15-826: Multimedia Databases and Data Mining

Lecture #7: Spatial Access Methods - Metric trees

C. Faloutsos

Optional material

Outline

Goal: ‘Find similar / interesting things’
- Intro to DB
- Indexing - similarity search
  - Data Mining

Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
  - problem dfn
  - z-ordering
  - R-trees
  - misc
- fractals
- text
SAMs - Detailed outline

- spatial access methods
  - problem dfn
  - z-ordering
  - R-trees
  - misc topics
    - metric trees
- fractals
- text, ...

Metric trees - problem

- What if we only have a distance function \(d(o1, o2)\)?
- (Applications?)
A: Metric trees

- M-trees = ‘ball-trees’: Minimum Bounding spheres

Metric trees

- (assumption: $d()$ is a metric: positive; symmetric; triangle inequality)
- then, we can use some variation of ‘Vantage Point’ trees [Yannilos]
- many variations (GNAT trees [Brin95], MVP-trees [Ozsoyoglu+] ...)

Copyright: C. Faloutsos (2019)
Metric trees

• Finally: M-trees [Ciaccia, Patella, Zezula, vldb 97]
• M-trees = ‘ball-trees’: groups in spheres
Metric trees

• Finally: M-trees [Ciaccia, Patella, Zezula, vldb 97]
• M-trees = ‘ball-trees’: Minimum Bounding spheres
Metric trees

- Search (range and k-nn): like R-trees
- Split?

Split? Several criteria:
  - minimize max radius (or sum radii)
  - (even: random!)

- Algorithm?
Metric trees

- Search (range and k-nn): like R-trees
- Split? Several criteria:
  - minimize max radius (or sum radii)
  - (even: random!)
- Algorithm?
- eg., similar to the quadratic split of Guttman

SAMs - Detailed outline

- spatial access methods
  - problem dfn
  - z-ordering
  - R-trees
  - misc topics
    - metric trees
- fractals
- text, ...
Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- Find cities within 100mi from Pittsburgh

Conclusions for SAMs

- z-ordering and R-trees for low-d points and regions – very successful
- M-trees & variants for metric datasets
- beware of the ‘dimensionality curse’
  - Estimate ‘intrinsic’ dimensionality (‘fractals’)
  - Project to lower dimensions (‘SVD/PCA’)

Copyright: C. Faloutsos (2019)
References