Scientific Director’s Message

We continue to grow our postdoctoral matching fund program. As of this fall, there are now seven exciting projects on diverse topics in data science and related applications. Four more postdoctoral fellows have joined the DSI; please read their short bios below. Our fellows will lead a journal club discussing state-of-the-art data science research. If you want to join, please contact us for more details.

Another exciting development this summer for the DSI was the Data Science for Social Good (DSSG) program. While more details are given below and on our website, we thank our collaborators - the eScience Institute at the University of Washington, the City of Surrey and the tourism unit in BC Government’s Ministry of Tourism, Arts & Culture. The program is sponsored by Microsoft, which provides technical mentoring and funding for our DSSG fellows. As successful as this program was in 2017, we will continue to fine tune to make the 2018 program even better. - Raymond Ng

New DSI Projects

Congratulations to these four projects on receiving funding from DSI and its partners:

**A platform for interactive, collaborative, and repeatable genomic analysis:** This project employs a whole system approach for providing a framework for genomic analysis by building on an existing pipeline and exploiting emerging high-density “rack-scale” computer hardware. This will facilitate sharing of datasets and analysis code across the research community and will provide sufficient capture of data and analysis provenance to encourage reproducibility of published results. Awarded to Drs. Loren Rieseberg and Andrew Warfield.

**Application of deep learning approaches in modelling cheminformatics data and discovery of novel therapeutic agents for prostate cancer:** This project is developing a deep neural network environment with a re-enforced learning component to capture all available information on 100s of millions of existing small molecules—including their interactions with proteins and other cell components. Such a platform will then be integrated with existing technologies of high-throughput synthesis to yield a paradigm-shifting ‘molecular printer’ to revolutionize life science. Awarded to Drs. Artem Cherkasov and William Welch. This project is sponsored by PHIX and VCHRI.

**Using text analysis for chronic disease management:** This project seeks to assess patients’ states (e.g. symptoms, emotions) by computationally modeling patients’ language use. This computational work involves speech and/or text analysis on patient-generated language and integrated learning of language and sensor-based physiological data. Ultimately, the project will help caregivers better care for patients with chronic diseases. Awarded to Drs. Kendall Ho and Giuseppe Carenini. This project is sponsored by PHIX and VCHRI.

**User Modeling and Adaptive Support for MOOCS:** This project aims to apply concepts and techniques from Intelligent Tutoring Systems research—a field that uses AI and machine learning to devise educational tools that provide instructions tailored for the individual learner—to massive open on-line courses (MOOCS). The benefit of this project is to make existing MOOCS more reactive to specific student needs to improve learning outcomes. Awarded to Drs. Cristina Conati and Ido Roll.
New DSI Postdoctoral Fellows

Dr. Hyeju Jang received her PhD in Computer Science from Carnegie Mellon University. Her research interests include natural language processing (NLP), computational linguistics, discourse analysis, and text mining in various domains.

Dr. Halyon Jeong received his PhD in Mathematics from NYU. His research interest is in mathematical signal processing including information theory and analog-to-digital (A/D) conversion, dynamical systems and stochastic processes, and geometry of high-dimensional data sets.

Dr. Ezequiel Smucler received his PhD in Mathematics from University of Buenos Aires in 2016. His research interests include statistics for high-dimensional data, dimension reduction for time series, robust statistics, and applications of natural language processing.

Dr. Michael Fernandez Llamosa received his PhD in Computer Sciences and Systems Engineering from the Kyushu Institute of Technology. His research interest is in machine learning and evolutionary computing strategies to identify structural-property relationship patterns in chemical, biomedical and advanced manufacturing data.

Data Science for Social Good

The Data Science Institute hosted the inaugural Data Science for Social Good—also known as DSSG—program this past summer (https://dsi.ubc.ca/dssg). This 14-week summer research program brought together 16 UBC undergraduate and graduate students from diverse backgrounds (e.g., computer science, economics, engineering, mathematics, physics, psychology, public policy and global affairs, statistics) to work on collaborative data science projects with municipal and provincial government organizations. This year the projects focused on the topics of early childhood education, economic development, transportation, and tourism.

The DSSG program is a key program of the Cascadia Urban Analytics Cooperative (CUAC), an applied, interdisciplinary, regional centre formed by the University of Washington and UBC in 2017 with funding from Microsoft. This initiative brings together academic researchers, students, and public stakeholder groups from both sides of the border to address topics, usually in an urban-context, affecting citizens of the Cascadia region. Specifically, CUAC leverages data science techniques embedded in interdisciplinary collaborations to explore these problems. These rich partnerships include researchers and professors across diverse fields (e.g., public policy, urban planning, population health, civil engineering, landscape architecture, sociology, real estate, computer science and law) working with non-profit organizations, city planners and government agencies. Examples of these projects include the “Transportation Data Collaborative” and “Neighborhood Research Collective”. For more information about these projects and CUAC, please visit http://cascadiadata.org/.