A NEW TRAINING PROGRAM IN DATA ANALYTICS & VISUALIZATION

JAMES H. ELDER

Department of Electrical Engineering & Computer Science, Lassonde School of Engineering
Department of Psychology, Faculty of Health
Centre for Vision Research
York University, Toronto, Canada

The growing appreciation for the role of Big Data in addressing important problems and societal needs has led to many welcome developments, including the new journal we are celebrating with its inaugural issue today. In Toronto, an interdisciplinary network of faculty has been working together on big data problems for a number of years through the Centre for Information Visualization and Data-Driven Design (CIVDDD). Founded and Led by Nick Cercone (York) and supported by the Ontario provincial government, this network has brought together researchers in data analytics and visualization from York University, OCAD University and the University of Toronto with industry partners to tackle challenging Big Data problems. A key emphasis of CIVDDD is the importance of fusing analytics with powerful visualization methods that allow the full value of the data to be extracted.

Also critical to exploiting the power of big data is the development and nurturing of young researchers in this area. In Canada we are fortunate to have a program operated by the Natural Sciences and Engineering Research Council (NSERC) that can help. The Collaborative Research and Training Experience (CREATE) Program is designed to support collaborative and integrative training, particularly of graduate students, to address significant scientific challenges. In addition to scientific training, the program emphasizes interdisciplinary training, the development of professional skills and linkages to the industrial sector that provide trainees with broader career options.

This is an ideal program for training in the area of Big Data research, and **in April 2015 a team of researchers from York University (lead institution), OCAD University and the University of Toronto, all in Toronto, with Concordia University (Montreal) were awarded a 6-year \$1.65M NSERC CREATE grant to establish a program in Data Analytics & Visualization (DAV)**. This program brings together researcher laboratories from these four institutions with a number of leading Canadian high-tech companies to support the training of more than 50 graduate and undergraduate students and postdoctoral fellows. Our goal is to provide training in both computational analytics and perceptual design methodologies for deriving value from big data and thus to produce multidisciplinary researchers who can combine signal processing, mathematical modeling, statistical analysis and machine learning methods with cognitive science and visual and design practices.

The core research team consists of researchers from Electrical Engineering & Computer Science, Mathematics & Statistics and Earth & Space Science & Engineering at York University, Electrical and Computer Engineering at Concordia University, Computer Science at the University of Toronto, and Design researchers from OCAD University. Each of the four institutions brings complementary expertise to the program:

- York University: Data mining, computer vision, geomatics, computer graphics
- OCAD University: Visualization and design
- Concordia University: Signal processing
- University of Toronto: Computational biology

The academic team consists of Aijun An, Nick Cercone, James Elder (lead applicant), Petros Faloutsos, Gunho Sohn and Jianhong Wu from York University, Peter Coppin and Sara Diamond from OCAD University, Amir Asif from Concordia University and Igor Jurisica from the University of Toronto. This team is augmented by a network of nine international academic collaborators.

Application areas vary widely but particular foci include:

1. **Cloud Computing**: Distributed, continuous and scalable data stream mining, platforms
2. **Digital Cultures**: Social media text analytics for attention, influence and affect, concept mapping and clustering, personal visualization, supply chain management, customer personalization
3. **Health Technology**: Medical imagery, blood flow, cognitive and neurological data, epidemiology, mobile apps for the disabled, patient / elderly monitoring, best practices decision-support
4. **Smart Cities**: Transportation, sustainability

Canadian industry collaborators include Autodesk Research (3D design), BBM Analytics (broadcast media), Biodiaspora (epidemiology), CaseBank (databases), Dapasoft (software engineering), Empress Software (databases), Esri Canada (GIS), Fugro Roadware (transportation), GestureTek (gesture control), IBM Canada (IT), Infersystems (marketing), Manifold Data Mining (marketing), Mircom (fire safety), NICE (contact centres), OMSignal (biometrics), Optech (3D sensing) and Side Effects (film).

We expect that the resulting enrichment in Data Analytics & Visualization expertise in Canada will yield many economic and social benefits. On the economic front, improving the utility of analytics extracted from raw data, and providing more compelling visualizations of the results will result in increased market share and opening of new markets for Canadian companies.

The benefits for society are numerous as well. For social and economic networks, inferring and visualizing the structure of community relations and friendships quantifies the connectedness of the world. For large-scale complex infrastructure, determining the relevance of particular components and their interactions leads to lower maintenance costs. In healthcare, determining and visualizing movements and interactions of people allows diffusion of disease to be modeled so that measures can be taken to reduce disease transmission. Researchers trained in DAV will also contribute to our understanding of transportation systems, the spread of news and information, voting trends, imitation and social influence and crowd behaviour.

Our objective is to produce world-class researchers and practitioners skilled in fusing data analytics with visualization methodologies to tackle a range of important

real-world problems. The proposed program consists of a number of unique training components that together will deliver both the interdisciplinary research skills as well as the professional skills that will allow trainees to translate this knowledge into productive collaboration with industry and public sector organizations. These include:

1. **Industrial Internship**. All graduate and postdoctoral trainees will complete a 5-month internship with one of our industrial partners.

2. **Interdisciplinary Rotation**. All PhD and Postdoctoral trainees will be required to complete a 3-month interdisciplinary rotation in a laboratory other than their own (optional for Masters students), in order to broaden their research skills.

3. **Unique Cross-Institutional Project Course**. To spearhead training, the 4 participating institutions are collaborating on the development of a new inter-institutional graduate course on Data Analytics & Visualization specifically for this initiative.

4. **Unique Interdisciplinary Curriculum**. The program will offer a broad, closely-coordinated selection of more than 46 graduate courses in sensing, signal processing, computer vision, data analytics, geomatics, computer graphics, visualization and design offered by the 4 participating institutions.

5. **Annual Boot Camp on Data Analytics & Visualization**. A three-day Boot Camp will be held annually, bringing trainees together with our industry partners, and focusing on key application areas as well as professional skills development.

6. **Site Visits**. Trainees will participate in annual visits to industry sites in the Toronto (e.g., IBM, MTO, Mircom, Baycrest Health Sciences, Autodesk, Manifold Data Mining) and Montreal (e.g., Bombardier Aerospace, Bell Helicopter Textron Canada Limited, Pratt & Whitney Canada Limited, Altair Engineering, Marinvent Corporation) regions.

7. **Summer School in DAV**. We plan to establish an annual three-day summer school for senior undergraduates, to be held at York University.

8. **Partnership with the Canadian Visual Analytics School (CANVAS)**. We will participate fully in the planning and delivery of this existing annual Canadian summer school.

We welcome inquiries from all interested parties, particularly potential graduate student or postdoctoral trainees (Canadian and international). Please contact James Elder at `jelder@yorku.ca` (`www.yorku.ca/jelder`) for more information.