Research Article *ISSN: 2319-507X*
Rahul Wasule, IJPRET, 2013; Volume 1(8): 1101-1111 *IJPRET*

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

MOBILE DATA STREAM MINING: ADAPTATION STRATEGIES WITH AOG

RAHUL WASULE, R.A. FADNAVIS

1. M.Tech, Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur.

2. Asst. Prof. , Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur.

Accepted Date:
27/02/2013

Publish Date:
01/04/2013

Keywords

Mobile Data

AOG

Corresponding Author

Mr. Rahul Wasule

Abstract

This paper presents an overview of the current state-of-the-art in mobile data stream mining and its applications. The paper presents the strategies and techniques for adaptation that are essential in order to perform real-time, continuous data mining on mobile devices. We present an overview of adaptation strategies for data stream mining and in particular for Resource-Aware Adaptation with Algorithm Output Granularity. Finally, we discuss the key challenges and future directions of mobile data stream mining.

INTRODUCTION

Mobile devices are increasingly becoming the central computing and communication device in people's lives. Devices today are equipped with a growing number of sophisticated embedded sensors such as an accelerometer, digital compass, gyroscope, GPS, microphone, light intensity sensor, and camera. [1]This creates the opportunity to develop applications that leverage on the sensing capability of these mobile devices, as well as data from other sensors such as bio/body sensors. Data from mobile users/devices is becoming increasingly important for numerous applications including urban modeling, transportation, and more recently for mobile crowd-sensing for citizen journalism and real-time traffic routing. While significant efforts are being focused towards the analysis of mobile user data, a key challenge that needs to be addressed in order to realize the full-potential of mobile user analytics is to address the scalability issues of real-time data collection. By scalability, we refer to both the challenges of data transmission from a large number of users, as well as the issues of energy consumed on individual devices as a result of that transmission.

Mobile data stream mining is a key technology for real-time analysis of data streams generated on-board the phone itself, for both data generated by sensors on the phone and/or in close proximity to the phone. The significant advantages that mobile data stream mining provides over traditional strategies for leveraging the phone as a "transmission device" for sensor data, are as follows: reduce the amount of data transmitted from the phone to servers/the cloud, as well as reduce the energy/battery usage on the phone due to transmission of sensor data. Mobile data stream mining is particularly significant for applications that need real-time analysis of continuous data streams such as such as mobile crowd sensing, mobile activity recognition, intelligent transportation systems, mobile healthcare, and so on. Mobile data stream mining techniques typically focus on adapting data stream mining algorithms to be operational in the context of mobile devices

In addition to such adaptation techniques for mobile data stream mining, numerous "light-weight" mining algorithms for varies types of analysis such as clustering, classification, concept drift detection, change detection, and frequent items analysis have been developed.

A.ADAPTATION STRATEGIES FOR ENABLING MOBILE DATA STREAM MINING

There has been much work on developing adaptation strategies for data stream mining algorithms that vary the accuracy levels according to available computational resource levels and incoming data rates. Adaptation strategies have been shown to significantly enhance the longevity of continuous real-time processing of data mining in mobile environments. Adaptation can enable, if not guarantee, the continuity, cost-efficiency and consistency of a mobile/ubiquitous data stream mining application.

As discussed in even efficient data stream mining algorithms such as Very Fast K-Means (VFKM) can cause device crashes on mobile devices when used without awareness to context such as variations in data rates and computational resource availability. Thus, there are primarily three different adaptation strategies that have been proposed for mobile data stream mining:

Situation-Aware Adaptation: With situation-aware adaptation, the focus of adaptation is with respect to the application's accuracy requirements changing based on the occurring situations. When the current situation warrants for less frequent monitoring/analysis, the algorithm accuracy can be moderately decreased to preserve resources.[1] On the other hand, in critical situations where there is a need for closer monitoring, it is important to increase the accuracy even if the resource availability is scarce. By situations we mean real-life situations such 'fire threat', 'heat-stroke', 'driving' and many others. There are certain situations in which applications do not need high accuracy (e.g. 'healthy' situation). When such situations occur, data mining algorithm settings can be adjusted to produce results with the low accuracy and thus consume less resource. On the other hand, when critical situations (e.g. 'fire threat') occur, the application requires a higher level of accuracy and the settings of the mining algorithm can be adjusted to achieve this according to the need. Situation-aware adaptation is typically driven by leveraging context-aware and situation-aware engines customized for specific mobile data mining applications.

mHealth (also written as m-health or mobile health) is a term used for the practice of medicine and public health, supported by mobile devices. The term is most commonly used in reference to using mobile communication devices, such as mobile phones, tablet computers and PDAs, for health services and information, but also to affect emotional states. The mHealth field has emerged as a sub-segment of eHealth,

A Scenario

John has had a heart attack and is released from hospital but there are concerns that he might be susceptible to another heart attack and is also experiencing blood pressure fluctuations. Constant monitoring of his vital signs could help him to reduce his anxiety, decrease the need for routine visits to medical facilities, and also detect early warning features of a possible impending event. He has a smart phone with SAAP installed on it and is willing to wear biosensors to measure his vital signs. The data is wirelessly sent to his mobile where SAAP detects any changes not only in his vital signs but in any contextual information that is related to the application (e.g. the battery level of the mobile phone). SAAP uses this information to reason about situations in real-time and according to inferred situations, it performs intelligent and cost -efficient analysis of data. When fluctuations of vital signs are within a specified acceptable threshold, there is no need for frequent measurement and use of resources can be reduced and moderated. However, when these fluctuations are over the threshold, this situation warrants a closer monitoring by the system and more frequent measurements. This type of adaptation requires factoring in both available resources and criticality of health situations.

According to the analyst firm Berg Insight, around 2.8 million patients worldwide were using a home monitoring service based on equipment with integrated connectivity at the end of 2012. The figure does not include patients that use monitoring devices connected to a PC or mobile phone. It only includes systems that rely on monitors with integrated connectivity or systems that use monitoring hubs with integrated cellular or fixed-line modems. Berg Insight forecasts that the number of home monitoring systems with integrated communication capabilities will grow at a compound annual growth rate (CAGR) of 26.9 percent between 2011 and 2017 reaching 9.4 million connections globally by the end of the forecast period. The number of these devices that have integrated cellular connectivity increased from 0.73 million in 2011 to about 1.03 million in 2012, and is projected to grow at a CAGR of 46.3 percent to 7.10 million in 2017.

Resource-Aware Adaptation:

The key focus of the adaptation strategies has been to factor in varying levels of computational resources on the mobile device and use this as a continuous control parameter to adapt the behavior of the stream mining algorithms that are operational on the mobile device. Generic granularity-based adaptation techniques that can be used with any data stream mining technique running on a resource-constrained device have been developed. This approach facilitates adaptation of data stream mining algorithms to varying data rates based on available computational resources in mobile devices by innovative strategies to perform knowledge integration, controlling the rate of learning and varying the accuracy levels of the discovered patterns. The granularity-based adaptation approach has three different variations. AOG (Algorithm Output Granularity) provides adaptability by adjusting the algorithm output rate (e.g. the number of clusters) according to the availability of memory, the remaining time to fill the available memory and the data stream rate. AOG uses a time threshold that is the time required to generate the results before any incremental integration according to some accuracy measure. AIG (Algorithm Input Granularity) is a process that adapts the data rates feeding into the algorithm according to the battery charge by using the methods of sampling, load shedding and synopsis. When the battery charge becomes critical, the adaptation process changes the sampling rate to reflect the current data rate according to the previous consumption pattern in the most recent time frame. APG (Algorithm Processing Granularity) performs adaptation of the processing settings of the algorithm with respect to the CPU usage. For example, in a a stream clustering analysis, it uses a novel approach termed Randomization Assignment. When the algorithm needs to make the micro-cluster assignment for a new data point, this method enables examining only a fraction of the current micro-clusters using the randomization factor. If the randomization factor is equal to 1, all micro-clusters need to be examined. However if the usage pattern of CPU increases this factor will be decreased; thereby reducing the number of micro-clusters examined and consequently the CPU consumption.

Mining data streams for mobile and embedded applications faces a major problem represented in the high rate of the streaming input with regard to the available computational resources. Common resources are battery charge, remaining memory, CPU utilization, communication buffer or bandwidth. The algorithm granularity setting that is responsible for adjusting mining algorithm parameters according to resource availability.

Hybrid Adaptation:

Hybrid adaptation strategies aim to integrate resource-aware strategies and situation-aware strategies to control the adaptation process

B. ALGORITHMS FOR MOBILE DATA STREAM MINING

Mobile data stream mining algorithms typically aim to perform the same data stream mining on-board a mobile device. In general, these algorithms are typically one-pass techniques that leverage sliding windows since they operate over streaming data. There are two strategies for development of mobile data stream mining algorithms. Basically, as explained a mobile data stream mining algorithm operates with variability of performance according to resource, situation and other constraints.

This variable performance is typically effected by specifying an upper and lower bound for accuracy levels that are acceptable. Thus, there are two ways in which algorithms are developed according to this principal. Firstly, the algorithms themselves are developed as light-weight techniques that have in-built strategies for adaptation such as

- Clusterers
  - Light-Weight Clustering
  - RA-Cluster and DRA-Cluster
- Change Detection
  - CHANGE-DETECT
- Classifiers
  - Light-Weight Class (LWC)
- Frequent Items and Associations
  - LWF (Light-Weight Frequent Items)
  - HiCoRE (Highly Correlated Energy-Efficient Rules)

Alternatively, the algorithms are general stream mining techniques that can be treated as a "black-box" (i.e. no modifications to the internal processing), but adaptation is leveraged by identifying an algorithm's potential control parameters

We have proposed to adapt the mining algorithm output according to resource availability and data stream rate. We have termed this approach as Algorithm Output Granularity *(AOG)* Based on AOG, we have developed data stream mining algorithms for clustering classification frequent items and change detection and classification This paper discusses the theoretical framework of AOG. The generalization of the approach using the Probably Approximately Correct (PAC) machine learning model is also discussed as a basis for theoretically applying AOG to other stream mining techniques. The choice of PAC learning is based on its acceptance as the fundamental machine learning model.

C. ALGORITHM OUTPUT GRANULARITY

AOG operates using three factors to enable the adaptation of the mining algorithm to memory availability.

ISSN: 2319-507X
IJPRET

1. The rate of the incoming data stream is the first factor.

2. The rate of the algorithm output represents the second one.

3. From these two, an estimated time duration to fill the available memory according to the logged history of data rate and algorithm output rate is calculated. This represents the last factor.

These three factors are used to adjust what we call the *algorithm threshold*. This threshold can control the output rate of the algorithm according to the mining technique. Fig. 1 shows how the algorithm threshold can control the output rate of a mining algorithm according to the three factors that AOG operates on. The data arrives sequentially and its rate is calculated.

The algorithm runs with an initial threshold value, and the rate of the output is observed. The threshold is adjusted periodically to conserve the available memory according to the relationship among the three factors.

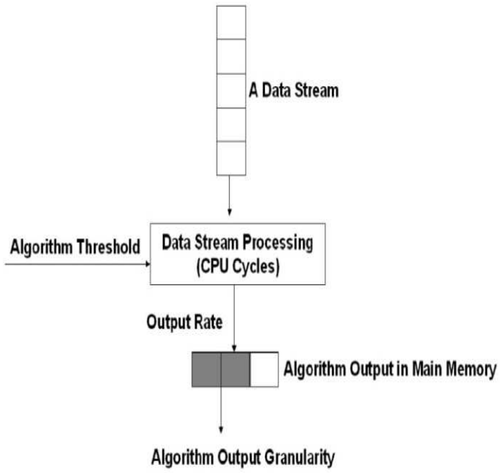

Figure 1: Algorithm Output Granularity Approach

AOG is a three-stage, resource-aware threshold-based data stream mining approach. The process of mining data streams using AOG starts with a mining phase. In this step, the algorithm threshold that can control the algorithm output rate is determined as an initial value set by the user of the system (or preset to a default initial value). In this stage, there are three variations in using this threshold according to the mining technique:

- *Clustering:* the threshold is used to specify the minimum distance between the cluster center and the data stream record.
- *Classification:* in addition to using the threshold in specifying the distance, the class label is checked. If the class label of the stored records and the new item/record that are close (within the accepted distance) is the same, the weight of the stored item is increased and stored along with the weighted average of the other attributes, otherwise the weight is decreased and the new record is ignored.
- *Frequent patterns:* the threshold is used to determine the number of counters for the frequent items.

The second stage in the AOG approach is the adaptation phase. In this phase, the threshold value is adjusted to cope with the data rate of the incoming stream, the available memory, and time constraints to fill the available memory with resultant knowledge structures.

The third and final stage in AOG approach is the knowledge integration phase. This stage represents the merging of produced results when the computational device is running out of memory. In clustering, we use the merging of clusters that are within short proximity. The merging of class representatives is used in classification. Releasing the least frequent items from memory is our strategy for frequent pattern mining. This integration allows the continuity of the data mining process. Otherwise the computational device would run out of memory even with adapting the algorithm threshold to its highest possible value that results in the lowest possible generation of knowledge structures. Fig. 2 shows a flowchart of AOG- mining process. It shows the sequence of the three stages of AOG.

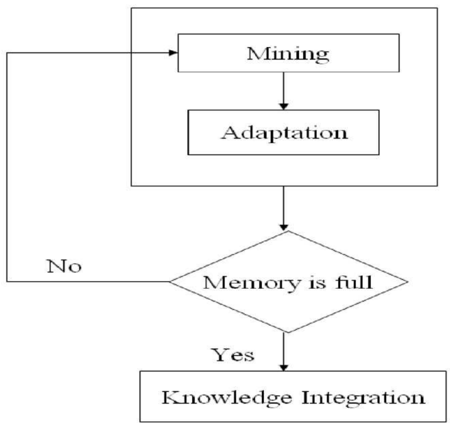

Figure 2: AOG Mining Approach Flow char

D.CONCLUSIONS AND FUTURE DIRECTIONS

With the significant interest in mobile users and applications, driven by the ever increasing sophistication and capabilities of today's mobile device, mobile data mining is emerging as a key technology. This paper gives an overview of the current state-of-the-art in terms of algorithms, adaptation strategies, and applications for mobile data stream mining. There are many key areas for future work including developing new application case studies that leverage mobile data mining such as in the network gaming area, as well as to develop activity recognition technologies that turn the phone to your "personal protection" device, as well as large scale gathering of data through mobile devices to sense urban phenomena and events. Furthermore, there is always need for more sophisticated analysis and visualization techniques. Finally, a key challenge is to take the next step from analysis to providing real-time decision making for mobile users.

E. REFERENCES

1. Shonali Krishnaswamy, Joao Gam, Mohamed Medhat Gaber "Mobile Data Strem Mining: From Algorithms to Applications" 2012 IEEE 13th International Conference on Mobile Data Management

2. Pari Delir Haghighi, Arkady Zaslavsky, Shonali Krishnaswamy, Mohamed Medhat Gaber "Mobile Data Mining for Intelligent Healthcare Support" Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009

3. Joao˜ Bartolo´ Gomes Ernestina Menasalvas and Pedro A. C. "Situation-Aware Data Stream Mining Service for Ubiquitous Applications" Eleventh International Conference on Mobile Data Management

4. Mohamed Medhat Gaber "Foundations of Adaptive Data Stream Mining for Mobile and Embedded Applications" Proceedings of the 2008 IEEE, CIBEC'08978-1-4244-2695-9/08/$25.00 ©2008 IEEE

5. Salim F.D., Loke, S. W., Rakotonirainy, A., Srinivasan, B.,Krishnaswamy, S., 2007, Collision pattern modeling and real-time

collision detection at road intersections, Proc of the 2007 IEEE Intelligent Transportation Systems Conference, USA, pp. 16.

6. Brett Gillick, Hasnain AlTaiar, Shonali Krishnaswamy, Jonathan Liono, Nicholas Nicoloudis, Abhijat Sinha, Arkady B. Zaslavsky, Mohamed Medhat Gaber: Clutter-Adaptive Visualization for Mobile Data Mining. ICDM Demo Paper 2010: 1381-1384