

# 데이터 스트림 시스템에서 이상 이벤트에 대한 연관 규칙 마이닝

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

## 요 약

최근에 데이터 스트림을 분석하여 잠재되어 있는 지식을 발견하기 위한 마이닝 방법에 대한 연구가 진행되고 있다. 그러나 대부분의 지지도 기반의 마이닝 방법들은 일정 주기 동안에 미리 정의된 지지도 이상의 발생 빈도를 갖는 이벤트만을 고려함으로써 발생 빈도에 비하여 중요도가 높은 이벤트를 간과하는 문제점을 가지고 있다. 본 논문에서는 이상 이벤트에 대한 연관 규칙을 탐사할 수 있는 SM-AF 방법을 제안한다. SM-AF 방법은 이상 이벤트가 감지된 윈도우만 고려하여 연관 정보를 탐사함으로써 자주 발생하지 않더라도 중요도가 높은 이벤트에 대한 연관 정보를 탐사할 수 있다. 또한 SM-AF 방법은 이상 이벤트에 대한 의미 있는 최소 항목 집합과 주기적인 이벤트 집합도 탐사한다. 그리고 다양한 실험을 통하여 SM-AF 방법이 기존의 연관 규칙 방법들에 비하여 우수함을 확인하였다.

키워드 : 데이터 스트림, 연관 규칙, 이상 이벤트, 의미 있는 최소 항목 집합, 지지도

## Mining Association Rule for the Abnormal Event in Data Stream Systems

Daein Kim<sup>†</sup> · Joon Park<sup>\*\*</sup> · Buhyun Hwang<sup>\*\*\*</sup>

### ABSTRACT

Recently mining techniques that analyze the data stream to discover potential information, have been widely studied. However, most of the researches based on the support are concerned with the frequent event, but ignore the infrequent event even if it is crucial. In this paper, we propose SM-AF method discovering association rules to an abnormal event. In considering the window that an abnormal event is sensed, SM-AF method can discover the association rules to the critical event, even if it is occurred infrequently. Also, SM-AF method can discover the significant rare itemsets associated with abnormal event and periodic event itemsets. Through analysis and experiments, we show that SM-AF method is superior to the previous methods of mining association rules.

Key Words : Data Stream, Association Rule, Abnormal Event, Significant Rare Itemsets, Support

### 1. 서 론

연관 규칙 탐사란 데이터베이스에 잠재되어 있는 지식을 발견하기 위한 마이닝 기법의 하나로 최근 데이터 스트림 시스템(DSMS : Data Stream Management System)에서 연관 규칙 탐사에 대한 연구가 활발하게 진행되고 있다[2, 4, 5, 9, 10]. 데이터 스트림 시스템의 응용은 ICU(Intensive Care Unit)와 같은 의료 분야, 생태계 및 환경 감시, 빌딩의 위험 감시, 이동 물체의 위치 추적 등 매우 다양하다[4, 8, 9, 13]. 그러나 데이터 스트림 시스템은 센서의 제한된 메모리와 배터리, 소형 프로세서, 그리고 제한된 통신 대역폭 등의 특성으로 인하여 센서를 통하여 수집되는 모든 데이터

스트림을 완벽하게 전송 및 분석하는 것은 불가능하다. 또한 센서를 통하여 수집되는 데이터 스트림은 연속적이며 그 크기가 무한하므로 데이터를 손실 없이 저장하고 처리하는 것은 많은 비용을 필요로 한다. 그러므로 데이터 스트림은 적용하는 응용 분야에서 관심을 갖고 있는 정보만을 추출하여 데이터의 특성에 맞게 변경하여 관리한다[6, 7, 9].

데이터 스트림 시스템에서 수집되는 데이터 스트림은 미리 정의된 값의 범위에 따라 기호화하여 저장하며, 각각의 기호는 가격 상승 및 하락과 같은 흐름(flow)과 같은 센서가 측정하는 객체에 대한 이벤트(event)를 의미한다[7, 9, 12]. 그리고 각각의 이벤트는 센서를 통하여 단위 시간 구간인 윈도우(window) 동안에 수집된 데이터 스트림이며 단일 윈도우 동안에 센서가 수집한 데이터 스트림은 하나의 트랜잭션에 포함된 데이터 항목으로 간주 할 수 있다[6, 7, 10, 12].

데이터 스트림 시스템의 응용의 하나인 ICU와 같은 의료 분야의 경우, 각각의 센서는 환자의 체온, 혈압, 맥박, 심장 박동 수 등 여러 종류의 데이터 스트림을 수집하며 각각의

\* 본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 지역문화산업연구센터(CRC) 지원사업의 연구결과로 수행되었음.

† 준 회 원 : 전남대학교 전자컴퓨터정보통신공학부 시간강사

\*\* 준 회 원 : 전남대학교 전자학과 박사과정

\*\*\* 종신회원 : 전남대학교 전자컴퓨터정보통신공학부 교수  
논문접수 : 2007년 1월 4일, 심사완료 : 2007년 4월 27일

데이터 스트림은 “체온과 맥박은 비례 관계이다.”, “출혈이 발생하는 경우 혈압은 내려가고 맥박은 빨라진다.”와 같이 이벤트 또는 환경의 변화에 연관성을 갖고 있다. 그러므로 이러한 데이터 스트림 간의 연관성을 분석함으로써 환자의 추후 발생 가능한 상태를 예측할 수 있다[9, 12].

데이터 스트림 시스템은 센서를 통하여 수집되는 단일 이벤트 관리 및 분석도 중요하지만 하나의 객체에 대한 여러 정보를 수집하는 다차원 센서에서 수집하는 이벤트 간의 연관 관계를 분석함으로써 향후 발생 가능한 이벤트를 예측하는 것도 매우 중요하며 이에 대한 연구가 진행되고 있다[7, 8, 9, 12]. 또한 이러한 응용에서 환자의 정상적인 체온 및 맥박에 대한 정보보다도 환자의 비정상적인 상태를 나타내는 체온 및 맥박의 상승과 같은 이벤트가 그 발생 빈도가 낮아도 중요성은 매우 높다[9]. 그러나 기존의 연구에서 적용하는 지지도 기반의 연관 규칙 탐사 기법들은 발생 빈도가 낮은 이벤트들은 연관 규칙 탐사 과정에서 제외하는 문제를 가지고 있다. 그러므로 본 연구에서는 데이터 스트림 시스템에서 객체의 비정상적인 상태를 나타내는 이상 이벤트(abnormal event) 간의 연관 규칙을 탐사하고 예측하기 위한 기법을 제안한다. 제안하는 방법에서 센서는 객체에 대한 여러 가지 이벤트를 수집하는 다차원 센서이며, 데이터 스트림은 윈도우 단위로 수집되어 처리된다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 데이터 스트림을 사용한 마이닝 기법에 관한 관련 연구를 기술하고 3장에서는 지지도 기반의 연관 규칙 탐사 방법의 문제점을 기술하며, 4장에서는 이상 이벤트에 대한 연관 규칙 탐사 알고리즘을 제안한다. 5장에서는 제안한 알고리즘의 성능을 분석하고, 끝으로 6장에서는 본 논문의 결론 및 향후 연구 방향을 기술한다.

## 2. 관련연구

데이터 스트림의 연관 규칙 탐사 방법은 기존의 데이터베이스 시스템에 비하여 여러 가지 제약사항을 가지고 있다[2, 5, 10]. 첫째, 마이닝 기법은 기본적으로 많은 데이터 스캔을 필요로 한다. 그러나 센서로부터 수집되는 데이터 스트림은 연속적이고 크기가 무한하므로 손실 없이 데이터베이스나 메인 메모리에 저장될 수 없으며 데이터가 수집되는 즉시 연관 규칙 탐사에 필요한 요약 정보를 추출하여야 한다. 둘째, 유용한 연관 규칙을 탐사하기 위하여 추출하는 후보 항목 집합(candidate itemsets)의 수가 무한하다. 데이터 스트림의 크기가 무한하고 연속적이므로 데이터의 조합으로 이루어지는 후보 항목 집합의 수는 무한히 증가한다[6, 7, 10]. 그리고 이러한 데이터 스트림의 특성으로 인하여 연관 규칙 탐사 방법은 많은 비용과 노력을 필요로 하며 데이터 스트림의 마이닝 방법으로 일정 시간 단위인 윈도우 동안에 수집된 데이터 스트림의 연관 정보를 분석하는 방법이 제안되었다[6, 7, 9, 10].

[10]에서는 윈도우 단위로 데이터 스트림간의 최대 빈발 항목 집합(maximal frequent itemsets)을 탐사하는 DSM-MFI(Data Stream Mining for Maximal Frequent Itemsets) 방

법을 제안하였다. DSM-MFI 방법은 트리 기반의 SFI(Summary Frequent Itemset)를 구축하여 사용자가 정의한 지지도  $X.CL$  이상의 발생 빈도를 갖는 데이터만을 추출하여 연관 규칙을 탐사한다. 또한 DSM-MFI 방법은 한 번의 스캔으로 요약 정보를 구축하고 발생 가능한 오류를 고려하여 0과 1 사이의 오류 임계값  $\zeta$ 를 고려한  $X.CL \times \zeta$  이상의 발생 빈도를 갖는 데이터를 포함하여 최대 빈발 항목 집합을 추출한다. 그러나 DSM-MFI 방법은 센서에서 수집되는 단일 데이터 스트림에 대한 연관 규칙만을 탐사하며 다차원 데이터 스트림 간의 연관 규칙 탐사는 고려하지 않는다.

[7]에서는 데이터 스트림의 연관 규칙을 탐사하는 MILE(Mining from multiple strEams) 방법을 제안하였다. MILE 방법은 윈도우 단위로 최소 지지도(minimum support) 이상의 발생 빈도를 갖는 토큰에 대한 트리 인덱스를 구축하여 다차원 데이터간의 연관 규칙을 탐사한다. 또한 MILE 방법은 [6]에서 제안한 PrefixSpan 방법의 단점인 반복적인 계산 과정을 줄이기 위하여 이벤트의 선행 관계로 구성된 해시 테이블을 유지함으로써 빠른 연관 규칙 탐사가 가능하다. 그러나 MILE 방법은 최소 지지도 이상의 빈도수를 갖는 데이터의 연관 규칙만 탐사하며 중요도가 높지만 발생 빈도가 낮은 이벤트에 대한 연관 규칙 탐사는 고려하지 않는다.

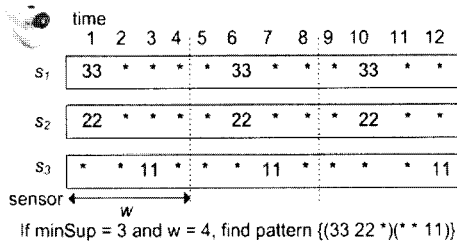
[9]에서는 윈도우 기반의 이벤트 발생 주기 탐사 대한 방법을 제안하였다. 제안 방법은 윈도우 단위로 이벤트의 발생 지지도를 계산하여 발생 주기를 탐사하고 최대 지지도를 만족하는 이벤트 시퀀스를 추출함으로써 추후에 발생 가능한 이벤트를 예측할 수 있다. 그러나 제안 방법은 이벤트 발생의 시간 간격만을 고려하며 이벤트의 발생 횟수와 센서로부터 동시에 수집되는 다차원 이벤트 간의 연관 관계는 고려하지 않는다.

## 3. 문제점

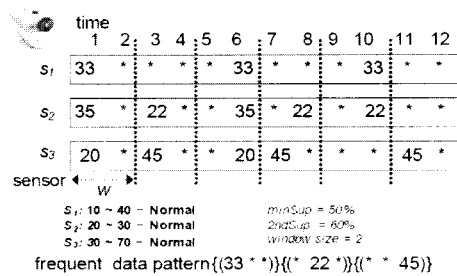
데이터 스트림 시스템에서 센서로부터 수집되는 데이터를 손실 없이 완전히 저장하는 것은 불가능하므로 일반적으로 데이터는 윈도우 단위로 기호화하여 저장된다. 그리고 기호화된 데이터 간의 연관 규칙을 탐사하여 변화를 분석함으로써 객체 및 이벤트 간의 연관 관계 탐사 및 향후 발생 가능한 이벤트 예측이 가능하게 된다[7, 9, 12].

[7]에서 제안한 MILE 방법은 데이터 스트림의 연관 규칙을 탐사하기 위하여 전체 윈도우 동안에 수집된 데이터 중에서 발생 빈도가 최소 지지도 이상인 데이터만을 추출하여 센서에서 수집되는 다차원 이벤트 간의 연관 규칙을 탐사한다.

(그림 1)은 세 종류의 데이터 스트림을 수집하는 다차원 데이터 스트림 시스템에서 MILE 방법을 사용하여 연관 규칙을 탐사하는 과정이다. (그림 1)에서 전체 시간은 12이고 단위 윈도우의 크기  $W$ 가 4이므로 전체 시간은 3개의 윈도우로 나누어진다. 그리고 같은 윈도우 동안에 수집된 각각의 데이터 스트림은 하나의 트랜잭션이 접근하는 데이터 집합으로 간주할 수 있다[7, 12]. 또한 (그림 1)에서 정의된 최소 지지도는  $3(3/3 = 100\%)$ 이므로 MILE 방법은 3개 이상의 윈도우에서 공통적으로 추출되는 데이터만을 빈발 항목 집



(그림 1) MILE 방법의 연관 규칙 탐사



(그림 2) 빈도만을 고려한 연관 규칙 탐사

함으로 추출한다. 그러므로 센서 S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>에서 최소 지지도 이상의 빈도를 갖는 빈발 항목 집합으로 각각 33, 22, 11이 추출된다. 또한 각 윈도우에서 33과 22는 센서 S<sub>1</sub>과 S<sub>2</sub>에서 동시에 발생하지만 센서 S<sub>3</sub>의 11은 그 이후에 발생하므로 {(33 22 \*) (\* \* 11)}과 같은 패턴 정보가 탐사된다.

일반적으로 같은 윈도우에서 함께 발생한 데이터들은 데이터 스트림 수와 센서에 감지된 이벤트로 구성된 내부 패턴으로 표현된다. 그러나 내부 패턴은 발생 순서만을 포함하므로 (그림 1)의 경우 패턴 (33 22 \*) 발생 후 패턴 (\* \* 11)은 발생하지만 패턴 (33 22 \*) 발생 후 즉시 패턴 (\* \* 11)이 발생하는 것을 의미하지는 않으며 두 패턴의 발생 시간에는 시간 간격이 존재할 수 있다[7]. 그러나 MILE 방법은 각각의 센서에 자주 감지되는 데이터만을 고려하여 연관 규칙을 탐사하므로 (그림 2)와 같이 데이터의 발생 빈도에 비하여 상대적으로 중요도가 높은 데이터의 연관 정보 탐사는 고려하지 않는다.

(그림 2)는 MILE 방법을 포함한 데이터의 발생 빈도만을 고려하는 연관 규칙 탐사 방법의 문제점을 보여준다. 그림 2에서 단위 윈도우 크기는 2이므로 전체 구간은 6개의 단위 윈도우로 나누어진다. 그리고 정의된 최소 지지도가 50%이므로 총 3개 이상(3/6 = 50%)의 단위 윈도우에서 발견되는 데이터가 빈발 항목 집합으로 추출된다. 그러므로 그림 2에서 MILE 방법은 센서 S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>에서 각각 33, 22, 45를 빈발 항목 집합으로 추출한다. 그러나 S<sub>2</sub>와 S<sub>3</sub>에서 수집된 35와 20은 발생 빈도는 최소 지지도보다 작지만 센서 S<sub>1</sub>에 빈발 항목 집합 33이 수집되는 경우 최소 지지도 이상의 확률로 함께 발생함을 알 수 있다.

일반적으로 ICU와 같은 의료 분야에서 환자의 정상적인 상태를 나타내는 체온, 맥박, 혈압 등의 정보는 센서를 통하여 자주 수집되어도 그 중요도는 높지 않다. 반면에 환자의 이상(abnormal) 상태를 나타내는 맥박, 혈압, 호흡 수, 체온

등의 정보는 환자의 이상 상태에 대한 바이탈 사인(vital sign)이므로 발생 빈도가 낮아도 그 중요도는 매우 높다[7, 8, 9]. 그러므로 발생 빈도가 낮더라도 중요도가 높은 이벤트 발생을 예측할 수 있는 연관 정보 탐사는 매우 중요하다.

본 연구에서는 센서로부터 수집된 다차원 데이터 간의 연관 규칙을 분석하여 이상 이벤트 예측을 위한 탐사 방법으로 SM-AF(data Stream Mining for Abnormal event Forecasting) 방법을 제안한다. SM-AF 방법은 객체의 위험 상황을 의미하는 비정상적인 데이터 스트림만을 추출하여 요약 정보를 생성한다. 이 때 이벤트 발생 간격에 대한 시간 정보를 포함하여 구축한다. 또한 자주 발생하지 않더라도 특정 이벤트와 빈번하게 함께 발생하는 의미 있는 희소 항목 집합(significant rare itemsets)을 추출함으로써 의사 결정에 필요한 다양한 연관 규칙을 탐사할 수 있다.

### 4. SM-AF 방법

#### 4.1 기본 개념

SM-AF 방법은 단순히 발생 빈도가 높은 데이터뿐만 아니라 중요도가 높은 데이터에 대한 연관 정보를 탐사한다. SM-AF 방법은 센서를 통하여 수집된 데이터 중에서 객체의 이상 상태를 의미하는 데이터가 감지된 윈도우만 추출하여 연관 규칙을 탐사한다. 그리고 하나의 단위 윈도우에 포함된 각각의 센서에서 수집된 데이터 스트림은 하나의 트랜잭션에 포함된 데이터 항목으로 간주된다.

SM-AF 방법에서 각각의 데이터 스트림은 시스템에 미리 정의된 값에 따라 객체의 정상적인 상태를 나타내는 NE(Normal Event) 상태의 집합과 비정상적인 상태를 나타내는 이상 이벤트 AE(Abnormal Event) 상태의 집합으로 나누어지며, AE 상태는 객체의 정상적인 상태의 범위에 포함되지 않지만 허용 가능 범위 ζ에 포함되는 TE(Tolerant Event) 상태와 ζ를 벗어나 객체의 위험 상태를 나타내는 RE(Risk Event) 상태로 나누어진다.

ICU와 같은 의료 분야에서 NE 상태는 환자의 체온 및 혈압이 정상인의 체온 및 혈압 범위에 속하는 경우이며 TE 상태는 환자의 체온이 정상인의 체온보다 높거나 낮지만 정의된 허용 범위 ζ에 포함되는 허용 가능한 범위의 데이터임을 의미한다. 그리고 RE 상태는 센서를 통하여 수집된 환자의 체온 및 혈압이 정상인에 비하여 매우 높거나 낮아서 위험을 가져올 수 있는 상태를 의미한다. 일반적으로 이벤트는 NE 상태에서 TE 상태로, 그리고 더욱 상태가 악화되면 TE 상태에서 RE 상태로, 호전되면 RE 상태에서 TE 상태로 전환된다. 또한 객체의 상태를 측정하는 센서에서 감지할 수 없는 외적인 환경 변화 및 발생으로 인하여 객체의 상태가 NE 상태에서 RE 상태, 또는 RE 상태에서 NE 상태가 될 수 있다. 즉 센서에서 수집되는 각각의 이벤트들은 센서에서 수집 가능한 내부의 변화뿐만 아니라 객체의 환경과 같은 외부 변화(external change)에 의하여 결정될 수 있다[9]. 그러나 외부 변화와 이벤트 발생에 대한 연관성은 기존의 데이터베이스 시스템에서 구축된 지식에 기반하며 SM-AF 방

법은 이벤트 간의 변화만을 고려한다.

SM-AF 방법은 센서에 TE 이벤트 또는 RE 이벤트에 해당되는 데이터가 감지되는 윈도우의 데이터 스트림만을 추출하여 빈발 항목 집합을 추출한다. 또한 SM-AF 방법은 [9]에서와 같이 이벤트 발생의 주기성을 고려하여 이상 이벤트 발생을 예측한다. 제안하는 SM-AF에서 연관 규칙 탐사에 적용하는 지지도는 다음과 같이 정의된다.

**[정의 1](지지도)**

- 1차 지지도(1st support) : 빈발 항목 집합 탐사에 적용되는 사용자가 정의한 최소 지지도로 전체 윈도우에서 이벤트 X가 발생한 단일 윈도우의 수를 의미한다.
- 2차 지지도(2nd support) : 의미 있는 최소 항목 집합 탐사에 적용되는 사용자가 정의한 지지도로 이벤트 X가 감지된 윈도우에서 동시에 이벤트 Y가 발생한 윈도우의 수를 의미한다.
- 3차 지지도(3rd support) : 이상 이벤트 발생의 주기 정보 탐사에 적용되는 사용자가 정의한 지지도로 센서 S에서 이상 이벤트 E 발생 후 시간 T안에 발생하는 이상 이벤트의 수를 의미한다.

$$1차\ 지지도 = \frac{|\sigma(X)|}{|\sigma(W)|}$$

$$2차\ 지지도 = \frac{|\sigma(X \cup Y)|}{|\sigma(X)|}$$

$$3차\ 지지도 = \frac{|T[\sigma(E)]|}{|\sigma(E)|}$$

$|\alpha W|$ 는 전체 윈도우 수를,  $|\alpha X|$ 는 데이터 X를 포함하는 윈도우 수를,  $|\alpha XUY|$ 는 데이터 X와 Y를 동시에 포함하는 윈도우 수를, 그리고  $|T[\alpha E]|$ 는 이상 이벤트 E 발생 후 시간 T안에 발생한 이상 이벤트 E'의 수를 의미한다. □

1차 지지도는 기존의 마이닝 방법에서 데이터간의 연관 규칙을 탐사하기 위하여 적용하는 최소 지지도로 전체 트랜잭션에서 데이터 x를 포함하는 트랜잭션 수가 1차 지지도보다 큰 경우 데이터 x는 빈발 항목 집합으로 추출된다. 그러나 SM-AF 방법은 중요도가 높은 이상 이벤트에 대한 연관 정보를 탐사하기 위하여 센서에서 수집된 전체 데이터가 아닌 이상 이벤트가 감지된 윈도우만을 추출하여 연관 규칙을 탐사하므로 1차 지지도는 이상 이벤트가 감지된 전체 윈도우에서 특정 이벤트가 감지된 윈도우의 수를 의미한다.

2차 지지도는 전체 윈도우에서 자주 발생하지 않더라도 특정 이벤트와 빈번하게 함께 발생하는 이벤트 탐사에 적용하는 지지도로 이벤트 X가 감지된 윈도우의 수에서 이벤트 X와 이벤트 Y가 함께 발생한 윈도우의 수가 2차 지지도보다 큰 경우  $\langle X, Y \rangle$ 는 의미 있는 최소 항목 집합으로 추출된다. 그리고  $\langle X, Y \rangle$ 는 X와 Y의 발생 빈도는 1차 지지도보다 작을 수 있지만 X 또는 Y가 발생하는 경우 두 이벤트는 빈번하게 함께 발생함을 의미한다.

3차 지지도는 주기적으로 발생하는 이상 이벤트를 탐사하기 위하여 적용하는 지지도로 전체 윈도우에서 발생한 이상

이벤트 E의 수에서 E가 발생한 후 시간 T안에 발생하는 이상 이벤트 E'의 수가 3차 지지도보다 큰 경우 주기적인 이벤트 집합으로 추출된다.

SM-AF 방법은 이상 이벤트가 감지된 윈도우만을 추출하여 저장한다. 그러므로 SM-AF 방법은 (그림 2)의 예에서 처음 윈도우 W1에서 센서 S2와 S3가 비정상적인 이벤트를 감지하므로 W1 시점에 센서에서 수집된 정보를 저장한다. 같은 원리로 이상 이벤트가 감지된 세 번째 윈도우 W3의 정보도 저장한다. 그러므로 (그림 2)의 예에서 SM-AF 방법은 센서로부터 전체 6개 윈도우 동안의 데이터를 수집하였지만 2개의 윈도우만을 고려하여 연관 정보를 분석함으로써 이상 이벤트를 포함하는 연관 규칙을 탐사한다. 그러므로 (그림 2)에서 SM-AF 방법은 이상 이벤트가 수집된 2개의 윈도우 동안에 각각의 센서에서 33, 35, 20이 수집되고 센서 S2와 S3에서 감지된 35와 20은 객체의 이상 상태를 나타내므로 MILE과 같은 기존의 방법에서 탐사할 수 없는 이상 이벤트가 포함된 연관 법칙을 탐사할 수 있다.

**4.2 단일 빈발 항목 탐사 알고리즘**

SM-AF 방법은 크게 세 단계의 과정을 통하여 이상 이벤트 예측을 위한 연관 규칙을 탐사한다. 첫 번째 단계는 센서를 통하여 수집된 데이터 중에서 이상 이벤트를 포함하는 윈도우만을 추출한다. 두 번째 단계는 정의된 1차 지지도와 2차 지지도 이상의 발생 횟수를 갖는 이벤트를 추출함으로써 이상 이벤트를 포함한 다차원 이벤트 간의 연관 규칙을 탐사하는 단계이다. 그리고 세 번째 단계는 3차 지지도를 사용하여 이상 이벤트 예측을 위한 이벤트 발생의 주기 정보를 탐사하는 단계이다. SM-AF 방법에서 센서를 통하여 수집된 데이터 중에서 이상 이벤트가 감지된 윈도우 추출 알고리즘은 다음과 같다.

```

Input : 센서를 통하여 수집된 다차원 데이터
Output : 이상 이벤트가 감지된 윈도우의 데이터

MakeSymbol( ); // 데이터 스트림의 기호화
For(w = 1:W |w|) // |W| 는 전체 윈도우 수
    For(s = 1:S |s|) // |S| 는 센서에서 수집하는 다차원 데이터 수
        If (s ≠ Normal_condition) then
            Add Data to Queue; // 이상 이벤트가 감지된 윈도우의 데이터를 큐에 추가
    
```

**알고리즘 1. SM-AF 방법에서 이상 이벤트가 감지된 윈도우 추출 알고리즘**

SM-AF 방법은 하나 이상의 센서에서 이상 이벤트가 감지되는 경우, 해당 윈도우 동안에 수집된 데이터만을 추출하여 각각의 센서별로 큐에 저장한다. 그리고 이상 이벤트가 감지된 윈도우의 데이터를 사용하여 1차 지지도 이상의 발생 빈도를 갖는 이벤트들을 단일 빈발 항목 집합으로 추출한다. 추출된 단일 빈발 항목 집합은 기존의 분산 데이터베이스 시스템에서 제안된 연관 규칙 알고리즘을 사용하여 다양한 연관 규칙을 탐사할 수 있다. SM-AF 방법에서 이상 이벤트가 감지된 윈도우만 추출하여 1차 지지도 이상의 발생 빈도를 갖는 단일 빈발 항목 집합 탐사 알고리즘은 다음과 같다.

```

Input : 추출된 이상 이벤트 발생 시점의 다차원 데이터(큐), 1차 지지도
Output : 단일 빈발 항목 집합

For(s = 1:|S|:s++)
  For(i = 1:|Ws|:i++) // |Ws| : 이상 이벤트가 발견된 윈도우 수
    count++: // 각 센서의 이상 이벤트 별로 동시에 발생한 이벤트의 빈도 기록
  For(s = 1:|S|:s++)
    For(d = 1:|D|:d++) // |D| : 센서에 감지된 이벤트 수
      If ( count[d] / |Ws| >= 1st support SUP1 ) then
        Add Data to frequent itemsets:
        // 1차 지지도 이상의 빈도를 갖는 이벤트를 단일 빈발 항목 집합으로 추출
    
```

알고리즘 2. SM-AF 방법에서 단일 빈발 항목 집합 탐사 알고리즘

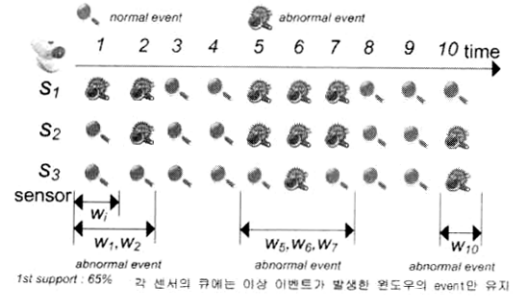
4.3 희소 항목 집합 탐사 알고리즘

SM-AF 방법은 이상 이벤트가 감지된 윈도우의 데이터만을 추출하여 연관 정보 탐사의 기초 자료인 단일 빈발 항목 집합을 추출한다. 그리고 (그림 3)과 같이 이상 이벤트가 감지된 윈도우 집합  $|W_s|$  동안에 미리 정의된 1차 지지도  $SUP_1$  이상의 확률로 동시에 발생하는 이벤트 간의 연관 정보를 탐사한다.

(그림 3)은 세 개의 센서가 전체 10개의 윈도우에서 데이터 스트림을 수집하는 경우 단일 빈발 항목 집합 탐사 과정을 보여준다. (그림 3)에서 데이터 스트림 시스템에 적용되는 윈도우 구간은 1이고 1차 지지도가 65%이므로 MILE 방법과 같은 기존의 데이터 마이닝 방법은 센서  $S_3$ 의  $NE_3$  이벤트만을 빈발 항목 집합(8/10=80%)으로 추출한다.

그러나 SM-AF 방법은 이상 이벤트가 감지된 윈도우만을 고려하므로 전체 10개의 윈도우 중 이상 이벤트가 감지된 윈도우  $W_1, W_2, W_5, W_6, W_7, W_{10}$ 만을 추출하여 연관 정보를 탐사한다. 그러므로 SM-AF 방법은 전체 10개의 윈도우 중에서 6개의 윈도우만을 고려하여 6개의 윈도우 중 4개 이상의 윈도우에서 감지된 이벤트(4/6=67%)를 1차 지지도 이상의 빈도수를 갖는 빈발 항목 집합으로 추출한다. 그러므로 SM-AF 방법은 센서  $S_1, S_2, S_3$ 에서 각각 이벤트  $AE_1, AE_2, NE_3$ 를 빈발 항목 집합으로 추출하며 [12]에서와 같은 연관 규칙 탐사 알고리즘을 적용하면 이상 이벤트가 감지된 전체 6개의 윈도우 중  $AE_1$ 과  $AE_2$ 가 4개의 윈도우 동안에 함께 발생함을 알 수 있다. 같은 원리로  $AE_1$ 과  $AE_3$ 가 4개의 윈도우에서 함께 감지되므로 최대 빈발 항목 집합으로 ( $AE_1, AE_2, *$ )와 ( $AE_1, *, NE_3$ )를 추출한다.

제안하는 SM-AF 방법은 전체 윈도우가 아닌 객체의 이상 상태를 의미하는 이벤트가 감지된 윈도우만 고려하여 연관 정보를 탐사하므로 발생 빈도가 낮더라도 이상 이벤트간의 연관 정보를 탐사할 수 있는 가능성이 커지게 된다. 또한 SM-AF 방법은 자주 발생하지 않더라도 특정 이벤트와 함께 빈번하게 발생하는 의미 있는 희소 항목 집합을 2차 지지도를 사용하여 추출한다. 2차 지지도는 데이터 간의 신뢰도(confidence)에 기반하며 발생 빈도에 상관없이 단일 윈도우 동안에 특정 이벤트와 동시에 발생하는 이벤트의 빈도를 측정하여 평가한다. SM-AF 방법에서 의미 있는 희소 항목 집합의 정의는 다음과 같다.



(그림 3) 이상 이벤트 추출 과정

[정의 2](의미 있는 희소 항목 집합)

의미 있는 희소 항목 집합 SRI는 각각의 센서에서 감지되는 이상 이벤트의 발생 빈도가 미리 정의된 1차 지지도보다 낮더라도 특정 센서에 이상 이벤트가 발생한 윈도우  $W$  동안에 정의된 2차 지지도 이상의 높은 비율로 동시에 발생하는 데이터를 의미한다.

$$SRI = (E_1, E_2, \dots, E_n)$$

SRI에서  $E_i$ 은 기준이 되는 이상 이벤트를 의미하며 이벤트  $E_i(2 \leq i \leq n)$ 는 2차 지지도 이상의 확률로 이벤트  $E_1$ 과 동시에 발생하는 이벤트를 의미한다. 그러나  $E_i(2 \leq i \leq n)$ 에 속하는 이벤트들이 2차 지지도 이상의 확률로 함께 발생함을 의미하지는 않는다. □

SM-AF 방법에서 2차 지지도 이상의 빈도를 갖는 이벤트에 대한 의미 있는 희소 항목 집합 탐사 알고리즘은 다음과 같다.

```

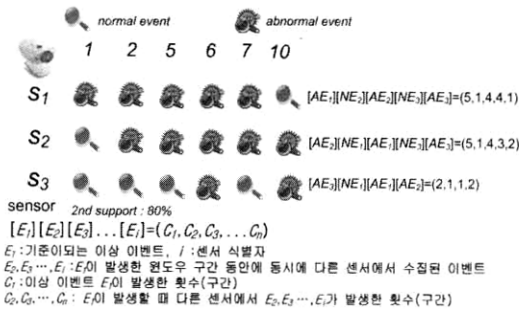
Input : 추출된 이상 이벤트 발생 시점의 다차원 데이터(큐), 2차 지지도
Output : 의미 있는 희소 항목 집합

For(a = 1:|A|:a++) // |A|: 감지된 이상 이벤트 수
  For(d = 1:|D|:d++) do // |D|: 이상 이벤트와 동시에 발생한 이벤트 수
    If ( count[d] / |E| >= 2nd support SUP2 ) then // |E| : 이상 이벤트 수
      Add Data to significant rare itemsets:
      // 2차 지지도 이상의 빈도를 갖는 이벤트 쌍을 희소 항목 집합으로 추출
    
```

알고리즘 3. SM-AF 방법에서 의미 있는 희소 항목 집합 탐사 알고리즘

SM-AF 방법은 (그림 3)의 예에서 기존의 방법에서 추출할 수 없는 최대 빈발 항목 집합 ( $AE_1, AE_2, *$ )과 ( $AE_1, *, NE_3$ )를 추출한다. 또한 SM-AF 방법은 2차 지지도를 사용하여 기존의 [7, 9, 10] 방법에서 추출할 수 없는 이상 이벤트에 대한 연관 정보를 추출한다. (그림 3)에서 센서  $S_1, S_2, S_3$ 에서 수집되는 이상 이벤트는  $AE_1, AE_2, AE_3$  이므로 전체 센서에서 감지된 이상 이벤트 수  $|A|$ 는 3이 되며 (그림 4)와 같은 정보를 유지한다.

(그림 4)는 SM-AF 방법에서 의미 있는 희소 항목 집합을 추출하기 위하여 (그림 3)의 예에서 유지하는 자료 구조를 보여준다. 자료 구조  $[E_1][E_{i+1}][E_{i-2}]...[E_n] = (C, C_{i+1}, C_{i-2},$



(그림 4) 희소 항목 추출을 위한 자료 구조

..., C<sub>n</sub>)에서 C<sub>i</sub>는 센서 i에서 수집된 이상 이벤트 E<sub>i</sub>의 발생 횟수를 의미하며 C<sub>j</sub>((i+1) ≤ j ≤ n)는 센서 i에서 이상 이벤트 E<sub>i</sub>가 발생할 때 같은 단위 윈도우 동안에 센서 j에서 동시에 발생한 이벤트 E<sub>j</sub>의 빈도를 의미한다.

(그림 4)에서 시스템에 정의된 2차 지지도가 80%이므로 센서 S<sub>1</sub>에 이상 이벤트 AE<sub>1</sub>이 발생하는 윈도우 동안에 센서 S<sub>2</sub>와 S<sub>3</sub>에 동시에 감지된 이벤트 중에서 2차 지지도 이상의 발생 빈도를 갖는 이벤트로 AE<sub>2</sub>(4/5=80%)와 NE<sub>3</sub>(4/5=80%)가 추출된다. 그리고 같은 원리로 각각의 센서에 대하여 적용하면 의미 있는 희소 항목 집합으로 (AE<sub>1</sub>, AE<sub>2</sub>, NE<sub>3</sub>), (AE<sub>2</sub>, AE<sub>1</sub>), 그리고 (AE<sub>3</sub>, AE<sub>2</sub>)가 추출된다. 특히 (AE<sub>3</sub>, AE<sub>2</sub>)에 포함된 AE<sub>3</sub>는 전체 윈도우 동안에는 자주 발생하지 않지만 센서 S<sub>3</sub>에 AE<sub>3</sub>가 감지되는 경우 센서 S<sub>2</sub>에 AE<sub>2</sub>가 빈번하게 높은 확률(2/2=100%)로 함께 발생함을 알 수 있다.

4.4 주기적인 이벤트 탐사 알고리즘

SM-AF 방법은 [9]에서와 같이 이벤트 발생의 주기성을 고려하여 하나의 센서에서 수집되는 이상 이벤트 간의 주기 정보를 탐사한다. SM-AF 방법은 특정 센서 i에서 감지된 이상 이벤트에 대한 발생 주기 |T|를 정의한다. 그리고 |T|안에 발생하는 이상 이벤트의 개수를 탐사하여 미리 정의된 3차 지지도 이상의 확률로 발생하는 이상 이벤트를 추출하여 주기적인 이상 이벤트 집합을 추출한다. SM-AF 방법에서 주기적인 이상 이벤트 탐사 알고리즘은 다음과 같다.

```

Input : 추출된 이상 이벤트 발생 시점의 다차원 데이터(큐), 이벤트 주기, 3차 지지도
Output : 주기적인 이상 이벤트 집합

For(s = 1;|S|;s++) do
  For(i = 1;|W|;i++)
    For(j= i +1;|W|;j++)
      If ( Equal(i, j) && and (i ∈ abnormal event) ) then
        // Equal(i, j) : 이벤트 i와 j가 같은 이상 이벤트
        If ( ( Interval(i, j) < Time interval |T| ) then
          // (Interval(i, j)) :이벤트 i와 j의 발생 간격
          count[a]++;
        For(a = 1;|A|;a++) do
          If ( count[a] / |E| = 3rd support SUP3 ) then
            Add Data to periodic abnormal itemsets;
          // 3차 지지도 이상의 빈도수를 갖는 토큰을 주기적인 이상 항목
          집합으로 추출
    
```

알고리즘 4. SM-AF 방법에서 주기적인 이상 이벤트 탐사 알고리즘

SM-AF 방법은 이벤트 발생의 주기성을 고려하여 하나의 센서에서 주기적으로 발생하는 이상 이벤트 집합을 탐사한다. SM-AF 방법은 센서에서 수집하는 이상 이벤트 AE의 발생 간격이 미리 정의된 시간 간격 |T|보다 작은 이벤트를 주기적인 이상 이벤트 집합으로 추출한다. 즉 AE 발생 후 재발생 간격이 |T|보다 작은 AE의 발생 횟수를 측정하여 그 비율이 미리 정의된 3차 지지도 SUP<sub>3</sub> 이상인 경우 센서 S에서 AE가 주기적으로 |T|안에 발생한다와 같은 정보를 탐사한다.

설명을 위하여 (그림 3)의 예에서 센서 S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>에 적용되는 시간 간격 |T|는 2이고 3차 지지도는 70%라고 정의하자. 전체 윈도우에서 S<sub>1</sub>은 이상 이벤트 AE를 5번 감지하였고 그 중 4번의 발생 시간 간격이 |T|보다 작으므로 주어진 3차 지지도(4/5=80%)를 만족한다. 그러므로 S<sub>1</sub>에서 수집하는 AE는 주기적으로 발생하는 이상 이벤트라는 정보를 추출할 수 있다.

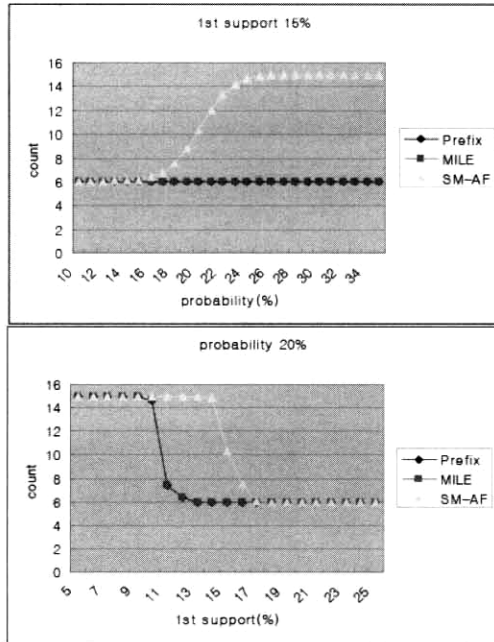
5. 성능평가

본 연구에서 제안한 SM-AF 방법은 센서로부터 연속적으로 수집되는 데이터 스트림 중에서 측정하는 객체의 이상 상태를 의미하는 이상 이벤트가 발생하는 윈도우만 선택하여 연관 정보 탐사를 위한 요약 정보를 구축함으로써 발생 빈도는 낮지만 중요도가 높은 이상 이벤트에 대한 연관 정보를 탐사한다.

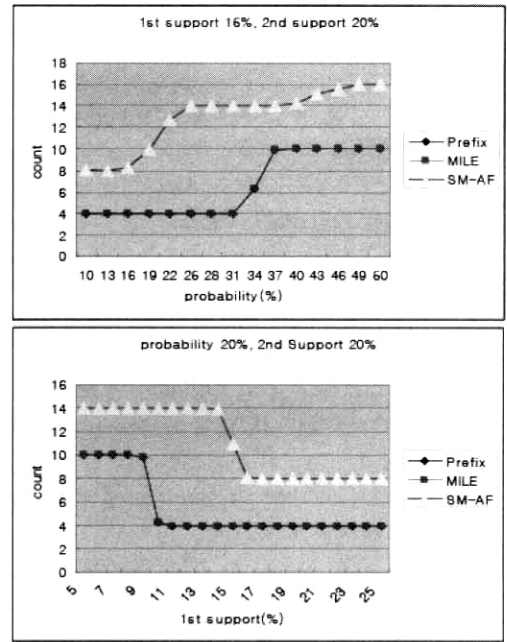
이 절에서는 SM-AF 알고리즘의 성능을 비교하며 모든 실험은 Microsoft Windows XP Professional 운영체제에서 2GB 메인 메모리, Pentium 4 3.0MHz 시스템에서 자바를 사용하여 작성하였다. 실험에서 사용하는 데이터 집합은 S, T, V, A의 네 가지 첨자로 표현된다. S는 데이터 스트림의 종류이며 T는 마이닝 작업을 수행하는 단위 시간, V는 각 센서에서 수집된 데이터에 대한 이벤트의 수, 그리고 A는 이상 이벤트의 수(V ≥ A)를 의미한다. 즉 S5T1000V5A3은 센서가 다섯 종류의 데이터 스트림을 단위 시간 1000 동안에 수집하며 수집된 데이터는 각 센서별로 각각 5개의 이벤트로 기호화되며 이 중에서 3개의 이벤트가 이상 이벤트임을 의미한다.

본 논문에서 실험은 세 종류의 데이터를 단위 시간 10000 동안 수집하여 5개의 이벤트로 기호화한 S3T10000V5A3 집합을 가지고 수행한다. 그리고 실험은 [6]에서 제안한 PrefixSpan 방법과 [7]에서 제안한 MILE 방법을 비교 대상으로 한다. 두 방법은 최소 지지도에 기반하여 데이터 스트림의 시퀀스 패턴 정보 분석 및 연관 정보를 탐사하므로 제안한 SM-AF 방법의 비교 대상으로 적합하다.

처음 실험은 S3T10000V5A3 집합을 사용하여 이상 이벤트의 발생 확률과 1차 지지도를 기준으로 추출되는 단일 빈발 항목 집합의 수를 측정한다. 그리고 실험의 정확도를 위하여 각각의 확률과 지지도에 대하여 100번씩 반복 테스트 한 후 그 평균 데이터를 실험 결과로 추출한다. 그리고 PrefixSpan



(그림 5) 단일 빈발 항목 집합의 수



(그림 6) 추출된 연관 정보 수

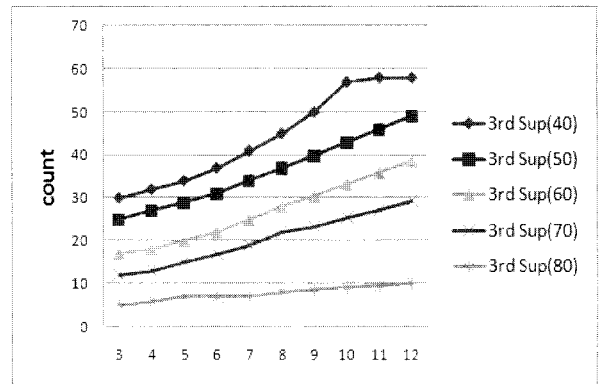
방법은 하나의 센서에 대한 연관 정보만 추출하므로 다차원 센서간의 연관 정보를 추출할 수 있도록 같은 데이터 스트림을 실험에서 적용하는 센서의 개수에 맞추어 반복 적용한다.

PrefixSpan 방법과 MILE 방법은 전체 윈도우 동안에 센서로부터 수집된 모든 데이터 스트림을 사용하여 빈발 항목 집합을 추출한다. 그리고 주어진 최소 지지도 이상의 발생 빈도를 갖지 못하는 데이터만을 추출하므로 발생 빈도가 낮지만 중요한 의미를 나타내는 이벤트들은 빈발 항목 집합 선택 과정에서 제외된다. 그러나 제안한 SM-AF 방법은 다차원 센서가 이상 이벤트를 감지하는 시점의 이벤트만 추출하여 연관 정보 탐사에 적용하므로 상대적으로 많은 수의 이상 이벤트들이 최소 지지도 이상의 발생 빈도를 갖게 되어 빈발 항목 집합으로 추출된다.

(그림 5)는 이상 이벤트가 발생하는 확률과 1차 지지도를 기준으로 추출되는 단일 빈발 항목 집합의 수에 대한 실험 결과를 보여준다. 실험 결과 PrefixSpan 방법과 MILE 방법은 같은 수의 단일 빈발 항목 집합을 추출하지만 SM-AF 방법은 이상 이벤트가 감지되는 윈도우만 추출하여 1차 지지도 이상의 발생 빈도를 갖는 항목을 추출하므로 상대적으로 이상 이벤트가 빈발 항목 집합으로 추출될 수 있는 가능성이 커짐을 알 수 있다.

PrefixSpan 방법과 MILE 방법은 최소 지지도 이상의 발생 빈도를 갖는 데이터 간의 연관 정보를 분석하므로 데이터의 중요성에 상관없이 일정한 확률 이상으로 발생하는 항목에 대한 연관 정보만 탐사한다. 그러나 SM-AF 방법은 최소 지지도 이하의 발생 빈도를 갖더라도 특정 이벤트와 2차 지지도 이상의 확률로 빈번하게 함께 발생하는 이벤트를 의미 있는 희소 항목 집합으로 추출하여 보다 다양한 연관 정보를 탐사할 수 있다.

실험 결과 SM-AF 방법은 (그림 6)과 같이 1차 지지도가



(그림 7) 추출된 주기적인 이벤트 수

작거나 이상 이벤트가 발생하는 확률이 클수록 보다 많은 연관 법칙을 추출한다. 그리고 이러한 결과는 SM-AF 방법이 단지 추출되는 연관 법칙의 수가 많다는 것을 의미하는 것이 아니라 기존의 발생 빈도 기반의 마이닝 방법에서는 추출할 수 없는 이상 이벤트간의 연관 정보를 추출하고 자주 발생하지는 않지만 특정 이벤트와 함께 빈번하게 발생하는 의미 있는 희소 항목 집합에 대한 연관 법칙을 추출함으로써 보다 중요도가 높고 의미 있는 연관 법칙을 추출할 수 있음을 보여준다.

또한 SM-AF 방법은 3차 지지도를 사용함으로써 주기적인 이벤트 집합을 추출함으로써 향후 발생 가능한 이벤트를 예측할 수 있다. (그림 7)은 3차 지지도를 사용하여 이벤트 발생 간격을 기준으로 추출되는 주기적인 이벤트 수를 보여준다. (그림 7)에서와 같이 SM-AF 방법은 이벤트의 발생 간격이 길수록, 그리고 3차 지지도가 작을수록 보다 많은 주기적인 이벤트에 대한 정보를 추출함을 알 수 있다.

## 6. 결 론

최근 데이터 스트림 시스템의 응용 분야는 환경 감시, 위치 추적, 의료 분야 등 매우 다양하며 이러한 응용 분야의 수는 빠르게 증가하고 있다. 특히 데이터 스트림은 크기가 무한하고 연속적인 특성을 가지고 있으며 센서가 수집하는 하나의 객체에 대한 여러 가지 정보들은 객체의 상태 변화에 따라 높은 연관성을 갖고 변화한다. 그리고 의료 분야의 경우 센서에서 수집되는 데이터는 측정하는 객체의 정상적인 상태를 의미하는 정보보다는 객체의 이상 현상에 대한 정보가 더욱 중요하다. 그러므로 센서에서 수집되는 이벤트의 중요도를 고려하고 객체의 상태 변화에 따라 향후 발생 가능한 이벤트를 예측할 수 있는 연관 법칙 탐사에 대한 연구가 필요하다. 본 연구에서는 데이터 스트림 시스템에서 이상 이벤트 간의 연관 정보를 탐사할 수 있는 SM-AF 방법을 제안하였다. SM-AF 방법은 다차원 센서에서 수집되는 데이터 중에서 이상 이벤트가 감지된 윈도우만 추출하여 이벤트간의 연관 정보를 탐사함으로써 보다 중요도가 높은 이벤트에 대한 연관 정보를 탐사할 수 있다. 그리고 SM-AF 방법은 2차 지지도를 사용하여 적용되는 1차 지지도 이하의 발생 빈도를 갖더라도 특정 이벤트와 빈번하게 발생하는 의미 있는 회소 항목 집합을 추출할 수 있다. 또한 SM-AF 방법은 3차 지지도를 적용하여 주기적으로 발생하는 이상 이벤트 정보를 탐사함으로써 시간 정보를 포함한 연관 규칙 탐사가 가능하며 향후 발생 가능한 이상 이벤트 예측도 가능하다. 그리고 성능 평가를 통하여 제안 방법이 기존의 연관 법칙 탐사 방법들에 비하여 보다 중요도가 높은 이상 이벤트에 대한 다양한 연관 규칙 정보를 탐사할 수 있음을 알 수 있었다. 앞으로의 연구 방향은 다차원 센서에서 수집되는 다중 이벤트 집합과 단일 이벤트 간의 연관 규칙을 탐사하고 다양한 요소를 고려한 다차원 이벤트 예측 시스템에 대하여 연구하고자 한다.

## 참 고 문 헌

[1] B. Brian, S. Babu, M. Datar, R. Motwani, and J. Widom, "Models and Issues in Data Stream Systems," In Proc. of PODS, March, 2002.

[2] D. Kifer, S. Ben-David, and J. Gehrke, "Detecting in Data Streams," In Proc. of VLDB, pp.180-191, 2004.

[3] M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy, "Mining Data Streams: A Review," SIGMOD Record, Vol.34, No.2, pp.18-26, June, 2005.

[4] L. Cohen, G. Avrahami-Bakish, M. Last, A. Kandel, and O. Kipersztok, "Real-time data mining of non-stationary data streams from sensor networks," Information fusion, pp.1-10, 2005.

[5] J. Chang, and W. Lee, "Finding frequent itemsets over data streams," Information and Software Technolgy. pp.606-618, 2006.

[6] J. Pei, J. Han, B. Mortazavi-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, Nov., 2004.

[7] G. Chen, X. Wu, and X. Zhu, "Mining Sequential Patterns Across Data Streams," Univ. of Vermont Computer Science

Technical Report(CS-05-04), March, 2005.

[8] H. Han, H. Ryoo, and H. Patrick, "An Infrastructure of Stream Data Mining, Fusion and Management for Monitored Patients," In Proc. of 19th IEEE International Symposium on CBMS 2006, pp.461-468, June, 2006.

[9] K. Kuramitsu, "Finding Periodic Outliers over a Monogenetic Event System," In Proc. of UDM05, pp.97-104, April 2005.

[10] H. Li, S. Lee, and M. Shan, "Online Mining (Recently) Maximal Frequent Itemsets over Data Streams," In Proc. of RIDE-SDMA'05, pp.11-18, April, 2005.

[11] H. Yun, D. Ha, B. Hwang, and K. Ryu, "Mining association rules on significant rare data using relative support," The Journal of Systems and Software, pp.181-191, 2003.

[12] 김대인, 박준, 김홍기, 황부현 "다차원 스트림 데이터의 연관 규칙 탐사 기법," 정보처리학회 논문지 D 제13-D권 제 6호, pp. 765-774, 2006.

[13] 한승철, 강현철, "XML 스트림 데이터에 대한 연속 질의 처리 시스템," 정보처리학회 논문지 D 제11-D권 제 7호, pp. 1375-1384, 2004.

### 김 대 인



e-mail : dikim@chonnam.ac.kr

1996년 동신대학교 전자계산학과 (학사)

1998년 전남대학교 전산통계학과 (이학석사)

2006년 전남대학교 전산통계학과 (이학박사)

2004년~현재 전남대학교

전자컴퓨터정보통신공학부 시간강사

관심분야: 스트림 데이터 마이닝, 이동 컴퓨팅, 분산 데이터베이스, 전자 상거래 등

### 박 준



e-mail : kingrion@empal.com

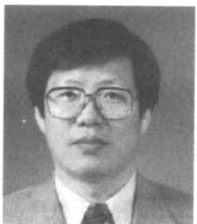
2001년 동신대학교 컴퓨터학과(학사)

2003년 전남대학교 전산학과(이학석사)

2004년~현재 전남대학교 전산학과 박사 과정

관심분야: 스트림 데이터 마이닝, FTTH, 이동 컴퓨팅, 객체 지향 등

### 황 부 현



e-mail : bhhwang@chonnam.ac.kr

1978년 숭실대학교 전산학과 (학사)

1980년 한국과학기술원 전산학과 (공학석사)

1994년 한국과학기술원 전산학과 (공학박사)

1980년~현재 전남대학교 전자컴퓨터정보통신공학부 교수  
관심분야: 스트림 데이터 마이닝, 이동 컴퓨팅, 분산 시스템, 분산 데이터베이스, 전자 상거래 등