http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (11): 1534-1543, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.1534.1543

Mining Data Generated by Sensor Networks: A Survey

Azhar Mahmood, Ke Shi and Shaheen Khatoon
School of Computer and Applied Technology,
Huazhong University of Science and Technology (HUST), Wuhan, China

Abstract: Sensor Networks (SNs) produces huge amount of data which offer promising prospect for the application of data analysis techniques to extract useful information for end users. Knowledge discovery from sensor data is an emerging research area due to their wide range of potential application of crucial importance to our society. Since sensors nodes are resource constraints in term of memory, communication bandwidth and battery power, therefore data mining techniques required to meet these constraints while applying to SN data. This study presents the classification and evaluation of Data Mining (DM) methods applied on sensor networks data, leading to more efficient techniques for resources optimization. Plenty of research work done on sensor networks but there is none of such study exist that focuses exclusively on data mining methods. The main contribution in this study is classifying and comparing the published data mining approaches in sensor networks and highlights the challenges as well as future trends in applying data mining methods on SNs data.

Key words: Data mining, sensor network, stream data, association rule mining, clustering, classification, sequential patterns, frequent pattern

INTRODUCTION

In information world data mining is gaining a rapid popularity because of its aim towards extraction previously unknown and potentially valuable information (such as knowledge rules, constraints, regularities) from large amounts of data where the data can be stored in databases, data warehouses, OLAP (On Line Analytical Process) or other repository information (Halkidi, 2000; Han and Kamber, 2001). It involves integration of techniques from multidiscipline such as database technology, statistics, machine learning, neural networks, information retrieval etc. (Fayyad *et al*., 1996; Hand *et al*., 2001). Data mining is used to solve wide range of problems in various applications such as product analysis, understanding consumer research marketing, e-commerce, demand and supply analysis, direct marketing, health industry, e-commerce, stocks and real estates, Customer Relationship Management (CRM), telecommunication industry and financial sector investment. Besides these general applications, nowadays data mining is also applied on domain specific applications such as biomedical and bioinformatics (Ling, 2008; Bajcsy *et al*., 2005), education (Ranjan and Khalil, 2008) software engineering (Khatoon *et al*., 2011a, b), fraud detection (Thiruvadi and Patel, 2011), distributed (Hui, 2011; Yao *et al*., 2012). The data set of these applications is different from general ones which call for special mining requirements.

With the advancement in sensors' technology sensor networks are increasingly finding its applications in many domains such as, habitat monitoring (Rozyyev *et al*., 2011) object tracking (Chauhdary *et al*., 2009), environment monitoring (Sabri *et al*., 2011; Lee and Chen, 2008; Kumar, 2011), military (Arampatzis *et al*., 2005; Yicka *et al*., 2008). An enormous amount of data has been collected from these applications. Data mining techniques can be applied to this data for solving various problems. The traditional architecture in SNs deployed over an area of interest to sample data and send the results to a data-collection point, i.e., base station (BS) through an onboard radio transceiver. Such technological development has encouraged researchers to visualize the limited capabilities of the individual sensors in a large scale network that can operate unattended (Estrin *et al*., 1999; Pottie and Kaiser, 2000; Akyildiz *et al*., 2002; Chong and Kumar, 2003; Wang *et al*., 2003a). In many applications SNs are deployed on hostile and difficult to access location with limited energy, storage, computational power and bandwidth. Development of algorithms that consider the characteristics of sensor networks such as energy and computation constraints, network dynamics etc. constitutes an active area of current research (Gao *et al*., 2010; Han and Gao, 2008).

Corresponding Author: Azhar Mahmood, School of Computer and Applied Technology,
Huazhong University of Science and Technology (HUST), Wuhan, China

The extraction of useful knowledge from raw sensor data is a difficult task and conventional tools might not be able to handle the massive quantity, the high dimensionality and the distributed nature of the data. For such reasons, in recent years a great interest emerged in the research community in applying data mining techniques to the large volumes of sensor data.

In this study, SNs techniques with respect to sensor data mining process are critically analyzed. As a result a qualitative analysis is presented in tabular form. This work not only provides significant contributions to the SNs research, also exposes how challenging to compare different techniques due to the unavailability of complete data set used to evaluate the study. Since most of the studies have only reported the results without the availability of implemented tools therefore it is difficult to compare the existing techniques quantitatively.

CLASSIFICATION OF DATA MINING TECHNIQUES FOR SENSORS' DATA

Data mining community has observed that extracting knowledge from SNs data imposes two kinds of challenges. At one hand, network side processing techniques require real-time analysis methodologies and systems to handle dynamic data streams or events; on the other hand centralized processing through high end computing is required for generating off-line predictive insights which facilitates real-time analysis. To address these challenges there are primarily two approaches for mining SNs data are adopted in literature. First, data from SNs is directed to central location where the entire processing take places at central side called centralize data models. Second, instead of sending the raw data to the central site, sensor nodes use their processing abilities to locally carry out mining processing and transmit only the required and partially processed data called local model. This paper attempts to characterize and classify the related approaches whether it applies mining techniques on network side or central side.

Mining techniques at network side: This section discusses the techniques that use the distributed nature of sensors and apply distributed mining algorithms on sensor data. In these techniques most of the work is pushed to the sensors themselves to build local models. The techniques are classified under following types on the basis of major data mining task it used for finding meaningful patterns.

Frequent pattern mining: Loo *et al.* (2005) proposed a framework for extracting association rules from sensor network. They proposed a model that analyzes the representation of sensor data streams for mining by proposing an Interval-List (IL). The analysis of IL based presentation of stream data showed favorable results using synthetic data set.

Romer (2006) proposed an in-network data mining technique to discover frequent patterns of events with certain spatial and temporal properties. Major issues in this approach are memory consumption of itemset discovery algorithms and the communication overhead of event collection.

Boukerche and Samarah (2007) proposed a distributed data extraction methodology for extracting association rules from SN by attaching a storage device to each node. Major issues in this methodology are increase in cost for node buffer and also delay in crucial messages in case of high support value.

Boukerche and Samarah (2008) proposed sensor association rules in which the event-detecting sensors are the main objects of the rules regardless of their values. A data structure called Positional Lexicographic Tree (PLT) is maintained to store the sensor's event-detecting status. The issue is multi database scans and the additional PLT update during mining limits the efficient use of this approach in handling SN data.

Chong *et al.* (2008) proposed a model that finds strong rules from sensor readings and use these learnt rules as a triggers to control sensor network operations or supplement sensor operations. To mine the rules *A. priori* is modify to count the number of transactions that are frequent instead the itemset with in transaction and transaction are processed in a batches $b_1, b_2, \ldots, b_X$.

Sequential pattern mining: Tseng and Lin (2007) and Tseng and Lu (2009) proposed an object tracking strategy named Multi Level Object Tracking (MLOT) to discover sequential patterns in object tracking sensor networks (OTSNs) by mining the movement log in sensor networks. A multi-level hierarchical structure is adapted by using the clustering mechanism that represents the hierarchical relations among sensor nodes to achieve the goal of keeping track of moving objects in a real-time manner. Through experimental evaluation of various simulated conditions, the proposed method is shown to deliver excellent performance in terms of both energy efficiency and timeliness.

Samarah *et al.* (2011) proposed Prediction-based Tracking technique by using Sequential Patterns (PTSPs). This technique helps to predict the future location of a moving object with the minimum number of sensor nodes while keeping the other nodes in sleep mode. The simulation results compared with other predictive techniques shows that PTSPs is much efficient in energy consumption and missing rate recovery mechanism.

Esposito *et al.* (2011) presented a relational framework to identify the hidden frequent temporal correlations between sensor nodes. The framework exploits the relational language to describe the temporal evolution of a sensor network along with contextual information. By taking into account interval-based temporal data along with contextual information about events it discovers interesting and more human readable patterns.

Clustering: Dai *et al.* (2008) proposed a cluster formation protocol which consumes less energy and achieves longer network lifetime called Algorithm of Cluster head Election by counting (ACE-C). Sensor nodes transmit information only to cluster heads and then heads will communicate the aggregated information to the center. Heinzelman *et al.* (2000), proposed a popular clustering protocol for SNs called Low Energy Adaptive Clustering Hierarchy (LEACH). It forms clusters based on the received signal strength and uses the Cluster Head (CH) nodes as routers to the base-station. LEACH forms clusters by using a distributed algorithm, where nodes make autonomous decisions without any centralized control. LEACH Centralized (LEACH-C) (Heinzelman *et al.*, 2002), improves LEACH in terms of data delivery by the use of an overlooking base station to evenly distribute cluster head nodes throughout the sensor network. The base station does this by computing the average node energy and allows only nodes over the average node energy to become cluster heads for the current iteration.

Younis and Fahmy (2004), proposed energy-based distributed clustering protocol named Hybrid Energy-Efficient Distributed clustering (HEED) which considers a hybrid of energy and communication cost when selecting Cluster Heads (CHs). The protocol probabilistically selects several nodes as CHs according to their residual energy. In HEED, each node is mapped to exactly one cluster and can directly communicate with its CH.

Taherkordi *et al.* (2008) proposed a communication efficient distributed protocol for clustering sensory data. A distributed version of *k*-means clustering algorithm is adopted which reduces communication transmission, time and power consumption of sensor nodes. Major issues are extra memory for cluster head and computation power for summarization of data before transmitting to sink. Guo *et al.* (2009) proposed H-Cluster a distributed algorithm to cluster sensory data. The results show that H-Cluster algorithm is much efficient in energy conservation and the quality of cluster data in small SN. Beyens *et al.* (2005) proposed a cluster-based architecture for wireless sensor networks. A prediction model is executed to predict the spatiotemporal correlation of cluster heads. Several centralized clustering techniques, such as *k*-means and hierarchical clustering (Han and Kamber, 2001), used in this method.

Yoon and Shahabi (2007) presented the Clustered Aggregation (CAG) algorithm that forms clusters of nodes sensing similar values. However, CAG is a lossy clustering algorithm (most sensory readings are never reported) which trades a lower result precision.

Classification: McConnell and Skillicorn (2005), presented a classification framework for building and deploying predictors in sensor networks. The proposed framework is useful especially for sensor networks limited power resources, computation, bandwidth and data confidentially problems. Major issues in this framework are we need an expensive CPU on each sensor node for computation and building local predictive model and also extra memory is required to store local predictive model.

Sharma *et al.* (2011) proposed a methodology for classifying the sensors data by using a Nearest Neighbor Trajectory Classification (NNTC). When a request to classify a query vector is made the closest training vector(s), according to a distance metric are located. The classes of these training vectors are used to assign a class to the query vector. The nearest-neighbor method predicts the class of a test example.

Malhotra *et al.* (2008) proposed a distributed classification scheme for classifying the audio signals from wireless sensor network. A distributed cluster-based algorithm for detection and classification of vehicles has been proposed. Two approaches are proposed: first combines extracted features and the second combines' individual decisions. Classification results using decision fusion and a Maximum Likelihood (ML) classifier led to the best results.

Flouri *et al.* (2006a, b) proposed a distributed and incremental techniques for training Support Vector Machines (SVMs) in sensor networks. Two distributed algorithms for training a SVM are applied to the classification problem in WSN. The challenge for these SVM formulations for this application is their computational complexity.

Mining techniques at central side: Plenty of centralized techniques are proposed in literature which is classified under major data mining method use to discover meaningful patterns.

Frequent pattern mining: Le-Gruenwald (2005), proposed a data model to store and updates the data generated from sensor networks. They adopt a centralized extraction methodology where all the data is collected at sink for further analysis. Data Stream Association Rule Mining (DSARM) framework is proposed to identify the missed readings that result from the loss and corruption of messages.

Jiang (2007) and Jiang and Gruenwald (2007) proposed a data estimation technique, called CARM (Closed Itemsets based Association Rule Mining) which can derive the most recent association rules between sensors based on the current closed itemsets in the sliding window. The technique is based on the Closed Frequent Itemsets mining algorithm in data streams called CFI-stream (Jiang and Gruenwald, 2006). It maintains an in-memory data structure, called Direct Update (DIU) to store closed itemsets. The results show the algorithm achieve time and space efficiency. However, the algorithms estimates the missing data according to the frequent patterns which are pre-computed based on the existing data. If the pattern containing the missing data does not appear in the frequent patterns, the missing data cannot be estimated.

Chi *et al.* (2004) proposed an algorithm called Moment to mine closed frequent itemsets over a data stream sliding window. Moment stores much more information other than the current closed frequent itemsets which consumes much memory, especially when the support threshold is low. Furthermore, the exploration and node type checking are time consuming.

Giannella *et al.* (2003) proposed FP-tree-based algorithm called FP - Stream to mine frequent itemsets at multiple time granularities over data streams. A novel tilted-time window is proposed to calculate the frequent patterns for the most recent transactions. A number of experiments are conducted to prove the algorithm's efficiency.

Tanbeer *et al.* (2009), proposed a tree-based data structure called Sensor Pattern Tree (SP-Tree) to generate association rules from SNs data with one database scan. The main idea of proposed approach is to obtain the frequency of all event-detecting sensors' data and construct a SP-Tree. Results shows SP-Tree based approach is less time consuming because it constructs the tree by scanning the database only once. Moreover, the mining phase of SP-tree is highly efficient due to the frequency-descending tree structure.

Sequential pattern mining: Cook *et al.* (2003) presented MavHome an intelligent home which perceives the state of the home through sensors and acting upon the environment through device controllers. An important component of proposed architecture is the ability to make decisions based on predicted activities. A key disadvantage is the fact that the entire action history must be stored and processed off line which is not practical for large prediction tasks over a long period of time.

Wu *et al.* (2001) proposed an algorithm for mining sequential alarm patterns from the alarm data of Global System for Mobile Communication (GSM). By specifying the referent urgent window d in sequential alarm pattern it predicts and controls the alarms. For example, if d is set as six hours, the sequential alarm pattern (a, b, c) indicates that a, b and c happen in order and that the time interval between a and b and between b and c is less than 6 h.

Rabatel *et al.* (2009) presented a strategy to detect anomalies from sensor data to improve the railway maintenance. It extract sequence pattern from real railway data and identify the abnormal behavior or patterns. PSP algorithm has been used to identify the sequential patterns. To tackle the environments conditions a contextual knowledge based method is proposed. If new data has any anomalies this method provide the reason and seriousness of each anomaly.

Guralnik and Haigh (2002) used sequential pattern mining to learn typical behaviors of humans in their homes. Human behavior is inferred by using motion sensors, pressure pads, door latch sensors and toilet flush sensors. By analyzing from room sequence in which the person was acting complex behavioral models are identified.

Clustering: Aggarwal *et al.* (2003) proposed a framework for clustering data streams, called the CluStream algorithm. The proposed technique divides the clustering process into two components. The online component stores summary statistics about the data streams and the offline one performs clustering on the summarized data. A number of experiments on real data sets have been conducted to prove the accuracy and efficiency of the proposed algorithm.

Ordonez (2003) proposed the improvements to the *k*-means algorithm to cluster binary data streams. Three variant of *k*-means algorithms includes: On-line *k*-means, Scalable *k*-means and Incremental *k*-means are proposed. The experiments were conducted on real data sets as well as synthetic data sets. They demonstrated that the proposed algorithm outperforms the scalable *k*-means in most of the cases.

O'Callaghan *et al.* (2002) proposed an algorithm called Stream and LocalSearch algorithms. The Stream algorithm starts by determining the size of the sample and then applies the LocalSearch algorithm if the sample size is larger than a pre-specified equation result. This process is repeated for each data chunk. Finally, the LocalSearch algorithm is applied to the cluster centers generated in the previous iterations.

Papadimitriou *et al.* (2003) proposed an algorithm named Arbitrary Window Stream Modeling Method (AWSOM) to discover interesting patterns and trends from sensors. AWSOM is a one-pass algorithm to incrementally update the patterns. It requires only O (log N) memory where N is the length of the sequence. Wavelet coefficients are used to compress information representation and correlation structure detection and linear regression model is applied in the wavelet domain. Experiments are conducted on real and synthetic data sets.

Chen and Tu (2007) proposed a framework D-Stream to efficiently identify outliers of clusters. In this framework each input data is mapped into a grid, density of each grid is computed and grids are clustered using a density-based algorithm. This method can discover clusters with arbitrary shapes and detect many outliers around original clusters.

Classification: Aggarwal *et al.* (2004) have applied the idea of micro-clusters introduced in CluStream (Aggarwal *et al*., 2003) for On-Demand classification. The technique uses clustering results to classify data using statistics of class distribution in each cluster. A classification system is created in which the training model is quickly adapted to the changes of the underlying data stream. On-demand classification process dynamically selects the appropriate window of past training data to build the classifier. The results show high classification accuracy in an evolving data stream.

Domingos and Hulten (2000, 2001), Hulten *et al.* (2001) presented a general method called Very Fast Decision Tree (VFDT) to mine high speed stream data. It is using a small sample of available examples when choosing the spilt attribute at any given node. Only the first example to arrive at data stream need to be used to choose split attribute at the root, subsequent ones are passed through induced portion of the tree until they reach a leaf.

Chikhaoui *et al.* (2010) proposed Decision Tree (DT) based classification technique for person identification in ubiquitous environment. In order to identify persons, the proposed approach first extracts frequent episodes from the datasets. Next, it assigns weights to these episodes. Finally, it applies Decision Tree (DT) to classify the frequent episodes. Gaber *et al.* (2005) proposed Algorithm Output Granularity (AOG) based on Light Weight Classification (LWClass) for on-board mining of data streams in sensor networks. The algorithm is empirically validated using synthetic streaming data and the resource-constrained environment of a common handheld computer.

Wang *et al.* (2003b) proposed a general framework for mining concept drifting data streams. The proposed framework use weighted classifier ensembles to mine data streams. They use synthetic and real life data streams to test their algorithm and compare between the single classifier and classifier ensembles. The proposed algorithm combines multiple classifiers weighted by their expected prediction accuracy. Also the selection of a number of classifiers instead of using all is an option in the proposed framework without losing accuracy.

ANALYSIS OF DATA MINING TECHNIQUES FOR SENSORS' DATA

In previous section data mining applications used for mining sensor network are critically analyzed. As a result a qualitative analysis is presented shown in Table 1 and Table 2. The analysis shows the efficiency of data mining applications in developing sensor data mining algorithm. Furthermore, it highlights the objectives of data mining function being used, its strength and open issues in glimpse. Additionally, it directs the future research work for interested researcher in this area of research.

It can be observed from analysis that different types of data mining algorithms are developed for different type of sensors' data. Data mining methods for the sake of association rule mining, sequential pattern mining; clustering and classification are widely used for extracting meaningful patterns from sensor data. On the basis of results drawn from analysis the amount of work done in each data mining method is also analyzed. Figure 1 shows

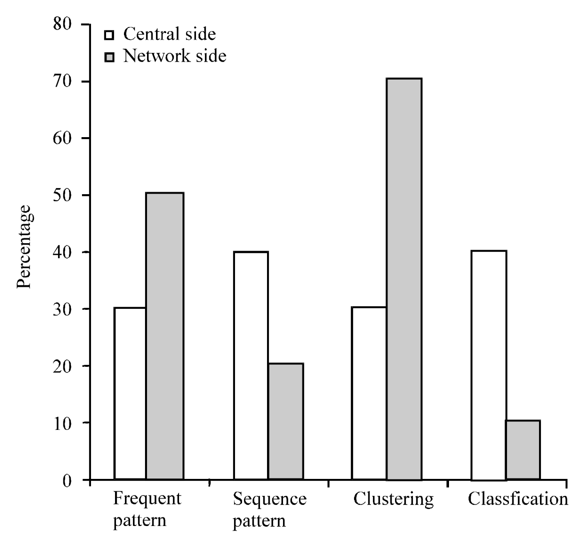

Fig. 1: Mining techniques applied for each side

Table 1: Critical analysis of data mining techniques for network side

Technique	Objectives	Strengths	Data source Real	Data source Synthetic	Limitations
Frequent pattern					
A priori	Extracting rules from sensor values	Memory efficient		x	Time consuming
Frequent patterns	Find spatiotemporal relationships	Data reduction	x		Communication overhead
Frequent patterns	Identify the correlated sensors	Message reduction Fault Prediction	x		Increases buffer cost Messages delayed
FP-growth	Extract behavioral patterns	Predict future events		x	Multi database scan
Frequent patterns	Infer values from previous	Efficient power management data reading		x	Information missing
Sequential pattern					
MLOT	Prediction of moving object	Energy efficient		x	Memory consuming
PTSPs	Prediction of moving object	Missing data recovery	x		Difficult to predict high speed objects
Relational pattern	Mining multi-dimensional patterns	Human readable patterns	x		Memory and time consuming mining
Clustering					
ACE-C	Save energy consumption	Less energy consumption		x	Data redundancy Data delay
LEACH	Save energy consumption	Self organizing clustering		x	Low scalability and stability
HEED	Save energy consumption	Prolong SN life time	x		Energy lost
k-means	Future analysis of sensory data	Reduced power consumption		x	Effects communication when number of sensor s' are high
H-Clusters	Sensory data clustering	Best cluster quality	x	x	High data loss rate
General architecture	Prediction based monitoring	Prolong SN life time	x		Clustering over heads
CAG	Network structure maintenance	Data reduction, Energy efficient		x	Sensor data loss
Classification					
J48 decision tree	Distributed predictive model	Data confidentially, Easy deployment		x	Extra Memory required, Increase power computation
NNTC	Data classification	General purpose		x	Not evaluated on real dataset
kNN, ML	Vehicle classification	High classification accuracy		x	High computation Large memory
SVM training method	Energy-efficient distributed algorithms	Incremental learning		x	Computational complexity

Table 2: Critical analysis of data mining techniques for central side

Technique	Objectives	Strengths	Data source Real	Data source Synthetic	Limitations
Frequent pattern					
DSARM	Estimating sensors' missing	Estimated value accuracy values		x	Ignore the sensor that reports different values
CARM	Missing value estimation	Compute exact sets of FIs		x	Inefficient for handling high speed streams
Moment	Discovering closed frequent itemset over sliding window	Incremental update		x	Memory and time consuming
FP-Stream	Mining time sensitive data streams	Incremental pruning and update of itemsets		x	FP-stream grow larger and larger over time
FP-growth	Mining tree structures	Memory efficient	x	x	Tree construction time and cost is high
Sequential pattern					
MSAP	Mining sequential alarm patterns	Fault prediction	x		Increase cost (multiple DB scan)
GSP	Intelligent home creation	Reduced operation cost	x	x	Fixed time interval
PSP	Extract sequence patterns and detect anomalies	More effective on offline anomaly prediction	x		Real time Anomaly prediction is missing
Tree projection	Infer human behavior human behavior	Automatic prediction of		x	Technique is partially validated
Clustering					
CluStream	Clustering large evolving data streams	Higher quality and scalability	x	x	Off line clustering
Incremental *k*-means	Clustering binary data streams	Fast and efficient	x	x	Only suitable for sparse binary data
Stream local search	Clustering data streams	Incremental learning	x		Low clustering quality in high speed streams
AWSOM	Pattern and trend prediction	Dynamic update single pass	x	x	High complexity
D-Stream	Density-based clustering	Concept drift detection	x		High complexity
Classification					
LWClass	Classifying data streams	High speed less memory		x	Time consuming and costly learning
Ensemble-based	Mining concept drifting Classification	Single pass dynamic update data streams	x	x	Low speed Storage memory problem costly learning
VKFM	Decision tree learning system	High speed less memory		x	Multi-pass
On-demand stream classifier	Dynamic classification model	Dynamic update high speed	x	x	High cost and time need for labeling
DT	Person identification in ubiquitous environment	Improved classification results	x		Increase cost (multiple DB scan)

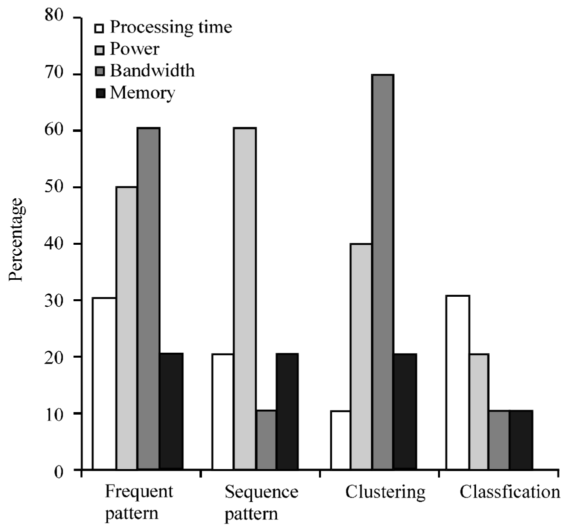

Fig. 2: Optimization at network side

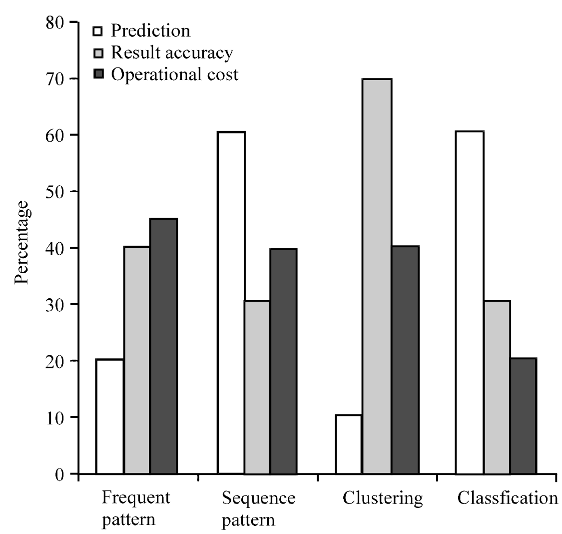

Fig. 3: Optimization at central side

the percentage of work done in literature using these data mining technique on both sides.

It is also observed from analysis that each data mining technique proposed in literature is especially concerned with optimization of certain resource constraint of sensor network such as processing time, memory, bandwidth or power optimization. Each data mining technique is also analyzed for specific optimization problem it addressed. Figure 2 and 3 show the percentage of optimization tasks addressed by each data mining method.

The critical analysis shows that common characteristic of all these approaches is integration of a specific data mining algorithm and a certain type of SNs data for the sake of solving a certain kind of problems. All these approaches shows an efforts to improve the performance of traditional mining algorithm by using additional knowledge and constraint nature of sensors for improving the quality of final results.

CONCLUSION

The emerging need for involvement of data mining algorithms in the field of SNs resulted in the development of numerous algorithms. Each one of these algorithms solves certain particular issues related to the appropriate SNs type. This study presented an overview of selected existing research approaches and classifies the past research either they use centralized or distributed data mining approach. From analysis it is observed that the techniques intended for mining sensor data at network side are helpful for taking real time decision as well as serve as prerequisite for development of effective mechanism for data storage, retrieval, query and transaction processing at central side, on the other hand centralized techniques are helpful in generating off-line predictive insights which in turn can facilitate real-time analysis. This study presented the strengths, limitations and an overall analysis of the past research which can provide insights for end-users in applying or developing an appropriate data mining techniques for sensor networks.

ACKNOWLEDGMENTS

This study was supported in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20110142110062 and International S&T Cooperation Program of Hubei Province under Grant No. 2010BFA008.

REFERENCES

Aggarwal, C.C., J. Han, J. Wang and P.S. Yu, 2003. A framework for clustering evolving data streams. Proceedings of the 29th International Conference on Very large Data Bases, September 12-13, 2003, Berlin, pp: 81-92.

Aggarwal, C.C., J. Han, J. Wang and P.S. Yu, 2004. On demand classification of data streams. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 22-25, 2004, USA., pp: 503-508.

Akyildiz, I.F., W. Su, Y. Sankarasubramamiam and E. Cayirci, 2002. Wireless sensor networks: A survey. Comput. Networks, 38: 393-422.

Arampatzis, T., J. Lygeros and S. Manesis, 2005. A survey of applications of wireless sensors and wireless sensor networks. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, June 27-29, Limassol, Cyprus, pp: 719-724.

Bajcsy, P., J. Han, L. Liu and J. Yang, 2005. Survey of Biodata Analysis from a Data Mining Perspective. Springer-Verlag, London, England, pp: 9-39.

Beyens, P., A. Nowe and K. Steenhaut, 2005. High-density wireless sensor networks: A new clustering approach for prediction-based monitoring. Proceedings of the 2nd European Workshop on Wireless Sensor Networks, February 2, 2005, Istanbul, Turkey, pp: 188-196.

Boukerche, A. and S. Samarah, 2007. An efficient data extraction mechanism for mining association rules from wireless sensor networks. Proceedings of the IEEE International Conference on Communications, June 24-28, 2007, Glasgow, pp: 3936-3941.

Boukerche, A. and S. Samarah, 2008. A novel algorithm for mining association rules in wireless ad hoc sensor networks. IEEE Trans. Parallel Distrib. Syst., 19: 865-877.

Chauhdary, S.H., A.K. Bashir, S.C. Shah and M.S. Park, 2009. EOATR: Energy efficient object tracking by auto adjusting transmission range in wireless sensor network. J. Applied Sci., 9: 4247-4252.

Chen, Y. and L. Tu, 2007. Density-based clustering for real-time stream data. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 12-15, 2007, San Jose, USA., pp: 133-142.

Chi, Y., H. Wang, P.S. Yu and R. Muntz, 2004. Moment: Maintaining closed frequent itemsets over a stream sliding window. Proceedings of the 4th IEEE International Conference on Data Mining, Nov. 1-4, Brighton, UK., pp: 59-66.

Chikhaoui, B., S. Wang and H. Pigot, 2010. A new algorithm based on sequential pattern mining for person identification in ubiquitous environments. Proceedings of the 4th International Workshop on Knowledge Discovery form Sensor Data (ACM SensorKDD'10), December, 2010, Washington, USA., pp: 20-28.

Chong, C.Y. and S.P. Kumar, 2003. Sensor networks: Evolution, opportunities and challenges. Proc. IEEE, 91: 1247-1256.

Chong, S.K., S. Krishnaswamy, S.W. Loke and M.M. Gaben, 2008. Using association rules for energy conservation in wireless sensor networks. Proceedings of the 2008 ACM symposium on Applied computing, March 16-20, 2008, Brazil, pp: 971-975.

Cook, D.J., M. Youngblood, E.O. Heierman, K. Gopalratnam, S. Rao, A. Litvin and F. Khawaja, 2003. Mavhome: An agent-based smart home. Proceedings of the IEEE International Conference on Pervasive Computing and Communications, March 23-26, 2003, USA., pp: 521-524.

Dai, S., P. Wang, L. Gao and S. Zheng, 2008. Mining clustering algorithm in wireless sensor networks. Proceedings of the IEEE International Conference on Granular Computing (GrC), August 26-28, 2008, Hangzhou, pp: 178-182.

Domingos, P. and G. Hulten, 2000. Mining high-speed data streams. Proceedings of the 6th ACM SIGKDD International Conference on KNOWLEDGE DISCovery and Data Mining, August 20-23, 2000, Boston, Massachusetts, United States, pp: 71-80.

Domingos, P. and G. Hulten, 2001. A general method for scaling up machine learning algorithms and its application to clustering. Proceedings of the 18th International Conference on Machine Learning (ICML), Jun. 28- Jul. 1, San Francisco, CA., USA., pp: 106-113.

Esposito, F., T. Basile, N. Di-Mauro and S. Ferilli, 2011. A Relational Approach to Sensor Network Data Mining. In: Information Retrieval and Mining in Distributed Environments, Soro, A., E. Vargiu, G. Armano and G. Paddeu, (Eds.). Springer, USA., pp: 163.

Estrin, D., R. Govindan, J. Heidemann and S. Kumar, 1999. Next century challenges: Scalable coordination in sensor networks. Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 15-19, 1999, Seattle, Washington, USA., pp: 263-270.

Fayyad, U., G. Piatetsky-Shapiro, P. Smyth and R. Uthursamy, 1996. Knowledge Discovery in Databases. AAAI/MIT Press, USA..

Flouri, K., B. Beferull-Lozano and P. Tsakalides, 2006a. Poster abstract: Energy-efficient distributed support vector machines for wireless sensor networks. Proceedings of the European Workshop on Wireless Sensor Networks, February 13-15, 2006, Switzerland.

Flouri, K., B. Beferull-Lozano and P. Tsakalides, 2006b. Training a SVM-based classifier in distributed sensor networks. Proceedings of the 14th European Signal Processing Conference, September 4-8, 2006, Florence, Italy, pp: 1-5.

Gaber, M., S. Krishnaswamy and A. Zaslavsky, 2005. On-board Mining of Data Streams in Sensor Networks. In: Advanced Methods for Knowledge Discovery from Complex Data, Maulik, U., L.B. Holder and D.J. Cook (Eds.). Springer, USA., pp: 307.

Gao, A., W. Wei and X. Xiao, 2010. Multiple hash Sub-chains: Authentication for the hierarchical sensor networks. Inform. Technol. J., 9: 740-748.

Giannella, C., J. Han, J. Pei, X. Yan and P.S. Yu, 2003. Mining Frequent Patterns in Data Streams at Multiple Time Granularities. In: Next Generation Data Mining, Kargupta, H., A. Joshi, K. Sivakumar and Y. Yesha (Eds.). CRC Press, USA., pp: 191-212.

Guo, L., C. Ai, X. Wang, Z. Cai and Y. Li, 2009. Real time clustering of sensory data in wireless sensor networks. Proceedings of the IEEE 28th International Performance Computing and Communications Conference (IPCCC), December 14-16, 2009, Scottsdale, pp: 33-40.

Guralnik, V. and K.Z. Haigh, 2002. Learning models of human behaviour with sequential patterns. Proceedings of the AAAI-02 Workshop Automation as Caregiver, July 29, 2002, Canada, pp: 24-30.

Halkidi, M., 2000. Quality assessment and uncertainty handling in data mining process. Proceedings of the EDBT Ph.D. Workshop, March 31-April 1, 2000, Germany..

Han, J. and J. Gao, 2008. Research Challenges for Data Mining in Science and Engineering. In: Next Generation of Data Mining, Kargupta, H., J. Han, P.S. Yu, R. Motwani and V. Kumar (Eds.). Chapman and Hall/CRC, USA..

Han, J. and M. Kamber, 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann, USA.

Hand, D., H. Mannila and P. Smyth, 2001. Principles of Data Mining. MIT Press, Cambridge, MA.

Heinzelman, W.B., A.P. Chandrakasan and H. Balakrishnan, 2002. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wireless Commun., 1: 660-670.

Heinzelman, W.R., A. Chandrakasan and H. Balakrishnan, 2000. Energy-efficient communication protocol for wireless microsensor networks. Proc. 33rd Int. Conf. Syst. Sci., 2: 1-10.

Hui, Y., 2011. Ad hoc networks based on rough set distance learning method. Inform. Technol. J., 10: 1038-1043.

Hulten, G., L. Spencer and P. Domingos, 2001. Mining time-changing data streams. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26-29, 2001, ACM Press, New York, pp: 97-106.

Jiang, N. and L. Gruenwald, 2006. Cfi-stream: Mining closed frequent itemsets in data streams. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, August 20-23, 2006, USA, pp: 592-597.

Jiang, N. and L. Gruenwald, 2007. Estimating Missing Data in Data Streams. In: Advances in Databases: Concepts, Systems and Applications, Kotagiri, R., P.R. Krishna, M. Mohania and E. Nantajeewarawat (Eds.). Springer, USA., pp: 981-987.

Jiang, N., 2007. Discovering association rules in data streams based on closed pattern mining. Proceedings of the SIGMOD Ph.D. Workshop on Innovative Database Research, June, 2007, China..

Khatoon, S., G. Li and R. M. Ashfaq 2011a. Framework for Automatically Mining Source Code J. Software Engin., 5: 64-77.

Khatoon, S., A. Mahmood and G. Li, 2011b. An evaluation of source code mining techniques. Proceedings of the 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), July 26-28, 2011, China, pp: 1929-1933.

Kumar, D., 2011. Monitoring forest cover changes using remote sensing and GIS: A global prospective. Res. J. Environ. Sci., 5: 105-123.

Le-Gruenwald, M.H., 2005. Estimating missing values in related sensor data streams. Proceedings of the 11th International Conference Management of Data (COMAD'05). January 6-8, 2005, Goa, India..

Lee, L.T. and C.W. Chen, 2008. Synchronizing sensor networks with pulse coupled and cluster based approaches. Inform. Technol. J., 7: 737-745.

Ling, H., 2008. Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana*. Pak. J. Biol. Sci., 11: 763-767.

Loo, K., I. Tong and B. Kao, 2005. Online algorithms for mining inter-stream associations from large sensor networks. Adv. Knowledge Discovery Data Min., 3518: 291-302.

Malhotra, B., I. Nikolaidis and M.A. Nascimento, 2008. Distributed and efficient classifiers for wireless audio-sensor networks. Proceedings of the 5th International Conference on Networked Sensing Systems, June 17-19, 2008, Kanazawa, pp: 203-206.

McConnell, S.M. and D.B. Skillicorn, 2005. A distributed approach for prediction in sensor networks. Proceedings of the Workshop on Data Mining in Sensor Networks, April 21-23, 2005, Newport Beach, CA., USA., pp: 1-10.

O'Callaghan, L., N. Mishra, A. Meyerson, S. Guha and R. Motwani, 2002. Streaming-data algorithms for high-quality clustering. Proceedings of the 18th International Conference on Data Engineering, February 26-March 1, 2002, San Jose, CA., USA., pp: 685-694.

Ordonez, C., 2003. Clustering binary data streams with *k*-means. Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, June 13, 2003, San Diego, CA., USA., pp: 12-19.

Papadimitriou, S., A. Brockwell and C. Faloutsos, 2003. Adaptive, hands-off stream mining. Proceedings of the 29th International Conference on Very Large Databases, September 9-12, 2003, Berlin, Germany, pp: 560-571.

Pottie, G.J. and W.J. Kaiser, 2000. Wireless integrated network sensors. Commun. ACM, 43: 51-58.

Rabatel, J., S. Bringay and P. Poncelet, 2009. SO_MAD: Sensor mining for anomaly detection in railway data. Proceedings of the 9th Industrial Conference Advances in Data Mining: Applications and Theoretical Aspects, July 20-22, 2009, Leipzig, Germany, pp: 191-205.

Ranjan, J. and S. Khalil, 2008. Conceptual framework of data mining process in management education in India: an institutional perspective. Inform. Technol. J., 7: 16-23.

Romer, K., 2006. Distributed mining of spatio-temporal event patterns in sensor networks. Proceedings of the 1st Euro-American Wkshop on Middleware for Sensor Networks, September 2006, UK., pp: 112-15.

Rozyyev, A., H. Hasbullah and F. Subhan, 2011. Indoor child tracking in wireless sensor network using fuzzy logic technique. Res. J. Inform. Technol., 3: 81-92.

Sabri, N., S.A. Aljunid, R.B. Ahmad, A. Yahya, R. Kamaruddin and M.S. Salim, 2011. Wireless sensor actor network based on fuzzy inference system for greenhouse climate control. J. Applied Sci., 110: 3104-3116.

Samarah, S., M. Al-Hajri and A. Boukerche, 2011. A predictive energy-efficient technique to support object-tracking sensor networks. Vehicular Technol. IEEE Trans., 60: 656-663.

Sharma, K., M. Rajpoot and L.K. Sharma, 2011. Nearest neighbour classification for wireless sensor network data. Int. J. Comput. Trends Technol., 2: 41-43.

Taherkordi, A., R. Mohammadi and F. Eliassen, 2008. A communication-efficient distributed clustering algorithm for sensor networks. Proc. 22nd Int. Conf. Adv. Infor. Networking Applic. Workshops, 3: 634-638.

Tanbeer, S.K., C.F. Ahmed, B.S. Jeong and Y.K. Lee, 2009. Efficient mining of association rules from wireless sensor networks. Proc. 11th Int. Conf. Adv. Communi. Technol., 3: 719-724.

Thiruvadi, S. and S.C. Patel, 2011. Survey of data-mining techniques used in fraud detection and prevention. Inform. Technol. J., 10: 710-716.

Tseng, V.S. and E.H.C. Lu, 2009. Energy-efficient real-time object tracking in multi-level sensor networks by mining and predicting movement patterns J. Systems Software., 82: 697-706.

Tseng, V.S. and K.W. Lin, 2007. Energy efficient strategies for object tracking in sensor networks: A data mining approach J. Systems Software 80: 1678-1698.

Wang, H., J. Elson, L. Girod, D. Estrin and K. Yao, 2003a. Target classification and localization in habitat monitoring. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 6-10, Hong Kong, China, pp: 1-4.

Wang, H., W. Fan, P. Yu and J. Han, 2003b. Mining concept-drifting data streams using ensemble classifiers. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24-27, 2003, Shanghai, China, pp: 226-235.

Wu, P.H., W.C. Peng and M.S. Chen, 2001. Mining sequential alarm patterns in a telecommunication database. Proceedings of the 1st Workshop on Databases in Telecommunications II, September 7-10, 2001, London, UK., pp: 37-51.

Yao, Y., L. Feng, B. Jin and F. Chen, 2012. An incremental learning approach with support vector machine for network data stream classification problem. Inform. Technol. J., 11: 200-208.

Yicka, J., B. Mukherjeea and D. Ghosal, 2008. Wireless sensor network survey. Comput. Networks, 52: 2292-2330.

Yoon, S. and C. Shahabi, 2007. The clustered aggregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks. ACM Trans. Sensor Networks, 3: 3307-3313.

Younis, O. and S. Fahmy, 2004. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mobile Comput., 3: 366-379.