

---

## DQB (Dynamic Query Band): 시계열 데이터의 효율적인 탐색을 위한 동적 쿼리 장치

DQB (Dynamic Query Band): Dynamic Query Device for Efficient Exploration of Time-series Data

조명수\*, 서진욱\*\*\*

---

**요약** 시계열 데이터(time series data)는 특정 대상의 시간적 변동을 지속적으로 관측하여 얻은 값이다. 시계열 데이터의 자료수가 많아짐에 따라 자료의 효율적인 탐색을 위한 장치들이 개발되고 있다. 이 중에서 대화형의 자료 탐색(interactive data exploration)을 위한 동적 쿼리(dynamic query)의 대표적인 장치로 Timeboxes를 들 수 있다. Timeboxes는 관심영역에 해당하는 자료들을 보여주는 사각형 쿼리 장치이다. 사용자 본인의 관심 구간을 간단한 마우스 조작을 통해 사각형 형태로 나타내면 이 사각형 영역 내에 해당되는 자료들이 실시간으로 나타난다. 그러나, 구체적인 쿼리 영역을 설정하기에는 한계가 있고 사용자의 사고 모델(mental model)과는 불일치되는 쿼리 영역을 시각화 한다는 문제점이 있다. 이를 해결하기 위해 본 논문에서는 DQB(Dynamic Query Band)를 제안한다. DQB는 시계열 데이터의 시각화 뷰에서 두께를 가지는 꺾은 선이 이루는 쿼리 영역을 말한다. 이 장치는 직관적이고 조작이 간편한 인터페이스를 제공할 뿐만 아니라, 구체적인 쿼리 영역 설정이 가능하고 좋은 개념 모델(Conceptual model)을 제공해 준다.

**Abstract** Time series data is a sequence of data points, measured typically at successive, spaced at time intervals. Many devices for an efficient exploration is developed according as the items of time series data increase. Among these devices, there is a Timebox widget as a representative device of dynamic query for interactive data exploration. Timeboxes are rectangular query region of interest. The users can draw the region of interest using simple mouse manipulation and the query result sets is displayed. But there is a limitation to represent the concrete query region and Timeboxes visualize the query region inconsistent with the mental model of users. To resolve these problems, we propose a new device called DQB(Dynamic Query Band). DQB is a query region consisting of user defined polyline with a thickness on time series data. This device is possible to concretely specify the query region. Also, it provides a simple and convenient interface and a good conceptual model.

**핵심어:** *Time series data, parallel coordinates, dynamic query, interactive data exploration, Timeboxes, conceptual model, pattern.*

---

본 논문은 2008년 BK21 연구비 지원에 의하여 연구되었음

\*주저자 : 서울대학교 컴퓨터공학과 박사과정 e-mail: msdanbi@cglab.snu.ac.kr

\*\*교신저자 : 서울대학교 컴퓨터공학과 교수 e-mail: jwseo@cse.snu.ac.kr

## 1. 서론

시계열 데이터(time series data)란 특정 대상의 시간적 변동을 지속적으로 관측하여 얻은 값이다. 이 데이터는 주식 시장, 금융시장 및 인구조사 등의 다양한 분야에 존재한다. 이러한 데이터의 분석에는 보통 통계적인 방법이 많이 사용되어 왔다. 컴퓨터의 대중화와 성능향상으로 대용량의 시계열 데이터의 탐색을 위한 가시화 방법이 주가를 비롯한 금융데이터의 분석에 많이 이용되기 시작하였다. 시계열 데이터를 위한 시각화 방법으로는 평행 좌표(parallel coordinates)를 이용한 방법이 효율적인 것으로 알려져 있다. 이 방법은 시간에 따른 자료의 변화를 한 눈에 파악할 수 있다는 장점이 있지만 자료의 개수가 늘어날수록 각 자료를 나타내는 꺾은선들(polylines)이 얽히고 겹쳐져서 자료들을 서로 구분하는 데 어려움이 있다. 이러한 자료의 중첩은 시계열 데이터 전체에 대한 개관(overview)을 파악하기 어렵게 만들 뿐 아니라 중요한 유형(pattern)들이 가려지는 문제점을 야기시킨다. 이러한 문제점을 해결하고 사용자들의 효과적인 자료 탐색을 지원하기 위하여 다양한 정보 시각화 방법들이 개발되었다. 특히 동적 쿼리(dynamic query)는 대용량 시계열 데이터에서 사용자들이 원하는 유형을 지정하고 탐색 결과를 즉각적으로 제시함으로써 사용자들의 탐색 욕구를 충족시켜준다. 평행좌표를 이용한 동적 쿼리의 대표적인 장치로 Timeboxes가 있다[1]. 이 방법은 실시간 쿼리가 가능하고 조작이 간편한 반면 사용자의 관심 구간이 구체적이고 자료의 밀집도가 높을수록 사용자가 그려야 할 Timebox의 수는 증가하게 되는 단점이 있다. 또한 Timebox의 좌우를 통과하는 자료뿐만 아니라 상하를 통과하는 자료까지도 쿼리의 결과에 반영시킴으로써 사용자의 사고 모델과 불일치되는 문제점을 유발한다. 이러한 Timeboxes의 문제점을 해결하기 위해 본 논문에서는 새로운 방식의 쿼리 장치인 DQB(Dynamic Query Band)를 제안한다.

2장에서는 Timeboxes를 비롯한 본 연구와 관련된 연구를 기술하고, 3장에서는 본 논문에서 제안하는 장치를 소개한다. 4장에서는 본 연구에 대한 방법 및 결과를 종합하고, 향후 연구 과제를 모색해 본다.

## 2. 관련 연구

QuerySketch는 사용자가 원하는 유형을 자유 형식 그래프(freehand graph)로 그리면 그와 유사한 유형을 갖는 자료를 찾아준다[2]. 사용자는 시각화 뷰에서 마우스 드래깅을 통하여 자유 형식 그래프를 그림으로써 쿼리를 만들 수 있고, 수정도 가능하다. 또한, 자유 형식 그래프를 특정한 시간대에만 부분적으로 정의할 수 있어 사용자에게 유연성을 제공해 준다. 이러한 QuerySketch는 특정 유형을 가진 자료를 찾아주는 직관적이고 간편한 쿼리 장치이다. 반면, 특정 시

간 범위 내에서 특정 값의 범위를 가지는 자료의 쿼리와 같이 구체적인 쿼리를 하기에는 부족하다.

Ceiling-and-Floor 쿼리는 특정 시간대에서의 데이터의 상한선 또는 하한선을 지정해 주면, 그 사이의 자료들을 표시해 주는 장치이다[3]. 상한선과 하한선은 각각 꺾은선 그래프로 표현된다. 사용자는 시각화 뷰에서 왼쪽 또는 오른쪽 마우스 버튼 클릭을 통하여 제어점을 추가할 수 있고, 이 제어점들은 각각 상한선과 하한선 꺾은선 그래프를 구성하게 된다. Ceiling-and-Floor 쿼리는 사용자가 특정한 시간대에 특정 유형과 값의 범위를 가지는 자료들을 찾기 원하는 경우 효율적인 장치이다.

TimeSearcher는 시계열 데이터의 시각화 및 쿼리, 탐색 등의 다양한 기능을 제공하는 정보 가시화 도구(information visualization tool)이다 [1][5]. TimeSearcher의 쿼리 기능은 Timeboxes에 의해서 제공된다. Timeboxes는 시계열 데이터가 표시된 영역에 사용자에게 의해 지정된 사각형 쿼리 영역을 말한다 [1][6]. Timeboxes 사각형의 가로는 관심 있는 시간의 범위를 나타내고, 세로는 주어진 시간 범위 내에서 값의 범위를 한정짓는 역할을 한다. 해당 영역이 정해지면 그 영역 내에 포함된 데이터 자료들이 표시된다. 사용자는 왼쪽 마우스 드래깅을 통하여 사각형을 만들고, 이것을 이동하고 수정할 수 있다. 이런 인터페이스는 기존의 그리기 프로그램에서 사각형을 그리는 방법과 동일하게 설계되어 사용자에게 친숙함을 제공한다. 또한, 마우스 이벤트가 발생할 때 마다 즉각적인 쿼리 처리가 이루어지고 그 결과가 화면에 반영됨으로써 대화형 데이터 탐색(interactive data exploration)이 가능하다.

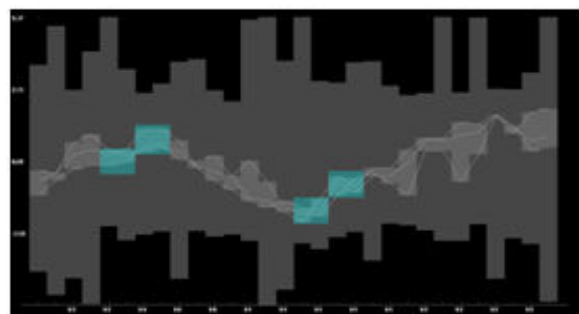


그림 1 Timeboxes를 이용한 쿼리

## 3. 본론

Timeboxes는 사용자에게 친숙한 인터페이스를 제공하면서 동적인 쿼리를 가능케 해 주는 반면 크게 두 가지의 문제점을 가지고 있다.

첫 번째는 사용자가 특정한 유형의 자료를 탐색하기 위해 연속된 시간대의 자료임에도 불구하고, 두 개 이상의 Timebox를 그려야 한다는 것이다.

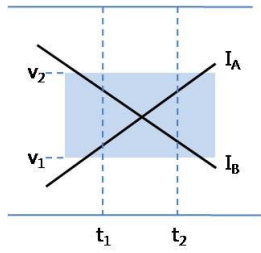


그림 2 Timeboxes를 이용한 잘못된 쿼리 결과

사용자는  $t_1$ 부터  $t_2$ 까지의 연속 시간 구간에서  $v_1$ 과  $v_2$  사이의 값을 가지면서 그 값이 상승하는 자료에 관심이 있다고 가정하자. 그러면 사용자는 그림 2와 같이  $t_1$ 과  $t_2$ ,  $v_1$ 과  $v_2$  구간에 해당하는 사각형을 그리게 된다. 그러나 사용자가 원했던 자료  $I_A$  외에도  $t_1$ 과  $t_2$  구간에서 그 값이 하강하는 자료  $I_B$ 가 함께 나타남을 알게 된다. 이에 따라 사용자는 처음에 그렸던 사각형의 구간을 줄이고, 또 다른 사각형을 그려 그림 3과 같이 구간을 설정하고 난 후, 제대로 된 탐색 결과를 얻게 된다.

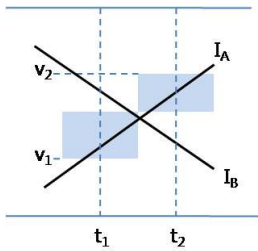


그림 3 원하는 쿼리 결과를 얻기 위해 2개의 timebox 사용

따라서 사용자의 관심 구간이 구체적이고, 시계열 데이터 자료의 밀집도가 높을수록 사용자가 그려야할 Timebox의 수는 증가하게 된다.

또 다른 문제점은 사용자가 그린 사각형과 쿼리 결과 자료들의 꺾은 선들을 함께 보여주는 과정에서 발생하는 시각적인 불일치이다. 그림 4의 경우  $t_2$ 와  $t_3$ ,  $v_1$ 과  $v_2$  구간에 해당하는 자료  $I_A$ ,  $I_B$ ,  $I_C$ 를 쿼리한 결과이다. 그러나 이산적인 데이터 특성상  $t_2$ 와  $t_3$ 의 구간을 표현하기 위해  $(t_1+t_2)/2$ 부터  $(t_3+t_4)/2$ 까지의 영역에 사각형을 보여줌으로써 그림 4의 화살표로 표시된 부분처럼 자료  $I_A$ 와  $I_C$ 가 쿼리 오류라는 착각을 줄 수가 있다. 이는 Timeboxes의 정의에 의하면 올바른 쿼리 결과이지만, 사용자 측면에서는 혼란을 야기시킬 수 있다. 즉, 그림 4의 하늘색으로 표시된 쿼리 영역은 사용자의 사고 모델(mental model)에 의한 쿼리 영역과의 차이로 인해 잘못된 개념 모델(Conceptual model)을 제공한다[4].

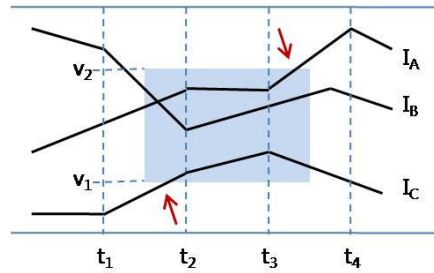


그림 4 Timeboxes의 잘못된 개념 모델

위의 두가지 문제점을 해결하기 위해 본 논문에서는 새로운 방식의 쿼리 장치인 DQB(Dynamic Query Band)를 제안한다. DQB는 시계열 데이터의 시각화 뷰에서 두께를 가지는 꺾은 선이 이루는 쿼리 영역을 말한다. 시간의 축을  $t$ , 값(value)의 축을  $v$ 라고 하고, 시계열 데이터  $D$ 에서의 각 자료를  $I_i$ , 특정 시간  $t_j$ 에서의 자료  $I_i$ 의 값은  $I_i(t_j)$ 라고 하자 (여기서  $i$ 의 범위는  $I$ 부터  $D$ 에 속한 자료의 개수까지 이다).  $(t_j, v_k)$ 에서의 쿼리 결과  $R$ 은 다음과 같이 쓸 수 있다.

$$R = \{I_i | v_k - d_j/2 \leq I_i(t_j) \leq v_k + d_j/2, \forall I_i \in D\} \quad (1)$$

여기서  $d_j$ 는  $(t_j, v_k)$ 에서의 두께를 의미한다.

DQB는 두 개 이상의 시간 구간에 대해서도 쿼리가 가능하다.  $(t_1, v_1)$ 과  $(t_2, v_2)$ 에서의 쿼리 결과  $R$ 은 다음과 같이 쓸 수 있다.

$$R = \{I_i | v_1 - d_1/2 \leq I_i(t_1) \leq v_1 + d_1/2, v_2 - d_2/2 \leq I_i(t_2) \leq v_2 + d_2/2, \forall I_i \in D\} \quad (2)$$

여기서  $d_1, d_2$ 는 각각  $(t_1, v_1), (t_2, v_2)$ 에서의 두께를 의미하고,  $t_1, t_2$ 가 연속 구간의 값일 필요는 없다.

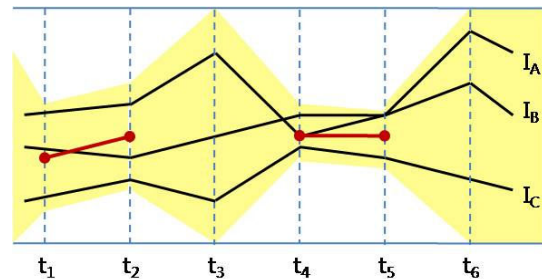


그림 5 두 개의 구간에서 DQB를 이용한 쿼리 결과

그림 5는  $t_1 \sim t_2$  구간과,  $t_4 \sim t_5$  구간에 서로 다른 두께를 적용시켜 쿼리 영역을 설정한 결과이다. 전체 구간에 대한 쿼리 영역은 옅은 노란색으로 표시된다.

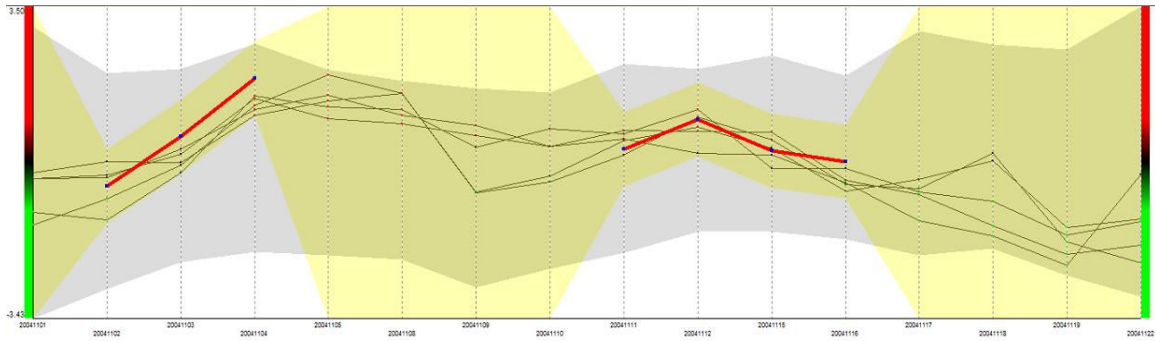


그림 6 DQB가 구현된 뷰

사용자는 특정 시간에 쿼리를 원하는 값에 해당되는 위치에 왼쪽 마우스 클릭을 통하여 제어점(control point)을 생성할 수 있다. 연속된 시간 구간에서의 제어점 생성을 통해 제어점들이 연결된 하나의 꺾은 선 그래프를 표현할 수 있고, 그렇지 않은 시간 구간에서의 제어점 생성을 통해 또 다른 꺾은 선 그래프를 추가할 수 있다. 제어점이 생성될 때 마다 쿼리 영역이 갱신되고, 그 결과는 뷰에 바로 반영된다. 꺾은 선을 그리는 방식은 기존의 그리기 프로그램에서 지원하는 인터페이스와 유사하여 사용자에게 편의를 제공해 준다.

각 제어점에서의 두께는 해당 점 위에서 마우스 휠을 통해 조절이 가능하며, 각 제어점 마다 서로 다른 두께를 설정할 수 있다. 연속적인 두 제어점 사이의 꺾은 선 위에서의 마우스 휠을 하면 두 제어점들의 두께를 동시에 조절하는 것이 가능하다.

또한 생성된 제어점들은 마우스 드래깅을 통해 상하로 위치 이동이 가능하고, Delete 키를 이용해 삭제도 할 수 있다.

그림 6은 DQB가 실제 구현된 뷰이다. 각 자료를 나타내는 꺾은 선들과 같은 방식으로 쿼리 영역을 설정함으로써 Timeboxes에 비해 보다 직관적이고 표현력(expressibility)이 높아졌다. 또한 옅은 노란색 배경을 통해 전체 데이터의 쿼리 영역을 한 눈에 알아볼 수 있게 되어 사용자의 이해도를 증가시켰다. Timeboxes와 달리 DQB의 쿼리 영역은 사용자의 사고 모델과 일치함으로써 좋은 개념 모델을 제공해준다.

#### 4. 결론

시계열 데이터는 다양한 분야에서 널리 사용되며 이를 시각화 하고 동적으로 데이터의 자료를 탐색하는 장치들에 대한 연구가 진행되어 왔다. DQB는 이러한 시계열 데이터의 동적 쿼리 기능을 효율적으로 제공해 주는 장치이다. 사용자들은 간단한 마우스 조작을 통해 꺾은 선 그래프를 그리고, 각 제어점의 두께를 조절함으로써 쿼리 영역을 구체적으로

설정할 수 있다. 사용자에 의해 설정된 쿼리 영역과 쿼리 영역에 대한 쿼리 결과는 실시간으로 나타난다. 또한, 대표적인 쿼리 장치인 Timeboxes와의 비교를 통해 DQB의 유용성을 확인할 수 있었다.

#### 참고문헌

- [1] Hochheiser H, Shneiderman B, "Dynamic query tools for time series data sets: Timebox widgets for interactive exploration", Information Visualization, Volume 3, Number 1, 2004 , pp. 1~18.
- [2] Wattenberg M, "Sketching a graph to query a time-series database", CHI '01 extended abstracts on Human factors in computing systems, March 31-April 05, 2001, Seattle, Washington
- [3] Seo J, Shneiderman B, "Understanding clusters in multidimensional spaces: Making meaning by combining insights from coordinated views of domain knowledge", Technical Report, HCIL-2004-03, 2004
- [4] Donald A. Norman, The design of everyday things, Basic Books, New York, USA, 2002, pp.12~17.
- [5] Hochheiser H, Shneiderman B, "Visual specification of queries for finding patterns in time-series data", Proceedings of Discovery Science, Springer, Berlin, 2001, pp. 441-446.
- [6] Keogh E, Hochheiser H, Shneiderman B, "An augmented visual query mechanism for finding patterns in time series data", Proceedings of the 5th International Conference on Flexible Query Answering Systems, Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, 2002, pp. 240-250.