DATA STREAM MINING IN WIRELESS DEVICES

SONTAKKE S.*, SHELKE P. AND GAWANDE A.D.

Computer Science and Engineering Department, Sipna's C. O. E. T, Amravati, MS, India.
*Corresponding Author: Email- srsontakke21@gmail. com

Abstract- Mining data streams is concerned with extracting knowledge structures represented in models and patterns in non-stopping stream of information. The data stream in a wireless environment is a time consuming process in pattern discovery which concerned with Wireless devices Data Mining (UDM). The motivation for an efficient data analysis tool capable of gaining these data stream continuously is the dissemination of data stream system, mobile devices and wireless network. In this paper, a data stream mining service is proposed for knowledge discovery in daily life that is wireless devices application, such as the mobile devices. The resource consumption increases the anytime data mining model quality, for adapting the data stream mining services. For achieving adaptation and autonomy of the data mining services, the main component for the adaptability is a decision mechanism on the machine learning. This general mechanism, which autonomously adapts the execution of the data stream mining process to each situation, using context and resource awareness.
Keywords- stream mining, data mining, data stream mining and services, UDM.

Introduction

Data stream have become wireless devices in recent years and handled on several of platforms. In the widespread, wireless devices computing environment are used, because of the small devices which helps in increasing the wireless networks and power capabilities [5, 7, 3]. The characteristics of these computing environments are needed to handle stream of data continuously and to use in different contexts in the resource constraints. In [2], the authors propose a resource-awarestream mining algorithm which adapts the data stream rate with respect to available resources. In this paper, a concrete scenario and situation-aware data stream mining service are presented. Also a decision mechanism for the configuration selection task to achieve autonomous adaptation of the data mining services. Data stream mining has been included as a service on an wireless devices device, providing anytime [1] anywhere [6, 8] intelligent data analysis to wireless devices applications. We formalize the problem and introduce the service components description based on the architecture in [4]. The main challenge is to allow the execution of the embedded process to be situation-aware.

This paper is organized as follows. The next section describes these work preliminaries. In section III we detail ourservice, its components and the adaptation algorithm. A possibleadaptability mechanism is explained in section IV. The experimental setting and results is described in section V. Finally we provide the conclusions and outline future working Section VI.

Scenario, Challenges and Requirements

A. Scenario

Wireless devices applications require intelligent data analysis; if they are running in pervasion mobile devices for their functionalities. The obtained knowledge is used by these applications:
i. Decision tasks internally;
ii. Delivered to its users; and
iii. Sent to other applications.
Examples which can be considered for our scenario of wireless devices applications are wireless sensor networks (WSN) [3] or

ambient intelligence applications [7], personal digital assistants (PDA) [10, 6], intelligent vehicles [5]. As the application domain is broad, it manages the data stream mining algorithm without one-pass and analyzing it continuously being generated at high speeds. The unreliability and costs in communication imply that the data mining process must run locally where the data is generated. To cope with application requirements and device constraints, this on-board data mining process needs to be aware of environmental dynamics. It is important to develop methods to analyze and process data stream for extracting knowledge or pattern from data stream. To fulfill the application data mining process, DSMP must adapt the changes in context and resource availability constantly, anytime anywhere which supports wireless devices applications' data mining requirements. These requests are defined as the data mining process high-level goals, which deals with what the process must do, but not how the system adapts the process to meet those high-level goals.

- A record is bounded for time to train or classify. Therefore the process will scale linearly with the number of records.
- The context and resource variables can be accessed using a known interface depending on the application domain and device used.
- The data to analyze is a data stream. There can be infinite records to analyze. This stream is preprocessed, the mining schema (attributes) is known and the number of classlabels is limited and also known.
- Depends on the device and application, memory used is bounded. There can be large numbers of records to analyze which could fit in the available memory.

B. Challenges

Our research is focused on the challenges which are not solved by data stream mining algorithms. It is designedto manage with the challenges in our scenario.

- Wireless devices application must give good quality of services.
- Provide an interactive DSMP environment to meet user requirements anytime.
- The process history can result in a non-optimal decision that leads to errors and unwanted behavior. So, to avoid unwanted behavior and errors, monitor the historical changes in the DSMP.
- Awareness of contextual information and resource constraints is needed. As the data stream mining algorithms compared with classic algorithms are light weighted, therefore the algorithms is not resource aware.
- Handling possible changes in strategies over time.
- Understanding how the sequence of configuration over time influences the process and how to minimize reliance on expert knowledge.

C. Requirements

Depending on the scenario data stream mining service requirements are:

i. Memory available should be maximum; Accuracy should minimum; Maximum admissible error; Upper bound on the processing time of each record(train and classify); Context varia-

bles. (Data rate. . .); Data schema of the stream is assumed to be fixed; Computational resources should be available. These are the input requirements.

ii. Using a one-pass mining algorithm; A record is processed only once; Adapt the data mining process to changes in service inputs; prediction should done anytime. These are the functional requirements.

Data Stream Mining Service Adaptation

The features of the Data stream mining service are:
Interactive adaptation of the DSMP to changes; this can be done by considering current context, resource availability and data mining requirements. Resource-awareness; Context-awareness; The process history as mean of adaptation; Adaptation mechanisms; Autonomous execution; Simplifying the wireless devices applications for data analysis.

Data stream mining service in wireless devices application adapts the execution process as described:

To look after the information that influences the process execution. According to current situation in the wireless devices application we can analyze and compose this information into a description.

Then based on this situation the process of data mining plan's the best result for DSMP.

The general architecture for context-aware adaptive data stream mining proposed in [4].

Figure 1 shows the service is divided into the several components:

i. We can access a known interface to tune parameters using the DSMP;

ii. The context-awareness, resource-awareness and functionality for inference and modeling of situations these are provided by the Strategy Manager; and

iii. Adjustment of DSMP parameters according to the situation and also includes a decision mechanism that defines how the parameters are tuned is managed by the Situation Manager.

Fig. 1- Service for adaptable UDM - Ubiquitous Data stream Mining

The service for adaptable UDM runs continuously and is able to classify or train records at anytime. When an event is triggered, the services are:

i. UDM evaluates the data mining process;

ii. Using context, resources, requirements and process evaluation information, the current situation is decided;

iii. According to the situation the data mining process parameters in order for execution of the process goals/requirements.

Wireless devices Data Mining (UDM) is the process of performing analysis of data on mobile, embedded and wireless devices [10]. It represents the next generation of data mining system that will

support the intelligent and time-critical information needs of mobile uses and facilitate "anytime anywhere" data mining [6, 10].

A. Data Stream Mining Process (DSMP)

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining).
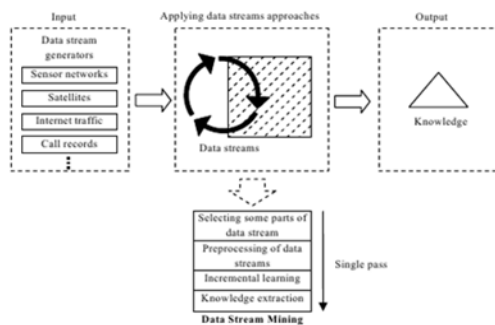

Fig. 2- General process of data stream mining

The fig2, shows the general process of data stream mining for extracting knowledge or pattern from data stream, it is important to develop methods that analyze and process streams of data in single pass, multi-dimensional, online manner and multilevel. This method should not be limited to data stream only, because they are also needed when we have large volume of data. Also some of the techniques in the computation theory can be used for implementing time space efficient algorithms. We can also use common data mining approaches by enforcing some changes in data stream [9].

The component is a concrete data stream mining classification algorithm that keeps an anytime model. The algorithm execution is defined by the specification of its parameters, generated model format, metrics of the process and preprocessing requirements of the data. Information regarding constraints of the input, such as data rates and data types and semantic definition is required.

B. Data Stream Mining Process of Adaptation.

The adaptation problem can be put formally as:

1. The set which defines the situation be S = { S1,. . . ,Sn }, where this set influence the execution of data mining process.
2. Considering, Ci = { P1=vj,. . . ,Pn=vk } as a data stream mining process configuration, where Pn=vk represents that the value vk is set to the parameter Pn of the DSMP.
3. Let St(S) = Ci be the function which maps the current situation to the best DSMP configuration.

C. Strategy Manager

Strategies define the mapping between the current situation and configuration of the process. This is shown in our formal definition where the strategy contains the semantics used to achieve the adaptive behavior. In a classic data mining process this is done by the expert data miner. However in wireless devices environments, like the ones described in our scenario, the process must execute autonomously and its adaptation doesn't require human intervention. Many solutions can be used to implement the strategy man-

ager component.

D. Situation Manager

The situation manager generates the situation features that influence the adaptation of the DSMP. The evolution of context, resource and process evaluation metrics over time. For example, monitoring the accuracy of the classifier and compute a situation feature that reflects its evolution over time. The requirement of adaptation to have persistence of action, the historical information is an important for adaptation. The approach over purely reactive situation features that won't represent the persistent behavior of the stream process for a given period, leading to errors and unstable adaptation.

The contextual information using the multidimensional and each dimension represents a context feature. This situation modeling is based on the Context Space model [8]. The combination of context, resources, process requirements and process evaluation results into a multidimensional space. Collecting useful information from the other subspaces generates into high-level semantic descriptors of the multidimensional features.

E. Adaptation Algorithm

Require: Data stream DS,

1. repeat
2. repeat
3. Get next record DSi from DS;
4. if DSi has classLabel then
5. DSMP. train(DSi);
6. else
7. DSiClass = DSMP. classify(DSi);
8. return DSiClass;
9. end if
10. updateProcessStats();
11. until ADAPT
12. evaluateProcess();
13. Si=SituationManager. GenerateCurrentSituation();
14. Ci= StrategyMananger. GetConfiguration(Si);
15. if Ci != NoChange then
16. DSMP. InvokeConfiguration(Ci);
17. end if
18. until END OF STREAM

The above algorithm is Service Adaptation Algorithm. In this algorithm, adaptation algorithm handles new data records from the stream in the first loop, If they have a class label, then records are used to train the classifier; otherwise, they are classified according to current model. In the second loop one flag is used named as ADAPT flag, which is used to decide when to run adaptation process, and this process can check for every second whenever an event is triggered. Depending on the application and its requirements the updateProcessStats() function updates the statistics needed to store about the algorithm. It can include information-such as data rate or time to train/classify a record. The evaluate-Process() function assesses the process evaluation metrics, such as accuracy or memory used, and updates these values. These values are subsequently used by Situation Manager. The Situation Manager uses the available information from context, resources, requirements, statistics and evaluation of the process to generate the current situation. Using the current situation the Strategy Man-

ager finds the best configuration. If the configuration is different from the default configuration (NoChange), it is then invoked in the DSMP, resulting in the adaptation of the process behavior.

Adaptability Mechanism
In [4], the DSMP make use of pre-defined strategies based on if-then adaptation rules. That is the strategy manager component use a table based on IF Situation THEN Configuration. In this if-then process, the current situation is matched with the available strategies, and the obtained configuration is then invoked in the DSMP. Some problems in this process can be as:

i.   It is a static solution because strategies do not change over time. To deal with this problem, we use methods of learned strategies.
ii. The storage size required grows linearly with the number of strategies; now, to deal with such problems a supervised learning technique is used for decision tree.
iii. Conflicts between strategies; This conflicting strategies is solved with the situation space is fully covered by tree without intersection, and this will avoids the need to deal with conflicts.
iv. The strategies are based on pre-defined heuristics;

A. Learning process
A possible approach to address the problem is to apply supervised learning to data collected from previous process execution. These execution records the configuration selected by experts for given situation. Appling supervised learning, particularly decision trees, to compute models of strategies for configuration selection which was previously logged. A situation space high-level feature, a configuration was then assigned to each record by a human expert. The obtained data set was used to learn strategies.
The data schema used in the learning process is in the format: {Situation set, Configuration (class)} = {S1,. . .,Sn, Conf}
This approach is not free from drawbacks. Time must be taken to create and train strategies before deployment. The advantage remains that once they are created they can be exploited indefinitely. This process must be repeated and this can be a problem if it is costly to send data to the wireless devices device in order to update the strategies used.

B. Experiment design
For simplicity while illustrating the instantiation of the service, the inputs are kept minimum.
Situation Definition: The input variables are enumerated the service requirements and constrains used in wireless devices application.
- Min Accuracy Required - Lower bound on the required accuracy;
- Max Time to Classify - Upper bound on the required timeto classify a new record;
- Max Time to Train - Upper bound on the required timeto train a new record;
- Memory available in the device - Upper bound on thememory available for the service;
- DataStream - The stream that is used as data source.
The resource state and the internal context of the processare

monitored using the variables: Memory consumed, The memory currently used by the service; Accuracy (2 variables), The current value and the value from the past evaluation.

Application of Data Mining
Spatial data mining is the application of data mining methods to spatial data. Spatial data mining follows along the same functions in data mining, with the end objective to find patterns in geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions and approaches. GIS have only very basic spatial analysis functionality. Recently, the task of integrating these two technologies has become critical, especially as various public and private sector organizations. Among those organizations are:
- offices requiring analysis or dissemination of geo-referenced statistical data
- public health services searching for explanations of disease clusters
- environmental agencies assessing the impact of changing land-use patterns on climate change
- geo-marketing companies doing customer segmentation based on spatial location.

Conclusion and Future Work
In this work, the growth of data stream phenomenon and the dissemination of wireless devices motivate the need for wireless devices data stream mining. The aim of knowledge discovery requirement of wireless devices application is provided by a data stream mining services. Also an adaptability mechanism has been presented. We will try to complete the open challenges: minimization of reliance on expert knowledge, the sequence of configuration over time, changes handled in strategies over time and also to explore additional adaptability mechanisms.

References
[1] Domingos P. and Hulten G. (2000) *Sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 71-80.
[2] Gaber M. M. , Krishnaswamy S. and Zaslavsky A. (2005) *Journal of Universal Computer Science*, 11(8), 1440-1453.
[3] Gama J. and Gaber M. M. (2007) *Learning from data streams: processing techniques in sensor networks*.
[4] Haghighi P. D. , Gaber M. M. , Krishnaswamy S. Zaslavsky A. and Loke S. (2007) *An Architecture for Context-Aware Adaptive Data Stream Mining*. In ecmlpkdd.
[5] Kargupta H. , Bhargava R. , Liu K. , Powers M. , Blair P. , Bushra S. , Dull J. , Sarkar K. , Klein M. and Vasa M. (2004)

SIAM International Conference on Data Mining, 334.

[6] Krishnaswamy S. and Loke S. , Zaslavsky W. and Towards A. *15th Australian Joint Conference on Artificial Intelligence (AI02).*

[7] Natalia Marmasse and Chris Schmandt (2000) *The 2nd international symposium on Handheld and Wireless devices Computing*, 157-171.

[8] Padovitz A. , Loke S. W. and Zaslavsky A. (2004) *The Second IEEE Annual Conference*, 38-42.

[9] Muthukrishnan S. (2003) *The fourteenth annual ACM-SIAM symposium on discrete algorithms*.

[10] Grossman R. (1998) *Supporting the Data Mining Process with Next Generation Data Mining System*.