

---

# 다차원 스트림 데이터 환경에서 이벤트 가중치를 고려한 시간 관계 탐사

## Discovering Temporal Relation Considering the Weight of Events in Multidimensional Stream Data Environment

---

김재인, 김대인, 송명진, 한대영, 황부현  
전남대학교 전자컴퓨터공학과

Jae-In Kim(sereno3@naver.com), Dae-In Kim(dikim@chonnam.ac.kr),  
Myung-Jin Song(audwls0324@nate.com), Dae-Young Han(abyo@naver.com),  
Bu-Hyun Hwang(bhhwang@chonnam.ac.kr)

---

### 요약

이벤트는 환자의 증상과 같은 시간 속성을 갖는 흐름을 의미하며 센서를 통하여 수집된 스트림 데이터는 시작과 종료 시점을 갖는 인터벌 이벤트로 요약 가능하다. 그러나 대부분의 시간 마이닝 기법은 빈발 이벤트를 고려하며, 빈발하지 않는 이벤트는 중요하더라도 제외되는 문제가 있다. 이 논문에서는 다차원 스트림 데이터 환경에서 인터벌 이벤트에 기초하여 의미있는 시간 관계에 대한 연관 규칙 마이닝 기법을 제안한다. 제안 방법은 이벤트 가중치와 이상 이벤트가 감지된 시점의 스트림 데이터만 고려하여 이벤트의 발생 횟수에 상관없이 의미있는 시간 관계에 대한 연관 규칙을 탐사한다. 그리고 성능 평가를 통하여 제안 방법이 기존의 방법에 비하여 보다 유용한 지식을 탐사함을 보인다.

■ **중심어** : | 다차원 스트림 데이터 | 인터벌 이벤트 | 가중치 | 의미있는 시간 관계 | 연관 규칙 |

### Abstract

An event means a flow which has a time attribute such as a symptom of patient. Stream data collected by sensors can be summarized as an interval event which has a time interval between the start-time point and the end-time point in multiple stream data environment. Most of temporal mining techniques have considered only the frequent events. However, these approaches may ignore the infrequent event even if it is important. In this paper, we propose a new temporal data mining that can find association rules for the significant temporal relation based on interval events in multidimensional stream data environment. Our method considers the weight of events and stream data on the sensing time point of abnormal events. And we can discover association rules on the significant temporal relation regardless of the occurrence frequency of events. The experimental analysis has shown that our method provide more useful knowledge than other conventional methods.

■ **keyword** : | Multiple Stream Data | Interval Event | Weight | Significant Temporal Relation | Association Rules |

---

\* 본 논문은 2009년도 한국연구재단 기본연구지원사업(2009-0076136)에 의하여 연구되었음

접수번호 : #090929-003

심사완료일 : 2009년 12월 08일

접수일자 : 2009년 09월 29일

교신저자 : 황부현, e-mail : bhhwang@chonnam.ac.kr

## I. 서론

센서와 무선 통신 기술의 발달로 인하여 시공간의 제약 없이 데이터를 수집할 수 있는 스트림 데이터 시스템 환경이 출현하였다. 그리고 이러한 환경의 응용 분야는 ICU(Intensive Care Unit) 및 U-health와 같은 의료 분야, 생태계 및 환경 분야, 빌딩 감시 및 이동 물체의 위치 추적 등 매우 다양하다[1-4]. 그러나 스트림 데이터 시스템은 전통적인 데이터베이스 시스템에 비하여 작은 용량의 배터리와 메모리, 소형 프로세서, 제한된 대역폭 등 다양한 제약사항이 존재한다. 따라서 스트림 데이터 시스템 환경에서는 센서에서 수집된 모든 데이터를 손실없이 전송하는 것은 많은 비용을 필요로 하므로 이러한 특성을 고려한 데이터 전송 및 관리 방법에 대한 연구가 필요하다[3][4].

[3][5]에서는 스트림 데이터를 시스템에 정의된 값의 범위에 따라 기호화하여 가격 상승 및 하락과 같은 하나의 흐름(flow)을 의미하는 이벤트로 요약하는 방법을 제안하였다. 스트림 데이터 시스템 환경에서 센서를 통하여 수집된 스트림 데이터는 감지(sensing) 시점과 같은 시간 속성을 갖는다. 그리고 수집된 스트림 데이터들 사이에는 다양한 시간 관계 정보가 잠재되어 있으며 이벤트들 사이에 존재하는 시간 관계에 대한 연관 규칙을 탐사함으로써 이벤트 발생에 대한 원인 요소를 분석할 수 있다[6][7].

다수의 센서를 사용하여 감지하는 객체에 대한 다양한 종류의 스트림 데이터를 수집하는 시스템을 다차원 스트림 시스템이라 한다[3]. ICU나 U-health와 같은 의료 분야에서는 한 명의 환자에 대하여 환자의 체온, 맥박, 혈압, 심장 박동 수와 같이 다양한 정보를 센서를 통하여 수집한다. 다차원 스트림 시스템에서 수집되는 이벤트들 사이에는 “체온과 맥박은 비례 관계이다.”, “출혈이 발생하는 경우 혈압은 내려가고 맥박은 빨라진다.”와 같이 다양한 인과 관계가 존재한다. ICU와 같은 의료 분야에서 환자의 정상적인 체온 및 맥박 상태를 의미하는 이벤트보다는 맥박 하락, 체온 상승과 같은 환자의 비정상적인 상태를 나타내는 이상 이벤트(abnormal event)가 보다 중요하다.

이상 이벤트에 대한 연관 규칙은 발생 횟수가 적더라도 중요도는 매우 높다[4]. 그러나 전통적인 데이터베이스 시스템에서 적용하는 이벤트 발생 빈도에 대한 최소 지지도(minimum support) 중심의 연관 규칙 탐사 방법에서 발생 빈도가 낮은 이벤트는 연관 규칙 탐사 과정에서 제외한다. 그리고 응용 및 시스템에 따라 이벤트의 중요도는 다양하지만 기존의 지지도 중심의 연관 규칙 탐사 방법들은 이러한 이벤트 응용의 특성을 고려하지 않는다.

이 논문에서는 다차원 스트림 데이터 환경에서 의미 있는 시간 관계에 대한 연관 규칙 탐사 방법을 제안한다. 제안하는 방법은 센서와 서버와의 통신을 최소화하기 위하여 센서는 이상 이벤트가 감지된 시점의 스트림 데이터만을 서버로 전송한다. 그리고 이벤트의 중요도에 따라 부여된 가중치(weight)를 적용함으로써 빈발하지 않더라도 중요도가 높은 이벤트에 대한 연관 규칙을 탐사한다. 또한 제안 방법은 이벤트 발생에 대한 원인 분석을 위하여 연관 규칙에 대한 이벤트 발생 영향력을 분석한다.

이 논문의 구성은 다음과 같다. 2절에서는 스트림 데이터 시스템 환경과 시간 데이터 마이닝에 대한 관련 연구를 기술하고, 3절에서는 기존 방법의 문제점, 제안 방법의 기본 개념 및 알고리즘을 기술한다. 4절에서는 실험을 통하여 제안 방법의 효율성을 분석하고, 끝으로 5절에서는 본 연구의 결론 및 향후 연구 방향을 기술한다.

## II. 관련 연구

스트림 데이터에 기초한 연관 규칙 탐사 방법은 전통적인 데이터베이스 시스템에서의 탐사 방법에 비하여 여러 가지 제약사항을 갖는다[1-4]. 첫째, 센서에서 수집되는 데이터는 연속적이며 모든 데이터를 손실 없이 저장할 수 없다. 따라서 스트림 데이터는 센서에서 수집되는 즉시 연관 규칙 탐사에 적용될 수 있는 이벤트로 요약되어야 한다. 둘째, 연관 규칙 탐사를 위하여 추출되는 후보 항목 집합(candidate itemsets)의 수가 무

한다. 센서를 통하여 수집되는 스트림 데이터의 크기는 무한하므로 연관 규칙 탐사에 적용되는 후보 항목 수 역시 무한하다. 이러한 특성을 고려한 데이터 요약 방법으로 일정 시간 구간으로 정의된 윈도우 동안에 수집된 스트림 데이터를 이벤트로 요약하는 방법이 제안되었다[8][9].

[9]에서는 스트림 데이터 환경에서 연관 규칙을 탐사하는 MILE(Mining from multiple strEams) 방법을 제안하였다. MILE 방법은 센서로부터 수집된 이벤트에 대한 트리 기반 인덱스를 구축하여 다차원 스트림 데이터들 사이에 존재하는 연관 규칙을 탐사한다. 또한 MILE 방법은 [8]에서 제안한 PrefixSpan 방법의 반복적인 계층인 계을 최소화하기 위하여 이벤트 선행 관계에 대한 해시 테이블을 유지함으로써 빠른 연관 규칙 탐색이 가능하다. 또한, PrefixSpan 방법은 하나의 센서에서 수집된 스트림 데이터에 대한 연관 규칙만을 탐사하지만 MILE 방법은 둘 이상의 센서에서 수집된 다차원 스트림 데이터에 대한 연관 규칙을 탐사한다. 그러나 MILE 방법은 스트림 데이터의 시간 및 인터벌 정보는 고려하지 않으며 정의된 윈도우 구간 동안에 수집된 이벤트에 대한 연관 규칙만을 탐사한다. 또한 최소 지지도 이상의 발생 빈도를 갖는 이벤트만을 대상으로 연관 규칙을 탐사한다.

[6][7][10]에서는 이벤트 시간 관계에 대한 연관 규칙 탐사 방법을 제안하였다. [10]에서는 정의된 윈도우 구간 동안에 최소 빈발 임계값을 만족하는 이벤트를 탐사하는 에피소드 방법을 제안하였다. 에피소드는 밀접히 관련된 이벤트 시퀀스로 이벤트 시퀀스( $A, B, D, A, C, B, D, A, D$ )에서 이벤트가 매초 발생하고, 윈도우 사이즈가 3초, 최소 빈발 임계값이 50%라면 시퀀스( $A, B$ )가 빈발 에피소드로 추출되며 이벤트  $A$ 와  $B$ 에 대한 연관 정보를 탐사한다.

[6][7]에서는 [10]의 에피소드와는 달리 이벤트 시작 시점과 종료 시점을 의미하는 인터벌 이벤트(interval event)를 구성하여 연관 규칙을 탐사한다. [6][7]의 방법은 빈발 이벤트만을 고려하여 시간 정보에 대한 연관 규칙을 탐사하므로 자주 발생하지 않는 이벤트에 대한 연관 규칙은 탐사하지 못한다.

[11]에서는 최소 지지도 이하로 발생하지만 특정 데이터와 높은 확률로 함께 발생하는 의미있는 희소 데이터 쌍에 대한 연관 정보 탐사 기법을 제안하였다. 제안 방법은 상대 지지도(relative support)에 기반하여 연관 규칙을 탐사함으로써 빈발 지지도 중심의 연관 규칙 탐사 방법에서는 탐사할 수 없는 희소 이벤트 쌍에 대한 연관 규칙을 탐사한다. 또한 [12]에서는 패턴의 중요도에 따라 가중치를 부여하여 패턴 사이에 존재하는 연관 규칙을 탐사하는 WIP(Weighted Interesting Patterns) 방법을 제안하였다. WIP 방법은 패턴의 중요도 및 관심도에 따라 가중치를 높게 부여함으로써 보다 중요한 패턴에 대한 연관 규칙을 탐사할 수 있다. 그러나 [11][12] 방법은 전통적인 데이터베이스 환경에서 수집된 모든 이벤트를 대상으로 연관 규칙을 탐사하므로 스트림 데이터 시스템의 특성을 고려하지 않고 있다.

이 논문에서는 다차원 스트림 데이터 환경에서 의미 있는 시간 관계에 대한 연관 규칙 탐사 방법을 제안한다. 제안하는 방법은 스트림 데이터의 특성을 고려하여 센서에서 수집한 이벤트 중 이상 이벤트 감지 시점의 이벤트만을 서버로 전송한다. 제안 방법은 [6][7]에서와 같이 이벤트 시퀀스에 기초하여 인터벌 이벤트를 구성하며, 신뢰도에 기초한 지지도와 중요도에 따라 이벤트 가중치를 정의함으로써 빈발하지 않더라도 상대적으로 중요한 이벤트에 대한 연관 규칙을 탐사한다. 또한 제안 방법은 탐사된 연관 규칙에 대한 이벤트 발생 영향력을 분석함으로써 이벤트 발생에 대한 원인과 규칙을 찾고 향후 발생 가능한 이벤트 발생을 미리 예측한다.

### III. 이벤트 가중치를 고려한 연관 규칙 탐사

#### 1. 기존 연구의 문제점

스트림 데이터 환경에서 다차원 센서로부터 수집된 스트림 데이터들은 시스템에 정의된 이벤트로 기호화되어 센서에서 서버로 전송한다[3][4]. 스트림 데이터 환경에서 센서에서 수집된 모든 이벤트를 서버로 전송하는 것은 제한된 대역폭의 사용을 증가시키고 서버의 처리 비용을 증가시키는 원인이 된다[1-4]. 또한 센서

에서 수집된 스트림 데이터는 일반적인 경우 객체의 정상적인 상태를 나타내는 이벤트가 대부분이며 이러한 이벤트는 발생 빈도가 높다고 하더라도 중요도는 상대적으로 높지 않다[4].

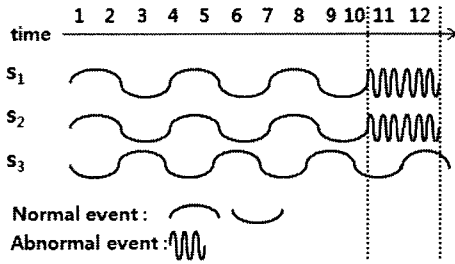


그림 1. 스트림 데이터 예

[그림 1]은 3 종류의 센서를 사용하여 스트림 데이터를 수집하는 시스템의 예이다. 설명을 위하여 [그림 1]에서 각 시간 단위는 초이고 윈도우를 2초로 적용하는 경우 전체 스트림 데이터는 6개의 이벤트로 기호화된다. 센서  $S_1, S_2, S_3$ 에서는 각각 이벤트  $NNNNA, NNNNA, NNNNN$ 을 서버로 전송한다( $N$ :Normal event,  $A$ :Abnormal event). [9]의 MILE 방법은 각 시점의 이벤트 집합을 트랜잭션으로 간주하므로 [그림 1]의 경우  $(NNN), (NNN), (NNN), (NNN), (NNN)$ , 그리고  $(NAA)$ 와 같은 6개의 트랜잭션에 대한 연관 규칙을 탐사한다. 탐사 결과 MILE 방법은 6개의 트랜잭션 중에서 5번 발생한 트랜잭션  $(NNN)$ 의 이벤트를 빈발 이벤트로 추출하며 센서  $S_1, S_2, S_3$ 에 대하여  $N-N-N$ 과 같은 연관 규칙을 탐사한다. 그러나 탐사된 연관 규칙은 객체의 정상적인 상태에 대한 이벤트로 이러한 정보는 센서에 자주 감지되더라도 중요도는 상대적으로 높지 않다. 오히려 발생 빈도는 적더라도 객체의 빈도상태를 나타내는 이벤트가 포함된  $(NAA)$ 에 대한 연관 규칙이 중요도가 높다. ICU 및 U-Health와 같은 의료분야 탐사해자의 정상적인 상태를 나타내는 체온, 맥박, 혈압 등의 정보는 센서에 탐사 자주 수집되더라도 중요도는 높지 않다. 반면에 해자의 빈도상태를 나타내는 맥박, 혈압 등에 대한 정보는 해자의 바이탈 사인으로 발생 빈도가 낮더라도 매우 정상적이며 이러한 이벤트들

의 발생 예측에 적용될 수 연관 연관 규칙 탐사관 보다 정상적해자의 빈MILE과 같은 스트림 데이터 시스템에서의상적인 상태에 대한 이값척이 중요도간 속성 델는 상대지 않는다. 따라서 탐사된 연관 규칙들은 발생된 이벤트는 알 수 있지만 발생하는 이벤트들 사이에 어느 정도의 시간 인터벌이 존재하는지와 같은 시간 관계 정보는 탐사할 수 없다. 그러나 이벤트들 사이에는 이벤트 발생 및 지속성 여부에 기초하여 다양한 시간 관계가 존재한다. 그리고 이러한 시간 관계는 발생된 이벤트에 기초하여 향후 발생 가능한 이벤트를 미리 예측함으로써 발생할 위험 이벤트를 미리 사전에 방지할 수 있다.

제안 방법은 센서에서 객체의 비정상적인 상태를 의미하는 이상 이벤트가 감지된 시점의 정보를 하나의 트랜잭션으로 간주하여 센서에서 서버로 전송함으로써 통신을 최소화하고 상대적으로 발생 빈도가 적은 이상 이벤트를 포함한 연관 규칙을 탐사할 수 있는 가능성을 높일 수 있다. 그리고 이상 이벤트에 보다 높은 가중치를 부여함으로써 보다 중요도가 높은 이벤트를 포함한 연관 규칙을 탐사한다.

## 2. 기본 개념

제안 방법은 센서를 통하여 수집된 스트림 데이터를 이벤트  $E$ 로 요약하며 센서 감지 시점에 대한 시간 속성  $t$ 와 함께 순서쌍  $(E, t)$  형태로 기술한다. 그리고 센서가 감시하는 객체 단위로 이벤트 지속에 대한 요약 정보인 인터벌 이벤트를 구축한다. 예를 들어 U-Health 분야에서 센서를 통하여 환자의 혈압, 맥박 정보를 수집하며 각 환자별로 “맥박이 2초부터 5초까지 빨라졌다”와 같은 이벤트 발생 시점 및 종료 시점에 대한 인터벌 이벤트를 구축한다. 인터벌 이벤트는 이벤트  $E$ 의 시작 시점  $us$ 와 종료 시점  $ve$ 를 포함하고  $(E, [us, ve])$ 로 기술하며  $E$ 의 발생 및 종료 시점은 각각  $E.us$ 와  $E.ve$ 로 기술한다. 인터벌 이벤트들 사이에는 [13]의 Allen 연산자에 근거하여 [표 1]과 같은 인터벌 관계가 존재한다[6][7]. [표 1]의 인터벌 관계 during과 overlap은 Allen 연산자와 달리 이벤트  $x$ 와  $y$ 의 시작 시간에 대한 조건에 등호(=)가 빠져있으며 이러한 이유는  $x$ 와  $y$ 의 시작 시점이

같은 경우 이벤트  $x$ 가 이벤트  $y$  발생에 영향을 주었다고 판단할 수 없기 때문에 본 논문에서는 시작 시점이 같은 이벤트에 대한 시간 관계는 제외하여 during과 overlap에 대한 조건을 [표 1]과 같이 정의한다.

표 1. 인터벌 관계

관계	조건
$before(x,y)$	$x.ve < y.vs$
$during(x,y)$	$(x.vs < y.vs) \cap (x.ve \leq y.ve)$
$overlap(x,y)$	$(x.vs < y.vs) \cap (x.ve \geq y.ve)$

[표 1]의 인터벌 관계는 [그림 2]와 같이 도식화 할 수 있다. 그리고 [그림 2]와 같은 인터벌 이벤트 포함 관계에 인하여 이벤트  $x$ 와  $y$  사이에 인터벌 관계  $IR$ 이 존재한다면  $x$ 는  $y$ 를 발생시키는 원인 이벤트이므로 이벤트  $x$ 가 발생하는 경우 이벤트  $y$  발생을 미리 예측할 수 있다. 예를 들어 인터벌 관계  $before(x,y)$ 는 before 관계에 의하여 이벤트  $x$  종료 후 이벤트  $y$ 가 발생됨을 의미하므로 이러한 관계가 빈발하게 발생한다면  $x$ 는  $y$ 를 발생시키는 원인 이벤트로 간주 할 수 있다. 그리고 인터벌 관계는 무시빈발하 $x$ 와  $y$  사이의 인과 관계 정보만을 추출하고자 하는 경우에는  $x$ 와  $y$  사이에 존재하는 모든 인터벌 관계(before, during, overlap)에 대한 발생 횟수를 모두 합하여 이벤트 발생에 대한 인과 관계 정보를 추출할 수 있다.

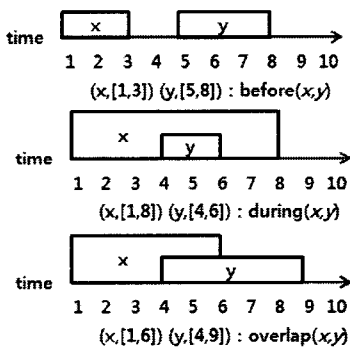


그림 2. 인터벌 이벤트 포함 관계

제안 방법은 다차원 스트림 데이터 환경에서 스트림 데이터들 사이에 존재하는 의미있는 시간 관계에 대한 연관 규칙을 탐사한다. 일반적인 스트림 데이터 응용에서 사용자는 객체의 일반적인 상태를 나타내는 이벤트 보다는 이상 상태에 대한 이벤트에 보다 관심이 있다. [6][7][10]과 같은 시간 데이터마이닝 기법들은 빈발하게 발생하는 이벤트들 사이에 존재하는 연관 규칙을 탐사한다. 그러나 이벤트의 빈발 정도가 이벤트의 중요도를 의미하는 것은 아니다. 또한, 센서를 통하여 수집되는 특정 이벤트는 자주 감지되지는 않더라도 발생한다면 특정 이벤트와 높은 확률로 함께 발생하는 경우가 존재한다. 또한 사용자는 특정 이벤트에 대한 연관 규칙 탐사에 관심이 있을 수 있다. 그러나 기존의 시간 데이터마이닝 기법들은 이러한 특성을 반영한 연관 규칙을 탐사할 수 없다. 따라서 제안 방법은 사용자 관심도 및 이벤트 중요도를 반영하고 자주 발생하지는 않지만 특정 이벤트와 함께 발생하는 이벤트 쌍에 대한 연관 규칙을 탐사하기 위하여 다양한 지지도를 연관 규칙 탐사 과정에 적용한다. 제안 방법은 가중치를 고려한 빈발 이벤트들 사이에 존재하는 연관 규칙을 탐사하기 위하여 정의 1과 같은 1차 가중 지지도(1st weight support)를 적용한다.

정의 1. 1차 가중 지지도

: 1차 가중 지지도는 사용자 관심도를 반영한 빈발 이벤트에 대한 연관 규칙을 탐사하기 위한 최소 지지도로 전체 고객 중 이벤트  $A$ 가 발생한 고객 비율과 전체 이벤트들의 평균 가중치에 대하여  $A$ 가 발생한 고객 수와  $A$ 의 가중치를 반영한 비율을 의미한다.

$$1차\ 가중\ 지지도 = \frac{|A| \times W_A}{|C_{usr}| \times Avg(W_{All})}$$

$|C_{usr}|$ 와  $|A|$ 는 이상 이벤트가 발생한 전체 고객 수와 이벤트  $A$ 가 발생한 고객 수를,  $W_A$ 와  $AVG(W_{All})$ 는 이벤트  $A$ 의 가중치와 전체 이벤트들의 평균 가중치를 의미한다.

전체 고객이 10명이고 이벤트  $A$ 와  $B$ 가 발생한 고객

수가 7명과 6명이라고 하자. 이 때 이벤트 가중치가 0.5와 0.9라면 전체 이벤트들의 평균 가중치는 0.7이 된다. 따라서 이벤트 A와 B에 대한 1차 가중 지지도는 50%와 77%가 되며 B가 발생한 고객 수는 A가 발생한 고객 수보다 적지만 중요도가 높아 상대적으로 높은 가중치가 적용되어 가중치를 반영한 1차 가중 지지도는 보다 높게 측정된다.

제안 방법에서 사용자는 관심도 및 이벤트 중요도에 따라 각각의 이벤트에 대하여 0과 1사이의 실수값을 가중치로 부여한다. 또한 제안 방법은 자주 발생하지는 않지만 높은 확률로 함께 발생하는 의미있는 최소 이벤트 쌍 탐사를 위하여 정의 2와 같은 2차 가중 지지도(2nd weight support)를 적용한다.

**정의 2. 2차 가중 지지도**

: 2차 가중 지지도는 사용자 관심도를 반영하여 의미 있는 최소 이벤트 쌍 사이에 존재하는 연관 규칙을 탐사하기 위한 최소 지지도로 이벤트 A가 발생한 고객 중 이벤트 B가 발생한 고객 수에 A와 B의 평균 가중치에 대하여 이벤트 A의 가중치를 반영한 비율을 의미한다. 이 때 A는 A와 B중 더 많은 고객에게서 발생한 이벤트로 설정한다.

$$\text{2차 가중 지지도} = \frac{|A \cup B| \times \text{Avg}(W_A, W_B)}{|A| \times W_A}$$

$|A \cup B|$ 와  $\text{AVG}(W_A, W_B)$ 는 이벤트 A와 B가 함께 발생한 고객 수와 A와 B의 가중치 평균을 의미한다.

전체 고객이 10명이고 이벤트 A가 발생한 고객 수와 이벤트 B가 발생한 고객 수가 모두 3명, A와 B가 함께 고객 수가 2명이라고 하자 이벤트 A, B, C의 가중치가 0.5, 0.7, 0.8이라면 전체 이벤트들의 평균 가중치는 0.67이며 이벤트 A와 B의 평균 가중치는 0.6이 된다. 따라서 이벤트 쌍 A와 B에 대한 2차 가중 지지도는 57%로 측정된다.

제안 방법은 1차 가중 지지도와 2차 가중 지지도를 사용하여 가중치를 적용한 빈발 이벤트와 의미있는 최소 이벤트 쌍을 추출하고 이들 사이에 존재하는 [표 1]과 같은 인터벌 관계에 대한 연관 규칙을 탐사한다. 그

리고 탐사된 연관 규칙에 대하여 정의 3과 같은 인터벌 이벤트 영향력을 측정함으로써 이벤트들 사이에 존재하는 인과 관계를 분석하여 향후 발생 가능한 이벤트 예측을 위한 정보로 활용한다.

**정의 3. 이벤트 관계 영향력**

: 이벤트 관계 영향력은 이벤트 A와 B 사이에 존재하는 인터벌 관계  $IR(A, B)$ 에 대하여 A가 B 발생에 어느 정도 영향을 미치는지에 발생 영향력 척도로 A에 대한 인터벌 관계  $IR(A, B)$ 의 발생 비율과 B에 대한  $IR(A, B)$ 의 발생 비율의 합을 의미한다.

$$\text{이벤트 관계 영향력} = \frac{|IR(A, B)|}{2} \times \left( \frac{1}{|A|} + \frac{1}{|B|} \right)$$

$|IR(A, B)|$ 는 인터벌 관계  $IR(A, B)$ 가 존재하는 고객 수를 의미한다.

[6][7][11][12]의 방법은 빈발 이벤트 A, B 사이에 존재하는 연관 규칙을 탐사할 수 있지만 이벤트들 사이에 인과 관계에 대한 영향력은 분석할 수 없다. 즉 지지도나 신뢰도 등의 기준은 고객에게서 발생한 이벤트 발생 횟수나 이벤트 A가 발생하는 경우 A와 함께 발생한 이벤트 B 발생 정도는 측정할 수 있다. 그러나 이러한 기준은 전체 고객에게서 발생한 B의 발생 횟수를 함께 고려하지 않으므로 A와 B가 어느 정도 영향을 주고 받았는지에 대한 영향력은 분석할 수 없다. 그러나 제안 방법은 정의 3과 같이 이벤트 A와 B가 발생한 각 고객에 대하여 인터벌 이벤트  $IR(A, B)$ 가 얼마나 발생하였는지를 측정함으로써 이벤트 발생에 대한 영향력을 분석할 수 있다.

설명을 위하여 전체 고객 10명 중 이벤트 A와 B가 발생한 고객이 5명과 8명이고 인터벌 관계  $before(A, B)$ 를 갖는 고객이 5명이라고 하자. 그리고 A와 B 사이에는 다른 인터벌 관계는 존재하지 않는다고 하자. 인터벌 관계  $before(A, B)$ 는 이벤트 A 종료 후 이벤트 B가 발생하였음을 의미한다. 그리고 이벤트 관계 영향력에서  $|IR(A, B)|/|A|$ 는 A가 B 발생에 얼마나 영향을 주었느냐를 의미하며  $|IR(A, B)|/|B|$ 는 B가 A에 의하여 얼마나 영향을 받았는가를 의미한다. 예에서 각 식의 계

산 결과는 1과 0.625이다. 이러한 수치는 이벤트 A가 발생하면 이벤트 B는 항상  $before(A,B)$  관계로 발생되지만 B는 A외에 0.375의 확률로 다른 이벤트에 의하여 발생됨을 의미한다. 그러므로 이벤트 A와 B 모두를 고려한 인터벌 관계  $before(A,B)$ 에 대한 이벤트 관계 영향력은 두 결과의 평균이  $0.8125((1+0,625)/2)$ 로 측정된다.

### 3. 연관 규칙 탐사 및 발생 영향력 알고리즘

제안 방법은 세 단계 과정을 통하여 다차원 스트림 데이터 환경에서 이벤트 가중치를 고려한 시간 관계를 탐사하고 그 영향력을 분석한다. 첫 번째 단계는 센서에서 수집된 스트림 데이터 중 이상 이벤트가 감지된 시점의 이벤트만을 추출하여 센서에서 서버로 전송한다. 두 번째 단계는 1차 가중 지지도와 2 가중 지지도를 적용하여 가중치를 고려한 빈발 이벤트와 의미있는 최소 이벤트 쌍을 추출하고 탐사된 이벤트들 사이에 존재하는 시간 요약 정보인 인터벌 이벤트를 구축한다. 그리고 [표 1]에 기초하여 이 들 사이에 존재하는 인터벌 관계를 탐사한다. 세 번째 단계는 탐사된 인터벌 관계 (연관 규칙)에 대한 이벤트 관계 영향력을 측정하여 이벤트 발생에 대한 인과 관계를 분석한다. 제안 방법에서 이상 이벤트 감지 시점의 이벤트만을 전송하는 과정은 알고리즘 1과 같다.

---

```

Input : 다차원 센서에 감지된 스트림 데이터
Output : 이상 이벤트가 감지된 시점의 스트림 데이터

For each 센서 s에 대하여
    If (시점 t에 s에서 감지한 스트림 데이터  $Str_s \in$ 
        객체의 정상 조건) then 시점 t에 감지한 각 센서의 스트림 데이터를 이벤트로 요약하여 서버로 전송
    
```

---

#### 알고리즘 1. 이상 이벤트 시점의 전송 알고리즘

제안 방법은 사용자 관심도 및 이벤트 중요도에 대한 가중치와 이벤트 발생 빈도를 고려하여 빈발 가중치 이벤트를 탐사한다. 그리고 탐사된 이벤트에 대하여 각

고객별로 이벤트 발생 시점과 종료 시점에 대한 요약 정보인 인터벌 이벤트를 생성하고 이 들 사이에 존재하는 인터벌 관계를 탐사한다. 인터벌 이벤트 구축 알고리즘과 빈발 가중치 이벤트에 대한 연관 규칙 탐사 과정은 각각 알고리즘 2와 알고리즘 3과 같다.

---

```

Input : 스트림 데이터(이벤트)
Output : 인터벌 이벤트 집합 IES

For each 센서 s에서 감지한 이벤트 E의 첫 이벤트  $E_1$ 과 마지막 이벤트  $E_n$ 의 시간 정보를 각각 인터벌 이벤트의 시작 시점  $us$ 와 종료 시점  $ve$ 로 설정하여 인터벌 이벤트 ( $E,[us,ve]$ )를 IES에 추가
    
```

---

#### 알고리즘 2. 인터벌 이벤트 구축 알고리즘

---

```

Input : 1차 가중 지지도  $SUP_{1st}$ , 관계 지지도  $SUP_{Rel}$ 
Output : 의미있는 인터벌 관계 집합 SIRS

For each 이벤트  $E_i$ 에 대하여 관심도에 따라 0과 1 사이의 가중치  $W_{E_i}$ 등록
    For each 고객  $C_{id}$ 에 대하여 이벤트 발생 빈도 기록
        For each  $E_i$ 에 대하여
            If ( $\frac{|E_i| \times W_{E_i}}{|C_{id}| \times Avg(W_{All})} \geq SUP_{1st}$ ) then 이벤트  $E_i$ 를 빈발 가중치 이벤트 집합 FWES에 추가
                else 이벤트  $E_i$ 를 최소 이벤트 집합 RES에 추가
            For each 고객  $C_{id}$ 에 대하여
                For each FWES의 이벤트에 대하여 알고리즘 2를 적용하여 인터벌 이벤트 및 인터벌 이벤트 관계 집합 IERS 구성
                For each IERS의 인터벌 관계  $IR_i$ 에 대하여
                    If ( $\frac{|IR_i|}{|C_{id}|} \geq SUP_{Rel}$ ) then  $IR_i$ 를 SIRS에 추가
                    
```

---

#### 알고리즘 3. 빈발 가중치 이벤트 연관 규칙 탐사

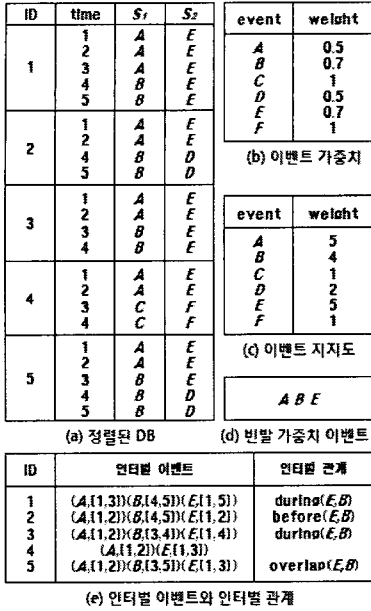


그림 3. 빈발 이벤트 인터벌 관계

제안 방법은 [그림 3(a)]와 같이 센서 S<sub>1</sub>과 S<sub>2</sub>에서 이상 이벤트가 감지된 시점의 이벤트만을 서버로 전송한다. 그리고 서버는 고객 아이디와 시간 순서로 이벤트를 정렬하고 가중치를 고려한 빈발 가중치 이벤트를 추출한다. 이벤트 가중치는 [그림 3(b)]와 같으며 [그림 3(a)]의 예에서 수집된 이벤트 지지도는 [그림 3(c)]와 같다. 1차 가중 지지도가 40%인 경우 추출된 빈발 가중치 이벤트는 [그림 3(d)]와 같으며 이들 사이에 존재하는 인터벌 이벤트와 인터벌 관계는 [그림 3(e)]와 같다. 또한 제안 방법은 [8][9]와 같은 기존의 스트림 데이터 마이닝에서 탐사할 수 없는 자주벌 뿔되지는 않지만 높은 확률로 함께 발생하는 이벤트에 대한 인터벌 관계를 탐사한다. 제안 방법은 2차 가중 지지도를 적용하여 자주 발생하지는 않지만 발생하는 경우 높은 확률로 함께 발생하는 희소 이벤트 쌍에 대한 연관 규칙을 알고리즘 4와 같이 탐사한다.

Input : 2차 가중 지지도 SUP<sub>2nd</sub>, 관계 지지도 SUP<sub>Rel</sub>

Output : 의미있는 인터벌 관계 집합 SIRS

For each 고객 C<sub>id</sub>에 대하여

For each 희소 이벤트 집합 RES에 포함된 이벤트 쌍(E<sub>i</sub>,E<sub>j</sub>)의 발생 빈도 기록

For each RES에 포함된 이벤트 쌍 (E<sub>i</sub>,E<sub>j</sub>)에 대하여

$$\text{If } \left( \frac{|E_i \cup E_j| \times \text{Avg}(W_{E_i}, W_{E_j})}{|E_i| \times W_{E_i}} \geq \text{SUP}_{2nd} \right) \text{ then}$$

(E<sub>i</sub>,E<sub>j</sub>)는 의미있는 희소 이벤트 집합 SRES에 추가

For each 고객 C<sub>id</sub>에 대하여

For each SRES의 이벤트에 대하여 알고리즘 2를 적용하여 인터벌 이벤트 및 인터벌 이벤트 관계 집합 IERS 구성

For each IERS의 인터벌 관계 IR<sub>i</sub>에 대하여

$$\text{If } \left( \frac{|IR_i|}{|E_i \cup E_j|} \geq \text{SUP}_{Rel} \right) \text{ then } IR_i \text{를 SIRS에 추가}$$

알고리즘 4. 의미있는 희소 이벤트 연관 규칙 탐사

[그림 3]에서 이벤트 C와 F는 높은 가중치를 갖지만 자주 발생되지 않으므로 [8][9]와 같은 데이터마이닝 기법들은 연관 규칙 탐사 과정에서 C와 F를 제외한다. 그러나 제안 방법은 2차 가중 지지도가 50%인 경우 이벤트 C와 F는 자주 발생되지 않더라도 C가 발생하는 경우 F가 함께 발생하는 확률이 정의된 2차 가중 지지도보다 크므로 이벤트 C와 F를 의미있는 희소 이벤트 쌍으로 추출하여 [그림 4]와 같이 C와 F 사이에 존재하는 인터벌 관계를 탐사한다.

ID	인터벌 이벤트	인터벌 관계
4	(C(3,4))(F(4,4))	during(C,F)

그림 4. 희소 이벤트 인터벌 관계

알고리즘 3과 알고리즘 4를 통하여 탐사된 연관 규칙은 이벤트 시간 순서에 대한 인터벌 관계 정보로 활용될 수 있지만 이벤트들 사이에 인과 관계가 존재하는 경우 이벤트 발생에 어느 정도 영향력을 주었는지는 알 수 없다. 따라서 제안 방법은 이벤트 관계 영향력을 측정하여 알고리즘 5와 같이 이벤트 발생에 대한 영향력을 분석한다.



Input : 의미있는 인터벌 관계 집합 *SIRS*

Output : 의미있는 인터벌 관계 영향력 집합 *SIRES*

For each 의미있는 인터벌 관계 집합 *SIRS*에 포함된 인터벌 관계  $IR(A,B)$ 에 대한 인터벌 관계 영향력  $IR_{eff} = \frac{|IR(A,B)|}{2} \times (\frac{1}{|A|} + \frac{1}{|B|})$  계산 후  $IR_{eff}$ 를 *SIRES*에 추가

**알고리즘 5. 인터벌 관계 영향력 분석 알고리즘**

[그림 3(e)]와 [그림 4]에서 탐사된 인터벌 관계는 이벤트 *E*와 *C*가 각각 이벤트 *B*와 *F*를 발생시키는 원인 이벤트임을 알 수 있다. 그리고 알고리즘 5를 적용하여 측정된 이벤트 발생 영향력은 [그림 5]와 같다.

인터벌 관계	인터벌 관계 영향력
during( <i>E,B</i> )	$(2/2) \times (1/5 + 1/4) = 0.45$
before( <i>E,B</i> )	$(1/2) \times (1/5 + 1/4) = 0.225$
overlap( <i>E,B</i> )	$(1/2) \times (1/5 + 1/4) = 0.225$
during( <i>C,F</i> )	$(1/2) \times (1/1 + 1/1) = 1$

그림 5. 인터벌 관계 영향력

[그림 5]의 인터벌 관계 during(*E,B*)에 의하여 이벤트 *B*는 이벤트 *E*가 발생하는 도중에 45%의 확률로 발생됨을 알 수 있으며 이벤트 *E*와 *B* 사이에 존재하는 세 개의 인터벌 관계 영향력을 통하여 *E*는 *B* 발생에 90%의 확률로 영향력을 미치는 원인 이벤트임을 알 수 있다. 또한 인터벌 관계 during(*C,F*)의 영향력 측정을 통하여 이벤트 *F*는 항상 이벤트 *C*로 인하여 발생됨을 알 수 있다.

**IV. 성능평가**

이 절에서는 제안 방법의 성능을 분석하며 실험은 윈도우 XP, 1GB RAM, P4 3.0MHz 시스템에서 C언어로 프로그램을 작성하여 수행한다. 실험 데이터는 3개의 센서에서 8개의 이벤트를 수집하는 다차원 스트림 시스템 환경에서 5,800명의 고객에 대한 32,200건의 트랜잭션을 적용한다. 실험에서 적용하는 이벤트 가중치는

[그림 6]과 같으며 이 중에서 이벤트 *B, C, E, F, H*는 이상 이벤트를 의미한다.

<i>S</i> <sub>1</sub>	<i>S</i> <sub>1</sub>	<i>S</i> <sub>1</sub>	weight
<i>A</i>	<i>D</i>	<i>G</i>	0.5
<i>B</i>	<i>E</i>	<i>H</i>	0.7
<i>C</i>	<i>F</i>		1

그림 6. 이벤트 가중치

실험 결과 전체 고객 5,800명중 5,462명의 고객에게서 이상 이벤트가 발생하였으며 빈발 가중치 이벤트와 의미있는 희소 이벤트 추출을 위하여 1차 가중 지지도와 2차 가중 지지도를 각각 30%와 40%로 적용하는 경우 가중치를 적용하지 않은 이벤트 발생 빈도에 대한 일반 지지도와 가중치를 적용한 1차 가중치 지지도는 [그림 7]과 같다.

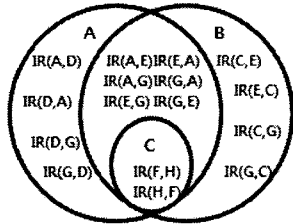
이벤트	가중치 지지도	지지도	이벤트	가중치 지지도	지지도
<i>A</i>	0.33	0.47	<i>E</i>	0.56	0.57
<i>B</i>	0.27	0.27	<i>F</i>	0.66	0.05
<i>C</i>	0.37	0.26	<i>G</i>	0.66	0.96
<i>D</i>	0.27	0.39	<i>H</i>	0.04	0.04

그림 7. 일반 지지도와 1차 가중치 지지도

실험 결과 [그림 7]과 같이 1차 가중 지지도 이상의 지지도를 갖는 이벤트 *A, C, E, G*는 빈발 가중치 이벤트로 추출된다. 특히 이벤트 *C*는 가중치를 고려하지 않은 경우에는 지지도가 1차 가중 지지도보다 작으므로 빈발 이벤트로 추출되지 않지만 *C*의 중요도에 따라 상대적으로 높은 가중치를 적용함으로써 빈발 가중치 이벤트로 추출된다. 그러나 이벤트 *D*는 가중치를 고려하지 않는 경우에는 지지도가 1차 가중 지지도 이상이므로 빈발 이벤트로 추출되지만 중요도가 낮아 상대적으로 낮은 가중치를 적용함으로써 빈발 가중치 이벤트로 추출되지 않는다.

또한 제안 방법은 자주 발생되지는 않지만 2차 가중치 지지도 이상의 확률로 함께 발생한 의미있는 희소 이벤트 쌍을 추출하며 실험 결과 이벤트 쌍 (*F, H*)가 2차 가중치 지지도를 만족하는 의미있는 희소 이벤트

쌍으로 탐사된다. 실험에서 최종적으로 탐사된 가중치를 고려하지 않는 경우에 탐사된 연관 규칙과 가중치를 고려한 경우 탐사된 연관 규칙들 간의 포함 관계는 [그림 8]과 같다.



A: 가중치를 고려하지 않은 빈발 이벤트 관계  
 B: 가중치를 고려한 빈발 이벤트 관계  
 C: 의미있는 희소 이벤트 관계

그림 8. 인터벌 관계

제안 방법은 이벤트 중요도에 따라 가중치를 부여함으로써 보다 중요도가 높은 이벤트에 대한 인터벌 관계를 탐사한다. [그림 8]에서 제안 방법은 상대적으로 중요도가 높은 이벤트 C에 높은 가중치를 부여함으로써 상대적으로 발생 빈도가 낮은 기존의 방법들에서 탐사할 수 없는 [그림 8]의 B 영역에 대한 연관 관계를 탐사할 수 있다. 또한 제안 방법은 의미있는 희소 이벤트 쌍에 존재하는 시간 정보를 탐사한다. [그림 8]에서 C 영역에 대한 연관 관계는 의미있는 희소 이벤트 쌍 사이에 존재하는 연관 관계로 적으로 발생 빈도만을 고려한다. 연관 규칙을 탐사하는 시간 데이터마이닝 기법에서는 탐사할 수 없는 연관 규칙을 제안 방법은 탐사할 수 있음을 보한다. 그리고 제안 방법은 탐사된 연관 규칙에 기초하나다적으로 발생에 대한 영향력 정도를 측정한다. 탐사된 연관 규칙 중 2 이 이상의 발생 지지도를 갖는 연관 규칙에 대한 이벤트 발생 영향력은 [그림 9]와 같다.

인터벌 관계	영향력	인터벌 관계	영향력
before(A,E)	0.24	overlap(A,G)	0.59
during(C,E)	0.20	overlap(G,E)	0.62
overlap(E,A)	0.39	overlap(F,H)	0.74

그림 9. 인터벌 관계 영향력

[그림 9]에 의하여 이벤트 E 발생에 대한 원인 이벤트에는 이벤트 A, C, G가 있으며 이중에서 가장 주요한 원인이 되는 이벤트는 인터벌 관계 영향력 측정을 통하여 이벤트 G임을 알 수 있다. 또한 인터벌 관계 overlap(F,H)에 대한 영향력 분석을 통하여 이벤트 H는 자주 발생되는 않지만 발생하는 경우 0.74의 확률로 이벤트 F와 함께 발생됨을 알 수 있다.

### V. 결론 및 향후 연구

이 논문에서는 다차원 스트림 데이터 환경에서 이벤트 가중치를 고려하여 시간 관계 규칙을 탐사하고 그 영향력을 분석하는 마이닝 기법을 제안하였다. 제안 방법은 스트림 데이터 환경의 특성을 고려하여 이상 이벤트가 감지되는 시점의 스트림 데이터만을 서버로 전송함으로써 센서와 서버와의 통신과 서버 처리 비용을 최소화한다. 그리고 중요도 및 사용자 관심도에 따라 각각의 이벤트에 가중치를 부여함으로써 중요도가 높은 이벤트에 대한 연관 규칙을 탐사한다. 또한 제안 방법은 다양한 지지도를 적용하여 이벤트 발생 빈도가 낮더라도 높은 확률로 함께 발생하는 의미있는 희소 이벤트들 사이에 존재하는 연관 규칙을 탐사한다. 그리고 탐사된 연관 규칙에 대한 영향력을 분석함으로써 이벤트 발생에 대한 주요 원인 이벤트를 분석할 수 있다.

향후 연구 방향으로 스트림 데이터들 사이에 존재하는 인과 관계를 보다 다양하게 분석하고 추론할 수 있는 평가 방법에 대하여 연구하고자 한다.

### 참고 문헌

[1] M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy, "Mining Data Streams: A Review," SIGMOD Record, Vol.34, No.2, pp.18-26, 2005.  
 [2] G. S. Manku and R. Motwani, "Approximate Frequency Counts over Data Streams," In Proc.

of Very Large Data Bases, pp.346-357, 2002.

[3] D. Kim, P. Park, H. Kim, and B. Hwang, "Mining Association Rules in Multidimensional Stream Data," Journal of Korea Information Processing Society, Vol.13-D, No.6, pp.765-774, 2006.

[4] D. Kim, P. 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] H. Li, S. Lee, and M. Shan, "Online Mining (Recently) Maximal Frequent Itemsets over Data Streams," In Proc. of Research Issues in Data Engineering: Stream Data Mining and Applications 2005, pp.11-18, 2005.

[6] D. Han, D. Kim, J. Kim, C. Na, and B. Hwang, "A Method for Mining Interval Event Association Rules from a Set of Events Having Time Property," Journal of Korea Information Processing Society, Vol.16-D, No. 2, pp.185-190, 2009.

[7] Y. Lee, J. Lee, D. Chai, B. Hwang, and K. Ryu, "Mining Temporal Interval Relational Rules from Temporal Data," The Journal of Systems and Software, Vol.82, No.1, pp.155-167, 2009.

[8] 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, 2004.

[9] G. Chen, X. Wu, and X. Zhu, "Mining Sequential Patterns Across Data Streams," Univ. of mont Computer Science Technical Report(CS-05-04), 2005.

[10] S. Laxman, P. S. Sastry, and K. Unnikrishnan, "Discovering Frequent Generalized Episodes where Events Persist for Different Durations,"

IEEE Transactions on Knowledge and Data Engineering, Vol.19, No.9, pp.1188-1201, 2007.

[11] H. Yun, D. Ha, B. Hwang, and K. Ryu, "Mining Association Rules on Significant Rare Data Using Relative Support," Journal of Systems and Software, Vol.67, No.3, pp.181-191, 2003.

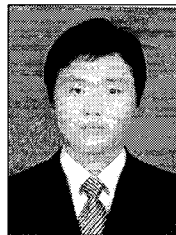
[12] R. J. Swargam, and M. J. Palakal, "The role of least frequent item sets in association discovery," In Proc. of International Conference on Digital Information Management 2007, Vol.1, pp.217-223, 2007.

[13] J. Allen, "Maintaining Knowledge about Temporal Intervals," Communications of the ACM, Vol.26, pp.832-843, 1983.

저 자 소 개

김 재 인(Jae-In Kim)

준회원



- 2008년 2월 : 전남대학교 전자컴퓨터공학과(공학사)
- 2008년 3월 ~ 현재 : 전남대학교 전자컴퓨터공학과 석사 과정

<관심분야> : 스트림데이터, USN, 디지털 콘텐츠, U-health, 데이터 마이닝

김 대 인(Dae-In Kim)

정회원



- 1998년 : 전남대학교 전산통계학과(이학석사)
- 2006년 : 전남대학교 전산통계학과(이학박사)
- 2004년 ~ 현재 : 전남대학교 전자컴퓨터공학부 시간강사

<관심분야> : 스트림 데이터, 데이터 마이닝, 이동 컴퓨팅, 디지털 콘텐츠

송 명 진(Myung-Jin Song)

준회원



- 2009년 : 전남대학교 전자컴퓨터공학과(공학사)
- 2009년 3월 ~ 현재 : 전남대학교 전자컴퓨터 공학부 석사 과정

<관심분야> : 스트림 데이터, 데이터 마이닝, 디지털 콘텐츠

한 대 영(Dae-Young Han)

준회원



- 2008년 2월 : 전남대학교 전자컴퓨터공학과(공학사)
- 2008년 3월 ~ 현재 : 전남대학교 전자컴퓨터 공학부 석사 과정

<관심분야> : 스트림 데이터, 데이터 마이닝, 멀티미디어, 디지털 콘텐츠

황 부 현(Bu-Hyun Hwang)

정회원



- 1978년 : 숭실대학교 전산학과(학사)
- 1980년 : 한국 과학기술원 전산학과(공학석사)
- 1994년 : 한국 과학기술원 전산학과(공학박사)

• 1980년 ~ 현재 : 전남대학교 전자컴퓨터공학부 교수

<관심분야> : 스트림 데이터, 데이터 마이닝, 이동 컴퓨팅, 분산 시스템, 디지털 콘텐츠