

다차원 스트림 데이터의 연관 규칙 탐사 기법

김 대 인[†] · 박 준^{††} · 김 흥 기^{†††} · 황 부 현^{††††}

요 약

연관 규칙 탐사는 데이터베이스를 분석하여 잠재되어 있는 지식을 발견하기 위한 기법으로 스트림 데이터 시스템에서 연관 규칙 탐사에 대한 연구가 활발하게 진행되고 있다. 그러나 대부분의 연구들은 센서에서 수집되는 단일 스트림 데이터에 관한 것이며 다차원 스트림 데이터 간의 연관 정보는 간과하고 있다. 본 논문에서는 다차원 스트림 데이터 간의 연관 규칙을 탐사할 수 있는 AR-MS 방법을 제안한다. AR-MS 방법은 한 번의 데이터 스캔으로 연관 규칙 탐사에 필요한 요약 정보를 구축함으로써 스트림 데이터의 특성을 반영하며, 자주 발생하지는 않지만 특정 이벤트와 빈번하게 발생하는 의미 있는 희소 항목 집합에 대한 연관 규칙을 탐사할 수 있다. 또한 AR-MS 방법은 구축된 요약 정보를 사용하여 다차원 스트림 데이터 간의 최대 빈발 항목 집합에 대한 연관 규칙도 탐사한다. 그리고 다양한 실험을 통하여 제안하는 방법이 기존의 방법들에 비하여 우수함을 확인하였다.

키워드 : 스트림 데이터, 스트림 데이터 마이닝, 연관 규칙, 의미 있는 희소 항목 집합, 최대 빈발 항목 집합

Mining Association Rules in Multidimensional Stream Data

Daein Kim[†] · Joon Park^{††} · Hongki Kim^{†††} · Buhyun Hwang^{††††}

ABSTRACT

An association rule discovery, a technique to analyze the stored data in databases to discover potential information, has been a popular topic in stream data system. Most of the previous researches are concerned to single stream data. However, this approach may ignore in mining to multidimensional stream data. In this paper, we study the techniques discovering the association rules to multidimensional stream data. And we propose a AR-MS method reflecting the characteristics of stream data since make the summarization information by one data scan and discovering the association rules for significant rare data that appear infrequently in the database but are highly associated with specific event. Also, AR-MS method can discover the maximal frequent item of multidimensional stream data by using the summarization information. Through analysis and experiments, we show that AR-MS method is superior to other previous methods.

Key Words : Stream Data, Stream Data Mining, Association Rule, Significant Rare Itemsets, Maximal Frequent Itemsets

1. 서 론

최근 센서와 무선 통신 기술의 발달로 인하여 시간과 공간의 제약 없이 데이터를 수집하고 분석하여 잠재된 지식을 발견할 수 있는 스트림 데이터 시스템(Data Stream Management System, DSMS) 환경이 대두되었다[1]. 그리고 스트림 데이터 시스템의 응용 분야는 생태계 감시, 지진, 대기 및 해양 기후 감시 등의 분야와 선박과 차량 및 철재의 이동 경로 추적 등의 분야, 그리고 ICU(Intensive Care Unit)와 같은 의료 분야 등으로 분류할 수 있다. 이러한 응용에서 센서를 통하여 수집되는 데이터들은 환자의 체온 및 맥박과 같이 상호 연관성을 가지고 있으며 이러한 연관성을

발견 및 분석함으로써 발생할 수 있는 위험 및 이상 현상을 탐지 및 예측할 수 있다[5, 10]. 그러나 스트림 데이터 시스템은 제한된 메모리와 배터리, 소형 프로세서 및 낮은 통신 대역폭의 특성으로 인하여 수집되는 모든 데이터를 완벽하게 전송 및 분석하는 것은 불가능하며, 스트림 데이터는 연속적이며 인접 노드로의 전송을 필요로 하므로 데이터 저장 및 처리에 많은 비용을 필요로 한다. 그러므로 스트림 데이터의 특성을 반영한 데이터 처리 및 분석에 대한 연구가 필요하다[1, 10, 14, 15].

스트림 데이터는 연속적으로 수집되는 무한한 크기의 데이터이다. 그러므로 스트림 데이터의 처리 및 관리를 위해서는 기존의 데이터베이스 시스템에서의 데이터 분석 및 처리 방법을 스트림 데이터 시스템의 특성을 고려하여 수정하여 적용하여야 하며 그러한 방법의 하나로 일정 주기 단위로 데이터를 필터링 및 축약 방법을 통하여 요약하는 방법이 제안되었다[6, 8, 9, 11, 14]. 그리고 스트림 데이터를 처

[†] 준 회 원 : 전남대학교 전자컴퓨터정보통신공학부 시간강사

^{††} 준 회 원 : 전남대학교 전산학과 박사과정

^{†††} 정 회 원 : 동신대학교 컴퓨터학과 부교수

^{††††} 종신회원 : 전남대학교 전자컴퓨터정보통신공학부 교수

논문접수 : 2006년 8월 17일, 심사완료 : 2006년 9월 11일

리하는 시간 주기는 하나의 윈도우로 정의되며, 하나의 윈도우 시간 동안에 센서를 통하여 수집되는 데이터는 하나의 트랜잭션에 포함된 항목으로 간주 할 수 있다[8, 9, 12].

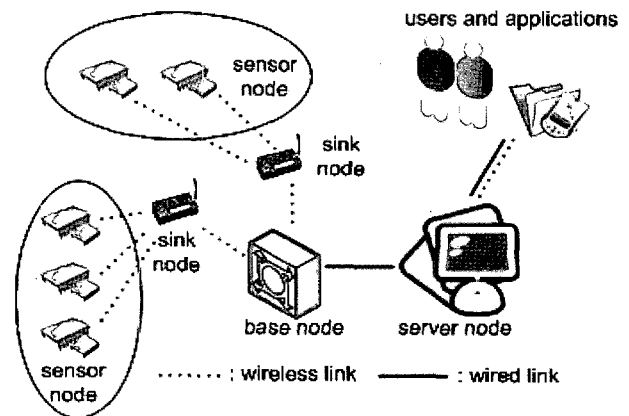
[4, 6, 11, 14]에서는 센서에서 수집되는 단일 스트림 데이터의 연관 정보 발견 및 요약에 초점을 맞추고 있다. 그러나 ICU와 같은 의료 분야에서의 센서는 환자의 체온, 혈압, 맥박, 심장 박동 수 등 여러 종류의 다차원 데이터를 수집한다[11]. 그리고 수집되는 각각의 데이터들은 “체온과 맥박은 비례 관계이다.”, “출혈이 발생하는 경우 혈압은 내려가고 맥박은 빨라진다.”와 같은 특정 이벤트에 대한 연관성을 갖고 있으며 이러한 연관성을 분석함으로써 환자의 현재 상태 및 추후 발생 가능한 상태를 예측할 수 있다. 따라서 단일 스트림 데이터에 대한 분석도 중요하지만 여러 종류의 다차원 스트림 데이터 사이의 연관 정보 탐사는 향후 발생할 수 있는 이상 현상에 대한 예측도 가능하며 이에 대한 연구가 필요하다.

본 연구에서는 스트림 데이터 시스템에서 다차원 스트림 데이터의 연관 규칙 탐사를 위한 마이닝 기법을 제안한다. 제안하는 방법은 ICU 분야와 같이 여러 종류의 다차원 스트림 데이터를 수집하여 윈도우 단위로 처리한다. 그리고 제안하는 방법은 한 번의 데이터 스캔 과정으로 연관 규칙 탐사에 필요한 요약 정보를 구축함으로써 스캔 횟수를 줄일 수 있으며 의미 있는 희소 항목 집합(significant rare itemsets)과 최대 빈발 항목 집합(maximal frequent itemsets)과 같은 연관 규칙을 발견함으로써 객체의 현재 상태 및 특정 이벤트의 발생 예측이 가능하다. 본 논문의 구성은 다음과 같다. 2장에서는 스트림 데이터 시스템 및 스트림 데이터 관리에 대한 관련 연구를 기술한다. 3장에서는 스트림 데이터 마이닝 기법에 대한 기존 연구의 문제점을 기술하고, 4장에서는 다차원 스트림 데이터의 연관 규칙 탐사 방법을 제시한다. 5장에서는 제안한 알고리즘의 성능을 분석하고, 끝으로 6장에서는 본 논문의 결론 및 향후 연구 방향을 기술한다.

2. 관련 연구

2.1 스트림 데이터 시스템

스트림 데이터 시스템은 기본적으로 센서 노드, 싱크 노드, 베이스 노드 그리고 서버 노드로 구성된다. 센서 노드는 제한된 배터리를 가지고 있으며 감지한 데이터를 인접한 센서 노드 및 싱크 노드로 전송한다. 그리고 센서 노드는 하나의 노드가 한 종류의 정보만을 측정하는 개별 센서 노드와 여러 종류의 다차원 정보를 수집할 수 있도록 여러 개의 센서를 포함하고 있는 센서 노드로 분류된다[2]. 싱크 노드는 일정 범위 안에 있는 센서 노드로부터 데이터의 수집, 관리 및 제어 기능을 하며 메인 메모리 기반의 데이터베이스를 가지고 있다. 일반적으로 센서 노드로부터 수집되는 데이터는 연속적이며 무한한 크기를 가지고 있는 데이터이므로 센서 노드에서 수집된 전체 데이터를 정확하고 완전하



(그림 1) 스트림 데이터 시스템

게 수집 및 저장하는 것은 불가능하며 싱크 노드는 수집된 데이터를 요약하여 처리한다[3]. 베이스 노드와 서버 노드는 싱크 노드로부터 요약된 데이터를 수신하여 데이터베이스에 저장하며 질의 처리를 통하여 수집된 정보를 사용자에게 제공한다. 이러한 요소들로 구성되는 스트림 데이터 시스템의 구조는 (그림 1)과 같다.

2.2 스트림 데이터 시스템에서의 데이터 관리

스트림 데이터는 연속적이며 크기가 무한하며, 스트림 데이터 시스템의 제한된 배터리, 무선 통신으로 인한 데이터 손실 등의 특성으로 인하여 잠재적인 에러를 가지고 있다. 그러므로 스트림 데이터 처리는 기존의 데이터베이스 시스템에서의 데이터 처리와는 다르게 축약 및 요약하여 처리된다. 또한 스트림 데이터 크기의 무한한 특성으로 인하여 스트림 데이터의 연관 규칙 정보를 탐사하기 위하여 마이닝 기법을 적용하는 경우 한 번의 데이터 스캔만으로 연관 규칙 정보를 분석하여야 한다[7, 12].

[6]에서는 센서 노드에서 전송되는 스트림 데이터의 크기를 줄이기 위하여 데이터를 집계(aggregation)하는 PGA (Potential Gains Adjustment) 알고리즘을 제안하였다. PGA 알고리즘은 스트림 데이터 정보를 트리 형태로 구축한 후 평균, 최대값, 최소값 등의 집계 연산을 통하여 요약함으로써 에러를 줄이고 전송되는 메시지 수를 줄일 수 있다. 그러나 PGA 방법은 스트림 데이터의 연관 규칙 탐사는 고려하지 않고 있다.

[4, 11]에서는 하나의 센서에서 일정 주기별로 발생하는 이벤트간의 연관 규칙을 탐사할 수 있는 윈도우 기반의 마이닝 방법을 제안하였다. [4]에서는 스트림 데이터 시스템에서 최소 지지도(minimum support)와 최대 허용 오차를 사용하여 빈발 항목 집합(frequent itemsets)을 계산하는 Lossy Counting 알고리즘을 제안하였다. Lossy Counting 알고리즘은 N 개의 데이터에 대하여 최소 지지도 s 를 만족하는 데이터를 검색하며, 이 때 허용 오차 ϵ 에 포함되는 $(s-\epsilon) \times N$ 개 이상의 발생 빈도를 갖는 데이터를 함께 추출하여 스트림 데이터의 연관 정보를 탐사한다. [11]에서는 하나

의 센서에서 수집되는 이벤트 중에서 정의된 지지도 이상의 이벤트들을 추출하여 주기적인 패턴 및 연관 규칙을 발견하며 데이터를 수집하는 시간 간격인 윈도우의 크기를 동적으로 구성함으로써 다양한 정보를 추출할 수 있는 마이닝 방법을 제안하였다. 그러나 [4, 11]에서 제안한 방법들은 단일 스트림 데이터에 대한 연관 규칙만을 탐사하며 환자의 체온, 혈압, 맥박과 같이 서로 연관된 다차원 스트림 데이터 간의 연관 규칙 탐사는 고려하지 않고 있다.

[8]에서는 다차원 스트림 데이터의 연관 규칙을 탐사할 수 있는 MILE(Mining from multiple strEams) 알고리즘을 제안하였다. MILE 알고리즘은 심장 박동 수, 주식 상승 가격과 같은 이벤트를 하나의 토큰으로 정의한 후 한 번의 스캔을 통하여 다차원 스트림 데이터 간의 연관 규칙을 탐사한다. 또한 MILE 방법은 [9]에서 제안한 PrefixSpan 방법에 비하여 보다 빠른 속도로 최소 지지도 이상의 지지도를 만족하는 이벤트 사이의 연관 정보를 추출하며 효율적인 메모리 사용을 위한 메모리 균형 분배 방법을 함께 제안하였다. 그러나 MILE 방법은 최소 지지도를 사용하여 빈번하게 발생하는 이벤트의 연관 규칙만을 탐사하며 자주 발생하지는 않지만 특정 이벤트와 빈번하게 발생하는 의미 있는 희소 항목 집합 탐사는 고려하지 않고 있다.

3. 문제점

3.1 스트림 데이터 시스템에서의 데이터 마이닝

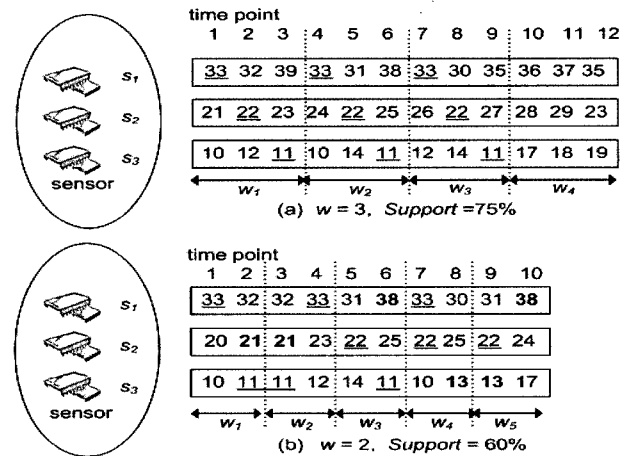
스트림 데이터의 연관 정보를 탐사하는 방법은 기존의 데이터베이스 시스템에 비하여 다음과 같은 제약사항을 가지고 있다[7, 9, 12]. 첫째, 센서를 통하여 연속적으로 수집되는 스트림 데이터는 많은 데이터 시퀀스를 생성한다. 데이터 시퀀스는 데이터의 조합으로 구성되므로 스트림 데이터와 같이 무한한 크기의 데이터에서 발생 가능한 시퀀스의 수는 무한히 증가할 수 있다. 둘째, 데이터 마이닝은 많은 데이터 스캔을 필요로 한다. 그러나 데이터의 크기가 무한한 스트림 데이터인 경우 데이터 전부를 저장하여 연관 규칙을 탐사하기 위하여 반복적으로 스캔하는 것은 현실적으로 불가능하다. 셋째, 스트림 데이터의 연관 규칙을 탐사하기 위하여 추출하는 후보 항목 집합(candidate itemsets)의 수는 무한하다. 후보 항목 집합 또한 센서에서 수집된 데이터의 조합으로 이루어지므로 추출되는 후보 항목 집합의 수는 무한히 증가할 수 있다. 또한 스트림 데이터의 연속성으로 인하여 발견하고자 하는 빈발 항목 집합의 길이에 비례하여 후보 항목 집합의 수는 무한히 증가할 수 있다. 이러한 스트림 데이터의 특성으로 인하여 스트림 데이터 시스템에서의 데이터 마이닝 기법은 많은 비용과 노력을 필요로 한다. 또한 센서로부터 수집되는 스트림 데이터 전부를 완전히 저장하는 것은 불가능하므로 일정 시간 단위인 윈도우 주기 동안에 센서를 통하여 수집된 데이터를 추상적 차원의 토큰으로 기호화 하며 토큰이 의미하는 이벤트 사이의 연관성을 분석한다[8, 9, 12].

3.2 기존 연구의 문제점

[8, 9]에서는 스트림 데이터의 연관 규칙을 탐사하는 방법으로 PrefixSpan 알고리즘과 MILE 알고리즘을 제안하였다. PrefixSpan 알고리즘과 MILE 알고리즘은 윈도우 주기 동안에 수집된 최소 지지도 이상의 윈도우를 갖는 토큰 정보를 트리로 구축하여 토큰이 의미하는 이벤트 간의 연관 규칙을 탐사한다. 그리고 MILE 방법은 PrefixSpan 방법에서 발생하는 반복적인 계산 과정을 줄이기 위하여 이벤트 사이의 선행 관계 정보를 해시 테이블에 유지함으로써 보다 빠른 연관 규칙 탐사가 가능하다. [8, 9]에서 제안한 알고리즘에서의 이벤트간의 연관 규칙 탐사 예는 (그림 2)와 같다.

(그림 2)는 ICU와 같은 의료 분야에서 세 종류의 센서를 통하여 환자의 혈압, 체온, 맥박과 같이 연관된 다차원 스트림 데이터를 수집하여 연관 규칙을 탐사하는 예이다. (그림 2(a))에서 윈도우 크기는 3, 그리고 최소 지지도는 75%로 정의되어 있다. 그리고 센서 S_1 , S_2 , 그리고 S_3 이 수집하는 데이터는 각각 $3x$, $2x$, 그리고 $1x(x=0..9)$ 형태의 토큰으로 표현되며 윈도우 크기가 3이므로 세 개의 단위 시간 동안에 센서를 통하여 수집된 데이터들은 하나의 트랜잭션에 포함된 데이터로 간주된다[8]. 그러므로 (그림 2(a))는 한 번의 데이터 스캔 과정을 통하여 트랜잭션 W_1 , W_2 , W_3 , 그리고 W_4 에 포함된 데이터의 연관 규칙을 탐사하는 것으로 간주할 수 있다. 센서 S_1 의 경우 토큰 $T(33)$ 만이 최소 지지도 75% 이상이며 센서 S_2 와 S_3 의 데이터는 각각 토큰 $T(22)$ 와 $T(11)$ 이 정의된 최소 지지도 이상의 지지도를 갖는다. 그러므로 [8, 9]의 알고리즘은 센서에서 수집된 토큰들 중에서 최소 지지도 이상의 지지도를 갖는 빈발 항목 집합인 토큰 $T(33)$, $T(22)$, 그리고 $T(11)$ 만을 고려하여 토큰이 의미하는 이벤트의 연관 규칙을 탐사한다. 또한 (그림 2(b))의 경우 주어진 최소 지지도 60%보다 큰 지지도를 갖는 토큰은 센서별로 각각 $T(33)$, $T(22)$, 그리고 $T(11)$ 이 된다.

[8, 9]의 방법은 (그림 2(a))의 예와 같이 최소 지지도 이상의 지지도를 갖는 빈발 항목 집합만을 사용하여 이벤트의 연관 규칙을 탐사한다. 그러나 (그림 2(b))의 경우 센서 S_1 과



(그림 2) 이벤트 사이의 연관 규칙 탐사 예

S_3 의 빈발 항목 집합 $T(33)$ 과 $T(11)$ 은 센서 S_2 에서 추출된 빈발 항목 집합인 $T(22)$ 보다는 주어진 최소 지지도 보다는 작지만 토큰 $T(21)$ 과 더욱 빈발하게 함께 발생함을 알 수 있다. 즉 토큰 $T(21)$ 이 의미하는 이벤트는 자주 발생하지는 않지만 센서 S_1 과 S_3 의 토큰에서 수집되는 특정 이벤트와 빈번하게 함께 발생함을 알 수 있다. 마찬가지로 센서 S_2 에서 추출된 토큰 $T(22)$ 는 최소 지지도 60% 이상을 만족하지는 못하지만 센서 S_1 과 S_3 의 토큰인 $T(38)$, $T(13)$ 과 빈번하게 함께 발생함을 알 수 있다. 그러나 [8, 9]에서 제안한 알고리즘은 주어진 최소 지지도 이상의 지지도를 갖는 토큰만을 추출하여 빈발 항목 집합을 구성하여 연관 규칙을 탐사하므로 (그림 2(b))의 예와 같은 희소 항목 집합은 탐사할 수 없는 문제가 있다.

본 연구에서는 센서로부터 수집되는 다차원 데이터 간의 연관 규칙 탐사 방법으로 AR-MS(Association Rule for Multidimensional Stream data) 방법을 제안한다. AR-MS 방법은 수집된 데이터를 토큰으로 기호화하여 이벤트의 연관 규칙을 탐사한다. 그리고 AR-MS 방법은 스트림 데이터 시스템의 특성을 고려하여 한 번의 데이터 스캔을 통하여 다차원 스트림 데이터의 연관 규칙을 탐사하며, 이 때 빈발 항목 집합뿐만 아니라 토큰 간의 신뢰도(confidence)를 고려하여 희소 항목 집합도 함께 추출함으로써 의사 결정에 필요한 보다 많은 유용한 정보를 탐사할 수 있다.

4. AR-MS 방법

4.1 의미 있는 희소 항목 집합 탐색

AR-MS 방법은 센서를 통하여 수집된 다차원 데이터의 연관 규칙을 탐사하기 위하여 스트림 데이터를 표 1과 같은 5개의 토큰으로 기호화한다.

의료 분야에서 센서를 통하여 일정 시간 주기 단위로 수집된 환자의 체온 및 혈압이 정상적인 사람의 체온 및 혈압의 범위에 포함되는 경우 스트림 데이터는 토큰 N 으로 기호화 되며, 토큰 H 와 L 은 각각 체온 및 혈압이 정상인보다

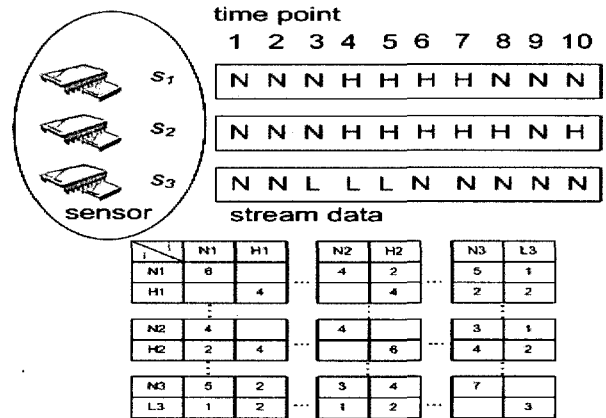
높거나, 또는 낮지만 미리 정의된 허용 범위 ζ_1 과 ζ_2 의 값임을 의미한다. 또한 토큰 O 와 U 는 센서를 통하여 수집된 환자의 혈압이 정상인의 혈압에 비하여 매우 높거나 낮아서 위험을 가져올 수 있는 경우를 의미한다. AR-MS 방법은 [8, 9]에서와 같이 스트림 데이터를 완전히 디스크에 저장하여 처리하는 것은 불가능하므로 센서로부터 데이터를 수신하는 즉시 한 번의 스캔 과정을 통하여 연관된 다차원 데이터의 연관 규칙 탐사에 필요한 요약 정보인 OS-FC(One Scan-Frequency Count)를 구축 한다. OS-FC는 다음과 같이 정의된다.

[정의 1] (OS-FC) OS-FC는 연관된 다차원 스트림 데이터의 연관 규칙을 탐사하기 위하여 각각의 토큰에 대하여 동일한 단위 시간에 다른 센서에서 수집된 토큰의 발생 횟수를 나타내는 $M \times M$ 크기의 2차원 배열이다. M 의 크기는 $T \times S$ 이며, 이 때 T 는 기호화되는 토큰의 종류이며 S 는 센서가 수집하는 다차원 데이터의 종류를 의미한다. OS-FC에서 $OS-FC[i][j](i=j)$ 의 값은 해당 센서에서 수집된 토큰의 발생 횟수를 의미하며 $OS-FC[i][j](i \neq j)$ 의 값은 해당 센서에 대하여 동시에 발생한 다른 센서에서 수집된 토큰의 발생 횟수를 의미한다.

(그림 3)은 센서가 혈압, 맥박, 체온과 같은 세 종류의 데이터를 수집하며 스트림 데이터가 표 1과 같이 5개의 토큰으로 기호화하는 스트림 데이터 시스템의 예이다. (그림 3)에서 센서는 서로 다른 세 종류의 데이터를 수집하고 수집된 데이터는 각각 5개의 토큰으로 기호화되므로 AR-MS 방법은 15×15 크기의 OS-FC를 구축한다. (그림 3)의 OS-FC에서 N, H, L 의 기호는 각각 표 1에서 정의된 토큰을 의미하며 다음 숫자는 센서의 식별자를 의미한다. 그리고 $OS-FC[N1][N1]$ 위치의 값은 정의 1에서 알 수 있듯이 센서 S_1 에서 토큰 N 의 발생 횟수를 의미하는 반면에 $OS-FC[N1][N2]$ 위치의 값은 센서 S_1 의 토큰 N 이 발생할 때 센서 S_2 에서 동시에 감지된 토큰 N 의 발생 횟수를 의미한다.

〈표 1〉 다차원 데이터의 변환 토큰

기호	약어	의미
N	Normal stream data	정의된 범위의 데이터
H	High stream data	정의된 범위의 데이터보다 크지만 정의된 허용 범위 ζ_1 에 포함되는 데이터
L	Low stream data	정의된 범위의 데이터보다 작지만 정의된 허용 범위 ζ_2 에 포함되는 데이터
O	Over stream data	정의된 허용 범위 ζ_1 의 값을 초과하는 데이터
U	Under stream data	정의된 허용 범위 ζ_2 의 값보다 작은 데이터



(그림 3) AR-MS 방법의 OS-FC 구조 예

제한하는 AR-MS 방법은 [8, 9]에서와 같이 구축된 OS-FC를 사용하여 최소 지지도 이상의 지지도를 갖는 빈발 항목 집합을 추출한다. 또한 AR-MS 방법은 [8, 9]에서 추출할 수 없는 의미 있는 희소 항목 집합을 추출하기 위하여 신뢰도와 2차 지지도를 사용한다. 스트림 시스템에서 의미 있는 희소 항목 집합은 다음과 같이 정의된다[13].

[정의 2](의미 있는 희소 항목 집합) 의미 있는 희소 항목 집합이란 센서를 통하여 수집된 데이터의 발생 빈도수가 미리 정의된 최소 지지도를 만족하지는 못하지만 특정 센서의 이벤트와 2차 지지도 이상의 높은 비율로 빈번하게 함께 나타나는 데이터를 의미한다.

2차 지지도는 [13]에서와 같이 의미 있는 희소 항목 집합의 탐색 과정에 적용하기 위하여 사용자가 정의한 지지도이

며 설명을 위하여 (그림 3)의 예에 최소 지지도와 2차 지지도를 각각 60%와 65%로 적용하자. (그림 3)에서 전체 시간 구간이 10이므로 하나의 센서에서 최소 6개 이상의 같은 토큰이 발생하여야 최소 지지도 60% 이상의 지지도를 갖는 빈발 항목 집합으로 추출된다. 그러므로 센서 S_1 은 토큰 N , 센서 S_2 은 토큰 H , 센서 S_3 은 토큰 N 이 6번 이상 발생하므로 각각 빈발 항목 집합으로 추출 된다. [8, 9]에서 제안한 방법들은 추출된 빈발 항목 집합만을 사용하여 이벤트의 연관 규칙을 탐사한다. 그러나 AR-MS 방법은 신뢰도와 2차 지지도를 적용하여 의미 있는 희소 항목 집합과 같이 자주 발생하지는 않지만 특정 데이터와 빈번하게 발생하는 정보에 대한 연관 규칙을 탐사한다.

AR-MS 방법은 (그림 3)의 경우 OS-FC를 사용하여 센서 S_1 의 빈발 항목 집합 N 은 6번 발생하였으며 동시에 센

```

Input : 토큰의 증감 범위, 센서를 통하여 수집된 다차원 데이터
Output : 배열 OS-FC, 단일 빈발 항목 집합, 의미 있는 희소 항목
        집합

Step 1. 센서를 통하여 수집된 데이터를 토큰 기호로 변환
Foreach window  $w = 1$  to  $W$  do //  $W$ 는 윈도우 수
  Foreach sensor  $s = 1$  to  $S$  do //  $S$ 는 센서에서 수집하는 다차원 데이터의 수
    MakeToken(Input Data[s]);
    // 단위 시간동안 각각의 센서로부터 데이터를 수집하여 정의된 토큰( $N, H, L, O, U$ )의
    // 범위에 따라 기호화하여 각 센서별로 Queue에 저장
  endfch
endfch

Step 2. 배열 OS-FC 구성 // OS-FC는  $M \times M$  크기의 2차원 배열
Foreach index  $i = 1$  to Queue.length do
  Foreach Queue  $q = 1$  to  $S$  do
    OS-FC[pivot_item][(q-1) * q.token]++; // pivot_item : 기준이 되는 센서의 토큰
  endfch
endfch

Step 3. 각 센서에 대한 단일 빈발 항목 집합 추출
Foreach index  $i = (q-1) * T + 1$  to  $q * T$  do //  $T$  : 토큰 인덱스( $N=1, H=2, L=3, O=4, U=5$ )
  If ( OS-FC[i][i] / Sum(OS-FC[i][j]) >=  $Sup_1$  ) then //  $Sup_1$  : 최소 지지도
    Add i.token to frequent itemsets;
    // i.token : OS-FC[i][j]위치의 값이 의미하는  $N_i$  형태의 토큰
  endif
endfch

Step 4. 의미 있는 희소 항목 집합 탐사
MoveFirst( $F$ ); //  $F$  : 단계 3에서 추출된 단일 빈발 항목 집합의 주소 정보
// 단일 빈발 항목 집합의 첫 번째 데이터 주소로 포인터 이동
While (  $F$  is not null ) do
   $f = F$ .Sensor_identifier; //  $f$  : 빈발 항목 집합  $F$ 의 센서 식별자 정보
  Foreach index  $i = 1$  to  $S * T$  do
    If ( (OS-FC[f][i]) !=  $f$  ) then
      If ( (OS-FC[f][i]) /  $F$ .frequency >=  $Sup_2$  ) then //  $Sup_2$  : 2차 지지도
        //  $F$ .frequency : 빈발 항목 집합의 빈도 수
        Add < $F$ .token.OS-FC[f][i].token> to significant rare itemsets;
        // 의미 있는 희소 항목 집합에 추가
      endif
    endif
  endfch
  MoveNext( $F$ ); // 다음 항목으로 이동
endw

```

(알고리즘 1) AR-MS 방법에서 의미 있는 희소 항목 집합 추출 알고리즘

서 S_2 에서는 토큰 N 과 H 가 각각 4번과 2번, 그리고 센서 S_3 에서는 토큰 N 과 L 이 각각 5번과 1번 발생하였음을 알 수 있다. 그러므로 센서 S_1 에서 토큰 N 이 발생하는 경우 센서 S_2 와 S_3 에서 토큰 N 의 신뢰도는 각각 66%와 83%이며 주어진 2차 지지도 65% 이상임을 알 수 있다. 특히 센서 S_2 에서 추출된 빈발 항목 집합은 토큰 H 이지만 센서 S_1 에서 토큰 N 이 발생하는 경우 S_2 에서는 H 보다 오히려 N 이 더욱 빈발하게 발생함을 알 수 있다. 마찬가지로 센서 S_2 의 빈발 항목 집합 H 와 센서 S_3 의 빈발 항목 집합 N 은 각각 센서 S_1 과 S_3 의 토큰 H 와 N , 그리고 센서 S_1 과 S_2 의 토큰 N 과 H 와 빈번하게 함께 발생함을 알 수 있다. 그러므로 최종적으로 $\langle N_1, N_2 \rangle$, $\langle N_1, N_3 \rangle$, $\langle H_2, H_1 \rangle$, $\langle H_2, N_3 \rangle$, $\langle N_3, N_1 \rangle$, $\langle N_3, H_2 \rangle$ 와 같은 희소 항목 집합을 탐사할 수 있다.

제안하는 AR-MS 방법은 4단계의 과정을 거쳐서 빈발 항목 집합 및 의미 있는 희소 항목 집합을 추출한다. 첫 번째 단계는 센서를 통하여 수집된 데이터를 정의된 데이터의 증감 범위에 따라 토큰 기호로 변환하는 단계이다. 각각의 토큰은 각 센서별로 큐에 저장된다. 두 번째 단계는 데이터의 연관 규칙을 탐사하기 위한 정보로 배열 OS-FC를 구축하는 단계이다. OS-FC는 전 단계에 구축된 각 센서에 대한 큐의 토큰들을 한 번의 스캔 과정을 통하여 구축된다. 세 번째 단계는 각 센서에 대한 단일 빈발 항목 집합을 추출한다. 추출된 단일 빈발 항목 집합은 각 센서별로 자주 발생하는 이벤트 분석 및 의미 있는 희소 항목 집합에 대한 연관 규칙 탐사를 위한 정보로 이용된다. 네 번째 단계는 세 번째 단계에서 구축된 단일 빈발 항목 집합에 대하여 의미 있는 희소 항목 집합에 대한 연관 규칙을 탐사하는 단계이다. AR-MS 방법의 단일 빈발 항목 집합 추출 및 의미 있는 희소 항목 집합 탐사 알고리즘은 (알고리즘 1)과 같다.

또한 AR-MS 방법은 의미 있는 희소 항목 집합을 탐사하기 위한 전 과정으로 각 센서별로 미리 정의된 최소 지지도를 만족하는 단일 빈발 항목 집합을 추출한다. 그러므로 [4, 11]에서 제안한 하나의 센서에서 발생하는 이벤트간의 연관 관계 정보도 각 센서별로 추출된 빈발 항목 집합 사이의 연관 관계 정보를 분석함으로써 추출할 수 있다.

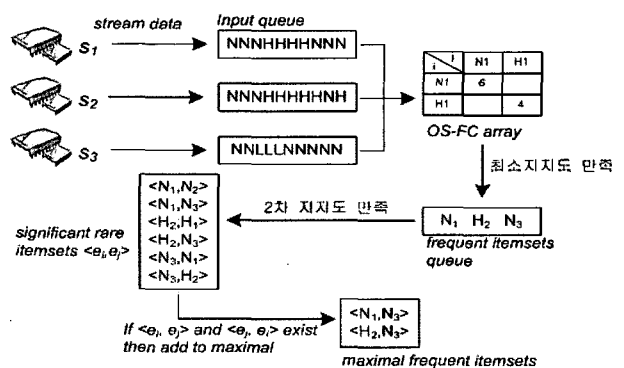
4.2 최대 빈발 항목 집합 탐사

제안하는 AR-MS 방법은 연관된 다차원 스트림 데이터의 연관 규칙을 탐사함으로써 특정 이벤트에 대하여 다른 센서에서 발생할 수 있는 이벤트를 예측할 수 있다. 그러나 이러한 발생 가능한 이벤트 예측뿐만 아니라 “체온과 맥박은 비례 관계이다.”와 같이 항상 일정하게 발생하는 이벤트 예측도 가능하다. 이러한 이벤트 간의 연관 관계를 발견하는 것은 기존의 전통적인 데이터베이스 시스템에서 최대 빈발 항목 집합을 찾는 것으로 간주할 수 있다. 최대 빈발 항목 집합이란 정의된 최소 지지도를 만족하는 빈발 항목 집합으로 구성된 최대 길이의 시퀀스를 찾는 과정으로 스트림 데이터 시스템에서의 최대 빈발 항목 집합 탐사는 최소 지지도 이상의 지지도를 갖는 이벤트 중에서 센서에서 동일한

시간 구간 동안에 수집되는 최대 수의 이벤트를 탐색하는 것을 의미한다.

[7]에서는 전통적인 데이터베이스 시스템에서 후보 항목 집합을 생성하여 최대 빈발 항목 집합을 추출하는 Apriori 알고리즘을 제안하였다. Apriori 알고리즘은 각 단계별로 빈발 항목 집합 생성 시 전 단계에서 추출된 빈발 항목 집합을 사용하여 다음 단계의 후보 항목 집합을 생성한다. 제안하는 AR-MS 방법은 OS-FC 정보를 사용하여 Apriori 방법과 유사한 방법으로 최대 빈발 항목 집합을 추출한다. AR-MS 방법은 알고리즘 1에서 추출된 의미 있는 희소 항목 집합을 사용하여 최대 빈발 항목 집합을 2-후보 항목 집합으로 바로 적용함으로써 Apriori 방법에서 발생하는 초기 추출 과정 단계를 줄일 수 있다. (그림 3)의 예에 AR-MS 방법을 사용하여 최대 빈발 항목 집합을 찾는 과정은 (그림 4)와 같다.

(그림 3)의 예에서 추출되는 빈발 항목 집합은 토큰 N_1, H_2, N_3 이다. 그러므로 AR-MS 방법은 N_1, H_2, N_3 에 대한 의미 있는 희소 항목 집합으로 (그림 4)와 같은 6쌍의 연관 규칙을 추출한다. 그리고 AR-MS 방법은 추출된 희소 항목 집합 $\{\langle e_i, e_j \rangle \dots \langle e_k, e_l \rangle\}$ 에 $\langle e_i, e_j \rangle$ 와 $\langle e_j, e_i \rangle$ 의 집합이 존재하는 이벤트 e_i 와 e_j 를 최대 빈발 항목 집합을 구하기 위한 2-빈발 항목 집합으로 적용한다. (그림 3)의 예에서 희소 항목 집합 중 $\langle N_1, N_3 \rangle$ 와 $\langle N_3, N_1 \rangle$, 그리고 $\langle H_2, N_3 \rangle$ 와 $\langle N_3, H_2 \rangle$ 가 존재하므로 $\langle N_1, N_3 \rangle$ 와 $\langle H_2, N_3 \rangle$ 는 최대 빈발 항목 집합을 구하기 위한 두 개의 항목으로 구성된 2-빈발 항목 집합이 된다. 그리고 그 이상의 길이를 갖는 빈발 항목 집합을 찾는 과정은 Apriori 방법과 같게 적용한다. 즉 AR-MS 방법은 (그림 4)에서 2-빈발 항목 집합 중 $\langle N_1, N_3 \rangle$ 와 $\langle H_2, N_3 \rangle$ 와 같이 공통된 토큰을 포함하는 집합을 찾는다. 그리고 공통된 항목 N_3 이 아닌 다른 항목들로 구성된 $\langle N_1, H_2 \rangle$ 가 2-빈발 항목 집합에 포함되어 있는지를 검색한다. 그러나 (그림 4)의 경우 $\langle N_1, H_2 \rangle$ 가 포함되어 있지 않으므로 최대 빈발 항목 집합은 현재 탐사된 빈발 항목 집합인 $\langle N_1, N_3 \rangle$ 와 $\langle H_2, N_3 \rangle$ 가 된다. 만약에 (그림 4)에서 $\langle N_1, H_2 \rangle$ 가 존재하면 최대 빈발 항목 집합은 2-빈발 항목 집합의 합집합으로 구성되는 $\langle N_1, H_2, N_3 \rangle$ 이 된다. Apriori 방



(그림 4) AR-MS 방법의 최대 빈발 항목 집합 탐색 예

```

Input : 의미 있는 희소 항목 집합
Output : 최대 빈발 항목 집합

Step 1. 의미 있는 희소 항목 집합에서 2-빈발 항목 집합 추출
Foreach rare itemsets  $r1 \in R$  do //  $R$ 은 의미 있는 희소 항목 집합
  Foreach rare itemsets  $r2 \in R$  do
    If (  $r1.item_1 = r2.item_2$  ) then
      Add  $r1$  to 2-frequent itemsets; // 선택된 2-후보 항목 집합을 break;
      // 2-빈발 항목 집합으로 선택
    endif
  endfch
endfch

Step 2. (k-1)-빈발 항목 집합을 사용하여 최대(k)-빈발 항목 집합 추출
// Step 2는 Apriori 방법을 적용,  $k \geq 3$ 
Insert into Candidate itemsets  $C_k$ 
  select  $c1.item_1, c1.item_2, \dots, c1.item_{k-1}, c2.item_{k-1}$ 
  from (k-1)-frequent itemsets  $c1$ , (k-1)-frequent itemsets  $c2$ 
  where  $r1.item_1 = r2.item_1, \dots, r1.item_{k-2} = r2.item_{k-2}, r1.item_{k-1} < r2.item_{k-1}$ ;
Foreach itemsets  $c \in C_k$  do
  Foreach (k-1)-frequent item subsets  $s$  of  $c$  do
    If (  $s \notin$  (k-1)-frequent itemsets ) then
      delete  $c$  from  $C_k$ ;
    endif
  endfch
endfch

```

(알고리즘 2) AR-MS 방법에서 최대 빈발 항목 집합 탐사 알고리즘

법은 한 번의 데이터베이스 스캔에 의하여 최대 빈발 항목 집합을 추출하므로 기존의 방법에 비하여 성능 향상은 있었지만 초기 단계에서 발생하는 후보 항목 집합의 수가 많은 문제가 있으며 이러한 문제를 해결하기 위한 여러 가지 방법들이 제안되었다. 제안한 AR-MS 방법 역시 추출된 의미 있는 희소 항목을 2-후보 항목 집합으로 사용하므로 최대 빈발 항목 집합을 탐사하기 위한 초기 추출 과정을 줄이고 초기 탐사 과정에서 발생하는 후보 항목 집합의 수를 줄일 수 있다.

제안하는 AR-MS 방법에서 최대 빈발 항목 집합을 찾는 과정은 크게 의미 있는 희소 항목 집합에서 2-빈발 항목 집합을 추출하는 단계와 Apriori 방법과 같이 2-빈발 항목 집합을 사용하여 길이 S 이하의 최대 빈발 항목 집합을 추출하는 두 단계로 나누어진다. S 는 수집되는 다차원 데이터의 종류를 의미하며 AR-MS 방법은 연관된 다차원 데이터 간의 최대 빈발 항목 집합을 탐색하므로 추출되는 최대 빈발 항목 집합의 길이는 S 보다 작거나 같다. AR-MS 방법에서의 최대 빈발 항목 집합 탐사 알고리즘은 (알고리즘 2)와 같다.

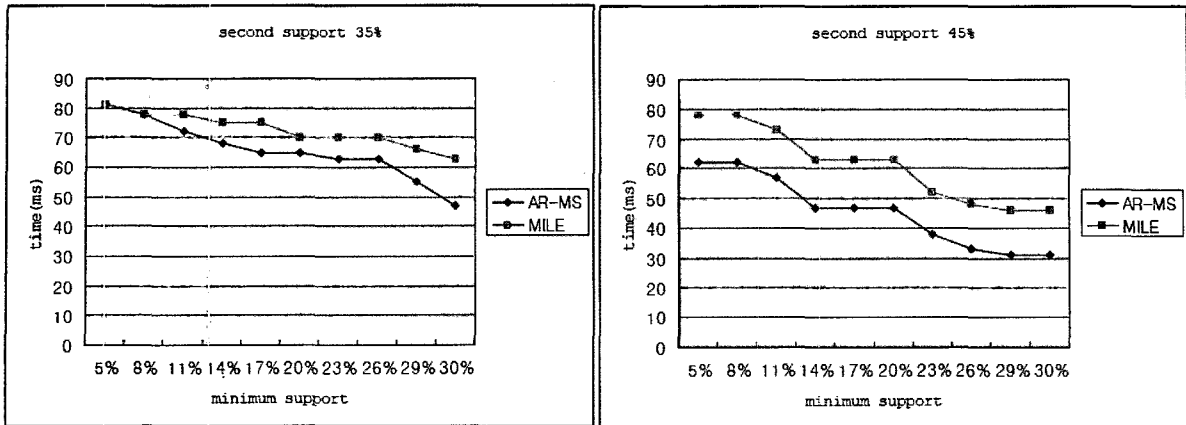
5. 성능 평가

본 연구에서 제안한 AR-MS 방법은 스트림 데이터 시스템에서 한 번의 데이터 스캔을 통하여 연관된 다차원 데이터의 의미 있는 희소 항목 집합과 최대 빈발 항목 집합을 탐사한다. 모든 실험은 Microsoft Windows XP Professional 운영체제에서 2GB 메인 메모리, Pentium 4 3.0MHz 시스템에서 자바를 사용하여 작성하였다. 그리고 [8]에서 제안한

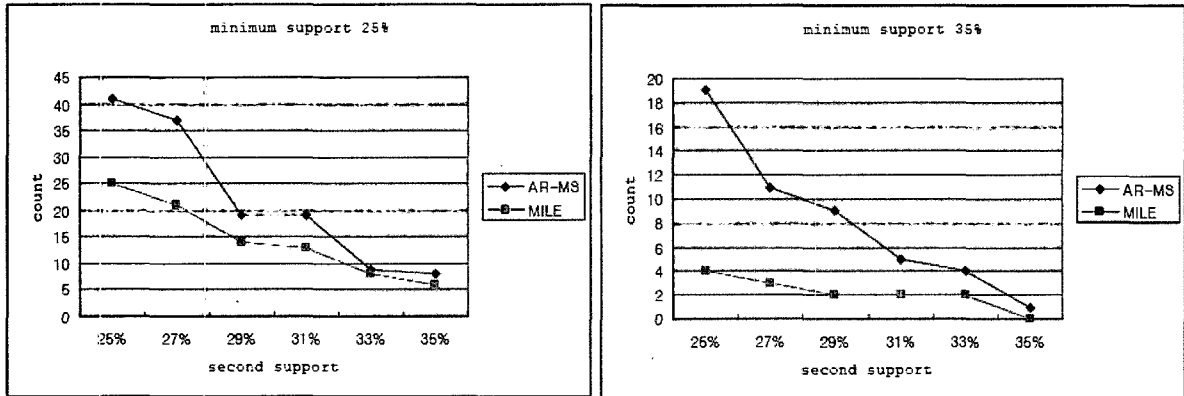
MILE 방법과 [7]에서 제안한 Apriori 방법을 비교 대상으로 한다.

[8]에서와 같이 실험에서 사용하는 데이터 집합은 s, t, v 의 세 가지 첨자로 표현된다. s 는 스트림 데이터의 종류이며 t 는 마이닝 작업을 수행하는 단위 시간, 그리고 v 는 각 센서에서 수집된 토큰 수를 의미한다. 즉 $s3t100v5$ 는 센서가 세 종류의 스트림 데이터를 단위 시간 100 동안에 수집하였으며 수집된 데이터는 5종류의 토큰으로 추상화되었음을 의미한다. 본 논문에서 실험은 $s3t1000v5$ 데이터 집합을 가지고 수행한다. 처음 실험은 주어진 데이터 집합을 사용하여 AR-MS 방법과 MILE 방법의 의미 있는 희소 항목 집합 탐사 시간을 최소 지지도를 기준으로 측정한다. 그리고 MILE 방법은 추출된 빈발 항목 집합만을 사용하여 데이터의 연관 정보를 분석하므로 의미 있는 희소 항목 집합을 탐사할 수 있도록 빈발 항목 집합과 동시에 발생하는 데이터 정보를 저장하는 별도의 자료 구조를 추가하여 실험하였다. (그림 5)의 실험 결과는 2차 지지도를 각각 35%와 45%로 적용한 경우에 의미 있는 희소 항목 집합의 추출 시간이다. MILE 방법은 의미 있는 희소 항목 집합을 탐사하기 위하여 추출된 단일 항목 집합의 모든 조합에 대하여 희소 항목 집합의 가능성을 살펴보아야 하는 반면에 제안한 AR-MS 방법은 구축된 OS-FC의 정보를 사용하여 보다 빠른 시간에 정보를 추출할 수 있음을 알 수 있다.

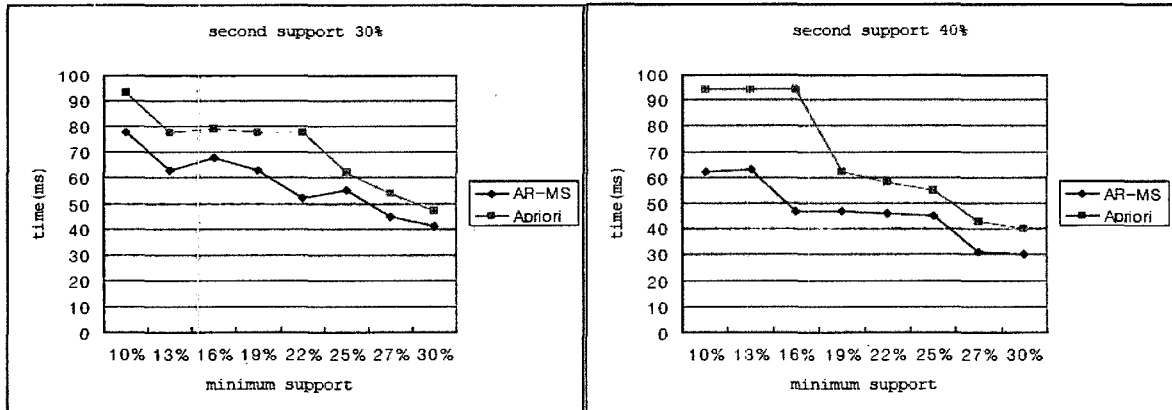
그리고 (그림 6)은 $s3t1000v5$ 데이터 집합을 사용하여 AR-MS 방법과 MILE 방법에서 추출되는 의미 있는 희소 항목 집합의 수를 2차 지지도를 기준으로 보여 준다. (그림 6)에서 최소 지지도가 25%와 30%인 모든 경우에 MILE 방법



(그림 5) AR-MS 방법과 MILE 방법의 성능 비교(회소 항목 탐사 시간)



(그림 6) AR-MS 방법과 MILE 방법의 성능 비교(회소 항목 집합 수)



(그림 7) AR-MS 방법과 Apriori 방법의 성능 비교(2-빈발 항목 집합 탐사 시간)

은 센서에서 수집된 스트림 데이터 중 주어진 최소 지지도 이상의 지지도를 갖는 항목에 대하여 의미 있는 회소 항목 집합을 추출하는 반면에 AR-MS 방법은 센서에 특정 데이터가 수집되는 경우 동시에 다른 센서에서 수집되는 데이터의 수를 요약 정보로 유지함으로써 보다 많은 의미 있는 회소 항목 집합을 추출한다. 그리고 (그림 6)에서와 같이 회소 항목 집합에 대한 2차 지지도가 커질수록 각각의 데이터들이 2차 지지도보다 더 큰 지지도를 갖는 확률이 낮아짐으로

탐사되는 회소 항목 집합의 수는 줄어든다. 또한 AR-MS 방법은 알고리즘 1에 의하여 발견된 의미 있는 회소 항목 집합을 사용하여 2-빈발 항목 집합을 구함으로써 Apriori 방법의 초기에 발생하는 후보 항목 집합의 수가 많아지는 문제를 해결할 수 있다. (그림 7)은 *s3t1000v5* 데이터 집합을 사용하여 2-빈발 항목 집합까지의 추출 시간을 2차 지지도를 각각 30%와 40%로 최소 지지도를 기준으로 측정 한 실험 결과이다. (그림 7)은 AR-MS 방법이 Apriori 방법에

비하여 최대 빈발 항목을 찾기 위한 2-빈발 항목 집합을 빠르게 생성함을 보여준다. 그리고 최소 지지도가 낮아질수록 추출되는 단일 빈발 항목의 수가 작아짐으로 AR-MS 방법과 Apriori 방법의 탐사 시간이 비슷해짐을 알 수 있다.

6. 결 론

최근 스트림 시스템의 응용 분야는 생태계 감시, 환경 감시, 차량 추적, 의료 분야 등 매우 다양하며 스트림 시스템에서의 데이터 관리 및 처리에 대한 연구가 활발히 진행되고 있다. 특히 스트림 데이터는 기존의 데이터와는 달리 연속적이고 무한한 크기를 갖으며 이러한 특성을 고려한 데이터 관리 방법에 대한 연구가 필요하다. 본 연구에서는 스트림 시스템에서 연관된 다차원 스트림 데이터의 연관 규칙을 탐사할 수 있는 AR-MS 방법을 제안하였다. AR-MS 방법은 스트림 데이터의 특성을 고려하여 한 번의 데이터 스캔으로 연관 규칙 탐사에 필요한 요약 정보를 구축한다. 그리고 AR-MS 방법은 자주 발생하는 이벤트에 정보뿐만 아니라 자주 발생하지는 않지만 특정 이벤트와 빈번하게 발생하는 의미 있는 희소 항목 집합에 대한 연관 규칙도 탐사할 수 있다. 또한 AR-MS 방법은 최대 빈발 항목 집합에 대한 연관 규칙 탐사에 필요한 후보 항목 집합을 의미 있는 희소 항목 집합을 사용하여 적용함으로써 초기 과정에서 많은 후보 항목 집합이 발생하는 기존의 문제점을 해결할 수 있다. 그리고 성능 평가를 통하여 제안 방법이 기존의 방법들에 비하여 보다 빠른 시간에 연관 규칙 탐사가 가능함을 알 수 있었다. 앞으로의 연구 방향은 시간 정보를 포함한 다차원 이벤트 간의 연관 규칙, 그리고 하나의 이벤트와 다차원 이벤트 간의 연관 규칙을 탐사하는 방법에 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] M. J. Franklin, S. R. Jeffery, S. Krishnamurthy, F. Reiss, S. Rizvi, E. Wu, O. Cooper, A. Edakkunni, and W. Hong, "Design Consideration for High Fan-in Systems: The HiFi Approach," In Proc. of CIDR, pp.290-304, Jan., 2005.
- [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] G. S. Manku and R. Motwani, "Approximate Frequency Counts over Data Streams," In Proc. of VLDB, pp.346-357, 2002.
- [5] R. C. Oliver, K. Smettem, M. Kran, and K. Mayer, "Fielding Testing a Wireless Sensor Network for Reactive," In Proc. of ISSNIP, pp.7-12, Dec., 2004.
- [6] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Hierarchical In-Network Data Aggregation with Quality Guarantees," LNCS(EDBT 2004), pp.658-675, March, 2004.
- [7] R. Agrawal and R. Srikant, "Fast Algorithms for mining association rule," In Proc. of VLDB, Sep., 1994.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] K. Kuramitsu, "Finding Periodic Outliers over a Monogenetic Event System," In Proc. of UDM05, pp.97-104, April, 2005.
- [12] 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.
- [13] 하단심, 황부현, "상대 지지도를 이용한 의미 있는 희소 항목에 대한 연관 규칙 탐사 기법," 정보과학회 논문지 데이터베이스 제 28권 제 4호, pp.577-586, 2001.
- [14] 장중혁, 이원석, "데이터 스트림에서 개방 데이터 마이닝 기반의 빈발항목 탐색," 정보처리학회 논문지 D 제10-D권 제 3호, pp.447-458, 2003.
- [15] 한승철, 강현철, "XML 스트림 데이터에 대한 연속 질의 처리 시스템," 정보처리학회 논문지 D 제11-D권 제7호, 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 데이터

베이스, 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, 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, 전자상거래 등



김 흥 기

e-mail : hkkim@dsu.ac.kr
1983년 전남대학교 계산통계학과(학사)
1986년 전남대학교 계산통계학과
(이학석사)
1996년 전남대학교 전산통계학과
(이학박사)

1991년~현재 동신대학교 컴퓨터학과 부교수
관심분야: 공간 데이터베이스, 컴퓨터 그래픽스, 멀티미디어
시스템, 스트림 데이터 마이닝 등