Welcome to

DS504/CS586: Big Data Analytics --Introduction & Logistics

Prof. Yanhua Li

Time: 6:00pm –8:50pm THURSDAY Location: AK 232 Fall 2016

Statistics

- 1. Registered
- 2. DS/CS
- 3. 2+nd year Graduate
- 4. DS/CS 2+nd year
- 5. PhD

Roadmap

- Logistics
 5 minutes break
- 2. Intro
 - 10 minutes break, talk to other students Self-intro (and group forming)
- Data Acquisition and Measurement Hand in your survey Email you for permission or not You will need to find your team and let me know

Projects

Timeline and Evaluation

- Self Introduction Session
- Who are you? Your expertise, such as programming experience, background knowledge of data mining, management, analytics.
- Experience on data analytics in any idea of the project 1 or II if any.

Who am I?



Yanhua Li, PhD Assistant Professor Computer Science & Data Science

PhD, Computer Science, U of Minnesota, 2013 PhD, Electrical Engineering, BUPT, 2009

Research Interests: Big data analytics, Smart Cities, Measurement, Spatio-temporal Data Mining

Industrial Experience: Bell-Labs, Microsoft Research, HUAWEI research Labs

What is DS504/CS586 about?

- A second Level DS/CS course (primarily) for graduates
 - CS/DS Ph.D students in big data analytics and related areas;
 - then other Ph.D students or MS students with
 - Experience in databases and/or in data mining, or equivalent knowledge.
 - Sufficient programming experience is expected so that you are comfortable to undertake a course project.

Course Prerequisite

Great if you have taken some couses on the list.

https://www.wpi.edu/academics/datascience/corecompetency.html

More importantly

- Willing to learn and work hard
- Love to ask questions and solve problems

What is DS504/CS586 about?

We'll learn about

- Advanced Techniques for Big Data Analytics

- Large scale data sampling and estimation,
- Data Cleaning,
- Graph Data Mining,
- Data management, clustering, etc.

- Applications with Big Data Analytics

- Urban Computing
- Social network analysis
- Recommender system, etc.

Learning outcomes

- Explain challenges and advances in the state-of-art in big data analytics.
- Design, develop and fully execute a big data analytics project.
- Communicate their ideas effectively in the form of a presentation and written documents to a technical audience.

Course Topics

- Large scale data sampling and estimation,
- Data Cleaning,
- Data management,
- Graph Data Mining,
- Data clustering,
- Applications with Big Data Analytics, etc

Course Mechanisms

- A seminar- and project-oriented course
- A series of (advanced) topics combining both theory and Practices in two "parallel" tracks:
 - Track 1: Seminar
 - Read, study and discuss research papers on Big Data Analytics.
 - Some presentations by the instructor, and the students.
 - In class discussion! The presenter functions primarily as the lead to facilitate discussion!
 - Track 2: Project
 - group students into "research teams"
 - investigate a selected research topic of interest.

Course Materials

- Textbooks
 - No Textbook.
- Assigned readings with each class:
 - Research papers will be posted on class website (tentatively, updated as we go along)
 - Optional papers for background, supplementary and further readings
- Slides
 - ✤ Will be posted on the class website after each class

Course Requirements

- Do assigned readings
 - Be prepared, read and review required readings on your own in advance!
 - ✤ Do literature survey: find and read related papers if any
 - Bring your questions to the class and look for answers during the class.
- Submit reviews/critiques
 - In myWPI before class
 - Bring 2 hardcopies to the class
 - ✤ Hand in one copy, and keep one copy with you.

Review Writing:

http://users.wpi.edu/~yli15/courses/DS504Spring16/Critiques.html

Attend and participate in class activities

- Please ask and answer questions in (and out of) class!
- Let's try to make the class interactive and fun!

Class Information

- Class Website :
 - http://users.wpi.edu/~yli15/courses/CS4516Fall15B/
- Announcement Page
 - Check the class web page periodically
- Class Mailing List for announcements, Q&As, discussions, etc.
 - cs586-ta@cs.wpi.edu (reaches instructor and TA)
 - cs586-all@cs.wpi.edu (reaches students and instructor)

Office Hours

- Professor Li's Office Hours:
 - ✤ Office: AK130
 - Email: yli15@wpi.edu
 - ✤ M,T, R, F 10:30-11AM
 - Others by appointments

TA



Hi Everyone,

My name is Chong. I'm teaching assistant for DS504. I'm very glad to work and study with you in this semester.

I would like to do my best to help you in my office hour. The office hour will be held on **Friday 2:00~4:00 p.m. AK013 Data innovation lab**. Besides, you can always contact me using email,

czhou2(a)wpi.edu

Thank you very much.

Workload and Grading

- Workload
 - Oral work (30%)
 - Written work (30%) (including a few quizzes)
 - Projects (40%);
 - Project 1: 10%
 - ✤ Project 2: 30%
- Focus more on critical thinking, problem solving, "heads-on/hands-on" experience!
 - Read and critique research papers
 - Understand, formulate and solve problems
 - Two Course Projects

A Few Words on Course Project I

Project I: Collecting and Measuring Online Data

- Team work; each team 2-4 students.
- Starting date: Week 3 (9/8 R)
- Proposal Due: Week 4 (9/17 R) 2 pages roughly
- Due date/time: Before Class on Week 8 (10/13 R) 8 pages rougly
- Requiring Programming in C/C++, Java, Python, and etc
- Choose one online site/service with APIs to download data.
- Examples:
- (1) estimate site statistics, or
- (2) applying machine learning methods to predict future trends, or
- (3) perform time-series analysis to capture dynamic patterns,
- or something else, as long as your work can potentially bring research value to the community.

Course Project II

- Projects will be in groups!
 - ✤ 2-4 students per group, depending on enrollment
- Topics on your choice (related to big data analytics)
 - Application-driven
 - Fundamental data analytics research (heterogeneous data)
 - Data sources on course website <u>http://wpi.edu/~yli15/courses/DS504Spring16/Resources.html</u>

Talk to me once you have an idea.

Course Project II

- Projects will be in groups!
 - ✤ 2-4 students per group, depending on enrollment

- "research-oriented" project timeline: (tentative!)
 - ✤ Group Project
 - Starting date: Week 7 (R):
 - Project Intent due date: Week 8 (R):
 - Project proposal due date: Week 10 (R):
 - Project proposal presentation: Week 11 (R):
 - Project Progress Presentation: Week 13 (R):
 - Project due date: Week 16 (R):
 - Project final Presentation: Week 17 (R):

Class Resources

Presentation

http://users.wpi.edu/~yli15/courses/DS504Spring16/ Presentation.html

Review / Critiques

http://users.wpi.edu/~yli15/courses/DS504Spring16/ Critiques.html

More resources

 http://users.wpi.edu/~yli15/courses/DS504Fall16/ Resources.html

Next Class: Data Acquisition and Measurement

10 Minutes Break