Challenges in Data Mining: A Domain Perspective

Peggy Agouris
Center for Earth Observing & Space Research
George Mason University
General Issue

- In addition to mainstream data mining (DM) challenges, the main domain-specific problem in DM is how to include domain knowledge so that queries, processes, and results can be improved.
Geospatial Data

- Large, constantly expanding volumes of data
- Heterogeneous, uncertain, missing, inconsistent data
- Increasingly temporal
- Scale variations in space, time
- Diverse user communities and interaction modalities
Data Mining Issues

- Extraction of patterns to translate measurements into events and activities:
  \[ g(x_i) \Rightarrow A(X, t_1) \]
- Forecasting future events, i.e.
  \[ A(X, t_1) + B(Y, t_2) = ? C(Z, t_3) \]
- Spatiotemporal event-based algebra
- 2-D vs. 3-D vs. 4-D
Domain Challenges

- Moving from *computer* queries to *user* queries: expressing and performing efficiently complex domain-specific queries
- Within-domain (multiple users) and across-domain collaborative DM (intelligence, emergency response)
  - Example: ontologies for cross-application and cross-domain use
Additional Issues

- Incorporation and gains from spatial data characteristics and relationships for knowledge discovery in other disciplines
  - spatial analysis in text data
  - spatial framework for multimedia datasets
- Identification and selection of
  - needed infrastructure
  - algorithms appropriate for geospatial data
**Desired Characteristics**

- Adaptive solutions
- Context awareness
- Novel query interfaces
- Privacy considerations
Contact Info

pagouris@gmu.edu

ceosr.gmu.edu