Project Title: Classification using Mass Estimation

Project Number: IMURA0231

Monash Supervisor(s): A/Prof. Kai Ming Ting

Monash Primary Contact: KaiMing.Ting@monash.edu
+61 3 990 26241

IITB Supervisor(s): Prof. Ganesh Ramakrishnan

IITB Primary Contact: ganesh@cse.iitb.ac.in
022 2576 7728

Research Academy Themes:

Highlight which of the Academy’s Theme(s) this project will address?
(Feel free to nominate more than one. For more information, see www.iitbmonash.org)

1. Advanced computational engineering, simulation and manufacture
2. Infrastructure Engineering
3. Clean Energy
4. Water
5. Nanotechnology
6. Biotechnology and Stem Cell Research

The Research Problem

Existing data mining methods are largely based on density estimation at their core. The problems associated with the use of density estimation are:
* limited to tasks which have low number of dimensions,
* cannot deal with large data sets, and
* high time and space complexities.

An alternative called mass estimation has attractive properties: it has constant time-and-space-complexities and has been shown to perform at least as good as and often better than a total of eight state-of-the-art methods in terms of task-specific performance measures in three tasks: information retrieval, regression and anomaly detection.

Mass estimation has the potential to be applied to a wide variety of applications, as density estimation has now, without the above-mentioned problems.

The first paper on mass estimation is published in proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2010 (go to www.gscit.monash.edu.au/~kmting to retrieve the SIGKDD paper.) The superiority of mass estimation has been established for anomaly detection and content-based image retrieval. See the SIGKDD paper for details. This project is part of a larger project investigating mass estimation’s applicability in all areas of data mining, including clustering, image mining, network mining, classification, and data stream.
### Project aims
This project specifically aims to develop the first mass-based classification method.

### Expected outcomes
The expected outcomes of the project are to:

* Establish mass estimation as a new base modelling mechanism for classification.
* Determine the superiority of mass estimation over density estimation and other mechanisms in terms of efficiency and efficacy for classification.

### Capabilities and Degrees Required
List the ideal set (up to 8) of background and capabilities required in a student for this project noting that the more specific you make it, the less likely that you will get a candidate that matches the requirements exactly.

- Proficiency in programming in C, Java, or Matlab.
- Have background in algorithmic methodology in data mining or machine learning (but not a must)