AAI Advanced Analytics Seminar Series on 21/09/2012

Seminar Title: Recent developments and some thoughts for the future of advanced analytics: The View from the bridge at Salford Systems
Speaker: Dan Steinberg, Ph.D., CEO and Founder, Salford Systems

Date and Time: 3:30pm to 5:00pm, the 21st of September 2012 (Friday)
Seminar Room: UTS Blackfriars Campus Building 05 CC05.GD.02 (5 minutes walk from the Tower Building CB01 of UTS)
Seminar Chairman: Professor Longbing Cao (longbing.cao@uts.edu.au)

Abstract: Salford Systems is best known for its pioneering products CART, MARS, and TreeNet, all invented and coded in proprietary software by legendary Stanford data mining visionary Jerome H. Friedman. TreeNet, whose first edition was written on Manly beach in 1999 ushered in the era of gradient boosting machines, now perhaps the most popular of all advanced learning machines. In the first part of this talk we will discuss the advances related to TreeNet due to subsequent work by Friedman and enhancements developed at Salford. The three key advances are (1) ensemble compression via regularized regression (ISLE, importance sampled learning ensembles, (2) rule extraction from learning ensembles via regularized regression (RuleSeker), and interaction detection via constrained tree ensembles (Interaction Control Language, ICL). These new tools will be explained and illustrated with examples using Salford' soon to be released SPM 7.0 (Salford redictive Modeler).

In the second part of the talk we will address some strategies for handling large data sets on single but powerful severs (32-200 cores, 512GB- 2 TB RAM) which have now become cheaper than many small cars. While still not "Big Data" running SPM 7.0 on such servers permits the analysis of data sets with data on every customer of most of the largest Australian corporations. We also will comment on our plans for 2013 which we hope will see the first release of H-TreeNet (TreeNet for Hadoop).

Short biography of the speaker: Dan Steinberg is CEO and founder of Salford Systems, the developer of the CART® decision tree, MARS® spline regression, TreeNet® gradient boosting, Breiman's RandomForests®, and other influential data mining technology. After earning a PhD in Econometrics at Harvard Dan began his professional career as a Member of the Technical Staff at Bell Labs, Murray Hill, and then as Assistant Professor of Economics at the University of California, San Diego. His consulting experience at Salford Systems has included complex modeling projects for major banks worldwide, including Citibank, Chase, American Express, Credit Suisse, and has included projects in Europe, Australia, New Zealand, Malaysia, Korea, Japan and Brazil. Dan led the teams that won first place awards in the KDDCup 2000, and the 2002 Duke/TeraData Churn modeling competition, and the teams that won awards in the PAKDD competitions of 2006 and 2007. Dan has published papers in economics, econometrics, computer science journals, and contributes actively to the ongoing research and development at Salford.

Overview to This Seminar Series

The Advanced Analytics Seminar Series presents the latest theoretical advancement and empirical experience in a broad range of interdisciplinary and business-oriented analytics fields. It covers topics related to data mining, machine learning, statistics, bioinformatics, behavior informatics, marketing analytics and multimedia analytics. It also provides a platform for the showcase of commercial products in ubiquitous advanced analytics. Speakers are invited from both academia and industry. It opens regularly on every Friday afternoon at the garden-like UTS Blackfriars Campus. You are warmly welcome to attend this seminar series.

Jinyan Li, Seminar Coordinator, Associate Professor
Advanced Analytics Institute, School of Software, Faculty of Engineering and IT
University of Technology, Sydney
P.O. Box 123, Broadway, NSW 2007, Australia
Tel: 02 95149264 (office); http://www-staff.it.uts.edu.au/~jinyli