Journal of Computer and System Sciences 79 (2013) 369–382

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Interactive self-adaptive clutter-aware visualisation for mobile data mining

Mohamed Medhat Gaber [a,*], Shonali Krishnaswamy [d,b], Brett Gillick [b], Hasnain AlTaiar [b], Nicholas Nicoloudis [b], Jonathan Liono [b], Arkady Zaslavsky [c]

[a] University of Portsmouth, Portsmouth, Hampshire, England, PO1 3HE, UK
[b] Monash University, Caulfield East, Melbourne, Victoria, 3145, Australia
[c] CSIRO ICT Centre, GPO Box 664, Canberra ACT 2601, Australia
[d] Institute for Infocomm Research, 1 Fusionopolis Way, 21-01 Connexis (South Tower), Singapore 138632, Singapore

ARTICLE INFO

Article history:
Received 5 March 2011
Received in revised form 19 January 2012
Accepted 11 September 2012
Available online 16 October 2012

Keywords:
Visualisation
Data streams
Data mining
Clustering
Mobile computing
Adaptive clutter reduction

ABSTRACT

There is an emerging focus on real-time data stream analysis on mobile devices. A wide range of data stream processing applications are targeted to run on mobile handheld devices with limited computational capabilities such as patient monitoring, driver monitoring, providing real-time analysis and visualisation for emergency and disaster management, real-time optimisation for courier pick-up and delivery etc. There are many challenges in visualisation of the analysis/data stream mining results on a mobile device. These include coping with the small screen real-estate and effective presentation of highly dynamic and real-time analysis. This paper proposes a generic theory for visualisation on small screens that we term *Adaptive Clutter Reduction ACR*. Based on *ACR*, we have developed and experimentally validated a novel data stream clustering result visualisation technique that we term *Clutter-Aware Clustering Visualiser CACV* and its enhancement of enabling user interactivity that we term *iCACV*. Experimental results on both synthetic and real datasets using the *Google Android* platform are presented proving the effectiveness of the proposed techniques.

1. 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 analysing 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 [14] 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 [15] in mobile devices such as *Personal Digital Assistants* (*PDAs*) [10,11,6] and there is a growing focus on "in-network" processing

* Corresponding author.
E-mail address: mohamed.gaber@port.ac.uk (M.M. Gaber).

0022-0000/$ – see front matter
http://dx.doi.org/10.1016/j.jcss.2012.09.009

using embedded devices such as sensor nodes [7,12,13]. Kargupta et al. [11] have developed a client/server data stream mining system: MobiMine which focuses on data stream mining applications for stock market data. Kargupta et al. in the *Vehicle Data Stream Mining System VEDAS* [10] develop algorithms that are operational on a *PDA*. *VEDAS* is a data stream mining system that allows continuous monitoring and pattern extraction from data streams generated on-board a moving vehicle. *MOLEC* [13] is a mobile cardiac monitoring system that aims to analyse ECG signals to identify a range of anomalies and arrhythmias using decision trees that were built off-line. In [7], real-time mining of information from in-vehicle sensors to minimise driver distraction is proposed through adaptation of the instance-selection process based on changes to the data distribution. Dynamic adaptation to data rates and fine tuning of processing parameters can significantly enhance the longevity of continuous real-time processing of data streams in mobile environments. Several mobile data stream mining algorithms have been proposed that operate using the principles of adaptation for Clustering, Change Detection, Classification and Frequent Items Analysis [6,8,9]. In summary, these techniques leverage the body of work that exists in mining data streams and aim to enable the operation of these algorithms in resource-constrained environments [9].

However, there are currently no general strategies for visualisation developed or available for mobile data mining. Thus, while much research has focused on developing novel ways of analysing data in real-time on mobile devices, there have not been specific techniques developed for visualisation of the analysis. We do note that visualising data streams has been the focus of recent research [25–27]. This can be primarily attributed to the fact that it is only now that even analysis is possible on mobile devices. Visualisation 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 visualisation 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 visualisation; and
3. the need to rapidly change the visualisation so that they capture and reflect accurately the current state of the underlying analysis process.

In this paper, we present a generic approach for real-time visualisation for mobile data mining – *Adaptive Clutter Reduction ACR*. This approach is based on the principles of clutter reduction [16]. In general, visualisation 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 visualisation techniques has been discussed in [16]. 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 visualisation is for a continuous, rapid and dynamically changing situation as is typical for mobile data mining applications such as monitoring heart-patient ECGs or analysing mobile police personnel locations in real-time.

Based on our *ACR* approach, we have developed the first clutter-aware data stream clustering visualiser 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 technique is further sophisticated by the inclusion of elements that allow this adaptive visualisation to be personalised to the user and the ability of the visualiser to adapt the process in relation to the available energy/battery levels on the device. There are three main variables 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 visualisation 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 visualisation algorithm allows changing at anytime the key thresholds that control the visualisation 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 visualisation is occurring, thereby enabling it to be dynamically instrumented for changing situations.

The rest of this paper is organised as follows: Section 2 presents related work in the area of clutter reduction techniques. Section 3 presents our *Adaptive Clutter Reduction ACR* theory. Section 4 details our *Clutter-Aware Clustering Visualiser CACV* technique based on the theory of *ACR*. An important enhancement to *CACV* that enables user interaction is given in Section 5. Section 6 shows detailed experimental study. Finally the paper is concluded in Section 7.

2. Related work

Our review of related work is discussed in the context of two areas: (1) the broad area of visual clutter reduction techniques, and (2) the existing techniques for *UDM* visualisation, which is in its nascent stage of growth and evolution. The problem of cluttered screens for information visualisation has long being studied [16] and a comprehensive review and taxonomy of such techniques is presented in [16]. Many clutter reduction techniques have been proposed in the literature and have been shown to improve the understanding of the underlying information visualised on the screen. These techniques have been categorised into three broad categories pertaining to the appearance, spatial distortion or temporal aspects of