Clutter-Adaptive Visualisation for Mobile Data Mining

Brett Gillick, Hasnain AlTaiar, Jonathan Liono, Shonali Krishnaswamy, NicholasNicoloudis, Mohamed Gaber, Abhijat Sinha and Arkady Zaslavsky
Centre for Distributed Systems and Software Engineering
Monash University

Contact Author: Shonali.Krishnaswamy@monash.edu

Abstract—There is an emerging focus on real-time data stream analysis on mobile devices. While many mobile data stream mining algorithms have been developed in recent times, generic and scalable visualization techniques have not been presented. This paper presents the demonstration of our innovative clutter-adaptive cluster visualization technique for mobile devices. We have fully implemented this technique on the Google Android platform and provide demonstrations for different datasets: location (both real and synthetic), and stock-market (real). The demonstration videos can be accessed from the links provided in Section III of this paper and on display will allow for user interaction with the phones to change various thresholds/settings as discussed in the paper.

Keywords: Mobile Data Mining, Visualisation

I. INTRODUCTION

The phenomenal growth of mobile devices coupled with their ever-increasing computational capacity presents an exciting new opportunity for real-time, intelligent data analysis in pervasive/ubiquitous environments. Ubiquitous/Mobile Data Mining is the process of analyzing data streams using mobile and/or embedded devices (e.g., sensors) to support critical applications such as mobile healthcare, intelligent transportation systems, and emergency/disaster management like bushfires.

The typical constraints that have to be addressed in performing mobile data mining are: (1) data streams are generated and sent in real-time in a stream format with little or no potential for persistent storage, (2) resource constraints include limited computational resources such as memory, processor speed, network bandwidth, battery power, and screen real-estate, (3) temporal constraints refer to real-time information and decision-making needs that, in turn, necessitate the analysis to be online, incremental and continuous, (4) mobility of users and devices and the connectivity issues thereof, and (5) adaptation of the analysis process to varying/dynamically changing resource-levels and user needs.

In the last few years, rapid strides have been made in accurately and efficiently mining high speed data streams in mobile devices such as Personal Digital Assistants (PDAs)/Smart Phones [1, 2, 3]. However, there are currently no general strategies for visualization developed or available for mobile data mining. Thus, while much research has focused on developing novel ways of analyzing data in real-time on mobile devices, there have not been specific techniques developed for visualization of the analysis. This can be primarily attributed to the fact that it is only now that even analysis is possible on mobile devices. Visualization of the results from analysis in real-time is therefore an emerging challenge - but one that is vital in order to effectively leverage the benefits of mobile data mining to enable real-time decision making by mobile users. The key challenges to visualization of mobile data mining are: 1) The small screen real-estate of mobile phones/PDAs and therefore the need to effectively use this limited screen space to present useful and easy-to-understand information; 2) The need to dynamically perform computations relating to visualization; and 3) The need to rapidly change the visualization so that they capture and reflect accurately the current state of the underlying analysis process.

In this paper, we present an adaptive approach for real-time cluster visualization for mobile data mining. This approach is based on the principles of clutter-reduction [4]. In general, visualization theory studies have established that reducing the clutter on the screen is a key to improving perception and understanding of information that is presented. An evaluation of clutter levels of traditional visualization techniques has been discussed in [4]. The issue of on-screen clutter becomes significant in the context of mobile devices which have much smaller screens than traditional desktops (on which the aforementioned studies have been conducted). In addition, it becomes much more acute when visualization is for a continuous, rapid and dynamically changing situation as is typical for mobile data mining applications such as monitoring heart-patient ECGs or analyzing mobile police personnel locations in real-time.

We have developed the first clutter-aware data stream clustering visualizer for mobile data mining that automatically considers the amount of information presented on a screen and dynamically adjusts the way this information is presented to reduce clutter and therefore increase comprehension/ease of understanding.

The rest of this paper is organized as follows: Section 2 presents our Clutter-Aware Clustering Visualizer CACV technique. Section 3 shows detailed experimental results and has links to the demonstration videos and software. Finally the paper is concluded in Section 4.

II. CLUTTER-AWARE CLUSTER VISUALIZER

We propose and develop a cluster visualizer that autonomously adapts the visualization according to changing screen-clutter. The technique is further sophisticated by the inclusion of elements that allow this adaptive visualization to be personalized to the user and the ability of the visualizer to
adapt the process in relation to the available energy/battery levels on the device. There are three main aspects that need to be factored. First, there are a large number of mobile devices each with varying screen sizes and computational capability. Second, the range of applications for which mobile data mining can be used and the consequent application specific needs for visualization is also highly variant. Thirdly, there is the variation in different users’ ability to process information present on the screen. Given these variables, our clutter-aware visualization algorithm allows changing at anytime the key thresholds that control the visualization process, such as how much of clutter is tolerable for a user and what frequency of information update is required by a user. Thus, the technique is not merely adaptive to clutter-levels on the screen, but is also flexible enough to be tailored to each user’s personal preferences. Furthermore, these preferences can be changed at anytime while the analysis and visualization is occurring, thereby enabling it to be dynamically instrumented for changing situations.

The visualization process in any mobile/ubiquitous data mining context occurs as follows. The data from various sources such as sensors is sent as a continuous stream to the mobile device. The data is analyzed by the mobile data mining algorithm on the phone. The results from the analysis process are sent to the visualization component which processes this incoming knowledge and presents the results on the mobile device’s screen. Clearly, the visualization process is closely linked with the analysis process as it has to display the output of the analysis process. Furthermore, there is a clear implication that the visualization strategy would also vary depending on the underlying analysis technique. Thus, while our proposed visualization technique in principle is general enough to support a range of mobile data analysis techniques, the algorithm presented here and its implementation are for a specific subclass of analysis techniques. In particular, we have focused on point-based clustering techniques such as LWC, RA-Cluster and Very-Fast K-Means. Point-based clustering techniques, in the context of mobile data mining, process incoming data streams and assign data items/records in real-time to existing clusters or create new clusters based on distance measures. The incoming data items are processed in a one-pass mechanism and are not persistently stored.

These clustering techniques in general are useful for mobile data analysis applications such as analyzing incoming calls to mobile field personnel during an emergency, analyzing taxi requests with respect to the context of the driver such as current drop-off location, analyzing heart rates to detect when the rates are outside the normal range, stock market data etc. While many clustering techniques only focus on clusters that grow, we have also extended this process to allow for shrinking of clusters which is relevant in applications such as emergency calls, taxi requests to represent the current tasks that need to be processed. An updated version of LWC to support both growing and shrinking clusters has been developed.

We measure clutter (c(v)) as in terms of two diverse attributes. One pertains to the percentage of the screen occupied by the clusters generated and shown on the screen. The second pertains to the percentage of clusters that overlap/intersect with one another on the screen. This approach to looking at clutter is consistent with how clutter is defined in the literature in terms of the amount of information presented on the screen. If this amount or some aspects of the information (i(v)) that is presented has the ability to reduce the understanding comprehension of the information that is presented (p(v)), then this can be treated as clutter. In the context of clustering, if two or more clusters intersect/overlap then it becomes unclear and indistinguishable. Thus, our approach to clutter-awareness is through allowing users to determine what level of screen occupancy and cluster overlap is tolerable for them given their context and dynamically adapt the visualization in real-time according to these preferences. We also allow these levels to be changed at any point of time while the analysis and visualization is happening. Our clutter-aware clustering visualization algorithm starts by accepting four settings from the user. The first is the acceptable level of coverage on the screen (CT). The second parameter is the allowable percentage of overlapping between clusters (OT). The third parameter is the duration between each two consecutive screen refresh events (RefSc). It is worth mentioning that such parameters are essential due to differences among applications, users, varying screen sizes of mobile phones which in turn results in differences in what is an appropriate amount of information presented on the screen. The last parameter is the maximum number of clusters to be presented on the screen (NocT hreshold). The actual process starts with drawing the clusters on the screen having each cluster size reflecting the number of points/data items in this cluster. This process is followed by the periodic assessment of the percentage of coverage (ScCov) and cluster overlapping (OvClu) against the preset thresholds (CT) and (OT) for coverage and overlap percentages respectively. This is done over fixed intervals of refreshing the screen (RefSc).

If one of these measurements exceeds the user set parameter, the visualization scheme changes dynamically to the first level of adaptation, scaling. Scaling is done such that all the clusters on the screen have their sizes reduced by a scaling factor (ScaleF actor) under the condition that the smallest cluster size is not less than a cluster with only one point in the normal mode. If after the first level of clutter-aware adaptation, the screen is still cluttered with the current occupancy being greater than the specified threshold, the percentage of overlapping clusters exceeds the acceptable overlap threshold, or the smallest cluster is underscaled (min(w0i) < 1), our visualizer switches to the next level of adaptation, shading. w0i is the number of points in cluster ci multiplied by the scaling factor (ScaleF actor). The shading is done by setting all the clusters to the same default size. Each cluster is drawn with different intensity of color/shade. The darkest cluster represents the stronger ones with highest number of points, while the white clusters represents the ones with lowest number of points. If one of the two conditions of cluttering or cluster overlapping still holds, or the number of clusters (NumOfClusters) on the screen is greater than a preset threshold(NocT hreshold), the visualizer...
switches to its highest level of adaptation, cluster selection. This is done by selecting only active clusters to be represented on the screen. We define active clusters as the ones that have attracted new data points in the most recent time intervals. The clutter-aware adaptive cluster visualization is shown below in Algorithm 1. In the context of mining data streams on mobile devices, energy consumption and management is a critical aspect. We have therefore included in our approach strategies for energy consumption ensuring that the application will last its preset lifetime. The energy adaptation happens based on parameters such controlling refresh rates and backlight according to changing battery levels.

III. EXPERIMENTAL EVALUATION

We have implemented our CACV algorithm with its clutter and energy adaptive strategies on a Google Android mobile phone. The implementation is a UDM application can support a range of scenarios discussed as follows where we will demonstrate the visualiser on a range of real and synthetic data sets:

- **Location Data:** This is primarily a stream of GPS coordinates from location sensors that is sent in real-time to the mobile device where the data is clustered using the LWC clusterer and visualized using the clutter-aware visualiser. Typical application scenarios include the following. A taxi driver picks up a person at the airport and is driving towards a particular suburb. By analyzing the data from GPS sensors of other taxis in that part of the city, the driver will be able to see where there is a current concentration of taxis and determine the location that he should be heading to. Similarly this can easily be mapped to identifying and analyzing the location of mobile police personnel or emergency personnel managing calls when a storm hits a city. Another potential is for viewing growing stress levels from fire-fighting personnel deployed to different areas, as well as for evacuating buildings when persons have RFID tags. We use Google Maps to obtain the maps for the GPS coordinates.

- **Stock Market Data:** The mobile device receives a stream of stock information such as price, volume of trade etc. which are analyzed by the LWC data stream clustering algorithm and visualized using our approach. The relevant underlay of the area maps for the GPS coordinates.

We have used both synthetic and real datasets to evaluate our techniques. Synthetic data has been used by randomly generating the latitude and the longitude of a location on the map within the range of suburbs in Melbourne South East in Victoria state of Australia. This demonstration can be viewed at: http://www.youtube.com/watch?v=8htWxdA5nf0 and http://www.youtube.com/watch?v=U7GuJ8U5r6c. The stock market data is a real data set obtained from the E-Research visualization challenge (http://www.eresearch.edu.au/vischallenge) and is over 1 GB of trade, volume and price information collected from that we stream as through it were a live stock feed from remote server. This demonstration is available at: http://www.youtube.com/watch?v=U7GuJ8U5r6c. Finally we also have a demonstration of the visualiser in a desktop environment which shows the value of the adaptive cluster visualization in an evacuation scenario (shown in Figure 1). This uses the evacuation dataset made available as part of the 2009 VAST Visualization Challenge (the data consists of a floor plan and people's x,y coordinates gathered using RFID). The cluster analysis is effectively able to show changing areas of bottlenecks as people leave the building. Since this is a desktop scenario we have made the Netbeans project available for download at: http://hercules.infotech.monash.edu.au/~shonali/ICDM2010.

---

**Algorithm 1 CACV Algorithm**

```plaintext
{User Settings} 
input CT, OT, NecThreshold, RfSe
repeat 
  if RfSe is reached then 
    {Draw Clusters} 
    for each cluster c_i do 
      Draw according SET(VIS) 
      SET(VIS) = Area(c_i)\text{area}_i 
      end for 
    end for 
    Calculate ScCor, OcClu 
  [Adaptation Level 1: Scaling] 
  if ScCor > CT + OcClu > OT then 
    ScaleFlag = False 
    for each cluster c_i do 
      w_i' = w_i \times ScaleFactor 
      if w_i' < 1 then 
        Break 
      end if 
    end if 
  end if 
  if StateFlag = False then 
    SET(VIS) = Area(c_i)\text{area}'_i 
    Calculate ScCor, OcClu 
  end if 
  if (Adaptation Level 2: Shading) 
  if ScCor > CT + OcClu > OT then 
    SET(VIS) = Color(c_i)\text{color}'_i 
    Calculate ScCor, OcClu 
  end if 
  [Adaptation Level 3: Cluster Selection] 
  if ScCor > CT + OcClu > OT \lor 
    NumOfClusters > NecThreshold then 
    SET(VIS) = Color(\text{active}(c_i))\text{area}'_i 
    end if 
  end if 
until END OF THE STREAM
```

The retrieval of property records is done one by one emulating the streaming data. We use the address of the property to locate its location on the map. We also demonstrate the interactive change of visualization in this demo. This demonstration can be viewed at: http://www.youtube.com/watch?v=8htWxdA5nf0 and http://www.youtube.com/watch?v=U7GuJ8U5r6c. The stock market data is a real data set obtained from the E-Research visualization challenge (http://www.eresearch.edu.au/vischallenge) and is over 1 GB of trade, volume and price information collected from that we stream as through it were a live stock feed from remote server. This demonstration is available at: http://www.youtube.com/watch?v=U7GuJ8U5r6c.

Finally we also have a demonstration of the visualiser in a desktop environment which shows the value of the adaptive cluster visualization in an evacuation scenario (shown in Figure 1). This uses the evacuation dataset made available as part of the 2009 VAST Visualization Challenge (the data consists of a floor plan and people’s x,y coordinates gathered using RFID). The cluster analysis is effectively able to show changing areas of bottlenecks as people leave the building. Since this is a desktop scenario we have made the Netbeans project available for download at: http://hercules.infotech.monash.edu.au/~shonali/ICDM2010.
Once imported the project can just be run to see the system in operation.

Our experimental analysis is aimed to demonstrate the visual effectiveness of the CACV approach in terms of being adaptive to screen clutter as well as cluster overlap. We also demonstrate the prolonged application lifetime that we achieve through energy adaptation. As shown in Figure 2, it becomes very evident that in a short span of a matter of minutes, the screen of a mobile device becomes completely overwhelmed and incomprehensible.

In the experiments shown in Figure 3, we have run the full CACV algorithm with CT = 50% and OT = 75% over the synthetic datasets. As shown, the algorithm adapts by scaling, but it did not reach the stage of coloring given that scaling has been able to keep both the coverage and overlapping percentages below the preset CT and OT. This gives a clear evidence of the full functionality of our CACV technique. The graph in Figure 4 clearly shows that the algorithm has kept the coverage and overlapping percentages below the respective thresholds by scaling down the cluster sizes. The graph also shows that the mode of visualization has changed from normal to scaling, but it did not reach the coloring nor the visualization of active clusters modes.

In order to demonstrate the effectiveness of our proposed energy adaptation process, experiments were conducted by comparing the runtime of the application with and without adaptation by starting it from 100% battery level until the phone is completely drained. The results from 5 experimental runs establish that the normal run of the application with clustering and visualization can last for approximately 3 hours. With the implementation of energy adaptation, the application was able to run for approximately 5 hours. Thus, it is clearly evident that our visualizer with energy management strategies enabled can almost increase the application runtime on a state of the art mobile device by around 40%. The results are shown in Figure 5.

We have presented a novel approach for the area of visualization for mobile and ubiquitous data mining. This technology has the potential to enable an entire new class of intelligent analysis applications for supporting mobile users.

V. REFERENCES


