



## Newsletter

Winter 2018

### MERIDIAN: Marine Environmental Research Infrastructure for Data Integration and Application Network

We are happy to announce that the MERIDIAN project has kicked off! Led by the Big Data Analytics Institute at Dalhousie University, the MERIDIAN is a three years project funded by the Canada Foundation for Innovation's (CFI). The CFI has invested a total of \$5 million to build a consortium of highly skilled professionals from all over Canada who, from now on, will have the opportunity to work together on a broad spectrum of marine data science problems and push mankind's knowledge about oceans.

By focusing on marine acoustic and satellite tracking data (e.g. the automatic identification system (AIS) - an automatic tracking system used on ships), MERIDIAN presents a unique opportunity to use data integration, management, and analysis expertise to create a world-class research infrastructure to effectively examine and quantify Canada's ocean soundscape and determine how Canada's marine activities can best mitigate their noise impact on the marine environment. MERIDIAN's focus on ocean data aims to become a reference for marine acoustic-related data "know-how" and resources, which will give us national and international visibility and establish Canada as a global leader in the field.



MARINE ENVIRONMENTAL RESEARCH  
INFRASTRUCTURE FOR DATA  
INTEGRATION AND APPLICATION NETWORK

The MERIDIAN Consortium builds on the scientific strength of leading ocean research organizations in Canada and will sustain Canada's current global leadership in ocean research covering virtually all natural and social science disciplines. The MERIDIAN team draws on the expertise in



academic institutions across Canada including computer scientists and ocean acoustics researchers from Dalhousie University, Simon Fraser University, University of British Columbia, University of Victoria, and Université du Québec à Rimouski, as well as from the Department of Fisheries and Oceans and organizations such as the Dalhousie-based Ocean Tracking Network (OTN), the Atlantic Computational Excellence Network (ACENET) and the Memorial University of Newfoundland (MUN). At the Institute of Big Data and Analytics at Dalhousie, the MERIDIAN's team will consist of a senior data manager, two data management experts, a lead data analyst, an acoustic data analyst visualization expert, an acoustic data analyst, and data software engineer, and the recruitment and hiring process has just started (see <http://www.dal.ca/careers>).

Over the course of three years, the MERIDIAN team will develop a research infrastructure to support data discovery, data integration and interoperability, data visualization, and data analysis. The MERIDIAN research infrastructure will assist the Canadian ocean research community in the use of advanced data science technologies. With MERIDIAN Canadian ocean researchers will maintain and extend their global leadership and level of excellence by fully exploiting Canada's ocean data, monitoring trends in the state of the ocean acoustic environment, and enabling more timely, effective and efficient protection of valued marine species and protected areas.

Above all, MERIDIAN is an opportunity to train the future generation of highly qualified people cutting edge of acoustic research technology.

We are looking forward to working with our exceptional team of experts and develop strong partnerships on the way – between people from different fields at MERIDIAN nodes, as well as across Canada and on an international level.



# News

## KDD 2017: An international Success for Both Experts and Students

In August the world's premier interdisciplinary conference on data science, data mining, knowledge discovery, large-scale data analytics, and big data, KDD2017, came to Halifax, chaired by Stan Matwin, Director of the Institute for Big Data Analytics in collaboration with Shipeng Yu of LinkedIn and a team of organizers including Local Chair Evangelos Milios. This was no small accomplishment; the conference has typically been hosted in some of the biggest cities in the world, including Beijing, Sydney, New York and Paris. Being invited to host this year's conference was a significant recognition of the strength of our team in this field and the strength of the proposal they had put forward. It was also a big event for the city of Halifax and the Province of Nova Scotia. In his opening remarks Mayor Mike Savage conferred upon Stan and Evangelos the honorary title of Ambassadors for the City of Halifax. As the conference progressed researchers in both academia and industry in this province benefited from access to the latest research and the leading practitioners in the field.



Dr. Evangelos Milios and Dr. Stan Matwin

For 5 days over 1600 participants from 51 countries met for a wide range of activities and networking opportunities. This was the largest KDD conference held outside of the United States and the third largest overall, after KDD2016 in San Francisco and KDD2014 in New York. It was held at the World Trade and Convention Centre on Argyle Street, with the largest events taking place in the Scotiabank Centre and the Cunard Centre on Marginal Road.

The conference continued its tradition of a strong tutorial and workshop program on leading issues of data mining during the first two days, including some hands-on tutorials on practical data science tools. The last three days were devoted to contributed technical papers, describing both novel, important research contributions and deployed, innovative solutions. An outstanding lineup of industry speakers shared their expertise in deploying industrial data mining solutions. Three keynote talks, by Cynthia Dwork, Bin Yu and Renée J. Miller touched on some of the hard, emerging issues before the field of data mining. A KDD panel brought together industry experts to discuss the future of artificially intelligent assistants. There was also a "Meet the Editors" Panel – to make KDD researchers aware of the academic journal opportunities for publishing. And a new feature this year was the series of "Meet the

Experts" sessions – informal roundtable meetings where graduate students and early career researchers could discuss research programs and career development plans with established experts.

Recognizing the importance of student participation, KDD2017 awarded a record amount of \$145,000 to support student travel. Students were also offered the opportunity to work at the conference, doing a range of essential tasks such as working at the registration desk, putting swag bags together, helping with events, or helping delegates to find their way around. In exchange for 8 hours of work, free access to all conference activities was provided. Some 40 graduate students from at least three Nova Scotia universities took this opportunity to participate in the conference. Many students also benefitted from the Broadening Participation in Data Mining (BPDMD) workshop, receiving mentoring in how to more effectively pursue their ambitions in the field. Zahraalsadat Alavizadeh, currently doing a Masters in Electrical Engineering at Dalhousie, spoke to industry representatives about trends, job requirements and how to become more eligible to employers. The Google mentor helped her to format and optimize her CV. She left with a fresh idea about changing her field of study to Computer Science for her PhD. Another Dalhousie student, Sara Khanchi, currently working on her PhD in Computer Science was given helpful advice on how to present a pitch to others and gained some valuable insight into the issue of whether to pursue a career in academia or industry.



Some of the research presented at KDD2017 has already begun to receive attention in the media. IT World Canada mentioned a research paper on embedding based online local event detection in geo-tagged Tweet streams. Business Insider published a feature on the presentation from Amazon Labs researchers on their work on an algorithm that can learn about styles from images and recreate similarly styled garments from scratch. The Daily California and Berkeley News highlighted the presentation by UC Berkeley student Rebecca Portnoff on her creation of two algorithms aimed to scan through online sex advertisements and identify human trafficking circles. ZDNet profiled the analogy gap research conducted by Carnegie Mellon University researchers and presented at the conference.

Halifax has proven itself to be ready to host premier international events in the high tech area; it is hoped that more such world-class events will follow. KDD2018 will be held in London, England, in August, 2018.



# Organizing Through Visualization



Dr. Fernando Paulovich arrived at Dalhousie in April to take up his position as a Tier 2 Canada Research Chair in the Faculty of Computer Science. Previously based at the University of São Paulo, in São Carlos, Brazil, he brings with him two students and a lot of research questions. Although the move meant leaving behind a well established research lab in a large faculty he welcomes the opportunities here in Canada to make closer research collaborations with industry and to be part of a more stable research funding environment.

Dr. Paulovich describes his research as Information Visualization and Visual Analytics. Much of his work so far has been motivated by problems he encounters in his own life. A music lover, he realized he had a personal collection of thousands of MP3s but only currently played a small selection of these, the remainder becoming more and more forgotten over time. The pieces of music that gradually slipped into this forgotten space were no less valuable than the ones that were still being currently played, but it would require too much cognitive effort to retain an active awareness of all of them. Applying the tools of Visualization to this collection Dr. Paulovich used machine learning to classify all of the pieces of music on a number of different parameters: genre, mood, gender, and intensity. He then created a visual interface which would plot all of the pieces of music according to these parameters, with similar pieces of music occupying positions

close to each other. It is then a simple thing, if you have been enjoying a song by Amy Winehouse, to be reminded of other adjacent songs in your collection that are similar. In this way we can see the whole collection at once and be reminded through a simple visualization of other songs that would appeal to us in our current mood. The visualization also enables the user to rotate the orientation of the parameters used in the multidimensional space so that any combination of parameters line up, grouping certain characteristics more definitely. If you want to see all the sad songs by female vocalists you can line up the parameters and cluster them all together.

Such techniques can apply to many different media collections. Another area with direct personal relevance is research papers. Conducting a literature review is becoming more and more difficult due to the larger and larger volumes of papers and an increasing production rate of new papers related to any given topic. Tools to organize such collections are increasingly needed.

Here at Dalhousie Dr. Paulovich is interested in the near future in investigating collections of data which evolve over time. He will be building up his research lab, recruiting new graduate students and making connections with other researchers and industry partners. In the long term he is interested in such problems as how to extract answers to common sense questions from the massive amounts of data available through search engines such as Google. If you want to decide whether or not to move from Brazil to Halifax there should be a simpler way to access the whole knowledge base and get a reliable answer based on the parameters that are important to you. That might take some time.

## MASTER Project on Trajectory Research Funded by EU

The MASTER (Multiple ASpects Trajectory management and analysis) project consists of an international and inter-sectoral consortium of organisations/partners working on a joint research programme in the field of holistic trajectories. From a scientific point of view, the partners build, manage and analyse holistic trajectories by considering as vital issues the privacy and big data related aspects. They propose ways to analyze and infer knowledge from multiple aspects, then, ultimately, use the resources in an innovative way, in practical and useful applications: tourism, sea monitoring, public transportation. The participants will exchange skills and knowledge which will allow them to progress towards key advances in the research field, and to strengthen research collaborations between different countries and sectors. Advances in holistic trajectory data management and analysis will have potential market opportunities for non-academic participants in the project and significant benefits for the economies of the participating countries. The staff members who participate in the project will develop new skills through the writing of scientific papers for top-level venues, and will be exposed to new international research, non-academic environments and training activities widening their career perspectives.

The consortium includes universities from Italy, Greece, France, Brazil and Canada (Dalhousie University) as well as the Municipality of Thira in Greece. Our membership in this project is based, in part, on our achievements in exploiting machine learning techniques to enrich the trajectory data of ship traffic and we look forward to collaborating in the further development of this research. The project will leverage the trajectory datasets that are available to each partner at the start of the project and should develop new and lasting research collaborations, achieve transfer of knowledge between participating organizations and contribute to the improvement of research and innovation.

The program has substantial funding from the European Union in the amount of 504,000 Euros. This grant is mainly a mobility grant, facilitating the travel of participants to the participating sites. We anticipate a total of 22 European researchers to be hosted here at Dalhousie University over a four year period. Visitors will stay with us for periods of time ranging from one to three months.



## CASCON 2017: Best Paper Award

Congratulations to Xiang Jiang, Erico Neves De Souza, Ahmad Pesaranghader, Baifan Hu, Danny Silver and Stan Matwin for the Best Paper Award at CASCON 2017 in Toronto in November. Their paper, "TrajectoryNet: An Embedded GPS Trajectory Representation for Point-based Classification Using Recurrent Neural Networks" is described in the abstract below:

Understanding and discovering knowledge from GPS (Global Positioning System) traces of human activities is an essential topic in mobility-based urban computing. We propose TrajectoryNet—a neural network architecture for point-based trajectory classification to infer real world human transportation modes from GPS traces. To overcome the challenge of capturing the underlying latent factors in the low-dimensional and heterogeneous feature space imposed by GPS data, we develop a novel representation that embeds the original feature space into another space that can be understood as a form of basis expansion. We also enrich the feature space via segment-based information and use Maxout activations to improve the predictive power of Recurrent Neural Networks (RNNs). We achieve over 98% classification accuracy when detecting four types of transportation modes, outperforming existing models without additional sensory data or location-based prior knowledge.

The paper: <https://arxiv.org/abs/1705.02636>

### Director

**Dr. Stan Matwin PhD, CRC**

### Executive Committee

**Dr. Sylvain Charlebois**  
*Dean of Management*

**Dr. David Anderson**  
*Dean of Medicine*

**Dr. Evangelos Milios**  
*Associate Dean, Research  
Faculty of Computer Science*

**Dr. Andrew Rau-Chaplin**  
*Dean of Computer Science*

**Dr. Nur Zincir-Heywood**  
*Faculty of Computer Science*

**Dr. Stan Matwin**  
*Director, (ex officio),  
Faculty of Computer Science*

### Advisory Board

**Sue Abu-Hakima Amika**  
*President and CEO,  
AmikaMobile, Ottawa, ON*

**Martin Davis**  
*Vice-President of IT, JD Irving,  
Moncton, NB*

**Peter Hickey**  
*Co-founder, Oris4, Halifax, NS*

**David Kasik**  
*Senior Technical Fellow,  
The Boeing Company, Seattle,  
Washington*

**Stephen Perelgut**  
*IBM Canada University Relations  
Manager, Markham, ON*

*Newsletter Editor:*  
**David Langstroth**  
*dll@cs.dal.ca*

*"Information is the oil of the 21st century, and analytics is the combustion engine"* - Peter Sondergaard, Senior Vice President, Gartner