EXPLORATORY GEOSPATIAL DATA ANALYSIS USING SELF-ORGANIZING MAPS

Case Study of Portuguese Mainland Regions

Main Topics
- Introduction
- Aim of the study
- Study Methodology
- Experimental Results
- Discussion
- Conclusion

KEYWORDS
Geographical Information Systems, Exploratory Data Analysis, Self-Organizing Maps.

Introduction

- Challenges in GIS
  - increasing amounts of data collected and stored
  - most traditional statistical methods not adequate
    - very restrictive assumptions on data
    - or high computational burden
    - complexity of geographic phenomena
- New approaches needed
  - data rather than theory driven
  - transform data into information, and ultimately, into knowledge

Portuguese Mainland Regions

- Delineation of regions
  - proposals presented by several authors
  - a region is characterized by internal homogeneity

Nowadays...

- Public domain geographical data
  - Atlas do Ambiente
    - Portuguese Environmental Agency
- Source thematic maps
  - experiments

Aim of this study

- evaluate the effectiveness of SOM in the exploratory analysis of Portuguese mainland physical geography data
  - Find evidence of Portuguese mainland regions
  - search for natural regions
- Justifications for SOM application
  - ordered mapping from a high-dimensional data space to a low-dimensional space
  - preserving the topological relations in the data
  - stressing of local, regional, factors
Study Framework

- Based on a Knowledge Discovery Process - KDD
  - Based on a Knowledge Discovery Process - KDD
    - Domain knowledge and Goal Definition
    - Data Set
      - Selection, cleaning and preprocessing
    - Algorithm application
      - Data Set Reduction and Projection
    - Exploratory Analysis
      - Interpreting the mined patterns...

Data

- Selection
  - based on literature review
- Preparation
  - Map overlay and intersection
  - Normalization
    - Min-max method

Self-Organizing Map (SOM)

- Also called...
  - Kohonen network, self-organizing feature map...
- Performs both clustering and projection
  - aiming to preserve the topological relations in the data (emphasis on local factors)
- Array (lattice) of elements (units or neurons)
  - arranged in a low dimensionality grid (1D or 2D), the map, for ease of visualization.
- Training:
  - units follow data distribution (DEMO)

Issues in SOM

- SOM application to geospatial data
  - Discrete nature of SOM
  - Enforce SOM geospatial unfolding
    - geo-initialization,
    - weighting of location and samples
    - SOM variant (Geo-SOM)
    - visualization aids to picture SOM grid unfolding.
  - methods to extract and present information from SOM
    - thematic maps of U-mat as groundwork for exploratory analysis.

Visualization aids

- SOM grid
  - projected on location plane
    - color and width of lattice is function of U-mat
- Thematic Maps of U-mat
  - U-mat
    - graphical representation of SOM output
    - each area gets the U-mat value from its BMU
    - high U-mat values (red) separate regions
- Component Planes

Experimental results

- Software
  - Simulation on MatLab with SOM Toolbox 5
  - Geo-processing: ArcGIS 8
- Standard SOM
  - sample weighting with area
    - location weight: 8
    - map size: 24x7
Discussion

- Geographic perspective in SOM
  - Geo-initialization
  - Location and sample weighting
- SOM output visualization
  - SOM grid
  - evaluate SOM unfolding
  - Thematic maps of U-mat
  - assist revealing local similarities between nearby areas
  - patterns reveal regions
- Component planes
  - characterize regions

Map Sketch of Regions

- Based on experiments with Geo-SOM

Conclusion

- No crisp borders delimiting most regions
  - fuzzy, vague, character of geographic phenomena.
  - agreement with reference maps of regions
- SOM
  - effectiveness of SOM in the exploratory analysis
  - physical geography data
  - search for regions