Introduction to Spatial Big Data Analytics

Zhe Jiang

zjiang@cs.ua.edu
Office: SEC 3435
What is Big Data?

- Examples

  - Facebook
  - Google
  - Amazon
  - LinkedIn
  - Alibaba
  - Netflix
  - eBay
  - Instagram

  Internet data (images from the web)

- Earth observation data (nasa.gov)

  - Pratt & Whitney's Geared Turbo Fan Engine
    - 5,000 sensors
    - 10 GB data per second
    - 12 hours of flight + 844 TB data

  - Healthcare data

  - Manufacture sensor data

Copyright of logos belong to the companies.
What is Big Data?

• Subjective, but some popular def.
  • Volume, velocity, variety, value
  • High dimensionality
  • Data characteristics exceed capability of existing computational techniques!

• Illusion or reality?
  • Business: mobile ads, personalization
  • Manufacture: driverless car
  • Health: diagnosis, personalized medicine
  • Science: bioinformatics, GIScience

• Shortage of skilled people in big data!
Spatial Big Data (SBD)

• Geo-referenced big data

• Examples:
  • GPS trajectories
  • Check-in records
  • Earth observation imagery
  • Spatial events, e.g., crimes, accidents
  • Climate model simulations

• Why “spatial” matters?
  • Impact everyday life
  • Computational challenges
Spatial Big Data Analytics

• SBD analytics
  • Process of identifying useful patterns or making predictions from SBD

• Types of patterns:
  • Spatial prediction: earth image classification
  • Colocation: crime and bars
  • Hotspot: disease outbreak
  • Spatial summarization, change, outlier, etc.

Descriptive v.s. Predictive?
Spatial Big Data Analytics Applications

Location based services

Precision agriculture

Disaster response (before and after flood)

Disease outbreak detection

Expedia

gps4us.com

innovationtoronto.com

arcgis.com

cdc.gov

Image from the web
Spatial Big Data Analytics: Challenges

Special challenges for spatial big data:
- **Autocorrelation**: interdependency across nearby locations
- **Anisotropy**: dependency varying across directions
- **Heterogeneity**: similar feature different classes
- **Scales**: multiple resolutions; local, regional, global scales
Spatial Big Data Analytics: Challenges

(a) Aerial photo (NIR,G,B) in spring
(b) Aerial photo (R,G,B) in summer
(c) Ground truth wetland map
(d) Decision tree prediction

Need novel spatial models, e.g., *spatial decision trees*

- Salt-and-pepper noise in decision tree prediction
- Require labor intensive pre/post-processing
- **General issue** for many models: random forest, SVM, neural network.
Spatial Data Scientist

A specialized doctor, e.g., Cardiologist

Spatial data scientist?
Process of Spatial Big Data Analytics

Input data → Preprocessing, Exploratory data analysis → Spatial Data Mining → Output patterns

Spatial statistics
Computational infrastructure

domain expert interpretation
Technical View of Spatial Big Data

Data analytics (effectiveness)

Question: what are the differences between “statistics”, “machine learning” and “data mining”?

Domain Knowledge
- Understand the problem
- Interpret results
- E.g., biology, agriculture

The main focus of this course is spatial data analytics!
Prerequisite

• Database
• Algorithm
• Basic statistic and probability
• Programming skills
• Additional data mining or machine learning background is preferred
Topics Covered in This Course

• Spatial data models
• Spatial statistics
• Spatial big data analytics tasks
• Spatial big data platforms
• Recent trends
Topics Covered in This Course

• Spatial data models

http://desktop.arcgis.com/

Field model

Object model

Spatial reference system

Spatial operations

Google map

http://desktop.arcgis.com/

Mississippi River

Des Moines River
Topics Covered in This Course

- Spatial data models
- **Spatial statistics**

Geostatistics, e.g., Kriging (arcgis.com)

Areal data model

Spatial point process (shooting in Chicago)
Topics Covered in This Course

• Spatial data models
• Spatial statistics

**Spatial big data analytics tasks**

• Spatial prediction
• Colocation
• Spatial outliers
• Spatial hotspot
• Spatial summarization
• Spatial change

Wetland mapping

Colocation patterns

Spatial outliers

Crime hotspot

Trajectory summarization
Topics Covered in This Course

- Spatial data models
- Spatial statistics
- Spatial big data analytics tasks
- **Spatial big data platforms**
- Recent trends
Topics Covered in This Course

• Spatial data models
• Spatial statistics
• Spatial big data analytics tasks
• Spatial big data platforms
• Recent trends
Course Website and Email

• Course website:
  • http://zhejiang.cs.ua.edu/Teaching/Spring17/SpatialBigData/
  • Class schedule, notes, homework, announcement, etc.

• Course email:
  • cs491591@gmail.com
  • Homework submission, questions
  • Please put “CS491/591: Spatial Big Data” in email title
Assignments

- *Four* question-and-answer homework
- *One* course project (see next slide)
- *One* news presentation
  - 5 minutes at the beginning of each lecture
  - Topic related to geospatial (big) data analytic

**Notes:**
- All assignments are done in teams of two. Find a partner
- Due by the beginning of the first lecture in each week
Course Project

• Topic
  • Real world spatial (big) data analytic problem
  • Students are welcome to bring their own topics
  • We provide some candidate topics
  • Discuss with the instructor in office hours to decide topics

• Progress management
  • Project assignments (deliverables)
    • Decide project topic and datasets
    • Literature survey (data preprocessing)
    • Proposed approach (implementation)
    • Experimental evaluation
    • Final report (w/ source codes, dataset)
  • Midterm project presentation
  • Final project presentation

• Come to my office hours often for feedback and guidance!
Exams

- One close-book midterm exam
- (Possibly one close-book final exam)
- Some questions will be different between CS 491 and CS 591
Grading Policy

• News presentation: 5%
• Question and answer assignment: 30%
• Project assignments: 40% (or 35% if two exams)
• Exams: 25% (or 30% if two exams)
Homework 0 Due Next Week

- A background survey
- Not graded
Questions?

• Introduce yourself to other people, and team up!