

# 윈도우를 이용한 스트림 데이터의 실시간 모니터링 기법

Xiaoyue Han, 최옥주, 이민수  
이화여자대학교 컴퓨터공학과  
sowolhan@gmail.com, pensica@naver.com, mlee@ewha.ac.kr

## Real-time Stream Data Monitoring Using Windows

Xiaoyue Han, Ok-ju Choi, Minsoo Lee  
Dept. of Computer Science and Engineering, Ewha Womans University

### 요 약

WBAN(Wireless Body Area Network)과 같은 스트림 데이터의 환경에서는 데이터가 아닌 질의들이 등록되어 있고 데이터들이 끊임없이 시스템에 도착한다. 때문에 도착한 데이터에 대해서 처리할 수 있는 질의만을 찾아 해당 질의들만을 수행하도록 해서 시스템의 질의 부담을 덜어주는 방법이 필요하다. 기존의 단순하고 단편적인 질의의 문제점을 해결하고자 본 연구에서는 Interval Skip List 자료 구조와 시간기반 윈도우를 이용하여 효율적인 실시간 모니터링 시스템을 구현하였다. 특히 산소포화도 생체 센서들로부터 연속적으로 전송되는 스트림 데이터에 대해 다양한 조건을 포함하는 질의들이 실행 되는데 이러한 실시간 모니터링 질의들을 효율적으로 식별하기 위한 질의 인덱스를 설계하였다

### 1. 서론<sup>1</sup>

기존의 많은 시스템들이 WBAN(Wireless Body Area Network) 환경에 기반해 센서들로부터 생체나 정보나 행위 정보를 제공받아 사용하지만, 대부분의 의료분야 응용에서는 데이터에 대한 단순하고 수동적인 모니터링 기법을 사용하고 있는 경우가 많다. 하지만 단편적인 해석뿐만 아니라 연속적인 스트림 데이터와 같은 다양한 조건에 대한 실시간 모니터링 방법이 필요하다.

따라서 본 논문에서는 WBAN 환경의 실시간 모니터링 시스템에서 보다 효율적인 질의 탐색을 위해 인터벌 스킵 리스트(IST, Interval Skip List) 인덱스 구조에 기반하고, 특정 연속된 시간에 대한 보다 윈도우 질의 기능을 비롯해서 다수의 조건으로 구성된 복합 질의를 지원하는 실시간 모니터링 시스템을 설계 및 구현하고자 하였다.

논문의 구성은 다음과 같다. 2 절에서 WBAN(Wireless Body Area Network)와 산소포화도, Interval Skip List에 대해서 각각 기술한다. 3 절에서 윈도우/비윈도우 질의 인덱싱과 Interval Skip List를 적용한 실시간 모니터링

구현 방법에 대해 설명한다. 4 절에서는 순차 탐색에 의한 모니터링과의 비교를 통한 성능 분석을 하고 마지막으로 5 절에서 결론과 향후 연구 과제를 기술한다.

### 2. 관련연구

#### 2.1 WBAN

WBAN은 사람이 착용하는 옷이나 인체에 부착된 여러 디바이스들로 구성된 네트워크를 통해 몸을 중심으로 센서와 구동 기기 간에 데이터 교환을 지원한다. WBAN을 의료 분야에 이용하여 질병에 대해 사전 검진하고 예방할 수 있으며, 모니터링 시스템을 사용해 만성적인 환자나 노약자에 대해 장기적인 건강 상태를 감지하거나 지속적인 상황을 확인할 수 있다.

#### 2.2 산소 포화도

산소 포화도(SpO<sub>2</sub>)는 총 헤모글로빈 농도에 대한 산소와 결합한 산화 헤모글로빈 농도의 백분율을 말하며, 산소 포화도는 그 수치에 따라 여러 단계로 상태를 구분할 수 있다. 일반적으로 산소 포화도 95% 이상을 정상적인 상태로 판단하며, 95% 이하는 저산소혈증으로 각 기준에 따라 그 증상의 정도를 나눌 수 있다. 90에서 94는 가벼운 편이며 75-89는

<sup>1</sup> 이 논문은 정부(교육과학기술부)의 재원으로 한국과학재단 중견연구자 사업(No. 2008-0061488)의 지원으로 수행된 연구결과임.

보통이고 75 미만은 심한 편이다. 산소 포화도 측정은 생명과 직결되는 피측정자의 상태를 빠르게 파악할 수 있다.

### 2.3 Interval Skip List

인터벌 스킵 리스트란 각 노드가 복수의 포인터를 가지고 있는 연결 리스트다. 각 노드의 포인터의 개수  $k$  는 각 노드의 최대 레벨과 같다. 각 노드의 최대 레벨은 아래와 같은 확률  $p$  에 대한 식으로 결정된다.

$$P(k) = \begin{cases} 0 & \text{for } k < 1 \\ (1-p) \cdot p^{k-1} & \text{for } k \geq 1 \end{cases}$$

인터벌 스킵 리스트를 구성하는 임의의 노드에 대해서, 최대 레벨  $k$  인 노드에는  $k$  개의 포인터가 존재한다. 이 때, 노드의  $i$  번째 포인터 ( $k \geq i$ ) 는  $i$  보다 크거나 같은 최대 레벨을 가지는 노드를 가리킨다. 또한, 노드와 포인터에 대한 간격(interval)에 대한 마커(marker)를 가진다.

### 3. 윈도우/비윈도우 질의 인덱싱과 Interval Skip List 를 실시간 모니터링에 적용

단편적인 데이터 해석뿐만 아니라 데이터 값들 간의 유기성을 파악 할 수 있는 시간 기반 슬라이딩 윈도우를 적용하였다. 스트림 데이터  $X$  에 대해 질의  $Q$  가 일정 기간 동안 지속적으로 적용되어 원하는 조건을 만족하는지를 효율적으로 알아내기 위하여 질의 언어와 질의 인덱스를 설계하였다.

그림 1 은 스트림데이터의 입력을 처리하는 과정을 단계적으로 보여주고 있다. 이과정은 제일 먼저 스트림데이터 입력의 원천을 식별하여 비윈도우 질이인가, 윈도우 질이인가, 아니면 두가지를 다 포함하는가를 판단하는것으로부터 시작한다. 일단 결정되면 해당하는 과정의 경로를 따라 다음 단계인 비윈도우 또는 윈도우 경로로 이동한다.

비윈도우질의의 처리과정은 다음과 같다. 입력된 스트림 데이터가 가지고 있는 조건(간격)을 얻기 위하여 비윈도우질의의 Interval Skip List 가 검색된다. 그리고 비윈도우질의의 여러개 조건단위들(Component Expression)을 구성한다. 입력 데이터를 받아 해당 데이터가 만족 시킬 수 있는 모든 C.E 에 대해 만족 플래그를 만족(1)상태로 있는지를 확인한다. 만약 질의의 모든 조건이 만족상태에 있으면 질의는 결과로서 출력된다.

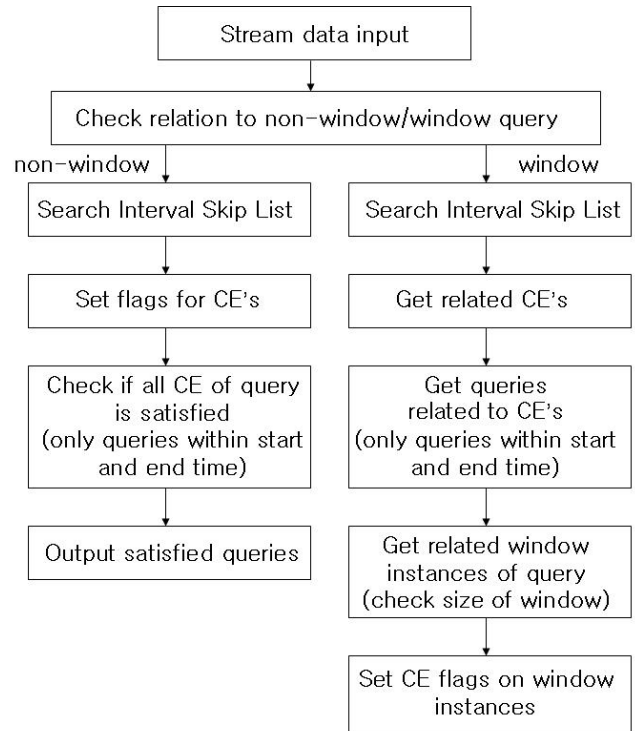


그림 1 윈도우/비윈도우 질의 처리 과정

윈도우질의의 처리과정은 다음과 같다. 입력된 스트림 데이터가 가지고 있는 조건(간격)을 얻기 위하여 윈도우질의의 Interval Skip List 가 검색된다. 위에서 얻은 조건들로 부터 조건단위들(Component expression)이 결정된다. 윈도우질의는 유효시간<Start Time, End Time>내에서만 생성 간격(Term)에 따라 윈도우를 생성한다. 해당 질의는 윈도우 객체들을 가지고 각 윈도우는 해당 윈도우 질의에 대한 윈도우 번호와 해당 윈도우 질의를 구성하는 조건에서 대상으로 삼는 모든 속성에 대해, 조건을 만족하는 모든 데이터 값을 저장하는 배열을 가진다. 윈도우 객체들의 조건들(C.E)을 만족 플래그로 표시한다.

본 연구에서는 사용자의 질의의 폭을 넓히기 위해 데이터 스트림을 이루는 각 속성에 대해 다양한 조건으로 이루어진 복합 질의를 사용하였고, 불필요한 비교 횟수를 줄이기 위해 각 속성마다 독립된 조건 탐색 과정을 거친 후 조건이 만족하는 질의에 대한 탐색만 수행하도록 하였다.

이 모니터링 윈도우 질의 언어의 윈도우  $W$  는 수행 시간<Start Time, End Time>동안 [Duration] 만큼의 데이터를 포함해 생성된 뒤, 질의의 조건문을 체크하는 윈도우이다.

비윈도우 질의의 경우, 어떠한 질의가 만족되면 그 질의의 플래그가 만족 상태로 바뀌게 되는데 한 번

만족 상태로 들어온 플래그는  $T(t)+1$ 분 동안 유효하다..

#### 4. 성능 분석

설계한 혈중산소포화도(  $PaO_2$  )에 대한 실시간 모니터링 시스템의 성능 분석은 순차 탐색에 의한 모니터링과 비교 평가로 수행하였다. 비윈도우 질의 수에 대한 비교 평가 결과를 바탕으로 센서에서 데이터를 보내는 시간을 제외하고 순수 질의 탐색 시간을 도출해서 비교해보았다.

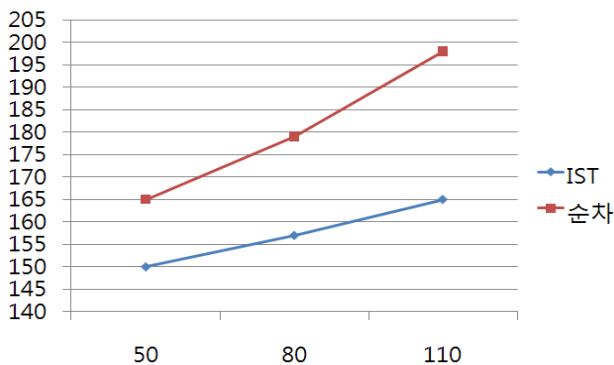


그림 2 수행시간 비교

그림 2 에서 종좌표는 시간(s)을 표시하고 횡좌표는 조건의 개수(개)를 표시한다. 결과그림을 통하여 총 조건 수가 증가함에 따라 수행시간이 큰 폭으로 증가하는 것을 알 수 있다. 따라서 점점 더 처리해야 할 데이터 양은 증가하고, 사용자가 시스템에 요구하는 질의는 복잡해지는 미래에 대비하기에 적합한 구조라고 볼 수 있다.

#### 5. 결론 및 향후 연구

앞으로 과학 기술이 발전함에 따라 센서 네트워크를 구성하는 센서의 종류는 증가하고 실시간 서비스 제공을 위해 서버에서 순간적으로 처리해야 할 데이터의 양은 점차 늘어날 것이며, 사용자가 시스템에 요구하는 질의 또한 복잡해질 것이다. 따라서 본 논문에서는 대량의 질의 탐색에 대해 뛰어난 성능을 보이는 인터벌 스킵 리스트를 적용한 모니터링 시스템 설계를 제안했다.

본 연구의 최종 목표는 WBAN 환경에서 효과적인 메디컬 서비스 제공을 위한 스마트 메디컬 스페이스(m-MediNet)요소 기술을 개발하는 것이다. 따라서 향후 연구는 본 논문에서 구현한 응용 계층으로서의 스트림 데이터 실시간 모니터링 기법을 통신의 물리 계층과 상위 계층과의 통합함으로써 완전한 메디컬 응용 프로토타입을 개발하고 성능 분석을 하여 최적화하는 것이다.

#### 참고문헌

- [1] Christian BÖhm, et al, "Efficiently Processing K-NN Queries on Data Stream", In Proc. the 23th ICDE'07, pp.156-165, Apr. 2007.
- [2] B. S. Kim, S. J. Kim, D. Y. Jung, "홈 헬스 케어를 위한 산소 포화도 모니터링 시스템 구현", 대학전자공학회 학술대회 논문집 II, vol.30, no.1, pp. 1033~1034, 2007
- [3] Eric N. Hanson, "The interval skip list: A data structure for finding all intervals that overlap a point", Algorithms and Data Structure vol. 519, 1991.
- [4] STREAM, <http://infolab.stanford.edu/stream>
- [5] M. Garofalakis, et al, "Quering and Mining Data Stream : You Only Get One Look", In Proc. the 28<sup>th</sup> VLDB'02, Hong Kong, China.