University of Portsmouth

Pocket Data Mining
The Next Generation in Predictive Analytics

Dr Mohamed Medhat Gaber
School of Computing
University of Portsmouth
E-mail: mohamed.gaber@port.ac.uk

Research Timeline

- 2003 – 2006
  - Adaptive resource-aware data stream mining approach and techniques
  - Algorithm Granularity (AG)
    - Algorithm Output Granularity (AOG)
    - Algorithm Input Granularity (AIG)
    - Algorithm Processing Granularity (APG)
- 2007 – 2010
  - Situation-aware data stream mining
  - Fuzzy Situation Inference (FSI)
- 2008 – 2011
  - Clutter-aware visualisation
  - Adaptive Clutter Reduction (ACR)
- 2010 – 2012
  - Distributed and collaborative mobile data stream mining
  - **Pocket Data Mining (PDM)**

Agenda

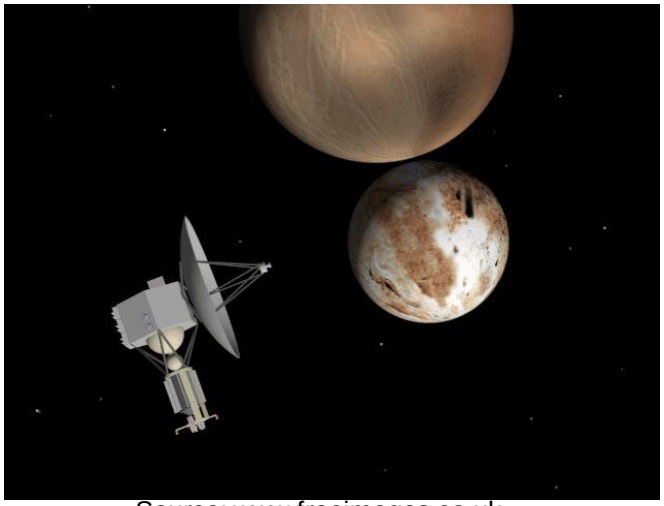
- Introduction to Data Streams
- Earlier work
  - Granularity-based Approach
  - Situation-aware Data Stream Mining
  - Clutter-aware Visualisation
- Pocket Data Mining
  - Background on Mobile Software Agents
  - PDM Architecture and Procedure
  - Hoeffding Tree Agent Miner
  - Naïve Bayes Agent Miner
  - Experimental Results
- Summary

Introduction to Data Streams

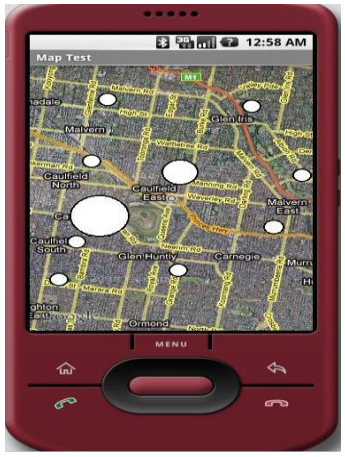
- The advances in data acquisition hardware and the emergence of applications that process continuous flow of data records have led to the data stream phenomenon.
- A data stream is *a continuous, rapid flow of data that challenge our state-of-the-art processing and communication infrastructure.*
- The general features of data streams are:
  - Very high rate input data
  - Read only once by an algorithm
  - Real time processing demand
  - Unbounded
  - Time varying.

Data Stream Processing in Resource-constrained Environments

- A wide range of data streams are generated in or sent to resource-constrained computing environments.
  - Spacecrafts
  - Wireless sensor networks
  - *PDAs* and smartphones



Source: www.freeimages.co.uk



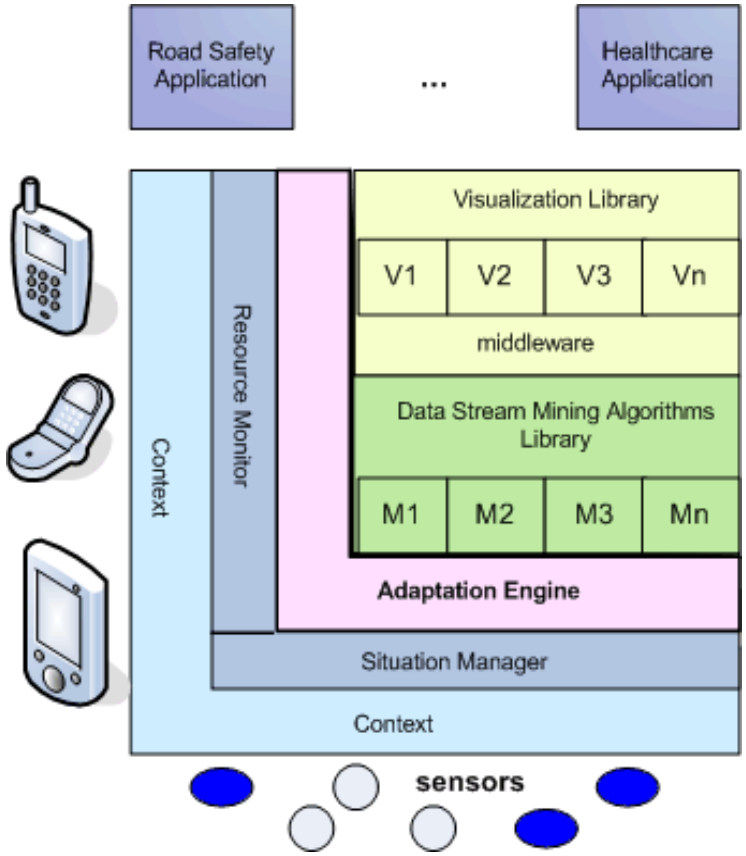
Gregory Piatetsky @kdnuggets
Data gravity principle - the bigger the data, the harder it is to move it, so logic need to come to big data #briefr

Research Issues

- Limited computational resources
- Limited bandwidth
- Limited screen realestate
- Change of the user's context

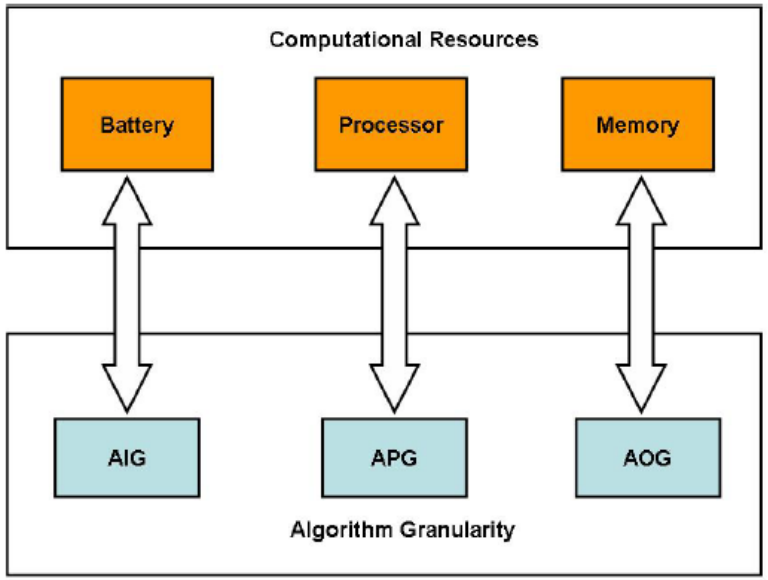
Our Approach

- Adaptability with regard to:
  - Computational resources
  - User's situation
  - Visual clutter

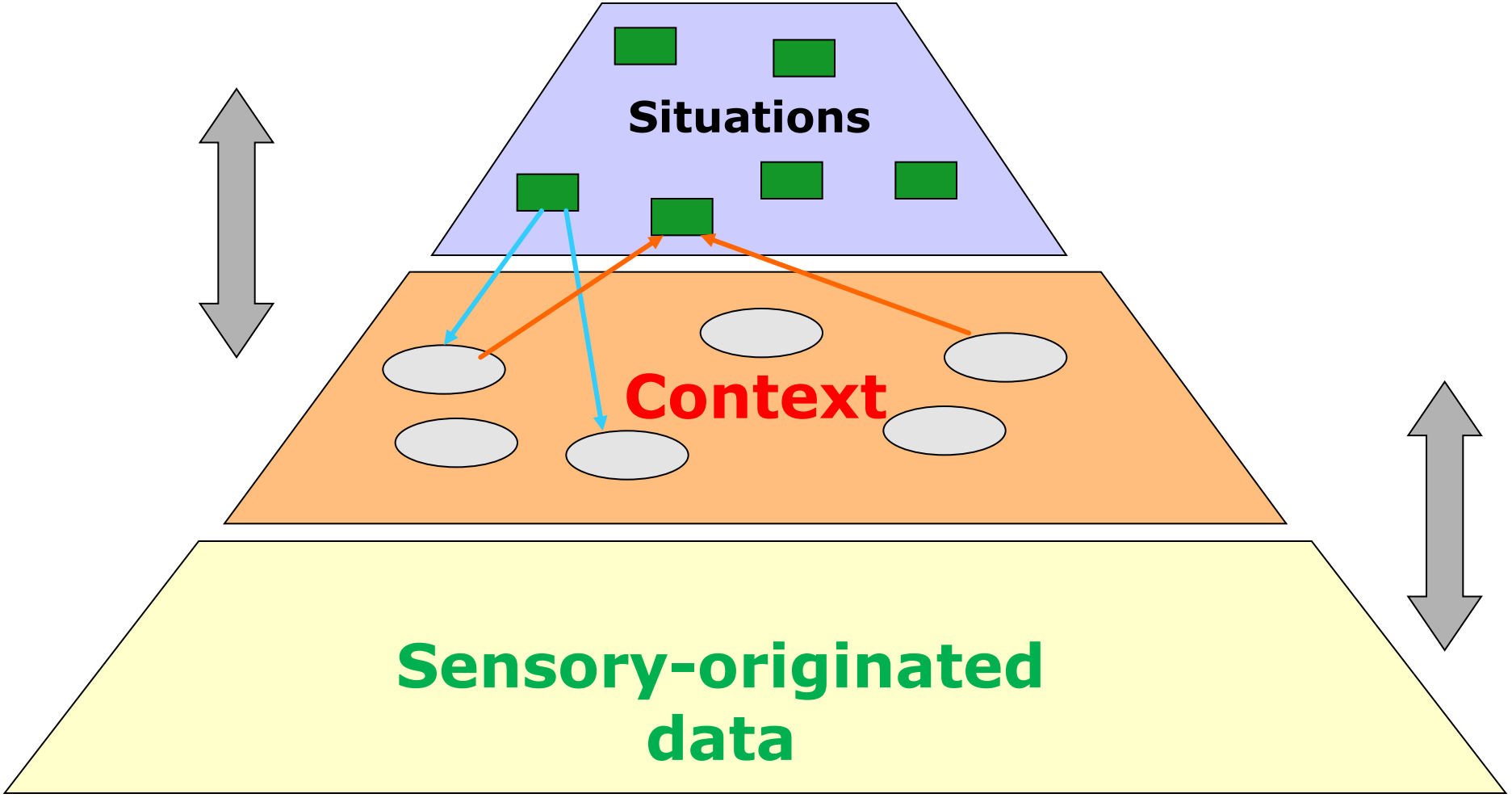


Granularity-based Approach

- Combining the three possible granularity-based adaptation, namely:
  - AIG: Algorithm Input Granularity
  - AOG: Algorithm Output Granularity
  - APG: Algorithm Processing Granularity

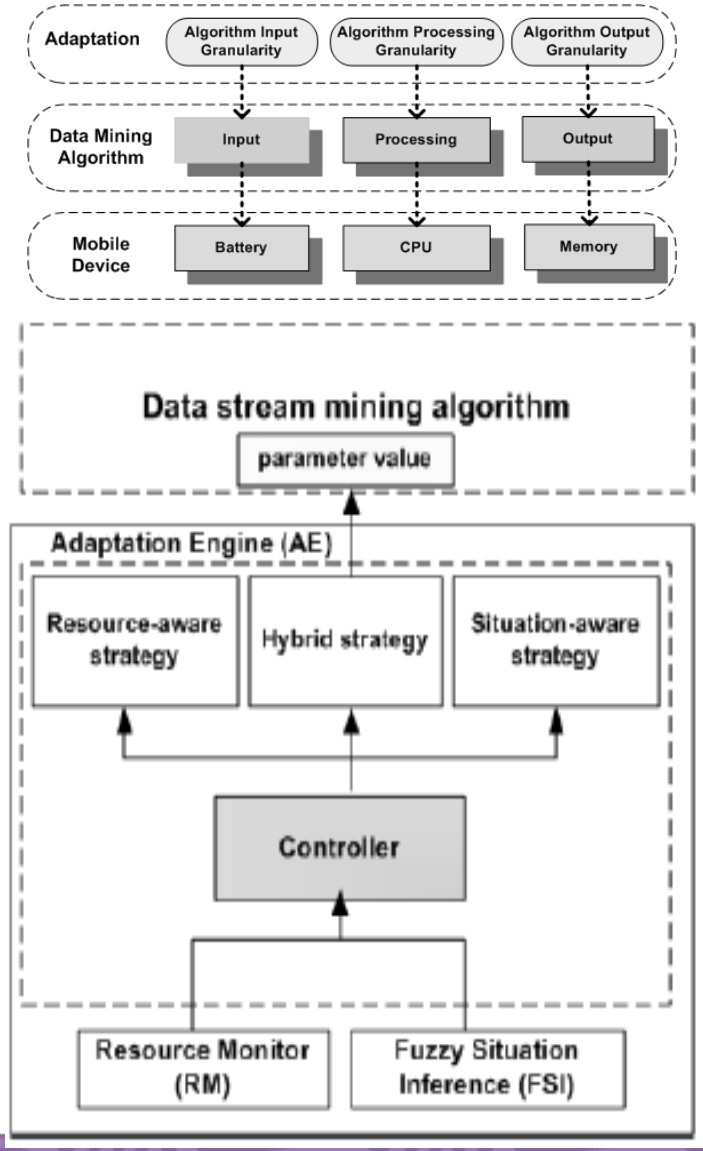


Situation-Aware Adaptation



Situation Inferencing

- Capture Application's "Situation"
- Fuzzy Context Spaces
- Enhance probabilistic situation inferencing with fuzziness
- Cope with changing situations
- Cope with unknown situations



Adaptive Clutter Reduction

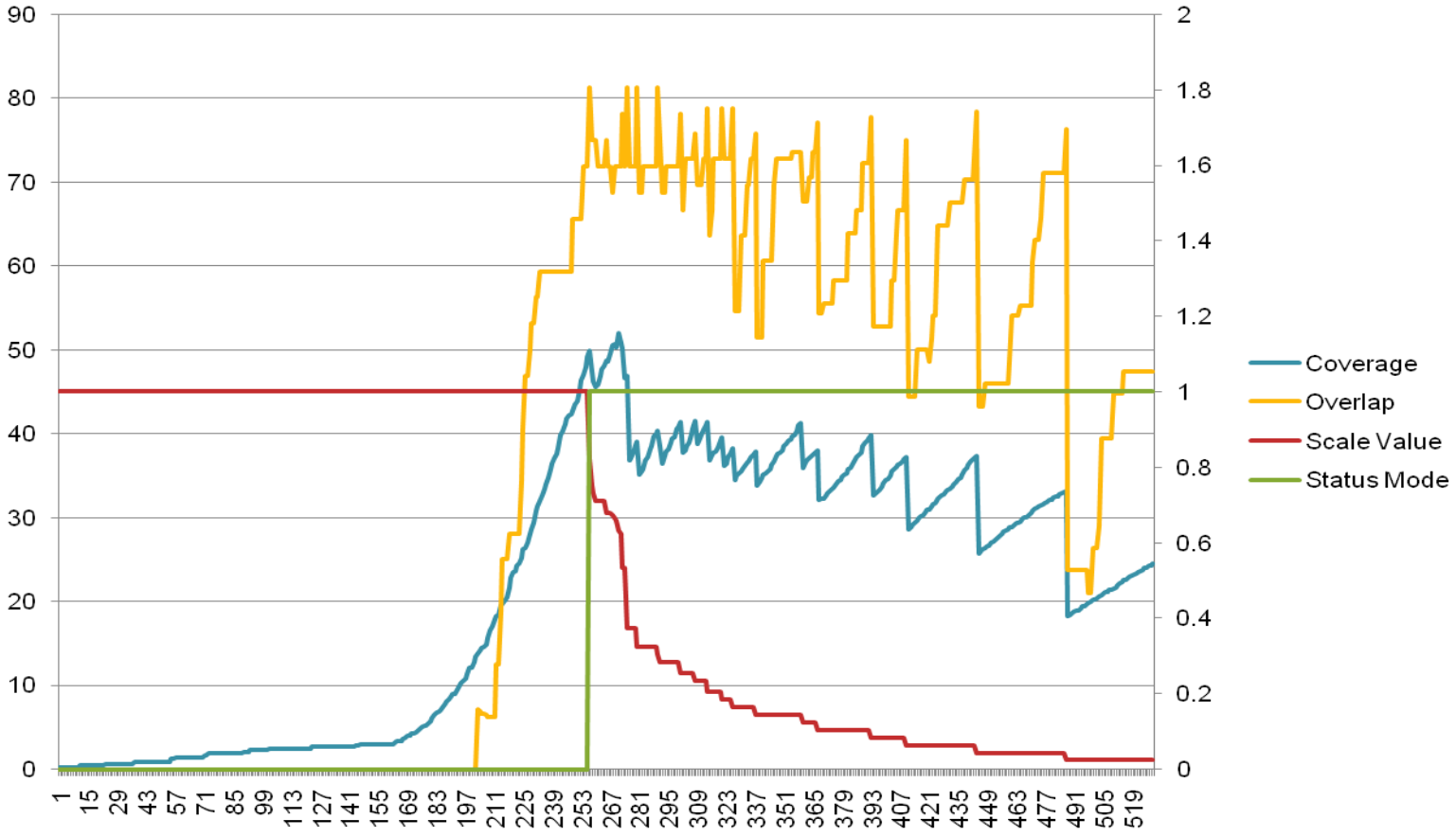
- Similar to resource-awareness and situation-awareness, we have developed a novel way to automatically reduce the clutter
- The new approach has many important applications (especially in disaster management)

Corollary 1. *Let $i(v)$ be the informative level of the visualizer v, and $c(v)$ be the level of clutter on the screen. It is established that $i(v)\ \alpha\ c(v)$.*

Corollary 2. *Let $p(v)$ be the level of perception of the visualised results. It is established that $p(v)\ \alpha\ \frac{1}{c(v)}$.*

Theorem *At any point in time t, ACR based technique minimize$(c(v))$ while maximize$(i(v), p(v))$*

Adaptive Clutter Reduction



50% Coverage and 80% Overlap

"Who is rich? He that is content. Who is that? Nobody" - Benjamin Franklin

TechEYE.net

Smartphones show disasters as they happen

For earthquakes, riots and heart monitoring

06 Jun 2011 10:50 | by Andrea Petrou | Filed in Science Smartphone

Computer Science geeks at the University of Portsmouth have found a way of making smartphones show a disaster unfolding in real-time on phone screens.

They've developed an application and prototype, which currently allows a range of different uses such as allowing docs to monitor heart patients' ECG right through to helping coppers in the central control unit to see where each PC plod is in an emergency.

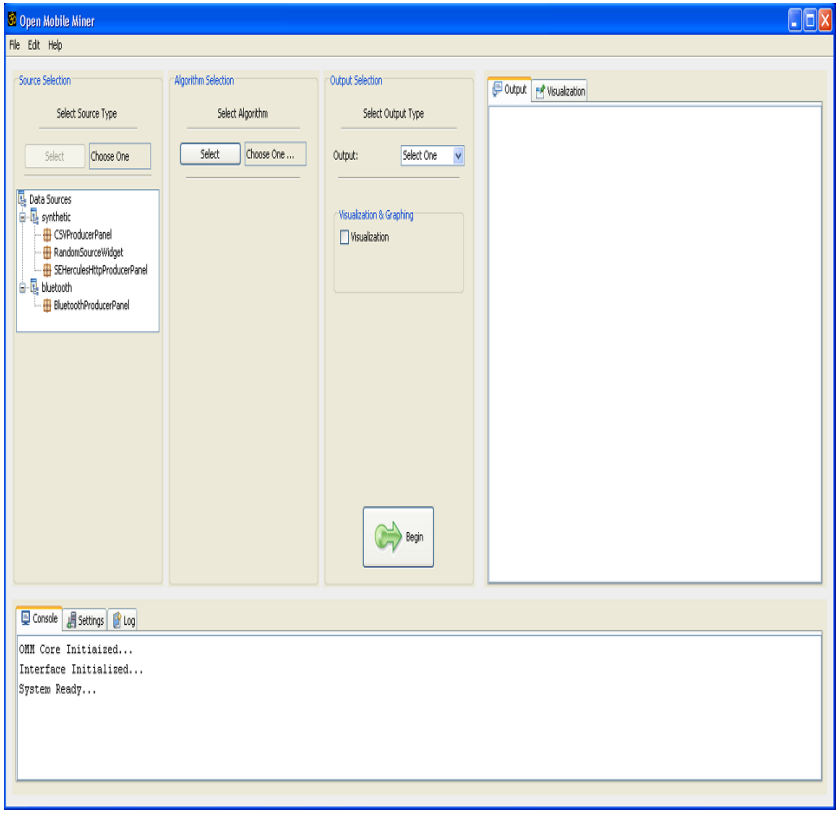
Dr Mohamed Gaber, of the University of Portsmouth's School of Computing, and geeks from Monash University have also said that the app can also be used in a natural disaster. One example given was to help those co-ordinating rescue efforts to use an electronic map on their phone screens with clusters showing which areas are worst affected.

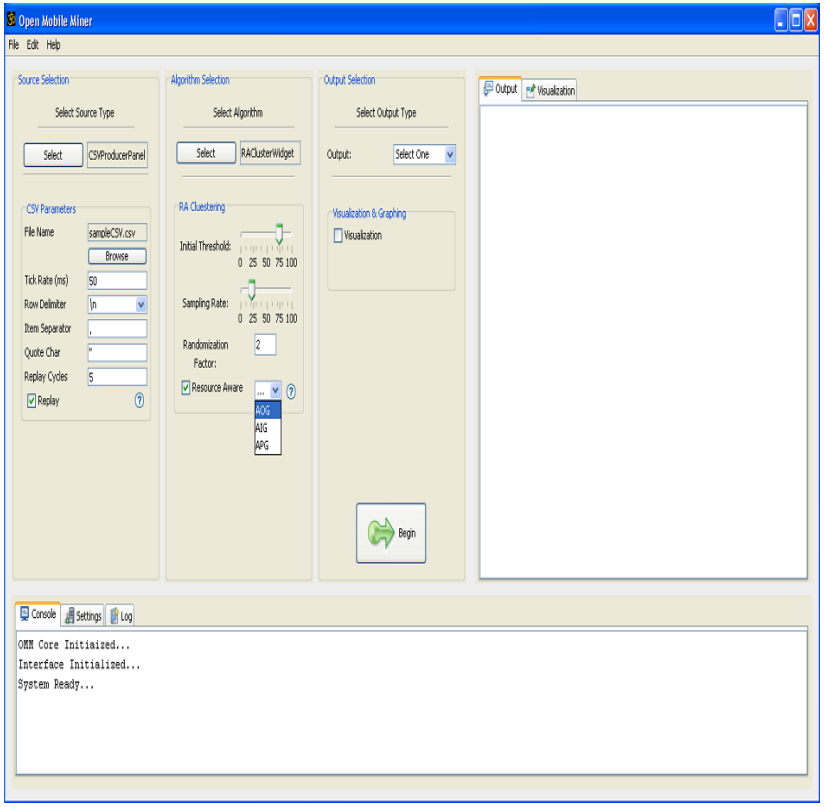
They said that because such information would constantly update as the disaster unfolds, the clusters would adjust automatically in size and scale as new clusters formed to stop the phone screen becoming over-crowded with information.

And the researchers are blowing their own trumpets claiming that this is the first time anyone has managed to develop a "clutter-aware visualisation for mobile data mining that automatically considers the amount of information presented on screen and dynamically adjusts the way this information is presented to avoid confusion and enhance ease of understanding."

They added that a lot of work had gone into making the application usable and interactive.
"The need for an application that knows when information overload is a threat is very important," they added.

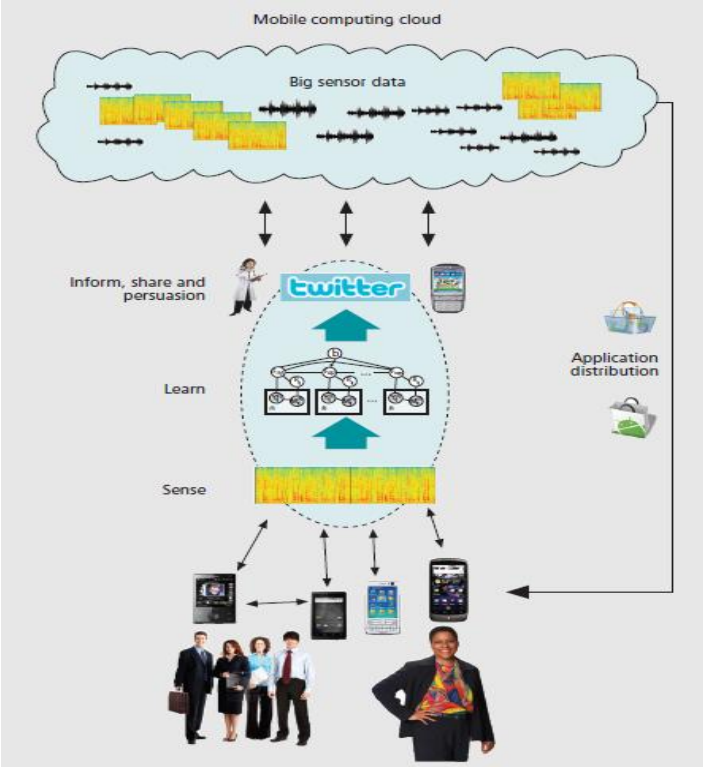
Open Mobile Miner - OMM





PDM: Pocket Data Mining

- **Pocket Data Mining (PDM)** is our new term describing collaborative mining of streaming data in mobile and distributed computing environments.
- With continuous advances in computational power and communication abilities for smartphones and tablet computers; and
- The sheer amounts of data streams that we subscribe to or acquire using the onboard sensing capabilities
- There is an unprecedented opportunity to perform complex data analysis tasks that can benefit mobile users



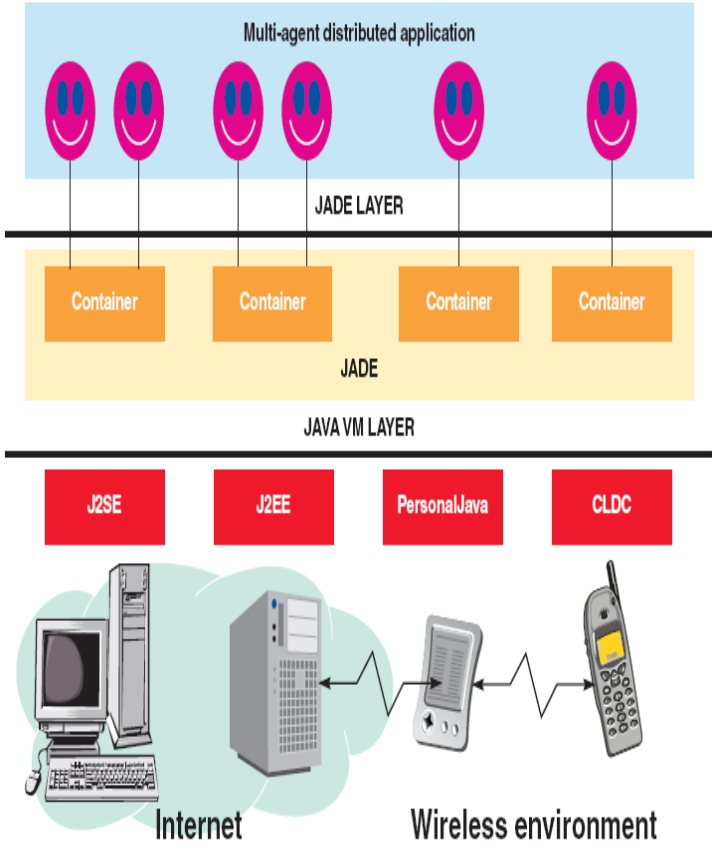
Source: Lane, N.D.; Miluzzo, E.; Hong Lu; Peebles, D.; Choudhury, T.; Campbell, A.T.; , "A survey of mobile phone sensing," Communications Magazine, IEEE , vol.48, no.9, pp.140-150, Sept. 2010

Technology Enablers

- This can be realised with the help of several established areas of study including:
  - data stream mining;
  - mobile software agents; and
  - programming for small devices.

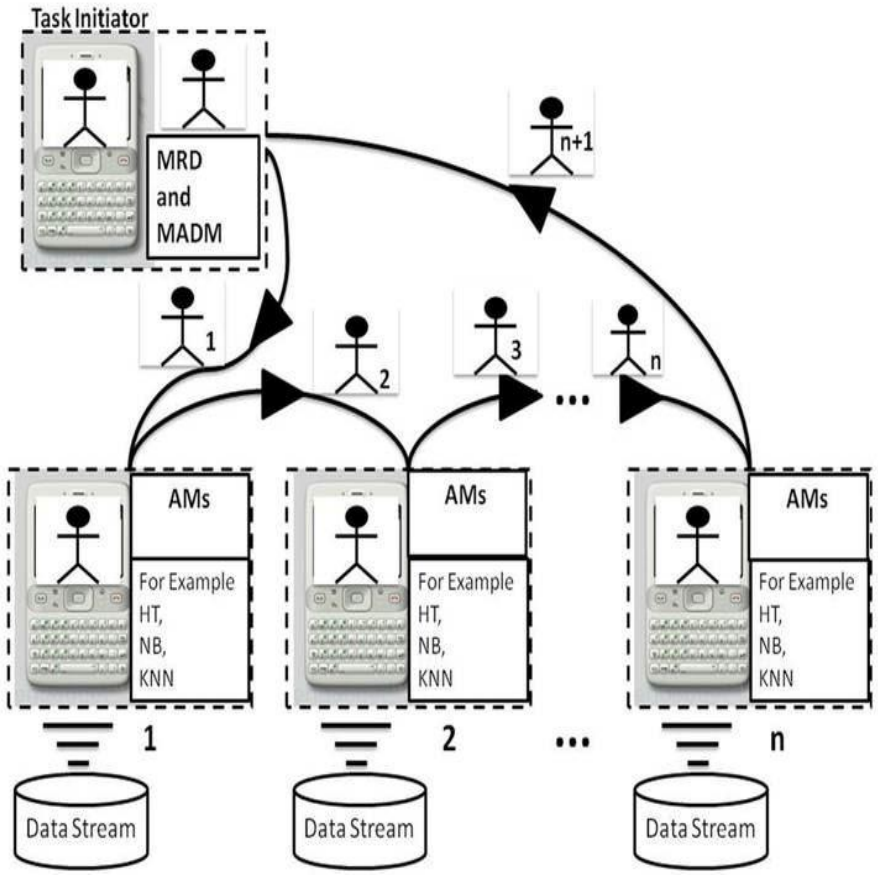
What is a Mobile Agent ?

- A software program
- Moves from machine to machine under its own control
- Suspends execution at any point in time, transport itself to a new machine and resume execution
- Once created, a mobile agent autonomously decides which locations to visit and what instructions to perform
- Continuous interaction with the agent's originating source is not required
- HOW?
  - Implicitly specified through the agent code
  - Specified through a run-time modifiable itinerary



JADE Architecture

PDM Architecture and Procedure



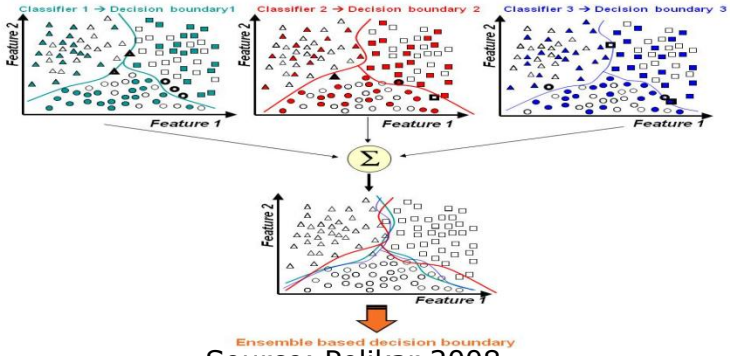
Algorithm 1 PDM's collaborative data mining workflow

Task Initiator: Form an ad hoc network of mobile phones;
Task Initiator: start *MRD* agent;
MRD: Discover data sources, computational resources and techniques;
MRD: Decide on the best combination of techniques to perform the task;
MRD: Decide on the choice of stationary AMs and deploy mobile *AMs*;
Task Initiator: start *MADM* agent with schedule provided by the *MRD*;
for $i = 1$ to $i =$ *number of AMs* do
  repeat
    AM_i: mine streaming data;
  until Use of the model by *MADM*
end for

PDM Agents

- (Mobile) agent miners (AM): these agents are either distributed over the network when the mining task is initiated or are already located on the mobile device.
  - Mobile data stream mining
- Mobile agent resource discoverers (MRD): these agents are used to explore the available computational resources, processing techniques, and data sources.
  - Mobile cloud
- Mobile agent decision makers (MADM): these agents roam the network consulting the mobile agent miners to collaborate in reaching the final decision.
  - Ensemble learning

Source: http://www.datacenterknowledge.com



Source: Polikar 2008

Agent Miners (AMs)

- We have used two stream classifiers, namely:
  - Hoeffding trees
    - Known for its statistically guaranteed accuracy
  - Incremental Naïve Bayes
    - Known for its computational efficiency and simplicity

Simple Weighted Majority Voting of the MADM

Y = 1.75 (0.55+0.65+0.55)

X = 1 .80 (0.95+0.85)

AM	Weight (Accuracy)	Class
1	0.55	Y
2	0.65	Y
3	0.55	Y
4	0.95	X
5	0.85	X

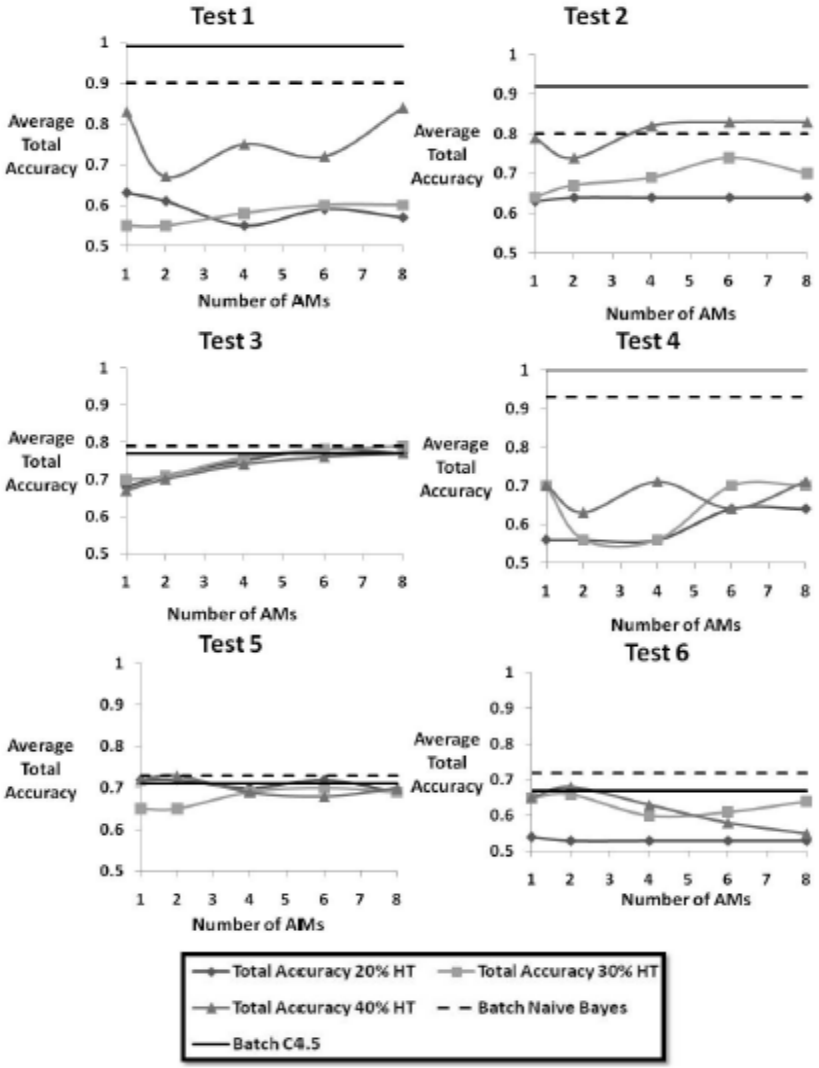
Experimental Study

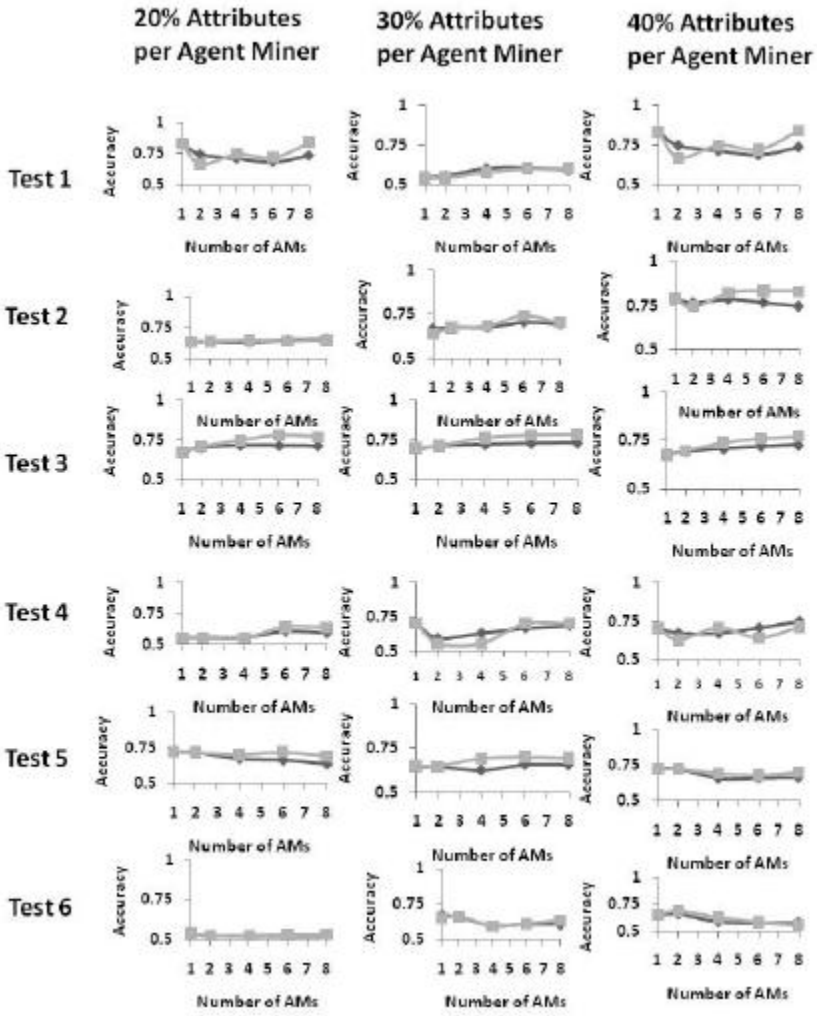
- Datasets

Test Number	Dataset	Number of Attributes	Number of Instances
1	kn-vs-kr	36	1988
2	spambase	57	1999
3	waveform-500	40	1998
4	mushroom	22	1978
5	infobiotics 1	20	$\approx$ 200000
6	infobiotics 2	30	$\approx$ 200000

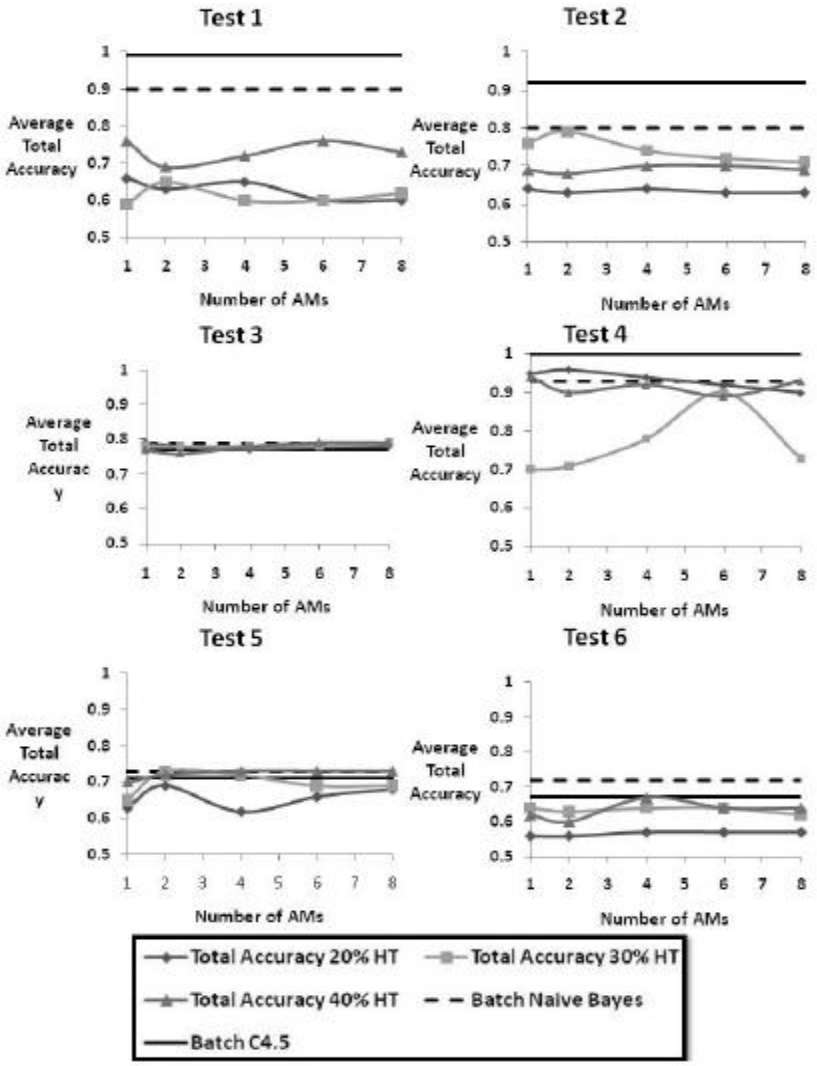
- Each AM has access to 20%, 30%, or 40% of the features (random vertical partitioning).

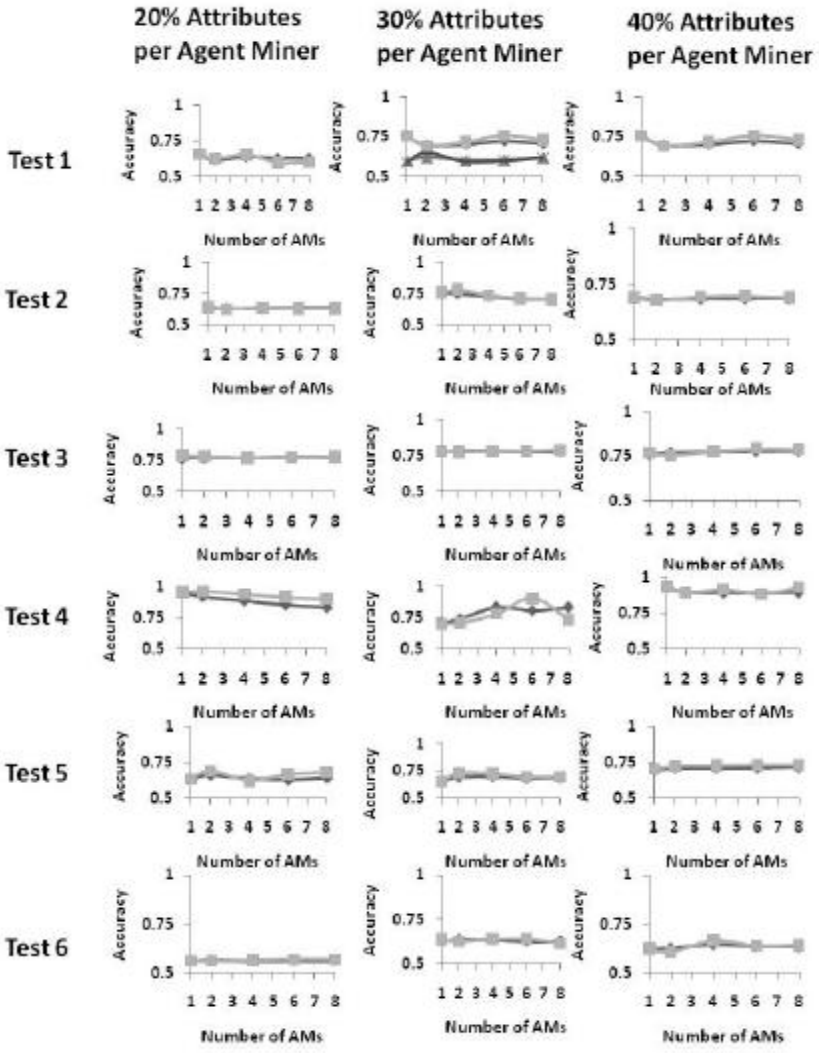
PDM with Hoeffding Trees



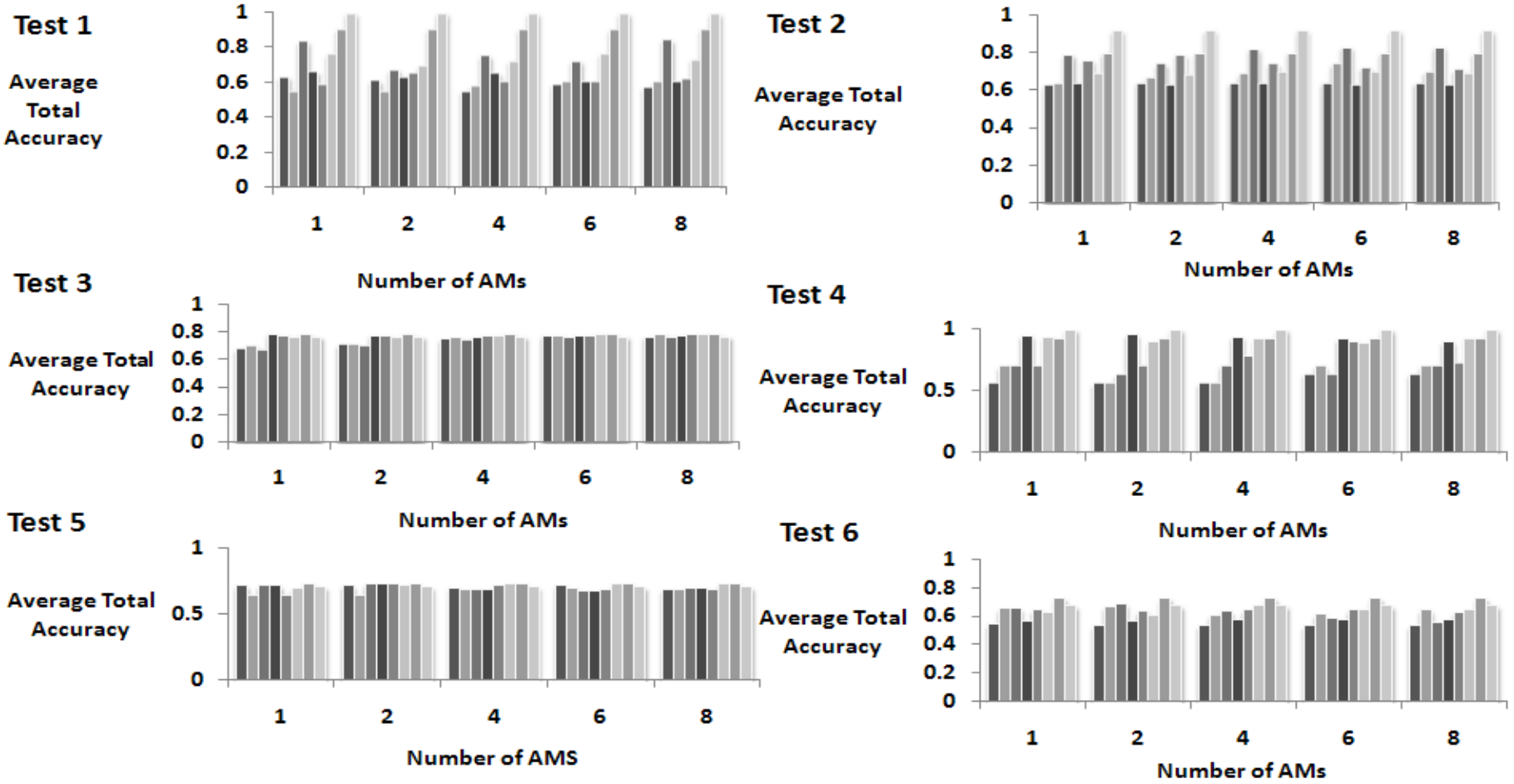


PDM with Naive Bayes



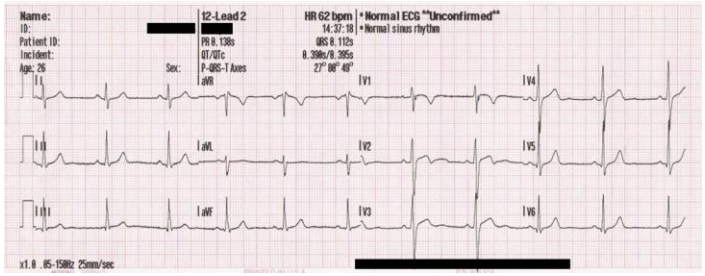


PDM with a Heterogeneous Setup



PDM Potential Applications

- Mobile ECG analysis



- Mobile social media analysis

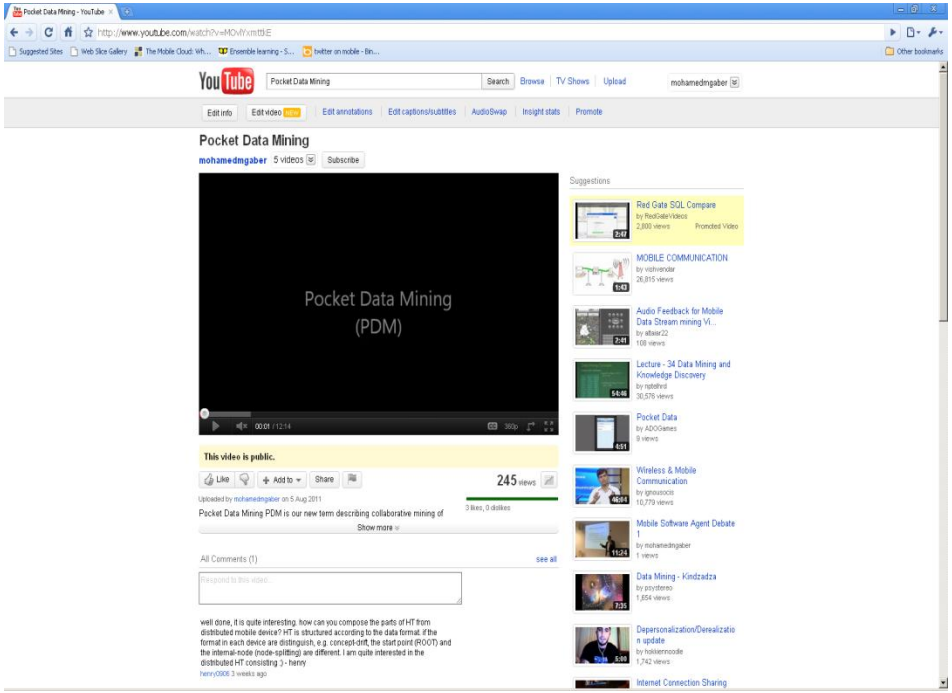
- Mobile policing

Source: YouTube videos

PDM Demonstration

YouTube video link:

http://www.youtube.com/watch?v=MOvlYxmttkE



Making News


The News
Phones 'can help solve crimes'
TMCnews
[January 18, 2011]
Research at the University
Smart use of mobile phone power
Tue, Jan 18, 2011
For the first time smart phones and tablet PCs could be used by police officers to help solve crimes faster and more cheaply.
Dr Mohamed Gaber
Latest News > 23 Jan 11 - Smartphones 'could revolutionise crime scene investigation'
Recent advances in smartphone technology could improve police crime-scene investigations, said Dr Mohamed Gaber of Portsmouth University.
Presenting at a Artificial Intelligence conference in Arras, France, he said that the meaty processing power in smartphones, coupled with its networking abilities and high-spec cameras could allow investigators to gain faster, greater insight into crime-scenes.
Phones could source fingerprints and other data locally, and communicate with other phones, without the need to sent anything back to a central computer for processing.
POLICE LINE DO NOT CROSS
COMMUNICATIONS OF THE ACM
Smartphones to fight crime
Call the iPolice: iPads and smartphones could help fight crime

Summary

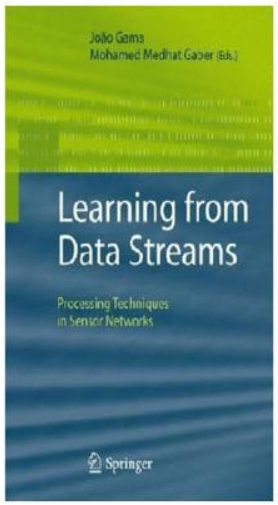
- Pocket data mining has been the outcome of earlier developments started in 2003.
- PDM is a mobile agent based framework for distributed and mobile ad-hoc data stream mining.
- PDM has proven its applicability experimentally with Hoeffding trees and Naïve Bayes classifiers.
- Many potential applications can benefit from PDM.

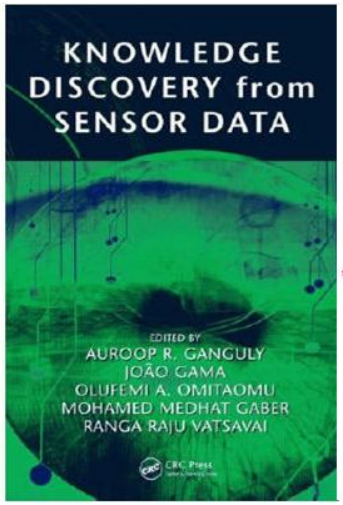
Main References

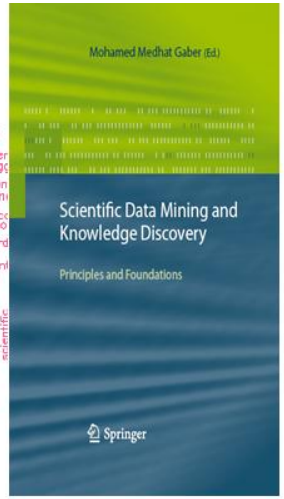
- Gaber M. M., **Data Stream Mining Using Granularity-based Approach**, a book chapter in Foundations of Computational Intelligence – Volume 6, (Eds.) Abraham A., Hassanien A., Carvalho A., and Snase V., Volume 206/2009, pp. 47-66, ISSN 1860-949X (Print) 1860-9503 (Online), ISBN 978-3-642-01090-3, Springer Berlin/Heidelberg, Germany, 2009.
- Gaber M. M., and Yu P. S., **A Holistic Approach for Resource-aware Adaptive Data Stream Mining**, Journal of New Generation Computing, ISSN 0288-3635 (Print) 1882-7055 (Online), Volume 25, Number 1, November, 2006, pp. 95-115, Ohmsha, Ltd., and Springer Verlag.
- Haghighi P. D., Krishnaswamy S., Zaslavsky A., Gaber M. M., **Reasoning About Context in Uncertain Pervasive Computing Environments**, Daniel Roggen, Clemens Lombriser, Gerhard Tröster, GerdKortuem, Paul J. M. Havinga (Eds.): Smart Sensing and Context, Third European Conference, EuroSSC 2008, pp. 112-125, Zurich, Switzerland, October 29-31, 2008. Proceedings. Lecture Notes in Computer Science 5279 Springer 2008, ISBN 978-3-540-88792-8.
- Gaber M. M., Krishnaswamy S., Gillick B., Nicoloudis N., Liono J., AlTaiar H., Zaslavsky A., **Adaptive Clutter-Aware Visualization for Mobile Data Stream Mining**, Proceedings of the IEEE 22nd International Conference on Tools with Artificial Intelligence (ICTAI 2010), pp. 304-311,Arras, France, 27-29 October, 2010.
- Stahl F., Gaber M. M., Aldridge P., May D., Liu H., Bramer M., and Yu P. S**, *Homogeneous and Heterogeneous Distributed Classification for Pocket Data Mining***, LNCS Transactions on Large-Scale Data- and Knowledge-Centered Systems, Springer-Verlag, 2011.

More available at: http://gaberm.myweb.port.ac.uk/publications.htm

Our Books in the Area







Acknowledgement

- I would like to thank the following:
  - Dr. Arkady Zaslavsky
  - A/Prof.. Shonali Krishnaswamy
  - Prof. Philip S. Yu
  - Dr. Bernhard Scholz
  - Dr. Uwe Roehm
  - Dr. Suan Khai Chong
  - Dr. Pari Delir Haghighi
  - Duc Nhan Phung
  - Iti Aggarwal
  - Brett Gillick
  - Osnat Horovitz
  - Rahul Shah
  - Dr Frederic Stahl
  - Prof. Max Bramer
  - Paul Aldridge
  - Han Liu
  - David May
  - Oscar Campos
  - Victor Mandujano

University of Portsmouth

Thanks for listening

Dr Mohamed Medhat Gaber
School of Computing
Faculty of Technology
University of Portsmouth
E-mail: mohamed.gaber@port.ac.uk, mohamed.m.gaber@gmail.com
Web: http://gaberm.myweb.port.ac.uk/