Knowledge Mining in Earth Observation Data Archives: A Domain Ontology Perspective

Surya S Durbha & Roger L King
Outline

- Introduction
- Research Objectives
- Architecture
- Ontology
- Feature Extraction
- Machine Learning
- Early Results
- Conclusions
Introduction

- Multiple Petabytes of data from EOS and Pre-EOS archived by NASA DAACs
  - Significant part of the archives have not been analyzed even once.
- NASA’s Earth Observing System Data and Information System (EOSDIS) has become one of the world’s largest and most active repositories of data.
- NOAA
  - National Climatic Data Center
  - National Geophysical Data Center
  - National Oceanographic Data Center
  - National Coastal Data Development Center
- USGS
  - Earth Resources Observation Systems (EROS) Data Center (EDC)
Data Transformation

Distributed Active Archive Centers (DAAC's)

Resource discovery, metadata access, browse data pool

Distributed Data Analysis Centers (Research labs, Universities, etc)

Resolve information heterogeneity (semantic, syntactic, format, etc)

Domain Specific knowledge building through ontological Modeling (OWL, DAML+OIL, etc)
Research Objectives

- Develop Middleware for Ontology Driven Brokering (MOB)
  - Translate metadata to semantic metadata
    - Enables identification of information and relevant knowledge (entities such as sensor type, geographic locations) and their relationships
    - Resource discovery, mediation and transformation
  - Ontology design, integration and deployment
    - Assert Inter-ontology relationships
    - Compute, integrate class hierarchy/consistency.
  - Provide tools for Image knowledge retrieval
    - Development of Application ontology (domain specific)
    - Image segmentation, primitive features, components extraction
    - Apply machine learning for feature classification
- Develop client side tools
  - Functionality to gather information at different levels of granularity, from the sub category to the specific data level
Ontology Integration - architectures

All Information Sources are related to global ontology

Each Ontology can be developed independently

Difficult 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

Easy to compare different source ontologies
Architecture

Internet

GeoPortal

DL Reasoner

Shared Ontology (Domain1)

Shared Ontology (Domain2)

Application1

Ontology (OWL-DL)

Segmentation

Primitive features

Components Extraction

Feature Classification

Repository

Application2

Ontology (OWL-DL)

Segmentation

Primitive features

Components Extraction

Feature Classification

Repository

Application3

Ontology (OWL-DL)

Segmentation

Primitive features

Components Extraction

Feature Classification

Repository

Middleware for Ontology driven Brokering (MOB)

Resource Discovery

Mediation

Transformation

Support for Ontology, design, integration, deployment

OGC Web Coverage Service (WCS)

Metadata

Indexing

Web Map server (OGC Compliant)

Data

Metadata

Indexing

Web Map server (OGC Compliant)

Data

Metadata

Indexing

Web Map server (OGC Compliant)

Data
Ontology Web language (OWL)

- Language for defining structured, web-based ontology”- OGC definition.
  - Richer integration
  - Interoperability of data across application domains
- OWL applications
  - Web portals
  - Agents and services
  - Ubiquitous computing
  - Multimedia collections

Inside Ontology

- Classes+class hierarchy
- Instances
- Slots/values
- Inheritance
- Restrictions on slots( type, cardinality)
- Properties of slots
- Relations between classes
Ontology Driven Applications

Ontology Representation Languages
- Learn
- Manipulate
- Evaluate
- Extract
- Merge
- Annotate

Query/Manipulation languages
- Version
- Transform
- Transfer
- Store
- Reason
- Secure

Ontology

Middleware

Application Client
- Share
- Knowledge Discovery
- Visualize
- Search
- Browse

Visualize
Browse
Search
Knowledge Discovery
Visualize
Share
The Web Coverage Service (WCS) supports electronic interchange of geospatial data as “coverages” - that is, digital geospatial information representing space varying phenomenon - OGC definition.

WCS provides:
- Spatial querying (grid spatial request)
- Reprojection
- Multiple output
- Range subsetting

Provide Coverage in different Formats, BBOX, SRS

Full description of one or more coverages

Web Coverage Service (WCS)
Feature Extraction

- Three level processing sequence consisting of Primitive Features Level (PFL), Intermediate Object Description Level (ODL) and a Higher Conceptual Level (HCL)

Segmentation

- Primitive features level
- Object description level
- Higher Conceptual level

- Color, Shape, Texture
- Object Ontology
- Domain Specific Ontology
Support Vector Machines

- Support Vector Machine (SVM) is a powerful classification method which has shown outstanding classification performance in practice.
  - Simple, and always trained to find global optimum
- In its simplest form an SVM is a hyperplane that separates the positive and negative training samples with maximum margin.
- In the nonlinear case the original feature space is mapped to some higher dimensional feature space where the training set is separable.

\[ \Phi : x \rightarrow \varphi(x) \]
As developed by Scholkopf et al., Kernel PCA (KPCA) is a technique for nonlinear dimension reduction of data with an underlying nonlinear spatial structure.

Kernel PCA is based on the formulation of PCA in terms of dot product matrix instead of covariance matrix.

It is possible to extract non-linear features using kernel functions by solving an eigenvalue problem like for PCA.

KPCA is used to extract structure from high dimensional data set.
Visualization of multiple features
Web Coverage Service

- Request: GetCoverage
- Layer: modis_wisc1
- SRS: EPSG:4326
- BBox: 44187.5, 447562.5, 747125
- Width: 674
- Height: 412
- Format: jpg
Visualization of multiple features
Conclusions

- Framework for content and semantic based information retrieval from remote sensing data archives
- Middleware for Ontology Driven Brokering (MOB)
- Web coverage service integration
- Machine learning methods for image information retrieval
- Early results from the prototype application using Landsat and MODIS data
Thank You!
Backup
Support Vector Machines

- Support Vector Machine (SVM) is a powerful classification method which has shown outstanding classification performance in practice.
  - Simple, and always trained to find global optimum

- It is based on a solid theoretical foundation—structural risk minimization.

- In its simplest form an SVM is a hyperplane that separates the positive and negative training samples with maximum margin.
  - The decision function of an SVM is \( f(x) = \langle w \cdot x \rangle + b \) where \( \langle w \cdot x \rangle \) is the dot product between \( w \) (the normal vector to the hyperplane) and \( x \) (the feature vector representing the example)
  - The margin for an input vector \( x_i \) is \( y_i f(x_i) \) where \( y_i \in \{-1,1\} \) is the correct class label for \( x_i \).
  - In the linear case, the margin is geometrically the distance from the hyperplane to the nearest positive and negative examples.
  - Seeking the maximum margin can be expressed as a quadratic optimization problem:
    \[
    \text{Minimizing } \langle w \cdot w \rangle \text{ subject to } y_i (\langle w \cdot x \rangle + b) \geq 1, \forall i
    \]
Support Vector Machines

Intuitively feels safest

Hyperplane is really simple

If we've made a small variation in the location of the boundary this gives us least chance of causing a misclassification

Robust to outliers since the model is immune to change/removal of any non-support vector data points
**Ontology**

**Why ontologies?**

- Share common understanding of domain
- Reuse domain knowledge
- Make domain assumptions explicit
- Separate domain knowledge from the operational knowledge
- Analyze domain knowledge

---

**Diagram Notes:**

- **S:** subclassOf
- **R:** range
- **D:** domain
- **T:** instanceOf

**Diagram:**

- Root
- Property
- RDF Schema
- Broad leaf/Needleleaf forest
- Needleleaf Evergreen/Deciduous
- Application Specific Data
- Temperate/Arctic
- N. Quebec, Yukon
- Evergreen Needleleaf

---

**Visualization:**

- Class
- Region
- hasLocation
- Location
- hasForest
- Ontology
- Application Specific Data