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
Research Issues

- Limited computational resources
- Limited bandwidth
- Limited screen real estate
- 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

Situations

Context

Sensory-originated data
**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 $i(v) \propto c(v)$.

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

**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
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

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
Algorithm 1 PDM’s collaborative data mining workflow

Task Initiation: Form an ad hoc network of mobile phones;
Task Initiation: 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 Initiation: start MADM agent with schedule provided by the MRD;
for $i = 1$ to $i = \text{number of AMs}$ 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
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**

\[ Y = 1.75 \ (0.55 + 0.65 + 0.55) \]

\[ X = 1.80 \ (0.95 + 0.85) \]

<table>
<thead>
<tr>
<th>AM</th>
<th>Weight (Accuracy)</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.55</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>0.65</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>0.55</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>0.85</td>
<td>X</td>
</tr>
</tbody>
</table>
Experimental Study

• Datasets

<table>
<thead>
<tr>
<th>Test Number</th>
<th>Dataset</th>
<th>Number of Attributes</th>
<th>Number of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kn-vs-kra</td>
<td>36</td>
<td>1988</td>
</tr>
<tr>
<td>2</td>
<td>spambase</td>
<td>57</td>
<td>1999</td>
</tr>
<tr>
<td>3</td>
<td>waveform-500</td>
<td>40</td>
<td>1998</td>
</tr>
<tr>
<td>4</td>
<td>mushroom</td>
<td>22</td>
<td>1978</td>
</tr>
<tr>
<td>5</td>
<td>infobiotics 1</td>
<td>20</td>
<td>≈ 200000</td>
</tr>
<tr>
<td>6</td>
<td>infobiotics 2</td>
<td>30</td>
<td>≈ 200000</td>
</tr>
</tbody>
</table>

• Each AM has access to 20%, 30%, or 40% of the features (random vertical partitioning).
PDM with Hoeffding Trees

Graphs showing the average total accuracy for different numbers of AMs in tests 1 to 6.
PDM with Naive Bayes
PDM with a Heterogeneous Setup

Test 1
Average Total Accuracy

Test 2
Average Total Accuracy

Test 3
Number of AMs
Average Total Accuracy

Test 4
Number of AMs
Average Total Accuracy

Test 5
Number of AMs
Average Total Accuracy

Test 6
Number of AMs
Average Total Accuracy

- 1st Bar: Total Accuracy 20% HT
- 2nd Bar: Total Accuracy 30% HT
- 3rd Bar: Total Accuracy 40% HT
- 4th Bar: Total Accuracy 20% NB
- 5th Bar: Total Accuracy 30% NB
- 6th Bar: Total Accuracy 40% NB
- 7th Bar: Batch Naive Bayes
- 8th Bar: Batch C4.5
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

Research at the University
Smart use of mobile phone power

Published: 28th January 2011

Dr Mohamed Gaber of the University of Portsmouth’s Centre for Computing and Information Sciences has discovered a way to use the power of mobile phones to help solve crime faster and more cheaply.

The research, now featured in the Journal of Telecommunications, involves using the power from mobile phones to perform calculations that can help solve crimes.

Dr Gabor said: “This is the first time a method has been found to enable crime investigations to be conducted using mobile phone technology.

Recent advances in smartphone technology could improve police crime scene investigations, said Dr Mohamed Gabor of Portsmouth University.

Presenting at a conference on artificial intelligence in France, he said that the increased processing power in smartphones, combined with their networking abilities and high spec cameras, could allow investigators to gain faster, greater insights into crime scenes.

Phones could source fingerprints and other data locally, and communicate with other phones, without the need to send anything back to a central computer for processing.

Latest News: 23 Jan 11 - Smartphones 'could revolutionise crime scene investigation'
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


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
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/