A Tutorial Abstract for WCCI 2014

A) Title: The Emerging "Big Dimensionality"

B) Proposer's name, title, affiliation, and email:

Dr. Ivor W. Tsang: Associate Professor, University of Technology, Sydney, Ivor.Tsang@gmail.com

Dr. Yew Soon Ong: Associate professor, Nanyang Technological University, asysong@ntu.edu.sg

C) Tutorial abstract:

The world continues to generate quintillion bytes of data daily, leading to the pressing needs for new efforts in dealing with the grand challenges brought by Big data. Today, there is consensus among the computational intelligence communities that data volume presents an immediate challenge pertaining to the scalability issue. However, when addressing volume in Big data analytics, researchers have taken a one-sided study of volume, which is the "Big instance size" factor of the data. The flip side of volume which is the dimensionality factor of Big data, on the other hand, has received much lesser attention.

In this tutorial, we will present an attempt to fill in this gap and places special focus on this relatively under-explored topic of "Big dimensionality", wherein features (variables) play somewhat of a trouble maker to computational intelligence. We begin with an analysis on the origins of "Big dimensionality". The evolution of feature dimensionality in the last two decades is then investigated using popular data repositories considered in the data analytics and computational intelligence research communities. Subsequently, the state-of-the-art feature selection schemes reported in the field of computational intelligence are reviewed to reveal the inadequacies of existing approaches in keeping pace with the emerging phenomenon of "Big dimensionality". Last but not least, the curse and bless of "Big dimensionality" are delineated and deliberated. Some recently developed feature selection algorithms to address Big Dimensionality and Big Data will be discussed.

D) URL address for eventual course materials:

The tutorial materials are mainly from the recent work published by the organizers and other research groups. We will try to make the Powerpoint slides publicly available at http://www.c2i.ntu.edu.sg/bigdata

E) Tutorial Duration:

The tutorial duration is 2 hour, each speaker deliveries 1 hour speech.

Biographies of the organizers

Ivor W. Tsang is an Associate Professor with the Centre for Quantum Computation & Intelligent Systems (QCIS), University of Technological, Sydney. He received his Ph.D. degree in computer science from the Hong Kong University of Science and Technology in 2007. His research focuses on kernel methods, transfer learning, feature selection, big data analytics for data with millions of dimensions, and

their applications to computer vision and pattern recognition. He has more than 100 research papers published in refereed international journals and conference proceedings, including JMLR, T-NN, T-PAMI, Neural Computation, NIPS, ICML, UAI, AISTATS, SIGKDD, IJCAI, AAAI, IJCNN, ICCV, CVPR, ECCV, etc.

Dr. Tsang received the prestigious IEEE Transactions on Neural Networks Outstanding 2004 Paper Award in 2006, the Australian Research Council Future Fellowship in 2013, and the second class prize of the National Natural Science Award 2008, China in 2009. His research also earned him the Best Paper Award at ICTAI'11, the Best Student Paper Award at CVPR'10, the Best Poster Honorable Mention at ACML'12 and the Best Paper Award from the IEEE Hong Kong Chapter of Signal Processing Postgraduate Forum in 2006. He was also conferred with the Microsoft Fellowship in 2005. He has delivered tutorials regarding *Domain Transfer Learning for Vision Applications* at CVPR 2012, and *Domain Adaptation for Real World Applications* at ACML 2012. He also gave an invited lecture about *Structural Feature Selection for Very High Dimensional Problems* in Machine Learning Summer School 2011. He has served as the workshop co-chair of NIPS 2009 workshop on Transfer Learning for Structured Data. He will serve as the local arrangement co-chair of Asian Conference of Machine Learning (ACML 2012) in Singapore.

Yew Soon Ong is currently an Associate Professor and Director of the Center for Computational Intelligence at the School of Computer Engineering, Nanyang Technological University, Singapore. He received a PhD degree on Artificial Intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, UK in 2003. His current research interests include memetic computing, evolutionary optimization, machine learning, multi-agent modeling & simulation as well as artificial intelligence in Games. He has coauthored over 130 refereed publications and his recent research grant in 'Big Data' is 19.7 million Singapore dollars.

Dr. Ong is founding Technical Editor-in-Chief of the Memetic Computing Journal, Chief Editor of the Springer book series on studies in Adaptation, Learning, and Optimization, Associate Editor of the IEEE Computational Intelligence Magazine, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, Information Sciences, Soft Computing and others. Presently, he chairs the IEEE Computational Intelligence Society Intelligent Systems and Applications Technical Committee. His research work on Memetic Algorithm was featured by Thomson Scientific's Essential Science Indicators as one of the most cited emerging area of research in August 2007. Recently, he also received the prestigious 2012 IEEE Transactions on Evolutionary Computation Outstanding Paper Award for his work pertaining to the modeling of Probabilistic Memetic Framework. In teaching, he has also received numerous awards including the Nanyang Excellence Award for Teaching in 2008, Most Popular Lecturer Award 2009, and recently invited as Fellow of Renaissance Engineering Programme at Nanyang Technological University.

Recent related publications by the organizers

Mingkui Tan, **Ivor W. Tsang**, Li Wang. Minimax Sparse Logistic Regression for Very High Dimensional Feature Selection. IEEE Transactions on Neural Networks and Learning Systems, 24(10): 1609-1622, Oct 2013.

Qi Mao, **Ivor W. Tsang**. A Feature Selection Method for Multivariate Performance Measures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(9): 2051 - 2063, Sept 2013.

Qi Mao, **Ivor W. Tsang**. Efficient Multi-Template Learning for Structured Prediction. IEEE Transactions on Neural Networks and Learning Systems, 24(2): 248 - 261, Feb 2013.

Zhongwen Xu, Yi Yang, **Ivor W. Tsang**, Alexander Hauptmann, Nicu Sebe. Feature Weighting via Optimal Thresholding for Video Analysis. Proceedings of the International Conference on Computer Vision (ICCV), 2013.

Yufeng Li, **Ivor W. Tsang**, James T. Kwok, Zhi-Hua Zhou. Convex and Scalable Weakly Labeled SVMs. Journal of Machine Learning Research (JMLR), 14:2151-2188, July 2013.

Yiteng Zhai, Mingkui Tan, **Ivor W. Tsang, Yew-Soon Ong**. Discovering Support and Affiliated Features from Very High Dimensions. Proceedings of the 29th International Conference on Machine Learning (ICML), Edinburgh, Scotland, 2012.

Qi Mao, **Ivor W. Tsang**. Optimizing Performance Measures for Feature Selection. Proceedings of the IEEE International Conference on Data Mining (IEEE ICDM 2011), Vancouver, Canada, Dec 2011.

Mingkui Tan, Li Wang, **Ivor W. Tsang**. Learning Sparse SVM for Feature Selection on Very High Dimensional Datasets. Proceedings of the 27th International Conference on Machine Learning (ICML 2010), Haifa, Israel, June 2010

Y. Liu, D. Xu, **I. W. Tsang** and J. Luo. Textual Query of Personal Photos Facilitated by Large-scale Web Data. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 33(5), pp. 1022-1036, May 2011.

L. Duan, D. Xu, **I. W. Tsang** and J. Luo. Visual Event Recognition in Videos by Learning from Web Data. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2010. (**Best Student Paper Award**)

Z. Zhu, **Y. S. Ong** and J. Zurada, "Identification of Full and Partial Class Relevant Genes", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 7, no. 2, pp. 263-277, 2010.

Z. Zhu, **Y. S. Ong** and M. Dash, 'Wrapper-Filter Feature Selection Algorithm Using A Memetic Framework', IEEE Transactions On Systems, Man and Cybernetics - Part B, Vol. 37, no. 1, pp. 70-76, 2007.

Z. Zhu, **Y. S. Ong** and M. Dash, 'Markov Blanket-Embedded Genetic Algorithm for Gene Selection', Pattern Recognition, Vol. 40, no. 11, pp. 3236-3248, 2007.

Ivor W. Tsang, James T. Kwok, Pak-Ming Cheung. Core vector machines: Fast SVM training on very large data sets. Journal of Machine Learning Research, 6:363-392, 2005.