TE-MING (DAVID) HUANG

BE (Hons.), Ph.D., Auck.
Email: david@yottamine.com
Homepage: www.learning-from-data.com/te-ming/
Language: English/Taiwanese/Mandarine

PROFILE

- **Expert in machine learning and data mining with over 5 years of experience. With commercial experience in both digital advertisements and telematics.**
- **Specializing in solving large-scale machine learning problem using Support Vector Machines and graph based semi-supervised learning.**
- **Strong numerical programming skills C++ and rapid prototyping skills in Matlab.**
- **Strong publication record, first author of the book "Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-Supervised and Unsupervied Learning".**
- **Team player with strong communication and collaboration skills**.

Work History

Current **Founder and CEO of Yottamine Analytics LLC**

Jan 2008-Jun 2009 **Senior Scientist in INRIX**

- Develop the Breakthrough Second Generation Traffic Prediction Platform for INRIX with huge accuracy improvement on congestive traffic state and low data density road segments.
- Working with huge and highly complex traffic data.
- Provide critical scientific support for various business activities. Lead the scientific effort for the multi-million dollars I95 Corridor Coalition vehicle probe project.

Oct 2006 – Jan 2008 **Applied Researcher in Microsoft AdCenter Lab**

- Develop a fast SVMs tool for solving large-scale text classification problem in contextual advertisement.
- Working with huge data sets and text classification problems.
- Perform analysis on various adCenter data to understand searchers' and advertisers' behaviour.
- Provide complete solutions to business problems using data analysis, data mining techniques, and statistics.

TECHNICAL SKILLS

Strong Background in Data Mining and Machine Learning:

- Research interests are to develop advance and practical data mining and machine learning algorithms for solving real-world applications, including, text classification, DNA microarray analysis and handwritten recognition.
- Developed algorithms for **support vector machines** and **graph-based semi-supervised learning** for solving **large-scale** problems.
- First author of the **monograph "Kernel Based Algorithms for Mining Huge Data Sets"** published by Springer (www.learning-from-data.com) .
- Published **one book chapter, two journal papers and five international conferences papers on the subject**. These published works can be found in my homepage (www.learning-from-data.com/te-ming/).
- **One of the conference papers also won the best paper and the best student contribution paper award in an international conference (see achievements in the education section)**.
- **One of the journal papers published is selected as the top 25 articles in the journal "Artificial Intelligence in Medicine" (see the publications section for more details).**

Expert in Developing Fast and Efficient Software Package for Mining Large Data Sets:

- Developed software SemiL (written in C++) which is **the first program in the world** that implements graph-based semi-supervised learning techniques for **large-scale** problems

(available: http://www.support-vector.ws/html/semil.html). SemiL receives approximately 100 downloads per month.

- Developed software ISDA (written in C++), a support vector machines package with GUI. ISDA was used to win the second place in pedestrian recognition competition organized by NiSIS (Nature-inspired Smart Information Systems).

Hands-on Experience in Various Machine Learning Techniques

- Expert in **classification, regression** and **novelty detection** using **support vector machines**.
- Expert in **graph-based semi-supervised** learning techniques.
- Experienced in **Decision Tree, Boosting, artificial neural network**, **nearest shrunken centroid**, and **ARX** model.

Hands-on Experiences with Text Classification and Information Retrieval:

- Successfully applied graph-based semi-supervised learning algorithm to text classification problems.
- Experienced with all the necessary steps of converting a text document into a readable representation for machine learning algorithms (e.g. TFIDF representation).
- Experienced with the Rainbow software package for text classification.
- Familiar with Lucent Search Engine.

Experience with a Variety of Programming Languages and Handling Large Data Sets

- Strong numerical programming skills in MATLAB and C++;
- Working with SQL Server on a daily basis for handling large data set.
- Experience with Perl for manipulating data.

EDUCATION

2002-2005 **Ph.D. of Mechanical Engineering** at the University of Auckland, New Zealand.

Thesis Topic Large-Scale Support Vector Machines and Semi-Supervised Learning Algorithms

Achievement

- **The Winners of the Best Paper Award and the Best Student Contribution Paper Award in 8th International Conference of Knowledge-Based Intelligent Information and Engineering Systems, 2004 (Out of 300+ papers).**

1998-2002 **Bachelor of Engineering** at the University of Auckland, New Zealand.

Achievement

- **Graduated as 1st Class Honors in Mechanical Engineering.**
- **Award for Final Year Project of Mechanical Engineering Student from Royal Aeronautical Society.**
- **Published one paper in the 5th UK Wind Engineering Society Conference**

1995-1997 **"A" Bursary**, Westlake Boys' High School, Auckland, New Zealand.

Achievement

- ***Top Scholar Award for New Zealand University Bursary Science Examination 1997. (Achieving the highest score for Bursary Science in NZ)***

REFEREES

Professor Vojislav Kecman,
VCU Engineering, Computer Science
P.O.Box 843019
Richmond, VA 23284-3019, USA

PUBLICATIONS

Book

1. Huang, T.-M., Kecman, V. and Kopriva, I., *Kernel Based Algorithms for Mining Huge Data Sets*: *Supervised, Semi-supervised, and Unsupervised Learning*, Ed. Kacprzyk, J., Series: Studies in Computational Intelligence Vol.17, Springer Verlag, 2006.

Book Chapters

2. Kecman, V., Huang T.-M., Vogt M., *Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance*, Chapter in 'Support Vector Machines: Theory and Applications, Ed. Wang, L., Series: Studies in Fuzziness and Soft Computing, Springer Verlag, Vol. 177, pp.255-274, 2005.

Journal Papers

3. Huang, T.-M., Kecman, V., *Semi-supervised Learning from Unbalanced Labeled Data – An Improvement*, 'International Journal of Knowledge-Based and Intelligent Engineering Systems', IOS Press, Vol 10, No. 1, pp. 21-27, 2006.

4. Huang, T.-M., Kecman, V., *Gene Extraction for Cancer Diagnosis using Support Vector Machines,* Artificial Intelligence in Medicine, Special Issue on Computational Intelligence Techniques in Bioinformatics, Vol 35, pp. 185-194, Elsevier Science, 2005. **Top 25 articles in AI in Medicine (Jun-Sept 2005).**

Conference Papers

5. Huang, T.-M., Kecman, V., *Performance Comparisons of Semi-Supervised Learning Algorithms*. Proceedings of the Workshop on Learning with Partially Classified Training Data, at the 22nd International Conference on Machine Learning, ICML 2005, W5,pp 45-49, Germany, 2005.

6. Huang, T.-M., Kecman, V., *Gene Extraction for Cancer Diagnosis using Support Vector Machines: An improvement and comparison with nearest shrunken centroid method.* Lecture Notes in Computer Science 3696, pp. 617-624, 2005.

7. Huang, T.-M., Kecman, V., *Semi-supervised Learning from Unbalanced Labeled Data – An Improvement*, 'Knowledge-Based Intelligent Information and Engineering Systems', Eds. Negoita, M. Gh., at al., Lecture Notes in Computer Science 3215, pp. 765-771, Springer Verlag, Heidelberg, 2004. ***Best Paper Award and Best Student Contribution Paper Award.***

8. Huang, T.-M. Kecman, V., *Bias Term b in SVMs Again*, 12th European Symposium on Artificial Neural Network, ESANN 2004, pp. 441-448, Bruges, Belgium, April 28-30, 2004.

9. Kecman, V., Vogt, M., Huang, T.-M., *On the Equality of Kernel AdaTron and Sequential Minimal Optimization in Classification and Regression Tasks and Alike Algorithm for Kernel Machines*, 11th European Symposium on Artificial Neural Networks, ESANN 2003, pp. 215-222, Bruges, Belgium, April 23-25, 2003.

10. Huang, T.-M. and Flay, R. G. J., *A Comparison of Recursive Filter and Spectral Methods for Digital Corrected Pressure Measurements Distorted by Tubing Response,* 5th UK Wind Engineering Society Conference, Nottingham, 2002.