

CSCI446/946 Big Data Analytics

Subject Overview

School of Computing and Information Technology
University of Wollongong Australia

Subject coordinator in CCNU:

Dr. Chao Sun

csun@ccnu.edu.cn

schao@uow.edu.au

Subject Description

- This subject covers the principles, techniques and applications of processing and analysing big data.
- This subject will equip students with the fundamental knowledge on big data analytics and the skills to appropriately choose and apply algorithms to resolve practical analytics problems.

Subject Description

- The subject is organized in three parts:
 - The **first** part introduces fundamental concepts, platforms, systems, and tools for Big Data analysis, to give students a basic understanding of this field.
 - The **second** part introduces main analytics algorithms that are used to uncover the underlying patterns and rules in big data; including classification, regression, clustering, recommendation and retrieval algorithms.
 - The **third** part focuses on case studies and practical applications of big data analytics to help students gain a better understanding of these algorithms.

Subject Learning Outcomes

- On successful completion of this subject, students will be able to
 - Understand basic concepts of big data analytics.
 - Correctly apply the related core algorithm(s).
 - Choose appropriate algorithms to design and build systems to complete a big data analysis task.
 - Interpret and explain results.
 - Integrate the knowledge and skills learned in this subject to resolve practical issues.

Relatedness to other Big Data subjects

- Big Data is a big field requiring algorithm design, implementation, deployment, management, and analytics.
- These aspects are covered the following subjects:
 - INFO411/911 (Data Mining and Knowledge Discovery)
 - Focuses of DM techniques and DM analytics.
 - ISIT312/912 (Big Data Management)
 - Focus on distributed BD storage, retrieval, and processing.
 - CSCI316 (Big Data Mining Techniques and Implementation)
 - Focus on BD algorithms and their implementation.
 - CSCI433/933 (Machine Learning Algorithms and Applications)
 - Focus on ML principles and theory.
 - CSCI446/946 (Big Data Analytics)
 - Focus on BD deployment and analytics.

Relatedness to other Big Data subjects

- Because these subjects are intertwined there is some intended overlap.
 - Many methods used in Data Mining can be adopted to Big Data and vice versa though the scope of application differs.
 - Machine Learning provides the tools of choice for Data Mining and for Big Data.
 - A subject with focus on implementation and a subject with focus on application could cover the same methods.
 - There are transitional elements that link these subjects.
 - No prerequisites on each other means that they can be studied independently and in any order. But this requires that each must cover the same or similar basics.

Attendance Requirements

- It is the responsibility of students to attend all lectures and labs.
- The amount of time spent on a 6 credit point subject should be at least 12 hours per week.
- Satisfactory attendance is deemed by the University, to be attendance at approximately 80% of the allocated contact hours.
- If you are present for less than 80% and would have otherwise passed you need to apply for student academic consideration, otherwise a TF (**technical fail**) grade will be recorded.

Attendance Requirements

- Attendance requirement can be met by attending live lectures/labs or by viewing the recorded lectures.
 - Active participation is expected. Level of participation may be recorded.
 - Labs may be recorded but recorded labs will not be made available.
- Consultation is available to students who attended relevant lectures
 - The lecturer does not offer private tuition!

Lecture Schedule

- Week 1: Introduction to Big Data Analytics
- Week 2: Big Data systems and platforms
- Week 3: Big Data Analytical Theory and Methods (I)
- Week 4: Big Data Analytical Theory and Methods (II)
- Week 5: Big Data Analytical Theory and Methods (III)
- Week 6: Big Data Visualisation

Lecture Schedule

- Week 7: Large-scale Text Processing: Theories
- Week 8: Large-scale Text Processing: Applications
- Week 9: Large-scale Image Processing: Theories
- Week 10: Large-scale Image Processing: Applications
- Week 11: Big Data and Social Media Analytics
- Week 12: Future Challenges of Big Data Analytics
- Week 13: Revision

Textbooks and References

There is no prescribed textbook.

Some topics of this subject can be found in the following books.

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, John Wiley & Sons, 27 Jan. 2015
2. Data Mining: The Textbook by Charu C. Aggarwal, Springer 2015
3. C.M. Christopher, P. Raghavan and H. Schutze, Introduction to Information Retrieval, Cambridge University Press, 2008
4. Computer Vision: A Modern Approach (2nd Edition), by David A. Forsyth and Jean Ponce, Pearson, 2011.

More references and reading materials will be recommended in the class.

Assessment

- Programming assignments x 3
 - Assignment 1 (10 marks)
 - Assignments 2 and 3 (40 marks in total)
 - Due date*: Weeks 5, 9, and 12, respectively
- Lab, quizzes, or exercises
 - 10 in total, 1 mark each (10 marks)
- Final exam (40 marks)

* Exact due date and time will be published on Moodle. Due date is subject to change. Changes to due dates will be published on the subject's Moodle site.

Submission of Assessment

- Unless notified by the subject coordinator, all written assignments must be submitted electronically via a corresponding submission link on Moodle and **before** the due date.
 - Submission of assessment items via email will not be accepted.
 - Details to the submission method will be given with the assignment instruction.

Submission of Assessment

- Penalties apply to all late work, except if academic consideration has been granted. Late submissions will attract a penalty of 25% per day (or part of it) late of the assessment mark. This amount is per calendar day thus includes weekends and public holidays.
- Work more than 3 days late will not be assessed and be awarded a mark of zero.
- Marked assignments will normally be returned within 2 weeks of the due date.

Technical Fail and Plagiarism

- To be eligible for a Pass in this subject a student must achieve a mark of **at least 50% in the final exam**. Students who fail to achieve this minimum mark & would have otherwise passed may be given a TF (Technical Fail) for this subject.
- **Plagiarism** will result in zero marks for the affected assessment item, the student will be reported to the school or faculty, and may be given a FAIL grade for this subject.
 - This applies to all assessed items: Assignments, labs, quizzes, and exam.

Important Message

- Students should check the subject's Moodle site regularly as important information, including details of changes in assessment requirements will be posted from time to time via Moodle.
- Any information posted to the Moodle site is deemed to have been notified all students.

Important Message

- Always maintain good communication with the subject lecturer and let me know
 - Your concerns and how I can help you
 - The positive aspects of this subject
 - The issues of this subject to be addressed
- Emergency procedures:
 - Please study the emergency procedures posted on the subject's Moodle site.

