Preface

The First International Workshop on High Dimensional Data Mining (HDM’2013)

Some 13 years ago, Stanford statistician David Donoho predicted that the 21-st century will be the century of data.

“We can say with complete confidence that in the coming century, high-dimensional data analysis will be a very significant activity, and completely new methods of high-dimensional data analysis will be developed; we just don't know what they are yet.” -- D. Donoho, 2000.

Indeed, unprecedented technological advances have lead to increasingly high dimensional data sets in all areas of science, engineering and businesses. These include genomics and proteomics, biomedical imaging, signal processing, astrophysics, finance, web, and market basket analysis, among many others. The number of features in such data is often of the order of thousands or millions—that is much larger than the available sample size. The properties of such high dimensional spaces are fundamentally different from those of low dimensional spaces, and this renders many of the existing data analysis methods inadequate, questionable, or inefficient at best, calling for new approaches.

Some of the manifestations of this curse of dimensionality are the following:

- High dimensional geometry defeats our intuition rooted in low dimensional experiences so that data presentation and visualisation become particularly challenging.
- Distance concentration is the phenomenon of high dimensional probability spaces where the contrast between pairwise distances vanishes as the dimensionality increases—this may make distances meaningless, and may affect all methods that rely on a notion of distance.
- Bogus correlations and misleading estimates will result when trying to fit complex models for which the effective dimensionality is too large compared to the number of data points available.
- The accumulation of noise may confound our ability to find low dimensional intrinsic structure hidden in the high dimensional data.
- The computation cost of processing high dimensional data is often prohibiting.

The HDM’2013 workshop is the first to bring together researchers and practitioners from data mining and related disciplines with the aim to advance and discuss new research and new directions that address the curses, and uncover and exploit the blessings of high dimensionality in data mining.

The papers included in these proceedings were selected from 18 initial submissions received from 11 different countries (including the papers forwarded from the main conference). Each paper went through a rigorous reviewing process, and was assessed by 3 to 4 members of the Programme Committee. The selected papers cover a range of aspects of high dimensionality in data mining, from theoretical foundations, to algorithms and implementation, to applications and empirical studies.
We hope that this workshop will see future instances, and that it will continue to instigate leading edge research that will bring to life the theory, the methods, and the applications that taken together will enable us to cope in an optimal way with the issue of dimensionality in data mining.

We would like to thank the authors of all the submitted papers for their valuable contributions, and to all members of the expert Programme Committee for their dedication and thoughtful reviews that ultimately resulted in a strong technical program. Special thanks are due to our three invited speakers: Dr. Bob Durrant (University of Waikato, NZ), Dr. Jo Etzel (University of Washington), and Dr. Stephan Günnemann (Carnegie Mellon University). Finally, we would like to thank the IEEE ICDM conference organisers for hosting this Workshop.

We look forward to welcoming you all at HDM’2013 on 7-th December 2013 in Dallas.

Ata Kaban
HDM’2013 Program Chair

Programme Committee
Adam Kowalczyk, Victoria Research Laboratory, NICTA, Australia
Arthur Zimek, LMU Munich, Germany
Barbara Hammer, Clausthal University of Technology, Germany
Ata Kaban, University of Birmingham, UK
John A. Lee, Universite Catholique de Louvain, Belgium
Laurens van der Maaten, Delft University of Technology, The Netherlands
Mark Last, University of the Negev, Israel
Milos Radovanovic, University of Novi Sad, Serbia
Pierre Alquier, University College Dublin, Ireland
Robert J. Durrant, University of Waikato, NZ
Stephan Gunnemann, Carnegie Mellon University
Yiming Ying, University of Exeter, UK