Knowledge Discovery from Sensor Data (Sensor-KDD)

Ranga Raju Vatsavai *
Oak Ridge National Laboratory, TN, USA
vatsavairr@ornl.gov

Olufemi A. Omitaomu
Oak Ridge National Laboratory, TN, USA
omitaomuoa@ornl.gov

Joao Gama
University of Porto, Portugal
jgama@fep.up.pt

Nitesh V. Chawla
University of Notre Dame, IN, USA
nchawla@cse.nd.edu

Mohamed Medhat Gaber
Monash University, Australia
mohamed.m.gaber@gmail.com

Auroop R. Ganguly
Oak Ridge National Laboratory, TN, USA
gangulyar@ornl.gov

ABSTRACT

Extracting knowledge and emerging patterns from sensor data is a nontrivial task. The challenges for the *knowledge discovery* community are expected to be immense. On one hand, dynamic data streams or events require real-time analysis methodologies and systems, while on the other hand centralized processing through high end computing is also required for generating offline predictive insights, which in turn can facilitate real-time analysis. In addition, emerging societal problems require knowledge discovery solutions that are designed to investigate anomalies, changes, extremes and nonlinear processes, and departures from the normal. Keeping in view the requirements of the emerging field of knowledge discovery from sensor data, we took initiative to develop a community of researchers with common interests and scientific goals, which culminated into the organization of Sensor-KDD series of workshops in conjunction with the prestigious ACM SIGKDD International Conference of Knowledge Discovery and Data Mining. In this report, we summarize the events of the Second ACM-SIGKDD International Workshop on Knowledge Discovery form Sensor Data (Sensor-KDD 2008).

1. INTRODUCTION

Wide-area sensor infrastructures, remote sensors, and wireless sensor networks, RFIDs, yield massive volumes of disparate, dynamic, and geographically distributed data. As such sensors are becoming ubiquitous, a set of broad requirements is beginning to emerge across high-priority applications including disaster preparedness and management, adaptability to climate change, national or homeland security, and the management of critical infrastructures. The raw data from sensors need to be efficiently managed and transformed to usable information through data fusion, which in turn must be converted to predictive insights via knowledge discovery, ultimately facilitating automated or human-induced tactical decisions or strategic policy based on decision sciences and decision support systems.

The challenges for the *knowledge discovery* community are expected to be immense. On one hand, dynamic data streams or events require real-time analysis methodologies and systems, while on the other hand centralized processing through high end computing is also required for generating offline predictive insights, which in turn can facilitate real-time analysis. The online and real-time knowledge discovery imply immediate opportunities as well as intriguing short- and long-term challenges for practitioners and researchers in knowledge discovery. The opportunities would be to develop new data mining approaches and adapt traditional and emerging knowledge discovery methodologies to the requirements of the emerging problems. In addition, emerging societal problems require knowledge discovery solutions that are designed to investigate anomalies, changes, extremes and nonlinear processes, and departures from the normal. The Sensor-KDD workshop seeks to bring together researchers from academia, government and the industry working in various aspects of knowledge discovery from sensor data.

*Corresponding author.

1.1 Motivation

The expected ubiquity of sensors in the near future, combined with the critical roles they are expected to play in high priority application solutions, point to an era of unprecedented growth and opportunities. The online knowledge discovery requirements described earlier imply immediate opportunities as well as intriguing short- and long-term challenges for practitioners and researchers in knowledge discovery. In addition, the knowledge discovery and data mining (KDD) community would be called upon, again and again, as partners with domain experts to solve critical application solutions in business and government, as well as in the domain sciences and engineering. The main motivation for the Sensor-KDD series of workshops stems from the increasing need for a forum to exchange ideas and recent research results, and to facilitate collaboration and dialog between academia, government, and industrial stakeholders. This is clearly reflected in the successful organization of the first workshop [3] along with KDD-2007, which was attended by more than **seventy** registered participants. The high quality of submissions allowed to us to put together an edited book [2] and a special issue in the 'Intelligent Data Analysis' journal [1].

2. SUMMARY OF THE WORKSHOP

Based on the positive feedback from the previous workshop attendees and our own experiences and interactions with the government agencies such as DHS, DOD, and involvement with numerous projects on knowledge discovery from sensor data, we organized the 2^{nd} Sensor-KDD workshop along with the KDD-2008 conference. As expected we received very high quality paper submissions which were thoroughly reviewed by a panel of international program committee members [4]. Based on a minimum of two reviews per paper, we have selected *seven full* papers and *six short* papers. In addition to the oral presentations of accepted papers, the workshop featured two invited speakers Dr. Kendra E. Moore, Program Manager, DARPA/IPTO and Prof. Jiawei Han, Department of Computer Science, University of Illinois at Urbana-Champaign. We now briefly summarize each of these presentations, for full details about each of the presented papers, please refer to the workshop proceedings [5].

2.1 Session 1

The first session was moderated by Dr. Auroop Ganguly. This session featured our first invited speaker, Dr. Kendra Moore, followed by two paper presentations. Dr. Moore presented the challenges of knowledge discovery from sensor data in defense applications. She touched upon wide variety of topics including heterogeneity, distributed sensors, real-time requirements, privacy, and applications of national importance.

The paper, "Anomaly Detection from Sensor Data for Real-time Decisions" by Olufemi A. Omitaomu, Yi Fang, and Auroop R. Ganguly, is presented by Olufemi. This paper is concerned about the detection of unusual profiles or anomalous behavioral characteristics from multiple types of sensor data. The authors presented a two-stage knowledge discovery process, where offline approaches are utilized to design online solutions that can support real-time decisions. They illustrated this innovative solution in the context detecting anomalous behavior of trucks using sensor-based measurements collected at truck weigh stations. This is a fine example of knowledge discovery application in the context of national security.

The paper, "Network Service Disruption upon Natural Disaster: Inference Using Sensory Measurements and Human Inputs" by Supaporn Erjongmanee and Chuanyi Ji, shows important role of data mining and machine learning in inferring large-scale network service disruption. Natural disasters, like recent Hurricane Katrina, can cause large-scale network service interruptions which leads to unreachability of networks. The authors presented a joint use of large-scale sensory measurements from Internet and a small number of human inputs for effective network inference through a clustering and semi-supervised learning algorithm. This approach is evaluated on network service disruption induced by Hurricane Katrina at subnet level. The results showed that clustering reduces the spatial dimensionality by 81%, and the subnet statuses inferred by the semi-supervised classifier showed interesting facts of network resilience.

2.2 Session 2

The second session was moderated by Dr. Olufemi Omitaomu. This session featured five paper presentations, covering variety of topics.

The paper, "Spatio-Temporal Outlier Detection in Precipitation Data" by Elizabeth Wu, Wei Liu, and Sanjay Chawla, was presented by Elizabeth. This paper address one of the core data mining techniques, outlier detection, from large volumes of spatio-temporal data. Current data mining techniques have several limitations in handling spatio-temporal data, therefore it is very important to develop new techniques or extend existing techniques to handle spatio-temporal data. The authors presented a spatio-temporal outlier detection algorithm called Outstretch, which discovers the outlier movement patterns of the top-k spatial outliers over several time periods. The top-k spatial outliers are found using the Exact-Grid Top-k and Approx-Grid Top-k algorithms, which are an extension of algorithms developed by Agarwal et al. [2]. These algorithms use Kulldorff spatial scan statistic, which is designed to discover all outliers, unaffected by the neighbouring regions that may contain missing values. After generating the outlier sequences, the authors shows how these sequences can be interpreted, by comparing them to the phases of the El Nino Southern Oscilliation (ENSO) weather phenomenon.

The paper, "Probabilistic Analysis of a Large-Scale Urban Traffic Sensor Data Set" by Jon Hutchins, Alexander Ihler, and Padhraic Smyth, was presented by Jon. It is very important to detect underlying patterns in large volumes of spatiotemporal data as it allows, for example, human behavior modeling, traffic planning, etc. However, real-world sensor time series are often significantly noisy and more difficult to work with than the relatively clean data sets that tend to be used as the basis for experiments in many research papers. In contrast, the authors report on a large case-study involving statistical data mining of over 100 million measurements from 1700 freeway traffic sensors over a period of seven months in Southern California. They discussed the challenges posed by the wide variety of different sensor failures and anomalies present in the data. The volume and complexity of the data precludes the use of manual visualization or simple thresholding techniques to identify these anomalies. The authors describe the application of probabilistic modeling and unsupervised learning techniques to this data set and illustrate how these approaches can successfully detect underlying systematic patterns even in the presence of substantial noise and missing data.

The paper, "WiFi Miner: An Online Apriori-Infrequent Based Wireless Intrusion Detection System" by Ahmedur Rahman, C.I. Ezeife, and A.K. Aggarwal, deals with intrusion detection in wireless networks. Their system, WiFi Miner, is capable of finding frequent and infrequent patterns from preprocessed wireless connection records using infrequent pattern finding Apriori algorithm. This online Apriori-Infrequent algorithm improves the join and prune step of the traditional Apriori algorithm with a rule that avoids joining itemsets not likely to produce frequent itemsets. An anomaly score is then assigned to each packet (record) based on whether the record has more frequent or infrequent patterns. Connection records with positive anomaly scores have more infrequent patterns than frequent patterns and are considered as anomalous packets. The authors described a solution that eliminates the need for hard-to-obtain training data in wireless network environments, and increases intrusion detection rate and reduces false alarms.

The paper, "Mobile Visualization for Sensory Data Stream Mining" by Pari Delir Haghighi, Brett Gillick, Shonali Krishnaswamy, Mohamed Medhat Gaber, and Arkady Zaslavsky, introduces an integrated architecture of situation aware adaptive data mining and mobile visualization techniques for ubiquitous computing environments. With the emergence of ubiquitous data mining and recent advances in mobile communications, there is a need for visualization techniques to enhance the user-interactions, realtime decision making and comprehension of the results of mining algorithms. To address this important problem, the authors proposed a novel architecture for situation-aware adaptive visualization that applies intelligent visualization techniques to data stream mining of sensory data. Their architecture incorporates fuzzy logic principles for modeling and reasoning about context/situations and performs gradual adaptation of data mining and visualization parameters according to the occurring situations. A prototype of the architecture is implemented using J2ME and tested in the area of health-care monitoring.

The paper, "Dense Pixel Visualization for Mobile Sensor Data Mining" by Pedro Pereira Rodrigues and Joo Gama, describes dense pixel visualization techniques for visualizing sensor data and as well as absolute errors resulting from predictive models. Sensor data is usually represented by streaming time series. Current state-of-the-art systems for visualization include line plots and three-dimensional representations, which most of the time require screen resolutions that are not available in small transient mobile devices. Moreover, when data presents cyclic behaviors, such as in the electricity domain, predictive models may tend to give higher errors in certain recurrent points of time, but the human-eye is not trained to notice these cycles in a long stream. To overcome some of these limitations, the authors proposed a simple dense pixel display visualization system, exploiting the benefits that it may represent on detecting and correcting recurrent faulty predictions. A case study is also presented, where a simple corrective strategy is studied in the context of global electrical load demand, exemplifying the utility of the new visualization method when compared with automatic detection of recurrent errors.

2.3 Session 3

This session which was moderated by Dr. Joao Gama, featured our second invited speaker, Prof. Jiawei Han, and two paper presentations. Prof. Jiawei's talk on "Data Mining in Sensor Network Systems: Trouble-Shooting and Shooting-Trouble" captures dual application of data mining and touches recent advances made in the filed. Abstract of Prof. Jaiawei Han's talk is given below.

"*Data mining will play an essential role in the development of robust sensor network systems. The talk will discuss our recent work in two research frontiers: (1) how to develop effective data mining methods for troubleshooting in the development of robust sensor network systems; and (2) how to develop new data mining methods for shooting troubles (i.e., anomalies) in data streams, which should be an essential function in sensor network systems.*

Handling the first task leads to a tool, called DustMiner, for uncovering bugs due to interactive complexity in networked sensing applications. Such bugs are not localized to one component that is faulty, but rather result from complex and unexpected interactions between multiple often individually non-faulty components. Moreover, the manifestations of these bugs are often not repeatable, making them particularly hard to find, as the particular sequence of events that invokes the bug may not be easy to reconstruct. Because of the distributed nature of failure scenarios, our tool looks for sequences of events that may be responsible for faulty behavior, as opposed to localized bugs such as a bad pointer in a module. An extensible framework is developed where a front-end collects runtime data logs of the system being debugged and an offline back-end uses frequent discriminative pattern mining to uncover likely causes of failure. The tool helped uncover event sequences that lead to a highly degraded mode of operation. Fixing the problem significantly improved the performance of the protocol.

For shooting troubles (i.e., anomalies) in sensor network systems, we propose a new approach to build predictive models for rare events in sensor data streams. The method estimates reliable posterior probabilities using an ensemble of models to match the distribution over under-samples of negatives and repeated samples of positives (i.e., anomalies). We formally show some interesting and important properties of the proposed framework, e.g., reliability of estimated probabilities on skewed positive class, accuracy of estimated probabilities, efficiency and scalability. Experiments are performed on several synthetic as well as real-world datasets with skewed distributions, and they demonstrate that our framework has substantial advantages over existing approaches in estimation reliability and predication accuracy."

The paper, "Monitoring Incremental Histogram Distribution for Change Detection in Data Streams" by Raquel Sebastio, Joo Gama, Pedro Pereira Rodrigues, and Joo Bernardes, addresses important problem of detecting changes in constructing histograms from time-changing high-speed data streams. Histograms are a common technique for density estimation and they have been widely used as a tool in exploratory data analysis. Learning histograms from static and stationary data is a well known topic. In this paper authors present algorithms to detect changes from high-speed time-changing data streams. The authors studied strategies to detect changes in the distribution generating examples, and adapt the histogram to the most recent data by forgetting outdated data. They used the Partition Incremental Discretization algorithm for this task. The authors compared the distributions using Kullback-Leibler divergence, defining a threshold for change detection decision based on the asymmetry of this measure and evaluated their algorithm on controlled artificial data.

2.4 Session 4

The final session was moderated by Dr. Nitesh Chawla. This session featured four paper presentations.

The paper, "Unsupervised Plan Detection with Factor Graphs" by George B. Davis, Jamie Olson, and Kathleen M. Carley, describes synchronous and asynchronous expectation maximization algorithms for unsupervised learning in factor graphs. Recognizing plans of moving agents is a natural goal for many sensor systems, with applications including robotic path finding, traffic control, and detection of anomalous behavior. However, plan recognition gets complicated in the absence of contextual information such as labeled plans and relevant locations. Authors introduced two unsupervised methods to simultaneously estimate model parameters and hidden values within a Factor graph representing agent transitions over time. These algorithms were evaluated by applying them on a GPS tracking dataset consisting of 1074 ships over 5 days in the English channel. Initial results indicated that a reasonable model may be inferred on this difficult problem.

The paper, "An Adaptive Sensor Mining Framework for Pervasive Computing Applications" by Parisa Rashidi and Diane J. Cook, presents adaptive data mining framework for detecting patterns in sensor data. Mining sequences of sensor events poses unique challenges to the KDD community especially when the underlying data source is dynamic and the patterns change. In this papers the authors have introduced an adaptive data mining framework that detects patterns in sensor data, and more importantly, adapts to the changes in the underlying model. The frequent and periodic patterns of data are first discovered by the Frequent and Periodic Pattern Miner (FPPM) algorithm; and then any changes in the discovered patterns over the lifetime of the system are discovered by the Pattern Adaptation Miner (PAM) algorithm, in order to adapt to the changing environment. This framework also captures vital context information present in pervasive computing applications, such as the startup triggers and temporal information. This data mining framework is evaluated using the data collected in the CASAS smart home testbed.

The paper, "Spatiotemporal Neighborhood Discovery for Sensor Data" by Michael P. McGuire, Vandana P. Janeja, and Aryya Gangopadhyay, describes a framework for the discovery of spatiotemporal neighborhoods in sensor datasets where a time series of data is collected at many spatial locations. The purpose of the spatiotemporal neighborhoods is to provide regions in the data where knowledge discovery tasks such as outlier detection, can be focused. As building blocks for the spatiotemporal neighborhoods, authors have developed a method to generate spatial neighborhoods and a method to discretize temporal intervals. These methods were tested on real life datasets including (a) sea surface temperature data from the Tropical Atmospheric Ocean Project (TAO) array in the Equatorial Pacific Ocean and (b) highway sensor network data archive and initial results were encouraging.

The paper, "Exploiting Spatial and Data Correlations for Approximate Data Collection in Wireless Sensor Networks" by Chih-Chieh, Hung Wen-Chih, Peng York, Shang-Hua Tsai and Wang-Chien Lee, describes algorithms for finding representative sensor nodes. Finding sensor nodes with similar readings is an important task as it allows data reduction. However, efficiently identifying the sensor groups and their representative nodes is a very challenging task. Authors proposed an algorithm, namely DCglobal, to determine a set of representative nodes that have high energy levels and wide data coverage ranges, where a data coverage range of a sensor node is the set of sensor nodes whose reading vectors are very close to the sensor node. Furthermore, they propose a maintenance mechanism to dynamically select alternative representative nodes when the representative nodes have less energy or representative nodes can no longer capture spatial correlation within their data coverage ranges. Experimental studies on both synthetic and real datasets, showed that DCglobal was able to effectively and efficiently provide approximate data collection while prolonging the network lifetime.

Closing remarks of the workshop were provided by Dr. Raju Vatsavai. He summarized the days proceedings by thanking invited speakers, sponsors, program committee members, and participants.

3. CONCLUSIONS

Extracting knowledge and emerging patterns from sensor data is a nontrivial task. The challenges for the *knowledge discovery* community are expected to be immense. As evidenced from the participation and quality of submissions to the first and second Sensor-KDD workshops, it is clear that the '*Knowledge Discovery from Sensor Data or Sensor-KDD*' is clearly a growing area and an important specialty (sub-area) within the knowledge discovery. The Sensor-KDD workshop is proven to be an attractive forum for the researchers from academia, industry and government, to exchange ideas, initiate collaborations and lay foundation to the future of this important and growing area. The workshop witnessed lively participation from all quarters, generated interesting discussions immediately after each presentation and as well as at the end of the workshop. All participants agreed for continued patronage for the Sensor-KDD workshop. In addition to the ACM workshop proceedings, extended papers will be published as post workshop proceedings in Springer's well-known 'Lecture Notes in Computer Science' series.

4. ACKNOWLEDGMENTS

We would like to thank the authors of all submitted papers and presenters. Their innovation and creativity has resulted in a strong technical program. We are highly indebted to the program committee members, whose reviewing efforts ensured in selecting a competitive and strong technical program. The program committee included: Michaela Black, Andre Carvalho, Sanjay Chawla, Francisco Ferrer Ray Hickey, Ralf Klinkenberg, Miroslav Kubat, Mark Last, Chang-Tien Lu, Elaine Parros Machado de Sousa, Sameep Mehta, Laurent Mignet, S. Muthu Muthukrishnan, Pedro Rodrigues, Josep Roure, Bernhard Seeger, Cyrus Shahabi, Mallikarjun Shankar, Alexandre Sorokine, Eiko Yoneki, Philip S. Yu, Nithya Vijayakumar, and Guangzhi Qu. We would like to thank our invited speakers, Dr. Kendra E. Moore, Program Manager, DARPA/IPTO and Prof. Jiawei Han, Department of Computer Science, University of Illinois at Urbana-Champaign, who despite their busy schedules, readily agreed and delivered highly motivating and informative talks. We would like to thank, Dr. Brian Worley, Director, Computational Sciences and Engineering Division (CSED), Oak Ridge National Laboratory (ORNL), for his encouragement, support, and continued patronage of Sensor-KDD workshop series, and Dr. Budhendra Bhaduri, Group leader of Geographic Information Science and Technology, CSED, ORNL, for his enthusiastic support and best paper award sponsorship. We would like to thank the SensorNet program (url: http://www.sensornet.gov) managed by the Computational Sciences and Engineering Division at the Oak Ridge National Laboratory and other collaborators.

5. WORKSHOP ORGANIZERS

Dr. Ranga Raju Vatsavai has been conducting research in the area of spatiotemporal databases and data mining for the past 15 years. Before joining the Oak Ridge National Laboratory (ORNL) as a Research Scientist, he worked at IBM-Research (2004-06; IIT-Delhi campus), U of Minnesota (1999-2004; Twin-cities campus, MN), AT&T Labs (1998; Middletown, NJ), Center for Development of Advanced Computing (1995-98; C-DAC, U of Pune campus, India), and National Forest Data Management Center (1990-95; FRI Campus, Dehradun, India). He has published over thirty peer-reviewed articles and served on program committees of several international conferences (KDD, ICTAI, SSTDM). He was also involved in the design and development of several highly successful software systems (UMN-MapServer - a world leading open source WebGIS, *Miner - a spatiotemporal data mining workbench, EASI/PACE classification modules, and first parallel softcopy photogrammetry system for IRS-1C/1D satellites). His broad research interests are centered on spatial, spatiotemporal databases and data mining, and computational geoinformatics; in particular he is interested in statistical pattern recognition, semi-supervised learning, multiple classifier systems, time series analysis and forecasting, information retrieval, uncertainty and error handling.

Dr. Olufemi A. Omitaomu is a research associate in the Computational Sciences and Engineering Division at the Oak Ridge National Laboratory. His research interests include data mining and knowledge discovery from distributed sensor data, infrastructure modeling and interdependencies, machine learning, and uncertainty analysis. He received Ph.D. in information engineering from the University of Tennessee. He has published in top peer-reviewed journals and conferences; co-organized and co-chaired workshop and sessions at professional conferences including the ACM Workshop on Knowledge Discovery from Sensor Data held in conjunction with ACM SIGKDD 2007 and ACM SIGKDD 2008. He previously worked as a data analyst with Mobil Exploration and Production Company for more than five years.

Dr. Joao Gama is a researcher at LIAAD-INESC Porto LA, the Laboratory of Artificial Intelligence and Decision Support of the University of Porto. His main research interest is Learning from Data Streams. He has published several articles in change detection, learning decision trees from data streams, hierarchical clustering from streams, etc. Editor of special issues on Data Streams in Intelligent Data Analysis, J. Universal Computer Science, and New Generation Computing. Co-chair of ECML 2005 Porto, Portugal 2005, Conference chair of Discovery Science 2009, and of a series of Workshops on Knowledge Discovery in Data Streams, ECML 2004, Pisa, Italy, ECML 2005, Porto, Portugal, ICML 2006, Pittsburg, US, ECML 2006 Berlin, Germany, SAC2007, Korea, and the ACM Workshop on Knowledge Discovery from Sensor Data held in conjunction with ACM SIGKDD 2007 and ACM SIGKDD 2008. Together with M. Gaber edited the book Learning from Data Streams-Processing Techniques in Sensor Networks, published by Springer.

Dr. Nitesh V. Chawla is an assistant professor at the University of Notre Dame. Dr. Chawla's research interests are broadly in the areas of data mining, machine learning, pattern recognition, and their applications. More specifically his research has focused on learning from massive datasets, distributed data mining/machine learning, ensemble techniques, cost/distribution sensitive learning, feature selection, and semi-supervised learning. His research has also focused on the inter-disciplinary applications such as intelligent scientific visualization, biometrics, bioinformatics, natural language processing, and customer analytics.

Dr. Mohamed Medhat Gaber is a research Fellow at Monash University, Australia. He has published more than 60 papers. Mohamed is the co-editor of the book: Learning from Data Streams: Processing Techniques in Sensor Networks, published by Springer in 2007 and the book: Knowledge Discovery from Sensor Data by CRC that is due to appear by end of 2008. His research interests include data stream mining, wireless sensor networks and context-aware computing. Mohamed has served in the program committees of several international and local conferences and workshops in the area of data mining and context-aware computing. He was the co-chair of the IEEE International Workshop on Mining Evolving and Streaming Data held in conjunction with ICDM 2006, International Workshop on Knowledge Discovery from Ubiquitous Data Streams held in conjunction with ECML/PKDD 2007, and the First and Second International Workshop on Knowledge Discovery from Sensor Data held in conjunction with ACM SIGKDD 2007/2008.

Dr. Auroop R. Ganguly is a research scientist within the Computational Sciences and Engineering division of the Oak Ridge National Laboratory since 2004. His research interests are climate change impacts, geoscience informatics, civil and environmental engineering, computational data sciences, and knowledge discovery. Prior to ORNL, he has more than five years of experience in the software industry, specifically Oracle Corporation and a best-of-breed company subsequently acquired by Oracle, and about a year in academia, specifically at the University of South Florida in Tampa. He has a PhD from the Civil and Environmental Engineering department of the Massachusetts Institute of Technology, several years of research experience with a group at the MIT Sloan School of Management, experience in private consulting, and a wide range of peer-reviewed publications spanning multiple disciplines. Currently, he is also an adjunct professor at the University of Tennessee in Knoxville.

6. INVITED SPEAKERS

Dr. Kendra Moore's research interests include automatic pattern learning and change detection in complex spatio-temporal data streams. This spans learning data representations, and activity and movement models, and adapting to changes as they occur. Dr. Moore is also interested in developing technology to understand, support, and assess peer production-based information fusion systems. Dr. Moore currently manages the Predictive Analysis for Naval Deployment Activities (PANDA) program. She also managed the Fast Connectivity for Coalition Agents Program (Fast C2AP) program, which transitioned to the US Navy's GCCS-M program in October 2007. Dr. Moore joined DARPA in 2005. Prior to joining DARPA, Dr. Moore was president and founder of Advestan, Inc., where she provided R&D consulting services to DoD customers and contractors in advanced estimation, analysis, and exploitation for large-scale information fusion applications. Before starting Advestan, Dr. Moore was the Director of Information Fusion at ALPHATECH, Inc. (now BAE Systems). She also served on the Problem-Centered Intelligence, Surveillance, and Reconnaissance (PCISR) study panel to develop recommendations for new all-source ISR architectures for a national intelligence agency. Prior to that, she developed, extended, and applied large-scale systems analysis techniques to a wide range of military command and control systems.

Dr. Jiawei Han is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. His research expertise include data mining, data warehousing, database systems, data mining from spatiotemporal data, multimedia data, stream and RFID data, social network data, and biological data. He has written over 350 journal and conference publications. He has chaired or served in over 100 program committees of international conferences and workshops, including PC cochair of 2005 (IEEE) International Conference on Data Mining (ICDM), Americas Coordinator of 2006 International Conference on Very Large Data Bases (VLDB). He is also serving as the founding Editor-In-Chief of ACM Transactions on Knowledge Discovery from Data. He is an ACM Fellow and has received 2004 ACM SIGKDD Innovations Award and 2005 IEEE Computer Society Technical Achievement Award. His book "Data Mining: Concepts and Techniques" 2nd ed., Morgan Kaufmann, 2006) has been popularly used as a textbook worldwide.

7. REFERENCES

[1] A. Ganguly, J. Gama, O. Omitaomu, M. Gaber, and R. R. Vatsavai, editors. *Intelligent Data Analysis, An International Journal*. IOS Press, Nieuwe Hemweg 6B, 1013 BG Amsterdam, The Netherlands, 2008.

[2] A. Ganguly, J. Gama, O. Omitaomu, M. Gaber, and R. R. Vatsavai, editors. *Knowledge Discovery from Sensor Data*. CRC Press, Boca Raton, FL 33487, 2008. http://www.crcpress.com/shopping_cart/products/product_detail.asp?sku=82329&isbn=9781420082326&parent_id=&pc=.

[3] Sensor-KDD 2007. First international workshop on knowledge discovery from sensor data (Sensor-KDD 2007). http://www.ornl.gov/sci/knowledgediscovery/KDD-2007-Workshop/.

[4] Sensor-KDD Program Committee. Second international workshop on knowledge discovery from sensor data (Sensor-KDD 2008). http://www.ornl.gov/sci/knowledgediscovery/SensorKDD-2008/organizers.htm.

[5] R. R. Vatsavai, O. Omitaomu, J. Gama, M. Gaber, and A. Ganguly. Second international workshop on knowledge discovery from sensor data (Sensor-KDD 2008). http://www.ornl.gov/sci/knowledgediscovery/SensorKDD-2008/index.htm, 2008. ACM.