INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

IMPLEMENTATION OF ALGORITHM OUTPUT GRANULARITY APPROACH IN CLUSTERING ALGORITHM FOR MOBILE DATA STREAM MINING

MR. RAHUL WASULE, MR. ANKUSH MAIND

Assistant Professor, Department of Information Technology, Rajiv Gandhi College of Engineering & Research, Nagpur, India.

Accepted Date: 05/03/2015; Published Date: 01/05/2015

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 memory conservation with Algorithm Output Granularity. For mining purpose, we uses k-means clustering algorithm.

Keywords: Mobile data mining, Algorithm Output Granularity, k-means

Corresponding Author: MR. RAHUL WASULE

Access Online On:

www.ijpret.com

How to Cite This Article:

Rahul Wasule, IJPRET, 2015; Volume 3 (9): 1032-1044

PAPER-QR CODE

INTRODUCTION

In the era of Information and technology, Mobile devices have become the prime source of communication as well as center of computing. These mobile devices today are not restricted for communication only but they are used for entertainment, data and information sharing, education purpose etc [1]. As utility of mobile devices is increasing, sophisticated embedded sensors are added to it 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.

In order to meet the above applications, analysis of data generated by mobile devices is of prime importance. 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. Mobile data stream mining is particularly significant for applications that need real-time analysis of continuous data streams such as mobile crowd sensing, mobile activity recognition, intelligent transportation systems, mobile healthcare, and so on. Though above applications have huge future scope, the computing devices has their own limitations such as limited memory size, limited computational resources, screen size and energy considerations of the mobile devices. Mobile data stream mining techniques typically focus on adapting data stream mining algorithms to be compatible with mobile devices so that it can overcome these limitations [2].

II.CONCEPT OF ALGORITHM OUTPUT GRANULARITY (AOG) APPROACH

In proposed system, we try to overcome the problem of limited memory size of mobile devices. For that purpose system uses the Algorithm Output Granularity *(AOG)* approach.

AOG operates using three factors to enable the adaptation of the mining algorithm to memory availability as follows:

1. The rate of the incoming data stream

2. The rate of the algorithm output

3. From these two, 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 [3].

These three factors are used to adjust what we call the *algorithm threshold*. This threshold is used to take the decision of swapping main memory content to virtually unlimited secondary memory. As the content of main memory is swapped to secondary memory, the capacity of RAM is restored for further mining of live data stream.

The data from real time data sources arrives sequentially and its rate is calculated which gives incoming data stream rate. As mining is continuously done on data stream the results are stored in main memory. Memory is filled with certain rate indicating the rate of the algorithm output. From these two, estimated time duration to fill the available memory according to the logged history of data rate and algorithm output rate is calculated.

Entire process is divided into three steps as

1. Data Mining
2. Adaptation by Algorithm Output Granularity
3. Knowledge Integration

In the first step, mining is done on incoming data by using various mining algorithm.

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 [4].

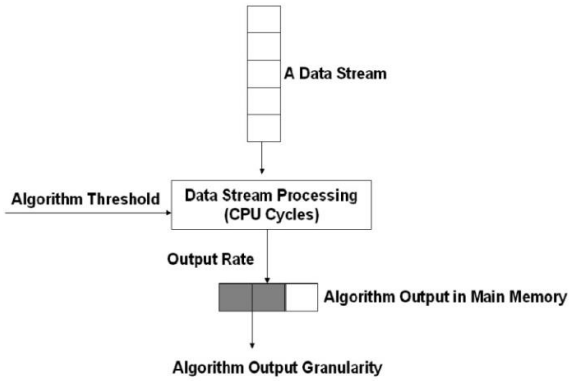

Fig 1: Algorithm Output Granularity Approach

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. This

integration allows the continuity of the data mining process. Otherwise the computational device would run out of memory.

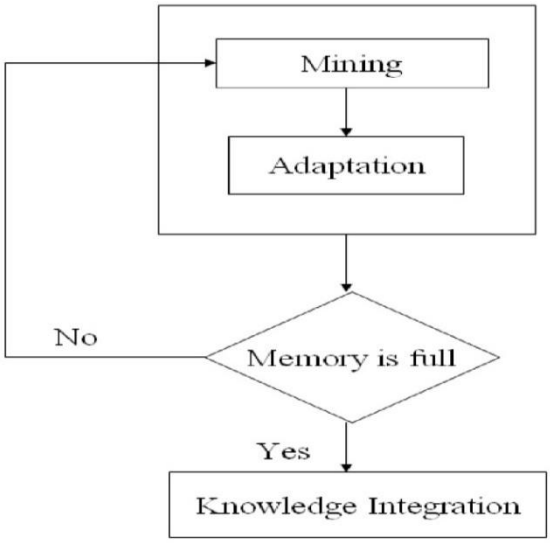

Fig. 2 shows a flowchart of AOG- mining process.

It shows the sequence of the three stages of AOG.

III. PROPOSED SYSTEM

We have divided the work into three modules as

A) Data Sources

B) Implementation of Algorithm Output Granularity approach

C) Knowledge Integration

A) Data Sources

Proposed system needs data sources which can generate real time & continuous data. We need data having stock information on the parameters stocks name, stock value & quantity being traded. Therefore proposed system uses real time data from http://www.steposol.com/stock_feed/get_data.php. [5] We apply clustering algorithm (k-means) on the parameter %change which is calculated by our system[6].

Time	Stock Name	Stock Value	Stock Quantity
T1	TCS	2330.80	980
T2	TCS	2398.20	1002

Our System is calculating %change as

%Change=(T2-T1)*100/T1

=2.89% as a parameter for k-means.

B) Implementation of Algorithm Output Granularity approach

To overcome problem of limited memory we are applying Algorithm Output Granularity approach operates upon three factors as follows-

1. Rate of incoming data stream

Proposed system fetches the real time data from

http://www.steposol.com/stock_feed/get_data.php. System has provision to control data stream rate by user.

2. Remaining available memory

System calculates the remaining available memory. It depends upon rate of incoming data stream. It is calculated as

memory_filled = memory_filled+number_of_total_bytes;

Memory_remain= (Memory Size - memory_filled);

3. Time required to fill available memory

Depending upon above two parameters, we calculate time required to fill available memory indicated by memory_time as follow

memory_time = memory_available / max_rate;

4. Algorithm Output Rate

System generates the results at certain rate to fill the main memory called as Algorithm Output Rate. It is indicated by algo_rate in outputs/second

It is given by

Float algo_rate = (float) (MaxCount) * (1000) / time t;

Where long time t = (end.getTime() - start.getTime());

1000 as time t is in ms.

MaxCount -Amount of data filled in memory in given time t

5. Deciding the threshold value

We calculate AOG threshold to trigger swapping of cluster results to secondary memory.

THi = (algo_rate) + CONSTANT_B;

THip1 = (arbdrp1 * THi) / arbdr;

Where arbdrp1 = algo_rate / memory_time;

Now if this threshold exceeds the time required to fill the remaining memory then swapping of Memory content begins to secondary memory. Because threshold is dependent on algorithm output rate, higher algorithm output rate means more data in main memory which will take less time to fill main memory. Therefore to prevent mobile device running out of memory content of main memory should be swapped to secondary memory as soon as threshold crosses the time required to fill the remaining memory.

C) Integration of Result

After clustering on parameter %change we draw following two conclusions-

I. Top Gainers-The stock which has gain most

i.e. Maximum positive change

II. Top Losers -The stock which has lost most

i.e. Maximum negative change

Depending upon the quantity of stock being sold or buy we decide most active stock

III. Most Active –The stock which has been traded most

i.e. Largest in quantity

If any stock is gaining continuously or traded in large quantity it means that it is in high demand. Now these three results will enable user to make a decision to buy or sell a stock.

Proposed system has been designed as generalized model for Algorithm Output Granularity approach which can be used for different applications and corresponding data sources.

IV. RESULTS:

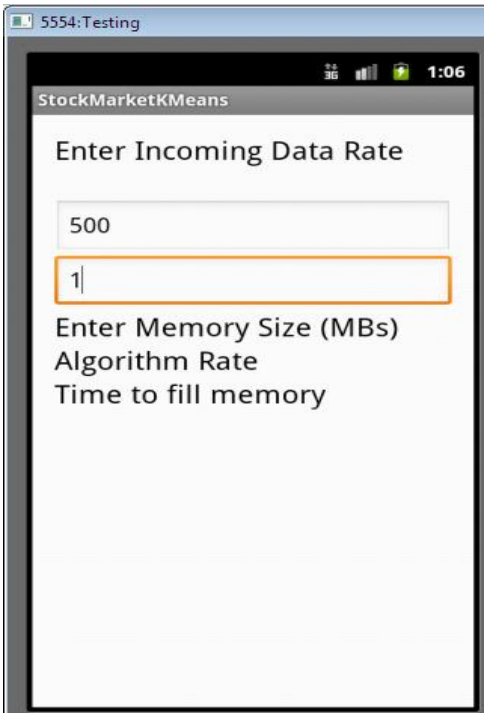

Fig 4: Input Screen

In proposed system, user is able to control the rate of incoming data stream rate. In given screen shot user has given incoming data stream rate equal to 500. User can vary the rate of data stream because for different applications incoming data stream rate is different. User has allocated memory equal to 1 MB. Again user is able to change allocated memory.

After clicking on Apply KMeans button, system starts fetching stock market data from

http://www.steposol.com/stock_feed/get_data.php. System calculates the various parameters for AOG as shown in Fig 5.

Total Memory filled: 3869.0 Bytes

Time required to fill the memory at this rate: 991.944s

Algorithm Rate:727.27 Output/second

Our system calculates % change on stock value and apply k-means algorithm on that calculated %change value. It forms the two clusters as shown in Fig 4.3: Clustering Results

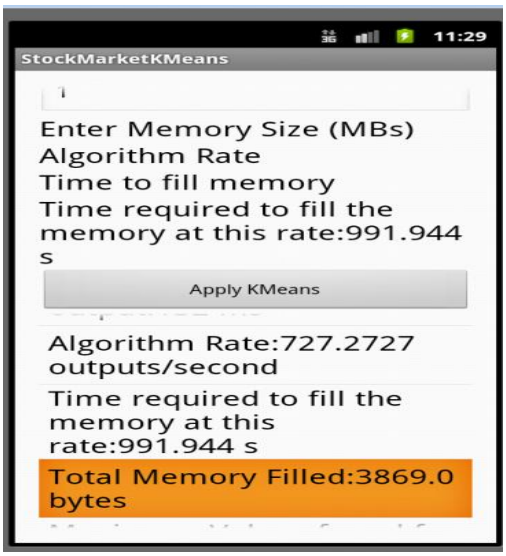

Fig 5: Calculation of AOG parameters

Fig 6: Clustering Results

After applying k-means algorithm on %change stock value, system gives results as top gainer and top loser by comparing % change.

By comparing the quantity of stock traded, our system find outs most active stock as shown in fig 7

Top Gainer: ATLSTEEL

Top Loser: AKITMETAL

Most Active Stock: DCMSRMCOS

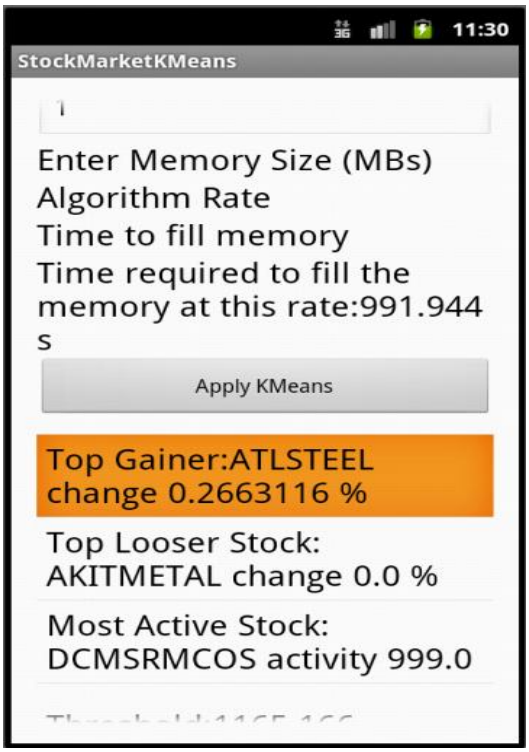

Fig7: Top gainer, Top loser & Most Active Stock

Again data is fetched from website, k-means is applied and results are generated. Every time k-means is applied our system checks the threshold condition. When the threshold value exceeds the time required to fill the remaining memory, our system transfer the memory content to secondary memory in the folder sd card/aog.txt as shown in fig 8 which can be used for future use. As memory contents are dumped to secondary storage, main memory is free which is available for continuous mining of live data stream.

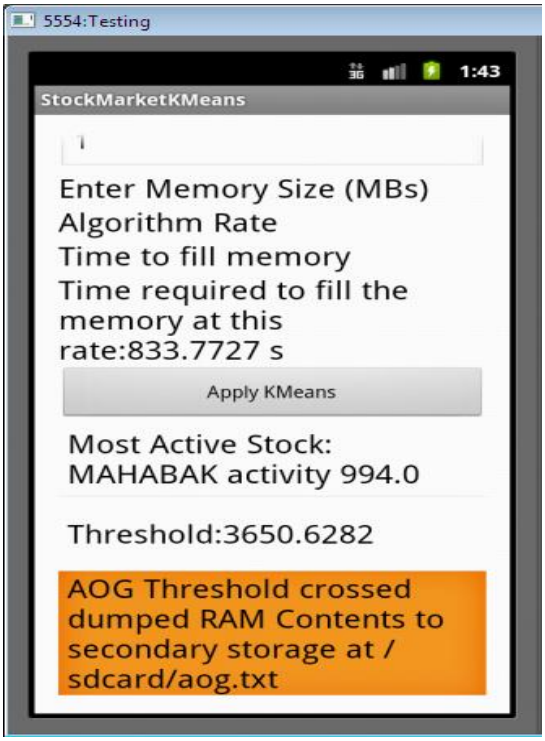

Fig 8: Memory contents are dumped to secondary storage when threshold is triggered

V. CONCLUSION AND FUTURE SCOPE

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.[12]

There are many key areas for future work including developing new application case studies that leverage mobile data mining.

Table 1: Data sources and corresponding application

Data sources	Applications
GPS	Transportation & real-time traffic routing
ECG Signal	A mobile cardiac monitoring system
iPAQ-Vehicle on board monitoring system	Supply chain management
Satellite camera	Weather forecasting
Tools employed in scientific laboratories.	Scientific application

The application and corresponding data sources are given in table 1. These data sources can be implemented with proposed system with necessary changes. 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.

Hence we have derived necessary parameters for Algorithm Output Granularity approach & implemented it for k-means clustering algorithm. Finally we draw conclusion according to our data source i.e. Stock market data.

Algorithm Output Granularity approach can be implemented with different data sources with different mining algorithm for the mining of real time data on mobile devices

REFERENCES

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

2. 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

3. Stahl, F., Gaber, Mohamed, Bramer, Max and Yu, (2010) "Pocket data mining: towards collaborative data mining in mobile computing environments". In: IEEE 22nd International Conference on Tools with Artificial Intelligence (ICTAI 2010), 27-29 October, 2010, Arras, France

4. Hillol Kargupta, Byung-Hoon Park, Sweta Pittie, Lei Liu, Deepali K shraj, “MobiMine: Monitoring the Stock Market from a PDA ”SIGKDD Exploration ,Volume 3, Issue 2-page

5. Emmanouil Koukoumidis, Student Member, IEEE, Margaret Martonosi, Fellow, IEEE, and Li-Shiuan Peh, Member, IEEE “Leveraging Smartphone Cameras for Collaborative Road Advisories” IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 11, NO. 5, MAY 2012

6. Shonali Krishnaswamy, Mohamed Medhat Gaber, Marian Harbach, Christian Hugues, Abhijat Sinha, Brett Gillick, Pari Delir Haghighi, and Arkady Zaslavsky "Open Mobile Miner: A Toolkit for Mobile Data Stream Mining"Centre for Distributed Systems and Software Engineering Monash University, Australia

7. Gaber, Mohamed, Krishnaswamy, S. and Zaslavsky, A. (2005)" Resource-aware mining of data streams" J. UCS (Journal of Universal Computer Science), 11 (8). pp. 1440-1453. ISSN 0948-6968 Journal of Universal Computer Science, vol. 11, no. 8 (2005), 1440-1453 submitted: 10/3/05, accepted: 5/5/05, appeared: 28/8/05 © J.UCS

8. Jo˜ao B´artolo Gomes Ernestina Menasalvas and Pedro A. C. Sousay "Situation-Aware Data Stream Mining Service for Ubiquitous Applications" Facultad de Informatica, Universidad Politecnica Madrid, Spain 978-0-7695-4048-1/10 $26.00 © 2010 IEEE DOI 10.1109/MDM.2010.27

9. Pari Delir Haghighi, Arkady Zaslavsky, Shonali Krishnaswamy, Mohamed Medhat Gaber "Mobile Data Mining for Intelligent Healthcare Support"Center for Distributed Systems and Software Engineeringn Monash University, Australia Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009

10. Mohamed Medhat Gaber, Shonali Krishnaswamy and Arkady Zaslavsky " Cost-Efficient Mining Techniques for Data Streams" School of Computer Science and Software Engineering, Monash University,900 Dandenong Rd, Caulfield East, VIC 3145, Australia In: Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), , January 2004, Dunedin, New Zealand

11. Mohamed Medhat Gaber[1], Shonali Krishnaswamy[1], Arkady Zaslavsky[1] "Adaptive Mining Techniques for Data Streams using Algorithm Output Granularity"[1] School of Computer Science and Software Engineering, Monash University,900 Dandenong Rd, Caulfield East, VIC3145, Australia

12. Mohamed Medhat Gaber "A Holistic Approach for Resource-aware Adaptive Data Stream Mining" Caulfield School of Information Technology, Monash University 900 Dandenong Road, VIC 3145, Australia mohamed.medhat.gaber@infotech.monash.edu.au Philip S Yu IBM T.J. Watson Research Center 9 Skyline Drive, Hawthorne, NY 10532, USA psyu@us.ibm.com

13. Nhan Duc Phung, Mohamed Medhat Gaber and Uwe Röhm " Resource-aware Online Data Mining in Wireless Sensor Networks" Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)University of Sydney, School of Information Technologies Sydney, NSW 2006

14. Greg Hamerly “Making k-means even faster” Department of Computer Science, Baylor University; greg hamerly@baylor.edu

15. Albert Bifet, Geoff Holmes, Richard Kirkby and Bernhard Pfahringer "DATA STREAM MINING"A Practical Approach May 2011

16. Behrad Bagheri, Hojat Ahmadi, Reza Labbafi," Application of Data Mining and Feature Extraction on Intelligent Fault Diagnosis by Artificial Neural Network and k-Nearest Neighbor" XIX International Conference on Electrical Machines - ICEM 2010, Rome

17. Amir Omidi, Esmaeil Nourani, Mehdi Jalili," Forecasting stock prices using financial data mining and Neural Network" 978-1-61284-840-2/11/$26.00 ©2011 IEEE

18. Baik-Woo Lee, Seong Woon Booh and Kunsoo Shin Samsung Advanced Institute of Technology, Yongin, Gyeonggi-do, 446-712 Korea baikwoo.lee@samsung.com, "Embedded Discrete Passives Technology for Bandage-Type Medical Sensors of E-Healthcare System" 978-1-61284-498 5/11/$26.00 ©2011 IEEE 1325 2011 Electronic Components and Technology Conference

19. Weili Zhou, Jingsheng Liao, Baopu Li, Jingrong Li, "A Family Medical Monitoring System Based on Embedded uC/OS-II and GPRS" Proceeding of the IEEE International Conference on Information and Automation Shenyang, China, June 2012

20. Hilol Kargupta, Ruchita Bhargava, Kun Liu, Micheal Powers, Patrick Blair, Computer Science and Electrical Engineering Department, University of Maryland at Baltimore Country, "VEDAS: A Mobile and Distributed Data stream Mining System for Real Time Vehicle Monitoring", Published in the proceeding of the 2003 International SIAM Data mining conference

21. http://www.steposol.com/stock_feed/get_data.php

22. http://en.wikipedia.org/wiki/K-means_clustering

23. http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

24. http://en.wikipedia.org/wiki/Real-time_data

25. http://www.moneycontrol.com