
센서 네트워크 환경에서의 질의 색인 기법

김동현

동서대학교 컴퓨터정보공학부

The Index Scheme for User Queries on A Sensor Network Environment

Donghyun Kim

Division of Computer Information Engineering, Dongseo University

E-mail : pusrover@dongseo.ac.kr

요 약

센서 네트워크는 센서 노드에서 수집된 최신의 현장 데이터를 기반으로 사용자 질의를 처리할 수 있는 시스템이다. 센서 네트워크에서 사용자 질의를 처리하기 위하여 먼저 사용자 질의를 데이터가 저장되어 있는 각 센서 노드에 전파하여 수행한 후에 질의 결과를 서버로 전송한다. 그러나 만약 연속 질의를 처리하면 반복적으로 사용자 질의를 센서 노드에 전파해야 하는 문제가 있다.

이 논문에서는 센서 네트워크에서 센서 스트리밍 데이터에 대하여 연속 질의를 처리하기 위한 질의 처리 기법에 대하여 제안한다. 이를 위하여 각 센서 노드는 자신의 노드에 질의 색인을 구축하여 사용자 질의를 처리한다. 그리고 센서 노드에서 발생 가능한 불확실 센서 데이터를 처리하기 위한 기법을 제시한다.

ABSTRACT

A sensor network system processes user queries using the recent field data collected by each sensor node. To process user queries, the system propagates the queries to the specific sensor nodes which have the relevant data and aggregates the results of the queries. However, if continuous queries are processed by the existing scheme, the system has the problem where the queries are propagated repeatedly.

In this paper, we propose the query processing scheme to process the continuous queries over the sensor streaming data. To do this, each sensor node builds its own query index on its node. And, we present the scheme to deal with the unexpected data rising on the sensor node

키워드

sensor databases, query index, sensor data, continuous query

1. 서 론

센서 네트워크(Sensor Network)는 유비쿼터스 환경을 구축하기 위한 기반 기술로서 각 노드에서 센서들을 통하여 데이터를 자동적으로 수집한 다음에 수집된 데이터는 무선 네트워크를 이용하여 서버로 전송되어진다. 센서 네트워크는 비교적 저렴한 비용으로 현장의 데이터를 수집하여 의사결정에 사용할 수 있기 때문에 다양한 응용에서 사용가능한 차세대 컴퓨팅 기술로서 주목받고 있다[1].

센서 네트워크에서 데이터를 검색하기 위한 기존의 방법은 크게 2가지로 구분되어진다[1]. 첫

번째는 서버에서 검색하는 방법으로 모든 센서 노드로부터 수집된 데이터는 서버로 보내어져서 검색을 수행한다. 두 번째는 네트워크 내부에서 검색하는 방법으로 각 센서 노드에서 자신의 수집 데이터만을 유지하여 검색하는 방법과 특정 위치의 센서 노드를 지정하여 해당 노드에 특정 범위의 데이터만을 저장하여 검색을 수행하는 방법이 있다. 그러나 기존의 방법은 서버 또는 노드에 저장된 데이터에 대하여 검색하는 방법으로 스트리밍 데이터인 센서 데이터를 효율적으로 검색하기 어려운 문제점이 있다.

[2][3][4]에서는 센서 네트워크에서 질의 처리를 하기 위한 기법을 제시하였다. 분산공산색인 기법

[2]은 센서 네트워크에서 공간 질의를 처리하기 위하여 분산공간색인을 구성하였다. peer-tree[3]는 센서 네트워크에서 최근접 질의와 기타 공간 질의를 수행하기 위한 기법을 제시하였다. 질의균등분할[4]은 센서 네트워크에 분산된 데이터를 수집하기 위한 XML 질의 처리 기법을 제시하고 Xpath 질의를 여러 개의 부분 질의로 나누어서 처리하였다. 그러나 이러한 기법들은 센서 노드에 저장되어 있는 데이터에 대하여 질의 처리를 수행하기 위한 기법으로 연속 질의를 처리하기 어려운 문제점이 있다.

본 논문에서는 스트리밍 데이터인 센서 데이터에 대하여 연속 질의를 처리하기 위한 처리 기법에 대하여 제안한다. 제안한 처리 기법은 분산 질의 색인을 기반으로 2단계로 나누어 질의를 처리한다. 또한 불확실 값을 처리하기 위한 2단계 질의 색인에 대하여 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 기술하고 3장에서는 본 논문에서 제안한 분산질의색인의 개념에 대하여 설명한다. 4장에서는 불확실 데이터를 처리하기 위한 2단계 질의 색인에 대하여 기술하고 5장에서는 결론 및 향후 연구에 대하여 기술한다.

II. 관련 연구

분산공간색인(distributed Spatial Index, SPIX) 기법[2]에서는 센서 네트워크에서 공간 질의를 처리할 때의 문제점을 기술하고 이를 해결하기 위하여 SPIX 색인을 제시하였다. SPIX는 센서 노드가 감지하는 범위의 위치를 최소경계사각형으로 구성하고 이를 이용하여 색인을 구성하였다. 그리고 센서 노드가 네트워크에 추가되거나 삭제될 때를 위한 색인 유지 기법을 제안하였다.

피어트리(peer-tree) 기법[3]은 질의를 수행할 때 질의와 관련된 노드만을 접근하기 위한 효율적인 접근 구조에 대하여 제안하였다. 그리고 peer-tree를 이용한 피어간 질의(peer-to-peer query) 수행 모델을 제시하고 peer-tree가 최근접 질의(Nearest Neighbor Query)와 기타 공간 질의를 수행하는데 효율적임을 설명하였다.

질의동일분할(Query Equal Decomposition, QED) 기법[4]은 센서 네트워크에서 분산되어 있는 데이터를 종합하기 위하여 XML 질의 처리 기법을 제안하였다. 이 기법은 Xpath 질의를 여러 개의 부질의로 분할하고 서버에서 검색되지 않는 데이터를 검색하기 위하여 해당 부질의만을 센서 네트워크로 전송한다. 그리고 부질의들의 결과를 종합하여 원 Xpath 질의의 결과를 구성한다. 그러나 이러한 기법들은 센서 노드에 저장되어 있는 데이터에 대하여 질의를 수행하기 때문에 연속 질의를 효율적으로 수행하기 어려운 단점이 있다.

연속 질의를 처리하기 위한 기법으로는 TelegraphCQ[5]와 NiagaraCQ[6]가 있다.

TelegraphCQ 모델[5]은 센서 네트워크에서 발생하는 데이터 흐름의 지연 현상과 사용자 요구에 맞추어 처리 방법을 즉시 변경 가능한 적응형 데이터흐름 구조를 제안하였다. TelegraphCQ는 연속질의를 크게 4가지의 질의로 구분하고 흐름 데이터 및 저장 데이터에 대하여 해당 질의를 처리하기 위한 적응형 튜플 전달 모듈(Eddy)와 상태 모듈(SteMs)을 제시하였다. Eddy는 전달 정책에 따라 튜플을 SteMs 모듈에 전달하고 SteMs은 지연이 발생한 경우에 튜플을 임시 저장하고 새로운 튜플이 입력될 때 조인 작업을 수행한다.

NiagaraCQ[6]는 인터넷 상의 XML문서에 대한 사용자 질의를 효율적으로 처리하기 위하여 XML 문서를 흐름 데이터로 가정한 질의 처리 모델을 제안하였다. 이를 위하여 사용자 질의를 변화기반 질의와 시간기반질의로 구분하고 질의시그니처(query signature)를 이용하여 유사 질의를 그룹화하였다. 그리고 그룹화된 질의간에는 질의 결과를 공유하는 질의 처리 기법을 제시하였다. 그러나 이러한 기법들은 모든 센서 노드의 데이터를 서버로 전송하여 처리하는 검색하기 때문에 센서 노드의 에너지 소비가 심한 단점이 있다.

III. 분산질의색인

연속질의는 데이터 흐름에 대하여 연속적인 검색을 수행하는 질의로 술어의 종류에 따라 크게 4가지로 구분될 수 있다. 첫 번째는 스냅샷 질의로 한 번만 수행되는 질의이고 두 번째는 구간 질의로서 고정된 시작시간과 종료시간을 가진다. 세 번째는 비고정종료 질의로 시작 시간은 고정되어 있으나 종료 시간은 계속하여 증가하는 질의이며 네 번째는 비고정 질의로 시작시간과 고정시간이 고정되지 않고 계속하여 증가하는 질의이다.

만약 연속 질의를 데이터 중심 저장(Data Centric Storage)방식 기반의 인터넷 질의 처리 기법으로 처리하면 해당 질의를 주기적으로 계속 센서 노드로 전파하여 처리해야 한다. 따라서 많은 메시지가 유발되기 때문에 센서 노드의 에너지 소비가 증가되어 효율적인 질의 처리가 어려운 문제점이 있다.

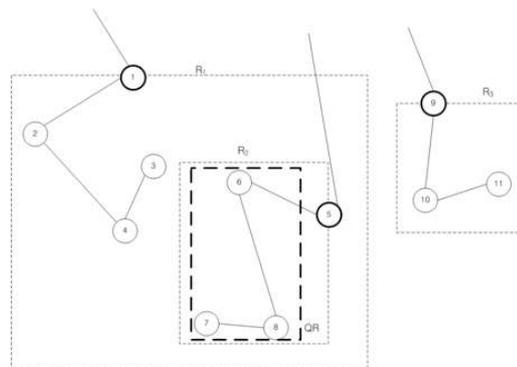


그림 1. 센서 네트워크 색인의 예

예를 들어 “지금부터 5일동안 온도 25℃ 이상, 습도 30% 이상의 온도와 습도를 검색하라”라는 질의범위(QR)가 그림 1과 같이 된 경우에 질의를 수행하기 위하여 노드 6, 7, 8에 질의를 전파해야 한다. 노드 7번을 기준으로 질의를 한 번 전파하기 위하여 최소 4번 이상의 메시지 전파가 필요하며 만약 해당 질의를 매 1분마다 전파한다면 질의 전파를 위해서만 최소 28,800번의 메시지 전송이 필요하다.

따라서 효율적인 연속 질의 처리를 위해서는 그림 2와 같이 혼합형 구조를 기반으로 각 센서 노드에 질의 색인을 구성하는 것이 필요하다. 이 구조에서 질의는 크게 두 단계로 나누어 처리된다. 첫 단계는 각 질의 술어를 해당 노드로 전파하는 단계이다. 두 번째 단계는 특정 노드로 전파된 술어를 질의 색인에 삽입하고 센서 노드에서 수집된 데이터 흐름에 대하여 검색하는 단계이다. 검색된 결과는 서버로 전송되어지며 질의 기간이 종료되면 질의 술어는 해당 노드의 질의 색인에서 삭제된다.

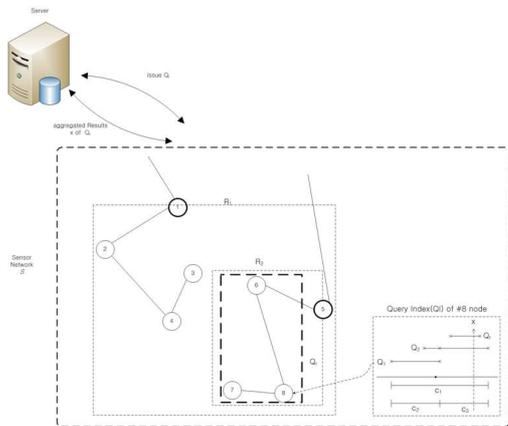


그림 2. 센서 노드의 질의 색인

예를 들어 그림 2와 같이 질의 술어 QR이 노드 6,7,8의 데이터 범위와 겹치는 경우 QR은 6, 7, 8번 노드의 질의 색인에 삽입되어 질의 처리 작업을 수행한다.

IV. 2단계 질의색인

각 노드의 데이터 범위에 속하지 않는 값이 생성될 수 있다. 예를 들어 센서 노드의 기계적인 오류이거나 또는 환경상의 변화가 발생하는 경우에 해당 노드의 데이터 범위에 속하지 않는 센서 값이 생성된다. 그러나 이러한 불확실 값은 발생 노드와는 다른 센서 노드의 질의 색인에 삽입된 술어의 결과 값이 될 수 있다.

불확실 값을 다루기 위한 방법은 크게 두 가지가 있다. 첫 번째는 각 센서 노드의 데이터 범위를 확장하는 것이다. 그러나 이 방법은 불필요한 메시지를 생성하고 정제화 단계의 비용을 증가시키는 문제가 있다. 두 번째는 불확실 값을 계층

구조상의 슈퍼 노드에 전송하여 검색하는 방법이다. 이 방법에서 슈퍼 노드는 서브 노드에서 전송된 값에 대하여 질의 색인을 이용하여 검색하고 만약 일치하면 결과 값으로 구성한다. 그러나 일치하지 않으면 계속하여 슈퍼 노드로 전송하며 만약 서버에서 일치하지 않으면 폐기된다.

이를 위하여 두 가지 종류의 질의 색인을 구축한다. 첫 번째는 자가 질의 색인(Self Query Index, SQI)로 자신의 노드에 삽입된 술어들을 색인하기 위한 색인이다. 노드 자체에서 수집된 데이터 흐름은 SQI를 이용하여 검색한다. 두 번째는 포함 질의 색인(Containing Query Index, CQI)이다. CQI는 데이터 범위의 계층구조 상에서 서브노드들의 술어를 색인하기 위한 색인이다. 서브 노드로부터 불확실 값이 전송되어지면 CQI를 이용하여 검사하고 만약 CQI에서 일치하는 질의 술어가 검색되면 결과 값으로 구성하여 서버로 전송한다.

그림 3은 2단계 질의 색인을 이용하여 질의 처리를 수행하는 예를 보여준다.

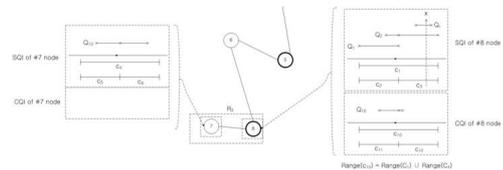


그림 3. 2단계 질의 색인을 이용한 사용자 질의 처리

질의 Q_1 이 사용자에게 의해 시작되면 Q_1 의 술어 영역과 센서 네트워크의 데이터 영역과 비교한다. 만약 Q_1 의 술어 영역이 R_3 와 8번 노드의 데이터 영역과 겹치면 Q_1 의 술어는 8번 노드의 SQI와 8번 노드의 슈퍼 노드들의 CQI들에게 삽입된다. 만약 그림 3과 같이 8번 노드에서 x 값이 발생하면 8번 노드의 SQI와 비교하고 Q_2 와 Q_1 과 일치하는 경우에 Q_2 와 Q_1 의 결과값으로써 서버에 전송한다.

V. 결론

센서 네트워크는 일정 범위의 데이터를 특정 노드에 저장하고 사용자 질의가 발생한 경우에 센서 노드의 데이터 범위에 따라 해당 노드로 질의를 전송하여 처리한다. 그러나 이러한 기법은 연속 질의를 처리하기 위해서 주기적으로 질의를 전파해야 하기 때문에 많은 메시지가 발생하고 노드의 에너지 소모가 심한 단점이 있다.

이 논문에서는 센서 네트워크 환경에서 연속 질의를 처리하기 위하여 분산질의색인 시스템을 제안하였다. 각 센서 노드에서는 데이터 대신에 사용자 질의의 술어를 이용한 질의 색인을 구축하고 수집되는 센서 데이터에 대하여 검색한다. 그리고 불확실 데이터를 검색하기 위하여 2단계 질의색인을 제시하였다. 향후 연구로는 각 센서 노드의 데이터 범위에 대한 색인 구축 기법에 대

한 연구가 필요하다.

참고문헌

- [1] 성동욱,여명호,유재수, "센서 네트워크에서 데이터 저장 및 질의처리를 위한 효율적인 분산 색인 기법", 한국컴퓨터종합학술대회 논문집, vol 36, No 1, pp.129-133, 2009.
- [2] Amir Soheili, Vana Kalogeraki, Dimitrios Gunopulos, "Spatial Queries in Sensor Netwrks", ACM GIS, pp.61-70, 2005.
- [3] Murat Demirbas, Hakan Ferhatosmanoglu, "Peer-toPeer Spatial Queries in Sensor Network", Int'l Conf. on Peer-to-Peer Computing, pp.32-39, 2003.
- [4] Guohua Liu, Shuzhi Zhang, Dongming Zhang, "Research of the Query Technology in Wide Area Sedor Databases", Int'l Conf. on Computational Intelligence for Modelling, Control and Automation, pp.128-133,2006.
- [5] Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Sailesh Krishnamurthy, Sam Madden, Vijayshankar Raman, Fred Reiss, and Mehul Shah, "TelegraphCQ: Continuous Dataflow Processing for an Uncertain World", Innovative Data Systems Research Conference, 2003.
- [6] Jianjun Chen, David J. DeWitt, Feng Tian, Yuan Wang, "NiagaraCQ: A Scalable Continuous Query System for Internet Databases", ACM SIGMOD, pp.379-390,2000.