

# U-Health에서 이벤트 상태 변화를 고려한 시간 마이닝 기법 개발

김 재 인<sup>†</sup> · 김 대 인<sup>\*\*</sup> · 황 부 현<sup>\*\*\*</sup>

## 요 약

U-Health는 다양한 종류의 센서로 환자 정보를 수집하며, 스트림 데이터는 시작 시점과 종료 시점을 갖는 인터벌 이벤트로 요약 가능하다. 그러나 대부분의 시간 데이터 마이닝 기법들은 이벤트 발생 시점만을 고려하며 스트림 데이터의 상태 변화는 간과하는 문제가 있다. 이 논문은 U-Health에서 이벤트 상태 변화를 고려한 시간 마이닝 기법을 제안한다. 제안 방법은 U-Health에서 관심이 있는 이벤트만을 센서에서 서버로 전송함으로써 환경의 제약 사항들을 극복하고 스트림 데이터에 대한 네 가지 이벤트 상태를 정의하여 상태 변화를 고려한 시간 마이닝을 수행한다. 최종적으로, 제안 방법은 이벤트들 사이에 존재하는 인과 관계를 시간 관계 시퀀스로 기술하여 탐사 규칙의 모호함을 제거한다.

키워드 : U-Health, 시간 마이닝, 이벤트 상태, 인과 관계, 시간 관계 시퀀스

## The Development of Temporal Mining Technique Considering the Event Change of State in U-Health

Jaeln Kim<sup>†</sup> · Daeln Kim<sup>\*\*</sup> · BuHyun Hwang<sup>\*\*\*</sup>

### ABSTRACT

U-Health collects patient information with various kinds of sensor. Stream data can be summarized as an interval event which has an interval between start-time-point and end-time-point. Most of temporal mining techniques consider only the event occurrence-time-point and ignore stream data change of state. In this paper, we propose the temporal mining technique considering the event change of state in U-Health. Our method overcomes the restrictions of the environment by sending a significant event in U-Health from sensors to a server. We define four event states of stream data and perform the temporal data mining considered the event change of state. Finally, we can remove an ambiguity of discovered rules by describing cause-and-effect relations among events in temporal relation sequences.

Keywords : U-Health, Temporal Mining, Event State, Cause-and-effect Relation, Temporal Relation Sequence

### 1. 서 론

센서와 무선 통신 기술의 발달로 출현한 U-Health는 생체 신호 감지 기술, 스트림 데이터 관리 기술 등을 의료 분야에 접목하여 원격 진료, 장애우 및 만성 질환자들의 질병 관리를 위한 새로운 서비스 분야이다. U-Health는 센서를 통하여 환자 증상 정보를 수집하며 수집된 데이터는 질병 발생에 대한 다양한 인과 관계 규칙을 포함한다. 따라서 U-Health에서 잠재된 규칙을 탐사하여 현재 이벤트에 기초

하여 향후 발생 가능한 위급 상황을 미리 예측하고자 하는 연구가 진행되고 있다[1,2,3].

U-Health는 전통적인 데이터베이스 시스템에 비하여 저용량의 배터리와 메모리, 소형 프로세스, 제한된 대역폭 등의 제약 사항이 존재하므로 수집된 모든 스트림 데이터를 서버로 전송하는 것은 불가능하며 이러한 제약 사항을 고려한 데이터 처리 기법에 대한 연구가 필요하다[4]. 스트림 데이터는 센서 감지 시점(sensing-time-point)이라는 시간 속성을 가지므로 이에 기초하여 이벤트 발생에 대한 인과 관계 분석이 가능하다[1]. U-Health의 스트림 데이터는 “환자의 맥박이 빠르다”, “환자의 체온이 낮다”와 같은 다양한 이벤트로 요약 가능하며 “체온이 올라가면 맥박이 빨라진다”, “출혈이 발생하면 혈압은 내려가고 맥박은 빨라진다”와 같이 다양한 인과 관계가 존재한다[2]. U-Health는 센서를 사용하므로 무한한 크기의 데이터를 연속적으로 수집하며 이

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2010-0005647).

† 준 회 원 : 전남대학교 전자컴퓨터공학과 박사과정

\*\* 정 회 원 : 전남대학교 전자컴퓨터공학과 시간강사

\*\*\* 중 심 회 원 : 전남대학교 전자컴퓨터공학과 교수(교신저자)

논문접수 : 2010년 9월 8일

수정일 : 1차 2011년 1월 26일, 2차 2011년 3월 24일

심사완료 : 2011년 3월 29일

중 환자의 정상적인 상태에 대한 이벤트보다 환자의 비정상적인 상태를 나타내는 이상 이벤트(abnormal event)가 더욱 중요하다. 그리고 이벤트 발생에 대한 인과 관계 규칙을 탐사함으로써 환자에게 향후 발생 가능한 긴급 상황을 미리 예측할 수 있는 인과 관계 규칙 탐사 연구는 중요하다[2,4].

이 논문은 U-Health에서 이벤트 상태 변화를 고려한 시간 마이닝 기법을 제안한다. 제안 방법의 아이디어는 다음과 같다. 첫째, 센서는 U-Health에서 관심이 있거나 환자의 이상 상태를 나타내는 스트림 데이터만을 서버로 전송하여 환경의 제약 사항을 극복한다. 둘째, 센서 감지 시점이라는 시간 속성과 윈도우를 적용하여 합리적인(reasonable) 인터벌 이벤트를 구축한다. 셋째, 스트림 데이터의 네 가지 이벤트 상태를 정의하고 이를 고려한 인과 관계 규칙 탐사를 수행한다. 넷째, 시간 관계 시퀀스를 구축하여 무한한 크기를 갖는 스트림 데이터를 전통적인 데이터베이스 시스템에서의 트랜잭션과 같은 처리 단위로 구분하고 탐사 규칙의 모호함(ambiguity)을 제거한다. 이 논문의 구성은 다음과 같다. 2절에서는 기존의 스트림 데이터 관리 및 시간 데이터 마이닝 기법에 대한 관련 연구를 기술하고 3절에서는 제안 방법의 기본 개념과 알고리즘을 기술한다. 4절에서는 제안 방법의 효율성을 분석하고, 끝으로 5절에서 결론 및 향후 연구 방향을 기술한다.

## 2. 관련 연구

스트림 데이터 시스템에서의 연관 규칙 탐사 기법은 환경적인 제약 사항으로 인하여 다음 사항을 고려하여야 한다. 첫째 무한한 크기의 스트림 데이터를 의미있는 이벤트로 요약하여야 한다. 둘째, 스트림 데이터를 적절한 데이터 처리 단위로 분리하고 실시간 처리하여야 한다[4,5].

[6]에서는 스트림 데이터에 기초하여 연관 규칙을 탐사하는 PrefixSpan 방법을 제안하였다. PrefixSpan 방법은 연관 규칙 탐사에서 적용되는 후보 항목 수가 무한하다는 점을 고려하여 윈도우 동안에 수집된 스트림 데이터를 이벤트로 요약하고 이에 대한 트리 기반 인덱스를 구축한다. PrefixSpan 방법은 하나의 센서에서 수집된 단일 스트림 데이터에 존재하는 연관 규칙만을 탐사하며 다차원 스트림 데이터의 규칙 탐사는 고려하지 않는 문제가 있다.

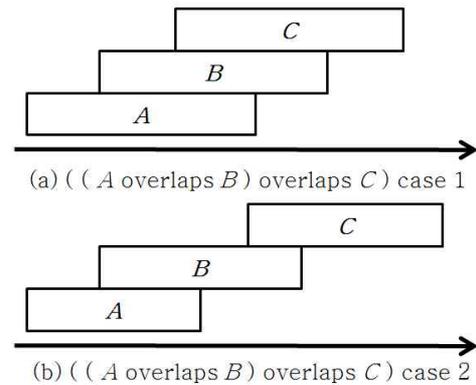
[7]에서는 다차원 데이터에 기초한 연관 규칙 탐사 방법을 제안하였다. 제안 방법은 데이터 수집 시점에 기초하여 일(day), 세션, 웹 페이지와 같은 3개의 차원에 존재하는 데이터 시퀀스를 탐사한다. 그러나 [7]의 방법은 데이터 발생 시점만을 고려하므로 데이터 발생에 대한 선행 관계는 알 수 있지만 인터벌 정보는 고려하지 않으므로 다양한 연관 규칙은 탐사할 수 없는 문제가 있다.

[8]에서는 인터벌 정보에 기초하여 시간 규칙을 탐사하는 TPrefixSpan 방법을 제안하였다. TPrefixSpan 방법은 [6]의 방법을 시간 규칙 탐사 방법으로 확장한 것으로 이벤트 발생 시점과 종료 시점을 포함하는 시간 시퀀스를 구성하여

시간 규칙을 탐사한다. TPrefixSpan 방법은 인터벌 이벤트 구성시 같은 이벤트 타입에 속하는 이벤트는 발생 간격에 상관없이 지속된 것으로 간주하므로 합리적인 인터벌 이벤트를 구축할 수 없는 문제가 있다.

[9]에서는 다수의 센서를 사용하여 스트림 데이터를 수집하여 상관 관계를 분석하는 다차원 스트림 데이터 상관 관계 분석 모델을 제안하였다. [9]의 방법은 스트림 데이터의 실시간성을 고려하여 동적으로 변하는 윈도우 개념을 적용하여 다차원 스트림 데이터들 사이에 존재하는 상관 관계를 분석한다. 그리고 [10]은 윈도우 개념을 적용하여 연관성이 높은 스트림 데이터를 클러스터링하는 기법을 제안하였다. [9,10]과 같은 연구들은 스트림 데이터 시스템에서 수집되는 데이터들 사이에 다양한 관계 및 규칙들이 존재함을 보여준다.

[2]에서는 이벤트 발생 시점에 기초하여 인터벌 이벤트를 구축하고 인과 관계를 탐사하는 시간 연관 규칙 탐사 방법을 제안하였다. [2]의 방법은 이벤트 지속 판단을 위한 임계값을 적용하여 합리적인 인터벌 이벤트를 구축하고 Allen 연산자에 기초하여 두 개의 인터벌 이벤트 사이에 존재하는 인과 관계 규칙을 탐사한다. 그러나 [2]의 방법과 같이 Allen 연산자를 적용하여 인과 관계 규칙을 탐사하는 경우 (그림 1)과 같은 모호함이 존재한다.



(그림 1) 규칙 분석의 모호함

(그림 1)은 세 개의 인터벌 이벤트 A, B, C에 대한 인터벌 관계를 의미한다. (그림 1(a))는 A, B가 지속되는 동안에 C가 발생함을 의미하지만 (그림 1(b))는 A 지속 동안에 B가 발생하고, A 종료 후 C가 발생함을 의미한다. 따라서 B만 발생하는 경우 (그림 1(a))와 (그림 1(b))의 규칙에는 서로 다른 두 개의 분석이 가능하다. (그림 1(a))는 A가 발생하지 않고 B만 발생하는 경우에 C 발생은 예측할 수 없지만 (그림 1(b))는 B 발생만으로 A 발생에 상관없이 C 발생을 예측할 수 있다. 이는 (그림 1(a))와는 달리 C는 A 종료 후 발생 되었으므로 B에만 영향을 받은 것으로 판단할 수 있기 때문이다. 또한 [2]의 방법은 정적인 데이터에 기초한 시간 연관 규칙 탐사 방법으로 동적으로 수집되는 스트림 데이터의 상태 변화와 같은 특성은 반영하지 못하는 문제가 있다.

### 3. 인과 관계 규칙 탐사

시간 속성을 갖는 이벤트들 사이에 존재하는 인과 관계 규칙을 탐사하는 대부분의 관련 연구는 이벤트 발생 시점(point)에 초점을 맞춘다[7,8]. 그러나 이벤트는 일정 시간동안 반복적으로 발생 가능하므로 같은 이벤트 타입에 속하는 이벤트들은 하나의 인터벌 이벤트로 구축 가능하다[1,2]. 인터벌 이벤트는 시점 기반 이벤트에 비하여 다양한 인터벌 관계를 추출할 수 있으며 이러한 관계 속에는 인과 관계 규칙이 존재한다[1,2].

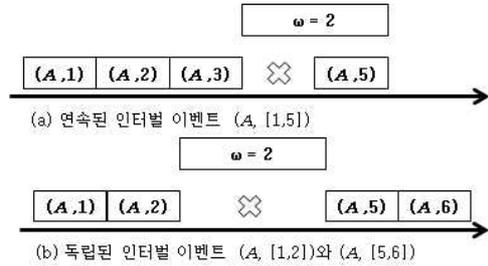
#### 3.1 이벤트 상태 변화

U-Health는 센서를 통하여 “환자의 맥박이 1시 5분부터 1시 20분까지 정상적인 수치보다 빨라졌다”와 같은 인터벌 이벤트를 수집하며, 이들 사이에는 “이벤트 A가 발생하는 동안에 이벤트 B가 발생한다”와 같은 다양한 인과 관계가 존재한다[1,2]. 센서에서 수집한 모든 환자 상태 정보를 전송 및 처리하는 것은 많은 비용을 필요로 한다. 제안 방법은 [4]에서와 같이 U-Health에서 관심이 있거나 환자의 이상 상태를 나타내는 중요 이벤트(significant event)만을 선별하여 센서에서 서버로 전송한다. 중요 이벤트 선정 기준은 개인 특성 및 질병 상태에 따라 센서별로 정의 가능하다[11]. 서버는 수신한 중요 이벤트와 이벤트 지속 임계값을 적용하여 인터벌 이벤트로 요약한다. 그리고 이벤트 관계 임계값을 적용하여 중요 이벤트 예측을 위한 인과 관계 규칙을 탐사한다. 중요 이벤트, 이벤트 지속 임계값, 그리고 이벤트 인과 임계값은 각각 정의 3.1, 정의 3.2, 그리고 정의 3.3과 같다.

**정의 3.1. (중요 이벤트)** 중요 이벤트는 U-Health에서 관심이 있거나 환자의 이상 상태를 나타내는 스트림 데이터를 이벤트  $(E, t)$ 로 기호화 한 것을 말한다. 이 때  $E$ 는 정의된 이벤트를,  $t$ 는 센서 감지 시점을 의미한다.

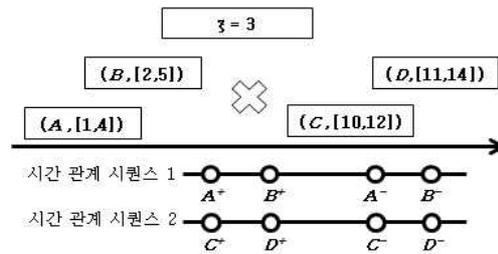
**정의 3.2. (이벤트 지속 임계값  $\omega$ )** 이벤트 지속 임계값  $\omega$ 는 시점 기반의 중요 이벤트를 합리적인 인터벌 이벤트로 요약하기 위한 시간 윈도우를 말한다. 즉  $\omega$  이내에 수신한 동일한 이벤트  $E$ 는 지속적으로 발생된 것으로 간주되며 인터벌 이벤트  $(E, E', E'')$ 로 표현한다. 이 때  $E$ 는 정의된 이벤트를,  $E'$ 와  $E''$ 는  $E$ 의 시작 시점과 종료 시점을 나타내며, 이는 이벤트  $E$ 가  $E'$ 부터  $E''$  구간까지 지속되었음을 의미한다.

**정의 3.3. (이벤트 인과 임계값  $\xi$ )** 이벤트 인과 임계값  $\xi$ 는 이전에 종료된 이벤트가 현재 시점에 지속중인 이벤트 발생에 영향을 미칠 수 있는 시간 윈도우를 말한다. 즉 이벤트  $A$ 가  $\xi$  이내에 발생된 이벤트  $B$ 에 영향을 준다는 것을 의미하여 이는 전통적인 데이터베이스 시스템에서 하나의 트랜잭션에 포함된 이벤트  $A$ 와  $B$ 를 의미한다.



(그림 2) 이벤트 지속 임계값

(그림 2)는 이벤트 지속 임계값  $\omega$ 를 적용하여 같은 타입의 이벤트가 중간에 단절되어 발생하는 경우 이를 연속된 인터벌 이벤트로 판단할 것인지, 아니면 독립된 인터벌 이벤트로 판단할 것인지를 보여준다. (그림 2(a))는 이벤트 지속 임계값  $\omega$ 가 2이고 같은 타입의 이벤트 A가 단위 시점 1, 2, 3, 5에 수집된 경우의 인터벌 이벤트 구성을 보여준다. 이벤트 A는 시점 3에 발생한 후 시점 4에 센서에서 감지되지 않았지만 다시 시점 5에 감지되었다. 적용하는  $\omega$ 가 2이므로 시점 5에 수집된 이벤트 A는 시점 1 이후로 지속되어 발생된 것으로 간주되며 따라서  $(A, [1,5])$ 와 같은 인터벌 이벤트가 구축된다. (그림 2(b))는 같은 타입의 이벤트 A가 단위 시점 1, 2, 5, 6에 수집된 경우의 인터벌 이벤트 구성을 보여준다. A는 단위 시점 2에 발생 후 감지되지 않았으며 다시 시점 5에 감지되었다. 이벤트 지속 임계값  $\omega$ 가 2이므로 초기에 수집된 A의 마지막 시점인 2와 재수집된 시점 5와의 간격이  $\omega$ 보다 크므로 시점 5에 발생된 A는 같은 타입에 속하는 이벤트가 독립적으로 발생된 것으로 간주 되고 두 개의 인터벌 이벤트  $(A, [1,2])$ 와  $(A, [5,6])$ 가 구축된다.



(그림 3) 이벤트 인과 임계값

U-Health에서 한 명의 환자에 부착된 다양한 종류의 센서를 통하여 수집하는 다차원 스트림 데이터는 환자가 센서를 떼어내기까지 수행되는 장기 트랜잭션(long transaction)이다. 제안 방법은 U-Health의 장기 트랜잭션을 의미있는 여러 개의 부트랜잭션(sub transaction)으로 분할하기 위한 시간 윈도우로 이벤트 인과 임계값  $\xi$ 를 적용한다. (그림 3)은 이벤트 인과 임계값  $\xi$ 를 적용하여 구축되는 시간 관계 시퀀스 구축 예를 보여준다. (그림 3)에서 4개의 인터벌 이벤트  $(A, [1,4])$ ,  $(B, [2,5])$ ,  $(C, [10,12])$ ,  $(D, [11,14])$ 가 구축되었고  $\xi$ 는 3이다. 따라서 인터벌 이벤트 A와 B는 하나의 부트랜잭션에 포함된 이벤트로 간주되며 (그림 3)과 같은

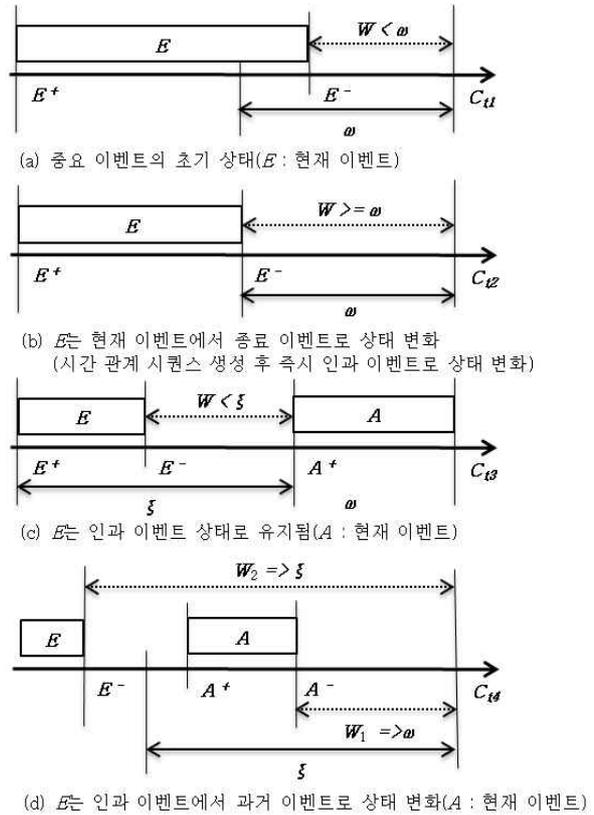
시간 관계 시퀀스 1이 구축된다. 시간 관계 시퀀스에 포함된 인터벌 이벤트 중 가장 마지막 시간인 시점 5 이후 이벤트 인과 임계값  $\xi$  이내에 발생된 다른 인터벌 이벤트가 있다면 이 인터벌 이벤트는 시간 관계 시퀀스 1에 확장되어 추가된다. 시간이 지나 인터벌 이벤트  $B$  종료 후 인터벌 이벤트  $C$ 가 감지되었고  $B$ 의 종료 시점과  $C$ 의 시작 시점과의 간격이  $\xi$ 보다 크므로 이벤트  $B$ 와  $C$ 는 영향력을 주는 관계가 아니므로  $C$ 는 새로운 시간 관계 시퀀스 2를 생성하여 저장된다. 그리고 이후에 발생된  $D$ 를 시간 관계 시퀀스 2에 포함하여 구성하여, 최종적으로 (그림 3)과 같은 두 개의 시간 관계 시퀀스가 구성되며 이는 전통적인 데이터베이스 시스템에서의 트랜잭션 개념과 유사하다.

제안 방법은 이벤트  $E$ 가 마지막 발생 시점  $E^-$  이후 이벤트 지속 임계값  $\omega$  동안 발생하지 않으면  $E$ 는 종료된 것으로 간주한다. 그리고 이벤트 인과 임계값  $\xi$ 를 적용하여 이벤트 종료 시점과 현재 시점에 지속되어 발생하는 인터벌 이벤트의 시작 시점과의 차이가  $\xi$ 보다 작은 경우 이벤트 발생에 영향을 준다고 판단한다. 제안 방법에서 센서에서 수집된 스트림 데이터는 정의 3.4와 같이 상태 변화한다.

**정의 3.4. (이벤트 상태 변화)** 스트림 데이터는 현재 이벤트, 종료 이벤트, 인과 이벤트, 그리고 과거 이벤트 순으로 상태 변화한다.

1. 현재 이벤트(current event) : 현재 이벤트는 초기 상태로 현재 시점과 마지막 발생 시점과의 시간 간격이 이벤트 지속 임계값  $\omega$  보다 작은 이벤트 상태를 의미한다. 즉 현재 시점  $C_t$ 를 기준으로  $\omega$  이내에 감지된 모든 이벤트는 현재 이벤트 상태이다.
2. 종료 이벤트(terminal event) : 종료 이벤트는 마지막 발생 시점 이후 이벤트 지속 임계값  $\omega$ 동안 발생하지 않은 이벤트를 의미한다. 즉 현재 이벤트 중에서  $\omega$ 동안 발생하지 않은 이벤트는 종료 이벤트로 상태 변화하며 이 시점에 종료 이벤트와 인과 이벤트 사이에 존재하는 시간 관계 시퀀스를 생성한다.
3. 인과 이벤트(cause and effect event) : 인과 이벤트는 이벤트 발생에 영향을 줄 수 있는 이벤트 상태로 인과 규칙 탐사 이후 종료 이벤트는 인과 이벤트로 상태 변화한다.
4. 과거 이벤트(past event) : 과거 이벤트는 다른 이벤트 발생에 영향을 주지 않는 이벤트 상태를 의미한다. 인과 이벤트 상태에 있는 이벤트  $E$ 의 마지막 발생 시점  $E^-$ 와 현재 이벤트 상태에 있는 이벤트 중에서 센서에서 가장 먼저 감지된(시작 시점이 가장 빠른) 이벤트의 시작 시점과의 시간 간격이 이벤트 인과 임계값  $\xi$ 보다 크면 인과 이벤트는 과거 이벤트로 상태 변화한다. 이는 이벤트  $E$ 가 더 이상 다른 이벤트 수행에 영향을 주지 않는다는 것을 의미하며 이 시점에  $E$ 에 대한 시간 관계 시퀀스에 포함된 인과 관계 규칙을 탐사한다.

(그림 4)는 정의 3.4의 이벤트 상태 변화를 보여주는 예이다. (그림 4(a))는 현재 시점이  $C_{11}$ 인 경우의 이벤트 상태이다. 이벤트  $E$ 의 마지막 발생 시점  $E^-$ 와 현재 시점  $C_{11}$ 과의



(그림 4) 이벤트 상태 변화

시간 간격  $W$ 가 이벤트 지속 임계값  $\omega$  보다 작으므로  $E$ 는 현재 이벤트 상태에 속한다. 즉 이벤트  $E$ 는 마지막 발생 시점  $E^-$  이후 이벤트 지속 임계값  $\omega$  동안은 현재 이벤트 상태이며,  $\omega$  이내에 다시 발생하면 시작 시점  $E^-$ 부터 마지막 발생 시점  $E^+$ 까지 지속중인 이벤트로 간주된다. (그림 4(b))는 현재 시점이  $C_{12}$ 인 경우에 현재 이벤트 상태에 있는 이벤트가 종료 이벤트로 상태 변화함을 보여 준다.  $C_{12}$  시점에서 이벤트  $E$ 는 마지막 발생 시점  $E^-$  이후 더 이상 발생하지 않아  $E$ 와 현재 시점  $C_{12}$ 와의 시간 간격  $W$ 가 이벤트 지속 임계값  $\omega$  보다 크거나 같게 되는 시점에  $E$ 는 현재 이벤트에서 종료 이벤트로 상태 변화한다. 인과 이벤트 상태에 있는 이벤트 중에서 이벤트의 마지막 발생 시점과 종료 이벤트  $E$ 의 시작 시점과의 시간 간격이 이벤트 인과 임계값  $\xi$ 보다 작은 이벤트  $E'$ 를 추출하여 종료 이벤트  $E$ 와 인과 이벤트  $E'$  사이에 존재하는 시간 관계 시퀀스를 생성한다. 연속적으로 이벤트  $E$ 는 인과 이벤트로 상태 변화한다. (그림 4(c))는 현재 시점이  $C_{13}$ 인 경우에 이벤트  $E$ 가 인과 이벤트임을 보여준다.  $C_{13}$  시점에 이벤트  $E$ 의 종료 시점  $E^-$ 와 현재 이벤트 상태에 있는 이벤트  $A$ 의 시작 시점  $A^+$ 와의 시간 간격  $W$ 가 이벤트 인과 임계값  $\xi$ 보다 작으므로 이벤트  $E$ 는 인과 이벤트 상태이다. 이는 이벤트  $E$ 가 다른 이벤트(그림 4(c)에서는 이벤트  $A$ )발생에 영향을 주는 이벤트임을 의미하며 이벤트  $A$ 가 종료 이벤트가 되는 시점에 이벤트  $E$ 와  $A$ 에 대한 시간 관계 시퀀스를 생성하기 위해 유지한다. (그

림 4(d))는 현재 시점이  $C_{t_i}$ 인 경우에 인과 이벤트 상태에 있는 이벤트가 과거 이벤트로 상태 변화함을 보여 준다.  $C_{t_i}$  시점에 현재 이벤트 상태에 있는 이벤트  $A$ 는 마지막 발생 시점  $A^-$  이후 더 이상 발생하지 않아 현재 이벤트 상태에서 종료 이벤트로 상태 변화되며( $W_1 \Rightarrow \omega$ )  $A$ 와 인과 이벤트 상태에 있는 이벤트  $E$ 에 대한 시간 관계 시퀀스를 생성한다. 그리고 이벤트  $E$ 의 종료 시점  $E^-$  이후 이벤트 인과 임계값  $\xi$  이내에 시작된 현재 이벤트가 없으므로( $W_2 \Rightarrow \xi$ )  $E$ 는 과거 이벤트로 상태 변화한다. 이는 정적으로 수행된 트랜잭션 완료와 같은 의미로 이 시점에 최종적으로 구성된 시간 관계 시퀀스에 포함된 이벤트 발생에 대한 인과 관계 규칙을 탐사한다.

3.2 이벤트 인과 관계 규칙 탐사

제안 방법은 U-Health에서 수집된 중요 이벤트를 이벤트 지속 임계값, 이벤트 인과 임계값, 그리고 이벤트 발생 시점에 따라 네 가지 이벤트 상태로 정의하고 이를 고려하여 이벤트 발생에 대한 인과 관계 규칙을 탐사한다. 또한 제안 방법은 [2]와 같이 Allen 연산자에 기초한 인터벌 관계 개념을 사용하지만 탐사 규칙의 모호함을 제거하기 위하여 인터벌 이벤트의 시작 시점과 종료 시점 정보로 구성되는 시간 관계 시퀀스를 구축한다. Allen 연산자를 사용한 인터벌 관계와 이에 대한 시간 관계 시퀀스는 <표 1>과 같다.

<표 1> 인터벌 관계 VS 시간 관계 시퀀스

No	인터벌 관계	시간 관계 시퀀스
1	$x$ before $y$	$x^+ < x^- < y^+ < y^-$
2	$x$ meets $y$	$x^+ < x^- = y^+ < y^-$
3	$x$ overlaps $y$	$x^+ < y^+ < x^- < y^-$
4	$y$ during $x$	$x^+ < y^+ < y^- < x^-$
5	$y$ finishes $x$	$x^+ < y^+ < x^- = y^-$

제안 방법은 7가지 Allen 연산자 중에서 equal 관계 연산자와 start 관계 연산자를 제외한 5가지 관계 규칙만을 고려하여 이벤트 발생에 대한 인과 관계 규칙을 탐사한다. 이는 equal 관계 연산자와 start 관계 연산자는 이벤트  $x$ 와  $y$ 에 대하여 발생 시점이 같은  $x$ 와  $y$ 에 대한 인터벌 관계이므로 이벤트 발생에 영향을 주는 관계라고 분석할 수 없기 때문이다.

제안 방법에서 센서는 중요 이벤트 감지 시점의 스트림 데이터만을 이벤트로 요약하여 서버로 전송한다. 서버는 이벤트 수신 시점에 이벤트 지속 임계값  $\omega$ 를 사용하여 현재 시점과 이벤트 종료 시점의 차가  $\omega$ 보다 큰 현재 이벤트는 종료 이벤트로 상태 변화시키며, 그렇지 않은 현재 이벤트는 센서 감지 시점을 사용하여 인터벌 이벤트를 구축 및 갱신한다. 즉, 서버는 수신한 이벤트가 현재 이벤트 상태이면 해당 인터벌 이벤트의 종료 시점을 센서 감지 시점  $t$ 로 갱신하며, 그렇지 않은 이벤트들은 시작 시점과 종료 시점을

모두  $t$ 로 갖는 인터벌 이벤트를 구축하고 현재 이벤트 리스트에 추가한다. 따라서 이벤트는 초기 감지 시점에 현재 이벤트 리스트에 추가되므로 이벤트 시작 시점에 대한 오름차순 순으로 정렬되어 유지된다. 제안 방법의 중요 이벤트 감지 시점의 이벤트 추출 과정과 서버의 인터벌 이벤트 구축 과정은 각각 [알고리즘 3.1]과 [알고리즘 3.2]와 같다.

---

Input : 센서에서 수집한 스트림 데이터  
 Output : 중요 이벤트(이벤트  $E$ , 센서 감지 시점  $t$ )

---

**For** 센서  $s$ 에 대하여 **do**

**If**(시점  $t$ 에 센서  $s$ 가 중요 이벤트  $E$ 를 감지) **then**

    센서는 시점  $t$ 의 스트림 데이터를 중요 이벤트로 ( $E, t$ )로 요약하여 서버로 전송

---

[알고리즘 3.1] 중요 이벤트 추출 알고리즘

---

Input : 중요 이벤트( $E, t$ ) 집합, 현재 이벤트 리스트  $CEL$ , 이벤트 지속 임계값  $\omega$   
 Output : 인터벌 이벤트 집합  $IES$ , 현재 이벤트 리스트  $CEL$ , 종료 이벤트 리스트  $TEL$

---

**For** 현재 이벤트 리스트  $CEL$ 의 이벤트  $E$ 에 대하여 **do**

**If** ( $|$ 감지 시점  $t - E$ 의 종료 시점  $E^-| <$  이벤트 지속 임계값  $\omega$ ) **then**

**If** ( $E$ 가 중요 이벤트 집합에 존재) **then**

    인터벌 이벤트의 종료 시점  $E^-$ 를  $t$ 로 갱신

**else**

    이벤트  $E$ 를 현재 이벤트 리스트  $CEL$ 에서 제거하고 종료 이벤트 리스트  $TEL$ 에 추가

**For**  $CEL$ 에 포함되지 않은 중요 이벤트  $E$ 에 대하여 **do**

$E$ 에 대한 인터벌 이벤트 ( $E, t, t$ )를 구축하고  $E$ 를 현재 이벤트 리스트  $CEL$ 에 추가

---

[알고리즘 3.2] 인터벌 이벤트 구축 알고리즘

[2]에서는 정적인 트랜잭션의 데이터에 기초하여 인과 관계 규칙을 탐사한다. 그리고 전체 고객에 대하여 인터벌 관계를 포함한 트랜잭션을 갖는 고객 수의 비율을 지지도로 적용한다. 그러나 [2]의 방법은 각각의 고객에 대한 트랜잭션 수가 일정하지 않아 트랜잭션 수가 많은 고객은 상대적으로 인과 관계 규칙을 포함할 확률이 커지는 문제가 있다. 또한 제안 방법은 U-Health와 같이 센서를 통하여 동적으로 수집되는 스트림 데이터에 기초하므로 [2]와 같은 정적인 규칙 탐사 방법의 지지도를 적용하는 것은 바람직하지 않다. 따라서 제안 방법은 시간 관계 시퀀스에 기초하여 이벤트  $A$ 와  $B$  사이에 존재하는 인과 관계 규칙 지지도를 정의 3.5와 같이 정의한다.

**정의 3.5. (인과 관계 규칙 지지도)** 인터벌 이벤트  $A$ 와  $B$ 에 대한 인과 관계 규칙 지지도( $A, B$ )는 전체 시간 관계 시퀀스 수에 대하여 시퀀스 " $A^+ < \dots < B^+$ "를 포함하는 시간 관계 시퀀스 수의 비율을 의미한다.

$$\text{인과 관계 규칙 지지도}(A, B) = \frac{|Sub\_s(A, B)|}{|S|}$$

인과 관계 규칙 지지도( $A, B$ )는 이벤트  $A$ 가 이벤트  $B$  발생에 영향을 주는 인과 관계가 존재함을 의미하며 이러한 규칙이 얼마나 자주 발생하는가를 측정하기 위한 것이다. 여기에서  $|S|$ 는 현재 시점까지 수집된 전체 시간 관계 시퀀스 수를,  $|Sub\_s(A, B)|$ 는 시퀀스 " $A^+ < \dots < B^+$ "를 부분 시퀀스로 포함하는 시간 관계 시퀀스 수를 의미한다.

<표 1>의 인터벌 관계는 이벤트  $x$ 가 이벤트  $y$  발생에 영향을 주는 경우에  $x$ 와  $y$  사이에 존재 가능한 인과 관계 규칙을 의미한다[2]. 그리고 모든 인터벌 관계는 <표 1>과 같이 제안 방법의 시간 관계 시퀀스로 표현 가능하며 공통적으로 " $x^+ < \dots < y^+$ "를 부분 시퀀스로 포함한다. 제안 방법에서 종료 이벤트에 대하여 수행되는 시간 관계 시퀀스 구성 과정은 [알고리즘 3.3]과 같다.

---

Input : 종료 이벤트 리스트  $TEL$ , 인과 이벤트 리스트  $EEL$ , 시간 관계 시퀀스  $TRS$ , 이벤트 인과 임계값  $\xi$

Output : 인과 이벤트 리스트  $EEL$

---

**For each** 종료 이벤트 리스트의 이벤트  $te$ 에 대하여 **do** {  
   **For each** 인과 이벤트 리스트의 이벤트  $ee$ 에 대하여 **do** {  
     **If** ( $te$ 의 시작 시점  $te^+ - ee$ 의 종료 시점  $ee^- < \xi$  인과 이벤트 인과 임계값  $\xi$ ) **then** {  
       **If** ( $te$ 를 포함하는 시간 관계 시퀀스  $TRS$ 가 존재) **then**  
         { 시간 관계 시퀀스  $TRS$ 에  $ee$ 와  $te$ 에 대한 시간 관계 규칙 추가 }  
       **else**  
         {  $ee$ 와  $te$ 에 대한 시간 관계 시퀀스  $TRS$  생성 } }  
     종료 이벤트  $te$ 를 인과 이벤트 리스트  $EEL$ 에 추가 }

---

[알고리즘 3.3] 시간 관계 시퀀스 구성

제안 방법은 시간 관계 시퀀스 구성 후 인과 이벤트 중에서 더 이상 다른 이벤트 수행에 영향을 주지 않는 이벤트를 검색하여 인과 이벤트에서 과거 이벤트로 상태 변화 시키며, 이는 시간 관계 시퀀스 구성 과정에 접근되는 인과 이벤트 수를 최소화하기 위함이다. 현재 이벤트 리스트는 이벤트 시작 시점에 따라 오름차순으로 정렬되어 있으므로 현재 이

벤트 리스트의 헤드 부분에 위치한 이벤트  $E$ 가 현재 이벤트 상태에 있는 이벤트 중에서 가장 빠른 시작 시점을 갖는 이벤트이다. 또한 알고리즘 3.3에 의하여 종료 이벤트는 종료 시점에 인과 이벤트에 순차적으로 추가되므로 인과 이벤트 리스트의 이벤트들은 종료 시점에 따라 오름차순으로 정렬되어 유지된다. 따라서 제안 방법은 현재 이벤트 리스트의 헤드에 위치한 현재 이벤트의 시작 시점과 인과 이벤트 리스트에 있는 이벤트의 종료 시점의 차가 이벤트 인과 임계값  $\xi$  보다 작을 때까지 원인 이벤트 리스트에 있는 이벤트들을 과거 이벤트로 상태 변화시키며 이러한 과정은 [알고리즘 3.4]와 같다.

---

Input : 현재 이벤트 리스트  $CEL$ , 인과 이벤트 리스트  $EEL$ , 이벤트 인과 임계값  $\xi$ , 시간 관계 시퀀스  $TRS$

Output : 인과 이벤트 리스트  $EEL$ , 과거 이벤트 리스트  $PEL$ , 시퀀스 데이터베이스  $DB(TRS)$

---

$Min\_CEL$  = 현재 이벤트 리스트  $CEL$ 의 헤드에 있는 이벤트의 시작 시점  
 $Min\_EEL$  = 인과 이벤트 리스트  $EEL$ 의 헤드에 있는 이벤트의 종료 시점

**While** ( $|Min\_CEL - Min\_EEL| > \xi$ ) {  
   인과 이벤트 리스트  $EEL$ 의 현재 위치에 있는 이벤트  $E$ 를 과거 이벤트 리스트  $PEL$ 로 이동  
    $E$ 가 포함된 시간 관계 시퀀스  $TRS$ 는 시퀀스 데이터베이스  $DB(TRS)$ 에 저장  
   인과 이벤트 리스트  $EEL$ 의 다음 위치로 이동  
    $Min\_EEL$  = 인과 이벤트 리스트  $EEL$ 의 현재 위치에 있는 이벤트의 종료 시점 }

---

[알고리즘 3.4] 인과 이벤트 재구성

제안 방법은 종료 이벤트를 사용하여 시간 관계 시퀀스 구성 후 최종적으로 과거 이벤트에 대한 시간 관계 시퀀스를 사용하여 이벤트 발생에 대한 인과 관계를 탐사한다. 인과 이벤트에서 과거 이벤트로의 상태 변화는 정적인 트랜잭션의 완료와 같은 의미로 이 때 수행되는 인과 관계 추출 과정은 [알고리즘 3.5]와 같다.

---

Input : 시퀀스 데이터베이스  $DB(TRS)$ , 이벤트 인과 임계값  $\xi$

Output : 인과 관계 집합  $CERS$

---

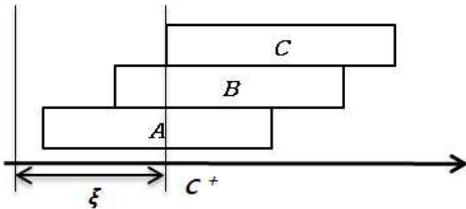
**For** 시퀀스 데이터베이스  $DB(TRS)$ 의 시간 관계 시퀀스  $TRS$ 에 대하여 **do**

시퀀스 TRS에 포함된 이벤트 E의 시작 시점( $E^+$ )을 탐사  
**IF** ( $E^+$ 의 Prefix 중 이벤트 인과 임계값  $\xi$  이내에 종료  
 시점 A가 존재) **then** // Case 1  
 인과 관계  $CER(A, E)$ 을 인과 관계 집합  $CERS$ 에  
 추가  
**IF** ( $E^+$ 의 Prefix 중 시작 시점만 있는 이벤트 집합  
 $\{A, \dots, B\}$ 가 존재) **then** // Case 2  
 인과 관계  $CER(A..B, E)$ 을 인과 관계 집합  $CERS$   
 에 추가

[알고리즘 3.5] 인과 관계 추출 알고리즘

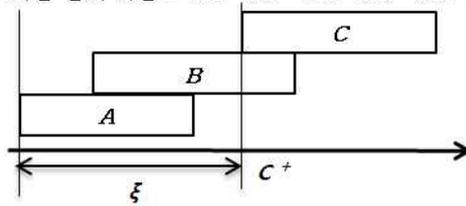
(그림 5)은 [알고리즘 3.5]를 사용하여 시간 [2]의 방법에서 발생하는 (그림 1)과 같은 세 개 이상의 이벤트들 사이에서 발생하는 인터벌 관계의 모호함을 해결함을 보인다. (그림 5(a))는 이벤트 C 발생에 대한 원인 이벤트가 A와 B인 경우에 대한 시간 관계 시퀀스를 보여준다. (그림 5(a))에서 이벤트 C의 시작 시점  $C^+$ 의 prefix는  $A^+ < B^+$ 이므로 이벤트 A와 B의 시작 시점  $A^+$ 와  $B^+$ 만이 존재한다. 따라서 이벤트 C 발생에 이벤트 A와 B가 함께 영향을 주었음을 의미하므로 인과 관계  $CER(AB, C)$ 가 추출된다(Case 2 만족). (그림 5(b))는 이벤트 C 발생에 대한 원인 이벤트로 이벤트 A와 B가 각각 독립적으로 존재함을 의미한다. (그림 5(b))에서 이벤트 C의 시작 시점  $C^+$ 의 prefix는  $A^+ < B^+ < A^-$ 이므로  $C^+$ 의 prefix에 종료 시점이 있는 이벤트 A는 이벤트 C 발생에 영향을 주었음을 의미하므로 인과 관계  $CER(A, C)$ 가 추출된다(Case 1 만족). 또한 prefix 부분에 이벤트 B의 시작 시점  $B^+$ 가 존재하므로 (그림 5(a))와 같이 이벤트 C 발생에 이벤트 B가 영향을 주었음을 의미하므로 인과 관계  $CER(B, C)$ 가 추출된다(Case 2 만족).

시간 관계 시퀀스 :  $A^+ < B^+ < C^+ < A^- < B^- < C^-$



(a) 그림 1(a)의 시간 관계 시퀀스

시간 관계 시퀀스 :  $A^+ < B^+ < A^- < C^+ < B^- < C^-$



(b) 그림 1(b)의 시간 관계 시퀀스

(그림 5) 시간 관계 시퀀스 구성

U-Health는 다양한 종류의 센서를 사용하여 환자 상태 정보를 실시간으로 수집하며, 노약자, 심장 질환, 만성 질환과 같이 환자 상태에 대한 지속적인 모니터링이 필요한 경우가 대부분이므로 현재 수집된 데이터에 기초하여 향후 발생 가능한 이벤트를 미리 예측하는 것은 중요하다[12]. 제안 방법은 스트림 데이터에 대한 이벤트 상태 변화를 실시간으로 고려하여 인과 이벤트에서 과거 이벤트로 상태 변화할 때 구축된 시간 관계 규칙 시퀀스에 포함된 의미있는 인과 관계 규칙을 탐사하며 이러한 과정은 [알고리즘 3.6]과 같다.

Input : 시퀀스 데이터베이스  $DB(TRS)$ , 최소 지지도  $M_{sup}$   
 Output : 인과 관계 규칙 데이터베이스  $DB(CR)$

**Call** 인과 관계 추출 알고리즘(알고리즘 3.5)

**For** 인과 관계 집합  $CERS$ 의 인과 관계  $CER(A, B)$ 에 대하여 **do** {

**For** 시퀀스 데이터베이스  $DB(TRS)$ 의 시간 관계 시퀀스  $TRS$ 에 대하여 **do** {

**IF** ( $TRS$ 가 인과 관계  $CER(A, B)$ 에 대한 시퀀스를 포함) **then**  
 $CER(A, B)$ 의 지지도  $Sup(CER(A, B))$  증가 }

**IF** ( $M_{sup} \leq \frac{Sup(CER(A, B))}{Sup(TRS)}$ ) **then**

인과 관계  $CER(A, B)$ 을 인과 관계 규칙 데이터베이스  $DB(CR)$ 에 추가 }

[알고리즘 3.6] 인과 관계 규칙 탐사

제안 방법의 시간 관계 시퀀스는 [2]에서와 같이 Allen 연산자에 기초한 규칙 표현의 모호함을 제거한다. 제안 방법의 시간 관계 시퀀스가 세 개 이상의 이벤트들 사이에 존재하는 인과 관계 규칙의 정확성을 보장한다는 것은 정리 3.1과 같다.

**정리 3.1** 시간 관계 시퀀스는 세 개 이상의 이벤트들 사이에 존재하는 인과 관계 규칙의 정확성을 보장한다.

(증명) 제안 방법은 인터벌 이벤트에 기초한 시간 관계 시퀀스를 구축하여 인과 관계 규칙을 탐사한다. 증명을 위하여 인과 관계가 존재하는 두 개의 인터벌 이벤트 A와 B에 대한 시간 관계 시퀀스  $A^+ < B^+ < A^- < B^-$  와  $A^+ < B^+ < A^-$  를 가정하자. 두 개의 인터벌 이벤트 사이에 존재하는 인과 관계 규칙은 종료 시점에 상관없이 선행하여 발생한 이벤트 A가 후에 발생한 이벤트 B에 영향을 줌을 알 수 있다. 제안 방법의 시간 관계 시퀀스에서도 이벤트 B의 시작 시점  $B^+$ 보다 이벤트 A의 시작 시점  $A^+$ 가 선행하므로 제안 방법의 [알고리즘 3.5]의 Case2에 따라 인과 관계  $CER(A, B)$ 를 탐사한다.

두 개의 인터벌 이벤트에 인터벌 이벤트  $C$ 가 추가되어  $A < B < C$  순서로 발생하는 경우 발생 가능한 인과 관계 규칙은 다음과 같다. 첫째, 이벤트  $A$ 와  $B$  종료 이전에 이벤트  $C$ 가 발생하는 경우, 둘째,  $A$  종료 후  $B$  종료 이전에  $C$ 가 발생하는 경우, 셋째,  $A, B$  종료 후  $C$ 가 발생하는 경우로 나누어진다. 첫 번째 경우에 제안 방법은 시간 관계 시퀀스  $A^+ < B^+ < C^+ < A^- < B^- < C^-$ 를 구성한다. 따라서 이벤트  $C$ 의 시작 시점  $C^+$ 의 prefix에 이벤트 시작 시점  $A^+, B^+$ 가 존재하므로 [알고리즘 3.5]의 Case 2에 따라 인과 관계  $CER(AB, C)$ 를 탐사하며 이벤트  $C$  발생에 이벤트  $A$ 와  $B$ 가 함께 영향을 줌을 의미한다. 두 번째 경우에 제안 방법은 시간 관계 시퀀스  $A^+ < B^+ < A^- < C^- < B^- < C^-$ 를 구성한다. 따라서 이벤트  $C$ 의 prefix에 이벤트 종료 시점  $A^-, B^-$ 가 존재하므로 [알고리즘 3.5]의 Case 1에 따라 인과 관계  $CER(A, C)$ 를 탐사한다. 또한 이벤트  $C$ 의 시작 시점  $C^+$ 의 prefix에 이벤트 시작 시점  $B^+$ 가 존재하므로 [알고리즘 3.5]의 Case 2에 따라 인과 관계  $CER(B, C)$ 를 탐사한다. 이는 이벤트  $C$  발생에 이벤트  $A$ 와  $B$ 가 각각 독립적으로 영향을 미침을 의미한다. 세 번째 경우에 제안 방법은 시간 관계 시퀀스  $A^+ < B^+ < A^- < B^- < C^- < C^-$ 를 구성한다. 따라서 이벤트  $C$ 의 시작 시점  $C^+$ 의 prefix에 이벤트 종료 시점  $A^-, B^-$ 가 존재하므로 [알고리즘 3.5]의 Case 1에 따라 인과 관계  $CER(A, C)$ 와  $CER(B, C)$ 를 탐사하며 이는 이벤트  $C$  발생에 이벤트  $A$ 와  $B$ 가 각각 독립적으로 영향을 미침을 의미한다.

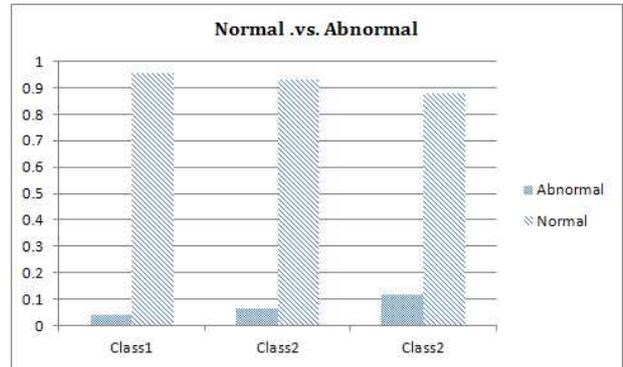
제안 방법은 두 개의 인터벌 이벤트 간에 존재하는 인과 관계 규칙에 다른 인터벌 이벤트가 추가되어 확장되어도 시간 관계 시퀀스에 포함된 이벤트 발생 시점과 종료 시점을 사용하여 이벤트 발생에 대한 인과 관계를 정확하게 탐사함으로써 탐사된 인과 관계 규칙의 정확성을 보장한다.

### 4. 성능 평가

제안 방법은 U-Health에서 관심 대상이 될 수 있는 중요 이벤트만을 선별하여 센서에서 서버로 전송하며 이벤트 발생 시점 및 종료 시점에 기초하여 네 가지 이벤트 상태에 따라 시간 관계 시퀀스를 구축하여 인과 관계 규칙을 탐사한다.

제안 방법은 [13]에서 제공하는 생물의학(biomedical)에 관련된 데이터베이스 샘플 중 에서 심혈관 질환 정보를 적용하여 성능을 평가한다. 심혈관 질환에서 심전도 센서에서 측정된 R파에 대한 지속 정보인 RR 인터벌은 심혈관 질환 발생 예측을 위한 기본 정보이다. 제안 방법의 성능 평가는 [13]에서 제공하는 심혈관 질환자에 대한 연령 및 치료 방법에 따라 세 개의 클래스에 대한 RR 인터벌 샘플을 적용한다. 각 클래스별로 환자에게서 발생하는 비정상적인 RR과 간격을 의미하는 이상 이벤트 발생 비율은 (그림 6)과 같다.

(그림 6)은 클래스 1, 클래스 2, 클래스 3은 각각 286명, 229명, 294명의 환자에 대하여 24시간 동안 수집된 12.5만 개의 RR 인터벌 이벤트 정보에 대한 이상 이벤트 발생 비



(그림 6) 이상 이벤트 발생 비율

율을 보여준다. (그림 6)과 같이 환자에 부착된 센서를 통하여 수집되는 대부분의 RR 인터벌 이벤트 정보는 대부분 환자의 정상적인 상태에 대한 이벤트(그림 6의 Normal 영역)이며 이상 이벤트(그림 6의 Abnormal 영역) 발생 비율은 매우 낮음을 알 수 있다. 스트림 데이터 시스템 환경에서 수집되는 전체 이벤트를 고려하여 규칙을 탐사하는 [6,8]과 같은 기존 연구는 (그림 6)의 Abnormal 영역과 Normal 영역의 모든 이벤트를 서버로 전송하여 처리하지만 제안 방법은 (그림 6)의 Abnormal 영역의 이상 이벤트만을 선별하여 서버에서 처리하므로 센서에서 서버로 전송하는 데이터양과 처리 비용을 최소화할 수 있다.

제안 방법은 이벤트 상태 변화를 반영하여 구축된 시간 관계 시퀀스를 사용하여 연관 규칙을 탐사하며 시간 관계 시퀀스에 포함된 이벤트 발생 시점과 종료 시점에 기초하여 세 개 이상의 이벤트들 사이에 발생하는 연관 규칙의 모호함을 제거한다. 실험을 위하여 3개의 센서에서 환자 정보를 수집하는 환경에서 3,000명의 환자에게서 발생하는 이벤트를 생성하여 실험하였다. 실험 결과 평균적으로 약 13,800여 개의 인터벌 관계가 탐사 되었으며 그 중 (그림 1)과 같이 Allen 연산자의 인터벌 관계 중의 하나인 중첩(overlap) 관계에 의하여 발생하는 모호한 인과 관계 규칙 AR은 이벤트 지속 임계값  $\omega$ 와 이벤트 인과 임계값  $\xi$ 에 따라 <표 2>와 같이 발생하였다.

<표 2> 모호한 인과 관계 규칙

$\xi$	AR 수		$\omega$	AR 수	
	$\omega$	AR 수		$\xi$	AR 수
고정 (8)	3	20	고정 (4)	4	41
	5	32		5	47
	6	39		6	51
	7	51		7	58
	8	52		8	58

<표 2>는 이벤트 지속 임계값  $\omega$  또는 이벤트 인과 임계값  $\xi$ 가 커질수록 시간 관계 시퀀스의 길이가 증가되므로 시퀀스에 포함되는 이벤트 수가 많아지므로 상대적으로 모호한 인과 관계 규칙 발생 빈도가 증가함을 보여준다. 제안

방법은 시간 관계 시퀀스를 구축하여 정리 3.1과 같이 정확하게 인과 관계 규칙을 탐사한다. 즉 [2]와 같은 기존의 연구에서 탐사하는 인터벌 규칙들은 <표 2>와 같이 발생 가능한 모호한 인과 관계 규칙을 포함하고 있지만 제안 방법은 시간 관계 시퀀스를 사용하여 탐사된 인과 관계 규칙의 정확성을 보장한다.

## 5. 결론 및 향후 연구

이 논문은 U-Health에서 이벤트 상태 변화를 고려한 시간 마이닝 기법을 제안하였다. 제안 방법은 센서와 서버로의 데이터 전송을 최소화하고 환자의 정상적인 상태를 나타내는 이벤트에 비하여 상대적으로 발생 빈도가 낮아도 의미 있는 이벤트에 대한 인과 관계 규칙을 탐사하기 위하여 센서는 중요 이벤트만을 서버로 전송한다. 또한 이 논문은 센서의 이벤트 수집 시점에 따라 네 가지 이벤트 상태를 정의하고 이벤트 지속 여부와 이벤트 발생에 영향을 줄 수 있는 이벤트를 결정하여 이벤트들 사이에 존재하는 인과 관계 규칙을 탐사하고 시간 관계 시퀀스를 구축하여 인과 관계 규칙의 모호함을 제거함으로써 정확성을 보장한다. 탐사된 규칙은 환자 개인의 상태 정보(이벤트)가 환자의 질병에 미치는 인과 관계 정보로 활용 가능하며 이를 사용하여 발생 가능한 위급 상황을 미리 예측하고 자가 진단, 환자 행동 개선 등의 분야에 적용 가능하다. 향후 연구 방향으로 의료 전문가의 자문을 받아 U-Health에서 보다 의미있는 인과 관계 탐사 방법을 결과를 실 의료 데이터에 적용하여 탐사 규칙의 유효성을 검증하고자 한다.

## 참 고 문 헌

- [1] H. Han, H. C. Ryoo, and H. Patrick, "An Infrastructure of Stream Data Mining, Fusion and Management for Monitored Patients," Proc. of 19th IEEE International Symposium on Computer-Based Medical Systems(CBMS06), pp.461-468, 2006.
- [2] Y. Lee, J. Lee, D. Chai, B. Hwang, and K. Ryu, "Mining Temporal Interval Relation Rules from Temporal Data," Journal of Systems and Software, Vol.82, pp.155-167, 2009.
- [3] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng, "A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing," IEEE Transactions on Information Technology in Biomedicine, Vol.14, No.3, May, 2010.
- [4] D. Kim, J. Park, and B. Hwang, "Mining Association Rule for the Abnormal Event in Data Stream Systems," Journal of Korea Information Processing Society, Vol.14-D, No.5, pp.483-490, 2007.
- [5] J. Pei, J. Han, B. M. Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach," IEEE Transactions on Knowledge and Data Engineering, Vol.16, No.11, pp.1424-1440, Nov., 2004.
- [6] H. Li, and S. Lee, "Mining Frequent Itemsets over Data Streams using Efficient Window Sliding Techniques," Expert Systems with Applications, Vol.36, pp.1466-1477, 2009.
- [7] C. Yu, and Y. Chen, "Mining Sequential Patterns from Multidimensional Sequence Data," IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.1, pp.136-140, Jan., 2005.
- [8] S. Wu, and Y. Chen, "Mining Nonambiguous Temporal Patterns for Interval-Based Events," IEEE Transactions on Knowledge and Data Engineering, Vol.19, No.6, pp.742-758, Jun., 2007.
- [9] Z. Wong, L. Xiaowei, and C. Chuntian, "Parallel Computing Model fo Multiple Dimensions Data Stream Canonical Correlation Analysis with GPU," Proc. of 2010 2nd International Conference on Information Engineering and Computer Science(ICIECS), pp.1-4, Dec., 2010.
- [10] H. Zhu, Y. Wang, and Z. Yu, "Clustering of Evolving Data Stream with Multiple Adaptive Sliding Window," Proc. of 2010 International Conference on Data Storage and Data Engineering(DSDE), pp.95-100, Feb., 2010.
- [11] J. Kim, C. Na, D. Han, D. Kim, and B. Hwang, "Design of U-Health System based on Embedding Event Filtering," Journal of The Korea Contents, Vol.9, No.2, pp.88-96, 2009.
- [12] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng, "A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing," IEEE Transactions on Information Technology in Biomedicine, Vol.14, No.3, pp.734-740, May, 2010.
- [13] <http://www.physionet.org/physiobank>



### 김재민

e-mail : sereno3@naver.com

2010년 전남대학교 전자컴퓨터공학과  
(석사)

2010년~현 재 전남대학교 전자컴퓨터  
공학과 박사과정

관심분야 : 스트림 데이터 처리, 스트림  
데이터 마이닝, USN응용



**김 대 인**

e-mail : dikim@chonnam.ac.kr  
1998년 전남대학교 전산통계학과  
(이학석사)  
2006년 전남대학교 전산통계학과  
(이학박사)  
2004년~현 재 전남대학교 전자컴퓨터  
공학부 시간강사

관심분야: 스트림 데이터, 데이터 마이닝, 디지털 콘텐츠



**황 부 현**

e-mail : bhhwang@chonnam.ac.kr  
1978년 숭실대학교 전산학과(학사)  
1980년 한국과학기술원 전산학과  
(공학석사)  
1994년 한국과학기술원 전산학과  
(공학박사)

1980년~현 재 전남대학교 전자컴퓨터공학부 교수  
관심분야: 스트림 데이터 마이닝, 분산 시스템, 분산 데이터베이스