HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Spatial Data Mining Summary

Antti Leino ‹antti.leino@cs.helsinki.fi›

Department of Computer Science

Quick summary of the course so far

- Data mining issues related to spatial data
  - Point vs. continuous vs. area data
  - Spatial (auto)correlation
- Co-location patterns
  - Mainly in point data
  - Background: frequent sets and association rules in transaction data
  - Extensions:
    - Spatiotemporal patterns
    - Combining point and area data

Quick summary of the course so far

- Clustering
  - Finding optimal *n*-centre configurations
    - ATM's, cellular phone networks and the like
    - Minimising the overall distance from point to centre
  - Finding contiguous high-density areas
- Modelling
  - Regression models
  - Finding outliers

What now?

- Regular sessions are finished
- Still to come:
  - Exam next Thursday (3.5., 16–19)
  - Course work by Wednesday, 16.5.
- So: what to do?

Exam

- Thursday, 3.5., 16–19 B123
- 3–6 questions
  - Short essays (up to one page or so)
  - Term definitions
  - Algorithm descriptions (not in detail)
  - »How would you start mining this data?»
- Material
  - Notes for introductory lectures
  - Main ideas and basic concepts of the articles

Examples of questions

- In a couple of sentences, define *participation index* and *participation ratio*
- Present a broad outline of the joinless co-location mining algorithm in Huang (2005)
- How is the discovery of co-evolving spatial patterns different from the discovery of purely spatial co-location patterns?
- You have a data set that consists of 16 point patterns and the measurements of 5 different continuous variables at various locations. How would you start mining this data?

Course work

- Submit by e-mail as a pdf file by Wednesday, 16.5.
- About five pages of text, either
  - Course diary:
    - About half a page per session
    - What did you learn this time?
    - What was good? What was bad?
  - Essay
    - Based on two or more articles not covered during the course
    - Links to suggested articles on the course page
  - Data mining exercise
    - Report on a hands-on experiment
    - Implementation of an algorithm
    - Mining and analysis of own spatial data

Thank you