Visual Analytics and Information Fusion

Abstract
Data in many real-world applications may arise from multiple sources, and can be viewed from different aspects. It is a significant analytical challenge to extract information from data and help people make right decisions in a timely fashion. To address this challenge, we believe visual analytics and information fusion are two important and interrelated scientific problems. The primary objective of the workshop is to bring together researchers in Computer Science, Mathematics, Statistics, and Visualization, who work on related problems in information fusion and visual analytics. The workshop should provide an opportunity to discuss recent developments and applications of visual analytics and information fusion.

Technical Scope
Visual analytics focuses on developing interactive and intuitive visualization methods that help users conduct data analysis effectively. Visual analytics provides a human-in-the-loop type of data mining paradigm, which goes beyond traditional automatic black-box approaches.

Information fusion deals with integrating data from multiple sources into a consistent and consolidated view, which subsequently can be used for analysis and decision-making. Many real-world applications have multiple sources or aspects on data, which need to be integrated and then analyzed. Fully automatic information fusion is often impossible because of various data challenges such as noise, inconsistency, ambiguity, and nonstationarity. To address this challenge, practical applications often require continuous data fusion and visual analytics, such as the following:

Healthcare informatics: Increasingly, patient data is being stored in electronic medical records (EMRs), which arise from different sources such as diagnosis, procedure, medication, lab results etc. Physicians need to make the decision quickly based on EMRs by integrating all the information about a patient.

Bioinformatics: Recent technological innovations have allowed us to collect massive amounts of biological data including gene expression data, protein-protein interaction data, amino acid sequences, etc. The challenge is how to effectively integrate various types of data with different representations for improved biological inference (e.g., protein function prediction).

Social media mining: Social network and social media continuously generate large amount of user data from various aspects. For example, users’ activities in a friendship network are encoded in Facebook, professional activities in emails and LinkedIn, users’ music and movie interest in Pandora and Netflix. It can be extremely challenging to link all those data and to visually analyze and present this information in a meaningful manner.
Spatio-temporal data mining: multiple aspects of sequential information may be captured over time at any given location, such as traffic patterns, video streams, temperature and sound, and other environmental sensors. How to link this data, to find insights from the data, and to visualize the insight in an intuitive fashion?

Planned activities
We intend to have four sessions in the one-day workshop. The first two sessions will include three invited talks, each half an hour long. The third session will be panel discussion. The panel discussion will focus on challenges and future directions. The panelists will be recruited from experts in the field, including the invited speakers. One or more of the organizers will moderate the panel. Questions will be taken from the audience and the moderator and panelists will be free to contribute additional questions. The last session will be a poster presentation session. All posters will be based on contributed submissions. The organizers will review the submissions and select high quality submissions for poster presentation. We plan to encourage participation by graduate students and junior researchers.

Expected outcome
We expect to inspire more active collaboration and research in the area of visual analytics and information fusion for solving real-world challenges in the targeting domains. We encourage students to work on this promising and emerging area for their PhD. Finally we plan to summarize the workshop through a white paper to publish in a top data mining and/or visualization journal or magazine.

Organizers
David Gotz, IBM TJ Watson Research Center, dgotz@us.ibm.com
Bio: David Gotz is a Research Staff Member at the IBM T.J. Watson Research Center where he conducts research in the areas of visual analytics and intelligent user interfaces. He is a member of a newly formed group at IBM Research focusing on innovative healthcare transformation technologies that combine data mining, statistical analysis and modeling, and visualization. David received his Ph.D. in Computer Science from the University of North Carolina at Chapel Hill. He has authored numerous peer-reviewed articles in the areas of visual analytics, information visualization, user interfaces, and multimedia. David has served on various organizing and program committees including IEEE VisWeek, ACM Multimedia, IEEE EMBC, and ISVC. He has also served as a reviewer for several leading conferences and journals. In particular, David has been active in the Visual Analytics community for several years both as a paper author and by serving on both the Program Committee and Organizing Committee for IEEE VAST.

Sridhar Mahadevan, University of Massachusetts, Amherst, mahadeva@cs.umasss.edu
Bio: Sridhar Mahadevan is a professor in the Department of Computer Science at the University of Massachusetts, Amherst. He received his PhD from Rutgers University under the direction of Professor Thomas Mitchell (now at Carnegie Mellon University). He is a recipient of many awards and honors, including the NSF
CAREER Award, the Michigan State Withrow Distinguished Scholar Award and the Teacher Scholar award, as well as best paper awards or nominations at several international conferences in artificial intelligence and machine learning. He co-directs the Autonomous Learning Laboratory at UMass Amherst, and is also on the editorial board of the Journal of Machine Learning Research (MIT Press). He is the author of three books and over a hundred articles in artificial intelligence and machine learning.

**Haesun Park, Georgia Tech, hpark@cc.gatech.edu**

Bio: Prof. Haesun Park received her B.S. degree in Mathematics from Seoul National University, Seoul Korea, in 1981 with summa cum laude and the University President's Medal for the top graduate, and her M.S. and Ph.D. degrees in Computer Science from Cornell University, Ithaca, NY, in 1985 and 1987, respectively. She was on the faculty of the Department of Computer Science and Engineering, University of Minnesota, Twin Cities, from 1987 to 2005. From 2003 to 2005, she served as a program director for the Computing and Communication Foundations Division at the National Science Foundation, Arlington, VA, U.S.A.

Since July 2005, she has been a professor in the School of Computational Science and Engineering at the Georgia Institute of Technology, Atlanta, Georgia where she is currently the associate chair. Her research interests include numerical algorithms, data analysis, visual analytics, bioinformatics, and parallel computing. She has published extensively in refereed journals and conferences in these areas. She is the director of the NSF/DHS FODAVA-Lead (Foundations of Data and Visual Analytics) project where the goal is to create mathematical and computational foundations for data and visual analytics. Prof. Park has served on numerous editorial boards including IEEE Transactions on Pattern Analysis and Machine Intelligence and SIAM Journal on Matrix Analysis and Applications, and has served as a conference co-chair for SIAM International Conference for Data Mining in 2008 and 2009.

**Jimeng Sun, IBM TJ Watson Research Center, jimeng@us.ibm.com**

Bio: Jimeng Sun is a Research Staff Member at IBM TJ Watson lab. He received the MS and PhD degree in Computer Science from Carnegie Mellon University in 2006 and 2007. His research interests include data mining for health care applications, medical informatics, social network analysis, visual analytics, and data streams. He has received the best research paper award in ICDM 2008, the KDD 2007 dissertation award (runner-up), the best research paper award in SDM 2007. He has published over 40 refereed articles and two book chapters. He filed eight patents and has given four tutorials. He has served as the program committee member of SIGKDD, ICDM, SDM and CIKM and a reviewer for AMIA, TKDE, VLDB, and ICDE. He has co-chaired the workshops on large-scale data mining: theory and applications in KDD’10 and ICDM’09, the workshop on large-scale Analytics for Complex Instrumented Systems on ICDM’10, and the workshop on Visual Analytics in Health
Care in VisWeek’10. He also co-edited the journal special issue on large-scale data mining at TKDD.

Jieping Ye, Arizona State University, jieping.Ye@asu.edu
Bio: Jieping Ye is an Associate Professor of the Department of Computer Science and Engineering at Arizona State University. He received his Ph.D. in Computer Science from University of Minnesota, Twin Cities in 2005. His research interests include machine learning, data mining, and biomedical informatics. He won the outstanding student paper award at ICML in 2004, the SCI Young Investigator of the Year Award at ASU in 2007, the SCI Researcher of the Year Award at ASU in 2009, the NSF CAREER Award in 2010, and the KDD best research paper award honorable mention in 2010.

Schedule

Duration: half day
Invited speakers

Talk 1
Speaker: Chris Ding
Title: Robust Nonnegative Matrix Factorization using L1, L21 Norms
Abstract
Nonnegative matrix factorization (NMF) related models are now widely used in text mining, bioinformatics, knowledge transfer, recommender systems, semi-supervised and unsupervised learning. However so far, the basic models use least square formulation which is prone to large noise and outliers. After reviewing major NMF models, here we present robust NMF models using L1 and L21 norms which exhibit stability and robustness w.r.t. large noises. We present computational algorithms for these models with rigorous theoretical analysis. These algorithm are as efficient as the algorithms for least square formulations, avoiding the significant computational complexities routinely associated with L1, L21 formulations. Experiments in image data demonstrate the strong robustness of the robust NMF models.
Bio: Chris Ding obtained Ph.D. from Columbia University, did research at California Institute of Technology, Jet Propulsion Laboratory, and Lawrence Berkeley National Laboratory, before joining University of Texas at Arlington as a professor in 2007. His research areas are data mining, bioinformatics, high performance computing, focusing on matrix/ tensor approaches. He served on NIPS, ICML, KDD, IJCAI, AAAI, ICDM, SDM conference committees, and reviewed research grants for National Science Foundations of U.S., Israel, Ireland, and Hong Kong. He has given invited seminars at Berkeley, Stanford, Carnegie Mellon, University of Waterloo, University of Alberta, Google Research, IBM Research, Microsoft Research. He published 180 research papers with 6180 citations.

Talk 2
Speaker: Daniel Keim
Title: Quality Metrics for Visual Analytics of High-Dimensional Data

Abstract
Visual Analytics of high-dimensional data is particularly challenging. In my talk I will present and discuss quality metrics which have been proposed to help in the visual exploration of patterns in high-dimensional data. In a number of recent approaches, quality metrics have been used to automate the demanding search through large spaces of alternative visualizations, allowing the user to concentrate on the most promising visualizations as suggested by the quality metrics. I will present a set of factors for discriminating the quality metrics, visualization techniques, and the process itself. The process can be described through a reworked version of the well-known information visualization pipeline. The usefulness of the model will be shown by applying it to several visual analytics approaches that are based on quality metrics.

Bio: Daniel Keim is full professor and head of the Information Visualization and Data Analysis Research Group at the University of Konstanz, Germany. He has been actively involved in information visualization and data analysis research for more than 20 years and developed a number of novel information visualization and visual analysis techniques for very large datasets. Dr. Keim got his Ph.D. and habilitation degrees in computer science from the University of Munich. Before joining the University of Konstanz, Dr. Keim was associate professor at the University of Halle, Germany and Technology Consultant at AT&T Shannon Research Labs, NJ, USA.

Talk 3
Speaker: Chang Wang
Title: Relation Extraction with Relation Topics
Abstract
Detecting semantic relations in text is very useful in both information retrieval and question answering because it enables knowledge bases to be leveraged to score passages and retrieve candidate answers. In this talk, I will present a novel transfer learning approach to the semantic relation detection problem.

Instead of relying only on the training instances for a new relation, we leverage the knowledge learned from previously trained relation detectors. Specifically, we detect a new semantic relation by projecting the new relation's training instances onto a lower dimension topic space constructed from existing relation detectors through a three step process. First, we construct a large relation repository of more than 7,000 relations from Wikipedia. Second, we construct a set of non-redundant relation topics defined at multiple scales using diffusion wavelets from the relation repository to characterize the existing relations. Similar to the topics defined over words, each relation topic is an interpretable multinomial distribution over the existing relations. Third, we integrate the relation topics in a kernel function, and use it together with SVM to construct detectors for new relations.
The experimental results on Wikipedia and ACE data have confirmed that background-knowledge-based topics generated from the Wikipedia relation repository can significantly improve the performance over the state-of-the-art relation detection approaches.

Bio: Chang Wang is a research scientist at IBM Research. He is currently working on DeepQA (Watson) project. The DeepQA project is to push question answering technology to levels of performance previously unseen and demonstrate the technology by playing Jeopardy! at the level of a human champion.

Chang’s research areas include Machine Learning (Manifold Learning, Representation Learning, Multiscale Analysis); Knowledge Transfer Across Domains; and Application of Machine Learning in Natural Language Processing (NLP) and Information Retrieval.

Talk 4
Speaker: Pak Chung Wong
Title: Visual Analytics and Knowledge Fusion in Multidisciplinary Scientific Discovery
Abstract
We discuss how we use visual analytics as a collaborative means to bring multidisciplinary expertise together and fuse cross-cutting knowledge into one scientific discovery environment. The discussion draws on our R&D experience in applying visual analytics to model and predict the impact of global climate change on the U.S. power grids and its wider implications on society and its critical infrastructure at the Pacific Northwest National Laboratory. The presentation focuses on the technical challenges and lessons learned from the collaboration that involves scientists from physical, social, and computation sciences.

Bio
Pak Chung Wong is a project manager and chief scientist in information analytics at the Pacific Northwest National Laboratory (PNNL). His research background is in visualization and visual analytics with focuses on extreme scale data analytics, graph analytics, multimedia analytics, and national security analytics. Currently, he serves on the IEEE Computer Graphics and Applications (CG&A) editorial board. In 2012, he will co-chair the IEEE VisWeek Conference in Seattle, WA and the SPIE Visual Data Analysis (VDA) Conference in San Francisco, CA. In the past, he has served as a conference chair, program chair, and papers chair at visualization conferences from IEEE Visualization (Vis), IEEE Information Visualization Conference (InfoVis), and IEEE Symposium on Visual Analytics Science and Technology (VAST). He received a PhD in computer science from the University of New Hampshire in 1997.
Talk 5
Speaker: Dr. David S. Ebert
Title: Predictive Interactive Visual Analytics
Abstract:
While less than ten year old, visual analytics has become crucial to enabling insight
and decision making in a wide range of fields. In this talk, I’ll characterize the
essentials of integrated, predictive, interactive advanced analytical and visualization
environments and show the potential of visual analytics to dramatically transform
scientific discovery, engineering development, public health, emergency response,
and safety. I’ll also describe our recent visual analytics work in applications ranging
from cancer biomarker discovery to public health and emergency response and
show the impact and potential of the field. Finally, I’ll conclude by discussing the
challenges and opportunities for visual analytics.
Bio:
David Ebert is the Silicon Valley Professor of Electrical and Computer Engineering at
Purdue University, a University Faculty Scholar, a Fellow of the IEEE, and Director of
the Purdue University Visualization and Analytics Center (PURVAC), which leads the
Visualization Science team of the Department of Homeland Security’s Command
Control and Interoperability Center of Excellence (VACCINE). Dr. Ebert performs
research in novel visualization techniques, visual analytics, volume rendering,
information visualization, perceptually-based visualization, illustrative
visualization, mobile graphics and visualization, and procedural abstraction of
complex, massive data. Ebert has been very active in the visualization community,
teaching courses, presenting papers, co-chairing many conference program
committees, serving on the ACM SIGGRAPH Executive Committee, serving as Editor
in Chief of IEEE Transactions on Visualization and Computer Graphics, serving as a
member of the IEEE Computer Society’s Publications Board, serving on the IEEE
Computer Society Board of Governors, and successfully managing a large program of
external funding to develop more effective methods for visually communicating
information.