

Tutorials

This year's meeting will start with two tutorials given by leaders in their fields.

Subjectivity and Sentiment Analysis

Jan Wiebe (University of Pittsburgh)

A growing area of research, “subjectivity analysis” is the computational study of affect, opinions, and sentiments expressed in text. Blogs, editorials, reviews (of products, movies, books, and so on), and even “objective” newspaper articles (which include many opinions and sentiments) are just some of the genres for which accurate identification and interpretation of opinions is critical for full text understanding. Subjectivity analysis will support developing tools for information analysts in governmental, commercial, and political domains who want to automatically track attitudes and feelings in the news and on-line forums. How do people feel about the latest iPod? Is there a change in the support for the new Medicare bill? A system able to automatically identify and extract opinions and sentiments from text would be an enormous help to someone sifting through the vast amounts of news and web data, trying to answer these kinds of questions.

This tutorial will focus on methods for subjectivity analysis of text. It will focus on fine-grained analysis at the sentence, phrase, word, and word-sense levels. Specifically, it will cover problem definition (for example, what is subjectivity? what is sentiment?) and manual annotations; methods for learning subjective language and identifying opinion-bearing words and phrases; methods for identifying polarity and orientation (positive, negative, or neutral); and methods for identifying subjective word senses, that is, dictionary definitions.

Jan Wiebe is a professor of computer science and director of the Intelligent Systems Program at the University of Pittsburgh. Her research with students and colleagues has been in discourse processing, pragmatics, word-sense disambiguation, and probabilistic classification in NLP. Her most recent work investigates automatically recognizing and interpreting expressions of opinions and sentiments in text, to support NLP applications such as question answering, information extraction, text categorization, and summarization. Her current and past professional roles include program committee chair of NAACL, executive board member of NAACL, editorial board member of Computational Linguistics and Language Resources and Evaluation, AAAI workshop cochair, vice chair of the ACM Special Interest Group on Artificial Intelligence (SIGART), and chair of the ACM-SIGART/AAAI Doctoral Consortium.

Graph Mining Techniques for Social Media Analysis

Mary McGlohon and Christos Faloutsos (Carnegie Mellon University School of Computer Science)

How do structures in social networks form and appear? How does information propagate through these networks? What tools can we use to analyze network structures? This tutorial will give a concise overview of the important conceptual and software tools for characterizing, modeling, and ana-

lyzing graph structures common in weblogs and social media, with pointers to active research areas, latest publications, and available software. We review fundamental tools and surprising patterns that recur in social networks, including applications toward ranking influence of blogs, identifying communities and expertise, predicting links, and analyzing trends.

We first focus on patterns, as we review the “six-degrees of separation,” the ideas behind power-law graphs, and more recent patterns on time-evolving graphs, like the “densification” power law. Next, we turn our attention to tools, as we cover singular- and eigen-value decomposition, which are the engines behind HITS and PageRank; a brief overview of tensors for time-evolving graphs; MDL (minimum description language) for community detection; and results on virus-propagation and epidemic thresholds. The emphasis is on the intuition behind all these tools and their practical impact for the analysis of large, real datasets from social media. Finally, we report on recent discoveries on the timing and shape of blog cascades and influence propagation.

Mary McGlohon is a Ph.D. student in the Machine Learning Department at Carnegie Mellon University. She has received a National Science Foundation Graduate Research Fellowship (2005). Prior to beginning her graduate work, she received B.S. degrees in computer science and mathematics from the University of Tulsa in Tulsa, Oklahoma. Her dissertation research focuses on graph mining, particularly with respect to properties of evolving graphs, information diffusion in networks, and link analysis.

Christos Faloutsos is a professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, eleven best paper awards, and several teaching awards. He has served as a member of the executive committee of SIGKDD; he has published over 160 refereed articles, 11 book chapters and one monograph. He holds five patents and he has given over 20 tutorials and 10 invited distinguished lectures. His research interests include data mining for streams and graphs, fractals, database performance, and indexing for multimedia and bioinformatics data.