dads 2009|2010 spring

Theme: ***Data Mining for Architecture and Urban Planning***

Lecture III

Data Mining in Context

Ahu Sökmenoğlu, M. Arch.

Ceyhun Burak Akgül, PhD in EE

In this lecture

- Who applies data mining?
- Who applies data mining on social and spatial data?
- A brief clarification of contextual concepts & domains
- In the Domain of Geography:
  - What is spatial data mining?
  - What is temporal data mining?
  - What is spatio-temporal data mining?
- In the Domain of Urban Research:
  - What is spatial/urban data?
  - What is social data?
  - An example of Urban Data Mining
- Introduction to Semester Project:
  - How we can identify the patterns and trends of socio-spatial Activities of Architecture Student Community in Istanbul by Data Mining?

Applications: Classical

- Marketing (e.g., sales analysis)
- Banking (e.g., credit and loan approval)
- Medicine / Biology / Pharmacology
- Manufacturing (e.g., yield analysis)
- Finance (e.g., stock prediction)
- E-Commerce / Web (e.g., hits analysis)

>>

Applications: Less classical

- Multimedia search engines
  - Content search by audiovisual similarity
- Multimedia content management
  - Automatic content categorization and annotation
  - Audiovisual concept detection
- Image-based medical diagnosis
  - Visual biomarker discovery from medical images

Our context:

Who applies data mining on social and spatial data?

- Geographers/Urban geographers
- Urban Planners and Architects
- Social scientists

What is …?

Geography* (from Greek *γεωγραφία - geographia*, lit. "earth describe-write") is the study of the Earth and its lands, features, inhabitants, and phenomena. Modern geography has two main branches;

- human geography
- physical geography

*Wikipedia definition

What is …?

Urbanism* is a focus on cities and urban areas, their geography, economies, politics, cultural, socialcharacteristics, as well as the effects on, and caused by, the built environment.

Urbanism** is a way of life associated with residence in an urban area.

When something is described as **urban** it is related to or concerned with a city or densely populated area that thing is located in a city or it is a characteristic of a city or city life.

*Wikipedia definition **Michael Pacione's definition in *Urban Geography*

What is …?

Urban geography* is concerned to identify and explain the distributions of towns and cities and the socio-spatial similarities and contrasts that exist within and between them. It is both the study of system of cities and the study of the city as a system.

*Michael Pacione's definition in *Urban Geography*

What is …?

Space* is the unlimited expanse in which everything is located.

Spatial means pertaining to or involving or having the nature of space.

*http://wordnetweb.princeton.edu definition

What is …?

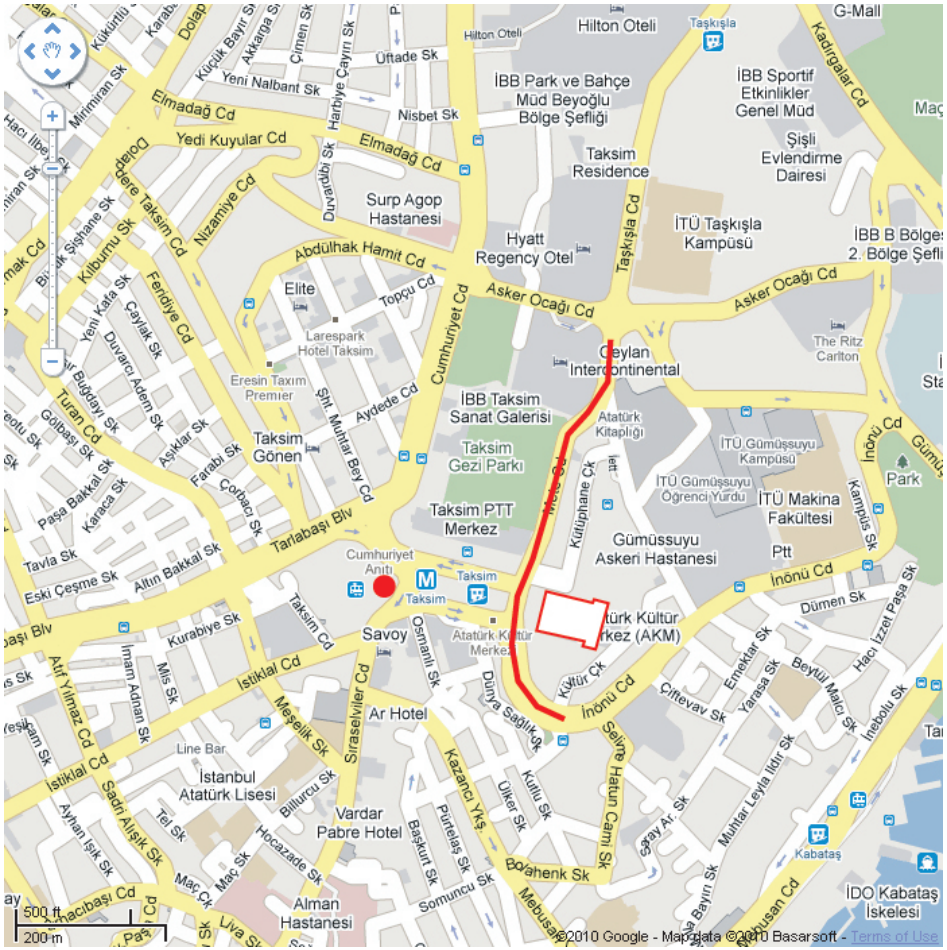
Social* relating to human society and its members

*http://wordnetweb.princeton.edu definition

In the domain of geography:
What is spatial data*?

Data that can be geometrically presented on a 2d plane as a point, line or surface in traditional cartography, topography, cadastral and relevant applications.

(*name of a city, population of a city, depth of a lake are not spatial data.)



What is spatial data?

Spatial data exhibits a unique property in that;

Tobler's (1979) first law of geography;

"everything else but nearby things are more related than distant things"

What is spatial data mining?

The goal of spatial data mining is to discover relationships between data and non-spatial data by using;

- spatial proximity relationships (such as topological relationships like intersects, overlap, disjoint);
- spatial orientation (such as left of, east of);
- distance information (such as close to, far away).

What is spatial data mining?

The patterns discovered from spatial data includes;

- Characteristics and discriminant rules
  - The description of general weather patterns in geographic regions
  - The comparison of the weather patterns in two geographic regions
- Spatial association rules
  - Large towns in British Columbia are closer to the sea
  - %80 of schools are close to parks
- Extraction and description of prominent structures or clusters, …
  - Smoke aerosols alter the likelihood of rainfall in a nearby region (co-location; instances that are located in the same neighborhood)

What is temporal data?

- Time series data;

- Sequence of real numbers that vary with time such as stock prices

ornekle

- Sequence data;

- Sequence of ordered events with or without concrete notions of time such as web page traversal sequences

ornekle

What is temporal data mining?

- Temporal data mining aims to discover and interrelationships of contextual and temporal proximity among the data.
- Patters that can be discovered from temporal data includes:

- temporal association patterns; *coffee and doughnuts have a strong tendency to occur together with high support during the time interval 7AM-9AM*
- sequential patterns; *a customer typically rents Star Wars, then Empire Strikes Back and then Return of the Jedi.*

What is spatio-temporal data mining?

- A spatio-temporal database embodies spatial, temporal and spatio-temporal concepts and captures simultaneously the spatial and temporal aspects of data.

- Spatio-temporal patterns does not only link events in different locations but also establish the sequence of changes of events in these locations.

What is spatio-temporal data mining?

- *Insert an example of spatio-temporal database see intoduction p. 7*
- **"There is a high incidence of earthquakes in a region during or soon after high atmospheric pressure in the nearby region."**
- **"Forest fire always occur at region R1 prior to the occurrence of haze in the nearby region R2, then a drop in the atmospheric pressure at region R3 then rainfall region R4."**
- **"From March to April, if there is a forest fire in a region in South Asia, haze and rainfall will subsequently occur in its Southeastern neighbors."**

The domain of urban research:
What is spatial/ urban data?

The data concerning the information about location, quantity, usage, orientation, dimensions, density, morphological characteristics, topological relations, continuity, connectivity, accessibility, etc. of urban entities.

What is social data?

The data concerning human related information such as population, gender, cultural and ethic identities, age, education level, occupation, employment status, economical status, personal histories, personal preferences, likes/dislikes, daily activities and paths, political opinions, etc.

An example for urban data mining

On the Discovery of Urban Typologies:

Data Mining the Multi dimensional Character of Neighbourhoods (Gil, J., Montenegro, N., Beirao, J. N., Duarte, J. P.,

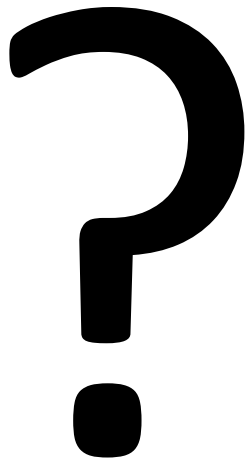
Attribute	Entity	Code	Calculation
Length	Street, Block	LEN	m
Width	Street, Block	W	m
Orientation	Street, Block	OIR	degrees
Solar Orientation	Street, Block	SOLO	N,S,E,W
Number of Buildings	Street, Block	BLDN	integer
Area	Block	TA	m2
Built-up area	Block	BA	m2
Perimeter	Block	PER	m
Proportion	Block	PROP	LEN/W
Compactness	Block	CMP	A/PER
Floor Area Ratio	Block	FAR	GFA / TA
Ground Space Index	Block	GSI	BA / TA
Layers	Block	L	GFA/BA
Open Space Ratio	Block	OSR	(TA - BA) / GFA
Private space area	Block	PRVA	m2
Public space area	Block	PUBA	m2
Pavement width	Street	PAVW	m
Pedestrian area	Street	PEDA	m2
Continuity	Street	CNT	Links
Connectivity	Street	CON	Degree
Global accessibility	Street	ACCG	Closeness
Local accessibility	Street	ACCL	Closeness
Global movement flow	Street	MOVG	Betweenness
Local movement flow	Street	MOVL	Betweenness

Semester project:

- What we are aiming to?
  - Our aim is to explore the patterns and trends of socio-spatial activities of architecture student community in Istanbul.

- What will be our final outcome?
  - The subject of the studio is to design a thematic city map of activity patterns and trends for different groups of architecture students.

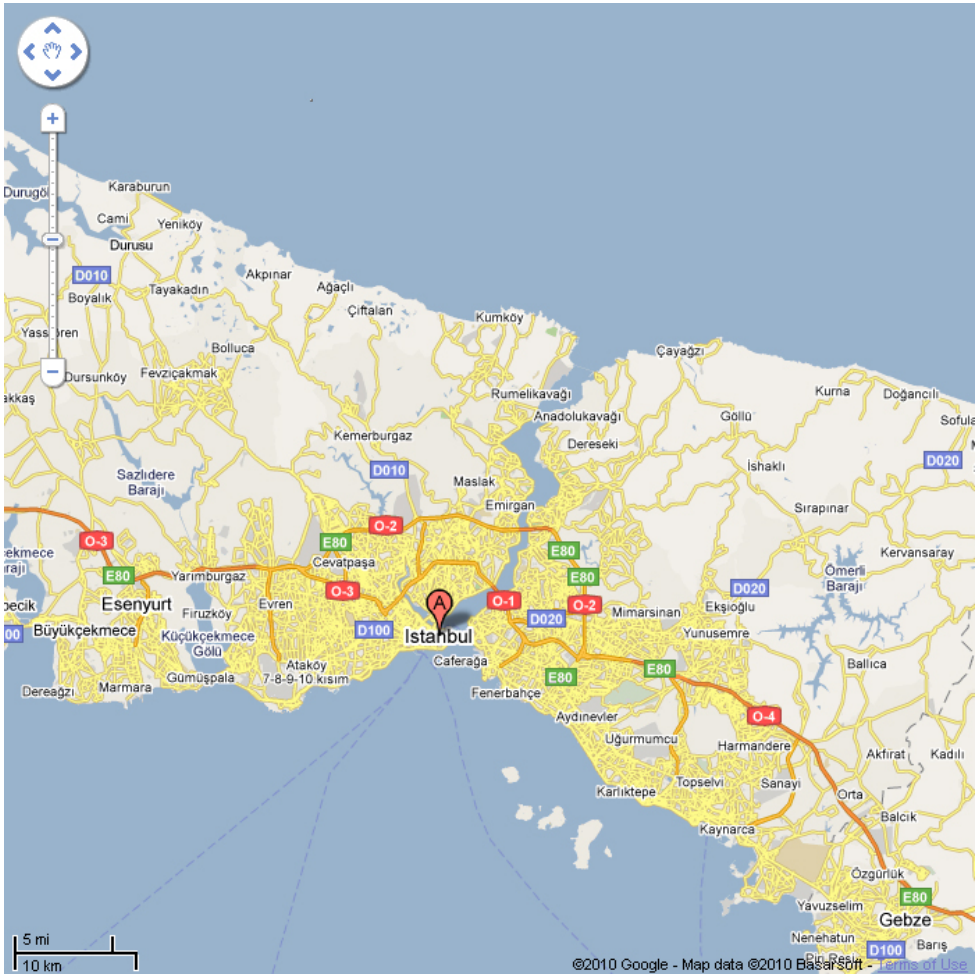
What is the motivation behind the formulation of this course



Semester project

Magritte's "La Trahison des Images" ("The Treachery of Images") (1928-9) or "Ceci n'est pas une pipe" ("This is not a pipe").

Semester project



"Ceci n'est pas l'espace"

"This is not space"

In reference to "Ceci n'est pas l'espace" Figure 11.1 in Massey, 2005

What is the motivation behind the formulation of this course?

No longer think of urban space as purely physical, and static; instead in the scope of this studio we propose to take it as social and relational.

- A map of roads and motorways, railway lines, topography fields and villages is not space. Odd because maps have become central to how we think about and imagine space.
- Hegemonic types of mapping represent space as a “completed horizontality” in which the dynamism of change is exorcised in favour of a totality of connections. Mapping is one of a number of ways in which the disruptiveness of space is tamed.

- An alternative non-euclidean imagination of space, that disrupts this
- Space is the product of interrelations, constituted through interactions from the immensity of the global to the intimately tiny
- Space is the sphere of possibility of the existence of multiplicity, space as the sphere in which distinct trajectories coexist
- Space is always under construction, always in the process of being made, never finished, never closed.

What is missing?

"Urban geography has not been particularly good at, or indeed often even interested in, making sense of many of the ***smaller elements that make up a city.*** *"*

What might be those small elements that make up a city?

"The day-to-day routines of shopping and household provisioning, the life of public places like parks, sidewalks and shopping malls, the networks of friendship and enthusiasm, and so forth that give urban life so much of its texture."

Semester project:

- How we will do that?

  1st step: We will collect raw data on the daily activities of students through dedicated questionnaires, surveys, and interviews.

  2nd step: We will then use data mining methodologies in order to understand "what this specific data says" about questions related to socio-spatial activities.

What are the questions that we might want to get answers to?

- How to find groups (clusters) of students with the same characteristics with regards to their socio-spatial activities?
- How to classify places where socio-spatial activities occur?
- Are there any spatial association rules describing the implication of one set of features by another set of features such as "most preferred cheap eating places are close to the university"?
- Are there any co-location patterns such as "shopping places that students prefer are frequently located in downtown"?
- Which variables concerning student's profiles matter the most for choosing specific activities?

What are the data types which we might be interested in using in the scope of the project?

What is data mining?

Sources of information

- Humans
  - Our senses
  - Our likes/dislikes, favorite books, movies, music, …
  - Our identity, political views, web browsing sessions, …
  - Our genome, medical records, …
- Earth and Universe
  - Stars and planets, celestial motions, …
  - Climate, geography, oceans, …
- Organizations/Companies
  - Expenses and revenues, stock rates, …
- …

2/23/2010 dads 2009|2010 spring 17

Cities
Districts
Neighborhoods
Streets
Blocks
Buildings
Transportation
…

What are the basic questions we might want to formulate in order to get the desired data?

- How do the students socially use/experience urban space and for what purpose?
- Where and why do they meet?
- Where and what do they eat/drink?
- Where do they have fun?
- Where are their favorite open-spaces?
- What do they feel in specific locations?
- How do they describe specific locations?
- In which proportion do they spend money/time on socio-cultural activities?
- …

This semester

week	date	studio
1	**9-Feb**	**-**
2	**16-Feb**	**Introduction: Data Mining in General**
3	23-Feb	Concepts in Data Mining
4	2-Mar	Data Mining Applications in Context Introduction to Semester Project
5	9-Mar	Statistics Primer
6	16-Mar	A Broad Picture of Data Mining Tools Jury Meeting; Semester Project's first concepts & ideas
7	23-Mar	Regression and Classification
8	30-Mar	Clustering, Exploratory Data Analysis, and Visualization Semester Project's review
9	6-Apr	Semester Project's review
10	13-Apr	Semester Project's review
11	20-Apr	Jury Meeting; Presentations
12	27-Apr	Semester Project's review
13	4-May	Semester Project's review
14	11-May	Jury Meeting; Final Presentations

Assignments for next week