

# RFID 데이터 스트림에서 이동궤적 패턴의 탐사

## (Mining Frequent Trajectory Patterns in RFID Data Streams)

서 성 보\* 이 용 미\*\* 이 준 욱\*\*\* 남 광 우\*\*\*\* 류 근 호\*\*\*\*\* 박 진 수\*\*\*\*\*  
(Sungbo Seo) (Yongmi Lee) (Junwook Lee) (Kwang Woo Nam) (Keun Ho Ryu) (Jin Soo Park)

**요약** 이 논문은 RFID 데이터 스트림의 변화 특성을 고려하면서 단일 패스로 이동궤적 패턴을 실시간 추출하는 새로운 기법을 제안한다. RFID, 센서와 무선 네트워크 기술의 발달로 인해 현실 세계에서 실시간으로 데이터를 수집하고 유용한 패턴을 탐사하는 연구에 많은 관심이 집중되고 있다. 스트림 데이터에서 순차 패턴 또는 이동궤적 패턴을 탐사하는 기존의 연구 기법들은 반복적으로 데이터베이스 또는 트리를 탐색하는 고비용 문제점과 시간의 변화에 따르는 동적 특성을 실시간으로 패턴에 반영하지 못하는 단점이 있다. 제안하는 기법은 시간에 따라 RFID 데이터 스트림의 변화를 정확히 반영하기 위해 시간진화 그래프를 이용하여 이진 시간관계 테이블에 빈발한 2-길이 항목간 정보를 유지한다. 또한 다중 패스의 문제점을 해결하기 위해  $t$  시점에 이진 시간관계 테이블을 이용하여  $k$ -길이의 후보 이동궤적 패턴을 추론하고,  $t+1$  시점에서 후보 패턴을 검증하는 과정을 통해  $k$ -길이 이동궤적 패턴을 단일 패스로 추출한다. 실험결과 제안하는 기법은 기존의 Apriori-계열 기법들과 비교하여 약 7% 정도 후보 패턴의 비율이 적게 생성되어 시간 및 공간 복잡도 측면에서 우수한 성능을 보였다.

**키워드** : RFID 데이터 스트림, 스트림 데이터 마이닝, 이동궤적 패턴

**Abstract** This paper proposes an on-line mining algorithm of moving trajectory patterns in RFID data streams considering changing characteristics over time and constraints of single-pass data scan. Since RFID, sensor, and mobile network technology have been rapidly developed, many researchers have been recently focused on the study of real-time data gathering from real-world and mining the useful patterns from them. Previous researches for sequential patterns or moving trajectory patterns based on stream data have an extremely time-consuming problem because of multi-pass database scan and tree traversal, and they also did not consider the time-changing characteristics of stream data. The proposed method preserves the sequential strength of 2-lengths frequent patterns in binary relationship table using the time-evolving graph to exactly reflect changes of RFID data stream from time to time. In addition, in order to solve the problem of the repetitive data scans, the proposed algorithm infers candidate  $k$ -lengths moving trajectory patterns beforehand at a time point  $t$ , and then extracts the patterns after screening the candidate patterns by only one-pass at a time point  $t+1$ . Through the experiment, the proposed method shows the superior performance in respect of time and space complexity than the Apriori-like method according as the reduction ratio of candidate sets is about 7 percent.

**Keywords** : RFID Data Stream, Stream Data Mining, Moving Trajectory Patterns

### 1. 서론

일반적인 RFID 응용에서 물품, 차량, 사람 등의 객체는 EPC(Electronic Product Code) 등의 유일 식별코드를 내장한 RFID 태그가 부착되어 있으며, RFID 리더를 통

해 그 유일 ID와 함께 객체의 상태 정보가 연속적으로 수집된다[1]. 이 때 유일 ID와 RFID 리더기의 공간적인 위치를 결합한다면 RFID가 부착된 객체의 시공간 정보를 수집할 수 있게 된다. 기존의 물류, 의료약품, 군사, 환경 분야 등 많은 RFID 응용들에서 이러한 단순 공간상태 변

† 이 논문은 2007년도 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구(KRF-2007-350-D00015)와 국토해양부 첨단도시기술포발사업 - 지능형국토정보기술혁신사업 과제에 연구비 지원(07국토정보C05)에 의해 수행되었습니다

\* 청주대학교 정보통신연구센터 학술연구교수, sbseo@cju.ac.kr

\*\* 충북대학교 전자계산학과 박사과정, ymlee@dblab.chungbuk.ac.kr

\*\*\* 한국전자통신연구원 선임연구원, junux@etri.re.kr

\*\*\*\* 군산대학교 컴퓨터정보공학과 조교수, kwnam@kunsan.ac.kr(교신저자)

\*\*\*\*\* 충북대학교 전기전자컴퓨터공학부 교수, khryu@dblab.chungbuk.ac.kr

\*\*\*\*\* 청주대학교 정보통신공학부 교수, parkjs@cju.ac.kr

논문접수 : 2009.01.12

수정일 : 1차 2009.02.23 / 2차 2009.03.12

심사완료 : 2009.03.13

화를 추적하는 기능은 가장 중요한 역할을 제공해 왔다. 그러나 최근 끊임없이 변화하는 공간 상태를 단순히 모니터링하는 응용 분야에서 객체의 시공간 이동 패턴을 실시간으로 분석하여 발견된 패턴을 활용하는 분야로 관심이 확대되고 있다.

RFID 데이터를 수집하는 관점에서 RFID 모델을 분류하면 크게 세 가지로 분류할 수 있다 [2]. 첫째, 물류와 같이 개별 상품을 팔레트나 컨테이너에 실고 이동하면서 데이터를 수집하는 방식으로 일정한 시간 간격마다 대량의 데이터를 수집한다. 둘째, 대형 마트와 같은 오프라인 매장에서 고정된 위치에 리더를 설치한 후 데이터를 수집하는 방식이 있다. 마지막으로 자동차에 RFID 태그를 부착하고 이동하면서 도로에 설치된 리더에 의해 실시간 감지되는 방법이다. 이 기법은 매우 짧은 시간간격마다 소량의 데이터를 수집하는 방식으로 이동 객체의 경로가 개별적으로 실시간 수집하는 방식이다. 본 논문에서는 [2]의 대량의 데이터 수집방식과 달리 마지막에 제시한 소량 수집 방식의 RFID 모델로 제한된 지역 내에서 이동 객체의 궤적 데이터가 실시간 수집되는 응용분야에 적용 가능하다.

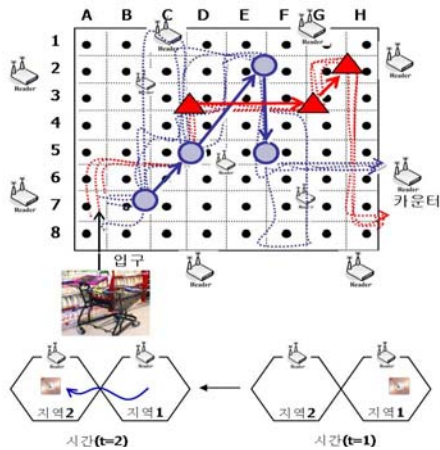


그림 1. 매장의 RFID 리더와 태그 예제

그림 1은 대형 매장에 RFID 리더가 일정한 간격마다 설치되어 있으며 쇼핑 카트에 태그와 리더가 설치되어 있는 RFID 응용 예제이다. 먼저 고객은 입구에서 카트에 개인 신원을 확인하고 쇼핑을 시작한다. 일정한 영역마다 RFID 리더를 설치했기 때문에 각각의 고객이 쇼핑을 하는 동안 이동하는 경로(그림 1의 점선들)는 리더를 통해 실시간으로 수집되며 특정한 영역을 단순히 지나갔는지, 구매를 하기 위해 머물렀는지(그림 1의 원과 삼각형 표시)는 리더에 의해 수집되는 객체의 시간 간격 데이터를 통해 유추할 수 있다. 그림 1의 예제에서 RFID를 통해 수집된 데이터는 고객의 ID(EPC), RFID 리더의 위치(location), 객체를 인식한 시간(timestamp)으로 이동 가능한 객체의 시공간 경로 데이터가 실시간 수집된다. 이

러한 RFID 응용에서 고객들이 1~2시간 동안 쇼핑하는 경로에 대해 단순히 이동궤적을 모니터링만 하는 것보다 고객들이 이동한 경로 궤적에 대한 데이터를 온라인으로 분석하여 마케팅에 활용한다면 매우 유용할 것이다. RFID 태그와 리더를 통해 일정한 시간 동안 고객들이 구매와 같은 의미 있는 행위를 하면서 특정한 공간을 순차적으로 이동한 경로 데이터에서 이동궤적 패턴(MT-Patterns: Moving Trajectory Patterns)을 추출한다. 발견된 패턴을 이용하여 고객의 유형을 분류하고 카운터에서 쇼핑을 마쳤을 때, 고객마다 차별화된 할인을 적용, 즉석 쿠폰 발행 등의 마케팅 전략을 적용한다면 매장이 더 많은 고객을 유치할 수 있다.

제한하는 MT-Patterns과 유사하게 순서를 이용하여 빈발하게 발생하는 사건들에서 k-길이의 의미 있는 패턴을 발견하는 대표적인 순차패턴 알고리즘들은 Apriori-All, GSP, DHP, FP-tree 등이 있다[3, 4]. 이들 Apriori-계열 알고리즘들은 k-항목들이 빈발하기 위해서 모든 부분 집합들이 빈발해야 하는 하향 닫힘 특성(downward closure)을 갖는다[5]. 따라서 k-길이의 순차 패턴을 탐사하는 과정에서 후보 항목 패턴과 패턴을 추출하기 위해 반복적인 데이터베이스 스캔을 해야 하기 때문에 오프라인 마이닝 기법에 적합하다. 하지만 RFID 및 센서 스트림 데이터에서 k-길이의 패턴을 온라인으로 탐사하기 위해서는 한번 또는 최소한의 데이터베이스 스캔을 해야 하는 제약성 때문에 Apriori-계열의 알고리즘을 적용하는 것은 매우 어렵다. 최근 스트림 데이터를 처리하기 위해 FP-stream[6], KSP[7], 샘플링[8] 알고리즘이 제안되었지만, RFID 및 센서에서 빠르게 수집되는 대용량의 스트림 데이터에서 윈도우 모델[7]을 적용하여 한번 또는 최소한의 데이터베이스 스캔으로 k-길이의 순차 패턴을 추출하고 제한된 메모리에서 처리하는 것은 여전히 매우 많은 비용이 소요되는 문제점이 있다.

데이터 스트림에서 반복적인 데이터 탐사 문제를 해결하기 위해, Wei는 [9]에서 일정한 슬라이딩 윈도우마다 단일 패스를 통해 빈발한 이동궤적 패턴을 탐사하는 기법에 대한 방식으로 t 시점에 t+1 시점에 발생 가능한 후보 순차 패턴을 추론하고, t+1 시점에 생성된 후보 순차 패턴을 검증하는 방식을 적용하였다. 하지만 [9]에서 제시한 기법은 k-길이의 후보 순차 패턴을 생성 단계에서 Apriori-계열 알고리즘과 같이 카디전 곱(cartesian product)를 이용하여 후보 패턴 집합을 생성하기 때문에 많은 후보 항목집합이 생성되는 문제점이 있다. 따라서 본 논문에서는 2-길이 항목들의 이진관계 테이블(binary relationship table)을 활용하여, 최적화된 k-길이 후보 순차 패턴을 추론한다. 이러한 추론 방식은 t 시점에 False Positive 오류를 범하지 않으면서 False Negative 오류를 최소화 하는 후보항목 집합을 생성하며 t+1 시점에 False Negative 오류를 제거한다. 예를 들어 P(1→5→9)와 같은 3-길이 후보 순차 패턴을 생성한다고 가정할 때, 순서를 고려한 부분 집합인 P(1→5), P(5→9), P(1→9)가 모두 빈발한 조건을 만족할 때만 P(1→5→

9)가 후보 순차 패턴으로 생성되며, 후보 항목 집합은  $t+1$  시점에 직접 빈발도를 계산하는 과정을 통해 검증한다. 이 아이디어는 일반적으로 빈발한  $k$ -길이의 순차 패턴들이 어느 시점에 갑자기 빈발하게 되거나 빈발하지 않게 될 확률이 적고, 시간의 흐름에 따라 부분 집합들의 빈발도가 완만하게 증가/감소하기 때문에 한 시점이 지연된 이후에  $k$ -길이의 MT-Patterns 탐사해도 전체적인 패턴을 탐사하는데 영향을 주지 않는다.

본 논문에서 제안하는 이동체적 패턴(MT-Patterns) 탐사 기법은 연속적으로 수집되는 RFID 스트림 데이터를 기반으로 슬라이딩 윈도우 기법을 적용하여, 한번의 스트림 데이터 탐색으로  $k$ -길이의 패턴을 추출하는 기법이다. 실험 결과 슬라이딩 윈도우 크기, 최소 지지도, 패턴의 최대 길이에 따라 수행 성능이 차이가 있지만, 일반적인 Apriori-계열 알고리즘의  $k$ -길이 후보 순차 패턴을 생성하는 방법에 비해 7% 정도 개수가 적게 생성되어 시간 및 공간 복잡도 측면에서 우수한 성능을 보였다. 추출된  $k$ -길이의 MT-Patterns의 응용분야로 동식물 감시 및 물류 시스템에서 센서와 리더의 배치 전략, 이동하는 경로 예측, 상품 배치 및 판매 전략에 적용될 수 있다. 아울러, 웹 페이지 이동패턴 분석, 비즈니스 프로세스, 허리케이션/태풍의 이동 경로 분석 등에 널리 활용 가능하다.

본 논문의 구성은 다음과 같다. 먼저 제 2장에서 관련 연구를 통해 기존의 연구와 본 논문에서 제안한 기법과의 차별성을 제시한다. 제 3장에서는 제안하는 MT-Patterns의 연구 동기와 문제를 정의한다. 제 4장에서는 시간 진화 그래프에 기반한 MT-Patterns의 탐사 알고리즘의 수행과정을 그림과 함께 설명한다. 제 5장에서는 실험 데이터 생성 과정과 실험 결과에 대한 성능 평가를 설명하고, 마지막으로 연구 결과를 요약한다.

## 2. 관련 연구

이 장에서는 제안하는 MT-Patterns 연구와 유사한 빈발패턴, 순차패턴, 프로세스 마이닝에 대한 기존 연구들에 대해 살펴보고 본 논문의 차별성 및 우수성을 제시한다.

빈발 패턴으로 가장 널리 알려진 Apriori-계열의 알고리즘은 Agrawal이 [5]에서 매장 고객들이 함께 구매하는 상품들의 연관규칙을 분석하기 위해 제안된 것이다. 이 기법은 하나의 트랜잭션에 다수 항목(items)들의 집합이 존재하고 각 트랜잭션에서 함께 구매되는 항목들의 집합을 추출하는 것이다. 이 알고리즘은 항목들의 발생 빈발도를 고려하여 최소 지지도(support) 이상을 만족하는 항목들을 패턴으로 고려하였다. 하지만 이 기법은 기본적으로 downward closure 특징[5], 다시 말해 어떤 항목들의 집합이  $X$ 에 포함된다면 이 항목들의 부분 집합 또한  $X$ 의 부분 집합을 의미하는 특성을 가지기 때문에  $k$  개의 주요 항목을 얻기 위해  $k$ 번 만큼 데이터베이스를 탐색하는 비용이 소요되는 단점이 있다. 이와 같은 단점을 극복하기 위해 GSP, DHP, FP-tree 등의 알고리즘으로 성능이 개선되었지만, 근본적으로 Apriori-계열의 알고리즘은

데이터베이스를 탐색하거나 트리를 탐색하는 비용이 매우 크다[3, 4, 7]. 따라서 위 알고리즘들을 이용하여 대용량의 스트림 데이터에서 단일 패스로  $k$ -길이의 빈발한 패턴을 탐사하는 것은 한계가 있다.

Jiawei Han은 [6]에서 데이터 스트림을 단일 패스로 처리하는 위해 FP-stream 알고리즘을 제시하였다. FP-stream은 FP-tree(Frequent Pattern tree)에 기반을 한다. FP-tree[10]는 Apriori-계열의 알고리즘과 달리 반복적인 데이터베이스 탐색 비용을 줄이고 공통된 접두사(prefix)를 이용하여 처리한다. FP-stream 기법의 장점은 FP-트리를 기반으로 하며 Tilted-time 윈도우 기법을 각 노드에 적용하여 연속된 스트림 데이터를 FP-growth 기법으로 업데이트 하면서 처리하는 기법이다. Karp는 또한 [7]에서 KPS 알고리즘을 제시하였다. 이 알고리즘은  $k$ -길이 항목 집합을 생성하기 위해 스트림 데이터에서 최적의 경계를 고려하여 패턴을 추출하는 기법이다. 하지만 이 기법들은 근본적으로 빈발 패턴을 찾는 알고리즘 특성상  $k$ -길이 항목에 대한 순차 또는 빈발 패턴에 대해 빈발 횟수를 계산하기 위해 트리나 디스크를 탐색해야 하기 때문에, 단일 패스로 스트림 데이터 처리를 위해서는 여전히 Apriori-계열과 유사한 문제점을 가지고 있다[7]. Wei는 [9]에서 단일패스로 패턴을 탐사하기 위해 사전에 후보항목을 추출하는 기법을 제안하였으나  $k$ -길이 후보항목을 추출하는 기법이 기존의 Apriori-계열과 동일한 방식으로 시간 및 공간 복잡도가 크다.

대용량의 스트림 데이터가 빠르게 입력될 경우 정확하게 처리하는 것은 현실적으로 불가능하다. 따라서 많은 연구자들은 정확한 지식 탐사보다 근사적 지식 탐사에 대한 연구를 수행하고 있으며, 대표적으로 샘플링(sampling) 기법[8, 13]을 적용한다. 샘플링은 데이터 스트림에서 가장 잘 데이터를 표현할 수 있는 샘플 집합을 샘플링하고 유지하는 기법이다. 현존하는 많은 샘플링 기법 중에서 Maku 와 Motwani가 제시한 샘플링 기법[8]은 수행 시간과 공간 측면에서 매우 유용한 기법으로 평가되고 있다. 하지만 스트림 데이터의 분포는 시간에 따라 변화하기 때문에 샘플링을 데이터 스트림의 변화 특성에 따라 수행해야 한다. 하지만 그러한 변화를 모니터링하기 매우 힘들고 샘플링 비율 및 메모리 공간도 중요한 문제이다.

데이터 스트림에서  $k$ -길이의 빈발한 패턴을 추출하기 위한 최근의 FP-stream, KPS, 샘플링 등의 연구가 전통적인 다중 패스 알고리즘에 비해 단일 패스 알고리즘으로 빠르고 정확하다고 할지라도, 실험 결과 여전히 수행 시간과 공간의 복잡도가 크다[7]. 왜냐하면 센서나 RFID에서 수집되는 데이터가 대용량이며 데이터를 수집하는 횟수가 빈발하기 때문에 스트림 데이터를 처리하는데 여전히 문제점으로 남는다. 따라서 대용량 데이터 스트림에서 단일 패스 또는 최적의 데이터 탐색으로  $k$ -길이의 패턴을 추출하는 새로운 알고리즘이 요구된다.

RFID를 이용한 관련 연구로서 [1]은 확률에 기반한 Flowgraphs를 이용하여 패턴을 탐사하는 기법을 제안하

었다. 이 연구는 RFID를 부착한 이동 객체가 이동한 경로의 상태(state)와 전이(transition)에 대한 확률 정보를 이용하여 순차 패턴을 고려하였지만 단순히 순차성만을 고려하였으며 데이터베이스를 반복적으로 탐사하는 오프라인 기법을 적용하였다. 따라서 스트림 데이터의 시간에 따라 변화하는 특성을 반영하지 못하며 반복성, 주기성 등의 다양한 패턴을 표현하지 못하는 단점이 있다.

본 논문에서는 기존 연구들의 문제점을 해결하기 위해 슬라이딩 윈도우를 기반으로 한번의 스트림 데이터 탐색으로 k-길이의 이동패턴을 탐사하는 기법을 제안한다. 제안하는 기법은 특정 t 시점에 2-항목의 이진관계 테이블을 이용하여 최적의 k-길이의 후보 패턴을 미리 추론하고, 한 시점 이후인 t+1 시점에 추론된 후보 패턴의 실제 발생 빈도를 단일 패스로 검증하는 방법을 반복적으로 적용하여 단일 패스로 패턴 탐사가 가능하다. 또한 시간의 흐름에 따라 슬라이딩 윈도우마다 2-항목 사이의 패턴의 변화에 대해 시간 진화 그래프(time-evolving graph)를 이용하여 이진관계 테이블에 주기적으로 업데이트하기 때문에 시간 변화 특성을 반영한 알고리즘이다.

### 3. 온라인 MT-Patterns 탐사 문제

이동 가능한 객체(예: 쇼핑카드, 동물)가 그림 2와 같이 특정한 지역을 순차적 또는 반복적으로 이동하는 특성이 있다고 가정하자. 이 시나리오에서 RFID 리더와 태그를 이용하면 각 객체들의 이동 경로에 대한 정보를 실시간으로 수집이 가능하며, 각 리더는 일정한 시간 간격마다 EPC, 위치(location), 시간(time) 데이터 스트림을 수집한다. 본 논문에서는 RFID 데이터 스트림을 이용하여 각 객체들이 빈발하게 이동하는 경로에 대한 패턴을 실시간으로 탐사하는데 관심이 있다.

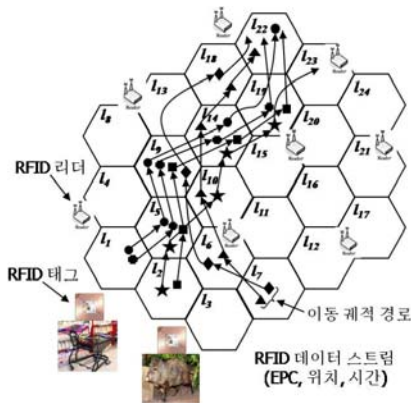


그림 2. RFID 태그와 리더 예제

이 장은 그림 2의 RFID 시나리오 예를 이용하여 연구를 수행하게 된 동기와 문제를 정의하고, 다음 장에서 이 예제를 이용하여 알고리즘 수행과정을 설명한다.

### 3.1 연구 동기

RFID 및 센서 기술과 유무선 통신 기술의 발달로 인해 현실 세계의 이동 가능한 객체가 움직이는 위치 경로에 대한 정보가 실시간으로 수집되고 있다. 일반적인 RFID 응용 시스템에 수집되는 데이터는 RFID = (EPC, location, time)와 같은 형태이다[1, 2]. EPC는 RFID 리더에 의해 읽혀진 유일한 ID(예: RFID 태그가 부착된 객체)이고, location은 RFID 리더에 의해 읽혀진 장소, time은 리더에 의해 읽혀진 시간을 의미한다[2]. 그림 2의 예에서 RFID 태그를 부착한 이동 객체들의 RFID 데이터 스트림을 이용하여 빈발하게 이동하는 경로에 대한 k-길이 순차 패턴 탐사가 가능하다. 하지만 기존의 k-길이 순차 패턴 탐사인 Apriori-계열과 트리 기반 탐사 기법은 downward closure 특성[5]을 만족하기 시키기 위해 반복적으로 데이터베이스를 탐색하거나 트리를 구성하고 탐색하는 비용이 소요된다. 따라서 기존의 기법으로 RFID 및 센서 스트림 데이터에서 단 한번 또는 최소한의 데이터 탐색과 저 비용으로 k-길이의 이동패턴을 탐사하는데 한계가 있다.

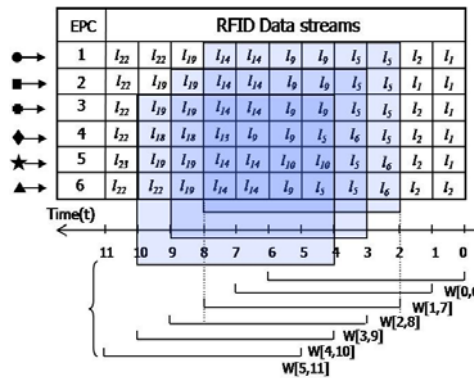


그림 3. 슬라이딩 윈도우 기반 RFID 데이터

그림 2의 시나리오 예제에서 이동 가능한 6개의 객체가 시간에 따라 특정한 지역을 이동하면서 수집하는 RFID 스트림 데이터는 그림 3과 같다. 스트림 데이터 특성상 모든 시간 크기를 고려하는 것이 불가능하기 때문에 대표적인 Landmark Window, Sliding Window, Damped Window 등 3가지 윈도우 모델[7] 중에서 슬라이딩 윈도우 모델을 고려하였다. 그림 3은 슬라이딩 윈도우 크기를 6이라 가정하고 6개의 객체가 이동한 경로 데이터를 수집하는 과정이다.

본 논문에서는 RFID 데이터 스트림에서 각 슬라이딩 윈도우마다 한번의 데이터 탐색으로 k-길이의 이동패턴을 탐사하는 새로운 알고리즘을 제시한다. 한번의 데이터 스캔으로 k-길이의 이동패턴을 탐사하기 위한 아이디어는 [9]에서 제시한 t 시점에 이진 패턴 사이의 관계를 고려하여 k-길이의 후보 이동패턴을 추론하고, t+1 시점에 직접 데이터를 탐색하여 패턴을 검증하는

연구에서 얻었다.

본 논문에서는 [9]에서 제시하는 2-길이 항목보다 긴 패턴에 대해 카디진 곱을 이용하여 모든 후보 항목을 고려하는 것이 아니라, 이진 항목간의 관계 패턴을 이용하여 False Positive 오류를 범하지 않으면서 False Negative 오류를 최소화한 최적의 후보 항목 집합을 생성한다. 예를 들어 그림 2에서  $l_1 \rightarrow l_5 \rightarrow l_9$ 와 같은 3-길이 후보 순차 패턴을 생성한다고 가정할 때 순서를 고려한 부분 집합인  $l_1 \rightarrow l_5$ ,  $l_5 \rightarrow l_9$ ,  $l_1 \rightarrow l_9$ 가 모두 빈발한 순차 패턴을 만족할 때만 후보 패턴으로 생성되며, 후보 항목 집합은  $t+1$  시점에 직접 빈발도를 계산하는 과정을 통해 검증한다. 이 아이디어는 일반적으로 빈발한 k-길이의 순차 패턴들이 갑자기 빈발하게 되거나 빈발하지 않고, 시간의 흐름에 따라 부분 집합들이 빈발도가 완만하게 변화하기 때문에 작은 시간간격은 k-길이 빈발 패턴 탐사에 크게 영향을 주지 않는다.

### 3.2 문제 정의

그림 2와 같이 RFID 데이터는 EPC, 위치와 시간 값으로 구성되며 이동 가능한 객체들의 이동궤적 데이터 스트림은 시간 값을 가진 위치들의 연속적인 값들을 의미한다. 실제 좌표를 가진 지리적인 위치 값은 응용 분야에 따라 지리적 또는 셀(cell-id)과 같은 위치로 변환되며, 시간 값도 데이터를 수집하는 주기를 고려한 단위 크기(granularity) 정수로 표시한다. [11] 연구와 같이 RFID를 부착한 객체가 이동한 경로도 일정한 범위 내부에서 이동하는 이동궤적(trajecory)을 의미한다. 본 논문에서는 순차적으로 수집되는 RFID 데이터 스트림을 다음과 같이 정의한다.

#### 정의 1: RF-stream

RFID 데이터 스트림 S는 EPC( $r$ ), 위치( $l$ ), 시간( $t$ ) 등 3개의 속성들로 구성된 튜플들의 시퀀스로서  $S = \langle (r_1, l_1, t_1), \dots, (r_n, l_n, t_n), \dots \rangle$ 로 표현된다. 이때  $t_i (i=0, \dots, \infty)$ 는 데이터 수집 시점이며,  $\forall 0 \leq i \leq \infty, t_i < t_{i+1}$ 를 만족한다.

그림 3과 같이 RF-stream은 연속적으로 수집되고 트랜잭션이 잠재적으로 무한하기 때문에 데이터 스트림을 효율적으로 처리하기 위해 슬라이딩 윈도우 기법을 적용한다. 윈도우는  $i$ 번째와  $j$ 번째 사이에 도착한 부분 순차 집합을  $i < j$ 일 때  $SW[i, j] = (t_i, t_{i+1}, \dots, t_j)$ 로 표시한다. 이때, 슬라이딩 윈도우(SW) 크기  $\omega$ 가 주어지고, 현재 시점이  $t$ 라면 빈발 순차 패턴을 탐사하는 범위는  $SW[t-\omega, t]$ 을 의미한다. 연속적인 RF-stream 데이터에서 이동 객체가 순차적으로 이동하는 궤적은 정의 2와 같이 정의한다.

#### 정의 2: k-길이 이동궤적(Moving Trajectory)

이동궤적은 시간  $t$ 가 흐름에 따라 RF-stream에서  $SW[t-\omega, t]$ 마다 객체가 k-길이만큼 순차적으로 이동하는 경로이며  $\forall 0 \leq i \leq \infty, t_i < t_{i+1} < t_{i+2}$ 일 때 k-길이 이동궤적 패턴은  $P(l_i \rightarrow l_j \rightarrow l_k)$ 로 표현한다.

이동궤적은 RFID를 부착한 이동 객체가 순차적으로

특정한 지역을 연속적으로 이동하는 경로의 집합이다. 모든 객체의 이동궤적 집합이 의미있는 패턴이 되기 위해서는 일정하게 반복되는 빈발도를 가져야 한다. 본 논문에서는 최소 지지도 이상의 빈발도를 만족하는 k-길이 이동궤적에 대해서만 k-길이 이동궤적 패턴(MT-Patterns)으로 간주하며, 지지도는 다음과 같이 정의한다.

#### 정의 3. 지지도(support = $\delta$ )

이동궤적인  $P(l_k \rightarrow l_m)$ 에 대한 지지도는 RF-stream의 슬라이딩 윈도우  $SW[t-\omega, t]$  사이에  $l_k$ 와  $l_m$ 이 순차적으로 발생한 이동 객체의 개수를 전체 이동 객체의 개수로 나누어진 값이다. 예를 들어 그림 3에서 슬라이딩 윈도우  $\omega$  값은 6이라 가정할 때  $SW[2, 8]$ 에서  $P(l_5 \rightarrow l_{14})$ 의 지지도는 5/6이다. 최소 지지도(minimum support)는 사용자가 정의하는 최소한의 빈발도 임계치를 의미한다.

정의 1~3에 따라 k-길이 MT-Patterns은 RF-stream에서 각 슬라이딩 윈도우마다 최소 지지도를 만족하는 k-길이 이동궤적 패턴을 말한다. 본 논문에서는 k-길이 MT-Patterns을 탐사하기 위해 기존의 반복적인 데이터 탐색 기법과 달리 각 슬라이딩 윈도우 데이터를 오직 한번 탐색하는 방법으로 패턴을 탐사하는 기법을 제안한다.

## 4. 시간 진화 그래프 기반 MT-Patterns 탐사

이 장에서는 3장에서 제시한 문제 정의를 기반으로 단일 패스로 MT-Patterns을 탐사하는 알고리즘 특성과 과정에 대해 설명한다. RFID 스트림 데이터에서 단지 한번의 데이터 탐색으로 MT-Patterns을 탐사하기 위해 기존의 반복적인 탐색 기법과 달리 시간 진화 그래프를 이용하여 k-길이의 후보 MT-Patterns을 추론 및 검증하는 과정과 알고리즘 수행 과정을 상세히 설명한다.

### 4.1 이진 관계를 이용한 k-길이 후보 MT-Patterns 탐사

RFID 스트림 데이터는 시간의 변화에 따라 데이터가 연속적으로 변화하는 시간 진화 특성을 가진다. 따라서 모든 시점에 동일한 패턴을 가지는 것이 아니라 시간의 변화에 따라 패턴의 특성이 변화되고 진화된다. 예를 들어 그림 1의 쇼핑센터에서 고객들이 이동하는 패턴은 오전, 오후, 저녁 시간대에 따라 변화되고, 평일, 주말에 따라 상이할 것이다. 본 논문에서는 데이터 스트림의 가장 큰 비용이 소요되는 반복적인 데이터 탐색의 문제를 해결하기 위해 시간 진화 그래프를 이용하여 사전에 최적의 k-길이의 후보 MT-Patterns을 추론하고 검증하는 방법을 적용하였다. 그림 2와 3의 시나리오에서 각 슬라이딩 윈도우( $\omega=6$ ,  $SW[0, 6]$ ,  $SW[1, 7]$ , ...)마다 객체가 이동하는 경로에 대해 노드와 연결선으로 구성된 방향성 그래프로 표현이 가능하다. 그림 4(a)는  $t=0$ 에서  $t=6$ 까지, 그림 4(b)는  $t=2$ 에서  $t=8$ 까지의 SW에서 이동 경로 그래프의 표현이다. 그림 4와 같이 RFID 스트림 데이터는 시간이 변화함에 따라 이동하는 궤적 경로가 동적으로 변화하며 이러한 경로 정보는 인접한 SW 사이(예: SW

[2,8]와 SW[3,9]에 패턴의 변화가 갑자기 크지 않기 때문에 사전에 변화에 대한 예측이 가능하다.

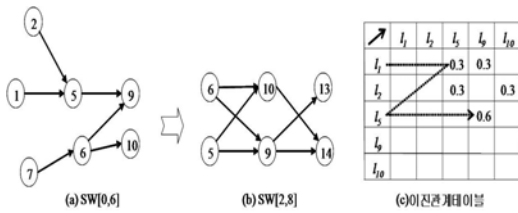


그림 4. 시간 진화 그래프와 이진관계 테이블

본 논문에서는 한번의 RFID 데이터 스트림 탐색으로 k-길이의 후보 MT-Patterns을 추출하기 위해 그림 4(c)의 이진 관계 테이블을 이용하였다. 이 테이블은 그림 4(a)의 이동 경로 그래프에서 두 개의 공간적 지역(예:  $l_1 \rightarrow l_5$ ,  $l_1 \rightarrow l_9$ ) 사이의 이동패턴의 최소 지지도를 만족하는 관계에 대한 정보를 유지한다. 예를 들어 최소 지지도를 0.3이라 할 때  $P(l_1 \rightarrow l_5)$ 의 지지도는 0.3이므로 테이블에 이진 관계 정보를 유지하고,  $P(l_1 \rightarrow l_9)$ 는 0.2로 이진 관계 테이블에 유지 되지 않는다. 이진 관계 테이블은 시간이 지남에 따라 각 슬라이딩 윈도우마다 계산된 지지도의 값을 기존에 저장된 값과의 평균값으로 갱신한다. 따라서 이진 관계 테이블은 시간이 변화함에 따라 객체들이 빈발하게 이동한 경로에 대한 최신 정보를 유지하고 있다. 그림 4(c)의 테이블은 이진 항목 사이에 단순순차적인 표현 이외에 반복성(repetition)과 주기성(cycle) 표현이 가능하다.

사전에 최적의 k-길이의 후보 MT-Patterns을 추출하기 위해 그림 4(c)의 이진 관계 테이블을 이용한다. 정의 3의 최소 지지도를 만족하는 이진 관계 패턴이 존재할 때 정의 4와 같이 k-길이 후보 MT-Patterns이 생성된다. 정의 4를 만족하는 후보 패턴은 False Positive 조건을 만족하며, False Negative 조건을 최소화 한 후보 패턴 추출 기법으로 [4, 5, 6]의 Apriori-계열의 추출 기법에 비해 약 7%정도 후보 패턴의 개수를 줄임으로써 수행 성능을 개선할 수 있다.

정의 4. k-길이 후보 MT-Patterns

이진 관계 테이블의 이진 순차 패턴들이 모두 최소 지지도를 만족한다면, 이진 순차 패턴들의 순열 조합에 의해 생성된 k-길이 패턴을 후보 MT-Patterns이라 한다. 예를 들어 그림 4(c)에서 최소 지지도( $\delta$ )를 0.3이라 가정할 때,  $P(l_1 \rightarrow l_5) > 0.3 \wedge P(l_5 \rightarrow l_9) > 0.6 \wedge P(l_1 \rightarrow l_9) > 0.3$ 이라면  $P(l_1 \rightarrow l_5 \rightarrow l_9)$ 는 후보 MT-Patterns이 된다.

정의 4에서 추출된 후보 MT-Patterns은 추출된 다음 슬라이딩 윈도우에서 최소 지지도를 만족하는지 검증하여 최종 MT-Patterns이 된다. 예를 들어 그림 3의 SW[2,8]에서 정의 4에 의해 추출된 후보 MT-Patterns,  $P(l_5 \rightarrow l_9 \rightarrow l_{14})$ 는 SW[3,9]에서 실제 빈발도를 계산하여 최소 지지

도를 만족할 때 MT-Patterns이 된다. 따라서 RFID 데이터 스트림에서 MT-Patterns은 유효한 시간간격과 정보와 함께 패턴 정보가 데이터베이스에 저장된다. 따라서 스키마는 <트랜잭션 ID(TID), MT-Patterns,  $t_s$ (시작 시간),  $t_e$ (종료 시간)>로 구성된다.

4.2 k-길이 MT-Patterns 탐사 알고리즘

RFID 데이터 스트림에서 슬라이딩 윈도우마다 그림 4(a)의 시간 진화 그래프를 이용하여 그림 4(c)의 이진 관계 테이블을 점진적으로 갱신한다면 시간의 변화에 따라 한번의 데이터 탐색으로 k-길이 MT-Patterns을 탐사하는 것이 가능하다. 이 절에서는 그림 3에서 제시한 예를 통해 k-길이 MT-Patterns을 탐사하는 알고리즘 수행 과정을 설명한다.

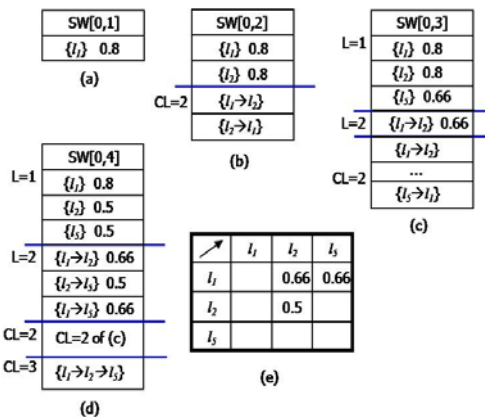


그림 5. MT-Patterns 알고리즘 수행단계

먼저 그림 3의 슬라이딩 윈도우 크기가 6이므로  $t=6$  이후에 한 시점씩 슬라이딩 윈도우가 변경된다. 그림 5는  $t=0$ 에서  $t=4$  시점까지 단계별로 k-길이 MT-Patterns을 추출하는 과정으로 상세한 설명은 다음과 같다.

- 1) 그림 5(a):  $t=1$  시점으로 SW[0,1]는 먼저 최소 지지도(정의 3 참조,  $\delta=0.3$ )를 만족하는 빈발한 1-길이 항목 집합(L=1)을 추출한다(예:  $l_1=5/6$ ). 추출된 1-길이 항목 집합( $l_1$ )이 1개이므로 2-길이 후보 항목 집합(CL=2)을 생성하지 못한다.
- 2) 그림 5(b):  $t=2$  시점으로 SW[0,2]는 그림 5(a)와 같이 최소 지지도를 만족하는 1-길이 항목 집합( $l_1, l_2$ )을 추출하고 지지도를 계산한다. 이 단계에서 1-길이 항목 집합을 이용하여  $CL=2(l_1 \rightarrow l_2, l_2 \rightarrow l_1)$ 을 생성한다. 이때, 생성된 CL=2의 후보 집합에 대하여 그림 5(e)와 같은 이진관계 테이블을 구축한다.
- 3) 그림 5(c):  $t=3$  시점으로 SW[0,3]은 L=1을 추출( $l_1, l_2, l_3$ )하여 지지도를 계산하고 CL=2을 생성한다. 아울러  $t=2$ 에서 생성된 CL=2의 집합에 대한 지지도를 계산(예:  $l_1 \rightarrow l_2=0.66, l_2 \rightarrow l_1=0.0$ )하여 최소 지지도 이상을 만족하면 그림 5(e)의 이진관계 테이블에 지지도를 저장한다.

4) 그림 5(d):  $t=4$  시점으로  $SW[0,3]$ 은  $L=1$ 을 추출( $l_1, l_2, l_3$ )하여 지지도를 계산하고  $CL=2$ 을 생성한다. 아울러  $t=3$ 에서 생성된  $CL=2$ 의 집합에 대한 지지도를 계산하여 최소 지지도를 만족하는  $l_2 \rightarrow l_3$ 와  $l_1 \rightarrow l_3$ 을 이진관계 테이블에 삽입한다. 이 단계에서  $L=2$ 을 이용하여  $CL=3$ 을 생성하기 위해 정의 4의  $k$ -길이 후보 MT-Patterns 조건을 만족하는  $l_1 \rightarrow l_2 \rightarrow l_3$ 을 생성한다. 생성된  $CL=2$ 와  $CL=3$ 은  $t=5$  시점에서 지지도를 계산하여 최소 지지도를 만족하는 패턴에 대해 MT-Patterns이 생성된다.

그림 5와 같이 제안하는  $k$ -길이 MT-Patterns 알고리즘은 한 시점 이전에  $L=2$ 의 이진 관계 테이블을 이용하여 정의 4의 조건을 만족하는  $k$ -길이 후보 항목 집합을 생성하고, 다음 시점에 후보 항목 집합을 검증하는 과정을 통해 단일 패스로 MT-Patterns을 발견한다. 제안하는 기법은 작은 시간 지연 이후에 패턴이 발견되는 문제점 있지만, 전체적인 MT-Patterns의 관점에서 작은 시간 차이는 영향이 크지 않다. 예를 들어  $l_1 \rightarrow l_2 \rightarrow l_3$  패턴이  $SW[0,3]$  윈도우에 추출되어야 하지만 두 시점 이후인  $SW[0,5]$ 에서 패턴으로 발견된다. 알고리즘 1은 그림 5의 과정에 대한 원시코드이다.

### 5. 실험 및 성능 평가

#### 5.1 실험 데이터 생성 및 구현환경

본 논문에서 제안하는 MT-Patterns 탐사 기법은 입력 데이터가 RFID 리더와 태그에 의해 실시간 수집되는 스트림 데이터를 기반으로 한다. 실제 RFID 리더와 태그를 이용한 실험 데이터를 이용하는 것이 현실적으로 불가능하여 데이터 생성기[12]를 이용하여 그림 6과 같이 모조

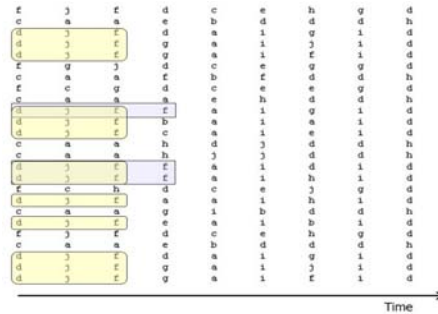
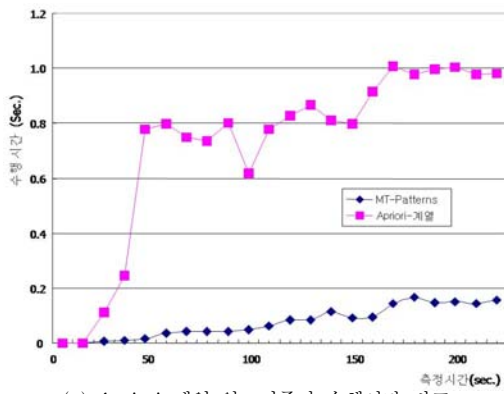
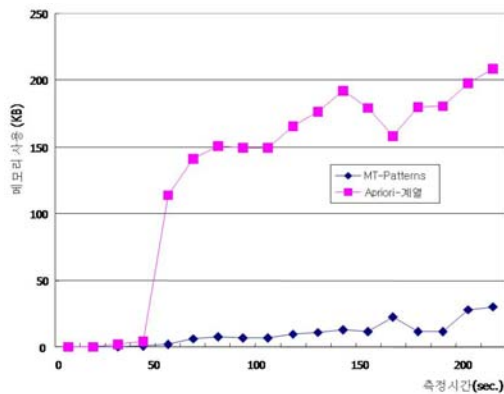


그림 6. RFID 데이터 스트림 생성기[12]

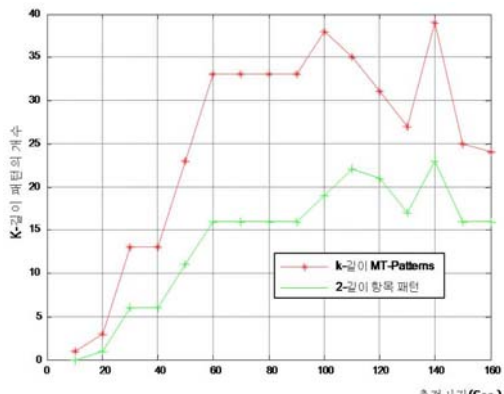
<p><b>알고리즘 1. Find_MT-Patterns()</b></p> <p><b>입력:</b> N개의 이동 객체에서 수집되는 RFID 데이터 스트림(정의 1), <math>\omega</math>(슬라이딩 윈도우 크기: 정의 1), <math>\delta</math>(최소 지지도: 정의 3)</p> <p><b>출력:</b> <math>k</math>-길이 MT-Patterns의 집합(정의 2)</p> <pre> 1.  t=0; 2.  buf = Null; //슬라이딩 윈도우 크기 만큼 저장하기 위한 윈도우 3.  L1={최소 지지도를 만족하는 1-항목 집합}; 4.  If (L1.length() &gt;= 2) 5.    CL2 = {1-항목 집합을 이용하여 2-길이 후보 항목집합 생성} 6.  While(true) { 7.    Wait until t = t+1; 8.    If(!isWindowFull) 9.      buf = sensingData(t); // n 객체에서 t 시점에 수집된 데이터 10.   If( L1 and CL2 exist) 11.     ScreeningSlidingWindow(buf); 12.   UpdateWindowSize(); //슬라이딩 윈도우 크기 증가 13.   Else 14.     Databuf = buf[t-<math>\omega</math>,t] // <math>\omega</math> 크기만큼 데이터 버퍼 유지 15.     ScreeningSlidingWindow (Databuf); 16.     MT-Patterns = MT-Patterns <math>\cup</math> {New MT-Patterns} 17.   } 18. // 슬라이딩 윈도우마다 L1, CL2 추출하고 이진관계 테이블을 업데이트하는 함수 19. Function ScreeningSlidingWindow (Buffer buf) { 20.   Finding-L1(buf); //buf에서 L1항목 추출 21.   Finding-L2(buf, CL2) in BinaryRelationTable; //CL2에서 L2 항목 추출 22.   UpdateBinaryRelationTable(L2); // 그림 5(e) 이진관계 테이블 업데이트 23. } 24. Function Finding-L1(Buffer buf) { 25.   Foreach(P) // P는 1-길이 항목들의 집합 26.     P.sup = Num. of Items/Num. of Objects in <math>\omega</math> // <math>\omega</math> 범위에서 객체의 발생빈도 27.   If (P.sup &gt; <math>\delta</math>) 28.     P <math>\in</math> L1; //최소지지도를 만족하는 항목 선택 29. }</pre>
---



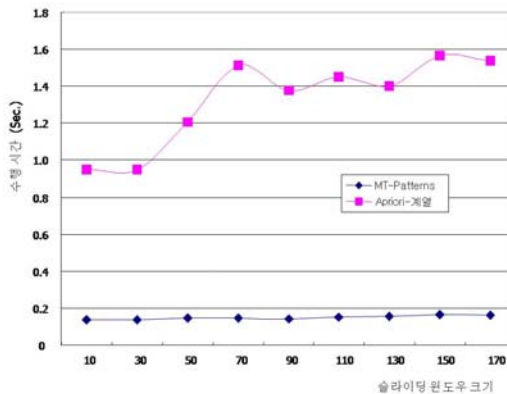
(a) Apriori-계열 알고리즘과 수행시간 비교



(b) Apriori-계열 알고리즘과 메모리 사용 비교



(c) 2-항목 패턴과 k-길이 후보항목 패턴 비율



(d) 슬라이딩 윈도우 크기 변화에 따른 비교

그림 7. Apriori-계열과 MT-Patterns 탐사 기법의 성능평가

데이터를 생성하였다. 이 생성기의 장점은 현실 세계에서 발생 가능한 데이터를 생성하기 위해 매우 엄격한 규칙을 적용하며 도메인, 데이터 분포, 순차성 및 주기성이 표현 가능한 데이터 생성기이다.

본 논문에서는 데이터 생성기를 이용하여 몇 개의 빈발한 패턴들을 생성하기 위해 결합 분류 규칙(conjunctive classification rules)을 적용하였다. 예를 들어 4개의 이동 객체를 분류하고 이들이 특정한 지역을 유사하게 순차적으로 하는 패턴을 생성하기 위해 그림 6과 같이 결합 규칙(IF A1 =d AND A2=j AND A3=f, THEN object =1)을 적용하였다. 가우시안 함수를 이용하여 랜덤 데이터를 생성하였으며 인위적으로 빈발 패턴을 발생시키기 위해 그림 6의 튜플들과 같이 각 이동 객체들이 d, j, f 등의 유사한 지역에서 발생되도록 데이터를 생성하였다. 실험 데이터는 100개의 객체가 26개(a~z) 지역을 순차적으로 이동한다고 가정하여 생성된 모조 데이터를 사용하였으며, 2GB 메모리를 가진 펜티엄 4 워크 스테이션에서 수행하였다.

프로그램 구현 언어는 Matlab 7.0 버전을 이용하였으며, 그림 6과 같이 입력 RFID 스트림 데이터는 파일에서

1초마다 각 객체에서 200개의 튜플이 수집된다고 가정하며, 입력 파라미터는 슬라이딩 윈도우 크기, 최소 지지도와 패턴의 최대 길이를 입력 받았다. 실험을 통해 우리는 시간이 진행하면서 수행시간을 관찰하여 단일 패스로 이동계적 패턴을 탐사가 가능한지 평가하였다. 또한 시간이 진행하면서 패턴의 개수가 증가하는 측면에서 메모리 사용량을 비교하였다.

### 5.2 성능 평가

실험 데이터를 이용한 성능 평가에서 제안하는 k-길이 MT-Patterns 탐사 기법은 연속적으로 수집되는 RFID 데이터 스트림을 단일 패스로 처리가 가능함을 알 수 있었다. 그림 7(a)는 제안하는 MT-Patterns과 Apriori-계열의 수행시간을 비교한 것이다. 제안 기법은 시간이 흐름에 따라 일정한 패턴이 검출되면서 안정적인 성능을 보이는 반면에 모든 2-길이 후보 항목을 고려하는 Apriori-계열은 급격하고 추가적인 수행시간 변화를 보였다. 그림 7(b)는 측정시간에 따른 메모리 사용량에 대한 비교이다. MT-Patterns과 달리 Apriori-계열 알고리즘은 일정한 시간 이후에 급격하게 메모리 사용량이 증가



하는 형태를 보인다. 그림 7(c)는 2-길이 항목 패턴과 k-길이 MT-Patterns의 개수를 비교하였다. 이 실험 결과와 같이 2-항목 패턴과 k-길이 MT-Patterns의 개수가 거의 동일한 비율로 증가하거나 감소하는 것은 정의 4에서 제안하는 2-길이 항목에서 k-길이 후보 MT-Patterns을 추출하는 기법이 최적화되었음을 알 수 있다. 마지막으로 그림 7(d)는 슬라이딩 윈도우 크기에 따른 수행시간을 비교하였다. 실험 결과 슬라이딩 윈도우는 패턴의 길이인 k 값이 크게 변화되지 않아서 시간이 변화함에 따라 큰 변화를 보이지 않았다.

실험을 통해 k-길이 MT-Patterns으로 발견된 패턴의 개수를 관찰한 결과 2-길이 항목 패턴의 관계를 이용하여 추출된 k-길이 후보 항목의 집합이 기존의 모든 항목들의 집합을 카디전 곱으로 생성한 후보 항목에 비해 약 7% 정도 절약되는 것을 알 수 있었다. 제안하는 기법은 스트림 데이터의 동적 변화를 실시간 반영하기 위해 시간진화 그래프를 이용하며 계산된 결과는 이진관계 테이블에 저장된다. 따라서 k-길이 후보 MT-Patterns을 추론하는 과정에서 최신의 정보를 반영하기 때문에 최적의 후보 항목 집합을 생성하여 수행 성능이 우수하다.

## 6. 결론

RFID 리더와 태그를 이용하여 끊임없이 수집되는 대용량 RFID 데이터 스트림에서 단일 패스로 k-길이의 빈발한 MT-Patterns을 탐사하는 기법을 제안하였다. 제안하는 기법은 2-길이 항목 패턴을 기반으로 사전에 k-길이의 빈발한 후보 MT-Patterns을 추론하고, 다음 시점에서 후보 패턴을 검증하는 새로운 기법이다. 이 기법을 통해 기존의 Apriori-계열 또는 트리 기반 패턴 탐사에서 가장 큰 비용이 소모되는 반복적인 데이터 스캔 비용과 고비용의 downward closure 문제를 해결하여 단일 패스로 데이터 스트림에서 MT-Patterns 탐사가 가능하다. 실험 결과 MT-Patterns은 시간 및 공간 복잡도 측면에서 슬라이딩 윈도우와 2-길이 항목 개수에 의존적이지만 우수한 성능을 보였고, k-길이 후보 패턴의 생성 비용이 기존 Apriori-계열에 비해 약 7% 정도 절약됨을 알 수 있다.

## 참고 문헌

- [1] Gonzalez, H., Han, J. and Li, X., "FlowCube: Constructing RFID FlowCubes for Multi-Dimensional Analysis of Commodity Flows," In Proc. of 32th Int. Conf. on Very Large Database, pp. 834-845, 2006.
- [2] Gonzalez, H., Han, J., Li, X. and Klabjan, D., "Warehousing and Analyzing Massive RFID Data Sets," In Proc. of 22nd Int. Conf. on Data Engineering, pp. 83-93, 2006.
- [3] Han, J., Cheng, H., Xin, D. and Yan, X., "Frequent pattern mining: current status and future directions," Journal of Data Mining and Knowledge Discovery, Vol. 14, No.1, pp. 55-86, 2007.
- [4] Zhao, Q. and Bhowmick, S. S., "Sequential Pattern Mining: A Survey," Technical Report Center for Advanced Information Systems, School of Computer Engineering, Nanyang Technological University, Singapore, 2003.
- [5] Agrawal, R. and Srikant, R., "Fast algorithms for mining association rules," In Proc. of 20th Int. Conf. on Very Large Database, pp. 487-499, 1994.
- [6] Giannella, C., Han, J., Pei J., Yan, X. and Yu, P. S., "Mining Frequent Patterns in Data Streams at Multiple Time Granularities," In H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (eds.), Next Generation Data Mining, Chapter 3, P. 191-212, AAAI/MIT, 2003.
- [7] Jin, R. and Agrawal, G., "Frequent Pattern Mining in Data Streams," In Charu C. Aggrawal (eds), Chapter 4, p. 61-84, Springer Science and Business Media, LLC., 2007.
- [8] Manku, G. S. and Motwani, R., "Approximate Frequency Counts Over Data Streams," In Proc. of 28th Int. Conf. on Very Large Database, pp. 346-357, 2002.
- [9] Teng, W. G., Chen, M. G., and Yu, P. S., "A Regression-Based Temporal Pattern Mining Scheme for Data Streams," In Proc. of 29th Int. Conf. on Very Large Database, pp. 93-104, 2003.
- [10] Han, J., Pei, J., Yin, Y. and Mao, R., "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," Data Mining and Knowledge Discovery, Vol. 8, No. 1, pp. 53-87, 2004.
- [11] Yunhao, L., Lei, C., Pei, J., Chen, Q. and Zhao, Y., "Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays," In Proc. of 5th IEEE Int. Conf. on Pervasive Computing and Communications, 2007.
- [12] Gabor Melli, Data Generator, <http://www.datasetgenerator.com/>
- [13] 김철연, 임종화, T. Ng Raymond, 심규석, "정량 정보를 포함한 순차 패턴 마이닝 알고리즘," 한국정보과학회 논문지 D-데이터베이스, Vol. 33, No. 5, pp 453-462, 2006.
- [14] 김의찬, 김계현, 이철용, 박은지, "비트 클러스터링을 이용한 빈발 패턴 탐사의 성능 개선 방안," 한국공간정보시스템학회 논문지 Vol.9, No.1, pp.105-115, 2007.
- [15] 유기현, 남광우, "공간 데이터스트림을 위한 조인 전략 및 비용 모델," 한국공간정보시스템학회 논문지,

Vol.10, No.4, pp.59-66, 2008.

- [16] 김동현, 이기형, 홍봉희, 반재훈, "RFID 태그 객체의 위치 추적을 위한 색인 구조의 설계 및 구현," 한국공간정보시스템학회 논문지, Vol.7, No.2, pp.67-79, 2005.



서 성 보

1999년 서원대학교 전자계산학과 졸업 (학사)  
 2001년 충북대학교 전자계산학과 졸업 (석사)  
 2001년~ 2003년 한국전자통신연구원 우정기술연구센터 연구원

2006년 충북대학교 전자계산학과 졸업 (박사)  
 2006년~2007년 미국 North Carolina State University 전산학과 Post-Doc.  
 2007년~현재 청주대학교 정보통신연구센터 학술연구교수  
 관심분야는 시공간 데이터베이스, 스트림 데이터마이닝, USN 미들웨어 시스템, u-GIS.



이 용 미

2002년 충북대학교 컴퓨터과학과 졸업 (학사)  
 2005년 충북대학교 전자계산학과 졸업 (석사)  
 2005년~현재 충북대학교 전자계산학과 박사과정

관심분야는 시공간 데이터베이스, 스트림 데이터 처리, 스트림 데이터 마이닝



이 준 옥

1997년 충북대학교 전자계산학과 졸업 (학사)  
 1999년 충북대학교 전자계산학과 졸업 (석사)  
 2003년 충북대학교 전자계산학과 졸업 (박사)  
 2003년~현재 한국전자통신연구원 융합

기술연구부문 RFID/USN연구부  
 관심분야는 USN 미들웨어 플랫폼, 시공간 데이터 마이닝, 센서 데이터 마이닝



남 광 우

1995년 충북대학교 전자계산학과 졸업 (학사)  
 1997년 충북대학교 전자계산학과 졸업 (석사)  
 2001년 충북대학교 전자계산학과 졸업 (박사)

2001년~2004년 한국전자통신연구원 텔레메틱스연구단  
 2004년~현재 군산대학교 컴퓨터정보공학과 조교수  
 관심분야는 데이터베이스, GIS, LBS 정책 및 기술, 데이터스트림, 지오 센서네트워크



류 근 호

1976년 숭실대학교 전산학과 졸업 (학사)  
 1980년 연세대학교 공학대학원 전산전공 (석사)  
 1988년 연세대학교 대학원 전산전공(박사)  
 1976년~1986년 육군군수지원사 전산실 (ROTC 장교), 한국전자통신연구원(연구원), 한국방송통신대 전산학과(조교수) 근무

1989년~1991년 Univ. of Arizona Research Staff(TempIS 연구원, Temporal DB)  
 1986년~현재 충북대학교 전자정보대학 교수  
 관심분야는 시공간 데이터베이스, GIS, 지식기반 정보검색 시스템, 유비쿼터스 컴퓨팅 및 스트림 데이터 처리, 데이터 마이닝, 데이터베이스 보안, 바이오인포메틱스



박 진 수

1977년 한양대학교 전자공학과 졸업 (학사)  
 1977년 한양대학교 대학원 (전자통신) 졸업 (석사)  
 1985년 한양대학교 대학원(디지털통신) 졸업 (박사)

1987년~1988년 University of Colorado Post-Doc.  
 1991년~1992년 IEEE Daejeon Section Chairman  
 1991년~1992년 대한전자공학회 충북지부 지부장  
 1998년~1999년 한국통신학회 충북지부 지부장  
 2000년~현재 한국정보처리학회 부회장  
 1999년~2007년 정보통신연구센터 소장  
 2006년~현재 충북 TIP 전자정보부품산업지원센터 기술자문위원  
 1978년~현재 청주대학교 정보통신공학부 교수  
 관심분야는 FHSS, DSSS, USN, 데이터스트림