Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (SensorKDD’10)

Held in conjunction with

July 25, 2010
Washington, DC

Workshop Chairs

Dr. Olufemi A. Omitaomu, Oak Ridge National Laboratory, TN, USA

Dr. Varun Chandola, Oak Ridge National Laboratory, TN, USA

Dr. Auroop R. Ganguly, Oak Ridge National Laboratory, TN, USA

Prof. Joao Gama, University of Porto, Portugal

Dr. Ranga Raju Vatsavai, Oak Ridge National Laboratory, TN, USA

Prof. Mohamed Medhat Gaber, University of Portsmouth, UK

Prof. Nitesh V. Chawla, University of Notre Dame, IN, USA
Fourth International Workshop on Knowledge Discovery from Sensor Data (SensorKDD’10)

Description
Wide-area sensor infrastructures, remote sensors, RFIDs, phasor measurements, and wireless sensor networks 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 adaptability to national or homeland security, critical infrastructures monitoring, disaster preparedness and management, greenhouse emissions and climate change, and transportation. 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 the one hand are dynamic data streams or events that require real-time analysis methodologies and systems, while on the other hand are static data that require high end computing 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, rare events, hotspots, changes, extremes and nonlinear processes, and departures from the normal.

According to the data mining and domain experts present at the NSF-sponsored Next Generation Data Mining Summit (NGDM ’09) held in October 2009, “finding the next generation of solutions to these challenges is critical to sustain our world and civilization” [1]. Some of the organizers of the SensorKDD workshop are part of the summit. The 4th International Workshop on Knowledge Discovery from Sensor Data (SensorKDD-2010) is the first step in bringing researchers together to address these challenges and moving toward the development of the next generation data mining solutions require to address these challenges.

Therefore, the SensorKDD-2010 seeks to bring together researchers from academia, government, and the industry working in the following areas and applications:

1. Offline Knowledge Discovery
   a. Predictive analysis from geographically distributed and heterogeneous data
   b. Computationally efficient approaches for mining unusual patterns, specifically, anomalies, extremes, nonlinear processes and change, from massive and disparate space-time data

2. Online Knowledge Discovery
   a. Real-time analysis of dynamic and distributed data, including streaming and event-based data
   b. Mining from continuous streams of time-changing data and mining from ubiquitous data
   c. Efficient algorithms to detect deviations from the normal in real-time
   d. Resource-aware algorithms for distributed mining

3. Decision and Policy Aids
a. Coordinated offline discovery and online analysis with feedback loops
b. Combination of knowledge discovery and decision scientific processes
c. Facilitation of faster and reliable tactical decisions as well as prudent and insightful longer term policies

4. Theory
   a. Distributed data stream models
   b. Theoretical frameworks for distributed stream mining

5. Case Studies
   a. Success stories in national or global priority applications
   b. Real-world problem design and knowledge discovery requirements

Motivation
The motivation for SensorKDD (http://www.ornl.gov/sci/knowledgediscovery/SensorKDD-Workshop/) in conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining 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. The expected ubiquity of sensors in the 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 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 first workshop was organized in 2007. Based on the positive feedback from the previous workshop attendees and our own experiences and interactions with the government agencies such as the United States Department of Homeland Security, United States Department of Defense, and involvement with numerous projects on knowledge discovery from sensor data, we strongly believe in the continuation of this workshop. We believe that the ACM SIGKDD conference is the right forum to organize this workshop as it brings the KDD community together in this important area to establish a much needed leadership position in research and practice in the near term, as well as in the long term.

Success of Previous SensorKDD Workshops
The previous three workshops – SensorKDD-2007 [2], SensorKDD-2008 [3], and SensorKDD-2009 [4] – held in conjunction with the 13th, 14th, and 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining respectively attracted several participants as well as many high quality papers and presentations. The 2007 workshop was attended by more than seventy registered participants. The workshop program included presentations by authors of six accepted full papers and four invited speakers. The invited speakers were Prof. Pedro Domingos of the University of Washington, Prof. Joydeep Ghosh of the University of Texas, Austin, Prof. Hillol Kargupta of the University of Maryland, Baltimore County, and Dr. Brian Worley of the Oak Ridge National Laboratory (ORNL). There were also poster presentations by authors of six accepted short papers. The extended versions of papers presented at the workshop were developed into a book [5], the first book published in this specific discipline. The four top accepted papers were awarded certificates and cash prize of $500 each. The prize money was donated by the Computational Sciences and Engineering Division (CSED) of the Oak Ridge National Laboratory and Information Society Technology Project KDUBiq-WG3 of the European Union. The top papers were also published in a special issue of the Journal of Intelligent Data Analysis. This workshop was partially sponsored by the Geographic Information Science and Technology Group of CSED at ORNL.

The SensorKDD-2008 workshop was attended by more than 60 registered participants. There were presentations by authors of seven accepted full papers and six accepted short papers; the workshop
program also included presentations by two invited speakers – Prof. Jiawei Han of the University of Illinois at Urbana-Champaign and Dr. Kendra Moore of the Defense Advanced Research Projects Agency. The extended versions of papers presented at the 2008 workshop were recently published as Springer's LNCS post-proceedings in 2009 [6]. The two top accepted papers were awarded certificates and cash prize of $500 each; the prizes were donated by the Computational Sciences and Engineering Division of the Oak Ridge National Laboratory. This workshop was partially sponsored by the Geographic Information Science and Technology Group of CSED at ORNL.

The 2009 workshop was attended by several registered participants. There were presentations by authors of eight accepted full papers, eight accepted short papers, two entries for the SensorKDD-2009 cup, and three invited speakers. The invited speakers were Prof. Carlos Guestrin of Carnegie Mellon University, Dr. Aurelie Lozano of IBM T.J. Watson Research Center, and Mr. Alessandro Donati of the European Space Agency. The extended versions of papers presented at the 2009 workshop are scheduled for publication as Springer's LNCS post-proceedings in 2010. The best paper, two best student papers, and two SensorKDD-2009 cup entries were awarded certificates and cash prizes. The prizes were donated by the Computational Sciences and Engineering Division of the Oak Ridge National Laboratory and Cooperating Objects Network of Excellence of the European Union. The workshop was partially sponsored by the Geographic Information Science and Technology Group of CSED at ORNL.

Workshop Sponsors
The SensorKDD-2010 workshop is sponsored by the Geographic Information Science and Technology (GIST) Group at Oak Ridge National Laboratory and the Computational Sciences and Engineering (CSE) Division at the Oak Ridge National Laboratory.

Appreciation
We would like to thank our sponsors for their kind donations. In addition, we thank the SIGKDD’10 organizers, the authors of the submitted papers, the invited speakers, and the members of the Program Committee for their respective and collective efforts to make this workshop possible.

The workshop proceedings were compiled by Dr. Olufemi A. Omitaomu of the Computational Sciences and Engineering Division at Oak Ridge National Laboratory. The workshop proceedings have been co-authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher by accepting the article for publication, acknowledges that the United States Government retains, a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References


Dr. Olufemi A. Omitaomu (omitaomuo@ornl.gov; +1-865-241-4310)
Dr. Varun Chandola (chandolav@ornl.gov; +1-865-576-6192)
Dr. Auroop R. Ganguly (gangulyar@ornl.gov; +1-865-241-1305)
Dr. Ranga Raju Vatsavai (vatsavairr@ornl.gov; +1-865-576-3569)
Oak Ridge National Laboratory, 1 Bethel Valley Rd, MS-6017, Oak Ridge, TN, USA.

Prof. Joao Gama (jgama@fep.up.pt; +351-22-339-2094)
LIAAD-INESC Porto LA, University of Porto, Rua de Ceuta, Porto, Portugal.

Prof. Mohamed Medhat Gaber (mohamed.m.gaber@gmail.com; +44 23 92 84 6416)
University of Portsmouth, School of Computing, Hampshire, UK

Prof. Nitesh V. Chawla (nchawla@cse.nd.edu; +1-574-631-8716)
University of Notre Dame, Dept. of Computer Science & Engineering, IN, USA.
Organizers

Workshop Chairs:
1. Olufemi A. Omitaomu, Oak Ridge National Laboratory, TN, USA.
2. Varun Chandola, Oak Ridge National Laboratory, TN, USA.
3. Auroop R. Ganguly, Oak Ridge National Laboratory, TN, USA.
4. Joao Gama, University of Porto, Portugal.
5. Ranga Raju Vatsavai, Oak Ridge National Laboratory, TN, USA.
6. Nitesh V. Chawla, University of Notre Dame, IN, USA.
7. Mohamed Medhat Gaber, Monash University, Australia.

Program Committee (In alphabetical order of last name):
1. Adedeji B. Badiru, Air Force Institute of Technology, Dayton, OH, USA.
2. Budhendra L. Bhaduri, Oak Ridge National Laboratory, TN, USA.
4. Albert Bifet, University Polytechnica, Catalunya, Spain.
5. Michaela Black, University of Ulster, Coleraine, Northern Ireland, UK.
6. Jose del Campo-Avila, Universidad de Malaga, Spain.
7. Andre Carvalho, University of Sao Paulo, Brazil.
8. Sanjay Chawla, University of Sydney, Australia.
9. Diane Cook, Washington State University, Pullman, WA, USA.
10. Alfredo Cuzzocrea, University of Calabria, Italy.
11. Jing (David) Dai, IBM Watson Research Center, USA.
12. Christie Ezeife, University of Windsor, Canada.
13. David J. Erickson III, Oak Ridge National Laboratory, TN, USA.
14. Yi Fang, Purdue University, West Lafayette, IN, USA.
15. Francisco Ferrer, University of Seville, Spain.
16. James H. Garrett, Carnegie Mellon University, Pittsburgh, PA, USA.
17. Joydeep Ghosh, University of Texas, Austin, TX, USA.
18. Bryan L. Gorman, Oak Ridge National Laboratory, TN, USA.
19. Sara Graves, University of Alabama, Huntsville, AL, USA.
20. Ray Hickey, University of Ulster, Coleraine, Northern Ireland, UK.
21. Forrest Hoffman, Oak Ridge National Laboratory, TN, USA.
22. Luke (Jun) Huan, University of Kansas, Lawrence, KS, USA.
23. Volkan Isler, University of Minnesota, Minneapolis, MN, USA.
24. Vandana Janeja, University of Maryland, Baltimore County, MD, USA.
25. Yu (Cathy) Jiao, Oak Ridge National Laboratory, TN, USA.
26. Ralf Klinkenberg, University of Dortmund, Germany.
27. Miroslav Kubat, University Miami, FL, USA.
28. Vipin Kumar, University of Minnesota, Minneapolis, MN, USA.
29. Mark Last, Ben-Gurion University, Israel.
30. Chang-Tien Lu, Virginia Tech., VA, USA.
31. Elaine Parros Machado de Sousa, University of Sao Paulo, Brazil.
32. Sameep Mehta, IBM Research, India.
33. Laurent Mignet, IBM Research, India.
34. S. Muthu Muthukrishnan, Rutgers University and AT&T Research, NJ, USA.
35. George Ostrowoch, Oak Ridge National Laboratory, TN, USA.
36. Guangzhi Qu, Oakland University, Rochester, MI, USA.
37. Rahul Ramachandran, University of Alabama, Huntsville, AL, USA.
38. Pedro Rodrigues, University of Porto, Portugal.
39. Josep Roure, Carnegie Mellon University, Pittsburgh, PA, USA.
40. Bernhard Seeger, University Marburg, Germany.
41. Cyrus Shahabi, University of Southern California, USA.
42. Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA.
43. Mallikarjun Shankar, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
44. Lucio Soibelman, Carnegie Mellon University, Pittsburgh, PA, USA.
45. Alexandre Sorokine, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
46. Eduardo J. Spinosa, University of Sao Paulo, Brazil.
47. Karsten Steinhaeuser, University of Notre Dame, IN and Oak Ridge National Laboratory, TN, USA.
48. Nithya Vijayakumar, Cisco Systems, Inc., USA.
49. Gary Weiss, Fordham University, NY, USA.
50. Peng Xu, ExxonMobil Corporate Strategic Research, NJ, USA.
51. Eiko Yoneki, University of Cambridge, UK.
52. Philip S. Yu, IBM Watson Research Center, Yorktown Heights, NY, USA.

Workshop website: http://www.ornl.gov/sci/knowledgediscovery/SensorKDD-2010
# Table of Contents

## Full Research Papers

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Recognition using Cell Phone Accelerometers</td>
<td>Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore</td>
<td>10</td>
</tr>
<tr>
<td>A New Algorithm Based on Sequential Pattern Mining for Person Identification in Ubiquitous Environments</td>
<td>Belkacem Chikhaoui, Shengrui Wang, and Helene Pigot</td>
<td>19</td>
</tr>
<tr>
<td>Activity Recognition Using Actigraph Sensor</td>
<td>Raghavendiran Srinivasan, Chao Chen, and Diane Cook</td>
<td>29</td>
</tr>
<tr>
<td>Network Comprehension by Clustering Streaming Sensors</td>
<td>Pedro Pereira Rodrigues, Joao Gama, Joao Araujo, and Luís Lopes</td>
<td>35</td>
</tr>
<tr>
<td>Energy Prediction Based on Resident’s Activity</td>
<td>Chao Chen, Barnan Das, and Diane Cook</td>
<td>45</td>
</tr>
</tbody>
</table>

## Short Research Papers

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi Home Transfer Learning for Resident Activity Discovery and Recognition</td>
<td>Parisa Rashidi and Diane Cook</td>
<td>53</td>
</tr>
<tr>
<td>Using Semantic Annotation for Knowledge Extraction from Geographically Distributed and Heterogeneous Sensor Data</td>
<td>Alexandra Moraru, Carolina Fortuna, and Dunja Mladenic</td>
<td>63</td>
</tr>
<tr>
<td>Random Kernel Perceptron on ATTiny2313 Microcontroller</td>
<td>Nemanja Djuric and Slobadan Vucetic</td>
<td>70</td>
</tr>
<tr>
<td>Anomalous Thermal Behavior Detection in Data Centers using Hierarchical PCA</td>
<td>Manish Marwah, Ratnesh Sharma, Wilfredo Lugo, and Lola Bautista</td>
<td>78</td>
</tr>
<tr>
<td>Self-Organizing Energy Aware Clustering of Nodes in Sensor Networks using Relevant Attributes</td>
<td>Marwan Hassani, Emmanuel Muller, Pascal Spaus, Adriola Faqolli, Themis Palpanas, and Thomas Seidl</td>
<td>87</td>
</tr>
<tr>
<td>Anomaly Localization by Joint Sparse PCA in Wireless Sensor Networks</td>
<td>Ruoyi Jiang, Hongliang Fei, and Jun Huan</td>
<td>97</td>
</tr>
</tbody>
</table>
FULL RESEARCH PAPERS