

Mining Social Structure on Context Objects Database

Tingting Fu*

School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: The advance of computing, hardware and wireless communication technologies has put the new words—Internet of Things, Social Networks *etc.*, into our daily life. Related areas such as Wireless Sensor Networks and Internet of Things have been changing our daily lives. Context awareness is the key to the success of prospective applications, such as location based services and social information delivery. This paper focuses on the hot topic of how to store and manage this context information efficiently. To support social information routing, this paper brought forward an unbalanced sliding window based social structure mining algorithm. We also developed new concept of Context Objects Database, COD, which stored and managed various context information and objects such as time, location, interaction and so on. The paper mainly discusses the unbalanced sliding window model, fundamental architecture of COD. We also discussed the application of the COD such as mining social structure using sliding window model. The simulation results show that the proposed method has better accuracy of social communities discovery and lower data amount processed.

Keywords: Context Objects Database, XML, Social structure, Sliding window.

1. INTRODUCTION

Database and data mining solutions are keys to the development of modern technology such as Internet of things, social network and big data. They are very useful in data analysis, data extraction, pattern recognition, and so on. Although a lot of works have been done on it, there remain many issues that need further exploration efforts. But related researches are rarely focusing on database technology. In this paper we try to combine the current off-the-shelf database technology and the specific needs of ubiquitous computing, design context object database and its management system which can store and process the context information. We utilize the mining technology and sliding window model to help find social structure so that there will be efficient information dissemination among users. We study the framework model of object-oriented database system, context information representation and storage.

At present, researches on the context object database are still in its infancy. Related research mainly focuses on small embedded mobile database technology such as the embedded database [1, 2], mobile database [3, 4] and moving object database [5]. Industrial companies have off-the-shelf products such as Berkeley DB, SQLite embedded database, and MySQL database management system [6]. Some scholars began researches on mobile database very early [7]. Grid-based database technology is also an important direction for future development [8]. The research of Context Object Database will greatly prompt database technology, the development and applications of ubiquitous computing.

Social networks have recently been a hot research topic and discussed how to use collected data to form social population behavior and evolution principle such as social media development, merchandise recommendation services, social communication applications and public sentiment surveillance. No matter how these social data are utilized, only through social network can we gather, disseminate, and analyze them. With the popularization of handheld smart digital devices, the Mobile Social Networks have been a key part of social networks. The notable attributes of Mobile Social Networks can be summarized as: i) sequent context aware sensing, ii) off-line interactive communications, iii) content sharing and inquiring. Among these three, content sharing and inquiring is often eventually the purpose of users.

From the viewpoint of technique, information must be delivered to the right users at the right time. In many researches [9-11], Mobile Social Network has been regarded as a kind of mobile opportunistic networks which are developed from delay tolerant networks. In such networks, end-to-end connectivity cannot be guaranteed at a single point of time but be intermittent so that the path of links between source and recipient is composed of several separated parts both in terms of time and physical communication channels. The most common way to deliver data in such networks is carry-and-forward technique. However, different from other networks such as Internet and wireless sensor networks, the network topology of Mobile Social Network is highly dynamic. It is hard to decide which node has more opportunities to transfer data to next hop that is closer to the destination node. To address this problem, flooding and broadcast is one of the simplest ways that can work in practice. Even in Wireless Sensor Networks, flooding was an important option before more efficient approaches were brought forward. However, social networks are often built on a scale of thou-

*Address correspondence to this author at the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China; Tel: +86-136-0661-1368; E-mail: ftt@hdu.edu.cn

sands to millions of people while in Wireless Sensor Networks even one thousand nodes can be regarded as a large scale network. The flooding or broadcast based method will consume huge amount of resources that can't be ignored. Furthermore, users can not be always willing to relay messages.

2. SOCIAL STRUCTURE MINING USING SLIDING WINDOW

Social structure is the key to implement efficient content sharing and inquiring because message can be forwarded through more reliable and delay predictable route. Social structure is not similar as network topology but can represent user social behavior and relationships both in spatial and temporal aspects. All these attributes play important roles in data relaying and exchanging. In this paper, we try to address the underlying requirements of social structure discovering, namely how to mine social contacts, friendship, kindred, user profiles and other elements to support decision making.

These social data, according to [12], is generated daily and can be accumulated to PB or EB level. It is really difficult to dig useful information out within such big data. In Mobile Social Networks, there are online environment and off-line environment. Online environment can include Social Network Site, such as Facebook, Twitter, Chatting tools such as MSN, and all other social related online applications. Off-line environment includes meetings, sports activity, traveling, even includes any activities of a person's daily life. Two persons may have intersection only because they meet each other every day on the way to work. Regarding online environment, the relationship between users can be mined through web log, users' friends list and many other ways. Data of Off-line activities can be gathered from wearable equipment, mobile phones and various kinds of sensors. All these data are continually generated and most with spatial and temporal attributes. Hence, we propose to use sliding window to mining incremental data to help forming social structure.

Data stream is kind of continuous data sequence as the time goes on. With continuous, high speed, infinite data flow, it requires data mining algorithm should satisfy several conditions, such as one time scan, online algorithm, and low spatial temporal complexity. Furthermore, it should be able to adapt to dynamic data stream. For example, various sensors are constantly producing data, movement of users is composed of location data. Actually, mining frequent itemsets is one of the most fundamental problems in data stream mining. Mining algorithms have to produce results within limited memory size and processing capability.

Sliding window is one of the most powerful tools to mine data stream. It can limit data to a certain amount. By adjusting the size of the sliding window it could adapt to different applications. In Mobile Social Networks, most recent data will have more attention where Sliding window (SW) exactly fits the requirement. Sliding window is a frame that contains several sub windows. Each Sub window holds certain amount of data. As the time goes on, data shift in sub windows. We propose to explore two aspects of social structures, Frequent Community and Frequent Location.

Frequent Community is to discover users who have frequent interactions. We record user contacts in each social event so that we can measure how many times users are co-located and interact within a given period of time. Here we emphasize on a given period of time and usually its most recent period of time because social structure changes dynamically. If we can find the most frequent interacted users or users above a given threshold it will greatly improve the routing efficiency.

Frequent Location is to discover locations that a specific user or group users are often visiting. Thus we can use this location to build communication between desired users such as using external message delivering devices which makes message delivery more efficiently.

The interaction of people will generate a huge amount of data so that it's difficult and time consuming for a thorough mining and storing. We use sliding window to mining on incremental data which can significantly reduce computing and communication cost. The importance of an interaction depends on location, time and participators. Time-to-time and location-to-location interaction changes dynamically as well. Hence we bring forward unbalanced sliding window model to solve the problem. That is, sliding window is user, time and location sensitive and will be adjusted accordingly.

3. PROBLEM STATEMENT OF MINING SOCIAL STRUCTURE USING SW

Let $I = \{i_1, i_2, \dots, i_n\}$ is a set of items. For Frequent Community, i_x stands for each individual user so that I is a set of all possible users. Within a given period of time, if any of the users have interactions then we can form a transaction $T = (tid, u_1, u_2, \dots, u_m)$, $u_i \in I$, $T \in I$, where m is the number of users in the transaction. With the time passing by, it will turn to be a sequence of transactions data stream $TDS = T_1, T_2, \dots, T_n$ in which T_n is most recent transaction (user interaction). In most occasions, we only need most recent data. We can divide TDS into several basic windows as W . Normally, we set $|W| = 1$. A sliding window SW is composed of a series of basic windows, $SW = [W_{i+1}, W_{i+2}, \dots, W_{i+k}]$, where k reflects the size of the sliding window. SW will move when a new transaction arrives from TDS. In the SW, we can find group of users that satisfy the requirement of least number of contacts. Furthermore, we can find users who have the highest contact history. To implement it, we use support count of sliding window method. The support count of itemset U is the number of transactions containing U in SW. We mark it as $sup(U)$. We call an itemset U is frequent if $sup(U) \geq min_sup$ where min_sup is the minimum support given by the analyzer. Using the support count we can find most frequent itemset and the closed frequent itemset. The prior one indicates the most frequent contacts between some users. The latter one implies that an itemset satisfies the min_sup and does not have any superset. Regarding Frequent Location, items mean locations. Each transaction represents locations of user in interaction. We can also find locations that have most or frequent relationship.

Social interaction data is not like other kind of data. It relies much on temporal and spatial facts. For example, most people act in the daytime while less people go out in the

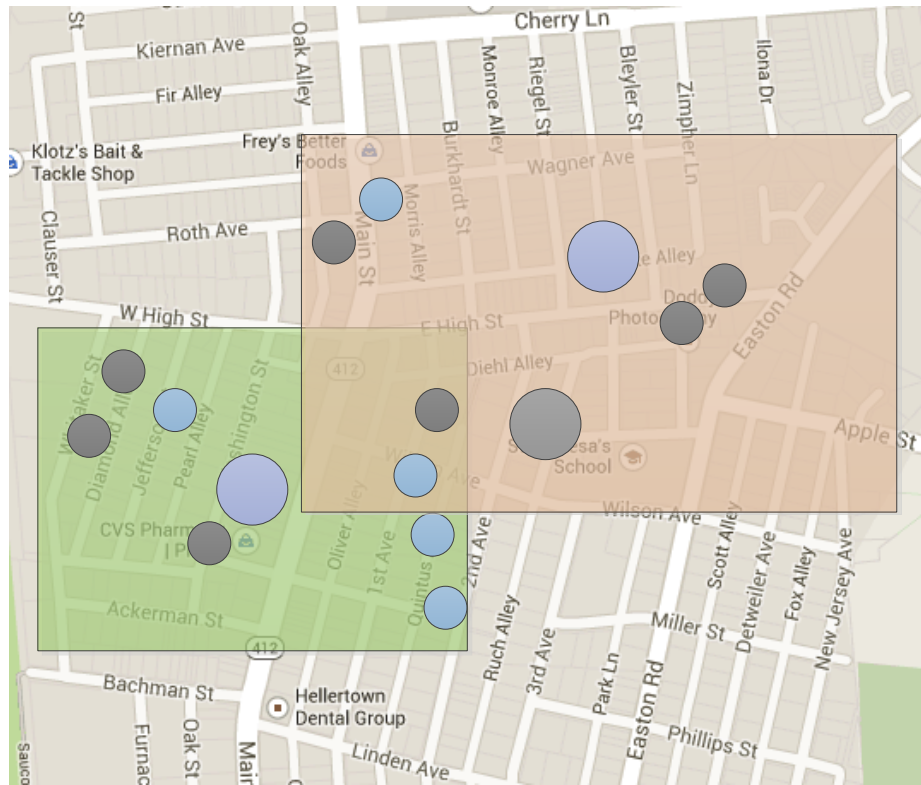


Fig. (1). Unbalanced sliding window at different location.

night. Most people attend social events during the weekend other than weekdays. In our work, we propose unbalanced sliding window based algorithm to mine social structures. The problem is transferred to mine frequent sequential patterns by SW. The main difference between it and the traditional sliding window model is that it dynamically adjusts the window size to adapt to scenarios. Namely, the mining results better meet the users' requirements.

It can be seen in Figs. (1 and 2). We have two kinds of unbalanced sliding windows. In Fig. (1), window size represents the geographic coverage. Green window stands for a hot area that has many interactions. We shrink the window size to focus on more special results. We enlarge the red window size to cover more users in which we could find useful social connections. Fig. (2) shows a time-scale unbalanced window. During night, we expend window to observe more hours while concentrate on two hours in the morning. We do exact the same to week days.

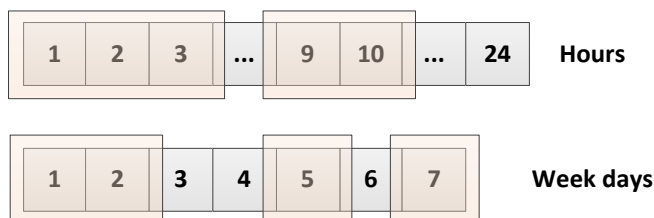


Fig. (2). Unbalanced sliding window at different time.

4. MINING ALGORITHM OF UNBLANCED SLIDING WINDOW

We use an example to explain our algorithm. There are social interactions in three periods of time as shown in Fig. (3). Seven users have different communications within transaction T1, T2 and T3. We record these three transactions in a sliding window. There are also following transactions which are allocated in sliding windows W2 and W3. The server requires to make statistics on every three sliding windows. The requirement is changeable subject to application, location and accuracy. The mapped data stream from interaction map is shown in the Table 1:

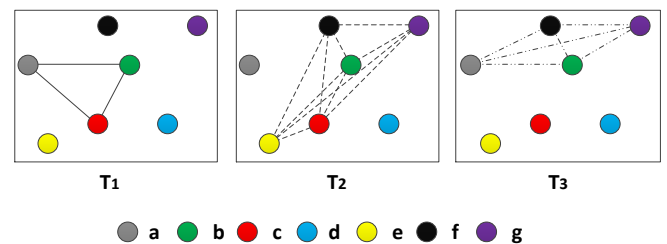


Fig. (3). User interaction map of different transaction.

To make it simple, we use English letter to name users and sort them in lexicographical sequence. Table 1 shows three sliding windows. Each window may have different transactions. There are many sliding windows based mining algorithms. In our paper, we adopted a table based method.

The algorithm forms two tables using above data, namely, head table and item Table 2.

Table 1. Transactions and sliding windows on user interactions in data stream manner.

SW	Transactions	Items
W ₁	T ₁	a b c
	T ₂	b c e f g
	T ₃	a b f g
W ₂	T ₄	b c f
	T ₅	a b c f
	T ₆	a b c d e f
W ₃	T ₇	c e f
	T ₈	d e
	T ₉	a b e
	T ₁₀	b c d f

Table 2. Item table.

Item	Count
a	5
b	8
c	7
d	3
e	5
f	7
g	2

From above Table 3, we can easily get the support count of 2-users itemset, $SUP(2\text{-users})=\{ab:5, ac:3, ad:1, ae:2,$

Table 3. Transaction bit table.

SW	Tran	a	b	c	d	e	f	g
W ₁	T ₁	1	1	1	0	0	0	0
	T ₂	0	1	1	0	1	1	1
	T ₃	1	1	0	0	0	1	1
W ₂	T ₄	0	1	1	0	0	1	0
	T ₅	1	1	1	0	0	1	0
	T ₆	1	1	1	1	1	1	0
W ₃	T ₇	0	0	1	0	1	1	0
	T ₈	0	0	0	1	1	0	0
	T ₉	1	1	0	0	1	0	0
	T ₁₀	0	1	1	1	0	1	0

af:3, ag:1, bc:6, bd:2, be:3, bf:6, bg:2, cd: 2, ce:3, cf:6, cg:1, de:2, df:2, dg:0, ef:3, eg:1, fg:2}. Given minimum support account as 3, the algorithm first deletes all 2-users which support count are less than 3. Then we get $SUP(2\text{-users})>3=\{ab:5, ac:3, af:3, bc:6, be:3, bf:6, ce:3, cf:6, ef:3\}$. From it, we can get closed frequent item set $\{(a, b, c), (a, b, f), (b, e), (b, c, f)\}$ and most frequent item set $\{(b, c), (b, f), (c, f)\}$. When there are more transactions move in, we will delete the oldest sliding window W₁ and add new sliding W₄ to accommodate new transactions, update two tables and compute new frequent item set. Thus, given any social interaction map, we can map it into data stream and use unbalanced sliding widow model to mine it.

Compared with other social structure finding algorithms which mostly are based on graph theory, our algorithm is light weighted. It only concentrates on a certain amount of data and updates the result *via* incremental mining. It fits for social structure discovery to enhance information delivery.

Those data stream and mined results can be stored in a new kind of database, Context Objects Database. The following section will give a brief introduction to the database.

5. CONTEXT OBJECTS DATABASE

With the rapid development of embedded and wireless communications technology, more and more temporal-spatial database are deployed in applications such as traffic management, geographic information systems, locating, urban planning and other fields. Temporal-spatial database as a modern, high-level object-oriented database technology can effectively organize and manage geographic data, attribute, spatial and temporal semantics. Temporal-spatial database, including time and space element is a three-dimensional or four-dimensional database by adding temporal dimensional on traditional spatial database. Compared with traditional database it has dynamic and comprehensive features so as to naturally support analysis and prediction adhere to the prediction systems, decision support systems and other analytical system services.

5.1. Temporal-Spatial Database

Context object database can be classified as a special temporal-spatial database; it stores the context object data. Context object is the context information in ubiquitous computing environment such as weather at a moment, or a place of traffic. Most of the context information includes the properties of space and time, and may be interrelated. The database should be ready for query such as "check my nearest station" which depends on the location of the user, and "check the urban weather at tomorrow noon" which depends on the time.

Context information storage and processing is an emerging technology. Research in this field is still in its infancy, there are many problems to be solved. The context object database proposed in this paper is a higher level abstraction based on an existing database technology. It has three data sources, one is classic relational database, and second, XML documents, both of which are based on storage. There is a wireless sensor network that belongs to a dynamic database. Users can query the data in a real-time manner and in the future.

5.2. Model of Context Object Database

As shown in Fig. (4), when the user needs context information, he does not care about where the data comes from. Neither does he want to use different tools/languages for different data sources. With respect to users, the easiest way is to use the simple system of natural language questions, for

example, using XML. Context information contains a lot of attributions which are naturally suitable for XML language to describe. Upon query request being sent to the affair handling agent, according to the query data (context or regular data), different query language and manner will be generated. If it is context data, XCQuery language will be generated, through the inquiry interface CI2, to obtain data from different data sources. Finally, when result is returned, it is not the original data, but the semantic data being processed at the affair handling agent.

5.3. XML and Database

XML (Extensible Markup Language) is a simplified subset of SGML which has rich features and ease of use combined with HTML to Web applications. Compared with the HTML XML have many advantages, such as half structural, self-description, extendable capability, and flexibility.

With rapid development of Internet-based business applications, XML has become the data representation and data exchange standard format of the Internet. As a result, the presence of large number of important XML documents on the Internet involves a series of new XML data management. However, under the current situation and development trend, after years of development, relational database technology is well established in the field of data management business and occupies a dominant position. Therefore, to meet these new XML based applications at the same time, we still need to rely on relational database technology, it is necessary to

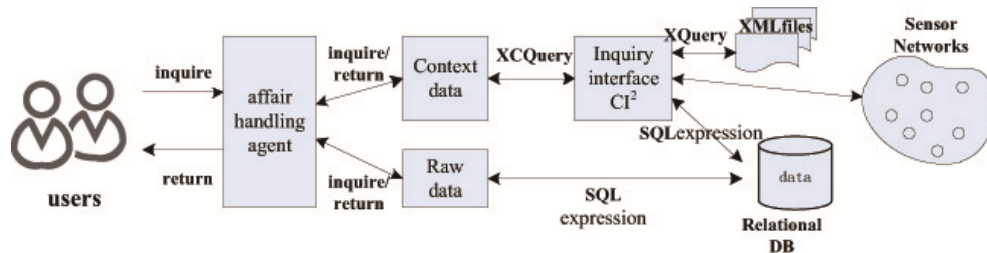


Fig. (4). Semantic information from sensor raw data.

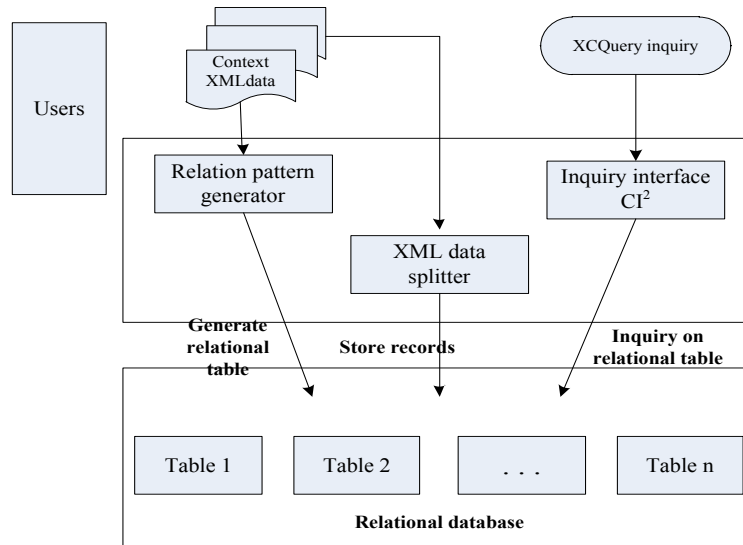


Fig. (5). XML and relational database storage and query.

find a way to implement XML data management using relational database. XML data based relational database management system is storing XML data in relational database systems, and with the maturity of XML technology to achieve a comprehensive data management. Main idea is: First, use XML document schema information to design relational table structure. Then, parse XML documents, combined with pattern information, the contents of the XML are stored in the corresponding relational table, users utilize XQuery query language to request data from the document through specific XQuery query, and then get the right results data from relational database, and the XML document structure is returned to the user. Structure of the system is shown in Fig. (5). By using XML, context information will be mapped to a traditional relational database. The challenge is to achieve the design of relational schema generator, XML data and inquiry interfaces CI2 which is different from ordinary XML documents.

5.4. XML Expression of Context Information

Certain context information may contain a lot of attributions which are naturally suitable for XML language to describe. For example, in this paper, user communication combined with time and location can be described using following XML expression:

```
<?xml?>
<userinteraction>
<!-below describe userinteraction at a location-->
<location name= 'Horton building'>
  <time>
<timestart>2013-11-15-13:24:33</timestart>
<timeend>2013-11-16-13:35:33</timeend>
  </time>
<users> 38</users>
</location>
</userinteraction>
```

6. EXPERIMENT

We conducted experiments to compare our unbalanced slide window method denoted by UbSW with fixed window method denoted by FIX. The tests were running on an open data set of social contact in an urban area. The evaluation metrics is first based on social community discovered and then the data processed by each method.

Fig. (6) shows our method is very close to actual communities. Fixed window separates big community to many so it discovers more than actual amount. Fig. (7) shows UbSW gets good results with small cost.

7. CONCLUSION

This paper brought forward unbalanced sliding window based data mining algorithm to discover social connections between target users. To use sliding window model, we can control the data scale and do incremental mining so as to save storage and reduce computing tasks. By making sliding

window unbalanced, the algorithm can fit the different requirements of system admins. This paper described how existing method can be mapped to social structure mining. It also discussed the representation, storage and inquiry methods of context information, proposed fundamental architecture of the system and some design strategies. The technique detail and some application specific solutions will be carried out in the future work. The Context Objects Database has lots of features that are in accordance with the advance of IT technologies. It can be widely used in logistics activities, smart transportation, firefighting, retail business, medical treatment, manufacturing industry, and so on. We also discuss the application of the COD such as mining social structure using sliding window model.

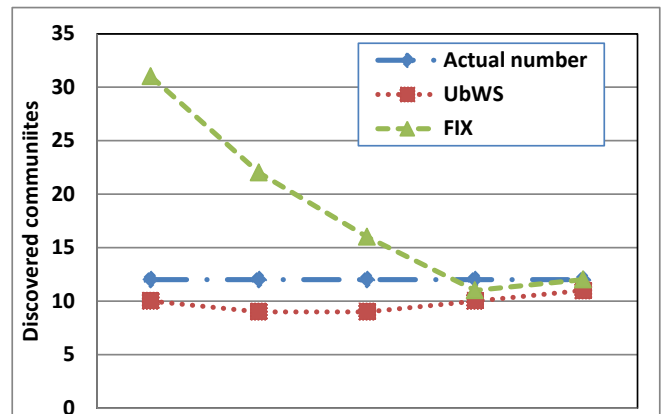


Fig. (6). Number of social communities discovered.

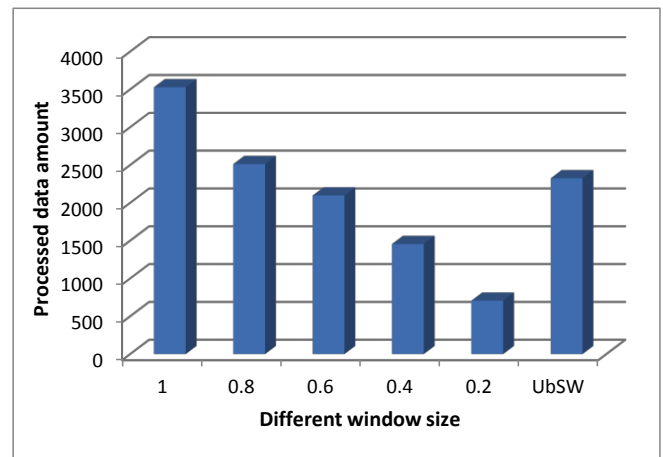


Fig. (7). Amount of data processed in mining.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is supported by the Zhejiang Province Natural Science Foundation (No. Y1110831), Scientific Research Foundation for the Returned Overseas Chinese Scholars,

State Education Ministry, the sub-foundation of Zhejiang Provincial Key Innovation Team on Sensor Networks under Grant No.2009R50046-4.

REFERENCES

- [1] Y.S. Liu, J.L. Xia, and G.P. Xu, "Transactions schedule of embedded database systems", *Journal of Software*, vol. 13, no. 2, pp. 1692-1697, 2002.
- [2] J.L. Xia, "Iacm: An admission control mechanism for embedded real-time database systems", *Chinese Journal of Computers*, vol. 27, no. 3, pp. 295-301, 2004.
- [3] Y.C. Feng, Y.Z. Wang, and Z.S. Gao, "An architecture of mobile database management system", *Journal of Computer Research and Development*, vol. 38, no. 5, pp. 620-624, 2001.
- [4] H.Z. Lin, C. Chen, B. Zhou, and Y.F. Pan, "A dynamic approach to transactional consistency in mobile environments", *Journal of Computer Research and Development*, vol. 39, no. 1, pp. 100-104, 2002.
- [5] J.D. Chen, X.F. Meng, and C.F. Lai, "Clustering objects in a road network", *Journal of Software*, vol. 18, no. 2, pp. 332-334, 2007.
- [6] M.Di Giacomo, "Mysql: lessons learned on a digital library", *IEEE Software*, vol. 22, no. 3, pp. 10-13, 2005.
- [7] D. Barbara, "Mobile computing and databases-a survey", *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 1, pp. 108-117, 1999.
- [8] C. Thompson, "Toward a grid-based DBMS", *IEEE Internet Computing*, vol. 10, no. 3, pp. 87-90, 2006.
- [9] H. Li, C. Ho, M. Shan, and S. Lee, "Efficient maintenance and mining of frequent itemsets over online data streams with a sliding window", In: *Proceedings of SMC*, pp. 2672-2677, 2006.
- [10] C. Xu, Y. Chen, and R. Bie, "Sequential pattern mining in data streams using the weighted sliding window model", In *Proceedings of ICPADS*, pp. 886-890, 2009.
- [11] M. Elfeky, W. Aref, and A. Elmagarmid, "STAGGER: periodicity mining of data streams using expanding sliding windows", In *Proceedings of ICDM*, pp. 1-12, 2006.
- [12] F. Nori, M. Deypir, M. Hadi, and K. Ziarati, "A new sliding window based algorithm for frequent closed itemset mining over data streams", In: *Proceedings of ICCKE*, pp. 249-253, 2011.

Received: June 16, 2015

Revised: August 10, 2015

Accepted: September 19, 2015

© Tingting Fu; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.