観測衛星概念（earth observation satellite）や航空機概念（airplane）が含まれる。

センサ概念は観測センサを定義するものであり、その観測原理によって様々な観測センサの下位概念が定義される。観測センサの概念体系は一般的な観測センサの分類に基づき定義されている [7]。

4.3.5 関係オントロジー

Fig. 4.5 は構築した関係オントロジーである。関係オントロジーには基本的関係概念（basic relation）、量的関係概念（quantitative relation）、幾何学的関係概念（topological relation）、時間的關係概念（temporal relation）、変化関係概念（changed）、対応関係概念（corresponding with）という6種類の関係概念およびそれらの下位概念が定義される。

基本的関係概念は概念同士の関係や概念とインスタンスの関係、インスタンス同士の関係にある基本的な関係を定義したものであり、同一性関係概念（sameAs）や非同一性関係概念（different）、概念とその下位概念のインスタンスの関係概念（subinstance of）などがある。

量的関係概念は数値同士の関係を定義したものであり、同値関係概念（equal）や非同値関係概念（not-equal）、大小関係概念（larger-than）などがある。

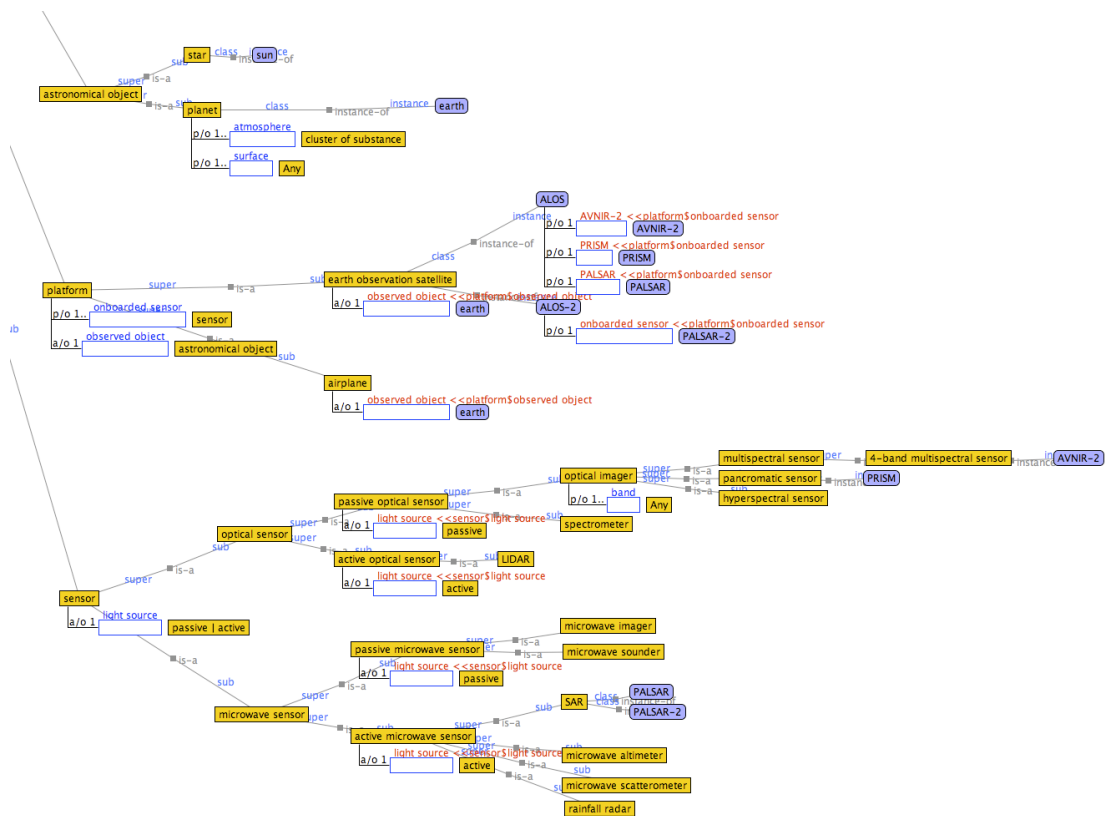


Fig. 4.4: 観測センサオントロジー

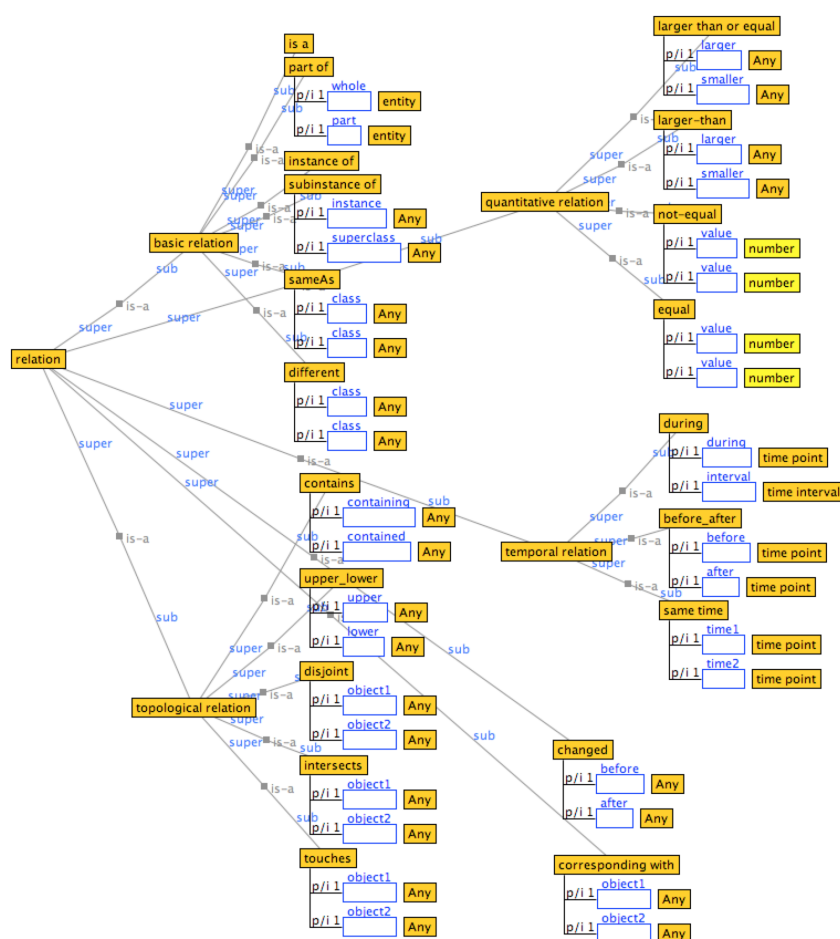


Fig. 4.5: 関係オントロジー

幾何的關係概念は物体や幾何形状同士の間幾何的關係を定義したものであり、包含關係概念 (contains) や交差關係概念 (intersects) などがある。

時間的關係概念は2つの時間の間關係を定義したものであり、ある期間内を表す關係概念 (during) や時間的前後關係概念 (before after) などがある。

また、変化關係概念はある2つのインスタンスの間に変化が生じていることを表し、対応關係概念はある2つのインスタンスの間に対応關係があることを表すものである。

4.3.6 データオントロジー

Fig. 4.6は構築したデータオントロジーである。データオントロジーは主に形式 (forms)、内容 (contents)、タイムスタンプ (timestamp) という3つの属性を持つ。形式にはその

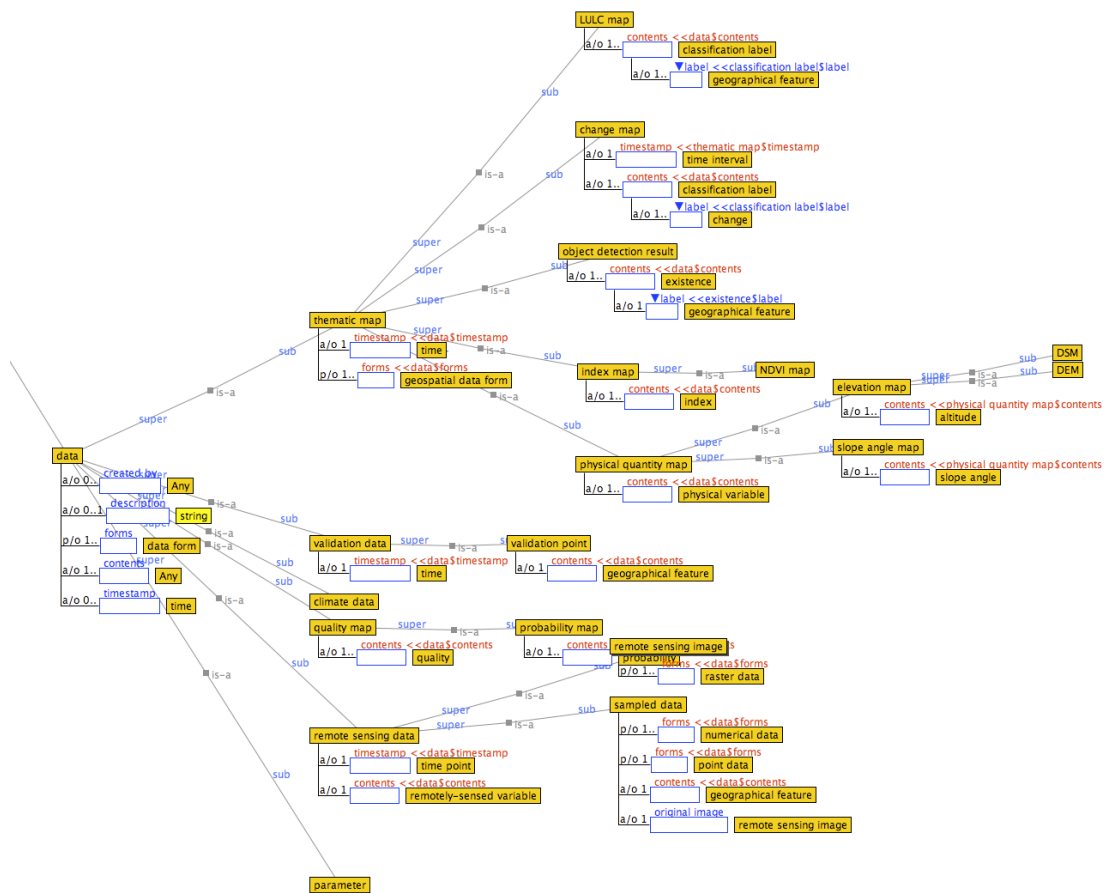


Fig. 4.6: データオントロジー

データが点データとして記述されたデータなのか，画像として記述されたデータなのかといった形式的な情報が記述される．形式的記述には Fig. 4.7 に示したデータ形式オントロジーの概念を用いられる．内容にはそのデータが地表面反射率を記述したデータなのか，土地被覆クラスラベルを記述したデータなのかといった意味的な情報が記述される．内容の記述には Fig. 4.8 に示した変数オントロジーの概念が用いられる．タイムスタンプには衛星データの観測時刻や土地被覆分類図生成に利用した情報の観測日時というように，そのデータが記述している対象の時間に関する情報が記述される．タイムスタンプの記述には Fig. 4.2 で示した時間オントロジーの概念が用いられる．

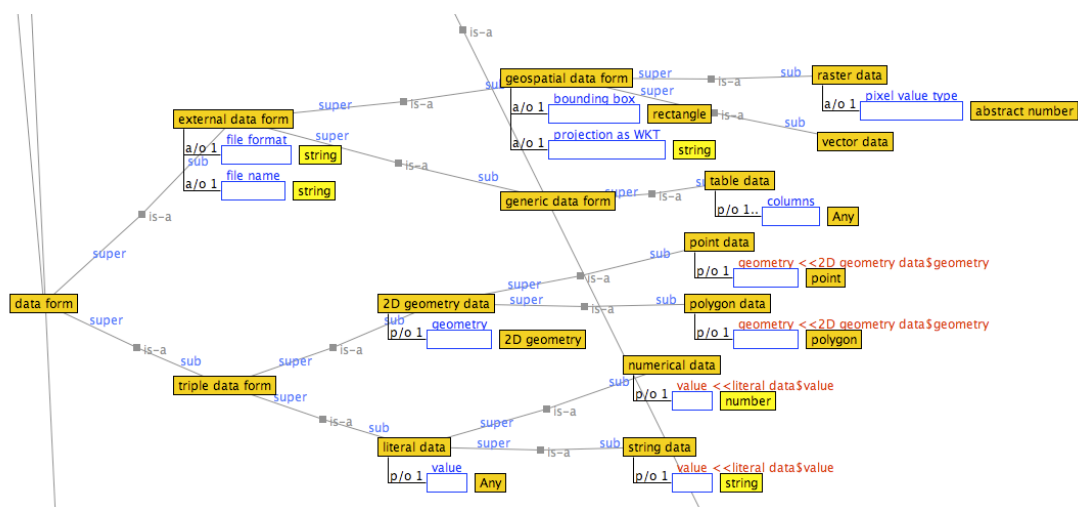


Fig. 4.7: データ形式オントロジー

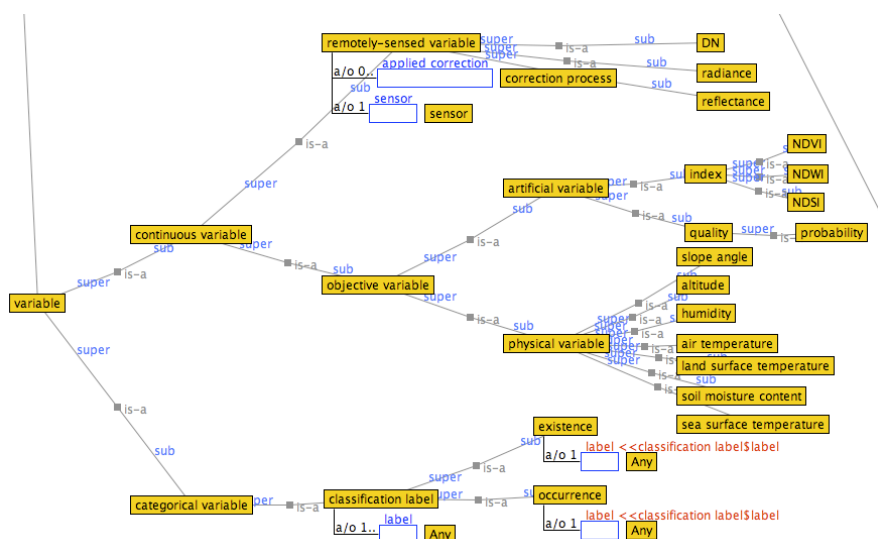


Fig. 4.8: 変量オントロジー

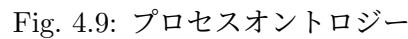


Fig. 4.9は構築したプロセスオントロジーである。プロセスオントロジーの定義は、WebサービスのRDFメタデータ記述のための枠組みであるOWL-SにおけるProcess概念の定義内容に基づいている[49]。プロセスオントロジーは主に入力データ(input)、出力データ(output)、前提条件(precondition)、結果(result)という4つの属性を持つ。

データストアやナレッジストアに置かれているデータやプロセスには、それぞれの意味が記述された RDF データがセットで用意される。システムは利用する全てのデータストアやナレッジストアへの参照を予め保持し、利用時には全データストアにある RDF デー

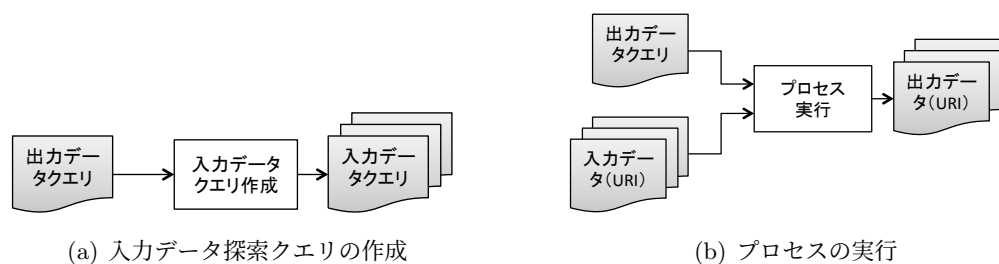


Fig. 4.10: プログラムモジュールの有する機能

タおよび全ナレッジストアにある RDF データを走査する。見つかった RDF データは合成され、仮想的な唯一のデータストア、唯一のナレッジストアが存在するかのように扱われる。そのため、個々のデータについてそれぞれ RDF データを用意するか、あるいは複数のデータについて記述した RDF データを用意するかについての規定はない。

ナレッジストアに格納されているプロセスは、実行可能なプログラムモジュールそのものか、あるいは API の形式で提供される。これらのプロセスは次の 2 つの機能の実装が必要となる。

1. 入力データ探索クエリの作成

出力予定のデータを表すクエリを入力し、そのデータを出力するために必要な入力データを推論し、必要な入力データのクエリのリストを返す。

2. プロセスの実行

出力予定のデータを表すクエリおよび必要な入力データの場所を示す URI のリストを入力し、プロセスを実行して得られた出力データをデータストアに格納し、その出力データの場所を示す URI のリストを出力する。

これらの機能は、次節で述べる解析ワークフロー自動構築・自動実行処理の中で利用される。

4.5 解析ワークフロー自動構築および自動実行のアルゴリズム

ユーザから受け取った目的の解析結果のデータを表すクエリを用いて、そのデータを得るための解析ワークフローを構築し、自動判読を行う処理のアルゴリズムについて説明す

る。提案するアルゴリズムでは、目的の解析結果を出力するために必要なデータおよびプロセスを、その解析結果をルートとした出力側からの深さ優先探索によって探索し、実行していく。Fig. 4.11 は解析ワークフローの自動構築・自動実行のアルゴリズムを表す擬似コードである。まとめると下記のようなアルゴリズムとなる。

Step 1:

クエリを用いてデータストアを探索する。該当するデータがマッチングしたならば、それを返す。

Step 2:

Step 1 で該当するデータが見つからなければ、クエリを用いてナレッジストアを探索し、該当するデータを出力可能なプロセスを調べる。

Step 3:

該当するデータを出力可能なプロセスがマッチングしたならば、そのプロセスに対してクエリを用いた入力データ探索クエリの作成を行い、入力データ探索クエリのリストを得る。

Step 4:

得られた入力データ探索クエリそれぞれについて Step 1 から行う。全ての入力データが揃ったならばプロセスを実行し、得られた出力データを返す。入力データが揃わなければ、空のデータを返す。

データストアの探索およびナレッジストアの探索では、クエリ側と被探索側の RDF グラフが適合するかどうかのマッチング処理が行われる。マッチングは次の 2 つの判定基準により評価される。

1. RDF グラフの構造が同型であるかどうか。
2. RDF グラフを構成するノードが同一のインスタンスを含み得るかどうか。オントロジーを用いた意味推論が行われ、関係概念による制約を考慮しつつクエリ側のノードと被探索側のノードが同じインスタンスを含むかどうかを判定する。

```

1 function ConstructWorkflow(query_for_output_data)
2     found_data_list = SearchData(query_for_output_data)
3     if found_data_list.length > 0 then
4         return SelectOneOf(found_data_list), Workflow()
5     else
6         found_process_list = SearchProcess(query_for_output_data)
7         if found_process_list.length > 0 then
8             available_output_data_list = List()
9             available_workflow_list = List()
10            for each process in found_process_list do
11                query_list_for_input_data = process.CreateInputQuery(query_for_output_data)
12                input_data_list = List()
13                workflow_for_all_input_data = Workflow()
14                for each query in query_list_for_input_data do
15                    input_data, workflow_for_input_data = ConstructWorkflow(query)
16                    if input_data is not None then
17                        input_data_list.Add(input_data)
18                        workflow_for_all_input_data.AddNextProcess(workflow_for_input_data)
19                    else
20                        continue to next process
21                    end if
22                end for
23                output_data = process.Execute(query_for_output_data, input_data_list)
24                available_output_data_list.Add(output_data)
25                available_workflow_list.Add(workflow_for_all_input_data.AddNextProcess(process))
26            end for
27            return SelectOneOf(available_output_data_list, available_workflow_list)
28        else
29            return None, None
30        end if
31    end if
32 end function

```

Fig. 4.11: 解析ワークフロー構築処理および自動実行のアルゴリズムを表す擬似コード

ここで 2 番目のオントロジーを用いた意味推論が行われることにより、クエリやデータ、プロセスの記述の中で関係概念による制約によって範囲を持たせた記述を行っていたものでも、マッチングを行うことが可能となる。これにより、例えば観測センサの種類をクエリに指定して具体的な観測センサによって得られた衛星データをマッチングさせることや、様々な問題に対して汎用的に適用可能なプロセスをシステムで利用することが可能となり、記述されたデータやプロセスに共有性や再利用性を持たせることができるようになる。

データストアの探索ではクエリとデータストアにある各データについて、クエリと被探索側のデータが有するデータ概念の `contents` 属性および `timestamp` 属性、`geometry` 属性のマッチング判定を行い、マッチしたデータを返す。同様にナレッジストアの探索ではクエリとナレッジストアにある各プロセスの出力データについて、出力データの持つ `contents` 属性のみマッチング判定を行い、マッチしたプロセスを返す。

提案するアルゴリズムはシンプルなものとなっており、現時点で未解決の課題が残されている。データストアやナレッジストアの探索時に何もマッチングしなかった場合、もしくは複数の候補がマッチングした場合が考えられるが、現時点ではこのときに特別な処理を行うことはなく、何もマッチングしなかった場合はその先の解析ワークフロー構築が打ち切られ、複数の候補がマッチングした場合はその全ての組合せについて解析ワークフロー構築が継続される。また、プロセスには本研究で扱うような入力データが揃った段階で実行できるプロセスの他に、衛星データ中のコンテキストのように何らかの条件が揃った場合にのみ適用できるプロセスも考えられる。これらの場合についてどう対処するかはさらなる検討が必要と考える。

4.6 本章のまとめ

本章では衛星データ解析に登場するデータやプロセスを機械可読な形で表現・蓄積し、それらを組み合わせた解析ワークフローを自動構築するための枠組みと、その枠組みに基づく自動判読システムの構成を提案した。

データやプロセスの意味情報記述の基盤として、Semantic Web 実現のために提案されたメタデータ記述の枠組みである RDF、および本研究で新たに構築した 6 種類のオント

ロジー（時間オントロジー、幾何オントロジー、観測センサオントロジー、関係オントロジー、データオントロジー、プロセスオントロジー）を採用し、データおよびプロセスの表現・蓄積の枠組みを提案した。

自動判読システム構築のため、データストア・ナレッジストアを持ったシステム構成および解析ワークフローの自動構築・自動実行のアルゴリズムを提案した。提案するアルゴリズムでは、目的の解析結果を出力するために必要なデータおよびプロセスを、その解析結果をルートとした出力側からの深さ優先探索によって探索し、実行していくものとなっている。途中の探索においてはオントロジーを用いた意味推論によるマッチングが行われ、それにより、例えば観測センサの種類をクエリに指定して具体的な観測センサによって得られた衛星データをマッチングさせることや、様々な問題に対して汎用的に適用可能なプロセスをシステムで利用することを可能とした。また、提案するアルゴリズムにおける未解決課題についても言及した。

提案手法による実現内容は下記のようにまとめられる。

- RDFによりデータやプロセスの意味を意味ネットワークとして形式的に記述する仕組み、およびデータやプロセスのオントロジーによりデータやプロセスの意味を一貫性を持って記述・解釈するための仕組みを提供することにより、データやプロセスの意味記述の機械可読性を実現した。
- 解析ワークフロー構築におけるデータやプロセスの探索時にオントロジーを用いた意味推論を行うことで、記述されたデータやプロセスを様々な問題に対して繰り返し利用可能とすることができ、データやプロセスの共有性や再利用性を実現した。
- RDFのトリプル形式による分散的なデータ構造やURIによる一意性を保証した参照方式を活用したデータストアやナレッジストアの構造、およびデータやプロセスの探索方法を考案し、記述されたデータやプロセスのモジュール性や分散管理性を実現した。
- RDFのトリプル形式によるデータ構造を活用し、データスキーマと記述されたデータそのものの依存関係を排した意味情報記述が可能となり、スキーマが更新されたとしても記述されたデータの永続性を保つことができるという柔軟性を実現した。

本章で構築した自動判読システムの検証実験結果は7章で言及する.

第5章 AVNIR-2 データを用いた災害被害域 の推論

5.1 実験の概要

本章では、2章で提案した確率的推論手法の検証実験結果を示す。2章で提案した変化検出システムを土石流跡の検出と洪水による冠水域の検出という2種類の災害状況把握解析事例に対してシステムに変更を加えずに適用し、システムが自動的に知識モジュールを選択してそれぞれの事例に適した情報抽出および推論を行った過程と得られた検出結果を示し、その精度検証を行う。変化検出の入力データはAVNIR-2によって観測された地上分解能10m、4チャンネルマルチスペクトルデータであり、オルソ補正済プロダクトに6SVコード[50]の標準大気モデルによる大気補正を施すことで地表面反射率画像に変換した。変化検出に用いられる変化前後の2時期画像同士は、剛体変換による画像レジストレーションによって0.5ピクセル(RMS)以下の幾何誤差で整合させた。

変化検出結果の精度検証を行うために、提案手法による変化検出結果として得られた確率画像を確率のしきい値処理によって2値化した後、変化検出結果の基準として事例ごとに用意した参照データとの比較を行った。確率のしきい値を5%刻みで変化させ、参照データで示された変化が検出できていた割合を表すPA、検出できていた変化が参照データでも変化とされていた割合を表すUA、そして検出結果と参照データが一致した割合を表すOAの3種類の精度がどのように変化するかを確認した。

本章で示す検証実験結果は著者らによる研究業績[10]にて報告したものである。

5.2 実験で利用した知識モジュール

本実験では、次の4種類の知識モジュールを用意した。

1. substance カテゴリに属する各概念を認識するための知識モジュール

この知識モジュールは、各ピクセルにおける地表面反射率の値から、substance カテゴリに属する概念の生起事象の事後確率を求めるピクセルベースの情報抽出および推論を行う。知識モジュールには概念の生起事象の事前分布と、その事象によって条件付けられた地表面反射率の分布が記述される。生起事象の事前分布にはベルヌーイ分布、地表面反射率の条件付き分布には多変量ガウス分布を仮定し、substance 概念のラベルが付けられた 1,900 点の地表面反射率データをトレーニングセットとして用いてそれぞれの分布のパラメータの最尤推定を行った。トレーニングセットは本実験に使用される衛星データを含む複数の衛星データから目視判読により地表面反射率データの抽出およびラベル付けを行うことで作成した。トレーニングセットは本実験に使用される衛星データを含む複数の衛星データから目視判読により抽出してラベル付けを行った。推定された確率分布を用いることで、それぞれのピクセルにおける各 substance 概念の事後確率がベイズの定理から求められる。例えば、あるピクセルの地表面反射率が x であったとき、そのピクセルにおける soil の事後確率は次のように求められる。

$$p(\text{soil}|x) = \alpha \cdot p(\text{soil})p(x|\text{soil}) \quad (5.1)$$

ここで、 α は正規化定数である。

2. cluster を認識するための知識モジュール

この知識モジュールは、概念の推論結果の確率画像のセグメンテーションによりセグメントを生成し、cluster の推論結果としての各セグメントの平均確率、および特徴量をそれぞれ出力するオブジェクトベース画像解析を行う。セグメンテーションには富田らが提案した反復型領域成長法 [51] を用いる。領域成長法は、画像の各ピクセルをセグメントの初期形状とし、画像全体を隣接ピクセル間の接続性判定をしきい値処理によって行いながらスキャンすることで、領域を拡張していくアルゴリズムである。反復型領域成長法は領域成長法の改良型アルゴリズムであり、領域成長法を、接続性判定のしきい値を徐々に緩めながら反復することで、値がなだらか

に変化している境界でも分離し易くなるという特徴がある。

3. slope を認識するための知識モジュール

この知識モジュールは、各ピクセルとその周辺の DEM/DSM の値から斜面傾斜を算出し、その値をもとに slope の確率画像を推論する。slope の確率は、slope の生起事象の事前分布および slope の生起事象によって条件付けられた斜面傾斜の分布と各ピクセルにおける斜面傾斜の値からベイズの定理を用いて求められる。

4. mudslide に隣接する soil layer を mudslide として認識するための知識モジュール

この知識モジュールは、変化後画像の soil layer の推論結果を用いて、変化後画像の mudslide の推論結果を修正する。この知識モジュールは、本研究の mudslide は急な斜面で起こる土砂の変化として定義されているが、土砂が流出する過程で局所的に緩い斜面を通過している場合や、変化前画像に裸地であった場所を通過している場合には検出できないことがあるため、このような場合でも mudslide として認識するために利用される。mudslide の推論結果は、soil layer の推論結果を用いて次のように更新される。

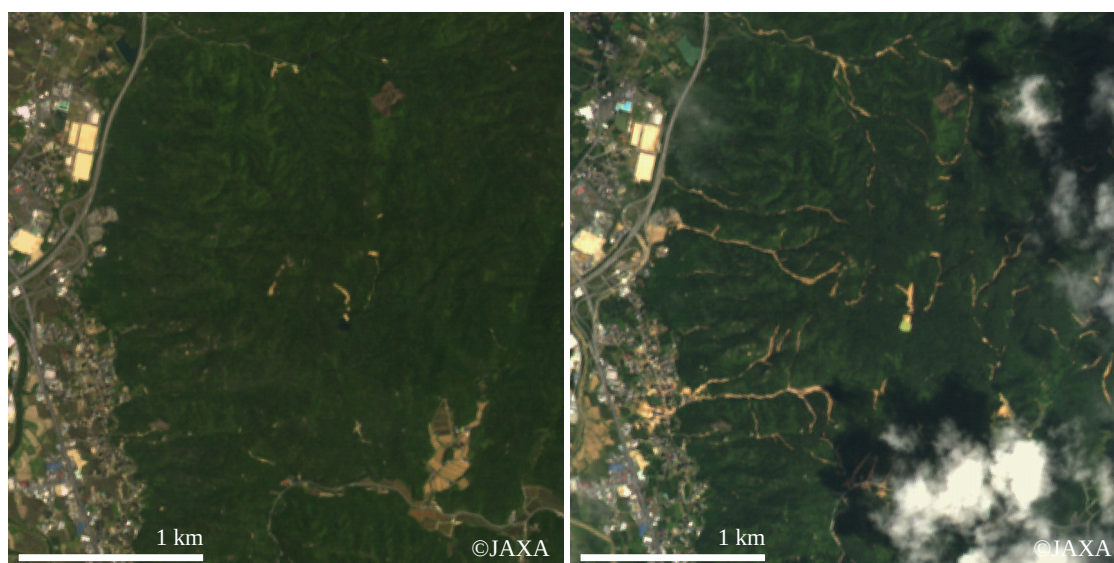
$$p(\text{mudslide})' = \max [p(\text{mudslide}), p(\text{soillayer}) \cdot \text{neighbor_max}\{p(\text{mudslide})\}] \quad (5.2)$$

ここで $\text{neighbor_max}(x)$ は隣接するセグメントの中で最大の x の値を出力する関数である。

5.3 事例 1：山口県における集中豪雨にともなう土石流跡の検出

5.3.1 解析事例

この事例では、2009 年 7 月 19 日から 7 月 26 日にかけて中国地方や九州北部の地域を襲った集中豪雨（気象庁命名「平成 21 年 7 月中国・九州北部豪雨」）により、山口県にて発生した土石流跡の検出を行う。変化検出用の 2 時期データには災害前の画像（Day 1）として 2009 年 6 月 14 日に観測された画像（Fig. 5.1(a)）、災害後の画像（Day 2）として



(a) 被災前 (2009 年 6 月 14 日)

(b) 被災後 (2009 年 7 月 30 日)

Fig. 5.1: 山口県における土石流発生前後の AVNIR-2 データ

AVNIR-2 データの True Color 合成画像 (RGB はそれぞれバンド 3, バンド 2, バンド 1 の地表面反射率)

2009 年 7 月 30 日に観測された画像 (Fig. 5.1(b)) を使用し, その中の北緯 $34^{\circ}15'$, 東経 $131^{\circ}54'$ を中心とした 400×400 ピクセルの領域を解析に使用する. また, 斜面傾斜を求める際に使用される DSM として, ALOS 搭載のパンクロマチック立体視センサ (PRISM) のデータを用いて生成された DSM を使用する.

変化検出結果の比較のための参照データとして, 山口県庁による報告書 [52] で示されている航空写真判読や現地調査で確認された山地災害発生箇所 (Fig. 5.2) と, JAXA EORC の専門家による Day 2 画像における土石流跡の手動解析結果 (Fig. 5.3) [53] の 2 種類を用いる. 山口県庁による報告書では, 山地災害発生箇所が地図上に印をつけた形式で公開されているが, 衛星データ上の土石流跡のピクセルと対応しておらず精度評価ができないため, これをディジタイズしたものを精度評価のための参照データとして使用する. なお, 本実験での解析範囲に存在する山地災害発生箇所 120 箇所のうち, 雲に覆われていたり, 小規模であったために衛星データ上では確認できなかった 42 箇所は除外している. また, JAXA EORC の専門家による手動解析結果は狭い範囲でしか公開されていないため, 精度算出による比較は行わずに目視による比較を行う.

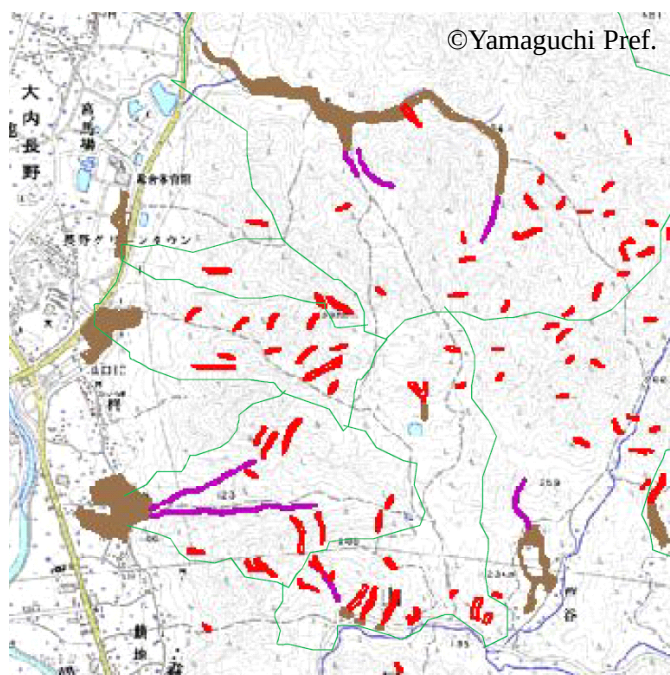


Fig. 5.2: 山口県庁による山地災害発生箇所の調査結果 [52]

赤色の領域は山腹被災地、紫色の領域は溪流被災地、茶色の領域は土砂流出区間を表す。

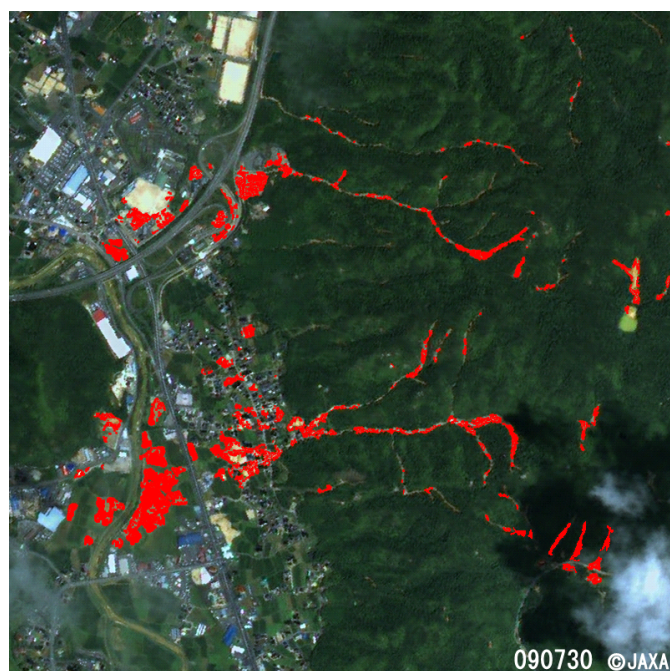


Fig. 5.3: JAXA EORC による山口県の土石流跡の手動解析結果 [53]

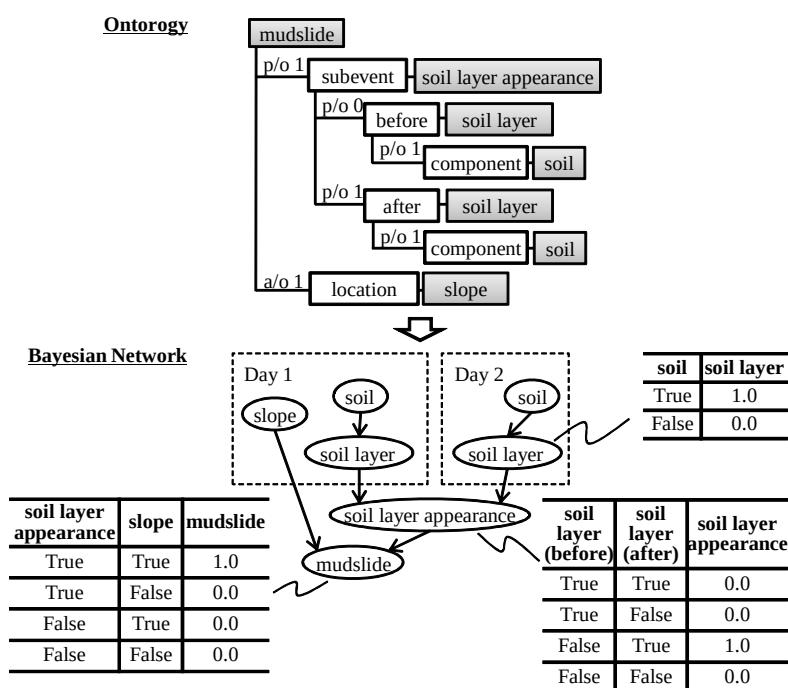


Fig. 5.4: 構築された mudslide 推論のためのベイジアンネットワーク

5.3.2 変化検出システムによる解析手順および解析結果

構築した変化検出システムに検出対象クエリとして土石流 “ mudslide ” を入力した場合の解析の過程および解析結果を示す。

土石流 “ mudslide ” を検出対象クエリとして入力すると、本提案システムはまず mudslide の定義をオントロジー中から探索し、それをもとに mudslide を推論するためのベイジアンネットワークを構築する。Fig. 5.4 はこのプロセスによって構築されたベイジアンネットワークを条件付き確率表とともに図示している。ここで、soil のノードには知識モジュール (1)、soil layer のノードには知識モジュール (2)、slope のノードには知識モジュール (3)、そして mudslide には知識モジュール (4) がそれぞれ関連付けられている。

その後、知識モジュールやベイジアンネットワークを用いた情報抽出および推論が行われ、Fig. 5.5, Fig. 5.6, および Fig. 5.7 に示す確率画像が得られる。比較のため、Fig. 5.7 には Fig. 5.1 で示した土石流発生前後の AVNIR-2 データを再掲する。

まず、知識モジュール (1) を用いた substance 推論の結果として soil の確率画像 (Fig. 5.5(a), 5.5(b)) が得られる。次に知識モジュール (2) を用いて soil の確率画像から soil

layer の確率画像 (Fig. 5.5(c), 5.5(d)) が得られる。さらに、2 時期の soil layer の確率画像から soil layer appearance の確率画像 (Fig. 5.6(a))、知識モジュール (3) を用いて DSM から slope の確率画像 (Fig. 5.6(b)) が得られる。そして最終的に、soil layer appearance の確率画像、slope の確率画像、知識モジュール (4) を用いて検出対象である mudslide の確率画像 (Fig. 5.7(c)) が得られる。

5.3.3 評価及び考察

Fig. 5.8 は、確率のしきい値を 5% ごとに変化させながら 2 値化した mudslide の確率画像と、山口県庁による参照データを比較して算出した PA, UA, および OA のグラフである。確率のしきい値を高くするほど PA は減少し、UA は増加する傾向が確認できる。検出結果として得られる確率は推論の信頼度を表すものであるため、確率のしきい値の増加と対応して UA が増加するという傾向は意図通りと考えられる。確率のしきい値が 50% のときに OA が最大値 99.0% をとり、その点での PA は 70.3%, UA は 68.5% となった。

Fig. 5.9 は土石流跡検出結果と山口県庁による山地災害発生箇所の差分を表すカラー合成画像である。本実験における誤検出の主な原因は人工物における 2 時期画像間のレジストレーション誤差の影響であった。人工物の中には土壌の地表面反射率との識別が難しいものがあり、レジストレーション誤差が山沿い等の斜面傾斜がある場所で発生していた場合に mudslide として誤検出されていた。Fig. 5.9 の左上に誤検出 (紫) された広い領域が確認できるが、これは領域の一部がレジストレーション誤差により mudslide として誤検出され、知識モジュール (4) による接続性判定で拡大されたことによるものである。この対策には、サブピクセルレベルでのレジストレーション精度の改善、ノイズ判定、人工物の幾何的なマッチング判定などが考えられる。

また、未検出の原因には次の 3 種類が確認された。まず 1 つめは、小規模で幅が狭い土石流や、大規模な土石流であっても境界部分では植生との混合被覆として観測されていたため、soil の確率が低くなって検出できなかったケースである。今回の実験で用いた被覆推定を行う知識モジュール (1) は混合被覆に対してロバストな認識ができていないため、混合被覆推定のための改善が必要である。2 つめは、雲影中であったために輝度が減

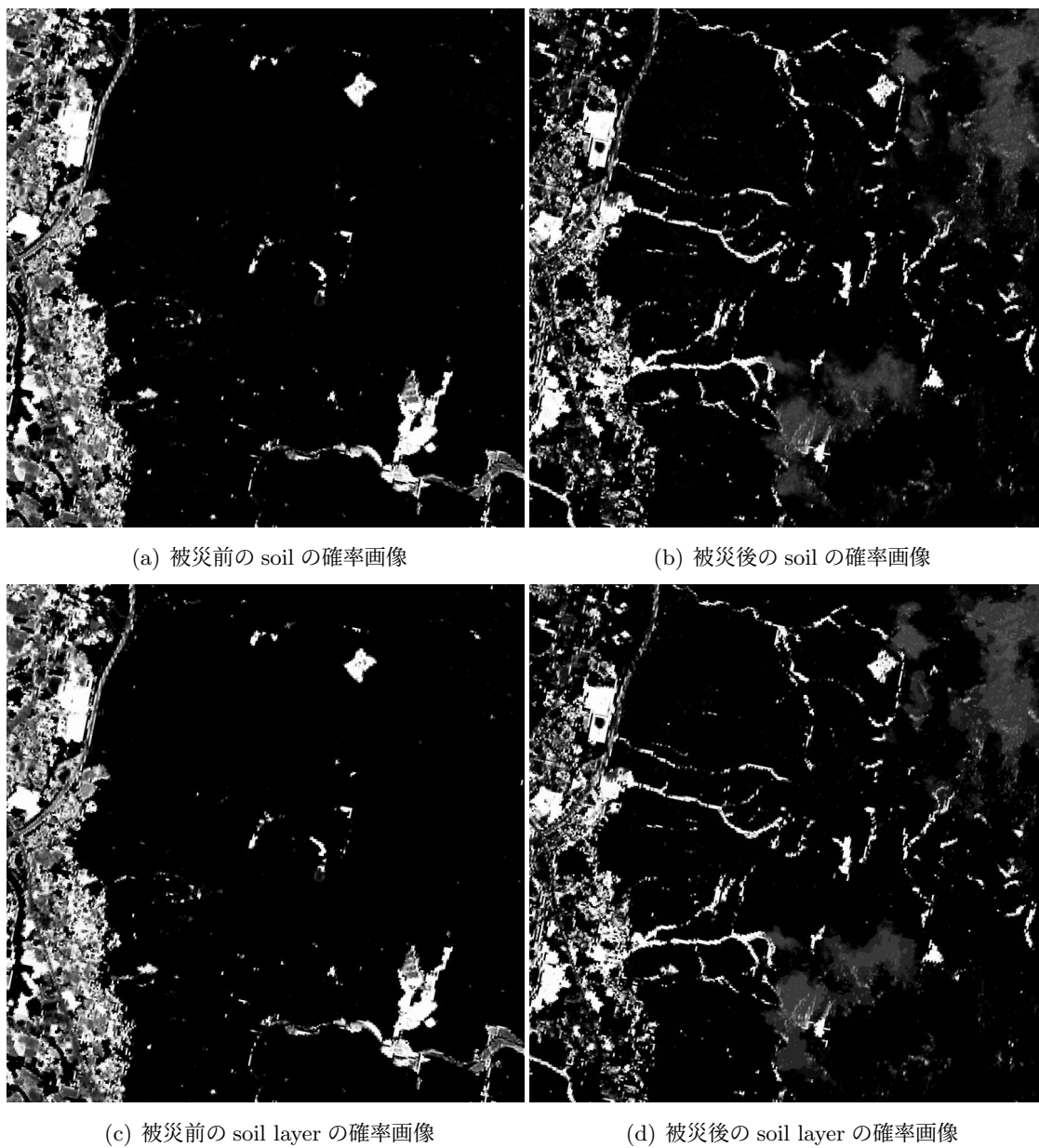


Fig. 5.5: 山口県の土石流跡の推論結果 1

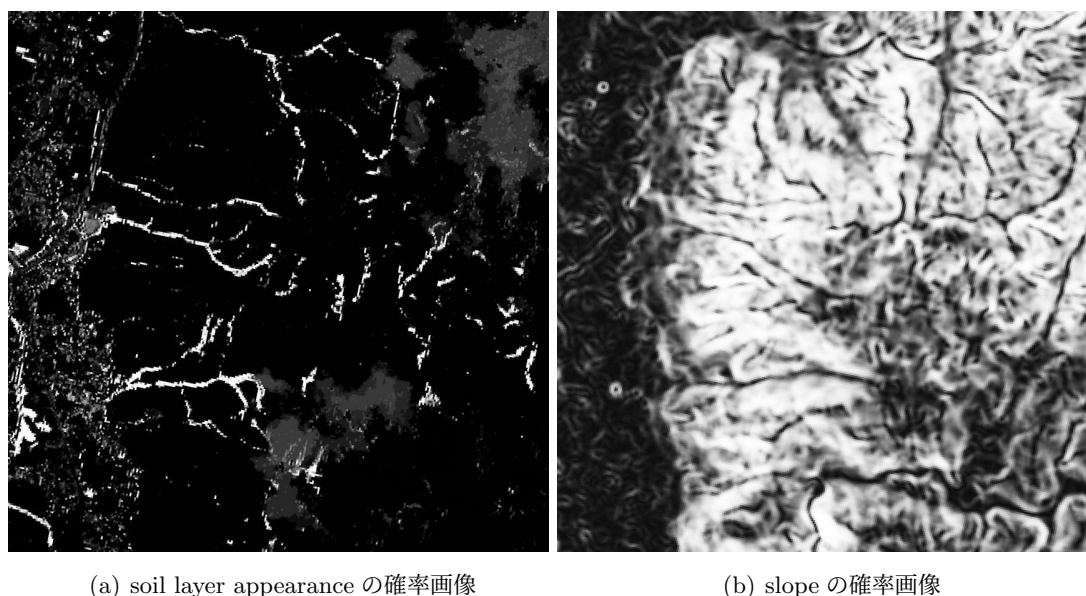
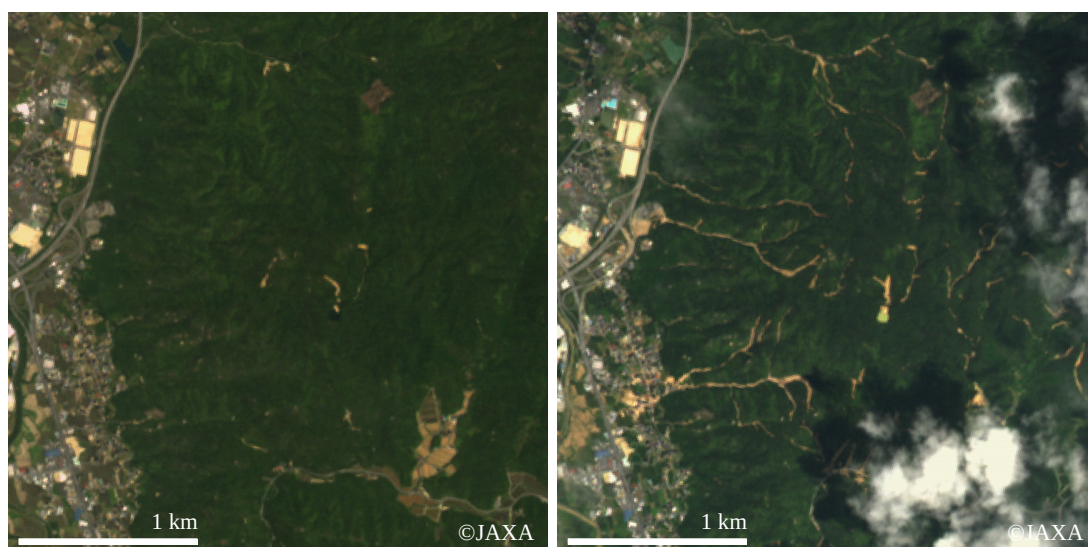


Fig. 5.6: 山口県の土石流跡の推論結果 2

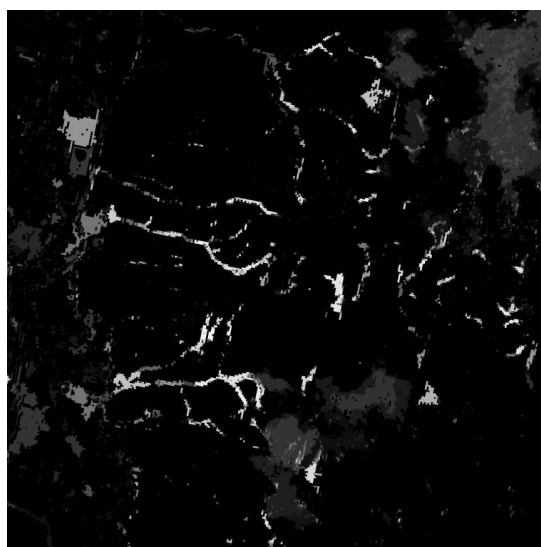
少し，soil の確率が低くなって検出できなかったケースである．今回の知識モジュールには雲影による影響を考慮に入れた処理は実装されていないため，雲との幾何的な位置関係から雲影を認識し，輝度の補正を行うといった知識モジュールの実装が必要と考える．そして3つめは，谷筋や山沿いの斜面傾斜が小さい場所に土砂が流出・堆積していた箇所を，mudslide として検出できなかったケースである．知識モジュール（4）では，斜面で発生した mudslide に隣接した soil layer は mudslide とする接続性判定を行なっているが，混合被覆や雲影によって soil layer が連続していない場合があった．混合被覆や雲影への対策を行う知識モジュールの実装や，連続性判定の改善が必要と考える．

EORC の専門家による手動解析結果との比較では，大規模な土石流では両者の結果は似通っていたが，小規模な土石流では専門家による手動解析結果で検出できていない箇所があることが確認された．専門家による手動解析結果は変化あり／なしの2値情報になっているため，小規模な変化の情報が切り捨てられてしまっていたが，本システムによる変化検出結果は土石流の推定結果の信頼度を連続値で表した確率画像であるため，小規模な変化であっても周囲の非土石流領域より確率が高くなっていることから判読が可能であった．また，山地に隣接していない平地部における土壌変化が，専門家による手動解析結果で検出されていたのに対し，本システムによる変化検出結果では他の土石流箇所よりも確



(a) 被災前 (2009 年 6 月 14 日)

(b) 被災後 (2009 年 7 月 30 日)



(c) mudslide の確率画像

Fig. 5.7: 山口県の土石流跡の推論結果 3 および土石流発生前後の AVNIR-2 データ (再掲)

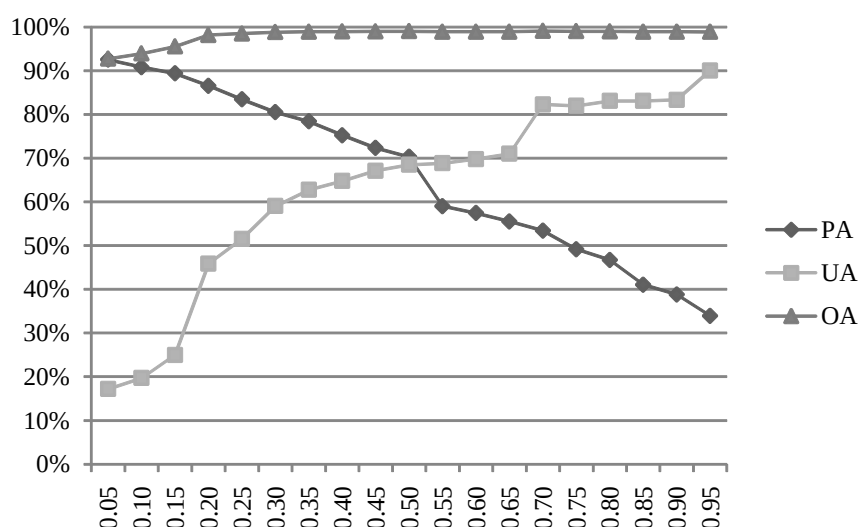


Fig. 5.8: 土石流跡検出結果精度の事後確率のしきい値変化に対する変化
横軸は事後確率のしきい値の値を表し、縦軸は土石流跡検出結果精度の値を表す。精度は山口県庁による山地災害発生箇所 [52] を基準として算出した。

率が低くなっていた。山地に隣接していない平地部における土壌変化であるため土石流であることを示す証拠が無く、土石流であるかどうかは専門家でも判断が分かれる事例であり、確率が低く推論されたと考察する。

5.4 事例2：パキスタンにおける豪雨による洪水冠水域の検出

5.4.1 解析事例

この事例では、2010年7月末からパキスタンで降り続いた季節的な豪雨により発生したインダス川の洪水の冠水域の検出を行った。変化検出用の2時期データには災害前の画像 (Day 1) として2009年10月14日に観測された画像 (Fig. 5.10(a))、災害後の画像 (Day 2) として2010年9月1日に観測された画像 (Fig. 5.10(b)) を使用し、その中の北緯 $30^{\circ}44'$ 、東経 $71^{\circ}54'$ を中心とした $2,000 \times 2,000$ ピクセルの領域を解析に使用した。

変化検出結果の比較のための参照データとして、JAXA EORC の専門家による Day 2 画像における冠水域の抽出結果 (Fig. 5.11) を用いた。

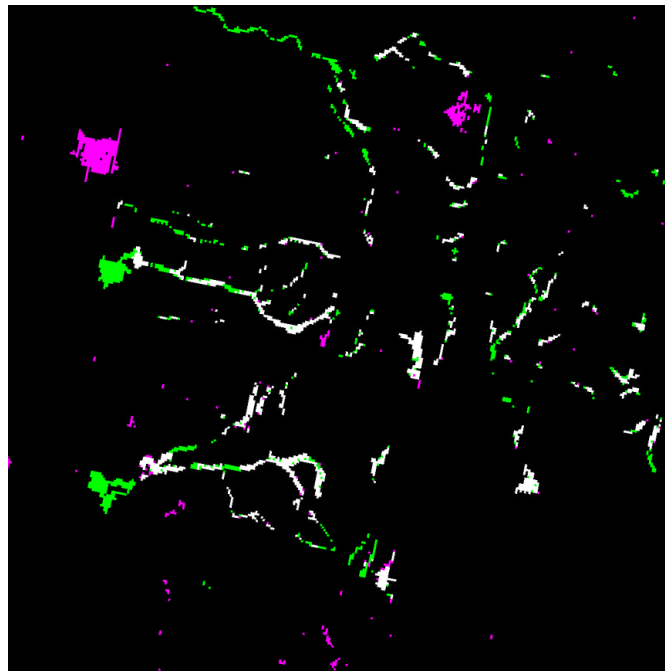


Fig. 5.9: 土石流跡検出結果と山口県庁による山地災害発生箇所の差分を表すカラー合成画像

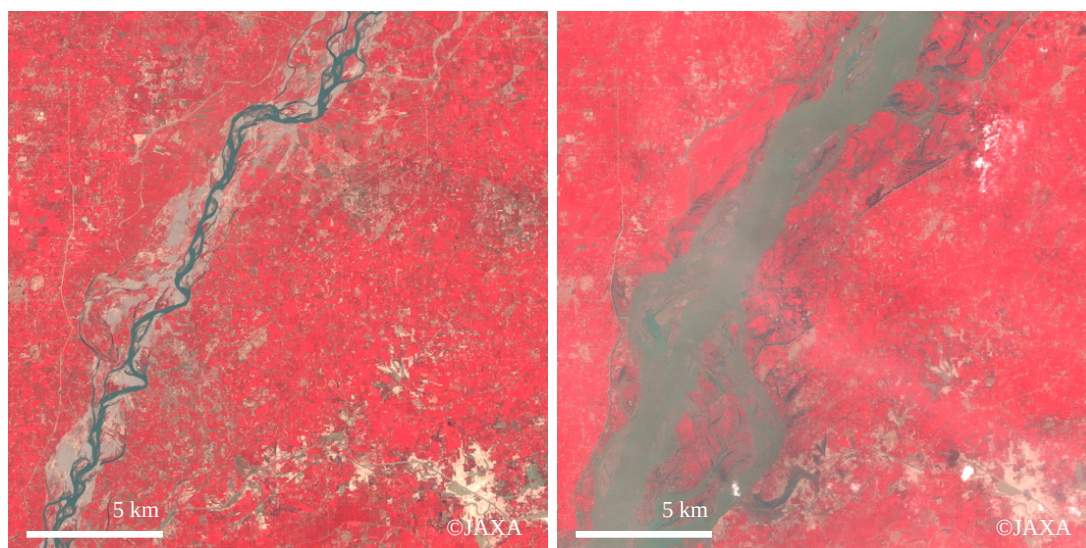
土石流跡検出結果において確率のしきい値を 50%とした場合の 2 値化画像と、山口県庁による山地災害発生箇所のカラー合成（R と B は提案システムによる土石流跡検出結果，G は山口県庁による山地災害発生箇所）を行った画像．図中の白い領域が両データで一致した変化箇所，緑の領域が未検出であった変化箇所，紫の領域が誤検出であった変化箇所をそれぞれ表している．

5.4.2 変化検出システムによる解析手順および解析結果

構築した変化検出システムに検出対象クエリとして水域の出現“ water layer appearance ”を入力した場合の解析の過程および解析結果を示す．

水域の出現“ water layer appearance ”をクエリして入力すると，システムはまず water layer appearance の定義をオントロジー中から探索し，それをもとに water layer appearance を推論するためのベイジアンネットワークを構築する．Fig. 5.12 はこのプロセスで構築されたベイジアンネットワークを条件付き確率表とともに図示している．ここで，water のノードには知識モジュール（1），そして water layer のノードには知識モジュール（2）がそれぞれ関連付けられている．

その後，知識モジュールやベイジアンネットワークを用いた情報抽出および推論が行われ，Fig. 5.13，および Fig. 5.14 に示す確率画像が得られる．比較のため，Fig. 5.14 には



(a) 被災前 (2009 年 10 月 14 日)

(b) 被災後 (2010 年 9 月 1 日)

Fig. 5.10: パキスタンにおける洪水発生前後の AVNIR-2 データ

AVNIR-2 データの False Color 合成画像 (RGB はそれぞれバンド 4, バンド 3, バンド 2 の地表面反射率)

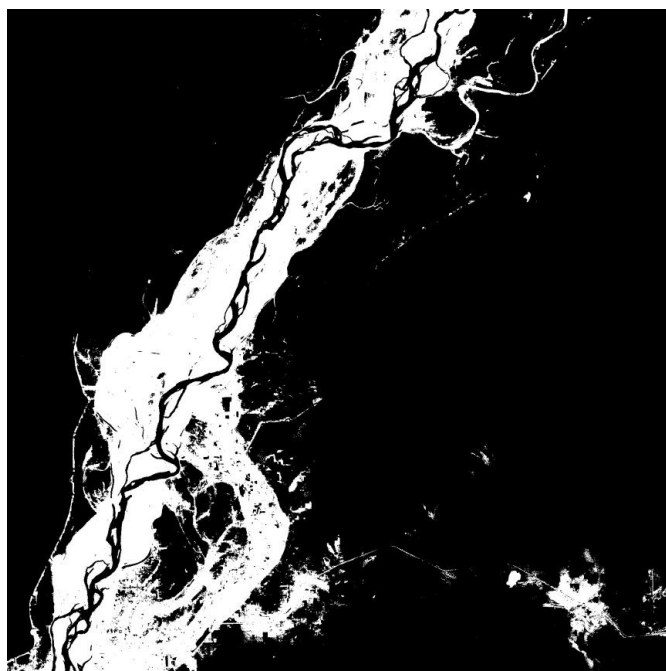


Fig. 5.11: JAXA EORC によるパキスタンの洪水冠水域の手動解析結果

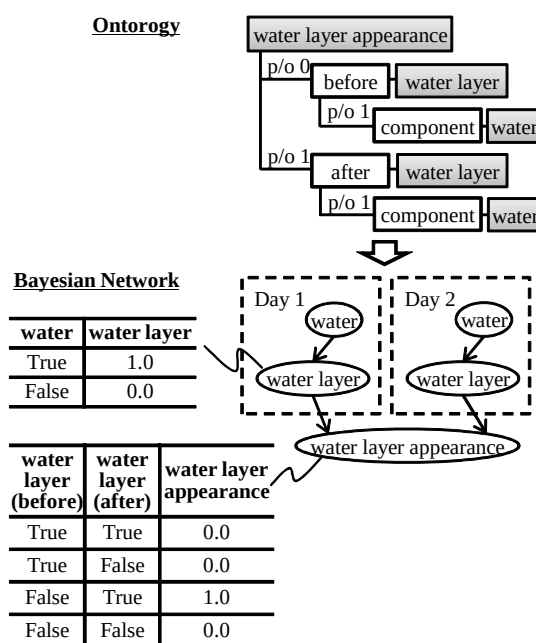


Fig. 5.12: 構築された water layer appearance 推論のためのベイジアンネットワーク

Fig. 5.10 で示した洪水発生前後の AVNIR-2 データを再掲する。

まず、知識モジュール (1) を用いた substance 推論の結果として water の確率画像 (Fig. 5.13(a), 5.13(b)) が得られる。次に知識モジュール (2) を用いて water の確率画像から water layer の確率画像 (Fig. 5.13(c), 5.13(d)) が得られる。そして、2 時期の water layer の確率画像から検出対象である water layer appearance の確率画像 (Fig. 5.14(c)) が得られる。

5.4.3 評価および考察

Fig. 5.15 は確率のしきい値を 5% ごとに変化させながら 2 値化した water layer appearance の確率画像と、専門家による解析結果を比較して算出した PA, UA, および OA のグラフである。事例 2 でも事例 1 と同様に、確率のしきい値を高くするほど PA は減少し、UA は増加する傾向が確認できる。また、確率のしきい値が 20% のときに OA が最大値 97.3% をとり、その点での PA は 91.3%, UA は 93.8% であった。事例 1 と異なる点として OA が最大値をとる点の確率のしきい値が低いことが挙げられる。これは、water layer appearance の確率値が全体的に低くなっていたためである。Day 2 での冠水域における水

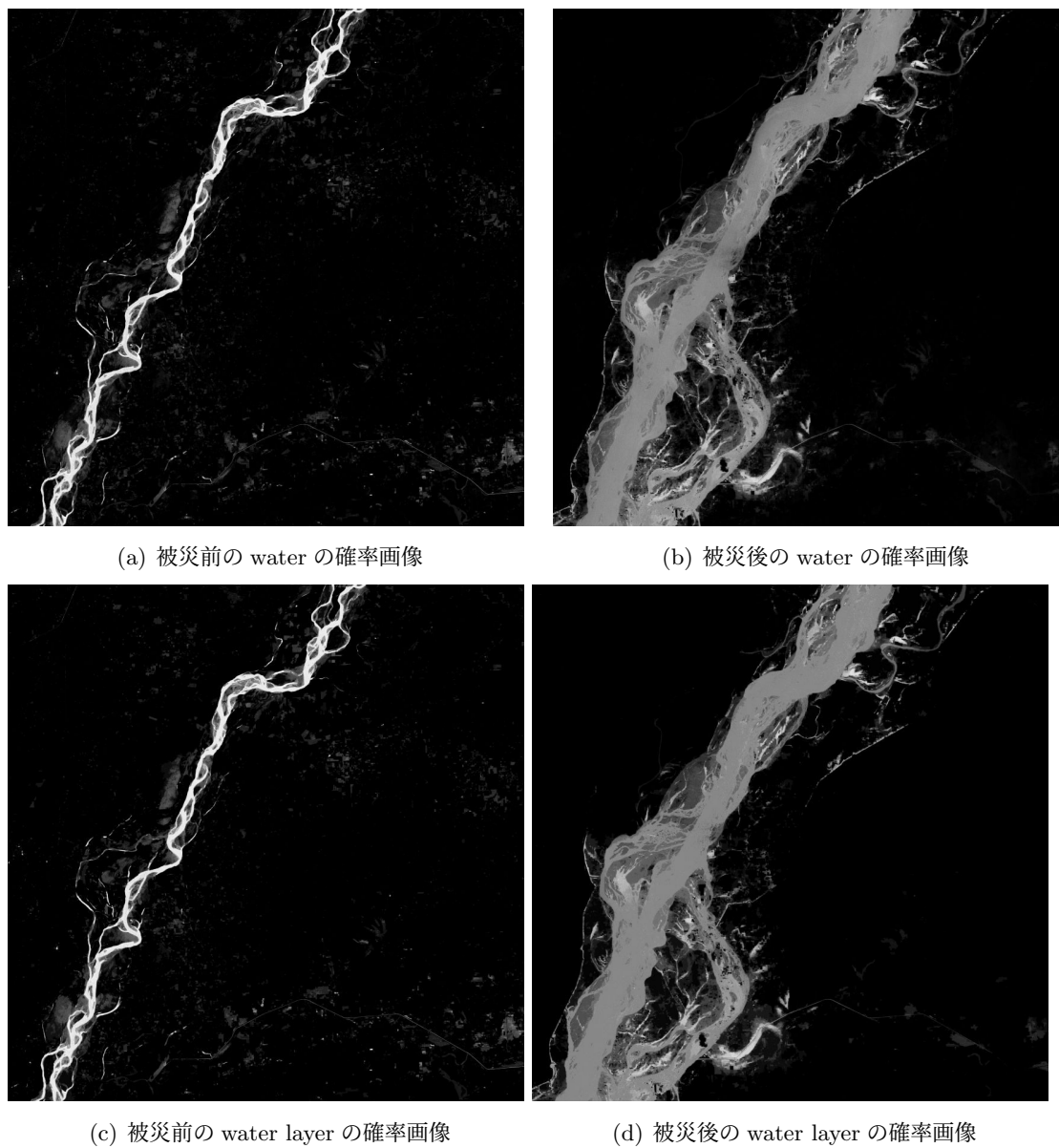
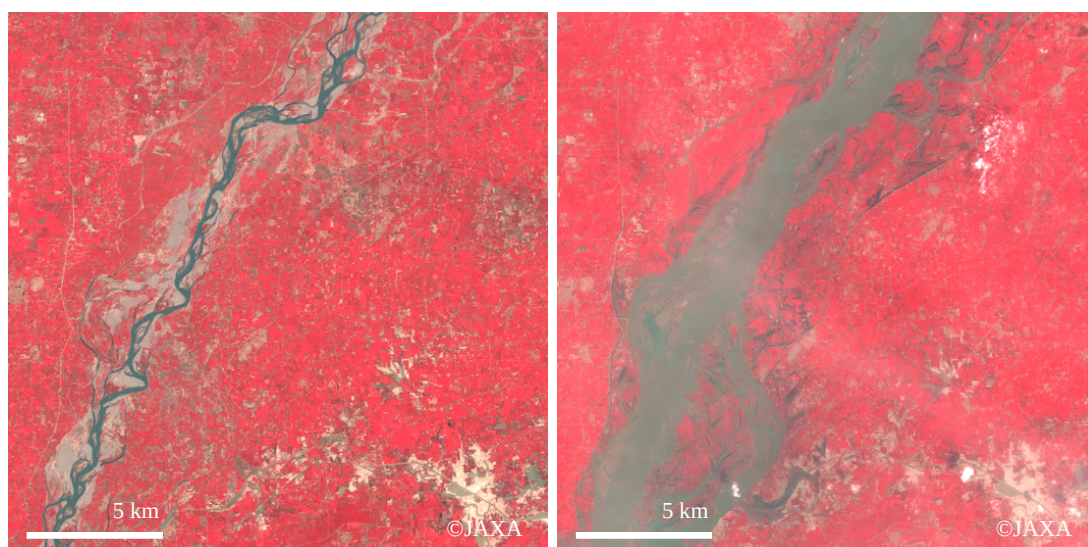
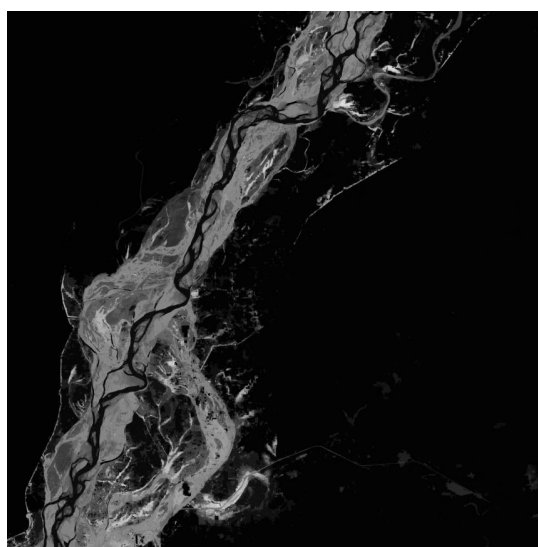


Fig. 5.13: パキスタンの洪水冠水域の推論結果



(a) 被災前 (2009 年 10 月 14 日)

(b) 被災後 (2010 年 9 月 1 日)



(c) water layer appearance の確率画像

Fig. 5.14: パキスタンの洪水冠水域の推論結果 2 および洪水発生前後の AVNIR - 2 データ (再掲)

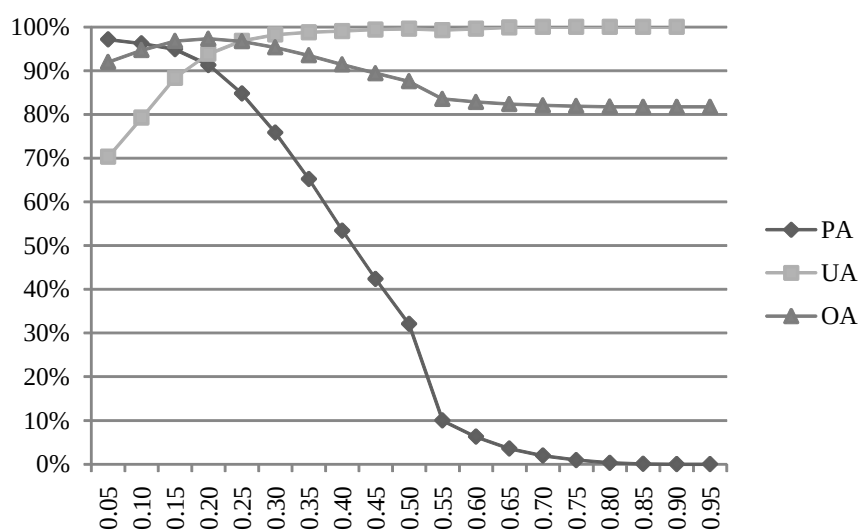


Fig. 5.15: 洪水冠水域検出結果精度の事後確率のしきい値変化に対する変化

横軸は事後確率のしきい値の値を表し、縦軸は洪水冠水域検出結果精度の値を表す。精度は JAXA EORC による洪水冠水域抽出結果を基準として算出した。

面が懸濁していたため、あるいは冠水域の水深が浅く底面の被覆が見えていたために、知識モジュール (1) による情報抽出で water の推定結果の信頼度が低かったことが原因である。しかし、冠水域における water の推定結果の信頼度は低くなっていたが、非冠水域における water の推定結果は更に低くなっており、Fig. 5.13(b) でも明確な違いが確認できる。この冠水域と非冠水域での water の推定結果の信頼度をベイジアンネットワークで伝播し、water layer appearance の推論に用いることができるため、water layer appearance の検出結果においても冠水域と非冠水域は高精度に分離できている。2 時期データの被覆分類後に非水域から水域へと変化した領域を抽出する従来の post-classification comparison では、この事例のように水域の分類が難しい場合には被覆分類時に間違った被覆に分類されることで精度の良い変化検出結果が得られない可能性が高く、専門家が被覆分類結果を確認する必要性が生じる。提案システムを用いることで、このような解析途中における判断を省き、最終的な解析結果の解釈のみを行うだけで済むようになるため、解析作業の負担を軽減することができる。

Fig. 5.16 は提案システムによる洪水冠水域検出結果と JAXA EORC による洪水冠水域抽出結果の差分を表すカラー合成画像である。図の右下に未検出 (緑) の広い領域が確認できるが、目視では冠水域は確認できないため、参照データが誤検出していたものと思わ

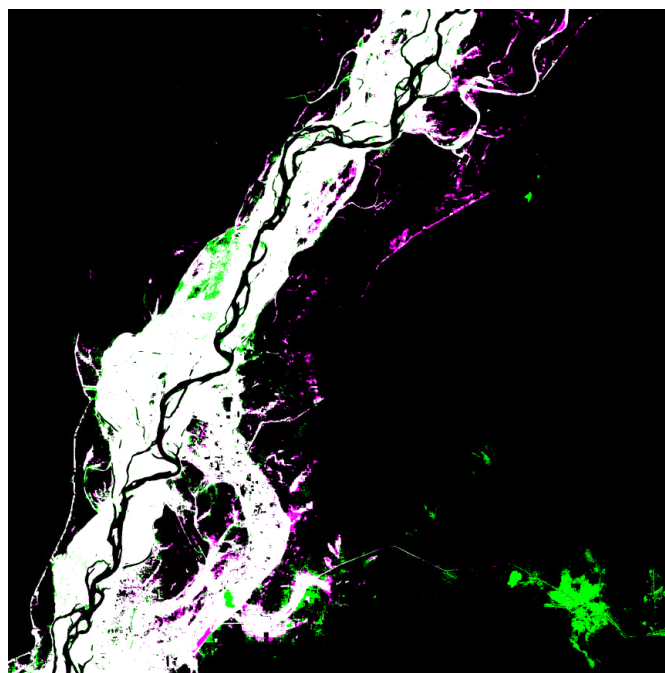


Fig. 5.16: 洪水冠水域検出結果と JAXA EORC による洪水冠水域抽出結果の差分を表すカラー合成画像

洪水冠水域検出結果において確率のしきい値を 20%とした場合の 2 値化画像と、JAXA EORC による洪水冠水域抽出結果のカラー合成（R と B は提案システムによる洪水冠水域検出結果，G は JAXA EORC による洪水冠水域抽出結果）を行った画像．図中の白い領域が両データで一致した変化箇所，緑の領域が未検出であった変化箇所，紫の領域が誤検出であった変化箇所をそれぞれ表している．

れる．両データで検出された洪水の冠水域は局所的な違いはあるものの，概ね一致していることが確認できた．

5.5 本章のまとめ

本章では，2 章で提案した変化検出システムを土石流跡の検出と洪水による冠水域の検出という 2 種類の災害状況把握解析事例に対してシステムに変更を加えずに適用し，システムが自動的に知識モジュールを選択してそれぞれの事例に適した情報抽出および推論を行った過程と得られた検出結果を示し，その精度検証を行った．

実験では，土石流跡と洪水冠水域の検出という 2 種類の変化検出アプリケーションについて，土石流と冠水域というそれぞれの概念の意味をオントロジーによって解釈し，適用可能な知識モジュールを選択して自動的に情報抽出および推論が行えることを示した．ま

た、土地被覆の推論結果を一つに絞る事が困難な場合であっても、推論結果の信頼度を伝播させて次の推論を行うことができるため、ロバストな変化検出が行えることを示した。

変化検出結果の高精度化のためには、混合被覆の推定、大気補正の精度や雲影による影響への対策、土石流跡の連続性判定の改善、サブピクセルレベルのレジストレーション誤差への対策といった知識モジュールの実装の課題があることが実験で確認された。よりロバストかつ高精度な変化検出には、これらの課題の解決が必要である。

第6章 AVNIR-2データを用いた日本周域における土地被覆分類図作成と精度評価

6.1 実験の概要

本章では、3章で提案した土地被覆推論手法の検証実験結果を示す。土地被覆分類の入力データはAVNIR-2によって観測された地上分解能10 m、可視・近赤外域4チャンネルのマルチスペクトルデータとし、日本周域を観測した雲域の割合が少ない1,876シーンのデータそれぞれに対して大気斜面補正およびオルソ幾何補正を施したデータセットを用いる。本実験では分類する土地被覆クラスは8クラスとし、衛星データの観測値とクラスの対応を示す教師データの取得には日本全国に分布する2,505点の検証点データを利用する。

まずAVNIR-2データセットから日本周域の高精度土地被覆分類図を作成し、その結果を示す。その後、作成した土地被覆分類図の精度評価として検証点データを用いた交差検定の結果を示す。精度の比較のため、最尤分類法を用いた土地被覆分類手法(MLC)および1対1ソフトマージンサポートベクターマシンを用いた土地被覆分類手法(SVMs)による同様の交差検定の結果も併せて示す。

本章で示す検証実験結果は著者らによる研究業績[14]にて報告したものである。

6.2 使用データおよび解析エリア

ALOSに搭載された光学センサAVNIR-2による観測データを用いて土地被覆分類を行った。解析には2006年から2011年の5年間に日本周域を観測し、雲域の割合が少ない1,876シーンのAVNIR-2データを利用した。これらのデータは大気斜面補正およびオルソ幾何補正が施された地表面反射率画像である[25]。Fig. 6.1は使用したデータセットにおける各地点での5年間のトータルでの観測頻度を、Fig. 6.2は月ごとの観測頻度をそれぞれ表し

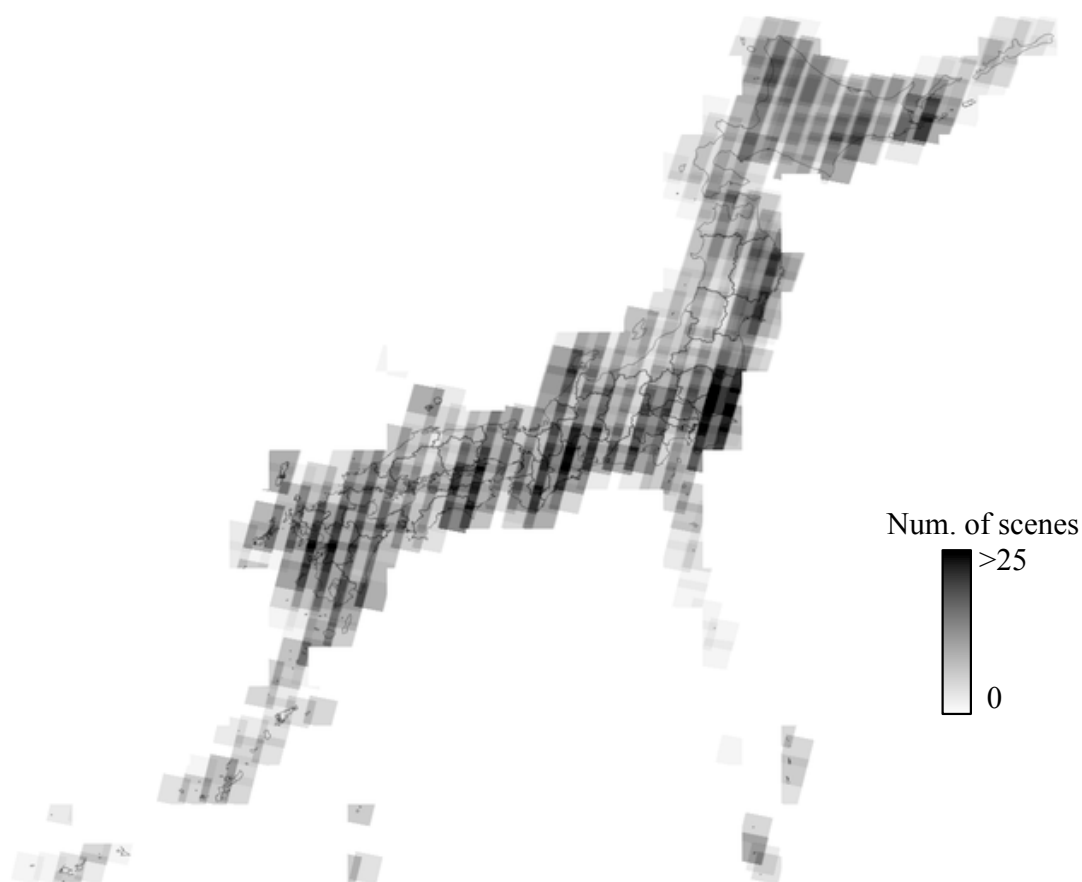


Fig. 6.1: 日本周域における AVNIR-2 画像の通算取得数

ている。広域高分解能データのトレードオフとなっている年間の観測頻度の少なさや、観測範囲のオーバーラップ、そして地域ごとの被雲率の影響などが起因して、使用するデータセットには観測頻度の偏りが多く含まれていた。例えば関東地方では、梅雨により雲に覆われることの多い6-7月の観測頻度が少なく、日本海側から山越えして乾燥した季節風が吹くことで晴天が続く12-3月の観測頻度が多くなっていた。

6.3 日本周域における土地被覆分類図作成

6.2 節で述べた衛星データ、および3章で紹介した土地被覆分類手法を用いて日本周域における土地被覆分類図を作成した。分類項目には、水面 (Water)、都市 (Urban)、水田 (Paddy)、畑 (Cropland)、草原 (Grassland)、落葉樹 (Deciduous Forest)、常緑樹 (Evergreen Forest)、裸地 (Bare Land) の8つの土地被覆クラスを設定した。分類時

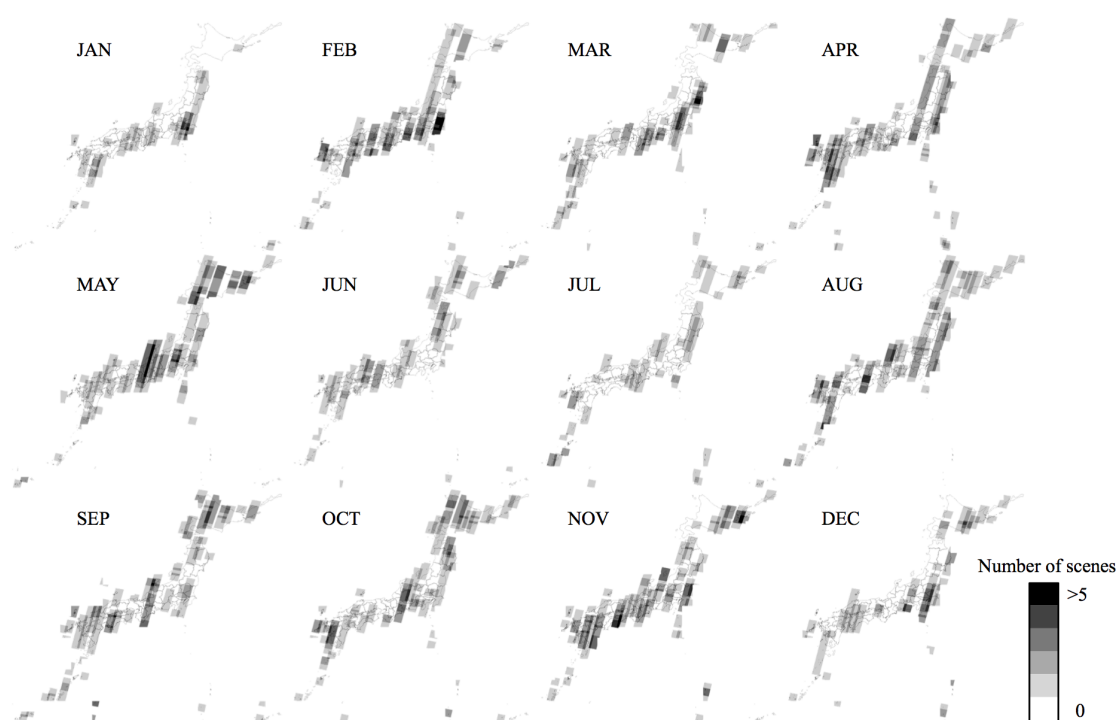


Fig. 6.2: 月毎の日本周域における AVNIR-2 画像の取得数

に使用する教師データには、2,505 点の土地被覆検証点データを基に AVNIR-2 データセットからサンプリングしたラベル付き観測値データを利用した。ここで用いた検証点データは JAXA EORC およびリモートセンシング技術センター (RESTEC) によって作成されたデータで、衛星画像や国土地理院の情報を用いて周囲が同一被覆であった地点を収集したものである [25]。Fig. 6.3 は検証点データの空間分布を図示したものである。各検証点上を観測した 4 バンドの地表面反射率データとその観測日を統合した 6 次元データを 1 つのラベル付き観測値データとし、計 28,252 点のラベル付き観測値データをサンプリングした。Table 6.1 はクラス毎の検証点データ数および得られたラベル付き観測値データ数と、ラベル付き観測値データのバンド毎の地表面反射率の平均・標準偏差の値を示したものである。

土地被覆分類図作成処理では、まず AVNIR-2 データセットの各シーンに対して教師データを用いた単時期解析を行い、クラス事後確率画像を作成した。この際、地表面反射率が全てのバンドで 0.2 以上であった観測値は雲・雪氷であるとし、値無しとした。その後、日本周域を 1 度 × 1 度のタイルに分割し、隣接画素間隔が AVNIR-2 データの地上分解能で

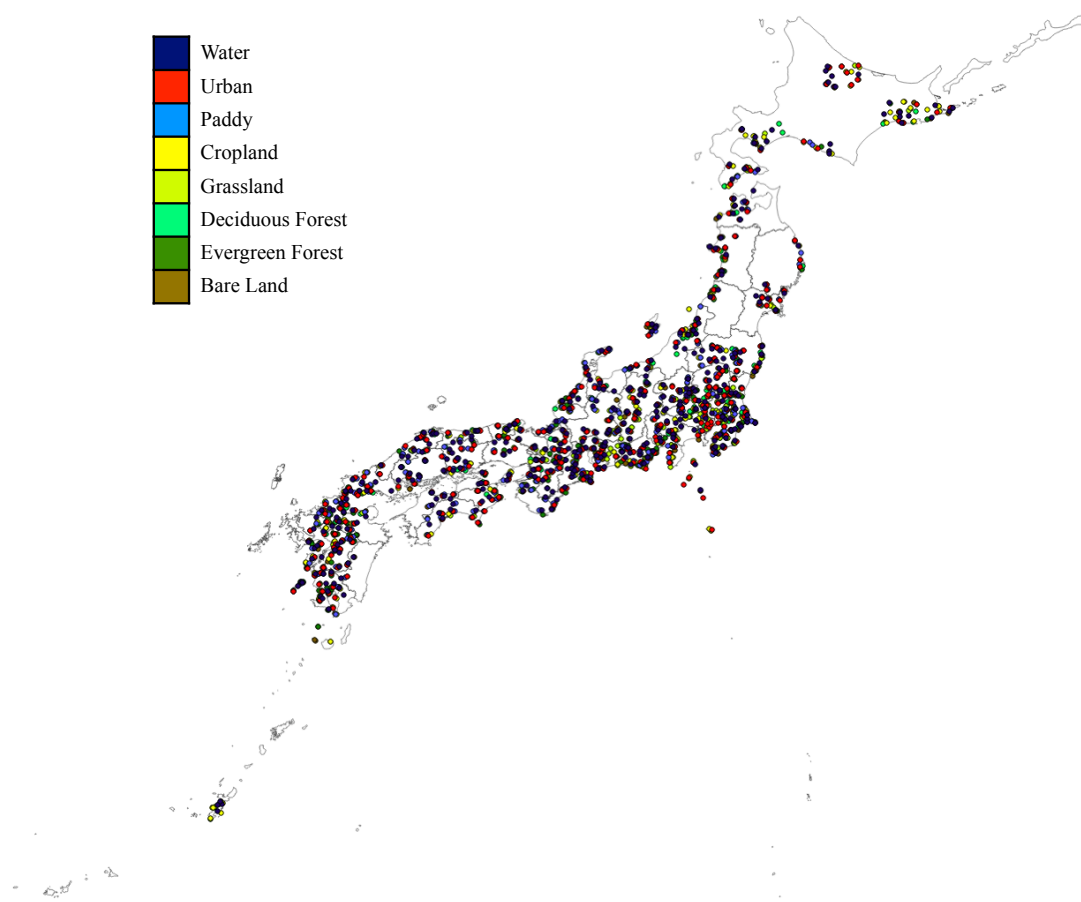


Fig. 6.3: 土地被覆検証点データの空間分布

Table 6.1: 土地被覆検証点データと観測値サンプルの要約

		Water	Urban	Paddy	Cropland	Grassland	Deciduous Forest	Evergreen Forest	Bare Land	Total
Num. of validation points		394	367	385	306	163	164	282	444	2,505
Num. of samples		4,273	4,066	4,439	3,462	1,901	1,758	3,350	5,003	28,252
Mean	Band 1	0.052	0.095	0.065	0.066	0.068	0.052	0.044	0.109	
	Band 2	0.059	0.109	0.089	0.091	0.097	0.066	0.055	0.160	
	Band 3	0.047	0.112	0.092	0.095	0.100	0.060	0.045	0.189	
	Band 4	0.057	0.150	0.225	0.257	0.304	0.261	0.251	0.250	
Standard Deviation	Band 1	0.026	0.029	0.036	0.034	0.058	0.027	0.021	0.044	
	Band 2	0.030	0.033	0.039	0.038	0.062	0.029	0.022	0.063	
	Band 3	0.028	0.036	0.045	0.044	0.071	0.032	0.021	0.082	
	Band 4	0.050	0.044	0.099	0.093	0.114	0.105	0.086	0.092	

Table 6.2: 土地被覆分類の検証実験における実行環境

OS	CentOS 6.5
CPU	Intel Xeon CPU E5-2650 × 2 (2.0 GHz 16 cores)
RAM	64GB
GPU	Nvidia Tesla K20 (compute capability 3.5, 2496 CUDA cores, and 5GB VRAM)

ある 10 m に近い値となるように各タイルを 12,000×12,000 画素に分割した格子点を設定し、各格子点上における多時期のクラス事後確率を総合して最終的な分類結果を得た。

提案手法はクラス事後確率算出時に観測値と全教師データの組み合わせに対してカーネル関数を求め、その総和を取る必要があることから計算コストが大きく、高速な計算環境を必要とした。そのため本研究では、GPU による並列処理を行う GPGPU のフレームワークである NVIDIA CUDA[54] を利用した計算環境を構築して処理を行った。Table 6.2 に示す環境で実行した結果、総計算時間は約 94 時間となった。同様の計算環境にて GPGPU による高速化を行わずに実行した場合の総計算時間は約 45,587 時間と推定され、GPGPU を用いることで約 485 倍の高速化を行うことができた。

Fig. 6.4 は提案手法である KPC-PROD で作成した土地被覆分類図である。なお、KPC-SUM による土地被覆分類図もほぼ同等の土地被覆分類図となったため、省略した。今回は前述した 8 つの土地被覆クラスを設定したが、実際には湿地帯、山岳地帯における疎林・低木林や岩盤帯、サンゴ礁というように、設定したクラスには分類が難しい土地被覆があった。これらの土地被覆は、今回の結果では全く別のクラスに誤分類されていた。その他の目立つ誤分類には、主に海上や積雪地域で雲・雪氷が除外しきれていなかった箇所が都市に誤分類されていたもの、高層建築物による陰影が水域に誤分類されていたものがあった。

Fig. 6.6 につくば市近郊における 3 種類の土地被覆分類図の比較を示す。Fig. 6.6(a) は同域における AVNIR-2 データの True Color 合成画像（RGB にそれぞれバンド 3, バンド 2, バンド 1 の地表面反射率を割り当てた画像）の一例である。Fig. 6.6(b) は提案手法である KPC-PROD による土地被覆分類図であり、比較のために Fig. 6.6(c) に国土地理院による数値地図 5000（土地利用）[55]、及び Fig. 6.6(d) に JAXA EORC による高解像度

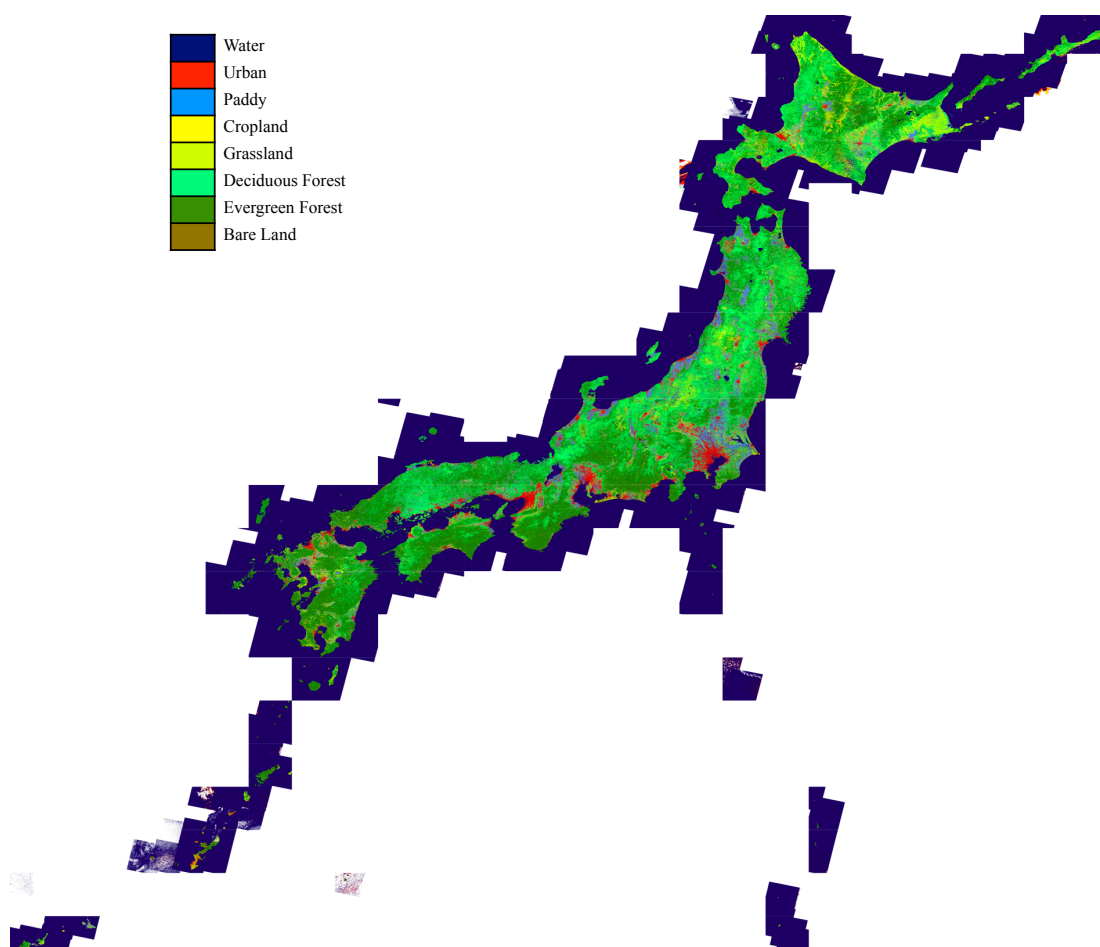


Fig. 6.4: KPC-PROD による日本周域土地被覆分類図

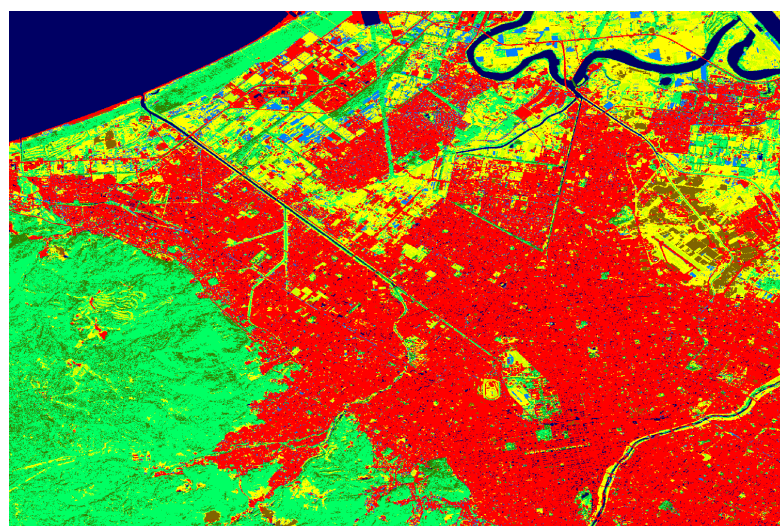


Fig. 6.5: KPC-PROD による日本周域土地被覆分類図における札幌市周辺の拡大図

土地利用土地被覆分類図バージョン 13.02[39] をそれぞれ示す。また、Fig. 6.6(e) に Fig. 6.6 の土地被覆分類図の凡例を示す。ここで、Fig. 6.6(b) および Fig. 6.6(c) は地上分解能 10 m で作成されているのに対し、Fig. 6.6(d) は地上分解能 30 m で作成されている点異なる。提案手法である KPC-PROD による土地被覆分類が、従来手法を利用して作成された JAXA EORC による高解像度土地利用土地被覆分類図に比べてより高精度であることが定性的に確認できる。

6.4 精度評価を行う手法および精度評価指標

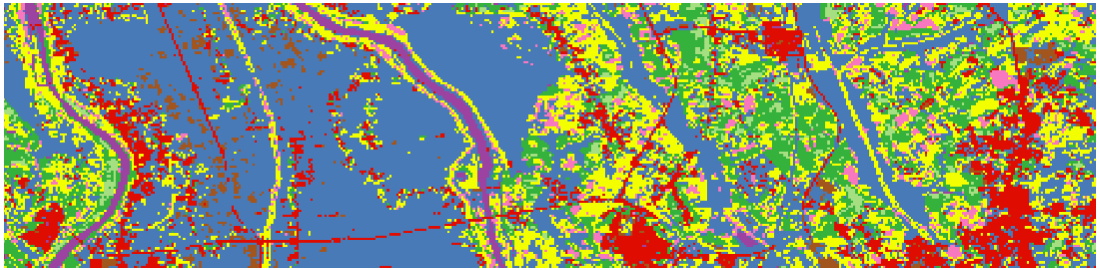
提案手法および作成した土地被覆分類図の精度評価のため、日本広域に分布した検証点を用いた精度検証結果を紹介する。

精度検証を行う手法は次の 5 手法とした。まず多時期解析におけるスコアの算出方法を変えた 3 種類の提案手法で、それぞれ総乗スコアを用いた手法 (KPC-PROD)、総和スコアを用いた手法 (KPC-SUM)、多数決を用いた手法 (KPC-FREQ) を評価した。加えて比較のために、最尤分類法 (MLC) [6] および 1 対 1 のソフトマージンサポートベクターマシン (SVM) を組み合わせた多クラス分類器 (SVMs) [56] の 2 種類の従来手法による多数決方式での多時期解析結果の評価も行った。これらの 2 手法では AVNIR-2 の 4 バンド地表面反射率を入力データとして用いた。SVM のカーネル関数には Kavzoglu らの解析 [57] で最も高精度であったガウシアンカーネル（もしくは放射基底関数カーネル, RBF kernel) を用い、パラメータの値はグリッドサーチにより決定して、スラック変数によるペナルティの重み $C = 1000$ 、ガウシアンカーネルの幅 $\gamma = 0.1$ とした。SVM は提案手法と同じくカーネルに基づくアルゴリズムであるため分類精度も同等になると予想されるが、分類結果に信頼度を含めることができないため多時期解析は多数決方式のみとなってしまうこと、また、基本的には 2 クラス分類器であるため多クラス分類を行うためにはクラス間の組み合わせ数分の SVM を学習して分類を行う必要があり、結果的に学習・分類の計算コストが非常に大きくなってしまうという問題点がある。

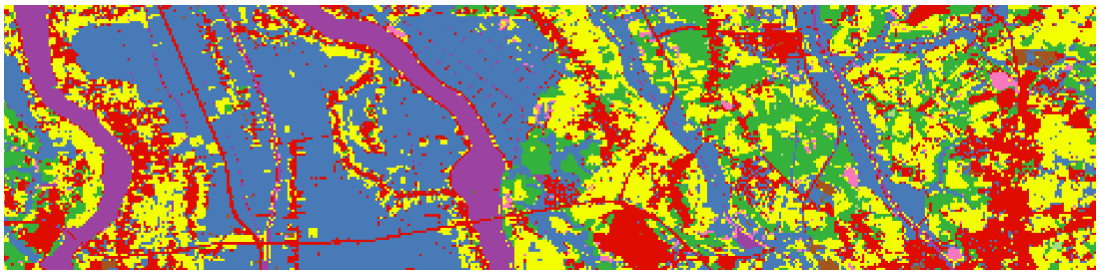
また、精度の定量的評価を行うために次の 4 種類の指標を算出した。総合的な精度指標として、クラスに依らず全てのテスト数に対する正答数の割合を表す OA や、結果の一致



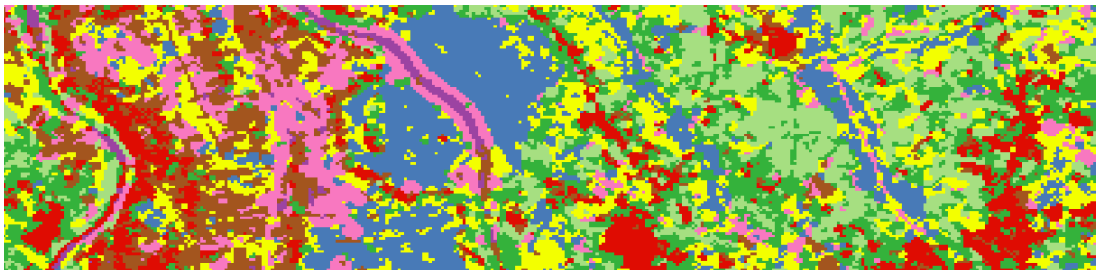
(a) 2009 年 9 月 6 日に観測された AVNIR-2 データの True Color 合成画像 (RGB はそれぞれバンド 3, バンド 2, バンド 1 の地表面反射率) の一例



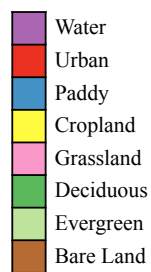
(b) KPC-PROD による土地被覆分類図



(c) 国土地理院による土地利用図 (首都圏 2005 年) [55]



(d) JAXA EORC による土地被覆分類図 (バージョン 13.02) [39]



(e) 凡例

Fig. 6.6: つくば市近郊における土地被覆分類図の比較

率を、偶然による一致率を勘案せず、0 を不一致、1 を一致で表した指標である Kappa 係数を算出した。またクラス別の精度指標として、クラス毎の全テスト数に対する正答数の割合を表す PA や、クラス毎の分類結果数に対する正答数を表す UA を求めた。PA と UA の間にはトレードオフ、すなわち PA は高いが UA が低いケース（そのクラスを過剰に分類していて、誤分類が多い状態）や UA は高いが PA が低いケース（そのクラスを過小に分類していて、出した結果は正しいが未検出が多い状態）があり、クラス別の精度を評価する際には PA と UA の両方がバランス良く高くなっていなければならない事に注意する必要がある。

6.5 日本広域に分布した検証点を用いた精度検証結果および考察

日本広域にて様々な時期に得られた衛星データが混在するデータセットを用いた際の土地被覆分類の精度を評価するため、6.3 節で示した日本中に分布する 2,505 点の検証点データをテストデータとして利用した精度検証結果を示す。それぞれの検証点を用いたテストでは、検証点を基にサンプリングされた多時期の観測値データを分類し、その結果のクラスと検証点のクラスを比較して正答できていたかを判断した。分類には教師データも必要とするが、教師データとテストデータを同じものにしてしまうのは精度検証として不適切であるため、検証点データを教師データとテストデータに分割しなくてはならないが、より正確な精度評価を行うためには教師データもテストデータも多くとる必要がある。そのため、ここでは交差検定 [58] を用いて精度検証を行った。この交差検定では、検証点データを 10 分割し、そのうちの 1 つをテストデータ、残りを教師データとした精度検証を 10 通り全ての組み合わせで行い、その結果を統合することで、検証点データ全体に対して教師データとの重複無く精度検証を行った。

Table 6.3 は比較した 5 手法の 10 分割交差検定による精度の算出結果である。また、Table 6.4 は KPC-PROD によるテスト結果の分類精度行列を示したものである。KPC-PROD の OA は 80.0% と最も高精度であり、次いで KPC-SUM が 79.2% という結果で、提案手法が従来手法と比較して高い精度であった。クラス別の精度を見ると、KPC-PROD や KPC-SUM は MLC や SVMs と比べて PA や UA がどのクラスともバランス良く改善され

Table 6.3: 検証点データを用いた 10 分割交差検定による精度評価結果

		KPC-PROD	KPC-SUM	KPC-FREQ	MLC	SVMs
Producer's Accuracy [%]	Water	97.7	98.0	98.0	97.0	98.0
	Urban	95.6	95.6	94.3	96.5	93.7
	Paddy	82.6	81.3	81.0	28.3	75.3
	Cropland	54.2	49.3	42.2	37.9	44.1
	Grassland	58.3	55.8	52.1	16.6	21.5
	Deciduous Forest	50.6	47.0	51.8	23.2	1.2
	Evergreen Forest	82.3	85.5	82.3	92.6	95.0
	Bare Land	84.2	84.0	80.4	73.0	82.4
User's Accuracy [%]	Water	98.2	97.2	95.8	95.0	98.0
	Urban	85.4	83.4	78.8	54.4	79.6
	Paddy	78.5	78.3	69.5	54.2	55.1
	Cropland	62.2	64.5	60.3	41.7	43.0
	Grassland	55.6	48.4	53.5	58.7	85.4
	Deciduous Forest	48.0	49.0	48.6	28.1	100.0
	Evergreen Forest	80.0	78.2	81.4	60.4	66.5
	Bare Land	94.4	93.3	93.7	90.0	93.1
Overall Accuracy [%]		80.0	79.2	77.1	64.3	72.9
Kappa Coefficient		0.768	0.759	0.735	0.586	0.683

ていた。KPC-PROD と MLC の差異に着目すると、水面、水田、畑、落葉樹、裸地の 5 クラスで精度が高く、都市、草原、針葉樹の 3 クラスについても PA・UA の一方が高くもう一方が低くなっており同等の精度であった。また KPC-PROD と SVMs の差異に注目すると、PA・UA ともに向上していたのは都市、水田、畑、裸地の 4 クラスであったが、草原と落葉樹の 2 クラスは PA がそれぞれ 21.5% から 58.3%, 1.2% から 50.6% と大きく向上しており、バランス良く改善されていた。

KPC-PROD のテスト結果では、畑-草原-落葉樹の間での誤分類が多く、その結果 PA が 50% 台に留まっていた。どれも植生被覆でフェノロジーに大きな変化があるという共通点があるため、観測値分布が大きく重なっている事が推察される。Table 6.1 のクラス毎のサンプルの記述統計量からも、畑、草原、落葉樹の平均値が近く、分布が重なっていることがわかる。畑はこの中でも季節変化に特徴があると思われるが良く識別できていなかった。その原因は次の 2 つが考えられる。1 つめは、提案手法では観測値分布の地域性を考慮していないことである。観測値分布には北方と南方では季節方向の分布が大きく異なるといった地域性が大きく含まれていると予想できるが、提案手法では日本全国の土地被覆を一律に扱って観測値分布を推定していることが、クラス間で観測値分布が重なってしま

Table 6.4: 検証点データを用いた 10 分割交差検定による精度評価における KPC-PROD の分類精度行列

	Reference								Total
	Water	Urban	Paddy	Cropland	Grassland	Deciduous Forest	Evergreen Forest	Bare Land	
Results									
Water	385	3	0	2	0	1	0	1	392
Urban	3	351	7	5	0	2	0	43	411
Paddy	2	7	318	55	9	1	0	13	405
Cropland	2	2	39	166	28	20	1	9	267
Grassland	1	0	12	45	95	13	2	3	171
Deciduous Forest	1	0	3	14	24	83	47	1	173
Evergreen Forest	0	0	1	7	6	44	232	0	290
Bare Land	0	4	5	12	1	0	0	374	396
Total	394	367	385	306	163	164	282	444	2,505

う原因の 1 つになっていると考察する。そして 2 つめには、提案手法が特定の地点における観測値の時系列変化を考慮していないことが挙げられる。畑の観測値の季節変化には人為的变化が含まれるため、特定の地点における観測値の時系列変化を調べると、ある時期に急激な変化が起こるといった特徴が見られることが予想される。そのような時系列変化を調べるためには高頻度かつ定期的に観測値が得られていることが必要と考えられるが、高分解能衛星データではそのようなデータを得ることは困難である。畑-草原-落葉樹の間の高精度な識別方法については今後、検討の余地がある。また、畑や草原は水田との間でも誤分類されていることが多かった。水田は灌水している時期だと高い精度で識別できるため、この種の誤分類は灌水時期の観測データが欠測していたことによる影響が大きい。また、落葉樹は常緑樹との間での誤分類が多かった。落葉樹は冬季に葉を落とすため冬季の観測データがあれば容易に識別可能であると思われるが、誤分類していた原因は、冬季の観測データが欠損していたことや、冬季の観測データはあるが落葉樹の林床に常緑の笹などが生い茂っていて観測データの違いに現れなかったことが考えられる。

次に、同一地点で利用可能な多時期データ数が及ぼす分類精度への影響を示す。Fig. 6.7 の折れ線グラフは、先ほどと同じ交差検定結果について KPC-PROD の分類精度を入力に使用した多時期データの数毎に示したものであり、棒グラフはテスト事例数を同様に多時期データ数毎に示したものである。テスト事例数が極端に少なく信頼出来る精度が算出できていない箇所もあるが、概ね、多時期データ数が増えるにつれて精度も向上している傾向が見られる。テスト事例数が比較的多い多時期データ数 10 前後の精度は 80%程度と

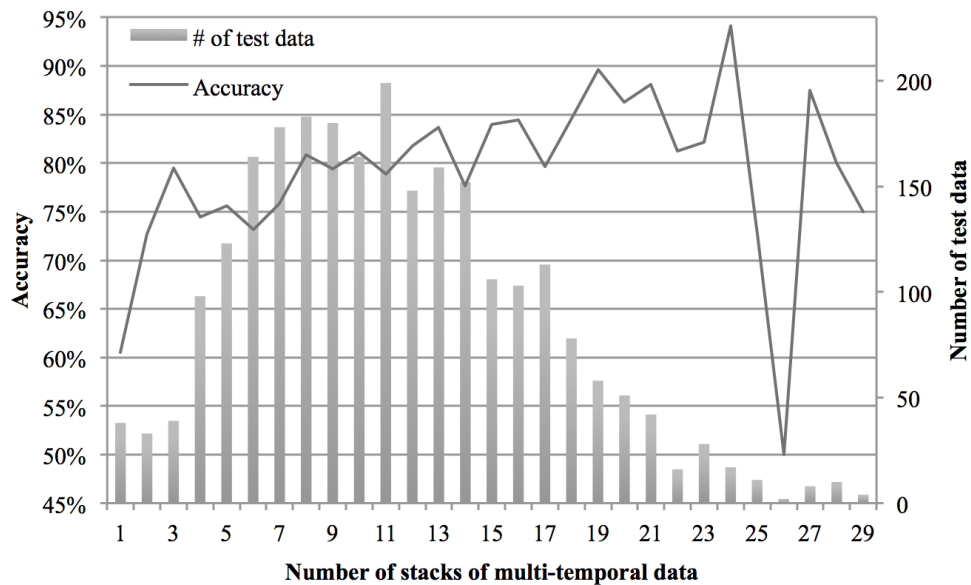


Fig. 6.7: 多時期データ数の増加に対する KPC-PROD の分類精度変化

なっており、Table 6.3 で示した全体精度に近い値となっている。それに対して多時期データ数 20 前後の精度は 85% から 90% となっている。このことから、より多くの多時期データ数を用いることで全体精度を更に高められると期待できる。一般的に、長期間のデータセットを利用することで多時期データ数を増やすことができるが、3.6 節で述べたように本手法では同一地点における多時期データ中での土地被覆変化は考えないという仮定のもとで分類を行っているため、これを踏まえた検討が必要である。他の方法には、他のセンサにより観測されたデータを併用した分類を行うことが考えられる。これは、センサ毎に単時期毎のクラス事後確率を求め、それらを統合したスコアを算出することにより可能であると考えられる。

最後に、教師データ数の増減による分類精度の差異を示す。Fig. 6.8 は 10 分割交差検定において、それぞれの組合せで教師とする検証点データを 100 点ずつ無作為抽出して増加させていった場合の OA の変動を示したものである。KPC-PROD と KPC-SUM の 2 手法は教師とする検証点データが増えていくにつれ精度が大きく下がることなく徐々に向上しており、MLC や SVMs に比べて高い安定性を示した。MLC や SVMs はモデルの推定過程に大域的最適化を含むため、新しく追加した教師データの傾向の影響を受けて分類境界が変化しやすいため、精度が大きく上下したと考えられる。それに対してカーネル密度

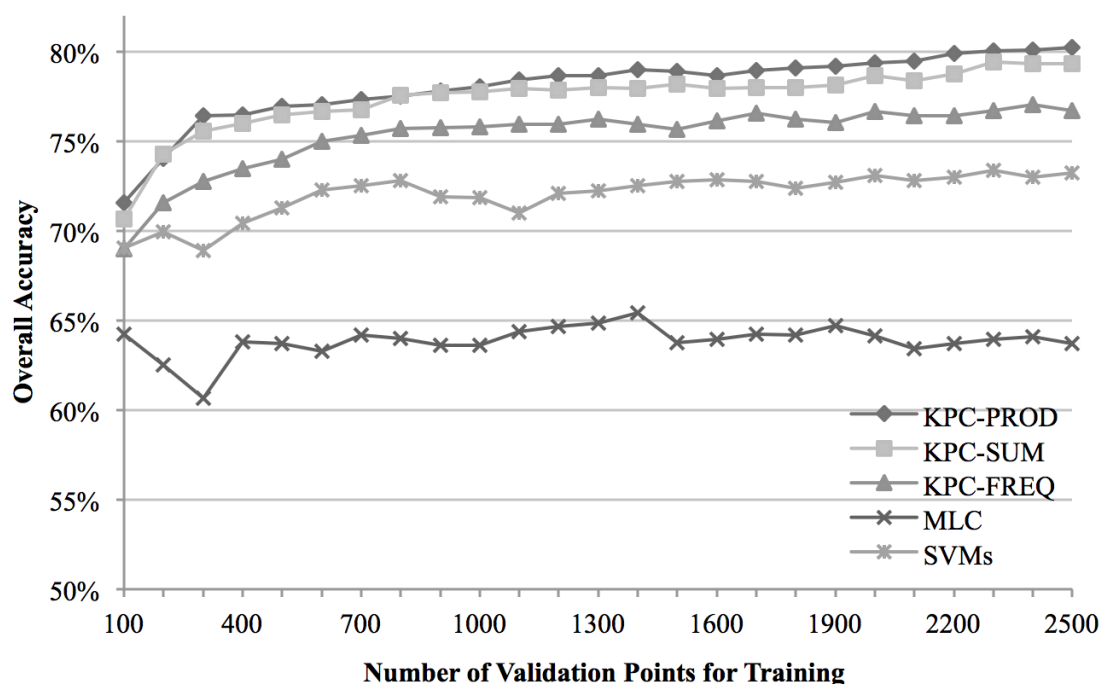


Fig. 6.8: 教師データ数の増加に対する各手法の分類精度変化

推定では、個々の教師データが推定後の密度分布に与える影響が局所的で、互いに影響しあうことなく多量の教師データを活かした推定が可能であり、この特徴が分類結果の安定性に寄与していると考えられる。

6.6 本章のまとめ

本章では、3章で提案した土地被覆推論手法の検証実験結果を示すことを目的とし、提案手法を用いて5年間分のAVNIR-2データセットから日本周域の10mメッシュ高精度土地被覆分類図を作成、その分類精度を算出し、既存手法による分類精度との比較を行った。

検証実験結果では、OAではKPC-PRODとKPC-SUMの精度に大きな差がなかったこと、MLCおよびSVMsを上回っていたことが確認された。KPC-PRODのOAは80.0%であり、同じ評価基準ではないため検証の余地は残るが、岩男ら[41]やRanら[42]による他の広域土地被覆分類図の精度評価結果ではOAが良くても60%弱であり、それよりも高精度であった。しかし、Thomlinsonら[40]が実用精度基準としたOAで最低85%かつPAで全クラス最低70%には届いておらず、依然として精度向上への課題が残ると考えられる。

しかし、検証では同一地点で利用可能な多時期データ数が増すことである程度の精度向上が望めることが確認でき、他センサの観測データを併用した分類による高精度化が期待できる。また、教師データ数の増減に対する分類精度の評価も行い、その結果、KPC-PRODとKPC-SUMの2手法はMLCやSVMsに比べて、少ない教師データでも高い精度の結果が得られたこと、教師データを増やしても精度が大きく下がることなく安定的に向上していく傾向が確認された。そのため、更に教師データを増加させることでもある程度の精度向上が期待できると考えられる。

第7章 多時期データと土地被覆解析に基づく 変化抽出の自動実行

7.1 実験の概要

本章では、4章で提案した解析ワークフローの自動構築および自動実行の枠組みの検証実験結果を示す。提案した枠組みをもとに自動判読システムを構築し、それを用いて変化抽出を行う実行例の結果を示す。システムへの入力は目的の変化抽出結果を表す解析クエリのみとし、解析クエリ入力から解析ワークフローが構築されて目的の変化抽出結果が得られるまでの流れを示す。

実験では、システムで利用するデータやプロセスは全て RDF データによって記述された意味情報を付加し、それを機械が読み取って解析ワークフローが構築できることを示す。本実験で用いる解析処理はある特定の解析を行うこと、あるいはある特定のセンサが観測したデータのみを入力として受け付けるのではなく、汎用的な解析を行えること、同じ種類のセンサであればどのセンサのデータであっても入力データとして利用できることを意図して記述されており、このような解析処理を用いた自動判読が行うことで記述されたデータやプロセスの共有性や再利用性を持つことを示す。また、システムで利用するデータやプロセスは全てデータストア及びナレッジストアに個々のモジュールとして置いた上で自動判読を行い、システムの持つモジュール性や分散管理性を示す。

7.2 データストアおよびナレッジストアの構成

本実験で利用するデータストア及びナレッジストアは、1台のマシン上にそれぞれ別のディレクトリとして設置した。ただし、システム内でデータストアやナレッジストアにアクセスするには URI を用いた参照 (e.g. http://ホスト名/data_store/AV20RI_GSI10/)

ALAV2A013712900-00RIRFU.RCRefG.GEO.rdf) を利用し、1 台のマシン上にデータストア及びナレッジストアを設置していても複数台のマシン上に設置しているものと同等のアクセス手段を取った。

Fig. 7.1 に本実験のために用いたデータストアの構成を表すディレクトリ／ファイルツリーを示す。このデータストアには 2006 年から 2011 年の 5 年間に日本域を観測した 1876 シーンの AVNIR-2 データ（放射輝度）、6.3 節で作成した KPC-PROD による日本全国の土地被覆分類図を 1 度毎に区切ったタイル画像、AVNIR-2 データの土地被覆ラベル付きサンプルデータ、土地被覆検証点データを置いた。4.4 節で説明したように、個々のデータにはそのデータのメタデータを記述した RDF データをセットで置いた。例として、Fig. 7.2 に AVNIR-2 データのメタデータを記述した RDF グラフを示す。このメタデータは 4.3.6 項で示したデータオントロジーの定義における data 概念が保持している属性に沿って記述されたものである。図中の楕円ノードはクラス概念のインスタンス、長方形ノードはリテラル、有向エッジは関係概念のインスタンスをそれぞれ表す。“rdf:”や“rso:”のような接頭辞は URI の一部を省略したもので、URI の先にそのインスタンスに関する詳細な記述があるという外部参照を表す。また、ランダムな 8 字の英数字で表されたノードは外部参照を持たない無名ノードである。Fig. 7.2 の RDF グラフには、具体的には次のような事柄が記述されている。

- AVNIR-2 (http://www.hozo.jp/owl/AVNIR_2) が観測センサである放射輝度 (rso: radiance) を内容として格納する。
- 観測日時は 2006 年 4 月 28 日 1 時 50 分である。
- ファイル名が ALAV2A013712900-00RIRFU.RcRefG.GEO, ファイル形式が hdr のラスタデータ (rso:raster_data) であり、画素値は実数値 (rso:real_number), 画像のバウンディングボックスは WKT で記述された長方形 (rso:rectangle) である。

```

1 data_store
2 |— AV2ORI_GSI10
3 |   |— ALAV2A013712900-00RIRFU.RCRefG.GEO      % AVNIR-2 データ本体
4 |   |— ALAV2A013712900-00RIRFU.RCRefG.GEO.hdr  % AVNIR-2 データヘッダ
5 |   |— ALAV2A013712900-00RIRFU.RCRefG.GEO.rdf  % AVNIR-2 データのメタデータ
6 |   |— ALAV2A013712900-00RIRFU.RCRefG.GEO.txt  % AVNIR-2 データの要約
7 |   |— ALAV2A015022730-00RIRFU.RCRefG.GEO
8 |   |— ALAV2A015022730-00RIRFU.RCRefG.GEO.hdr
9 |   |— ALAV2A015022730-00RIRFU.RCRefG.GEO.rdf
10 |  |— ALAV2A015022730-00RIRFU.RCRefG.GEO.txt
11 |  .
12 |  .
13 |  .
14 |  |— ALAV2A278172690-00RIRFU.RCRefG.GEO
15 |  |— ALAV2A278172690-00RIRFU.RCRefG.GEO.hdr
16 |  |— ALAV2A278172690-00RIRFU.RCRefG.GEO.rdf
17 |  |— ALAV2A278172690-00RIRFU.RCRefG.GEO.txt
18 |— JAPAN_LULC_KPCPROD
19 |   |— N24E122_class.tif      % 日本全国土地被覆分類図
20 |   |— N24E122_class.tif.rdf  % 日本全国土地被覆分類図のメタデータ
21 |   |— N24E123_class.tif
22 |   |— N24E123_class.tif.rdf
23 |   .
24 |   .
25 |   .
26 |   |— N45E148_class.tif
27 |   |— N45E148_class.tif.rdf
28 |— av2_samples.rdf      % AVNIR-2データのラベル付きサンプル(1)
29 |— av2_samples_extend.rdf % AVNIR-2データのラベル付きサンプル(2)
30 |— validation_points.rdf % 土地被覆検証点データ
31
32 2 directories, 7757 files

```

Fig. 7.1: データストア内のディレクトリ／ファイルツリー

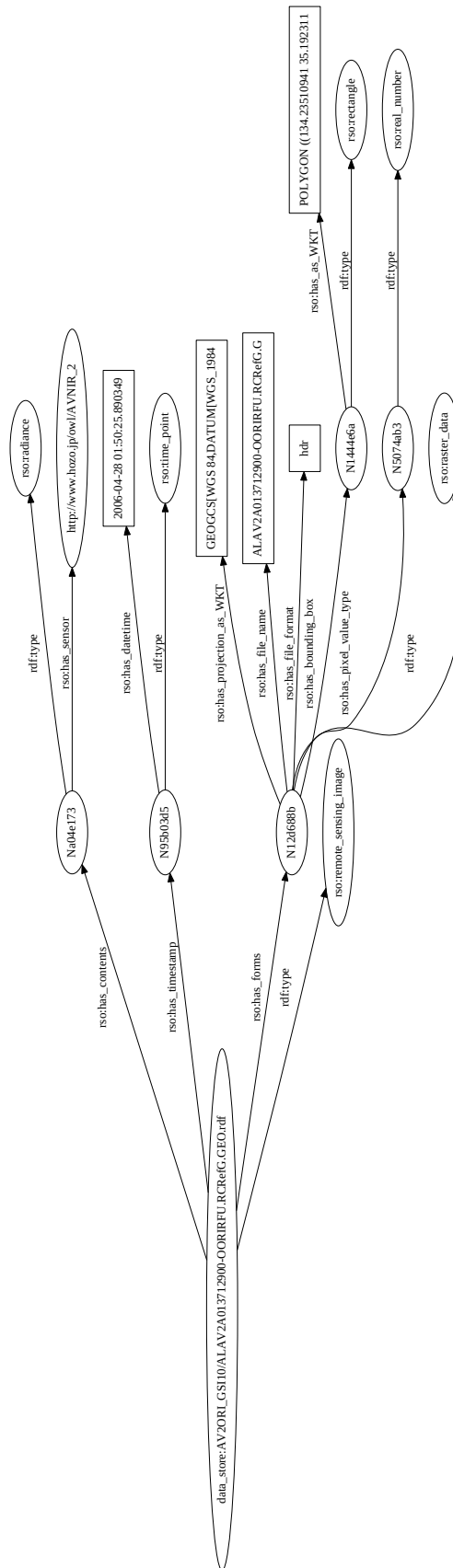


Fig. 7.2: AVNIR-2 データのメタデータを記述した RDF グラフ

また、Fig. 7.3 に本実験のために用いたナレッジストアの構成を表すディレクトリ／ファイルツリーを示す。ナレッジストアには 6SV を用いた大気補正プロセス、画素値の差分に基づく変化検出プロセス、変化前後の土地被覆分類図の変化を検出する土地被覆変化検出、3 種類の手法（KPC, MLC, SVMs）による土地被覆分類プロセス、KPC-PROD による多時期土地被覆プロセス、NDVI 算出プロセス、道路検出プロセス、斜面傾斜算出プロセスを置いた。こちらもデータストアと同様に、個々のプロセスにはそのプロセスのメタデータを記述した RDF データをセットで置いた。例として、Fig. 7.4 に土地被覆変化検出プロセスのメタデータを記述した RDF グラフを、Fig. 7.5 に土地被覆分類プロセスのメタデータを記述した RDF グラフをそれぞれ示す。このメタデータは 4.3.7 項で示したプロセスオントロジーの定義における process 概念が保持している属性に沿って記述されたものである。

Fig. 7.4 に示した土地被覆変化検出プロセスのメタデータには、次のような事柄が記述されている。

- 変化検出（`rso:change_detection`）を行うプロセスである。
- プロセスモジュールのファイル名は `lulc_change` である。
- 出力データは地物（`rso:geographical_feature`）のサブクラスのインスタンス（`rso:subinstance_of`）から地物（`rso:geographical_feature`）のサブクラスのインスタンス（`rso:subinstance_of`）への変化（`rso:chnage_of_existence`）の発生（`rso:occurrence`）を記述した変化マップ（`rso:change_map`）であり、形式はラスターデータ（`rso:raster_data`）である。
- 入力データは 2 つあり、どちらも土地被覆分類図（`rso:LULC_map`）で形式はラスターデータ（`rso:raster_data`）である。片方は出力データの変化前の地物（`rso:geographical_feature`）と同一（`rso:sameAs`）であり、もう片方は出力データの変化後の地物（`rso:geographical_feature`）と同一（`rso:sameAs`）である。
- 2 つの入力データの画像のバウンディングボックスは、出力データの画像のバウンディングボックスを表す長方形（`rso:rectangle`）と同一（`rso:sameAs`）である。

- 2つの入力データの作成日時には前後関係 (rso:before_after) がある.

また Fig. 7.5 に示した土地被覆分類プロセスのメタデータには, 次のような事柄が記述されている.

- 単時期土地被覆分類 (rso:single_scene_LULC_classification) を行うプロセスである.
- プロセスモジュールのファイル名は lulc_kpc である.
- 出力データは地物 (rso:geographical_feature) のサブクラスのインスタンス (rso:subinstance_of) の土地被覆分類図 (rso:LULC_map) であり, 形式はラスターデータ (rso:raster_data) である.
- 入力データは2つあり, 1つは地表面反射率 (rso:reflectance) を記述したリモートセンシング画像 (rso:remote_sensing_image) であり, もう1つはサンプルデータ (rso:sample_data) の集合 (rso:bag) で, リモートセンシング画像 (rso:remote_sensing_image) と同一 (rso:sameAs) の観測センサで得られた地表面反射率 (rso:reflectance) と, 出力データの地物 (rso:geographical_feature) と同一のインスタンス (rso:sameAs) とを対応付けたものである.
- 入力となるリモートセンシング画像 (rso:remote_sensing_image) の形式はラスターデータ (rso:raster_data) である.
- 入力となるサンプルデータ (rso:sample_data) の形式は点データ (rso:point_data) である.

```

1 knowledge_store
2 |-- 6sv_atmoscorrect      % 大気補正プロセス
3 |-- 6sv_atmoscorrect.rdf
4 |-- imgdiff              % 差分による変化検出プロセス
5 |-- imgdiff.rdf
6 |-- lulc_change          % 土地被覆変化検出プロセス
7 |-- lulc_change.rdf
8 |-- lulc_kpc             % KPCによる土地被覆分類プロセス
9 |-- lulc_kpc.rdf
10 |-- lulc_mlc            % MLCによる土地被覆分類プロセス
11 |-- lulc_mlc.rdf
12 |-- lulc_svm            % SVMによる土地被覆分類プロセス
13 |-- lulc_svm.rdf
14 |-- mt_lulc_kpcprod     % KPC-PRODによる多時期土地被覆分類プロセス
15 |-- mt_lulc_kpcprod.rdf
16 |-- ndvi_calc          % NDVI算出プロセス
17 |-- ndvi_calc.rdf
18 |-- road_detect        % 道路検出プロセス
19 |-- road_detect.rdf
20 |-- slope_angle        % 斜面傾斜算出プロセス
21 |-- slope_angle.rdf
22
23 0 directories, 20 files

```

Fig. 7.3: ナレッジストア内のディレクトリ／ファイルツリー

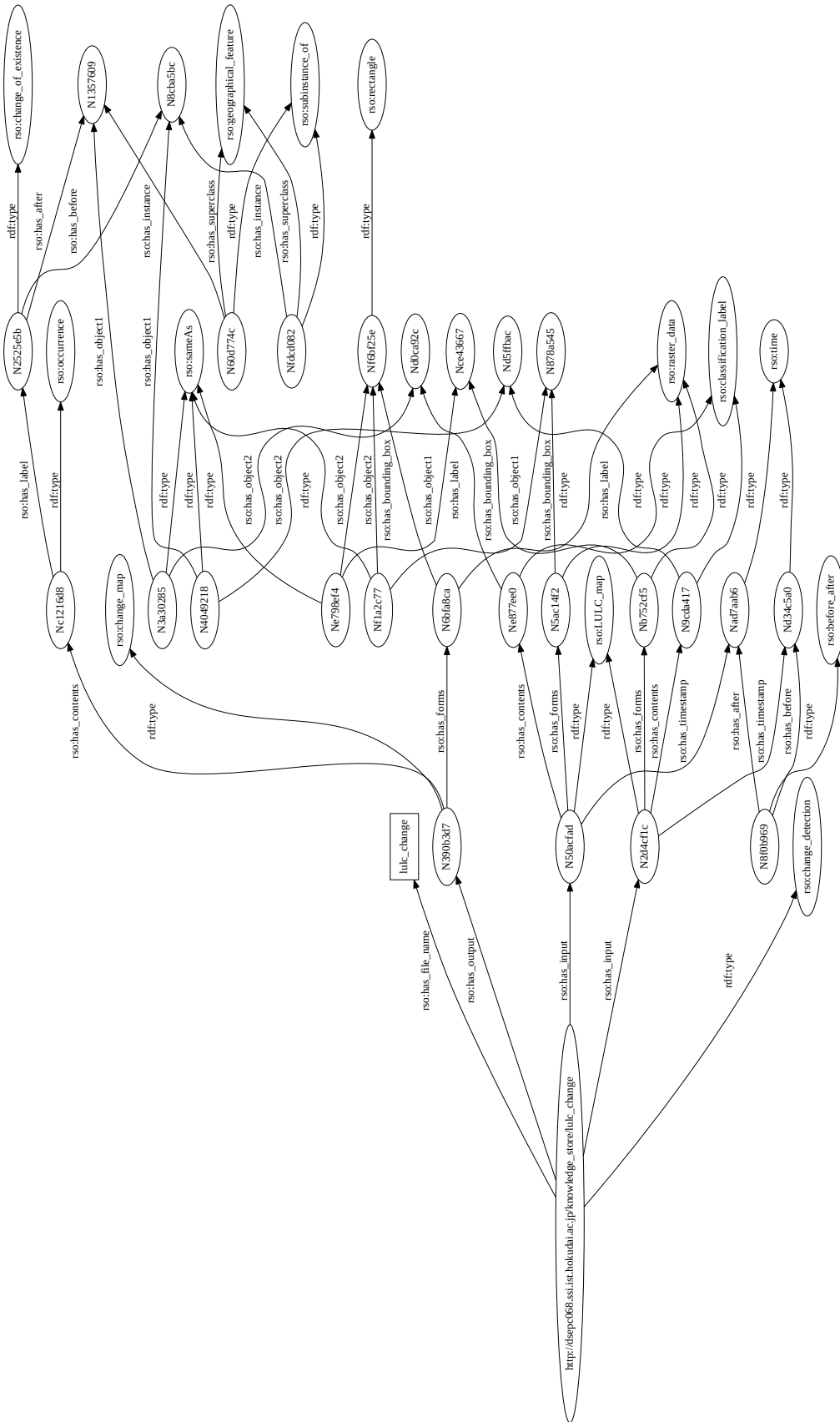
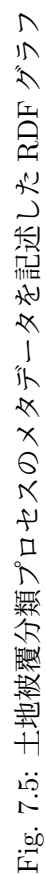


Fig. 7.4: 土地被覆変化検出プロセスのメタデータの記述したRDF グラフ



7.3 森林から裸地への変化検出の実行

森林から裸地への土地被覆変化の検出結果を表す解析クエリをシステムに入力し、自動判読の途中経過ならびに出力結果を示す。システムに解析対象の情報や解析場所、時間等の必要な情報を入力すると、それらの情報を元にシステム内部で利用される解析クエリが作成される。Fig. 7.6 はシステムへ入力した森林から裸地への土地被覆変化の検出結果を表す解析クエリの RDF グラフである。このクエリには主に次のような事柄が記述されている。

- 出力として得たいデータが森林 (`rso:forest`) から裸地 (`rso:granular_layer`) への変化検出結果 (`rso:change_map`) である。
- 解析エリアの範囲が WKT で指定した長方形 (`rso:rectangle`) である。
- 変化前の土地被覆の作成時間が 1 年 1 月 1 日～9999 年 12 月 31 日の期間内 (`rso:during`) であり、同じく変化後の土地被覆の作成時間が 2009 年 7 月 21 日～8 月 21 日の期間内 (`rso:during`) である。

Fig. 7.7 に、この入力クエリを用いて得られた解析ワークフローの概略図を示す。入力クエリから解析ワークフローが構築されるまでの行程を以下に示す。

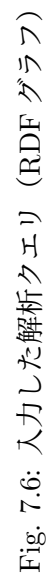




Fig. 7.7: 構築された解析ワークフロー (RDF グラフ)

1. 利用者から受け取った入力クエリを用いてデータストアを探索し、森林から裸地への土地被覆変化検出結果がデータストアに存在しない事を確認する。
2. 入力クエリを用いてナレッジストアを探索し、森林から裸地への土地被覆変化検出結果を出力可能なプロセスである `lulc_change` とマッチングする。`lulc_change` は特定の土地被覆変化のみを検出するものではなく、様々な土地被覆間の変化を検出する汎用的なプロセスであり、そのメタデータには地物 (`rso:geographical_feature`) から地物 (`rso:geographical_feature`) への変化検出結果 (`rso:change_map`) を出力するというように曖昧に記述されている。ここではオントロジーに基づく意味推論が行われ、森林 (`rso:forest`) から裸地 (`rso:granular_layer`) への変化検出結果 (`rso:change_map`) とマッチングする。
3. マッチングした `lulc_change` に対して入力データ探索クエリの作成を行い、変化前の森林の土地被覆分類図を探索するクエリ (Fig. 7.8) 及び変化後の土地被覆分類図を探索するクエリ (Fig. 7.9) を得る。

4. 変化前の森林の土地被覆分類図のクエリを用いてデータストアを探索し、日本全国土地被覆分類図である JAPAN_LUL_KPCPROD/N34E131_class.tif とマッチングする。日本全国土地被覆分類図には落葉樹 (`rso:deciduous_forest`) 及び常緑樹 (`rso:evergreen_forest`) の土地被覆クラスが含まれており、オントロジーに基づく意味推論によってこれらと森林の土地被覆分類図がマッチングし、これらを統合した土地被覆分類図が作成されて返される。
5. 変化後の裸地の土地被覆分類図のクエリを用いてデータストアを探索し、目的のデータが存在しない事を確認する。
6. 変化後の裸地の土地被覆分類図のクエリを用いてナレッジストアを探索し、4つの土地被覆分類プロセス (`lulc_kpc`, `lulc_mlc`, `lulc_mlc`, `mt_lulc_kpcprod`) とマッチングする。これらのプロセスも `lulc_change` と同様に特定の土地被覆のみに対するプロセスではなく、汎用的なプロセスとなっており、オントロジーに基づく意味推論によってマッチングする。
7. マッチングした4つのプロセスのうち、まず `lulc_kpc` に対して変化後の裸地の土地被覆分類図のクエリを用いた入力データ探索クエリの作成が行われ、サンプルデータ (`rso:sampled_data`) を探索するクエリ (Fig. 7.10) 及びマルチスペクトルセンサ (`rso:multispectral_sensor`) による地表面反射率 (`rso:reflectance`) のリモートセンシング画像 (`rso:remote_sensing_image`) を探索するクエリ (Fig. 7.11) を得る。
8. サンプルデータを探索するクエリを用いてデータストアを探索し、複数のサンプルデータとマッチングし、これを返す。
9. マルチスペクトルセンサによる地表面反射率のリモートセンシング画像を探索するクエリを用いてデータストアを探索し、目的のデータが存在しないことを確認する。
10. マルチスペクトルセンサによる地表面反射率のリモートセンシング画像を探索するクエリを用いてナレッジストアを探索し、大気補正を行うプロセスである `6sv_atmoscorrect` とマッチングする。

11. 6sv_atmoscorrect に対してマルチスペクトルセンサによる地表面反射率のリモートセンシング画像を探索するクエリを用いて入力データ探索クエリの作成が行われ、マルチスペクトルセンサによる放射輝度 (rso:radiance) のリモートセンシング画像を探索するクエリ (Fig. 7.12) を得る.
12. マルチスペクトルセンサによる放射輝度のリモートセンシング画像を探索するクエリを用いてデータストアを探索し、AVNIR-2 による放射輝度データとマッチングし、これを返す. ここではマルチスペクトルセンサと AVNIR-2 をオントロジーに基づく意味推論によってマッチングする.
13. 6sv_atmoscorrect の入力データが揃い実行可能となったため、その出力データとして AVNIR-2 による地表面反射率のリモートセンシング画像を lulc_kpc での処理に返す.
14. lulc_kpc の入力データが揃い実行可能となったため、その出力データとして変化後の裸地の土地被覆分類図を lulc_change での処理に返す.
15. lulc_kpc に対して行われた 7~14 までの処理が同様に他の 3 つのプロセス (lulc_mlc, lulc_mlc, mt_lulc_kpcprod) に対して行われ、それぞれ出力データを得る.
16. ここで利用者が 4 つのプロセスの出力データの中から 1 つを選択を行い、選択されたデータが lulc_change の入力データとして採用される.
17. lulc_change の入力データが揃い実行可能となったため、その出力データとして森林から裸地への土地被覆変化検出結果を利用者へと返す.

以上の行程により作成された解析ワークフローの RDF グラフは Fig. 7.7 で示したものとなるが、システム内部では Fig. 7.13 で示すように実行順序を表す制御構造が付け加えられた RDF グラフとして保持され、解析ワークフロー構築処理終了後にこの制御構造に従って実際の解析が実行される.

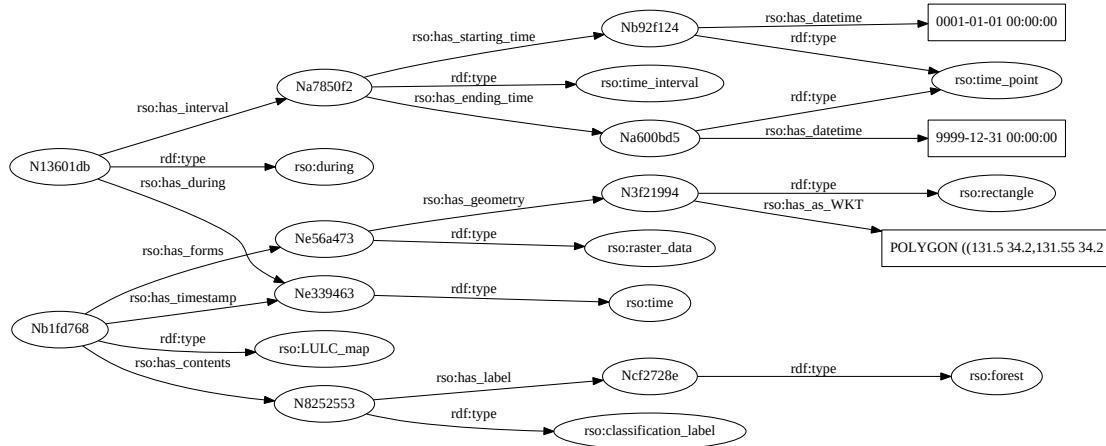


Fig. 7.8: 変化前の森林の土地被覆分類図を探索するクエリ (RDF グラフ)

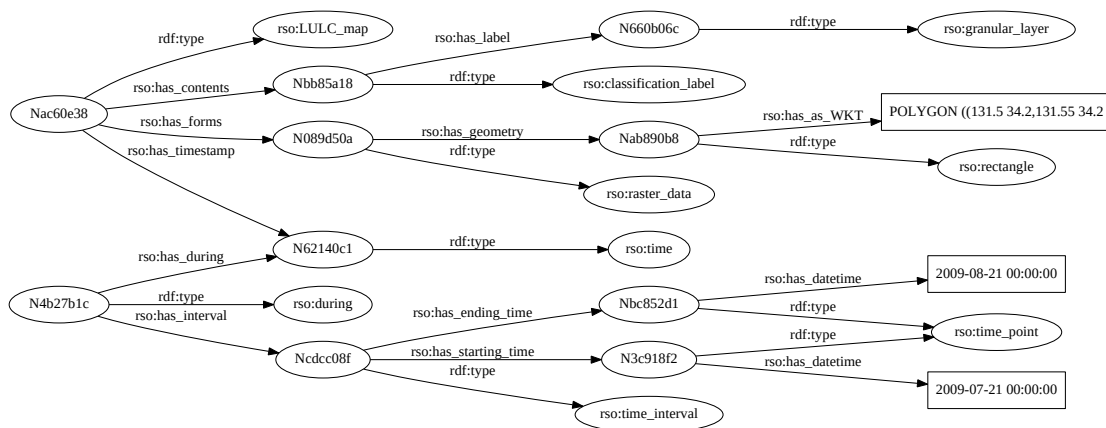


Fig. 7.9: 変化後の裸地の土地被覆分類図を探索するクエリ (RDF グラフ)

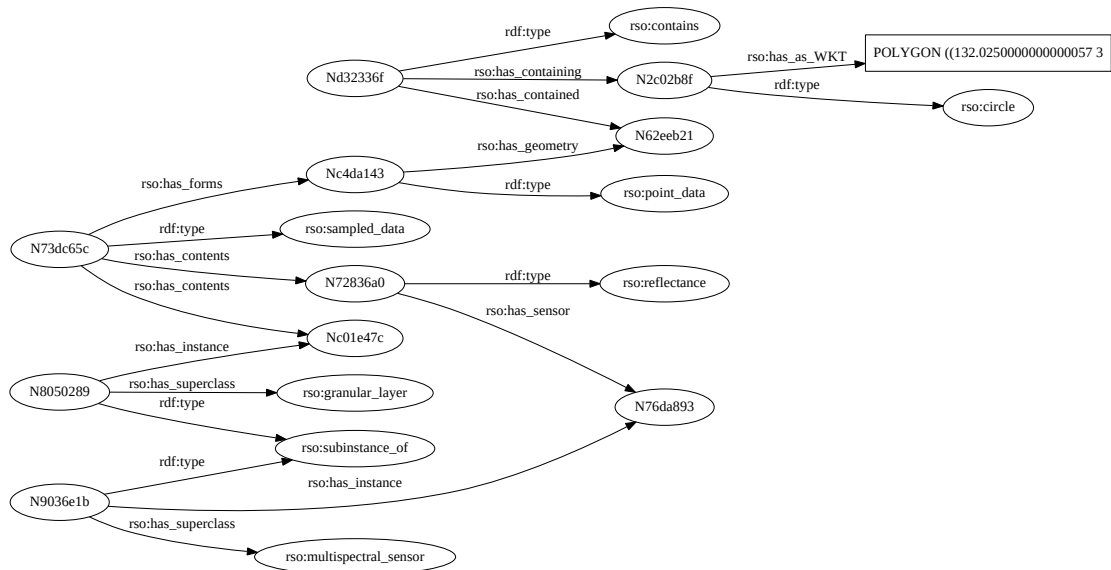


Fig. 7.10: サンプルデータを探索するクエリ (RDF グラフ)

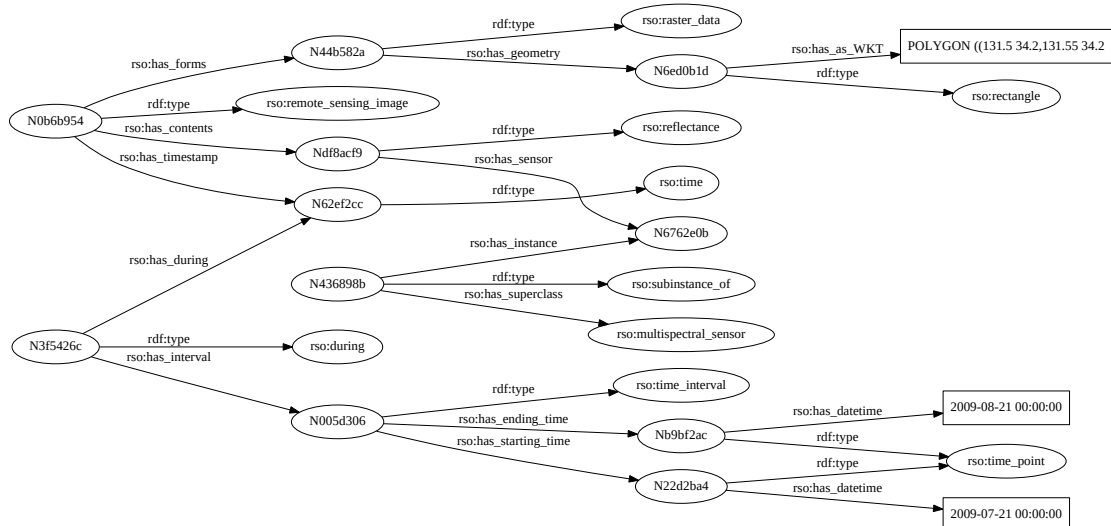


Fig. 7.11: マルチスペクトルセンサによる地表面反射率のリモートセンシング画像を探索するクエリ (RDF グラフ)

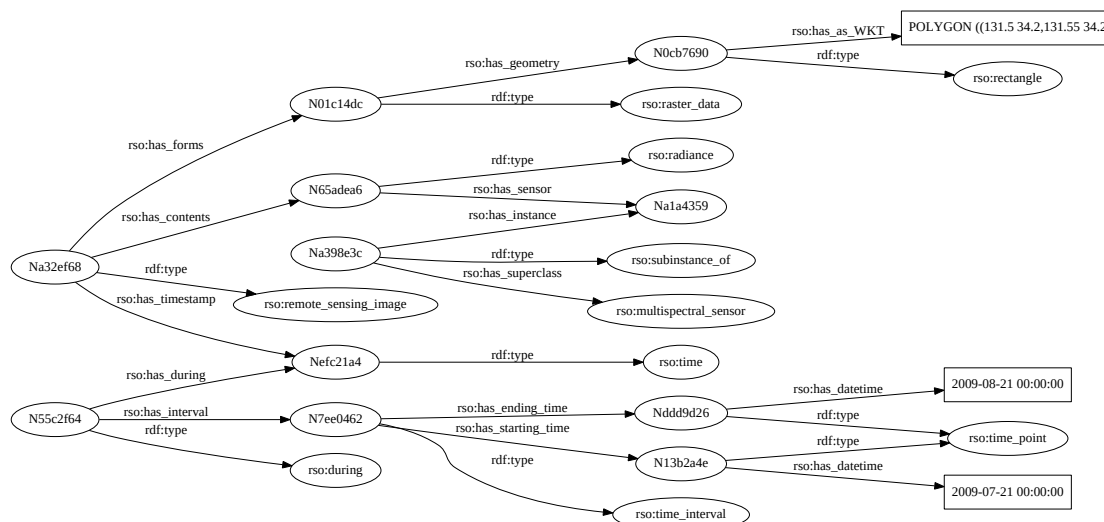


Fig. 7.12: マルチスペクトルセンサによる放射輝度のリモートセンシング画像を探索するクエリ (RDF グラフ)

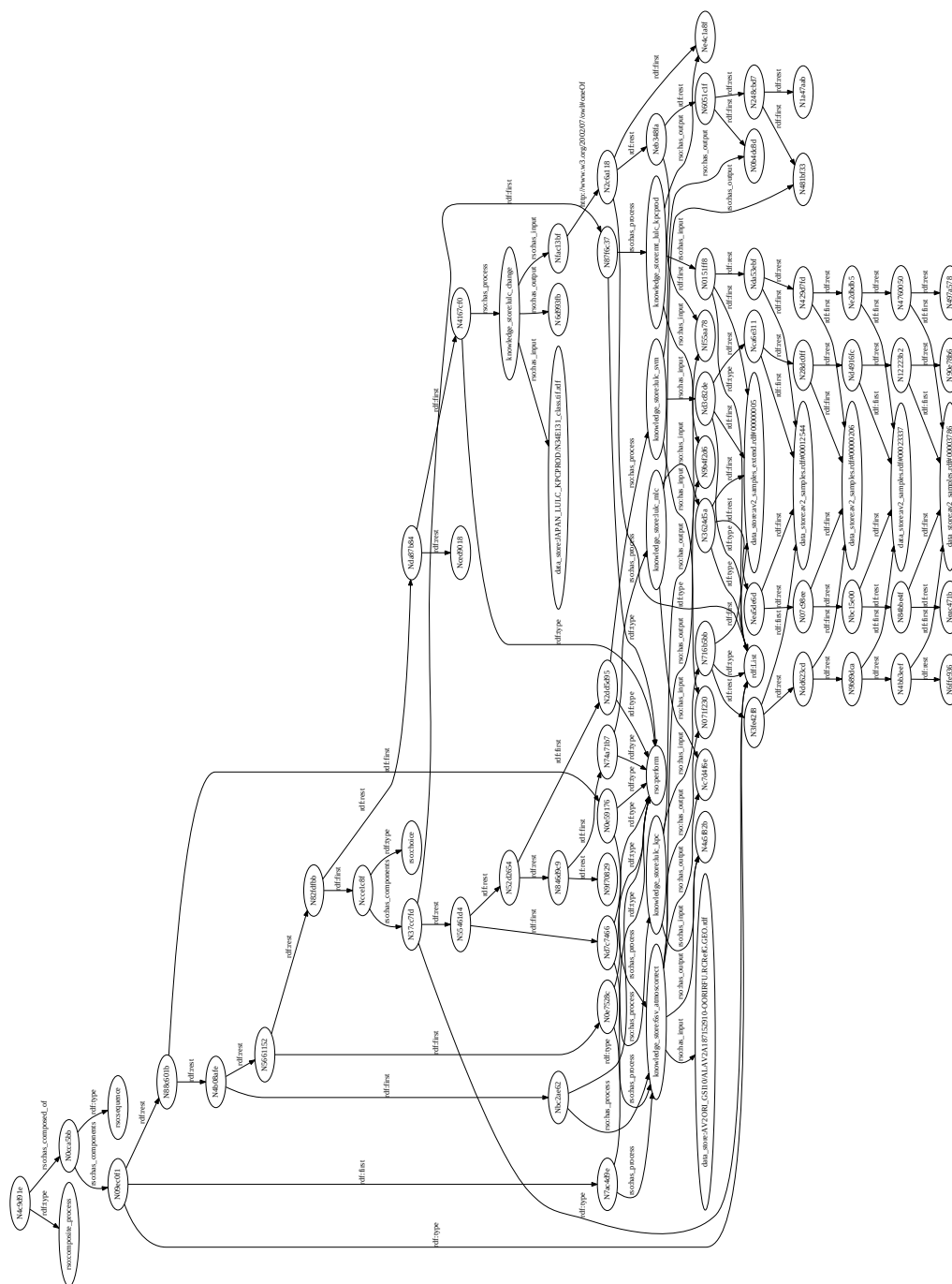


Fig. 7.13: 構築された解析ワークフローとその制御構造 (RDF グラフ)

Fig. 7.14(a) に解析の入力データとなった変化前の日本全国土地被覆分類図の一部を示す。同様に、Fig. 7.14(b) に同じく解析の入力データとなった変化後の AVNIR-2 データの一部を示す。Fig. 7.14(c) は解析実行中に Fig. 7.14(b) から土地被覆分類プロセスである `lulc_kpc` によって生成された裸地の土地被覆分類図である。そして Fig. 7.14(d) は最終的に得られた森林から裸地への土地被覆変化検出結果である。解析クエリとして入力した目的の解析結果が自動で得られていることが確認できた。土地被覆分類プロセスに `lulc_kpc` を選択した場合、Table 6.4 の分類精度行列から計算される裸地の土地被覆図の OA は 96.3%、森林の土地被覆図の OA は 96.1%であるため、両者の差分の結果である森林から裸地への変化検出結果の OA は同時確率値である 92.5%と推定される。

ここで示した実行例の途中では、次のような場面でオントロジーを用いた意味推論によってマッチングが行われた。

- 森林から裸地への変化検出結果というクエリを用いて、地物から地物への変化検出結果を出力するプロセスとマッチングしたとき。
- 森林の土地被覆分類図というクエリを用いて、落葉樹と常緑樹の土地被覆クラスが含まれる土地被覆分類図とマッチングしたとき。
- 裸地の土地被覆分類図というクエリを用いて、地物の土地被覆分類をおこなうプロセスとマッチングしたとき。
- マルチスペクトルセンサによる放射輝度のリモートセンシング画像を探索するクエリを用いて、AVNIR-2 による放射輝度のリモートセンシング画像とマッチングしたとき。

ここでは、定義上の包含関係を用いることで記述されたデータやプロセスを探索できること、またその定義上の包含関係はオントロジーの参照によって整合性が保証されていることが確認できる。このような特徴から、記述されたデータやプロセスの共有性や再利用性の実現されていることがわかる。

実験で用いたプログラムのうち、解析ワークフロー構築のプログラムは Python にて RDF ライブラリである `RDFLib`[59] を利用して記述し、ナレッジストアに格納した個々のプロセスモジュールは C++ にて記述した。また本実験は Table 7.1 に示す環境で実行した。

Table 7.1: 自動判読の検証実験における実行環境

OS	OS X v10.8
CPU	Intel Core i7-3615QM (2.30 GHz 4 cores)
RAM	16GB
GPU	NVIDIA GeForce GT 650M (compute capability 3.0, 384 CUDA cores, and 1GB VRAM)

クエリの入力から解析結果が得られるまでの処理時間は約 510 秒であり、そのうちデータストアおよびナレッジストアに格納されているデータやプロセスの読み込みに約 144 秒、解析ワークフローの構築に約 355 秒、解析ワークフローの実行に約 10 秒を要した。データやプロセスの読み込み、及び解析ワークフローの構築の処理時間が長くなっている主な原因は Python の繰り返し処理の遅さが考えられ、実用段階ではより高速な実装方法を検討する必要がある。

7.4 本章のまとめ

本章では、4 章で提案した解析ワークフローの自動構築および自動実行の枠組みの検証実験として、提案した枠組みをもとに構築した自動判読システムによる変化抽出の実行例を示した。データストアおよびナレッジストアにあらかじめ複数のデータやプロセスを格納しておき、それらを組み合わせることで解析クエリで指定した「森林から裸地への変化検出」という判読結果を得るための解析ワークフローを構築し、推定精度 92.5% で目的の変化検出結果が得られることを示した。解析ワークフロー構築の途中では、「森林から裸地への変化検出結果」というクエリに対して「地物から地物への変化検出結果」と曖昧な表現で記述されたプロセスをオントロジーに基づく意味推論によってマッチングできたことや、「マルチスペクトルセンサによる放射輝度のリモートセンシング画像」という曖昧な表現で記述されたクエリに対して「AVNIR-2 による放射輝度のリモートセンシング画像」をオントロジーに基づく意味推論によってマッチングできたことを示し、様々な問題に対して汎用的に適用可能で再利用性の高いプロセスを導入可能であることを示した。

本実験では導入はしなかったが 2 章にて提案した地物・地物変化推論を行う解析処理も

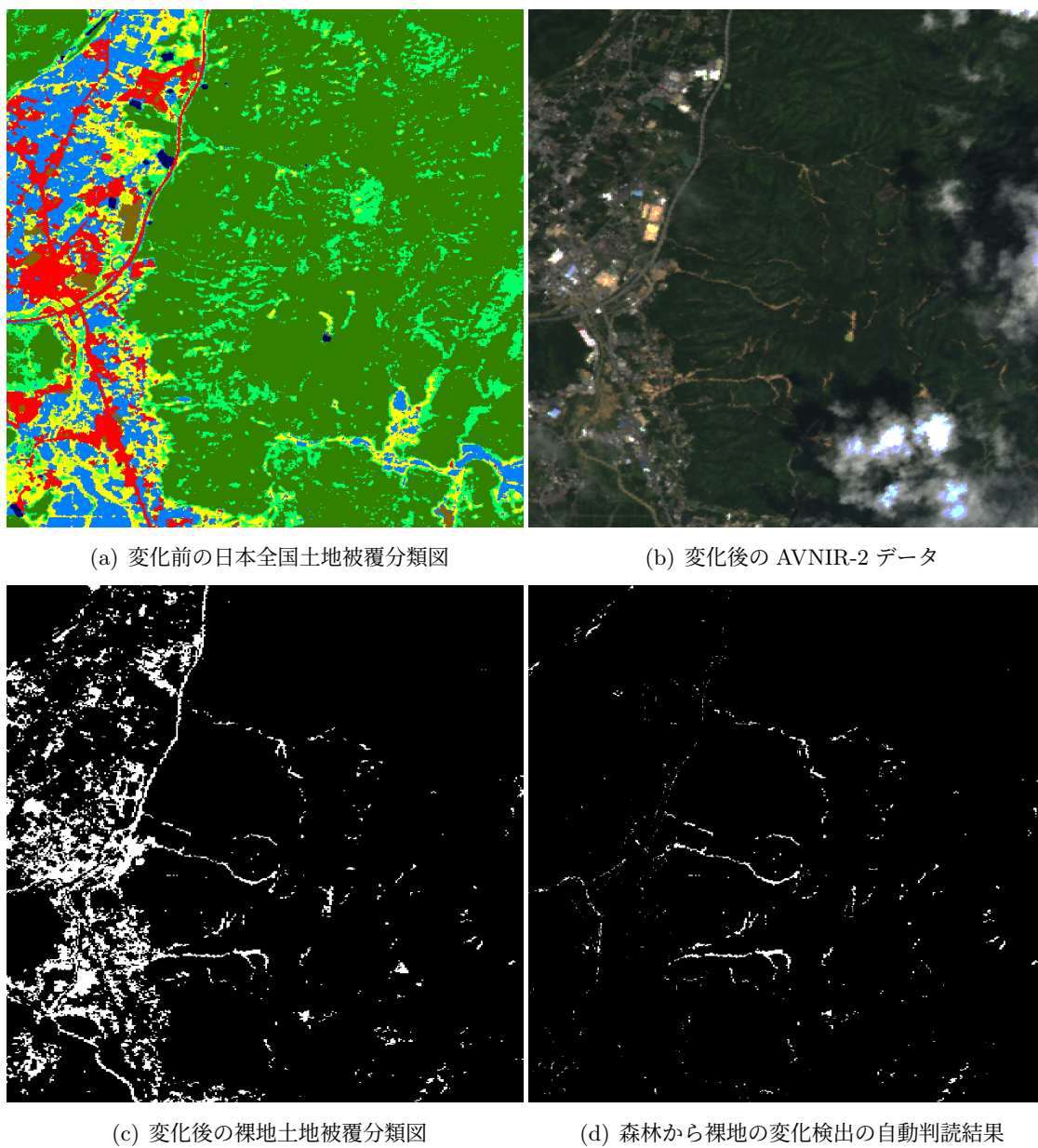


Fig. 7.14: 自動判読の実行結果

提案した自動判読の枠組みで利用可能であり，2 章および 3 章の提案内容を組み合わせた自動判読システムも構築可能である．

第8章 結論

8.1 本論文の成果

光学衛星データ解析において、データ選択から高次解析までを高精度に自動化するための自動判読システムの枠組みの提案を行った。本論文における成果を下記にまとめる。

1. 情報の不確実性を考慮しつつ複数のプリミティブな情報からより複雑な情報を推論するため、ベイズ確率による情報表現及びベイジアンネットワークによる確率的推論を採用した枠組みを提案した。またそれを応用することで衛星データの災害状況把握解析における専門家の判断処理／作業を軽減するための自動変化検出システムを提案した。提案した変化検出システムを土石流跡の検出と洪水による冠水域の検出という2種類の災害状況把握解析事例に対してシステムに変更を加えずに適用し、システムが自動的に知識モジュールを選択してそれぞれの事例に適した情報抽出および推論を行った過程と得られた検出結果を示し、土地被覆の推論結果を一つに絞る事が困難な場合であっても、推論結果の信頼度を伝播させて次の推論を行うことができるため、ロバストな変化検出が行えることを示した。
2. 大量の教師データを活かした土地被覆推論手法として、単時期の土地被覆推論手法である KPC、およびそれを多時期土地被覆推論手法に拡張して高精度化を行った KPC-PROD を提案した。KPC-PROD は分類クラス間の重なりや教師データの密度といった不確実性を定量的に評価可能な事後確率として分類結果を得ることができること、分布に対する先見的な仮定や制約を置くことなく、実データに忠実な分布を推定でき、かつ外れ値には影響されにくい高精度かつロバストな推論ができること、検証点データおよび衛星データセットを用意するだけで自動処理を行うことができるスケーラブルなアルゴリズムであること、低頻度観測かつ欠測が多いため

季節変動を追いつらいデータセットであっても、得られているデータのみから季節性を考慮した土地被覆推論が可能であり、高分解能衛星データを用いた土地被覆分類に適していること、という特徴を持つ。KPC-PROD を用いて日本周域 10m メッシュ 8 クラス土地被覆分類図を作成し、現行版に対して大きく精度向上できたことや、従来手法である MLC の OA:64.3%, SVMs の OA:72.9% に対して KPC-PROD の OA:80.0% と高精度であること、教師データを増やしても精度が大きく下がることなく向上し、更に精度向上が期待できることを確認した。

3. データやプロセスを機械可読に表現・蓄積し、それらを組み合わせた解析ワークフローを自動構築・実行するための枠組みを提案した。RDF とオントロジーを意味情報記述基盤とする枠組みの提案や、衛星データ自動判読に用いるための 7 種類のオントロジーの構築、およびデータ・プロセスを用いて解析ワークフローを自動構築・実行する処理を提案した。個々のデータや補正・解析プロセスを RDF 及び独自構築したオントロジーに基づき蓄積することで、データやプロセスの表現・蓄積を、機械可読性、共有性、再利用性、モジュール性、分散管理性、柔軟性という性質を持たせて実現することができた。
4. 提案した枠組みを用いてクエリ駆動型の自動判読システムを構築し、例として森林から裸地への変化検出結果を求める解析クエリを入力した。その結果、自動でデータやプロセスを組み合わせた解析が実行され、92.5% の推定精度で目的の解析結果が得られたことを確認した。

謝辞

本研究を進めるにあたり、多くの方々に御世話になりました。ここに深く感謝の意を表します。

システムセンシング情報学講座の田殿武雄客員教授、ならびにシステム環境情報学研究室の小野里雅彦教授には修士課程進学時から現在にいたるまでの5年間、終始、研究指導だけに留まらず様々な面からサポートしていただきました。この博士論文があるのも、また今日の自分があるのも、お二方の厳しくも温かいご指導のおかげだと思います。

システムセンシング情報学講座の堀雅裕客員准教授および塩見慶客員准教授、ならびにシステム環境情報学研究室の田中文基准教授にはゼミやミーティング等の場で積極的なご指導をいただきました。研究内容に対しての指摘は的確なものばかりで、議論を通じて博士論文をより洗練されたものにすることができました。

システム制御情報学研究室の金子俊一教授には本論文の審査にあたり、様々なご助言をいただきました。これまで行ってきた研究内容をまとめ、発表するにあたって、先生のご助言は非常に意義深いものでした。

JAXA EORC の皆さま、特に ALOS グループの皆さまにはインターンシップ期間中やリサーチ・アシスタントとして在籍していた際に大変お世話になりました。暖かく受け入れていただくことができ、また業務で忙しい中でも研究に対するご助言をいただけたこともあり、心から感謝しております。

JAXA および RESTEC からは実験で使用した AVNIR-2 データや検証点データを提供していただきました。

研究活動費においては、北海道大学大学院情報科学研究科グローバル COE「知の創出を支える次世代 IT 基盤拠点」ならびに日本学術振興会 平成25年度科学研究費補助金(特別研究員奨励費)からのご支援を頂戴しました。

システム環境情報学研究室の同期および先輩方や後輩達，そしてシステムセンシング情報学講座の本田謙一氏（現・国際航業株式会社）には普段の研究室生活の中で，またときには研究の議論を通じて大変お世話になりました．皆さんがいたおかげでこの5年間で非常に充実した，楽しいものにすることができたと思います．

最後になりましたが，ここまで支えてくれた両親および祖母に深く感謝の意を表します．

参考文献

- [1] 滝口 太. 衛星の防災・災害マネジメントへの貢献. **日本リモートセンシング学会誌**, 31(2):223–226, 2011.
- [2] 宇宙航空研究開発機構. 東日本大震災対応報告書 ～地球観測衛星及び通信衛星による対応の記録～. Technical report, 2011. URL <http://www.sapc.jaxa.jp/antidisaster/20110311report.html>. Accessed: 2013-12-12.
- [3] 大木 永光 and 三原 莊一郎. 衛星地球観測ハードウェア-技術動向. **日本リモートセンシング学会誌**, 31(2):249–251, 2011.
- [4] Yanhua Zhuang, Xingjian Liu, Thuminh Nguyen, Qingqing He, and Song Hong. Global remote sensing research trends during 1991-2010: a bibliometric analysis. *Scientometrics*, 96(1):203–219, November 2012.
- [5] Jixian Zhang. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, March 2010.
- [6] 日本リモートセンシング研究会. **改訂版 図解リモートセンシング**. 日本測量協会, 東京, 2004.
- [7] 日本リモートセンシング学会. **基礎からわかるリモートセンシング**. 理工図書, 2011.
- [8] JAXA/EORC. ALOS ユーザハンドブック, 2007. URL <http://www.eorc.jaxa.jp/ALOS/doc/jhandbk.htm>. Accessed: 2014-01-22.
- [9] JAXA. 陸域観測技術衛星「だいち」(ALOS) プロジェクトについて. URL http://www.jaxa.jp/press/2012/01/20120111_sac_daichi.pdf. Accessed: 2014-01-22.

- [10] 人工知能学会. **デジタル人工知能学事典**. 共立出版, 2008.
- [11] Takashi Matsuyama. Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3):305–316, May 1987.
- [12] Daniel Crevier and Richard Lepage. Knowledge-Based Image Understanding Systems: A Survey. *Computer Vision and Image Understanding*, 67(2):161–185, 1997.
- [13] Riichiro Mizoguchi. Tutorial on ontological engineering. *New Generation Computing*, 21(4):363–364, 2003.
- [14] M. Datcu, K. Seidel, S. D’Elia, and P.G. Marchetti. Knowledge-driven information mining in remote-sensing image archives. *ESA Bulletin*, 110:26–33, 2002.
- [15] Mihai Datcu, Herbert Daschiel, Andrea Pelizzari, Marco Quartulli, Annalisa Galoppo, Andrea Colapicchioni, Marco Pastori, Klaus Seidel, Pier Giorgio Marchetti, and Sergio D’Elia. Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2936, December 2003.
- [16] Mihai Datcu and Klaus Seidel. Human-centered concepts for exploration and understanding of Earth observation images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):601–609, 2005.
- [17] Surya S. Durbha and Roger L. King. Semantics-enabled framework for knowledge discovery from Earth observation data archives. *IEEE Transactions on Geoscience and Remote Sensing*, 43(11):2563–2572, 2005.
- [18] Surya S. Durbha, Roger L. King, and Nicolas H. Younan. An Information Semantics Approach for Knowledge Management and Interoperability for the Global Earth Observation System of Systems. *IEEE Systems Journal*, 2(3):358–365, September 2008.

-
- [19] S.S. Durbha, R.L. King, V.P. Shah, and N.H. Younan. A framework for semantic reconciliation of disparate earth observation thematic data. *Computers & Geosciences*, 35(4):761–773, April 2009.
- [20] Deshendran Moodley, Andrew Terhorst, Ingo Simonis, Graeme McFerren, and Frans van den Bergh. Using the sensor web to detect and monitor the spread of wild fires. In *Proceedings of the 2nd International Symposium on Geo-information for Disaster Management*, pages 27–30, 2006.
- [21] Deshendran Moodley and Ingo Simonis. A New Architecture for the Sensor Web: The SWAP Framework. *Architecture*, 5:18–38, 2006.
- [22] Deshendran Moodley, Anwar Vahed, Ingo Simonis, Graeme McFerren, and Terence Van Zyl. Enabling a new era of Earth observation research: scientific workflows for the Sensor Web’. *Ecological Circuits*, 1:20–23, 2008.
- [23] Deshendran Moodley. *Ontology driven multi-agent systems: an architecture for sensor web applications*. Phd thesis, University of KwaZulu-Natal, South Africa, 2009.
- [24] Deshendran Moodley and Jules Raymond Tapamo. A semantic infrastructure for a Knowledge Driven Sensor Web. In *Proceedings of the 4th International Workshop on Semantic Sensor Networks 2011*, pages 30–45, Bonn, Germany, 2011.
- [25] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, and 堀 雅裕. 衛星データによる災害状況把握のための知識駆動型自動変化検出システムの提案. **日本リモートセンシング学会誌**, 33(1):13–26, 2013.
- [26] C. M. ビショップ (著), 元田 浩 (訳), 栗田 多喜夫 (訳), 樋口知之 (訳), 松本 裕治 (訳), and 村田 昇 (訳). **パターン認識と機械学習 上 - ベイズ理論による統計的予測**. 丸善出版, 東京, 2007.
- [27] Judea Pearl. Bayesian Networks: A Model of Self-Activated Memory for Evidential

- Reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, number CSD-850021, R-43, pages 329–334. UCLA Computer Science Department, 1985.
- [28] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, Burlington, MA, 1988.
- [29] 溝口 理一郎. **オントロジー工学**. オーム社, 東京, 2005.
- [30] 長井 正彦, 小野 雅史, and 柴崎 亮介. 地球観測データのためのリモートセンシング
オントロジーの構築. **写真測量とリモートセンシング**, 48(1):32–40, 2009.
- [31] 古崎 晃司, 来村 徳信, 佐野 年伸, 本松 慎一郎, 石川 誠一, and 溝口理一郎. オントロ
ジー構築・利用環境「法造」の開発と利用：実規模プラントのオントロジーを例とし
て. **人工知能学会論文誌**, 17:407–419, 2002.
- [32] アリストテレス（著） and 出 隆（訳）. **形而上学 全 2 巻**. 岩波書店, 東京, 1959.
- [33] Zhongli Ding, Yun Peng, and Rong Pan. BayesOWL: Uncertainty modeling in
semantic web ontologies. In *Soft Computing in Ontologies and Semantic Web*,
pages 3–29. Springer, Berlin Heidelberg, 2006. URL <http://www.springerlink.com/index/03667368557J7776.pdf>.
- [34] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, 堀 雅裕, and 塩見 慶. 多時期光学観測データを用いた高精度土地被覆分類手法の開発. **日本リモートセンシング学会誌**, 34(2), 2014.
- [35] M.A. Friedl, D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1-2):287–302, November 2002.
- [36] Mark A. Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. MODIS Collection 5 global land

-
- cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168–182, January 2010.
- [37] E. Bartholomé and A. S. Belward. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26(9):1959–1977, May 2005.
- [38] T.R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang, and J.W. Merchant. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21(6 & 7):1303–1330, 2000.
- [39] JAXA/EORC. 高解像度土地利用土地被覆図ホームページ, . URL http://www.eorc.jaxa.jp/ALOS/lulc/lulc_jindex.htm. Accessed: 2013-12-01.
- [40] John R. Thomlinson, Paul V. Bolstad, and Warren B. Cohen. Coordinating Methodologies for Scaling Landcover Classifications from Site-Specific to Global. *Remote Sensing of Environment*, 70(1):16–28, 1999.
- [41] Koki Iwao, Kenlo Nishida, Tsuguki Kinoshita, and Yoshiki Yamagata. Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters*, 33(23):L23404, December 2006.
- [42] Youhua Ran, Xin Li, and Ling Lu. Evaluation of four remote sensing based land cover products over China. *International Journal of Remote Sensing*, 31(2):391–401, January 2010.
- [43] B.W. Silverman. *Density estimation for statistics and data analysis*. Number 1951. Chapman and Hall, London, 1986.
- [44] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, 1992.

- [45] W3C. RDF Primer, . URL <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>. Accessed: 2013-12-05.
- [46] W3C. RDF Vocabulary Description Language 1.0: RDF Schema, . URL <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>. Accessed: 2013-12-05.
- [47] W3C. OWL 2 Web Ontology Language Document Overview (Second Edition), . URL <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>. Accessed: 2013-12-05.
- [48] OGC. OpenGIS Simple Features Specification For SQL Revision 1.1. URL http://portal.opengeospatial.org/files/?artifact_id=829. Accessed: 2014-01-13.
- [49] W3C. OWL-S: Semantic Markup for Web Services, . URL <http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/>. Accessed: 2013-12-16.
- [50] E.F. Vermote, S.Y. Kotchenova, J.C. Roger, D. Tanre, J.L. Deuze, M. Herman, and J.J. Morcrette. Second Simulation of a Satellite Signal in the Solar Spectrum Vector (6SV) Code. URL <http://6s.ltdri.org/>. Accessed: 2013-12-08.
- [51] Fumiaki Tomita, Yoshiaki Shirai, and Saburo Tsuji. Description of textures by a structural analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):183–191, 1982.
- [52] 山口県庁. 「平成 21 年 7 月 21 日豪雨山地災害対策検討委員会」報告書, 2009. URL <http://www.pref.yamaguchi.lg.jp/cms/a17800/sanntisaigai/20100225001.html>. Accessed: 2013-12-15.
- [53] JAXA/EORC. 陸域観測技術衛星「だいち」(ALOS) による『平成 21 年 7 月中国・九州北部豪雨』にともなう緊急観測 (3), . URL http://www.eorc.jaxa.jp/ALOS/img/_up/jdis_opt_090730_gouu.htm. Accessed: 2013-12-15.

-
- [54] NVIDIA Corporation. CUDA Parallel Computing Platform. URL http://www.nvidia.com/object/cuda_home_new.html. Accessed: 2013-12-01.
- [55] 国土地理院. 数值地図 5000 (土地利用) . URL <http://www1.gsi.go.jp/geowww/LandUse/lum-5k.html>. Accessed: 2013-12-16.
- [56] C. Huang, L. S. Davis, and J. R. G. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, January 2002.
- [57] T. Kavzoglu and I. Colkesen. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5):352–359, October 2009.
- [58] Ron Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the 14th international joint conference on Artificial intelligence*, pages 1137–1143, Montreal, Canada, 1995.
- [59] Rdfliib dev group. RDFLib/rdfliib - GitHub. URL <https://github.com/RDFLib/rdfliib>. Accessed: 2014-01-22.

研究業績

- [1] 橋本 秀太郎, 田殿 武雄, 堀 雅裕, 森山 隆, and 小野里雅彦. ピクセルベースの変化検出アプローチの比較及び災害事例への応用に関する一検討. In **日本リモートセンシング学会第 47 回（平成 21 年度秋季）学術講演会**, pages 115–116, 2009.
- [2] 橋本 秀太郎, 小野里 雅彦, 田殿 武雄, 堀 雅裕, and 森山 隆. 画像オブジェクトに基づくコンテキスト認識を用いた変化情報推論. In **日本リモートセンシング学会第 48 回（平成 22 年度春季）学術講演会**, pages 63–64, 2010.
- [3] Shutaro Hashimoto, Masahiko Onosato, Takeo Tadono, Masahiro Hori, and Takashi Moriyama. Case studies of automatic change detection using AVNIR-2 onboard ALOS. In *2010 IEEE International Geoscience and Remote Sensing Symposium*, pages 3644–3647, Honolulu, July 2010. IEEE.
- [4] 橋本 秀太郎, 田殿 武雄, 堀 雅裕, 森山 隆, and 小野里雅彦. オントロジーに基づく知識情報処理によるリモートセンシングデータからの変化判読の自動化. In **日本リモートセンシング学会第 50 回（平成 23 年度春季）学術講演会**, pages 113–114, 2011.
- [5] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, Masahiro Hori, and Takashi Moriyama. An approach for automatic change inference in high resolution satellite images. In *34th International Symposium on Remote Sensing of Environment*, 2011.
- [6] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, Masahiro Hori, and Takashi Moriyama. A framework of ontology-based knowledge information processing for change detection in remote sensing data. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 3927–3930. IEEE, July 2011.

- [7] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, 堀 雅裕, and 森山 隆. 衛星データ判読のためのベイジアンネットワークによる知識表現・学習・推論モデル. In **日本リモートセンシング学会第 51 回 (平成 23 年度秋季) 学術講演会**, pages 127–128, 2011.
- [8] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, Masahiro Hori, and Takashi Moriyama. Probabilistic land cover classification approach toward knowledge-based satellite data interpretations. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 1513–1516. IEEE, July 2012.
- [9] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, and 堀 雅裕. 生成モデル推定に基づく土地被覆の確率的推論手法の検討. In **日本リモートセンシング学会第 53 回 (平成 24 年度秋季) 学術講演会**, pages 77–78, 2012.
- [10] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, and 堀 雅裕. 衛星データによる災害状況把握のための知識駆動型自動変化検出システムの提案. **日本リモートセンシング学会誌**, 33(1):13–26, 2013.
- [11] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, and 堀 雅裕. カーネル法を用いた生成モデル推論による屋島の土地被覆分類. In **第六回四国 GIS シンポジウム**, 2013.
- [12] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, and Masahiro Hori. Land cover classification of entire Japan by kernel-based probabilistic inference using multitemporal AVNIR-2 data. In *International Symposium on Remote Sensing 2013*, 2013.
- [13] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, and Masahiro Hori. Land use and land cover inference in large areas using multi-temporal optical satellite images. In *2013 IEEE International Geoscience and Remote Sensing Symposium*, pages 3333–3336, 2013.
- [14] 橋本 秀太郎, 田殿 武雄, 小野里 雅彦, 堀 雅裕, and 塩見 慶. 多時期光学観測データを用いた高精度土地被覆分類手法の開発. **日本リモートセンシング学会誌**, 34(2), 2014.