International Journal of Hybrid Information Technology
Vol.8, No.3 (2015), pp.155-164
http://dx.doi.org/10.14257/ijhit.2015.8.3.15

Challenges and Issues in DATA Stream: A Review

Muhammad Arif [1,2], Khubaib Amjad Alam[1,2] and Mehdi Hussain[1,3]

[1]Faculty of Computer Science and Information Technology, University of Malaya 50603 Kuala Lumpur, Malaysia
[2]Computer Science Department, Comsats Institute of Information and Technology Islamabad Pakistan
[3]School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad Pakistan
arifmuhammad36@siswa.um.edu.my, khubaibalam@siswa.um.edu.my
mehdi141@siswa.um.edu.my

Abstract

Data stream is a continuous, time varying, massive and infinitely ordered sequence of data elements. The streaming data are fast changing with time, it is impossible to acquire all the elements in a data stream. Therefore, each data element should be examined at most once in data streams. Memory usage for mining data stream should be limited due to the new data elements are continuously generated from the streams. It is essential to ensure that newly arrived stream should be immediately available whenever it is requested made this task much challenging and necessary for fraud detection in stream, taking out knowledge, for business improvement and other applications where data arrived in stream. This paper tries to highlight important issues and research challenges of data stream by means of a comprehensive review.

Keywords: *Data Mining, Data Stream, Data Set, Memory, Efficiency, Time Complexity*

1. Introduction

Data streams, consist of unbounded data in continues fashion and it can be regarded as real time data. Continuous data coming from different areas with a high speed and massive size is called data stream. Computer network traffic, ATM, phone conversations, web searches, transactions and sensor data are the real examples of streaming data. Now question arises that how to tackle these data streams and how to get valuable information. Researchers have developed different techniques in this regard. Some techniques depend on the nature of problem, while some algorithms have been developed for time constraint that one pass algorithm some for memory management that is for limited memory to reduce the usage of memory. But focus has been on specific problems related to data streams. In this paper we have discussed issues related to adverse impact of Radio frequency interference in real time streaming discussed, i.e. How to improve the performance of predictive stream learning algorithms, Load shedding problem of join operator, overloading of Real time system, handling unpredictable network stream, exact processing over data streams, mining online data stream, frequent items in data stream, computational issue, frequency count issue, maximum frequent item sets issue, In-core mining of streaming, online closed frequent items issue, classifier selection, frequency estimation over Sliding Windows, memory consumption, close weighted frequent item sets, distributed privacy, high data speed stream and outlier detection in distributed data streams.

ISSN: 1738-9968 IJHIT

2. Application of Data Mining in Crime Detection

This section is aimed at discussion and analysis of existing research on data streams and tries to highlight significant challenges and issues in this domain. In [1] the authors have discussed the issue of power reduction of radio frequency interference (RFI) effect on the high-energy impulses to find it and receive it from different sources. For this the purpose radio astronomy has been used along with an example of Allen Telescope Array (ATA) to collect data. This article further describes how to develop statistical methods which are perfect for removing RFI (Radio frequency interference) from ATA (Allen Telescope Array) data in real time and then implement those methods into real-time setting with the help of developing a prototype Real time streaming data system (RTSDS) is used in this regard. Developing such system need some important considerations which include Operating system, double buffering, ring buffers and programming strategies. Using real time operating system RTOS the amount of time to process for ATA data will take less variability than it will if using standard OS. Another study [2] discusses about how to improve the performance of predictive stream learning algorithms. A framework has been proposed that utilizes pre-quential methods by applying both sliding windows and fading factors mechanism. The authors argues that fading factors is better than sliding windows because of best utilizing of memory that is using less memory and faster. [3] aims at performance improvement of query system and processing efficiency to get best and correct result during real time streaming .For this purpose semantic-base load shedding is used and attributes value of tuples follow by it that is learn attribute value of data stream, classify tuples and then executing to solve the following problems.

1. Load shedding time and anti-load shedding time
2. Amount
3. Location
4. Selecting predicate.

Another study by [4] focuses on an importing issue of overloading of real time DSMS. An algorithm named RLS-EDA has been designed in this regard. [5] Aims at solving the critical issue of memory adaptively inside a system where sliding window queries are running temporary over a continuous data stream. Two techniques have been used, one for the adjustment of window sliding and another for the adjustment of time granularity. These techniques are also compatible with semantics stream and do not crash with query optimization at run time. [6] Introduced an online algorithm to solve the latest issue of mining of frequent tunes and piece of music pattern in musical data and the named that algorithm as BVMDS which uses a two level method to produce candidate pattern set (temporal patterns). An important issue of how to get best efficiency algorithm that must handle not only reliable network stream but as well as unreliable stream has been addressed by [7]. An algorithm has namely PHM show better efficiency in term of unpredictable network as well as growing sizes problem.

Acquisition of exact result of processing of joining streams has been focused by [8]. Authors have introduced an algorithm (EWJ) which produces correct and exact result of joining streams with limited memory.

In [9] the author elaborated an important issue in online data stream mining, for this purpose introduces an algorithm named MRFI-SW it is based on apriori algorithm and have three phases window initialization, sliding and mining frequent item sets. But one pass algorithm so it mining frequent stuffs from data stream of sliding window on transaction base in accurate form and give correct result as well as it consumes less memory as well as time when compare to other algorithm like SWFI-stream algorithm.[10] This paper is about to solve the critical issue of frequent items of data stream and how to tackle this issue so for this purpose author introduce tow techniques called TCAM-conscious Lossy counting and TCAM- conscious space saving but TCAM-CLC which is faster than TCAM-SS and also discuss to integrate NPU and TCAMs have tremendous performance for high data stream. In this [11] the author focus on an important issue of real time data coming online and how to control online data means their computational issue like limitless data streaming and speed so for this purpose introduce an algorithm called which is one pass algorithm using top-k paths for mining for traversing pattern where k is the demand of paths want to traverse and to keep information in best way using path-k forest so it is a best algorithm to handle data stream in one pass. This paper [12] focus on an importing issue of frequent frequency which change pattern when user specify some threshold value that is by addition and removal of items in dynamic data stream so how to identify this frequent frequency which in order to change pattern so for this purpose introduce an algorithm which solve this problem very efficiently by taking 2 dynamic data stream and produce max or minimum change depend on threshold. This paper [13] focus on an importing issue of how to mine maximum frequent pattern sets from huge data of data streams which is sparse as well as for dense one so for this purpose introduce an algorithm named DSM-MFI tackle this issue in a best way that is make a data structure called SFI-forest which store latest and frequent items sets and this algorithm is based on sliding windows. This paper [14] focus on an importing issue of computing the frequency count only assigned by user in sliding window of real time data stream so to solve it by designing an algorithm named approximation algorithm which working dynamically that is finding frequency count in sliding windows dynamically and to keep and maintain the recent frequency of items in data structure called exponential histogram and delete useless element from data structure by performing compression mechanism periodically. This paper [15] focus on an importing issue to mine frequent item sets from huge data sets for data stream so for this purpose design an algorithm named AICFI, this algorithm very better than other algorithm (RLS-U) and used the idle time O(N) of system in a very better way and produce throughput larger than other algorithm comparably and does not require out of core (summary structure) ,and support large data and low level support. Another study by [16] focuses on an importing issue of outlier detection in data stream so for this purpose introduce an algorithm named ODABK this algorithm based on classification method that is on K-NN that is on statistics based and instance base store all training samples so its detection rate of outlier is high as well as take less time compare to other algorithm (LOADED). Another study by [17] focuses on an importing issue of keeping information of music which is in continuous streams that is to keep latest information of music which is necessary in frequent temporal fashion of pattern through online system so for this purpose design an algorithm named FTP-MQS so it is one pass algorithm and its working consists of three phases 1st take to initialize the window, 2nd sliding of window, 3rd one and last phase to generate frequent pattern temporally so it is a best algorithm in term of efficiency to mine frequent pattern temporary for music stream. In this [18] the author focus on an importing issue of to mine the closed frequent pattern of streaming data in data streaming through online Streams. So for this purpose introduce an algorithm named new-movement so working of this algorithm to move items in sensitive sliding window based on bit-sequence using a data structure called newcet prefix based tree keep the recent information that is oldest one and append new coming information therefore it is better algorithm in term time and memory compare to other algorithm like movement algorithm. This study [19] paper focuses on an importing issue of that how to mine the closed frequent pattern of streaming data through online Streams with sliding windows. So for this purpose introduce an algorithm named MFI-Trans SW so working of this algorithm in sensitive sliding window having three phase 1stinitialize the window, 2nd widow sliding and 3rd and last phase is generate frequent item sets therefore it is better algorithm in term time and memory compare to other algorithm. This paper focus [20] on an importing issue to mine data stream for learning process and to classify according to their desire place that is how to select the most appreciate classifier so to tackle large data and to group them according to correct position in database so for this purpose introduce algorithm named Dynamic incremental SVM learning so it's working is during initialization classifier number is constant if similarity in attributes found then add to training data otherwise into another dataset of training data after exam all data then assemble to test data stream. This paper focus [21] on an importing issue that in online data stream how to find frequent vibrated points means items so for this purpose introduce two algorithm MVI and FMVI now MVI cover each window vibrated items but vibrated period is concerned due to this reason introduce MFVI to tackle this problem it have track list which will list vibrated items in track list and have starting point, count for points in the list so correctly find change support in count and quick report. Another study by [22] focuses on an importing issue that how to find estimate almost near and closest frequency in DataStream through sliding window So for this purpose introduce an algorithm named deterministic algorithm so working of this algorithm is designed snapshots which divided into two snapshots one is complete snapshot and another one is partial snapshots and to reduce space and time of processing designed a data structure called doubly link list so all points inserted to determine tail or head of list S that's why by deleting items takes the processing time O(1)time and take space O(I/ε). Another study by [23] improves the performance of sensor network introduce an improved algorithm which decrease reconstruction and reduce hidden variable so this algorithm based adaptive filtering technique and working in three phase 1st training weight using training step to get weights vectors, 2nd orthogonalization to check each iteration on weight vector,3rd and find hidden variables to find principles components that how much needed so less vibration in unstable phase and strong correlation in data streams thus it is best algorithm for sensor network. This paper [24] focus on an importing issue that how to use less memory in data stream, it is a big problem for researcher how to reduce memory So for this purpose introduce an algorithm named MFPBI as well as a data structure called IFP_Stream so working of IFP_Stream is based on CP_Tree to reduce the usage of memory and MFPBI algorithm based on IFP_Stream to determine frequent pattern at multiple time graininess over DataStream. Therefore it is better algorithm in term time and memory compare to other. Another study by [25] focuses on an importing issue to mine the closed frequent pattern of streaming data in data streaming and to remove the drawbacks of previous algorithms which is based on Apriori that is to remove redundancy of nodes and useless test for meaningless item-sets So for this purpose introduce an algorithm named MFCIDS working of this algorithm is based on data structure called FULL-CET and according to it those nodes or points are sibling both belong from a single parent so by this way processing time decrease and support large data sets and support minimum threshold that's why MFCIDS have best efficiency than other algorithms like MOVEMEN and A-MOVEMENT.

An importing issue in data stream is instant recognition of outlier pattern. [26] Introduced an algorithm based on association rules mining. This algorithm utilizes prefix-tree that regularly monitors the frequent items and traversal tree is used to find all the association rules.

Another importing issue in data streams is how to mine frequent pattern which have closed weight in data stream. An algorithm DS _ CWFP by [27] is based on sliding windows and a data structure called DS _ CWFP. DS_CWFP structure uses composite structure of compressed tree. This algorithm utilizes divide and conquer method to get closed weighted frequent items in data stream. [28] Introduced a protocol for efficiency in communication and data security. This protocol is based on an IACA making it appropriate to find center cluster than other algorithms like k-mean. [29] Introduced an algorithm named FIKOCFram to tackle high speed data of data streams. [30] Investigated the issues of the outlier detection in distributed data streams. An algorithm namely n-IncLOF is presented in this regard which is based on LOF and extension of it working dynamically using sliding window model and adjusts the n-threshold value automatically when there is a change in sliding window by outliers change number. This study has used standards deviation to get more accurate detection of anomaly as well as to reduce bugs.

Table 1. Analysis of Algorithms and Techniques

Ref No.	No of algorithms	Algorithm	memory	Time complexity	Data set	tools	Methods used
[1]	1	Computational Algorithm	1GB	O(n)	Synthetic data	ATA(Allen Telescope Array)	Grid based method
[2]	2	VFDT ,The Page Hinkley Algorithm.	2GB	O(1)	Non – stationary data	McNemar Test, Page Hinkley test	Pre-quential method
[3]	0	×	1GB	O(n)	Real Time	×	Time statistical, sliding windows
[4]	2	RLS-EDA, RLS-U	1.5GB	O(n2)	Real time	×	×
[5]	0	×	×	O(n)	Real time	×	window size adjustment, Time Granularity Adjustment
[6]	2	BVMDS and FTP-Stream algorithms	2GB	O(n)	Real time	×	Based on candidates to k-cond
[7]	2	PHMJ, RPJ	×	O(n2)	Real time	×	MM-Join and external join
[8]	4	EWJ,XJoin,PMJ,HMJ	×	O(n)	Real time	×	Age base and frequency model
[9]	2	MRFI-SW, SWFI-stream	×	O(n2)	Real time	×	Sliding window model technique
[10]	0	×	×		Real time	×	TCAM-conscious Lossy counting, TCAM-conscious space saving
[11]	1	DSM-TKP	×	O(n2)	Real time	×	Traversing method
[12]	1	MFC-append	×	×	Real time	×	×
[13]	1	DSM-MFI	×	×	Real time	×	Sliding window technique
[14]	2	Approximation algorithm, Arasu's algorithm.	×	×	Real time	×	Sliding window technique
[15]	1	AICFI	2.5MB	×	Real and Synthetic	×	×
[16]	1	ODABK	2GB	O(n2)	Real world data	×	Classification method
[17]	1	FTP-MQS	2.1MB	O(n)	Synthetic data	Visual c++	Sliding window technique
[18]	1	New-movement	15MB	O(n)	Synthetic data	C++,visual c++.net	sensitive sliding window
[19]	1	MFI-TransSW	20MB	×	Synthetic data	Visual c++	Sliding window technique
[20]	1	Dynamic incremental SVM learning algorithm	30MB	O(n)	Synthetic data	Simulator tool	Classification technique
[21]	2	MVI,MFVI	×	×	Syndata	Sim tool	Window modeling
[22]	1	?-approximate algorithm	×	O(1)	Synthetic data	×	Sliding window technique
[23]	1	AIDSPA	×	O(n)	Synthetic data	×	Adaptive filtering technique
[24]	1	MFPBI	25mb	O(1)	Synthetic data	C#, Visual studio	CP_Tree based
[25]	1	MFCIDS	1GB	×	Synthetic data	C++	FULL-CET technique
[26]	1	association rules mining algorithm	2GB	×	Synthetic data	DEV C++	Association rule
[27]	1	DS _CWFP	2GB	×	Both real,syn	C++	Sliding window technique
[28]	1	IACA	×	O(log n)	Both Syn,real	×	Yao,s protocol technique
[29]	1	FIKDR, FIKOCFram	2GB	O(n)	Synthetic data	Dev c++	IKPCA Technique
[30]	1	n-IncLOF	×	O(n)	Synthetic data	Simulator tool	Sliding window model

Table 2. Core Competency of the Papers

Ref no	PROBLEM	SOLUTION	ADVANTAGE
[1]	Adverse impact of Radio frequency interference in real time streaming.	Develop a Real time streaming data system(RTSDS) to reduce the impact of RFI in real time streaming	Removing RFI from real time streaming data.
[2]	How to improve the performance of predictive stream learning algorithms?	Predictive sequential methods that is prequential method by applying fading factors to improve performance	Design a framework for predictive stream learning algorithm
[3]	Load shedding problem of join operator?	Semantic-base load shedding technique.	Best design model to get correct result of Query system
[4]	Issue of overloading of Real time system?	Develop an RLS-EDA algorithm to solve the problem.	Buffers more tuples to execute instead of dropping them
[5]	Memory adaptivity inside a system for continues data stream.	Used two techniques 1. window size adjustment 2. Time Granularity Adjustment	Solve collision problem of query optimization at run time and compatible with semantics stream
[6]	How to mining Frequent tune pattern from music data?	To solve this problem introduce an algorithm called BVMDS which efficiently solve this problem	BVMDS based on 3-condidate to k-cond due to which it is more efficient and faster than other algorithm
[7]	How to handle unpredictable network streams?	Introduce an algorithm named PHMJ which solved it efficiently and effectively.	Not only handle reliable network but also handle unreliable network
[8]	How to get exact processing and sliding windows among data streams?	Design a framework and introduce an algorithm (EWJ) to solve the problem	To get exact processing of joining streams with limited memory
[9]	How to mining online data stream?	Introduce an algorithm (MRFI-SW) which solves this problem effectively and efficiently.	One pass algorithm consumes less memory and time
[10]	How to get frequent items in data stream?	Best technique for it is TCAM-conscious Lossy counting.	Getting frequent elements without collision.
[11]	Issue of computational issue during online web clicking	DSM-TKP(algorithm)	Reduce time for large data using top-k path technique
[12]	Issue of online frequent frequency which changes items in data stream?	Develop an MFC-append algorithm to solve the problem.	Identify change frequency pattern dynamically.
[13]	Issue of how to mine maximum frequent itemsets over data stream?	Develop a DSM-MFI algorithm to solve the problem.	Mining maximum frequent items from large datasets using SFI-forest data structure.
[14]	Issue of frequency count must not be less than specified threshold by user.	Develop a one pass-approximation algorithm to solve the problem.	Dynamically determine frequency count.
[15]	Issue of In-core mining of streaming data	Develop an algorithm to AICFI solve the problem.	In-core mining for high data stream.
[16]	Issue of outlier detection in data stream?	ODABK algorithm.	Required less time and high detection rate
[17]	Issue of latest information of music of continuous stream	FTP-MQS algorithm to solve the problem.	Keep latest information of online music query streams.
[18]	How to mine online closed frequent items over data stream efficiently?	A new- movement algorithm to solve the problem.	Get efficiency in term of time and memory to mine closed frequent itemsets over DataStream.
[19]	How to mine online frequent items over data stream efficiently?	Develop a MFI-TransSW algorithm to solve the problem.	Efficient in term of time and memory to mine frequent itemsets over sliding window in DataStream
[20]	How to deal with large data in learning data how to select appropriate classifier in data stream and its classification?	Dynamic incremental SVM learning algorithm to solve the problem.	Selection of suitable classifier to learn according to the characteristic of datasets.
[21]	How to find Frequent Vibrated Items over Online Data Streams?	MVI,MFVI algorithms to solve the problem.	monitor and Detect variation of items between two data streams.

[22]	How to find Frequency Estimation in data stream over Sliding Windows	Develop ε-approximate algorithm to solve the problem.	Take less time and space for estimation of frequency over sliding Window.
[23]	How to improve the processing of sensor networks?	IDSPA for solution.	Get efficiency in term of processing over data streams.
[24]	How to reduce the consumption of memory in data stream?	MFPBI algorithm to solve the problem.	Get efficiency in term of time and reduce usage of memory to find frequent pattern in DataStream.
[25]	How to mine frequent itemsets which are closed in data stream?	MFCIDS algorithm to solve the problem.	Remove redundancy of nodes and speedup processing time.
[26]	How to find instant recognition of outlier pattern in data stream?	association rules mining algorithm to solve the problem.	Get efficiency in term of time and memory to mine closed frequent itemsets over DataStream.
[27]	How to mine closed weighted frequent Pattern in data stream efficiently?	DS _ CWFP algorithm to solve the problem.	Determine frequent items based on closed weight in DataStream.
[28]	How to keep distributed privacy for clustering of data streams.	Developed a protocol for solution.	Get efficiency in term security and communication over distributed data streams.
[29]	How to tackle high data speed stream?	Developed a FIKOCFram technique for solution.	Reduce the limit dimensionality reduction.
[30]	How to detect outliers in distributed data streams accurately?	Developed an n-IncLOF algorithm for solution	Dynamically find local outliers over data streams.

3. Conclusion and Future Work

From this review we concluded that data streams still have to face certain challenges *i.e.*, time constraint, space limitation high dimensional data. Numerous methods have been introduced to tackle some of these issues. Algorithms to handle noise are not fully capable to address the noise issues of for high dimensional. Space limitation of data stream signifies that data stream volume is large and it is impossible to store and process all the data. Some solutions have divided the space into grids and just retain and maintain the gridded data and new data is just mapped with specified grid and the remaining data is excluded. Finally we discuss the time constraint that is time limits how to handle data stream which coming with high speed. One pass algorithm and techniques are significantly important in this regard having lowest time complexity and fast processing time but they are not able to cope with other challenges of data stream. If we want to get more efficiency and effectiveness in term of limited time, space limitation and high dimensional data, researcher should developed improved techniques of integration that is integration cluster and classification techniques which would emphasize on current issues in data streams not just to overcome one type of constraint but to resolve maximum issues combined by one technique and it is possible to solve all the issues in data streams by one technique up to some extent but need more work on it so integration of classification and cluster techniques would be a best choice in future to resolve the current issues in data streams.

References

[1] S. Michalak, A. DuBois, D. DuBois, S. V. Wiel and J. Hogden, "Developing systems for real-time streaming analysis", Journal of Computational and Graphical Statistics, vol. 21, no. 3, **(2012),** pp. 561-580.
[2] J. Gama, R. Sebastião and P. P. Rodrigues, "Issues in evaluation of stream learning algorithms", In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, **(2009)** June, pp. 329-338.
[3] L. Ma, Q. Zhang, K. Wang, X. Li and H. Wang, "Semantic load shedding over real-time data streams", In Computational Intelligence and Design, 2008. ISCID'08. International Symposium on IEEE, vol. 1, **(2008)** October, pp. 465-468.
[4] L. Ma, X. Li, Y. Wang and H. A. Wang, "An Approach to Handle Overload in Real-Time Data Stream Management System", In Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on IEEE, vol. 3, **(2008)** October, pp. 3-8.

[5] M. Cammert, J. Kramer, B. Seeger and S. Vaupel, "An approach to adaptive memory management in data stream systems", In Data Engineering, ICDE'06 Proceedings of the 22nd International Conference on IEEE, **(2006)** April, pp. 137-137.
[6] H. F. Li, M. H. Hsiao and H. S. Chen, "Efficiently mining frequent patterns in recent music query streams", In Multimedia and Expo, 2008 IEEE International Conference on IEEE, **(2008)** June, pp. 1269-1272.
[7] G. Chen, G. Li, B. Yang, X. Tang and H. Chen, "Progressive Hash-Merge Join Algorithm. In Computational Intelligence and Industrial Application", PACIIA'08. Pacific-Asia Workshop on IEEE, vol. 2, **(2008)** December, pp. 117-121.
[8] A. Chakraborty and A. Singh, "Processing exact results for sliding window joins over time-sequence, streaming data using a disk archive", In Intelligent Information and Database Systems, ACIIDS 2009 First Asian Conference on IEEE, **(2009)** April, pp. 196-201.
[9] J. D. Ren and K. Li, "Online data stream mining of recent frequent item sets based on sliding window model" In Machine Learning and Cybernetics, 2008 International Conference on IEEE, vol. 1, **(2008)** July, pp. 293-298.
[10] N. Bandi, A. Metwally, D. Agrawal and A. El Abbadi, "Tcam-conscious algorithms for data streams", In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on IEEE, **(2007)** April, pp. 1342-1344.
[11] H. F. Li, S. Y. Lee and M. K. Shan, "DSM-TKP: Mining top-k path traversal patterns over web click-streams", In Web Intelligence, Proceedings The 2005 IEEE/WIC/ACM International Conference on IEEE, **(2005)** September, pp. 326-329.
[12] H. F. Li and S. Y. Lee, "Single-pass algorithms for mining frequency change patterns with limited space in evolving append-only and dynamic transaction data streams", In e-Technology, e-Commerce and e-Service, 2004. EEE'04. 2004 IEEE International Conference on IEEE, **(2004)** March, pp. 215-222.
[13] H. F. Li, S. Y. Lee and M. K. Shan, "Online mining (recently) maximal frequent item sets over data streams", In Research Issues in Data Engineering: Stream Data Mining and Applications, RIDE-SDMA 2005 15th International Workshop on, **(2005)** April, pp. 11-18.
[14] G. Nie and Z. Lu, "Approximate frequency counts in sliding window over data stream", In Electrical and Computer Engineering, 2005. Canadian Conference on IEEE, **(2005)** May, pp. 2232-2236.
[15] R. Jin and G. Agrawal, "An algorithm for in-core frequent itemset mining on streaming data", In Data Mining, Fifth IEEE International Conference on IEEE, **(2005)** November, pp. 8.
[16] F. Han, Y. M. Wang and H. P. Wang, "Odabk: An effective approach to detecting outlier in data stream", In Machine Learning and Cybernetics, 2006 International Conference on IEEE, **(2006)** August, pp. 1036-1041.
[17] H. F. Li, C. C. Ho, M. K. Shan and S. Y. Lee, "Online mining of recent music query streams", In Multimedia and Expo, 2006 IEEE International Conference on IEEE, **(2006)** July, pp. 1985-1988.
[18] H. F. Li, C. C. Ho, F. F. Kuo and S. Y. Lee, "A new algorithm for maintaining closed frequent item sets in data streams by incremental updates", In Data Mining Workshops, 2006. ICDM Workshops 2006 Sixth IEEE International Conference on IEEE, **(2006)** December, pp. 672-676.
[19] H. F. Li, C. C. Ho, M. K. Shan and S. Y. Lee, "Efficient maintenance and mining of frequent item sets over online data streams with a sliding window", In Systems, Man and Cybernetics, 2006. SMC'06 IEEE International Conference on IEEE, vol. 3, **(2006)** October, pp. 2672-2677.
[20] Z. W. Li, J. Yang and J. P. Zhang, "Dynamic incremental SVM learning algorithm for mining data streams", In Data, Privacy, and E-Commerce, 2007. ISDPE 2007. The First International Symposium on IEEE, **(2007)** November, pp. 35-37.
[21] G. Lee and Q. T. Chen, "A single pass algorithm of finding frequent vibrated items over online data streams", In Digital Information Management, 2007. ICDIM'07. 2nd International Conference on IEEE, vol. 1, **(2007)** October, pp. 206-211.
[22] L. Zhang and Y. Guan, Frequency estimation over sliding windows. In Data Engineering, ICDE IEEE 24th International Conference on IEEE, **(2008)** April, pp. 1385-1387.
[23] W. Cheng and H. Shi, "An improved data streams processing algorithm for sensor networks based on training weights", In Computer Engineering and Technology (ICCET), 2010 2nd International Conference on IEEE, vol. 3, **(2010)** April, pp. V3-606.
[24] J. D. Ren and Y. L. Wang, "Mining Frequent Patterns Based on IFP_Stream over Data Stream", In Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on IEEE, **(2008)** June, pp. 548-548.
[25] G. Mao, X. Yang and X. Wu, "A new algorithm for mining frequent closed item sets from data streams", In Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on IEEE, **(2008)** June, pp. 154-159.
[26] L. J. Kao and Y. P. Huang, "Association rules based algorithm for identifying outlier transactions in data stream", In Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on IEEE, **(2012)** October, pp. 3209-3214.
[27] W. Jie and Z. Yu, "An efficient algorithm for mining closed weighted frequent pattern over data streams", In Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on IEEE, **(2012)** June, pp. 153-156.

[28] G. Jagannathan, K. Pillaipakkamnatt and D. Umano, "A secure clustering algorithm for distributed data streams", In Data Mining Workshops, ICDM Workshops Seventh IEEE International Conference on IEEE, **(2007)** October, pp. 705-710.
[29] W. Feng, Z. Yan, L. Ai-ping and W. Quan-Yuan, "Online classification algorithm for data streams based on fast iterative Kernel principal component analysis", In Natural Computation, 2009. ICNC'09. Fifth International Conference on IEEE, vol. 1, **(2009)** August, pp. 232-236.
[30] K. Gao, F. J. Shao and R. C. Sun, "n-INCLOF: A dynamic local outlier detection algorithm for data streams", In Signal Processing Systems (ICSPS), 2010 2nd International Conference on IEEE, vol. 2, **(2010)** July, pp. V2-179.

Authors

Muhammad Arif is a PhD student at Faculty of CS and IT, University of Malaya. Currently he is working on Medical image Processing. His research interests include image processing, E learning, Artificial intelligence and data mining. He joined UM as a Bright Spark Scholar in September 2013 for the period of 3 years. Before this he completed masters and bachelor degrees in Pakistan. He received his BS degree in Computer Science from University of Sargodha, Pakistan in 2011. He obtained his MS degree in Computer Science from COMSATS Islamabad 2013 Pakistan.

Khubaib Amjad Alam is a Ph.D. candidate at Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, under the prestigious Bright Spark Fellowship. He received Master of Computer Science degree in 2013 from Comsats Institute of Information Technology, Pakistan. His current research interests are Business Process Management (BPM); Service oriented architecture (SOA), Web services, distributed systems, Change management, Maintenance and Evolution issues of Enterprise level software systems and engineering applications of soft methodologies.

Mehdi Hussain received his BS degree in Computer Science from The Islamia University Bahawalpur, Pakistan in 2005. He obtained his MS degree in Computer Science from SZABIST Islamabad 2011 Pakistan. He has 8 years working experience in a renewed Software House (Streaming Networks (private)). He recently selected as funded scholar at National University of Science and Technology (NUST) under faculty development program 2014. Presently he is research scholar at University of Malaya, Malaysia. Research interests are multimedia security, steganography, data mining. He can be reached at *mehdi141@hotmail.com*.