## Foreword

Welcome to the eighth edition of the JMLR Conference and Workshop Proceeding Series for the Asian Conference on Machine Learning. This volume contains 29 papers accepted to the Eighth Asian Conference on Machine Learning (ACML 2016) held in Hamilton, New Zealand, from 16 to 18 November 2016. This volume continues the ACML tradition of having high-quality and original research papers in the area of machine learning following seven previous successful events held in China, Japan, Taiwan, Singapore, Australia, Vietnam, and Hong Kong respectively. ACML aims at providing a leading international forum for researchers in machine learning and related fields to share their original research findings, new ideas and achievements. Despite originating in the Asia-Pacific region ACML has become a worldwide conference: Submissions from regions other than the Asia-Pacific were encouraged and this years proceedings include papers with authors based in Europe, Scandinavia, India, and the Northeast of the USA, as well as the Asia-Pacific region.

This year, we continued the tradition of having two cycles and double-blind review for the proceedings track; each cycle had its own submission deadline. Papers that could not be accepted in the first cycle, but received positive feedback, were correctable and could be accepted after careful revision, were encouraged to resubmit in the second cycle, allowing the reviewers comments to be addressed. In total there were 113 submissions to the conference tracks, of which 29 were accepted into the main program, for an acceptance rate of 25.7%. A strict double-blind reviewing process was enforced, and each paper was assigned with one meta-reviewer, and at least 3 reviewers. To maintain quality, each reviewer was allocated no more than 6 papers in any one round. In total, there were 31 senior program committee members and 73 program committee members to provide expert opinion and reviews. Their contributions were essential to the quality and standard of papers selected for the conference and, without them the conference would have not been possible. Their names are acknowledged in the following pages. Finally, the Program Co-chairs considered all the reviews and meta-reviews by senior program committee members to make the final decisions for the papers.

As a new innovation this year ACML also ran an additional journal track, with a sequence of submission dates. The journal track attracted 28 papers of which 5 were accepted for final publication in the Springer journal Machine Learning, for an acceptance rate of 17.9%.
For ACML this year the overall number of accepted papers, from both the journal and proceedings tracks, was 34 from 141 submissions for a 24.1% total acceptance rate.

All accepted papers from the two tracks received both an oral and poster presentation, and those accepted after the two cycles of the proceedings track are published in this volume. The submissions covered a broad range of topics, including theoretical analyses, probabilistic models, deep learning, and applications to real world problems.

In addition to the submitted papers, we were very pleased to have two keynotes from leading experts, John Shawe-Taylor (University College London) and Vincent Tseng (Na-

tional Chiao Tung University), and invited talks from Tie-Yan Liu (Microsoft Research Asia, Beijing), Albert Bifet (Telecom ParisTech) and Aish Fenton (Netflix). In addition to the main program, we had a full day of tutorials and workshops prior to the main conference organized by the Tutorial and Workshops Chairs. Four tutorials were delivered on: Mass Estimation: Enabling density-based or distance-based algorithms to do what they cannot do, Recent Advances in Distributed Machine Learning, Bayesian Nets from the ground up, and Deep Approaches to Semantic Matching for Text. In parallel to the tutorials, we had three workshops: The Asian Workshop on Reinforcement Learning, the First New Zealand Text Mining Workshop, and the ACML Workshop on Learning on Big Data. We thank all of the speakers and organizers for putting together such a fantastic program.

ACML Steering Committee Chair Zhi-Hua Zhou provided valuable advice and support during the whole process. The General Chairs Geoff Holmes and Stephen Marsland took care of organizing issues and many other things to make sure the event ran smoothly. Stephen Marsland also took the role of Publications Chair, and has done a great job in producing the conference proceedings. Our special thanks go to the Local Co-Chairs Hannah Te Puia, Peter Reutemann and Lyn Hunt, as well as the local teams whose contributions were indispensable in making the event run so well. A special thanks goes also to Alvin Yeo for fundraising.

Last but not least, a big thank you to all participants of ACML 2016 who made it such a great event!

*November 2016*

ACML 2016 Program Co-Chairs

Bob Durrant
University of Waikato, New Zealand
bobd@waikato.ac.nz

Kee-Eung Kim
KAIST, South Korea
keeeung.kim@kaist.edu

## Invited Talks

Title: STRUCTURED OUTPUT LEARNING WITH A MARGIN

Presenter: John Shawe-Taylor (University College London, UK)

Abstract:

Structured output learning has been developed to borrow strength across multidimensional classifications. There have been approaches to bounding the performance of these classifiers based on different measures such as microlabel errors with a fixed simple output structure. We present a different approach and analysis starting from the assumption that there is a margin attainable in some unknown or fully connected output structure.

The analysis and algorithms flow from this assumption but in a way that the associated inference becomes tractable while the bounds match those attained were we to use the full structure. There are two variants depending on how the margin is estimated. Experimental results show the relative strengths of these variants, both algorithmically and statistically.

Bio:

John Shawe-Taylor is a professor at UCL where he directs the Centre for Computational Statistics and Machine Learning and heads the Department of Computer Science. His research has contributed to a number of fields ranging from graph theory through cryptography to statistical learning theory and its applications. However, his main contributions have been in the development of the analysis and subsequent algorithmic definition of principled machine learning algorithms founded in statistical learning theory. He has co-authored two influential text books on kernel methods and support vector machines. He has also been instrumental in coordinating a series of influential European Networks of Excellence culminating in the PASCAL networks.

Title: BIG DATA LEARNING FOR INTERDISCIPLINARY APPLICATIONS: IN-DEPTH VIEW OF SOME KEY CHALLENGES

Presenter: Vincent S. Tseng (National Chiao Tung University, Hsinchu, Taiwan)

Abstract:

Nowadays, large volume of data is being collected at unprecedented and explosive scale in a broad range of application areas. Analytics on such big data deliver amazing value and can drive interdisciplinary applications in various aspects of our life, including healthcare, retail, financial services, mobile services, life sciences, etc. Decisions that previously were based on hypothetical models or just unreliable guesswork can now be made effectively and efficiently by learning from the big data itself. New wave of revolutions in various domains has jumped into this Big Data era with new opportunities and challenges arisen. In this

talk, I will investigate some key challenges in Big Data Learning for interdisciplinary applications through in-depth observations from various aspects covering data preprocessing, key feature discovery, learning and modeling, post-processing, etc. Experiences from practical projects in different domains including biomedicine, social media, e-commerce, mobile sensing, etc., will be shared. Finally, some emerging research topics and potential opportunities underlying this topic will also be addressed accordingly.

Bio:

Dr. Vincent S. Tseng is currently a Distinguished Professor at Department of Computer Science and Director of Center for Big Data Technologies and Applications in National Chiao Tung University Taiwan, R.O.C.. He received the PhD in Computer Science from National Chaio Tung University in 1997 and then joined EECS Computer Science Division of UC Berkeley as a research fellow during 1998-1999. He was the Chair for IEEE Computational Intelligence Society Tainan Chapter during 2013-2015 and the President of Taiwanese Association for Artificial Intelligence during 2011-2012. Dr. Tseng has a wide variety of research interests covering data mining, machine learning, biomedical informatics, mobile and Web technologies. He has published more than 300 research papers in referred journals and conferences as well as 15 patents (held and filed). He has been on the editorial board of a number of top-tier journals including IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Knowledge Discovery from Data, IEEE Journal of Biomedical and Health Informatics, etc. He has also been overseeing the directions and architecture of big data technical platforms and interdisciplinary applications for governmental and industrial units in Taiwan. He is also the recipient of 2014 K. T. Li Breakthrough Award and 2015 Outstanding Research Award by Ministry of Science and Technology Taiwan.

## Title: MASSIVE ONLINE ANALYTICS FOR THE INTERNET OF THINGS (IoT)

Presenter: Albert Bifet (Telecom ParisTech, Universite Paris-Saclay, France)

Abstract:

Big Data and the Internet of Things (IoT) have the potential to fundamentally shift the way we interact with our surroundings. The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in stream mining. In this talk, I will present an overview of data stream mining, and I will introduce some popular open source tools for data stream mining.

Bio:

Albert Bifet is Associate Professor at Telecom ParisTech and Honorary Research Associate at the WEKA Machine Learning Group at University of Waikato. Previously he worked at Huawei Noah's Ark Lab in Hong Kong, Yahoo Labs in Barcelona, University of Waikato and UPC BarcelonaTech. He is the author of a book on Adaptive Stream Mining and Pattern Learning and Mining from Evolving Data Streams. He is one of the leaders of MOA and Apache SAMOA software environments for implementing algorithms and running experiments for online learning from evolving data streams. He is serving as Co-Chair of the Industrial track of IEEE MDM 2016, ECML PKDD 2015, and as Co-Chair of BigMine (2015, 2014, 2013, 2012), and ACM SAC Data Streams Track (2016, 2015, 2014, 2013, 2012)

## Title: TACKLING BIG-DATA AND BIG-MODEL CHALLENGES OF DEEP LEARNING

Presenter: Tie-Yan Liu (Microsoft Research, China)

Abstract:

The success of deep learning could be attributed to the availability of very big training data, the expressiveness of big deep models, and the computational power of GPU clusters. However, they are double-edged swords: it is costly or sometimes impossible to acquire sufficient labeled data for training; big models are usually hard to train and might exceed the capacity of GPU devices; it is non-trivial to distribute the training onto multiple nodes, with linear speed up and without accuracy loss. In this talk, I will introduce our recent research to address these challenges. First, I will introduce a technology called dual learning, which leverages the fact that many AI tasks have dual forms to create a closed feedback loop to enable the effective learning from unlabeled data. Second, we study the case that deep learning model is large due to its fat output layer (i.e., with many categories to predict), and propose to map the outputs onto a 2-dimensional table to effectively compress the model. By taking recurrent neural networks (RNN) as example, we show that our technology can lead to better accuracy and several-orders-of-magnitude smaller model. Third, we discuss the embarrassment of parallel computation  synchronous parallelization is slow due to synchronization barrier; asynchronous parallelization hurts accuracy due to communication delay. We then introduce a novel technology that leverages Taylor expansion of the gradient function to compensate the delay in asynchronous parallelization. It can achieve linear speed up and an accuracy comparable to sequential algorithms. All the technologies introduced in this talk will soon be open-sourced through Microsoft CNTK.

Bio:

Tie-Yan Liu is a principal researcher of Microsoft Research Asia. He is very well known for his pioneer work on learning to rank and computational advertising, and his recent research interests include deep learning and distributed machine learning. As a researcher

in an industrial lab, Tie-Yan is making his unique contributions to the world. On one hand, many of his technologies have been transferred to Microsofts products and online services. On the other hand, he has been actively contributing to academic communities. He is an adjunct/honorary professor of Carnegie Mellon University (CMU) and University of Nottingham. His papers have been cited for tens of thousands of times in refereed conferences and journals. He has won the best student paper award at SIGIR (2008), the most cited paper award at Journal of Visual Communications and Image Representation (2004-2006), the research breakthrough award at Microsoft Research (2012), and Top-10 Springer Computer Science books by Chinese authors (2015). He has been invited to serve as general chair, program committee chair, or area chair for a dozen of top conferences including SIGIR, WWW, KDD, NIPS, IJCAI, and AAAI, as well as associate editor/editorial board member of ACM Transactions on Information Systems, ACM Transactions on the Web, Information Retrieval Journal, and Foundations and Trends in Information Retrieval. Tie-Yan Liu is a senior member of IEEE and ACM.

## Tutorials

Title: Mass Estimation: Enabling density-based or distance-based algorithms to do what they cannot do

Presenter: Kai Ming Ting

Abstract:

This tutorial provides an overview of mass estimation, an alternative data modelling mechanism to density estimation; and details how it can overcome fundamental weaknesses of density-based or distance-based algorithms to enable them to do what they cannot do previously.

Mass estimation is attractive because the basic measure, mass, is not only more fundamental than density, but also more versatile—mass can be used to do density estimation, as a means for subspace selection and to find multi-dimensional median, and can be extended to measure dissimilarity of any two points. Example advantages of mass over density or distance are given as follows:

- DEMass–Density estimator based on mass–runs orders of magnitude faster than kernel and kNN density estimators
- Mass has been used, in place of density, as an effective means for subspace selection.
- Half-space mass is the maximally robust and efficient method to find multi-dimensional median. Existing methods such as data depth are either less robust or computationally more expensive.
- Simply replacing mass-based dissimilarity (a data dependent measure) with distance measure (a data independent measure) overcomes key weaknesses of density-based and distance-based methods in clustering, anomaly detection, information retrieval and classification.

This tutorial draws upon recent work on mass estimation and previous work which was also mass- based but was incorrectly categorised as density-based.

Title: Recent Advances in Distributed Machine Learning

Presenter: Taifeng Wang, Wei Chen

Abstract:

In recent years, artificial intelligence has demonstrated its power in many important applications. Besides the novel machine learning algorithms (e.g., deep neural networks), their distributed implementations play a very critical role in these successes. In this tutorial, we will first review popular machine learning models and their corresponding optimization

techniques. Second, we will introduce different ways of parallelizing machine learning algorithms, i.e., data parallelism, model parallelism, synchronous parallelism, asynchronous parallelism, and so on, and discuss their theoretical properties, advantages, and limitations. Third, we will discuss some recent research works that try to overcome the limitations of standard parallelization mechanisms, including advanced asynchronous parallelism and new communication and aggregation methods. Finally, we will introduce how to leverage popular distributed machine learning platforms, such as Spark MlLib, DMTK, Tensorflow, to parallelize a given machine learning algorithm, in order to give the audience some practical guidelines on this topic.

Title: Deep Approaches to Semantic Matching for Text

Presenter: Yanyan Lan, Jiafeng Guo

Abstract:

Semantic matching is critical in many text applications, including paraphrase identification, information retrieval, question answering, and machine translation. A variety of machine learning techniques have been developed for various semantic matching tasks, referred to as learning to match. Recently, deep learning approaches have shown their effectiveness in this area, and a number of methods have been proposed from different aspects of matching. In this tutorial, we will give a systematic and detailed survey on newly developed deep learning technologies for semantic matching. We will focus on the descriptions on the fundamental problems, as well as the novel solutions from bridging the word level semantic gap and conducting sentence level end-to-end semantic matching. We will also discuss the potential applications and future directions of semantic matching for text.

## Workshop

### Asian Workshop on Reinforcement Learning (AWRL 2016)

Organisers: Jianye Hao, Paul Weng, Yang Yu, Zongzhang Zhang

The first Asian Workshop on Reinforcement Learning (AWRL 2016) focuses on both theoretical models and algorithms of reinforcement learning (RL) and its practical applications. In the last few years, we have seen the growing interest in RL of researchers from different research areas and industries. We invite reinforcement learning researchers and practitioners to participate in this world-class gathering. We intend to make this an exciting event for researchers and practitioners in RL worldwide, not only for the presentation of top quality papers, but also as a forum for the discussion of open problems, future research directions and application domains of RL. AWRL 2016 will consist of keynote talks (TBA), contributed paper presentations, discussion sessions spread over a one-day period.

Reinforcement learning (RL) is an active field of research that deals with the problem of (single or multiple agents') sequential decision-making in unknown possibly partially observable domains, whose (potentially non-stationary) dynamics may be deterministic, stochastic or adversarial. RL's objective is to develop agents' capability of learning optimal policies in unknown environments (possibly in face of other coexisting agents) by trial-and-error and with limited supervision. Recent developments in exploration-exploitation, online learning, planning, and representation learning are making RL more and more appealing to real-world applications, with promising results in challenging domains such as recommendation systems, computer games, or robotics systems. We would like to create a forum to discuss interesting results both theoretically and empirically related with RL. The ultimate goal of this workshop is to bring together diverse viewpoints in the RL area in an attempt to consolidate the common ground, identify new research directions, and promote the rapid advance of RL research community.

### First New Zealand Text Mining Workshop

Organisers: Parma Nand, Rivindu Perera

In recent times, there has been an astronomical surge in demand for data scientists with Harvard Business Review naming Data Scientist as The Sexiest Job of the 21st Century.

The workshop will aim to foster collaboration among Data Science academics and practitioners focussing on text data. We have reached a point in Data Science where there is an increasing demand to integrate information from text data into models. This workshop calls for recent advances made both in the area of theoretical Text Processing dealing with lower level algorithms as well applications in Text Mining. The workshop aims to foster collaboration between academic researchers and practitioners so that the two groups could be able to integrate new advances in approaches into real world innovations being worked on by the practitioners.

## ACML Workshop on Learning on Big Data

Organisers: Ivor Wai-Hung Tsang, Ling Chen, Ying Zhang, Joey Tianyi Zhou

With the advance of data storage and Internet technology, data becomes more massive, noisier and more complex, which also brings good opportunities and challenges for machine learning. Learning technologies on Big Data have attracted many attentions. They have been successfully applied to many machine learning applications, including text mining, natural language processing, image categorization, video analysis, recommendation systems, sensor-based prediction problems, software engineering and so forth.

The aim of this workshop is to document recent process of Big Data technologies (e.g. Big Data Infrastructure, Distributed optimization, Stochastic optimization, MapReduce and Cloud Computing, etc.) in different real-world applications, to understand how computational bottlenecks trade-off with statistical efficiency for Big Data analysis tools, and also to stimulate discussion about potential challenges that may open new directions of learning on Big Data. We appreciate not only the manuscripts that dedicate to handle learning on Big Data, but also those which aim to discuss the approaches and/or theories for handling the new Big Data issues when exploiting massive data of different formats or structures.

## Conference Organisation

### Organising Committee

**General Co-Chairs:**
Geoff Holmes (University of Waikato, New Zealand) and Stephen Marsland (Massey University, New Zealand)

**Program Co-Chairs:**
Bob Durrant (University of Waikato, New Zealand) and Kee-Eung Kim (KAIST, South Korea)

**Workshop Co-Chairs:**
Yang Yu (Nanjing University, China) and Russel Pears (Auckland University of Technology, New Zealand)

**Tutorial Co-Chairs:**
Ichiro Takeuchi (Nagoya Institute of Technology, Japan) and Bernhard Pfahringer (University of Waikato, New Zealand)

**Local Arrangement Co-Chairs:**
Lyn Hunt (University of Waikato, New Zealand), Peter Reutemann (University of Waikato, New Zealand) and Hannah Te Puia (University of Waikato, New Zealand)

**Publication Co-Chairs:**
Stephen Marsland (Massey University, New Zealand)

**Steering Committee**

- Wray Buntine (Monash University, Australia)
- Tom Dietterich (Oregon State University, USA)
- Tu Bao Ho (Japan Advanced Institute of Science and Technology, Japan)
- Wee Sun Lee (National University of Singapore, Singapore)
- Chih-Jen Lin (National Taiwan University, Taiwan)
- Masashi Sugiyama (Co-Chair, RIKEN/The University of Tokyo, Japan)
- Zhi-Hua Zhou (Chair, Nanjing University, China)

**Senior Program Committee**

- Alice Oh
- Bernhard Pfahringer
- Bohyung Han
- Steven Hoi
- Dat Tran
- Dinh Phung
- Eibe Frank
- Hang Li
- Hoai An Le Thi
- Hsuan-Tien Lin
- Hwanjo Yu
- James Kwok
- Junmo Kim
- Junping Zhang
- Jun Xu
- Hisashi Kashima
- Kee-Eung Kim
- Irwin King
- Kristian Kersting

- Zhengdong Lu
- Shoude Lin
- Shirish Shevade
- Stephen Gould
- Masashi Sugiyama
- Tao Qin
- Truyen Tran
- Liwei Wang
- Wray Buntine
- Eunho Yang
- Yuh-Jye Lee
- Min-Ling Zhang

## Sponsors

We would like to thank 11ants Analytics, Parrot analytics Ltd, University of Waikato and Asian Office of Aerospace Research and Development (AOARD).