2015 Spring CS591 Seminar Presentation Schedule

CS591-Han Seminar: Advanced Topics on Data Mining (Presentation Schedule: Spring 2015) 4-5pm Thursdays at 3403 Siebel Center

- Major theme: Advanced Data Mining, with emphasis on (1) information network mining; (2) construction of heterogeneous information networks by data mining
- Papers can be selected from this year or previous several year's conference proceedings or journals. We welcome students who would like to present tutorials and/or writing topic survey articles related to this theme.
- Two students per unit (20 minutes presentation and 5 minutes discussion for each research paper, i.e., two papers will be covered per class unit). We can book one whole slot for you if you would like to present tutorials or surveys and need more time.
- Please book your time slot early. Also, please upload your papers and slides to be presented at least one day before its presentation!
- Please write down the presentation title, paper, venue info to help others to see what you will be presenting!
- Please link to the papers and slides at least 8 hours before the presentation so others can read them beforehand!
- First meet: Jan. 29, 2015 at 3403 SC!

Week 2 (1/29/15): Class organization and one presentation

Honglei Zhuang: ICDM 2014 Report (slides)

Doris Xin: Active Learning for Networked Data Based on Non-progressive Diffusion Model, WSDM '14 (paper) (slides)

Week 3 (2/5/15): Two presentations

Jingbo Shang: A parallel and efficient algorithm for learning to match, ICDM’14 (paper) (slides)

Hongkun Yu: FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering, WSDM '15 (paper) (slides)

Week 4 (2/12/15): One presentation

Speaker: Juan Liu
Title: Anomaly Detection and the Information Triage Problem

Abstract:
Anomaly Detection is a challenging machine learning/data mining problem. It is ill-posed in nature because anomaly (deviation from an expected norm) lacks a clear definition and may take diverse forms. Yet the ability to detect anomalies is often critical in practical applications. In this talk I will give two real-world anomaly detection scenarios: (1) the DARPA/ADAMS (Anomaly Detection at Multiple Scales) program, detecting malicious insider threats within organizations where information security need to be ensured, and (2) industrial research of detecting fraudulent claims in medical insurance programs. In these two scenarios, the intended users -- analysts (intelligence analysts or fraud investigators) face an information triage problem: too much data, and yet limited resources for investigation. We will explain the richness of real-world anomalies and how an ensemble of techniques, including social network analysis, natural language processing, user modeling, and spatial/temporal analysis are integrated to facilitate the information triage problem.

Bio:
Juan Liu got her PhD from University of Illinois at Urbana-Champaign in 2001. After graduation she joined Palo Alto Research Center (PARC) as a Research Scientist. She had also taken a research manager role and founded/managed the Learning, Inference, and Data Solution (LIDS) Area at PARC. She is currently a Lead Research Scientist at Medallia, developing a data science theme to understand and manage consumer experience. Her research interests include machine learning, data mining, statistical modeling, information theory, and applications to real-world scenarios. She has led the effort on a number of data analytics projects. She was the co-Principal Investigator of DARPA Anomaly Detection at Multiple Scales (ADAMS) program, leading the team effort (4yrs, $8.4M) developing data mining schemes to detect malicious insider threats. She holds over 30 USPTO patents and 75 publications.
Week 5 (2/19/15): Two presentations
Yucheng Chen: Distance Metric Learning Using Dropout: A Structured Regularization Approach (KDD 2014) Slides
Honglei Zhuang: Leveraging In-Batch Annotation Bias for Crowdsourced Active Learning (WSDM 2015) Slides

Week 6 (2/26/15): Two presentations
Chao Zhang: Mining Topics in Documents: Standing on the Shoulders of Big Data (KDD 2014) Slides

Week 7 (3/5/15): Two presentations
Huan Gui Negative Link Prediction in Social Media slides
Ahmed El-Kishky: Reducing the Sampling Complexity of Topic Modeling

Week 9 (3/12/15): Two presentations
Jingjing Wang Joint Topic Modeling for Event Summarization across News and Social Media Streams slides
Shi Zhi Automatic Gloss Finding for a Knowledge Base using Ontological Constraints (WSDM 2015) slides


Week 11 (3/19/15): Two presentations
Hao Luo Analyzing Expert Behaviors in Collaborative Networks (KDD’14) slides
Brandon Norick Ranking-based Clustering on General Heterogeneous Information Networks by Network Projection slides

Week 12 (4/2/15): Two presentations
George Brova “Effective Named Entity Recognition for Idiosyncratic Web Collections” paper [slides]
Xiang Ren Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers paper

Week 13 (4/9/15): Two presentations
Honglei Zhuang: EDBT 2015 Report (slides)
Jingbo Shang: Diffusion Component Analysis: Unraveling Functional Topology in Biological Networks. (slides)
Fangbo Tao: CatchSync: Catching Synchronized Behavior in Large Directed Graphs (slides)

Week 14 (4/16/15): Two presentations
Min Li: A Discriminative Latent Variable Model for Online Clustering. ( paper slides)
Wei Zhang: Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis (slides)

Week 15 (4/23/15): Two presentations
Haoyan Cai: SiGMa: Simple Greedy Matching for Aligning Large Knowledge Bases(slides)
Yi Zhou Knowledge Base Completion via Search-Based Question Answering ( paper slides)

Week 16 (4/30/15): Group Summary Meeting
will move to Tue 5/5 3-5pm 3405SC
Title: Building Blocks of Data Science
Abstract: In this talk I will try and flesh out the core components of the emerging discipline of Data Science using several examples from my group's research. Data Science is becoming a melting pot for a whole host of techniques from computer science, statistics, information theory but the underlying question is simple: how to extract useful information from multi-modal secondary data where the number of degrees of freedom is typically a small fraction of the dimensionality of the raw data.
Bio: Sanjay Chawla is a Professor at the School of IT, University of Sydney. He is currently on extended leave at the Qatar Computing Research Institute (QCRI). His main area of research is data mining and machine learning. More specifically he has been working on three problems of contemporary interest: outlier detection, imbalanced classification and adversarial learning. He served as the Department Chair from 2008-2011 and was an academic visitor in Yahoo! Labs in Bangalore India in 2012. He received his PhD from the University of Tennessee in 1995.