2014 Fall CS591 Seminar Presentation Schedule

CS591-Han Seminar: Advanced Topics on Data Mining (Presentation Schedule: Fall 2014) 4-5pm Thursdays at 0216 Siebel Center (Note: Classroom Change)

- Major theme: Advanced Data Mining, with emphasis on (1) information network mining; (2) data mining in cyber-physical systems
- 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!

Week 1 (8/29/14): Class organization and one presentation
  - Chao-Yuan Wu (CMU): Jointly Modeling Aspects, Ratings and Sentiments for Movie Recommendation (JMRS): based on KDD’14 paper [slides]

Week 2 (9/5/14): One conference report + one presentation
  - Ahmed Elk-Kishky and Xiang Ren: KDD’14 conference report [slides]
  - Fangbo Tao: Enterprise Taxonomy Construction [slides] (Summer Intern Work)

Week 3 (9/11/14): Two presentations
  - Honglei Zhuang: Clustering by fast search and find of density peaks [slides]
  - Jialu Liu: Large Scale High-Precision Topic Modeling on Twitter [slides]

Week 4 (9/18/14): Two presentations
  - Chao Zhang: PRESS: A Novel Framework of Trajectory Compression in Road Networks [slides]
  - Jingjing Wang: Insight for news link slides

Week 5 (9/25/14): Two presentations
  - Xiang Ren: Synonym Discovery for Structured Entities on Heterogeneous Graphs
  - Ahmed El-Kishky: Topic-Link LDA: Joint Models of Topic and Author Community. Slides

Week 6 (10/2/14): Two presentations
  - Jingbo Shang: Unsupervised Query Segmentation Using Clickthrough for Information Retrieval [slides]
  - Yu Shi: Typed Tensor Decomposition of Knowledge Bases for Relation Extraction [slides]

Week 7 (10/9/14): One presentation
  - Qi He (LinkedIn) slides
  - Title: Personalized Feed Recommendation in LinkedIn
**Abstract:**

Users on LinkedIn generate a large number of heterogeneous activities, ranging from connecting with other users, to sharing content, to updating their profiles. The set of activities within a user's network neighborhood forms a stream of updates (feed) for the user's consumption. In this talk, I discuss the personalized feed recommendation problem in the LinkedIn homepage feed. In particular, I describe our modeling efforts on generating activities' relevance scores with two kinds of personalization granularity: user segments and individual users. At the user segment level, users are assigned a mixture of social-driven segment and content-driven segment, and each user segment has its unique relevance model. At the individual user level, each pair of viewer and actor in a specific activity type has a personalized affinity score, named eAffinity. More than 4 billions of eAffinity scores are computed in LinkedIn everyday. Experimental results from online bucket tests are reported.

**Short Bio:**

Qi He is a staff applied researcher at consumer relevance science in LinkedIn since 2013. Prior to LinkedIn, he was a research staff member at IBM Almaden Research Center since 2010. He finished PhD at Nanyang Technological University in 2008, followed by a 2-year postdoc in CiteSeer at Pennsylvania State University. He is interested in developing effective and efficient data analysis techniques for state-of-the-art data intensive applications like recommender systems, heterogeneous information networks, social networks and event detection etc.

**Week 8 (10/16/14): Two presentations**

- Shi Zhi: From Data Fusion to Knowledge Fusion (VLDB, 2014) [paper] [slides]
- Haoyan Cai: Dynamics of News Events and Social Media Reaction (KDD, 2014) [slides]

**Week 9 (10/23/14): Two presentations**

- Huan Gui: Modeling Topic Diffusion in Multi-Relational Bibliographic Information Networks [paper] [slides]
- Yanglei Song: Tensor Decompositions: Exploiting Structure in Observed Correlations (AAAI tutorial part 1, 2014) [slides]

**Week 10 (10/30/14): Two presentations**

- Min Li: Sumblr: Continuous Summarization of Evolving Tweet Streams (SIGIR, 2013) [paper] [slides]
- Chenguang Wang: Programming with Personalized PageRank: A Locally Groundable First-Order Probabilistic Logic (CIKM, 2013) [paper] [slides]

**Week 11 (11/6/14): One presentation**

- Yuanyuan Zhou (UCSD)
  
  **Title:** Applying learning/mining techniques in computer systems and beyond
  
  **Abstract:** As computer systems become ever-so complex to manage and optimize, various machine learning or data mining techniques have become popular in analyzing a large amount of system data. In this talk, I will share my limited experience and mistakes we have made when applying these techniques to solve system problems in our research projects and also commercial products. Additionally, we will also share some of the challenges in extracting and analyzing people-related information in our recent project.

  **Bio:** Yuanyuan Zhou is currently a Qualcomm Chair Professor at UC-San Diego. She is an ACM Fellow and obtained her Ph.D from Princeton. Before UCSD, she was a tenured associate professor at University of Illinois at Urbana Champaign. Her research interests span the areas of operating systems, software engineering, system reliability and maintainability. She has co-founded three startups. Her recent startup, PatternInsight, has successfully deployed software quality assurance tools in many companies and in 2012 its Log Insight business line was acquired by VmWare and now Log Insight is a VmWare Product offering to its many data center customers for data center management. Recently she has launched her third startup, Whova, that helps professional networking intelligently offline at events and meetings. She has the great fortune of working with many talented students and post-docs, many of whom are currently professors at top universities such as Univ. of Wisconsin, Madison, and risk-taking startup founders.

**Week 12 (11/13/14):**

- George Brova: Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion (KDD 2014) [paper] [slides]
- Hongkun Yu: The Inverse Regression Topic Model (ICML, 2014) [paper] [supplement] [slides]

**Week 13 (11/20/14):**

- Doris Xin: Building High-level Features Using Large Scale Unsupervised Learning [paper] [slides]
- Brandon Norick: Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network [paper] [slides]

**Week 14 (11/27/14): Thanksgiving break, no presentations**
Week 15 (12/4/14): Group Summary Meeting