

Big Data Analytics

Rationale

In addition to constantly growing volumes of proprietary transaction, product, inventory, customer, competitor, and industry data collected from enterprise systems, organizations are also faced with overwhelming amounts data from the Web, social media, mobile sources, and sensor networks that do not fit into traditional databases in terms of volume, velocity and variety (the three Vs of Big Data). This Big Data flood poses challenges as well as opportunities, if managed and analyzed properly, to derive new actionable knowledge and intelligence in a timely manner. This course will explore existing and emerging methods to manage, integrate, analyze and visualize domain-specific Big Data, to identify and provide domain specific solutions.

Description

This course covers the research issues and practical methods of managing and analyzing Big Data to gain and discover insights, patterns, and knowledge nuggets that can support decision makers.

Topic List

Topics may include but are not limited to:

- Introduction
 - Environment, Challenges, and Opportunities
 - Analytics Platform: architecture, process, and analytic tools
 - Multiple data source management and data integration
- Structured Data Analytics
 - Structured Big Data: Issues and Approaches
 - Transportation Data Analytics
 - Financial, Banking, Web-based Transaction Data Analytics

- Semi/Unstructured Data Analytics
 - Textual Data Analytics
 - Social Media Data Analytics
 - Short Text Classification/Clustering
 - Real-time Big Data Processing
- Media Data Analytics
 - Fundamentals of Image/Video Data Analytics
 - Cultural Analytics and Visualization
 - Statistical Inference and Real-time Classification
- Network and Graph Data Analytics
 - Social Network/Graph Data Analytics
 - Semantic Web and Linked Data Analytics
- Societal Impacts on Big Data Analytics
 - Security and Privance Issues
 - Accountability Issues: Open Government Data

Learning Goals

- To expose students to Big Data as a scientific or engineering problem. Students will be guided to focus on a particular domain specific area, identify research challenges or application utilities, and present existing and/or innovative methods and algorithms to design a solution. Students are expected to submit a conference paper and/or a demonstration paper to a conference related to Big Data Analytics by the end of this seminar, in collaboration with the faculty member(s). A series of student presentations are expected at the end of the semester.

Assessment

Each student will present a critical review, summarizing the problems and solutions given in a selected research paper. Students will select one application domain area and collect a repository of data sets. Collectively, the data sets will serve a wider community for Big Data Analytics experiments and tests.

Students will identify a Big Data Analytics research problem related to their domain application and dataset, and write a research paper discussing the existing solutions and design/propose a potential new applied solution that can be used by the domain area decision makers.

With the given dataset, each student can analyze and design a specific use-case related to her/his research problem, and design (and possibly implement) his/her proposed solution as a tool.

The final presentation of the research paper and a demo will be given in the form of a workshop/poster presentation at the end of the semester with an audience of invited faculty, students and industry leaders. Top paper awards will be given and students will have the chance to work with a faculty or industry mentor on a conference paper and/or journal publication.