다차원 센싱 데이터 스트리밍의 연속 질의 처리*

박영욱O 류우석 장해붕 홍봉희
부산대학교 컴퓨터공학과
ywpark@pusan.ac.kr, wsryu@pusan.ac.kr, jsjzhp@pusan.ac.kr, bhhong@pusan.ac.kr

Continuous Query Processing over Multi-Dimensional Sensing Data Stream

Yongxu PiaoO Wooseok Ryu Haipeng Zhang Bonghee Hong
Department of Computer Science and Engineering, Pusan National University

1. Introduction

Currently, sensor techniques are widely applied in real world applications, such as security, healthcare and industrial automation applications. In such applications, sensor device measure the environment's status. The multi-dimensional sensing information continuously generated as data source to application or database in form of data stream. Many researchers have paid attentions on data stream processing. Main research issues are discussed in [1] [2]. Among the research issues, the continuous query processing is very important to abstract useful information from data stream. Continuous query [3] [4] [5] specifies the filtering conditions over data stream. It evaluates each incoming data item and verifies whether the data is user requested. This paper focuses on continuous query processing over multi-dimensional sensing data stream. In the under control environments, the sensor device is responsible for detecting the unexceptional situation. Each sensor device measure is presented as a multi-attributes tuple. Continuous query check this multi-attributes values to identify the unexceptional situation. A considerable factor in such data stream is that the sensing value change gradually. This observation motivated us to examine the continuous query processing and suggest a predicates scheduling strategy for continuous query processing. Based on the under control sensing data gradual changing property. The predicates scheduling strategy try to firstly evaluate the predicate which is not satisfied with the data item. By this way, reduce the number of predicate evaluations.

2. Body

We examine the continuous query processing over sensing data stream. To process sensing stream data and provide good service to end user, the system should handle a lot of continuous queries efficiently. In this study, continuous query defined as conjunction of several predicates. Each predicate specifies filtering condition on one sensing data dimension. As Figure 1 shown, three continuous queries are registered in host. First continuous query CQ 1 defines three predicates (temperature >10, humidity >50, illumination > 2000). Second continuous query define two predicates (temperature >5, humidity > 80) and third continuous query define three predicates (humidity <30, illumination >4000 and temperature >5). To evaluate these continuous queries, the measured sensing tuple should evaluate through every pipeline structure. In this example, the sensing measure tuple with ID 1 firstly evaluates illumination predicate. Then this sensing tuple evaluates humidity predicate and at last it evaluates temperature predicate. Tuple 1 is result of CQ 1, since this sensing tuple 1 satisfy all three predicates defined in CQ 1. But this sensing tuple is not satisfied with predicate (humidity > 80). It is not CQ 2 result. This sensing tuple will evaluate CQ3 till all continuous queries are evaluated.

Our main contribution is reducing number of the predicate evaluations with predicate scheduling strategy. Proposed predicates scheduling strategy try to firstly evaluate query predicate which is not satisfied by sensing tuple. i.e., the predicate with low selectivity should be evaluation firstly in order to avoid the following predicate evaluation. Every continuous query evaluation plan can be simply considered as a list of predicates. This predicate stores all predicates which included in continuous query. Predicates scheduling strategy is responsible to order this list of predicates. Instead of random predicates evaluation order, this strategy put the predicate which is not satisfied by following sensing tuple on top of list predicates. However, there is no way

*This work was supported by the Grant of Korean Ministry of Education, Science and Technology" (The Regional Core Research Program/Institute of Logistics Information Technology)

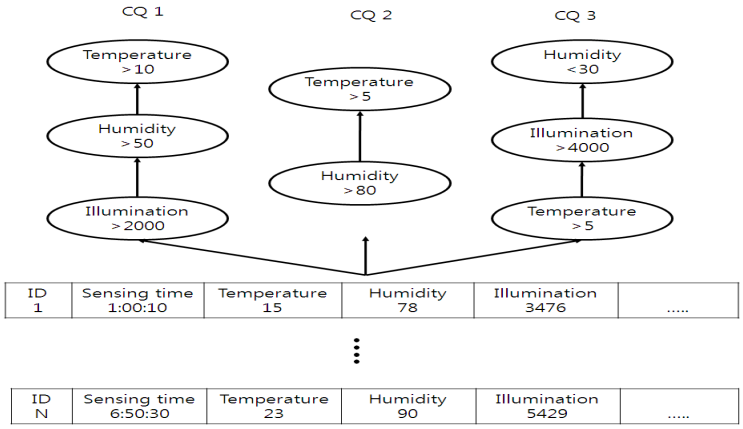

Fig 1. Continuous query processing over multi-dimensional sensing data stream.

to know following sensing values. It is difficult to find the predicate which is not satisfied by following sensing tuples. In our approach, we consider that one sensing value is not satisfied with predicate then the following sensing value has low probability to satisfy this predicates, since the sensing values change gradually. Therefore, we infer that this query predicates has high selectivity for following sensing tuples. To reduce unnecessary predicate evaluations, this predicate is put on bottom of query evaluation plan. That means, this predicate will be evaluated firstly by following sensing tuples. Actually, this approach is not always efficient. In case of great change of adjacent sensing values, even the current sensing value is not satisfied with the specific predicate, we can't sure that the following sensing values also are not satisfied with the predicate. The predicate scheduling can't improve performance of continuous query processing. However, in most environments, the sensing values are change gradually, the predicates scheduling can applied for continuous query processing.

3. Conclusion

In this paper, we mainly consider continuous query processing over multi-dimensional sensing data stream. Based property of the sensing values, adjacent sensing values change gradually, we suggest query predicates scheduling technique. This approach tries to evaluate the query predicate which is not satisfied by the sensing tuple. in this way, system can determine whether this sensing tuple is the result of query or not as early as possible. The redundant predicates evaluations are avoided. At last we compare our proposed predicates scheduling approach with random predicates evaluation order approach. The experiments result show that our approach improve continuous query processing greatly

Reference

[1] L. Golab, M. T. Ozsu, Issues in Data Stream Management, SIGMOD Record, 2003, vol.32.
[2] B. Babcock, S. Babu, M. Datar, R. Motwani, J. Widom, Models and Issues in Data Stream Systems, in Proceedings of the twenty-first ACM SIGMOD- SIGACT-SIGART symposium on Principles of database systems, 2002, p.1-16.
[3] Douglas Terry, David Goldberg, David Nichols, Brian Oki. 1992. Continuous Queries over Append-Only Databases. ACM SIGMOD.
[4] Ron Avnur, Joseph M. Hellerstein. 2000. Eddies: Continuously Adaptive Query Processing. MOD, Dallas, TX USA.
[5] Shivnath Babu, Jennifer Widom. Continuous Quer over Data Streams. SIGMOD Record, Vol.30, No. 3, September 2001