GEOG 482 / 582 : GIS Data Management

Lesson 15: Web GIS Applications – Data Issues

Overview

Learning Objective Questions:

1. What is spatial data mining?
2. What types of data mining techniques exist?
3. What are the steps for data mining implementation?
4. What are different types of decision problems?
5. What is a general approach to decision workflow?
6. What database systems are needed for decision support?

Lesson Preview

Learning objective questions act as the lesson outline.

Questions beg answers.

Spatial Data Mining

1. What is spatial data mining?

- Early implementations of web systems were used for data management and display.
- Data management of large databases now being used for data mining as a pre-cursor to decision support.
- Spatial data mining – process of extracting interesting and previously unknown information from complex data stored in databases or warehouses
- Commonly called knowledge discovery in databases (KDD); now also goes by name of machine learning within artificial intelligence (AI)
- Data mining techniques are now embedded within various commercial DBMS software

Key Terms
Spatial data mining

Data mining driven by

- Availability of database technology and software tools to search and filter through large databases to detect patterns
- Conventional processing approaches like SQL, statistical analysis, and OLAP techniques are not designed to detect and extract knowledge
- Surge in data processing power, e.g. parallel computers and cloud computing
- Advances in principles coming together from many fields: DBMS, machine learning, information theory, decision science

Data mining differs from SQL and OLAP in several ways, including:

- Designed specifically for very large databases (millions of records), i.e. a big data concern
- Designed more like analysis, than simple retrieval
- Discover patterns, relations, trends not previously seen, e.g., high resolution land cover and stream networks within imagery
- Use machine learning to apply patterns
- Detect characteristics of and correlations among large number of attributes in the dataset, e.g. what feature is related to what feature within what context

2. What types of data mining techniques exist?

- Framework for data mining - See Y&H Fig 11-4 p. 419 for concepts and techniques of data mining

Two general types of data mining: supervised and un-supervised

- Supervised (predictive) data mining
  - Classification
  - Prediction

- Unsupervised (descriptive) data mining
  - Time series analysis
  - Class concept hierarchies
  - Association
  - Clustering
  - Cluster analysis

Data Mining Using Machine Learning

- Learning algorithm discovers the relationships among data in a training set, e.g., streams within unmanned aerial system (drone) imagery
- Use those relationships to examine the larger data set
- Outcomes: classification, patterns, predictions and trends can be computed
- Supervised learning – requires data analyst to identify target field, or attributes to be mined; uses algorithms
- Unsupervised learning - find associations, clusters, trends in data without aid of pre-stated hypotheses or tests.
- See Y&H Fig 11-6 p. 422 for framework of machine learning for spatial data

Classification – as supervised learning for feature category formation

Decision tree – branching from one discovery to another
what attributes belong to what categories as a way to form categories

Neural Networks – build a network of associations
what attributes are associated with what other attributes to form categories

Bayesian classifiers – discovery using context based on conditional probabilities; what is the likelihood (probability) of a given attribute being associated with another attribute given the collection of attributes

Prediction – missing data or forecasts

Ordinary least squares simple linear regression – a dependent variable is predicted by one independent variable. It is based on minimizing errors of the average difference of a trend line among the data values.

Ordinary least squares multiple regression - multiple independent variables are used.

Nonlinear least squares regression – higher power equation than simple linear equation used for independent variables.

Unsupervised (descriptive) data mining

- Time series analysis – relationships examined over time; sequence of observations that repeat as pattern
- Class concept hierarchies – low-level and high-level concept formation based on detection of data values
- Association – what is associated with what in regards to attributes
- Clustering – what is in close relationship to what across space, over time, or as detected through processing of attribute data values
- Cluster analysis – use statistical method to determine closeness, e.g., correlation analysis, analysis of variance, discriminant analysis

Visualization

- Many of the techniques use visualization to support investigation
- Seeking to interpret complex relationships requires high-dimensional visual bandwidth
- Use graphic design to unpack the relationships
- All principles of graphic design apply, looking for relationships

3. What are the steps for data mining implementation?

Cross-Industry Standard Process for Data Mining (CRISP-DM)

- Business understanding – define business objectives for the mining
- Data Understanding – identify data sources to be mined
- Data preparation – characterize the data to be mined
- Modeling – apply the techniques
- Evaluate – how well did the techniques work
- Deployment – provide solution to user

(See CRISP-DM flow Y&H Figure 11-7 p. 441 for an overview.)

Case study: land cover data development in the Chesapeake Bay as an example of machine learning for land cover classification using Esri's ArcGIS tools and Microsoft's Cognitive toolkit

http://chesapeakeconservancy.org/conservation-innovation-center/high-resolution-data/land-cover-data-project/

Spatial Decision Support

4. What are different types of decision problems?

Type of Decision Problem	**4 Decision Problem Components in an Open System**			
	3 Decision Problem Components in a Closed System			
	Content	Structure	Process	Context
Simple	X			
Difficult (semi-structured)	X	X		
Complicated (ill-structured)	X	X	X	
Complex (wicked)	X	X	X	X

Closed system has a finite number of phenomena (parts) and relationships that need be addressed.

Open system has many as yet to be known phenomena (parts) and relationships that need be estimated, as they are uncertain in the computation.

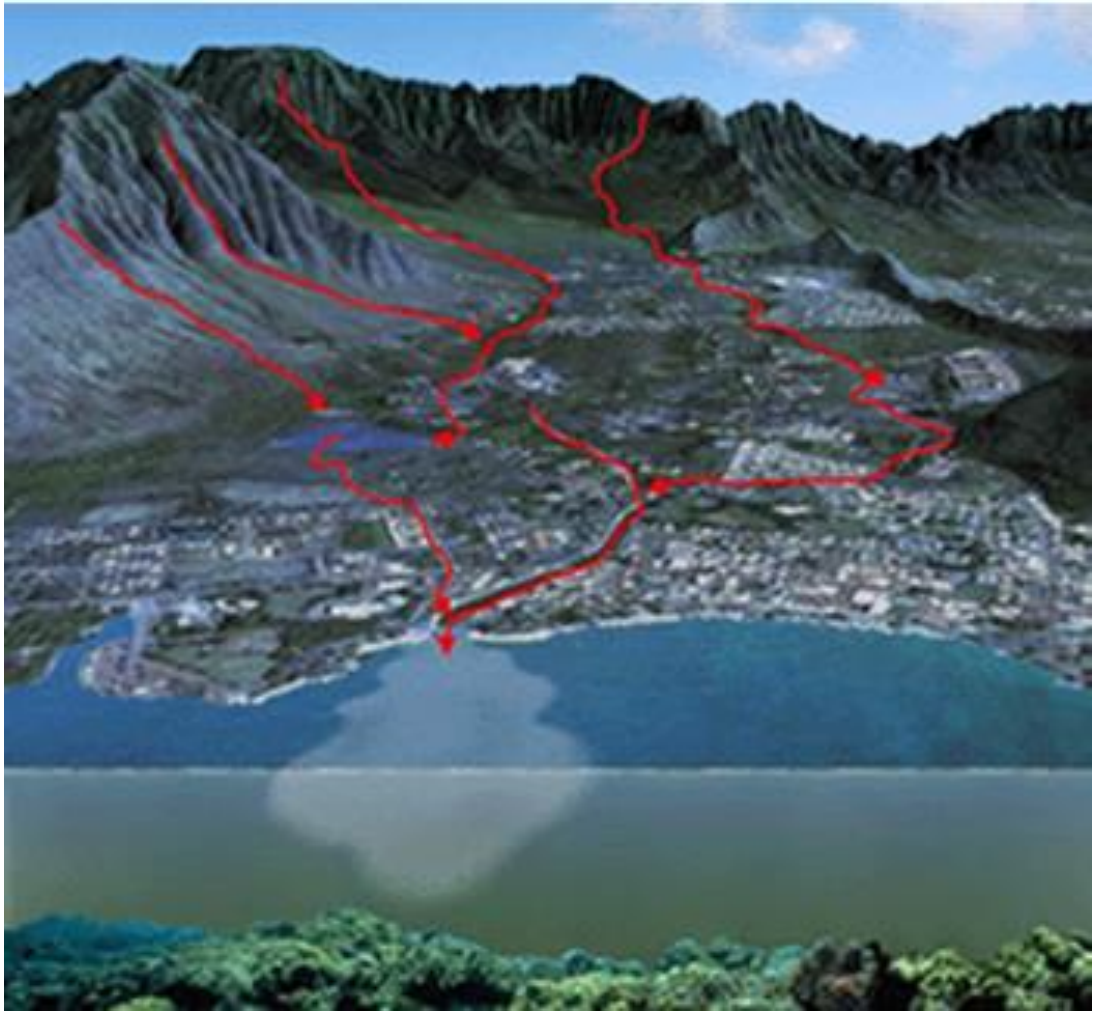
Complex problem situation assessment – data problem

System Content
1.1 Existence/identity as awareness of potential observables for dimensions.
1.2 Observations sampled in terms of units of measurement within dimensions.
1.3 Similarities among observations form a class of fields for observations.
1.4 Object classes specified in a database, with domain delimited for elements within reference systems.
System Structure
1.5 Composite of two or more space-time elements provided by relationships as core of sustainable systems.
System Process (dynamic)
1.6 Functional sustainability relationships within the context of a social-ecological setting.
System Context
1.7 Purpose of functional activity being performed, including expected outcome of activity.

Nyerges et al. 2014, Foundations of sustainability information representation theory: spatial–temporal dynamics of sustainable systems
http://www.tandfonline.com/doi/abs/10.1080/13658816.2013.853304#.Un0G5CcVGSo

Green stormwater infrastructure as a complex systems problem…

What content, structure, process, and context to consider?

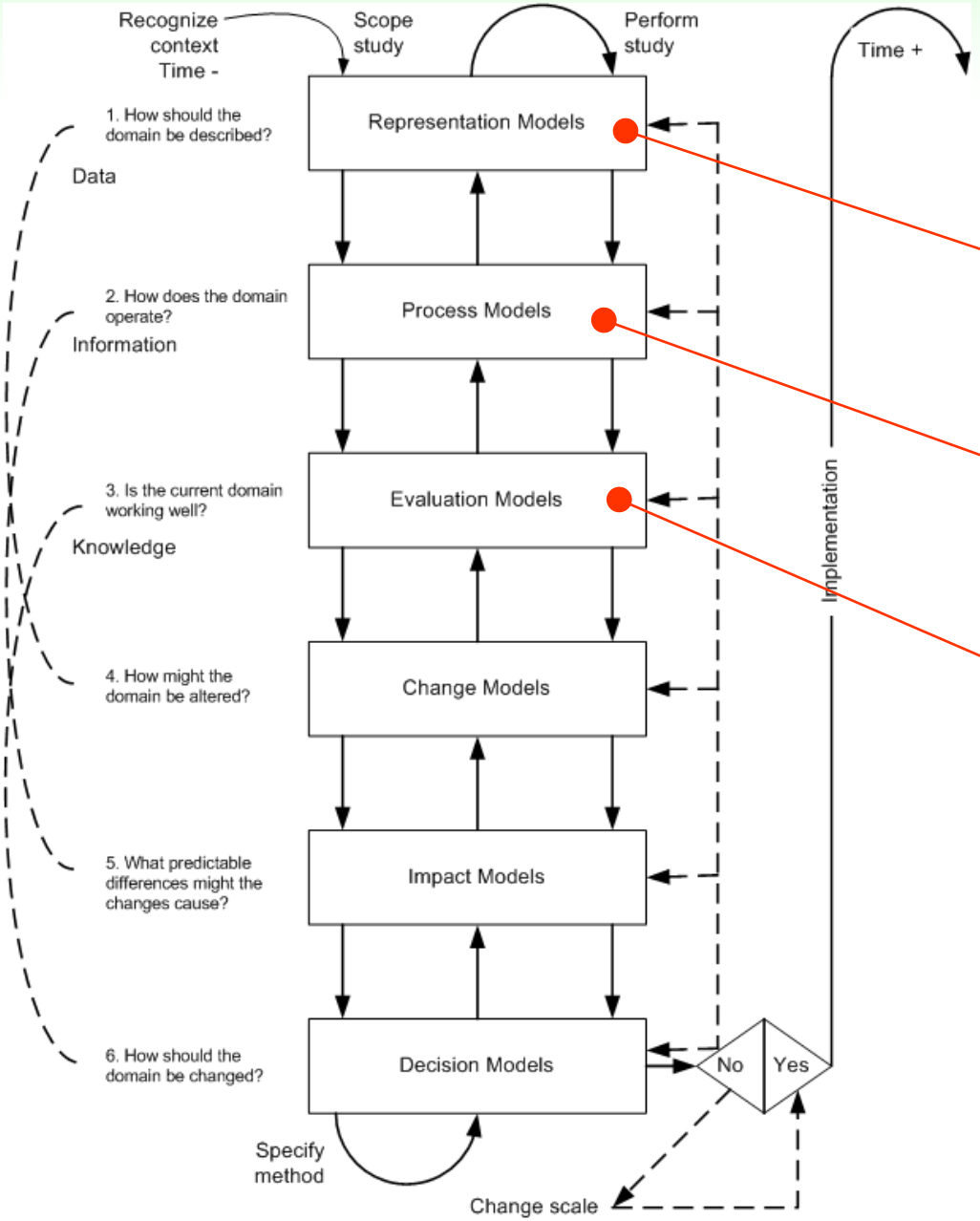


How is i related to j?

1st law of geography…
"Everything (i) is related to everything else (j), but near things are more related than distant things."
(Waldo Tobler 1970)

…near in terms of space, time, and function

Representation in Geodesign Workflow



[Data representation challenge in geodesign workflow about urban watershed management](#) ...

Modeling Step 1: Representation model is fundamentally content and structure.

Modeling Step 2: System process model is a spatial-temporal process.

Modeling Step 3: Evaluation model characterize conditions of the world that motivate decision making: Do nothing or do something to change the world?

5. What is a general approach to decision workflow?

Workflow can involve a considerable number of sequenced tasks

- Macro-micro workflow process
    - Macro stages (steps) for overall flow
        - Intelligence, Design, Choice e.g. according to Simon (1977)
    - Micro activities (substeps),
        - Gather, Organize, Select, Review

Workflow involves geospatial information technology use, possibly for all micro activity steps

Key Term

Geospatial workflow

Macro-micro workflow strategy—general approach

Micro- Activities in a Decision Strategy	Macro-Stages in a Decision Strategy		
	1. **Intelligence** about values, objectives and criteria	2. **Design** of a set of feasible options	3. **Choice** about recommendations
A. Gather...	issues to develop & refine **value trees** as a basis for objectives	**primary criteria** as a basis for option generation	**values, criteria, and option list scenarios** for an evaluation
B. Organize...	**objectives** as a basis for criteria and constraints	and apply approach(es) for **option generation**	approaches to **priority and sensitivity analyses**
C. Select...	**criteria** to be used in analysis as a basis for generating options	the **feasible option list**	**Recommendation** as a prioritized list of options
D. Review...	criteria, **resources, constraints,** and **standards**	**option set(s)** in line with resources, constraints and standards	**recommendation(s)** in line with original **value(s), goal(s)** and **objectives**

Start at Stage 1 and work down the rows A through D.

Then, move to Stage 2 and work down the rows A through D.

Jankowski and Nyerges, 2001 *Geographic Information Systems for Group Decision Making,* Taylor & Francis: London. Table 2.1, p. 17.

Contexts and tasks influence the nature of workflow

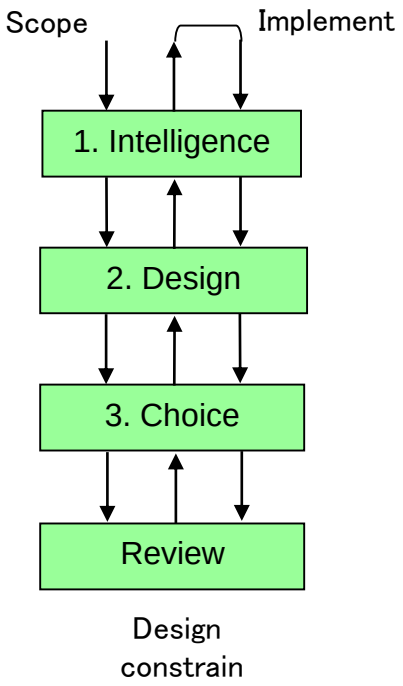
Three-passes (scope, design constrain, implement) undergird the challenge in workflow.

Planning tasks—commonly use longer-term and macro-scale perspectives in a community

Improvement programming tasks—commonly use medium-term and meso-scale perspectives

Implementation tasks—commonly use shorter-term and micro-scale perspectives

Scale is relative to conditions within organization and community



Transportation Improvement Programming workflow

Example workflow using a web-based participatory GIS tool called LIT…

Macro stages (steps)
Micro activities (substeps)

1. Discuss concerns
- 1a: Brainstorm concerns
- 1b: Review summaries

2. Assess improvement factors
- 2a: Discuss factors
- 2b: Weigh factors

3. Create transportation packages
- 3a: Discuss projects
- 3b: Discuss funding options
- 3c: Create your own package

4. Select a package for recommendation
- 4a: Discuss candidate packages
- 4b: Vote on package recommendation

5. Prepare group report
- 5a: Discuss report
- 5b: Vote on report endorsement

Let's Improve Transportation web site for large-scale scale experiment about transportation improvement decision making. The site is no longer available to the public.

The LIT experiment results are described in the following Nyerges_&_Aguirre_2011

6. What database systems are needed for decision support?

Planning tasks need what kind of database system support?

- Considerable number of data elements, managed over long periods of time; granularity of descriptions are rather coarse.

Improvement programming decision tasks need what kind of database system support?

- Few number of features management over time, but with highly interactive, costs and benefits enumerated.

Implementation project tasks need what kind of database system support?

- Considerable number of data elements tracked over fine-grained timeframe.

Summary

In this lesson, you learned about…

1. Spatial data mining
2. Types of data mining techniques that exist
3. Steps for implementing data mining
4. Different types of decision problems
5. General approach to decision workflow
6. Database systems needed for decision support

Contact me at nyerges@uw.edu if you have questions or comments about this lesson.

GEOG 482/582: GIS Data Management

END Lesson 15: Web GIS Applications – Data Issues