Design and Implementation of a USN Middleware for Context-Aware and Sensor Stream Mining

Cheng Hao Jin* Yang Koo Lee**
Seong-Ho Lee*** Unil Yun****
Keun Ho Ryu*****

Abstract Recently, with the advances in sensor techniques and network computing, Ubiquitous Sensor Network (USN) has been received a lot of attentions from various communities. The sensor nodes distributed in the sensor network tend to continuously generate a large amount of data, which is called stream data. Sensor stream data arrives in an online manner so that it is characterized as high- speed, real-time and unbounded and it requires fast data processing to get the up-to-date results. The data stream has many application domains such as traffic analysis, physical distribution, U-healthcare and so on. Therefore, there is an overwhelming need of a USN middleware for processing such online stream data to provide corresponding services to diverse applications. In this paper, we propose a novel USN middleware which can provide users both context-aware service and meaningful sequential patterns. Our proposed USN middleware is mainly focused on location based applications which use stream location data. We also show the implementation of our proposed USN middleware. By using the proposed USN middleware, we can save the developing cost of providing context-aware services and stream sequential patterns mainly in location based applications.

Keywords : Context-aware, Mining, Sequential Pattern, Stream, USN Middleware

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 2010-0001732) and by a grant (#07-KLSG-C05) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Land, Transport and Maritime Affairs of Korean government and by the grant of the Korean Ministry of Education, Science and Technology(The Regional Core Research Program/Chungbuk BIT Research-Oriented University Consortium).

* Ph.D. Candidate, Computer Science, Chungbuk National University, kimsungho@dblab.chungbuk.ac.kr
** Post-doc, Computer Science, Chungbuk National University, leeyangkoo@dblab.chungbuk.ac.kr
*** Senior Member of Engineering Staff, Electronics and Telecommunications Research Institute, sholee@etri.re.kr
**** Professor, Computer Science, Chungbuk National University, yunei@chungbuk.ac.kr
***** Professor, Computer Science and RICIC, Chungbuk National University, khryu@dblab.chungbuk.ac.kr(corresponding author)

1. Introduction

In recent years, with advances in sensor technique, sensor node becomes to be much smaller, lighter, lower cost, longer lifecycle and easily deployable so that it accelerates the USN computing services [11, 12, 13]. A large number of sensor nodes distributed at a wide area form a USN environment which has led a new paradigm for sensing information to monitor various physical phenomena such as temperature, light, humidity, atmospheric pressure, fluid level, location, etc. Examples of USN applications in different areas include logistics, intelligent transportations, industrial automation, health monitoring, environmental surveillance, retail stores, disaster protection and so on.

One of attempt in USN applications is to provide users a fast and reliable context-aware service and useful information such as sequential patterns. However, a large number of sensor nodes are widely distributed in the sensor network and they tend to continuously generate a huge number of stream data. Stream data arrives at a high speed rate and it is time-varying, unpredictable and unbounded. Therefore, it is impossible to accommodate the whole incoming data. Due to such characteristics of sensor stream data, the traditional data processing methods are not fit for stream data. One of solutions to solve above problem is to use sliding window to process a certain length of stream data at each time and then slide to process the next certain length of stream data. By using sliding window technique can successfully deal with continuous data and most previous researches [1, 2, 3, 6, 7, 8, 9, 10, 14, 16] mainly focus on query processing. However, there still has been a lack of USN middleware for diverse stream-based applications to support context-aware service, prediction and some other useful services.

For example, old people or patients can meet an emergency when they are moving from one place to another place and for another example, workers who work at the dangerous places such as work in coal mine can be suffered from gas poisoning. Therefore, they should be well monitored when they are moving with time advances. By receiving the stream data of their body temperature, heart rate, blood pressure, gas strength, their location, etc., administrator can monitor and predict the occurrence of emergency and also can take corresponding actions to prevent from any loss or damage. However, there is little attention on a USN middleware for both context-aware service and stream pattern mining, specially on the stream-based applications. Therefore, in this paper, we propose a USN middleware which can provide both fast and reliable context-aware service and useful stream sequential patterns using stream location data. To support the functionality of stream data processing, the proposed USN middleware employs sliding window technique.

The rest of the paper is organized as follows. In the following section, we review several Data Stream Management Systems(DSMS). The proposed USN middleware is presented in Section 3 and implementation of proposed USN middleware is shown in Section 4. Conclusion and future work are given in Section 5.

2. Related Work

Recently, there are many data stream management systems have been developed to meet the characteristics of stream data [1, 2, 3, 6, 7, 8, 9, 10, 14]. However, due to the limited space, in this paper we review 3 systems.

STREAM [1] is a general-purpose prototype of data stream management system developed at Standford University. STREAM provides declarative Continuous Query Language (CQL) over continuous streams. It focuses on memory management and approximate answering and targets the following modern applications such as network monitoring, financial analysis, manufacturing and so on.

TelegraphCQ [2] is a sensor based data stream management system developed at Berkeley University. It extends the open source relational DBMS PostgreSQL to support continuous streams.

IBM system S [3] is a stream processing middleware that it can process a large volume of both structured and unstructured continuous stream data. S can process stream data from sensor, cameras, news feeds and many other sources.

All of above systems mainly focus on continuous query processing. Although they can successfully process stream data, none of them can provide both context-aware service and useful patterns hidden in the stream data. Therefore, in this paper, we propose a general USN middleware for both context-aware service and stream sequential patterns mining.

3. USN Middleware Structure

In this section, we introduce a novel USN middleware used to provide the users context-aware service and stream sequential patterns which are extracted from stream location data. Figure 1 shows the structure of proposed USN middleware. As shown in Figure 1, USN middleware consists of Sensor Network Common Interface, Context Analysis Component, Sensor Stream Mining Component, Database and Open API Component.

Sensor Network Common Interface plays an intermediate role between middleware and sensor networks such that defines the common message between many diverse sensor devices or different sensor networks. To provide independence between different sensor networks and different applications, a set of common messages are defined such as command action request/response, connection information request/response and so on. These messages are represented as XML data format because XML has become a common tool used for data exchange.

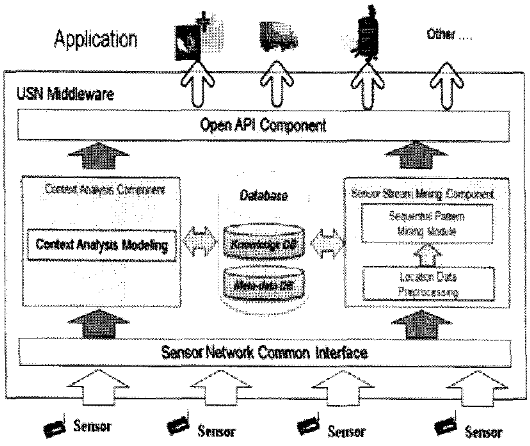


Figure 1. USN middleware structure

Context Analysis Component provides rapid and reliable context analysis service to the users based on sensed surrounding stream data and stream location data. For example, doctors are receiving an old man's heart rate from his heart rate sensors. When the heart rate is under a certain threshold for some time long, he may have a serious heart disease at that time. Therefore, after getting the alerting from he mayituaave a the doctor can take actionse fore,event him from an emergency. In our USN middleware, WHEN-DO model [15] is uherefore, - aftercontext analysis service.ng, he mamodel, 'WHEN'rclause specifies context codition and 'DO' clause specifies action to be performed when sensing data satisfies 'WHEN' condition.

Sensor Stream Mining Component is used to extract frequent sequential patterns from stream location data. Sequential pattern mining is the mining of frequently occurring ordered events or subsequences as patterns [4]. For example, there are many patients in the hospital and they often walks around the garden. By monitoring their stream location data, the doctor may want to know to which place they frequently go and also want to predict where the next place they will go. By doing so, patients can be well monitored and diagnosed at an urgent disease. However, due to the aforementioned characteristics of stream data, we could not directly mine sequential patterns from the entire continuous stream data so that we employ sliding window to mine frequent sequential patterns over stream location data. The stream sequential pattern mining algorithm over stream location data is shown in Section 4.

Database Component stores information about context-awareness and the extracted stream sequential patterns. For context-awareness, we stores the context occurred time, the value of threshold for invoking that context and some other surrounding sensed value related to that event such as temperature, humidity and so on. And also stores the time of mined stream sequential patterns and corresponding stream sequential patterns.

Open Application Programming Interface (API) Component is used to provide a communication interface (parameter, query, threshold and so on) to the different applications.

4. Stream Sequential Pattern Mining

Mining sequential patterns over stream data is much more complex than mining sequential patterns in a static dataset since it suffers from time and space restriction. Considering this point, we modify PrefixSpan [5] algorithm combining with time-based sliding window which keeps the latest information in the stream. Details of stream sequential pattern mining algorithm is shown in Table 1.

First of all, the data type used to mine sequential pattern should be categorical data. However, the location data received from remote sensors is numeric data. Therefore, in order to mine sequential

Table 1. Stream sequential pattern mining

Algorithm(SequentialPatternExtraction)
Input : A sequent dataset S, minimum support threshold min_sup
Output : The complete set of sequential patterns
Method : Call SequentialPatternExtraction(<>, 0, S)
Subroutine SequentialPatternExtraction(□, l, S|□)
Parameters :
□ : a sequential pattern;
l : the length of sliding window
S|□ : the □-projected database, if <>; otherwise the sequence dataset S.
Method :
1. convert numeric location data to categorical data.
2. scan S|□ once in the current sliding window, find the set of frequent items b such that
   (a) b can be assembled to the last element of □ to form a sequential pattern; or
   (b) <b> can be appended to □ form a sequential patterns.
3. for each frequent item b, append it to □' to form a sequential pattern □', and output □'
4. for each □', construct □'-projected dataset S|', and call SequentialPatternExtraction(□', l+1, S|□')

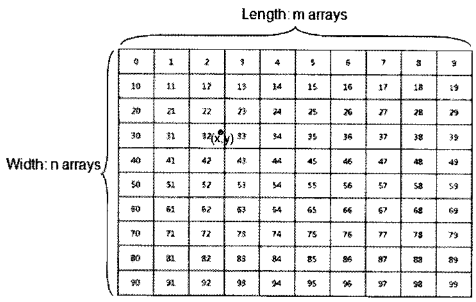


Figure 2. Grid based area

patterns from such stream location data, numeric location data is need to bis nnverted to categorical data. To siumeify this nnversion, we assume that moving objects (such as person, car and so on) are moving within a rectangle area and consists of m × n equal sized arrays as shown in Figure 2. The number in each array means the address label. Hence, when receiving the (x, y) coordinates, according to Eq. (1), we can easily get the address label of the place where the moving object is now in.

Address Label = (n*((int)y/width) + (int)x/length); (1)

5. Implementation

This section describes the implementation of proposed USN middleware. USN middleware is implemented on a Mobile Dual Core Intel Merom 2666MHz with 2G DDR2 SDRM memory, running Microsoft Window XP professional with service pack 3. We use Microsoft Visual Studio 2008 and C# language using Microsoft .NET Framework Version 3.5 SP1 and use ORACLE 10G Expression Edition to store meta data, context information and extracted stream sequential patterns.

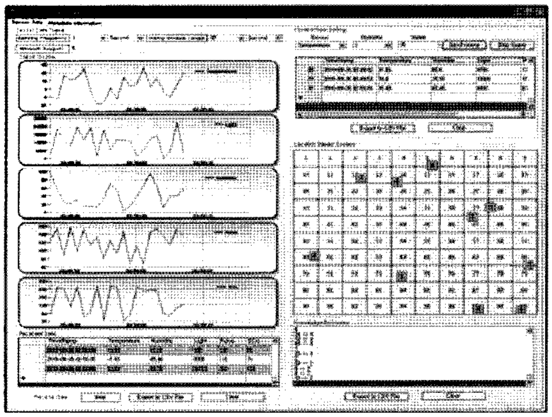


Figure 3. USN middleware implementation

The left top side of the Figure 3, there are 3 textboxes and two are used to set the values of sliding window and another one is used to set the value of minimum support threshold for mining stream sequential patterns.

Assume there are temperature, light and humidity sensors are fixed in the wide area of hospital and each patient wears pulse, ECG and Global Positioning System (GPS) sensors to collect pulse, ECG and location data. Also assume many patients (blue buttons on the big rectangle in Figure 3) are taking a walk within the hospital garden (big rectangle in Figure 3). Temperature, light, humidity and each patient's pulse and ECG data are received through corresponding sensors. These sensing data signals are displayed in 4 charts and the exact sensing values are displayed in a datagrid at the bottom of left side. Components on the right top side are used to set the context condition.

For example, a patient's pulse value may be bellow at a threshold and he could not move any more if he has a serious heart disease. When the sensing values satisfies this conditions, the event occurred time and other surrounding sensing information sensed at that time are displayed at the datagrid on the right top side of Figure 3 and the corresponding button on the big rectangle will stop moving. Then actions predefined in the WHEN-DO context model are performed according to this context so that an ambulance will appear near that patient. Using stream location data, frequent stream sequential patterns can be mined with stream sequential pattern mining algorithm in Section 4 and displayed at the listbox at the bottom of right side in Figure 3. These patterns help doctors to know which place patients frequently go and also to predict which the next place they will go. Therefore, patients can be well monitored.

Figure 4 shows meta data information about sensor nodes such as ID, sensor type, location, remaining battery, last sensing time and so on.

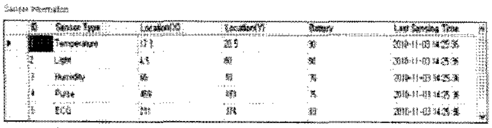


Figure 4. Meta data information

6. Conclusion

In this paper, we proposed a USN middleware. This USN middleware is composed of Sensor Network Common Interface, Context Analysis Component, Sensor Stream Mining Component, Database and Open API Component. The key aspect of this USN middleware is that it can provide context-aware service and frequent stream sequential patterns to diverse location-based applications. The implementation of proposed USN middleware shows that it can be efficiently used in several application domains such as patient monitoring, u-healthcare and so on. Therefore, with the help of our proposed USN middleware, we can save the developing cost of providing context-aware services and stream sequential patterns mainly in location based applications.

Although our proposed USN middleware can provide context-aware service and stream sequential pattern, in this paper, we assume the area is a rectangle when we mine the stream sequential patterns. However, in real-world applications, it may be much more complex than our assumption. Therefore, as a future work, we will mine stream patterns without this assumption, add more functions to support complex context-aware services and add performance evaluation comparing with other related systems.

References

[1] A. Arasu, B. Babcock, S. Babu, M. Datar, K. Ito, R. Motwani, I. Nishizawa, U. Srivastava, D. Thomas, R. Varma, and J. Widom, 2003, "STREAM: The Stanford .Stream Data Manager," IEEE Data Engineering Bulletin, vol 26, No. 1, pp. 19-26.

[2] S. Chandrasekaran, O. Cooper, A. Deshpande, M. Franklin, J. Hellerstein, W. Hong, S. Krishnamurthy, S. Madden, V. Raman, F. Reiss, and M. Shah, 2003, "TelegraphCQ: Continuous Dataflow Processing for an Uncertain World," In Proc. of the Innovative Data Systems Research Conference (CIDR).

[3] http://download.boulder.ibm.com/ibmdl/pub/software/data/sw-library/ii/whitepaper/SystemS_2008-1001.pdf

[4] J. W. Han and M. Kamber, 2006, Data Mining: Concepts and Techniques, 2nd ed, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[5] J. Pei, J. W. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, 2001, "PrefixSpan: Mining Sequential Patterns Effi-

ciently by Prefix-Projected Pattern Growth," In Proc. of the International Conference on Data Engineering (ICDE), pp. 215-226.

[6] D. Carney, U.Cetintemel, M.Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and S. Zdonik, 2002, "Monitoring Streams: A New Class of Data Management Applications," In Proc. of International Conference on Very Large Data Bases(VLDB), pp. 215-226.

[7] C. Cranor, T. Johnson, O. Spatscheck, and V. Shkapenyuk, 2003, "Gigascope: A Stream Database for Network Applications," In Proc. of the ACM SIGMOD International Conference on Management of Data, pp.647-651.

[8] Y. K. Lee, K. H. Ryu, 2008, "Historical Sensor Data Management using Temporal Information," Journal of Korea Spatial Information System Society, vol. 10, no. 4, pp. 143-813.

[9] J. Chen, D. J. DeWitt, F. Tian, and Y. Wang, 2000, "NiagraCQ: A Scalable Continuous Query System for Internet Database," In Proc. of the ACM SIGMOD International Conference on Management of Data, pp. 379-390.

[10] M. Sullivan, 1996, "Tribeca: A Stream Database Manager for Network Traffic Analysis," In Proc. of the 22th International Conference on Very Large Data Bases, pp. 594.

[11] M. Weiser, 1991, "The Computer for the Twenty-First Century," Scientific American, vol. 265, no. 3, pp. 94-104.

[12] B. Schilit, N. Adams, R. Want, 1994, "Context-Aware Computing Applications," In Proc. of the 1st International Workshop on Mobile Computing Systems and Applications, December, pp. 85-90.

[13] A. Dey, D. Salber, G. Abowd, and M. Futakawa, 1999, "The Conference Assistant: Combining Context-Aware with Wearable Computing," In Proc. of the 3rd International Symposium on wearable Computers, pp. 21-28.

[14] Y. Yao and J. Gehrke, 2002, "The Cougar Approach to In-Network Query Processing in Sensor Networks," In Proc. of the ACM SIGMOD Record, vol. 31, No. 3, pp. 9-18.

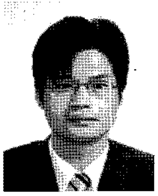
[15] C. H. Jin, Y. Lee, G. Shon, Hi-Seok Kim, K. H. Ryu, 2009, "Design of Context Analysis System on USN Environment," In Proc. the Sixth International Conference on Information Technology: New Generations (ITNG 2009), pp. 1061-1066.

[16] L. Wang, T. H. Zhou, Kwang-Deuk, Y. K. Lee, K. H. Ryu, 2009, "Using Skylines on Wavelet Synopses for CKNN Queries over Distributed Streams Processing," Journal of Korea Spatial Information System Society, Vol. 11, No. 2, pp. 7-11.

Manuscript received : 2010.11.11
Revised : 2010.01.18
Accepted : 2010.01.21

Cheng Hao Jin

2004 Dept. of Computer, Electronic & Commnunication Engineering, Yanbian University of Science and Technology, China (B.S.)
2008 Dept. of Computer Science, Chungbuk National University, Korea(M.S.)
2008~Present. Ph.D Candidate, Dept. of Computer Science, Chungbuk National University, Korea
Research Interest : Stream Mining, Spatial-temporal Database, Discriminative Pattern Mining

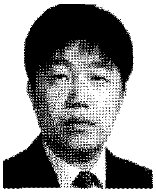


Yang Koo Lee

2002 Dept. of Computer Information Engineering, Chongju University of Korea(B.S.)
2004 Dept. of Computer Science, Chungbuk National University(M.S.)
2010 Dept. of Computer Science, Chungbuk National University(Ph.D.)
Research Interests : Spatial-temporal Database, Sensor Data Stream, u-GIS, and Multimedia Ddatabase

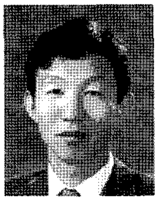
Seong-Ho Lee

1997 Dept. of Computer Science, Chungbuk National University(B.S.)
2000 Dept. of Computer Science, Chungbuk National University(M.S.)
2000~Prresent, Senior Member of Engineering Staff, Electronics and Telecommunications Research Institute
Research Interests : Spatial-temporal Database, GIS, GeoWeb



Unil Yun

1997 Dept. of Computer Science and Engineering, Korea University(M.S.)
2005 Dept. of Computer Science, Texas A&M University, USA(Ph.D)
2005~2006 Post-doctoral associate, Dept. of Computer Science, Texas A&M University, USA
Research Interests : Data Mining, Database Systems, Information Retrieval, Artificial Intelligence, Digital Libraries.



Keun Ho Ryu

1976 Dept. of Computer Science, Soongsil University(B.S.)
1980 Dept. of Computer Science, Yonsei University(M.S.)
1998 Dept. of Computer Science, Yonsei University(Ph.D.)
1976~1986 Post-doc of Arizona University, Research Scientist of Electronics & Telecommunications Research Institute of Korea
1989~1991 Research Staff of Arizona University
1986~Present Professor, Dept. of Computer Science,, Chungbuk National University
Research Interests : Temporal Database, Spatial-temporal Database, Temporal GIS, Ubiquitous Computing and Stream Data Processing, Knowledge-base Information Retrieval, Database Security, Data Mining, and Bioinformatics