Scientific Challenges in Information Retrieval from Earth Observation (EO) Imagery

Nicolas H. Younan
Department Head and James Worth Bagley Chair
Department of Electrical and Computer Engineering
Mississippi State University

Image Information Mining Conference
Bucharest, Romania, 2014
Background

• Earth Observations (EO) data are obtained from a multitude of sources and requires tremendous efforts and coordination among researchers and user groups to come to a shared understanding on a set of concepts involved in a domain.
• The goal is to formulate an end-to-end process that enables the collection and distribution of accurate, reliable EO data, information, products, and services to both suppliers and consumers worldwide.
• Also the knowledge about the background of the originating datasets can be encoded to address the diversity on a global scale.
• One of the critical components in the development of such systems is the ability to obtain seamless access of data across boundaries.
End-to-End Model of the Remote Sensing Information Channel

Sensor System

Scene (reality) Observable Phenomena

Sensor subsystem

Data Package Raw or calibrated

Physical linkages to signatures

Analysis System

Pre-processing

Image processing
Analysis Algorithms

Decision Criteria Image interpretation Classification

Delivery System

Archive & Distribution

Search Distributed archives

Information products of socio/economic, security, and/or scientific merit in the marketplace

Inputs

Outputs

Outcomes

Remote sensing analysis
Outcomes from EO Imagery used by Decision Makers

• Weather forecasts
• Establishment of protocols regulating the emission of greenhouse gases
• Selection of a corridor for a new highway
• Yield of agricultural products

Moving more toward the need for a local decision
Pattern Recognition and Problem Solving in the EO Domain

- EO based solutions require expertise and/or data sources from globally distributed resources.
- These same solutions may require processing of data into information from specific resources.
- How do we bring together geospatially diverse resources to facilitate sharing and knowledge discovery?
- How do we mine data from a variety of sources (sensors, databases, images) into information products and actionable intelligence?
- Constrained – compute horsepower, bandwidth, etc.
NASA’s Earth Science Discipline Focused DAACs

DAAC - Distributed Active Archive Centers

Cumulative archive growth of major EO sensor systems
Use of Pattern Recognition for Data Mining in an EO Context

- **Spreadsheet context**
  - A process of sorting through a large amount of data and picking out relevant information.
  - Data created by user.
  - Data => information => knowledge => actionable intelligence

- **EO context**
  - Data created by sensor.
  - EO sensor Data => models working on data to create information (content) => contextual knowledge about geopolitical and socioeconomic factors => actionable intelligence at the local level (understanding)
EO Contextual Hyper-map for Creating Actionable Intelligence

- Socioeconomic context
- Geopolitical context
- Physical processes
- EO data sets

Geospatial reasoning process

Actionable intelligence at the local level
Challenges - EO Pattern recognition Community

• Reasoning in the geospatial domain is needed for achieving understanding of global processes
• Requires modeling of physical processes
  – Models require assimilation of measurements and content/context of measurement
  – Global modeling is understood to be low spatial resolution
• Context (spatial & temporal context)
  – Google Earth presents global data, but only considers point information in a GIS layered format
  – Lacks connections
• High spatial resolution information is necessary for local decision making with its inherent challenges
Low Spatial Resolution EO Imagery

• EO content in low resolution scales results in parameters (sense of geophysical parameters)
High Spatial Resolution EO Imagery

- EO content in high resolution scales results in *objects* (sense of semantics)
Challenges in EO Imagery

- Multiple dimensional challenges
  - Hyper-spectral
  - Hyper-spatial
  - Hyper-temporal
  - Hyper-sensors
Challenges: Hyper-spectral

- Physical basis for spectral properties
  - Photons will interact with the target on a molecular scale.
- Interdisciplinary research is required to develop the knowledge base for deriving the information content from EM radiation interaction with targets.
Challenges: Hyper-spatial

Zoysia Grass

Ridged Soil - Marietta fine sandy loam

Disked Soil - Leeper silty clay loam

Long Term Fallow Soil - Leeper silty clay loam

Wet Leeper silty clay loam

Standing Water
Challenges: Hyper-temporal

USDA Ultraviolet Radiation Network: MS01 September 2000

- Total Horizontal Radiation
- Diffuse Radiation
- Direct Beam Radiation
Challenges: Hyper-sensors

- The overlapping science instruments of the A-train give a comprehensive picture of the Earth weather and climate.
- Challenge is to develop actionable intelligence from multiple sensors operating at multiple frequencies at multiple spatial resolutions.
Challenges – 6 Platforms & 16 Sensors

- Aura: Atmospheric Chemistry
  - HIRDLS: High Resolution Dynamics Limb Sounder
  - MLS: Microwave Limb Sounder
  - OMI: Ozone Monitoring Instrument
  - TES: Tropospheric Emission Spectrometer

- Parasol: Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar
  - POLDER: Polarization and Directionality of the Earth’s Reflectances

- Calipso: Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
  - CALIOP: Cloud-Aerosol Lidar with Orthogonal Polarization
  - WFC: Wide Field Camera
  - IIR: Imaging Infrared Radiometer

- CloudSat: Cloud Profiling
  - CPR: Cloud Profiling Radar

- Aqua: Earth's water cycle
  - AIRS: Atmospheric Infrared Sounder
  - AMSU-A: Advanced Microwave Sounding Unit
  - HSB: Humidity Sounder for Brazil
  - AMSR-E: Advanced Microwave Scanning Radiometer for EOS
  - MODIS: Moderate Resolution Imaging Spectroradiometer
  - CERES: Clouds and the Earth's Radiant Energy System

- OCO: Orbiting Carbon Observatory
  - Observatory Instrument: Three high resolution grating spectrometers
Challenges – Interoperability of Language

• There is a requirement for automation if the local decision making is going to be available in time
• Communication between machines versus communication between individuals

Earth Observation Semantics
What do we need?

• Making data available timely
  – required for emergency response tasks.
• Better categorization / aggregation of content
  – formulation of custom products and provision of subsetting tools at spatial and temporal levels.
• Interoperability between various formats of data within an archive and between them.
• Package products based on
  – meaning and knowledge about the measurements
  – context of the information sources.
• Creation of machine understandable semantic metadata
  – intelligent search engines / agents can automatically process and index the content.
Information Semantics to Manage Complex Systems

- Provide semantic representation (meaning) for our systems, our data, our documents, our software agents.
- Focus on machines more closely interacting at human conceptual level.
- Spans Ontologies, Knowledge Representation, Semantic Web, Natural Language Processing, Knowledge Management (Reasoning).
- Ontologies (formal models) provide the glue to hold everything together.

Collaborations
- Accelerate research process
- Maximize knowledge discovery
  - Minimize data handling
  - Contribute to both fields
Knowledge Discovery in EO Data

- Services require a rich set of metadata
  - Syntactic (structural) metadata: to provide full description of the physical parameters of the data file
  - Semantic metadata: to provide meaning of the data along with a context, to allow application to understand what it has read and how to use it

- Intelligence/Knowledge— to be able to make decisions via ontologies and a reasoning engine or via a machine learning algorithm or via heuristic algorithm
Resolving Semantic Heterogeneities

Earth Science Models makes Predictions

Semantic Translation

Data

Metadata (FGDC)
generates

Earth Observing Systems

hasSensorType {satellite, airborne, in situ}

hasTemporalResolution

hasSpatialResolution

...

Decision Support Tools

ASCII / binary, HDF/ HDFEOS Data

Semantic Translation

Observations

makes
Semantic Conflicts

- Three main causes of semantic heterogeneities are:
  - *Confounding conflicts* occur when information items seem to have the same meaning, but differ in reality, e.g., different temporal contexts.
  - *Scaling conflicts* occur when different reference systems are used to measure a value, e.g., different currencies.
  - *Naming conflicts* occur when the naming schemes of the information differ significantly, i.e., presence of synonyms.
EO Semantics: Confounding Conflicts

- Instruments sample the environment at various intervals, such as 1 hour or 1 day. Later, the data may be averaged over a longer time period, such as 5 days or a month.
  - Temporal and spatial gaps are thus introduced into the data and it is up to the analyst to resolve these conflicts manually as they are associated with an application specific context for which the data is acquired.

- Necessary to identify whether a value is an intrinsic and permanent property of some instance, or if it depends on some evaluation context (storm surge, coastal ecology, etc) and, in the latter case, by associating this value with its context it is possible to achieve interoperability.
EO Semantics: Scaling Conflicts

The data from coastal sensor systems are highly heterogeneous in syntax, structure, and semantics.

- Coastal-Marine Automated Network (C-MAN) station data typically include
  - barometric pressure,
  - wind direction,
  - speed and gust, and
  - air temperature;
  - some C-MAN stations are designed to also measure sea water temperature, water level, waves, relative humidity, precipitation, and visibility

- Granularity of parameters served by certain other stations is much finer. Semantic conflicts may arise when data from such sources are used in an overall decision-making scenario.
EO Semantics: Naming Conflicts

Commonly observed conflict in coastal sensor data.

- Example: the parameter *windspeed* differs in naming convention between DODS served data (wind_speed) and NDBC (WindSpeed).
- The interpretation of temperature *AirTemp* is defined as 1 meter above sea level for GoMOOS buoy data, whereas it is 3 meters above sea level for the Tropical Atmosphere/Ocean (TAO) array.
- Similar distinctions exist between several coastal sensor networks, that severely constrains the access of the right dataset for a give scenario.
Ontology Integration (Shared)

All Information Sources are related to global ontology

Difficult to compare different source ontologies

Each Ontology can be developed independently

Easy to compare different source ontologies

Shared Vocabulary

- Contains basic terms of a domain which are combined in the local ontologies to describe more complex semantics.
- Easy to add new sources
- Supports acquisition and evolution of ontologies
Framework for Knowledge-base Development

Knowledge-Driven Search and Access Client

- New Knowledge

Combine Ontologies to generate new knowledge

OWL statements about resources (knowledge representation)

Instantiation of the knowledge base (KB) with tuples from the sources

Data Access (Servlets/Wrappers)

- Machine understandable

- OWL Knowledgebases

- Machine readable

Heterogeneous Data Sources

- Text, Spreadsheets etc

- Relational Databases

- XML web services data
Need for Modularization

- Ontologies are built independently and desirable in large heterogeneous environment.
- Changes are not contained locally, but can affect large parts of the model.
- Modularization is desirable and inevitable because it facilitates flexible and efficient reuse of existing ontologies.
Modularization

• Modularization is significant in that it addresses copyright, privacy or security concerns, to make the entire ontology visible to the outside, while willing to expose partial ontology to certain subsets of users.

• This is particularly relevant in EO context as data providers from different sources, while conforming to various policies, can still share knowledge.
Case Study 1 – RS Imagery Retrieval

Semantic Querying

“Retrieve all images from sensor X which contains wetlands near a coast in the Eastern part of country Y”.

• This query requires problem specific discovery of knowledge that is responsive to the needs of an analytical task.
• Middleware is required that provides tools to browse and access the data resources for resolving the heterogeneity problems.
Know-A-Coast Portal

Middleware for Ontology Driven Image Mining (MODIM)

Query result
Ingest Query

DL Reasoner
Query Processing Service

Shared Vocabulary Coastal Zone

Segmentation
Primitive Features
Components Extraction

Application Ontology (Imagery)
Feature Classification

Segmentation
Primitive Features
Components Extraction

Application Ontology (Hydrology)
Feature Classification

Repository

OGC Web Coverage Service (WCS)

Metadata
Data

Metadata
Data

Metadata
Data
Results of a Semantic Query
Semantic Class: Forest
Case Study 2 – Landcover Classification

- Information is normally disseminated through classification.
- Classification systems exist in several domains and also unique to different countries.
- The problem is finding the right data that matches a given criteria.

- **Land Cover**
  - International Geosphere Biosphere Programme (IGBP)
  - United States Geological Survey (USGS)
  - Olson Global Ecosystems (OGE)
  - Simple Biosphere model (SiB)
  - Simple Biosphere 2 (SiB2)
  - Biosphere Atmosphere Transfer Scheme (BATS)

- **Soils**
  - Natural resources conservation service (NRCS)
  - Canadian soil classification system
  - Unified soil classification system

- **Wetlands**
  - U.S. Fish and wildlife service
  - USGS wet land classification
  - Ramsar classification system
  - Cowardian system
Semantic Conflicts between Classification Systems (IGBP and SiB)

International Geosphere Biosphere Programme (IGBP)

Simple Biosphere model (SiB)
Framework for Semantics Enabled Thematic Data Integration (SETI)

- **Semantics Enabled Thematic Data Integration (SETI)**
  - Ontology Layer
    - Application Ontology (IGBP)
    - Application Ontology (BATS)
    - Application Ontology (SiB)
  - Shared Ontology (Land Cover)
  - DL Reasoning Service
  - Query Processing Service
    - Return Concepts
- **Current Systems**
  - Syntactic Level
    - Land Cover Data
      - IGBP
      - BATS
      - SiB
  - Invoke Service
    - Catalog Service
  - Earth Science Gateway

**Miscellaneous**
- Semantic Web Portal
- Electrical & Computer Engineering
- Bagley College of Engineering
- Mississippi State University

*Note: Diagram depicts the integration framework with various components and services.*
Results Retrieved
Case Study 3 – RIIM Coastal Disaster

Know-A-Coast Portal

Middleware for Ontology Driven Image Mining (MODIM)

Query result
Ingest Query

DL Reasoner

Query Processing Service

Shared Vocabulary Coastal Zone

Segmentation
Primitive Features
Components Extraction
Repository

Application Ontology (Imagery)
Feature Classification

Segmentation
Primitive Features
Components Extraction
Repository

Application Ontology (Hydrology)
Feature Classification

OGC Web Coverage Service (WCS)

Metadata
Data
Results Retrieved
Case Study 4 - Integrated Ocean Observing System

Buoy ontology that conceptualizes *in situ* sensors and measurements
Portion of ontology for a Buoy; each station forms an instance of the class Station and is related to OwnedAndMaintainedBy Class through the property hasOwner. Each sensor is associated with parameters (wind speed, SST, salinity etc) which themselves are defined concepts.
Semantic Conflicts

GoMOOS buoy data (DODS)

National Data Buoy Center (NDBC)
Sensor Web Architecture

Electrical & Computer Engineering

Sensor Web Enablement (CosemWare)

- Observables Dictionary
- Observation XSD

SensorML mapping:
- <Sensor Group>
- <TempSensors>
- <Salinity sensors>
- ...

Definition of the geometric, dynamic, and observational characteristics of a sensor

Current System

- Metadata
- Metadata
- Metadata

CosemWare (AJAX Client)

- DescribeSensor
- GetCapabilities
- GetObservation

Observations and Measurements (O&M)

Registry/Catalog (CS-W)

- Sensor Planning Service (SPS)
- Sensor Alert Service (SAS)

Access to Collection of Sensors

Observables Dictionary

- SensorML mapping:
  - <Sensor Group>
  - <wind direction sensor>
  - <wind speed sensor>
  - <salinity sensor>
  - <Waves sensor>

Sensor Observation Service (SOS)

Observables Dictionary

- SensorML mapping:
  - <Sensor Group>
  - <...>

Observables Dictionary

- SensorML mapping:
  - <Sensor Group>
  - <...>

Buoy Sensors Gateway

References

ConstrainedBy
Summary

• Grand challenges to the EO pattern recognition community is to merge geophysics and geoinformation at high resolution, in the right local context, to enable local decision making for socioeconomic benefit.

• How can EO (satellite remote sensing) imagery be used in instantiating a variety of physical process models and in assessing impacts on geopolitical and socioeconomic factors at a scale that impacts me locally?

• What is the role of the EO pattern recognition community in moving beyond research to enable the local decision maker?