

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM – 2010)

Spatial and Spatio-Temporal Data Mining

Vania Bogorny

Universidade Federal de Santa Catarina, Brazil
Department of Informatics and Statistics
www.inf.ufsc.br/~vania
vania@inf.ufsc.br

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

1

Outline

- Part I
 - ✦ Introduction to Spatial Databases
- Part II
 - ✦ Spatial Data Mining Methods
- Part III
 - ✦ Trajectory Data Mining

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

2

Outline

Part I

- Spatial/Geographic Databases
 - ✦ Geographic Data
 - ✦ Spatial Relationships / Spatial Operations
 - ✦ Spatial Query Language
- Introduction to Spatial Data Mining
- Spatial Data Mining Methods
 - ✦ Co-location
 - ✦ Outliers
 - ✦ Spatial Association Rules
 - ✦ Classification
 - ✦ Clustering

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

3

Introduction to Spatial Databases

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

4

What is a Spatial/Geographic Database?

Database that:

- Stores spatial objects
- Provides data types for spatial objects
- Provides operations to manipulate spatial objects
- Manipulates spatial objects just like other objects in the database

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

5

Spatial Databases

- Oracle Spatial
- IBM DB2 Spatial Extender
- Informix Spatial DataBlade
- PostGIS / PostgreSQL

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

6

What is a Spatial/Geographic Data?

- Data which describe a location or a shape
 - e.g. House, Hospital, Road, River, Forests, Parks, Soil
- Is something that describes objects or phenomena that happen on the Earth and that have associated a geographic position

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

7

What is a Spatial/Geographic Data?

- Three main characteristics describe a geographic object:
 - Non-spatial attributes (what):** describe either quantitatively or qualitatively a geographic entity.
 - These attributes may be treated by non-spatial databases;
 - Spatial attribute (where):** describe the spatial location and representation of the geographic object, considering the geometry and a coordinate system.
 - This aspect requires a specific data type not available in conventional DBMS;
 - Spatial relationships (how):** neighbourhood relationships (e.g. topology, distance).
 - Requires special operations that are not available in conventional DBMS;

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

8

How can we classify geographic data?

- Objects
- Phenomena (field)

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

9

Spatial Representation: OBJECT

- Discrete objects (Features): well defined border/limit

0-dimensional

- representation: point
- E.g.: school, hospital

1-dimensional

- representation: line
- E.g.: river, road

2-dimensional

- representation: polygon
- E.g.: state, city

3-dimensional

- representation: surface

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

10

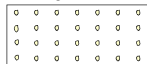
Spatial Representation: FIELD

Continuous Data

Irregular points (e.g. temperature)



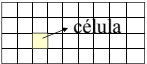
Regular Points



Isoline (e.g. relief)



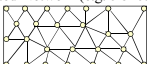
Grid (e.g. satellite image)



Adjacent polygons (e.g. soil)



Triangulated Network (e.g. the floor of a valley)



15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

11

Geographic Data and Geographic Databases: an example

Street

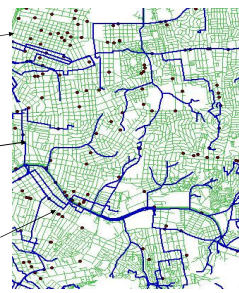
Gid	Name	The_geom
1	Ijuí	Multiline [(x1,y1),(x2,y2)...]
2	Lavras	Multiline [(x1,y1),(x2,y2)...]

WaterResource

Gid	Name	The_geom
1	Jacui	Multiline [(x1,y1),(x2,y2)...]
2	Guaíba	Multiline [(x1,y1),(x2,y2)...]
3	Uruguai	Multiline [(x1,y1),(x2,y2)...]

GasStation

Gid	Name	VolDiesel	VolGas	The_geom
1	BR	20000	85000	Point[(x1,y1)]
2	IPF	30000	95000	Point[(x1,y1)]
3	Eso	25000	120000	Point[(x1,y1)]



15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

12

Spatial Relationships

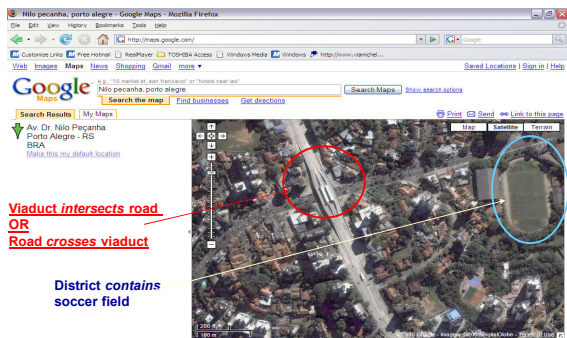
- Main characteristic which differs spatial data from non-spatial data
- Main aspect to be considered in SPATIAL DATA MINING

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

13

Spatial Relationships: an example

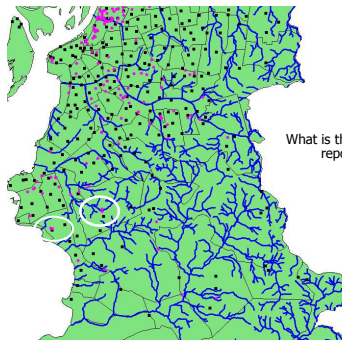


15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

14

Spatial Relationships



What is the average distance between industrial repositories and water collecting points?

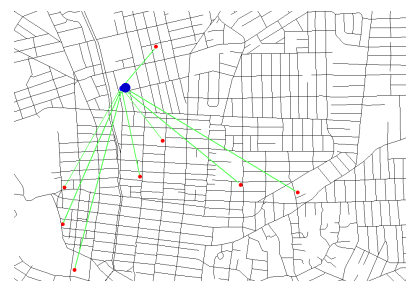
- Water collecting points
- Districts
- Industrial Repositories
- Water Body

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

15

Spatial Relationships



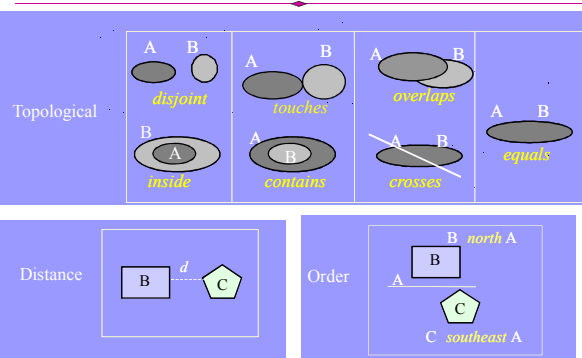
Which is the closest pub to my house?

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

16

Main Spatial Relationships

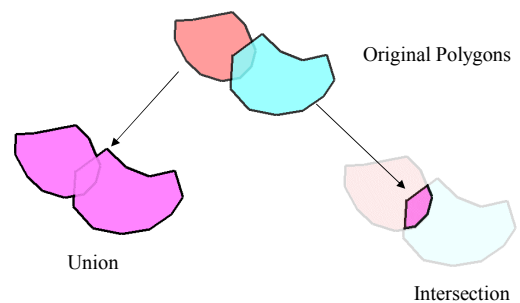


15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

17

Examples of Spatial Operations

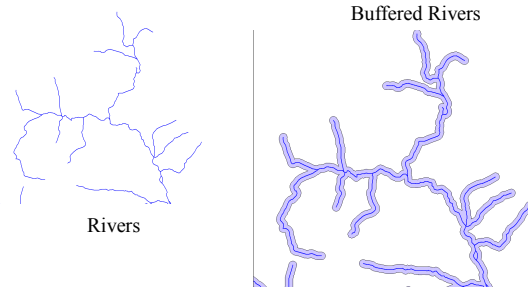


15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

18

Examples of Spatial Operations



15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

19

Spatial Query Example

Q1: Retrieve the rivers and countries that have the relationship "crosses".

Query

```
SELECT r.name, c.cntry_name
FROM river r, country c
WHERE crosses (r.the_geom,c.the_geom) = 'True'
```

Answer

name	cntry_name
Pembina	United States
Pembina	Canada
Rainy	United States
Rainy	Canada
Souris	United States
Souris	Canada
Red River of the North	United States
Red River of the North	Canada

(8 rows)

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

20

References

- GUTING, R. H. An Introduction to Spatial Database Systems. The International Journal on Very Large Data Bases, [S.l.], v.3, n.4, p. 357 – 399, Oct. 1994.
- RIGAUD, P.; SCHOLL, M.; VOISARD, A. Spatial Databases: With Application to GIS. San Francisco: Morgan Kaufmann, 2002.
- SHEKHAR, S., CHAWLA, S. Spatial databases: a tour. Upper Saddle River, NJ: Prentice Hall, 2003.
- OPEN GIS CONSORTIUM. Topic 5, the OpenGIS abstract specification—OpenGIS features—Version 4. 1999a. Available at <<http://www.OpenGIS.org/techno/specs.htm>>.
- OPEN GIS CONSORTIUM. OpenGIS simple features specification for SQL. 1999b. Available at <<http://www.opengeospatial.org/docs/99-054.pdf>>.
- OPEN GIS CONSORTIUM. Feature Geometry. 2001. Available at <<http://www.opengeospatial.org/specs>>.

15/12/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

21