

# 다변량 스트림 데이터 축소 기법 평가

정 훈 조<sup>†</sup> · 서 성 보<sup>\*\*</sup> · 최 경 주<sup>\*\*\*</sup> · 박 정 석<sup>\*\*\*\*</sup> · 류 근 호<sup>\*\*\*\*\*</sup>

## 요 약

센서 네트워크는 애플리케이션 분야에 따라 데이터 특성과 사용자의 요구사항이 다양함에도 불구하고, 현존하는 스트림 데이터 축소 연구는 데이터의 본질적인 특징보다 특정 축소 기법의 성능 향상 측면에 중점을 두고 있다. 이 논문은 계층/분산형 센서 네트워크 구조와 데이터 모델을 소개하고, 선택적으로 축소 기법을 적용하기 위해 데이터 특성과 사용자의 요구에 적합한 다변량 데이터 축소 기법을 비교 평가한다. 다변량 데이터 축소 기법의 성능을 비교 분석하기 위해, 우리는 웨이블릿, HCL(Hierarchical Clustering), SVD(Singular Value Decomposition), 샘플링과 같은 표준화 된 다변량 축소 기법을 이용한다. 실험 데이터는 다차원 시계열 데이터와 로봇 센서 데이터를 사용한다. 실험 결과 SVD와 샘플링 기법이 상대 에러 비율과 수행 성능 측면에서 웨이블릿과 HCL기법에 비해 우수하였다. 특히 각 데이터 축소 기법의 상대 에러 비율은 입력 데이터 특성에 따라 다르기 때문에 선택적으로 데이터 축소 기법을 적용하는 것이 좋은 성능을 보였다. 이 논문은 다차원 센서 데이터가 수집되는 센서 네트워크를 디자인하고 구축하는 응용 분야에 유용하게 활용될 것이다.

키워드 : 센서 네트워크 애플리케이션, 다변량 스트림 데이터, 스트림 데이터 축소

## Evaluation of Multivariate Stream Data Reduction Techniques

Hun Jo Jung<sup>†</sup> · Sungbo Seo<sup>\*\*</sup> · Kyung Joo Cheoi<sup>\*\*\*</sup> · Jeong Seok Park<sup>\*\*\*\*</sup> · Keun Ho Ryu<sup>\*\*\*\*\*</sup>

## ABSTRACT

Even though sensor networks are different in user requests and data characteristics depending on each application area, the existing researches on stream data transmission problem focus on the performance improvement of their methods rather than considering the original characteristic of stream data. In this paper, we introduce a hierarchical or distributed sensor network architecture and data model, and then evaluate the multivariate data reduction methods suitable for user requirements and data features so as to apply reduction methods alternatively. To assess the relative performance of the proposed multivariate data reduction methods, we used the conventional techniques, such as Wavelet, HCL(Hierarchical Clustering), Sampling and SVD (Singular Value Decomposition) as well as the experimental data sets, such as multivariate time series, synthetic data and robot execution failure data. The experimental results shows that SVD and Sampling method are superior to Wavelet and HCL with respect to the relative error ratio and execution time. Especially, since relative error ratio of each data reduction method is different according to data characteristic, it shows a good performance using the selective data reduction method for the experimental data set. The findings reported in this paper can serve as a useful guideline for sensor network application design and construction including multivariate stream data.

Key Words : Sensor Network Application, Multivariate Stream Data, Stream Data Reduction

## 1. 서 론

소형 무선 디바이스와 센서로 구성된 Wireless Sensor Networks(이하WSN) 환경에서 다양한 센서에서 수집된 대용량 스트림 데이터를 관리하는 연구가 활발히 진행되고 있

다[1, 2]. 대부분의 연구는 WSN의 제한된 통신 대역폭, 소형 프로세서, 적은 전력과 실시간으로 데이터를 처리해야 하는 제약성을 기반으로 한다[1, 2, 3, 5]. 그리고, 이 연구들은 데이터 수집, 전송과 질의 모델에 대한 고려가 하드웨어 및 네트워크 기반의 제약성을 해결하기 위한 연구에 집중되었다. 하지만 WSN의 응용 분야는 사용자 요구사항과 애플리케이션의 데이터 특성에 따라 다양하기 때문에, 이에 적합한 데이터 수집 방식, 데이터 전송 및 질의처리 방식이 함께 고려되어야 한다. 아울러 WSN의 구성 요소인 센서 노드, 싱크 노드, 베이스 스테이션과 서버의 각 레벨에서 데이터 저장 방식, 질의 처리 방법과 데이터 전송 방식을 WSN의 애

\* 본 연구는 산업자원부·한국산업기술평가원 지정 청주대학교 정보 통신 연구센터의 지원에 의한 것입니다.

† 중신회원 : 한서대학교 컴퓨터정보학과 교수

\*\* 준 회원 : North Carolina State University 박사후 연수

\*\*\* 비 회원 : 충북대학교 전기전자컴퓨터공학부 전임강사

\*\*\*\* 중신회원 : 충주대학교 전기전자 및 정보공학부 부교수

\*\*\*\*\* 중신회원 : 충북대학교 전기전자컴퓨터공학부 교수

논문접수 : 2005년 9월 12일, 심사완료 : 2006년 9월 6일

플리케이션에 따라 적절하게 적용하는 것이 바람직하다.

센서 네트워크의 일반적인 구조는 계층 또는 분산형 모델이다[4, 5]. 이 모델의 장점은 데이터 수집 측면에서 대용량의 원본 데이터가 센서 노드에서 생성되기 때문에, 생성된 데이터는 센서 네트워크 내부 단계에서 정제, 필터링, 집계 연산을 수행하는 것이 효율적이다. 또한 지리적 측면에서 노드간에 데이터 전송시 다중-홉 방식과 데이터 중심의 클러스터 기반 통신 방식이 적용됨으로 지역성을 고려한 데이터 수집 처리 방식이 유리하다.

센서 네트워크의 응용분야는 크게 환경 모니터링(environmental monitoring), 객체 추적(object tracking), 객체 감시(object guarding)로 분류된다[4, 6, 7, 11]. 환경 모니터링은 홍수 감시, 빌딩 모니터링, 동식물/생태 감시 분야이며, 저주파 데이터 특성을 가지며 장기적인 데이터 분석이 가능하다. 객체 추적은 이동객체 추적, 군사분야, 물류분야이며, 고주파 데이터 특성과 다차원 객체 속성을 가진다. 객체 감시는 응급의료 관리, 침입탐지와 지진 탐사 분야이며, 데이터의 비정상적인 변화와 같은 이상치의 특성과 실시간 처리 특성이 있다. 이와 같이 WSN의 응용 분야에 따라 다양한 특성을 가지므로 WSN의 구조와 응용에 따라 데이터 모델을 수립하고 데이터 특성에 따라 데이터 수집, 전송과 저장 방법에 대한 연구가 함께 고려되어야 한다.

하나의 센서 노드는 하나 이상의 센서(예: 온도, 습도, 조도 등)를 가질 수 있으며, 여러 개의 센서에서 주기적으로 수집되는 센서 데이터는 다차원 속성을 가진다. 기존의 센서 데이터 수집 및 전송 방식은 대부분 단일 또는 몇 개의 속성을 고려한 축소기법을 제안하고 있다[2, 8]. 하지만 객체 추적과 감시 분야는 여러 센서 노드로부터 수집되는 객체를 관리하고 다양한 속성을 분석해야 하기 때문에, 다차원 속성이며 각 센서 노드의 정보가 수집되는 싱크 노드는 대용량 다차원 속성을 포함하게 된다. 따라서 단순 속성의 스트림 데이터와 WSN의 하드웨어 제약성 해결을 위한 데이터 축소 연구[9, 13] 이외에 WSN의 애플리케이션 특성에 적합한 다변량 스트림 데이터 축소 연구가 수행 되어야 한다[26].

이 논문에서는 WSN 환경의 다변량 스트림 데이터에 대한 데이터 축소 기법 평가를 위해 다음과 같은 연구를 수행한다. 첫째, WSN의 계층/분산 구조와 데이터 모델을 정의한다. 둘째 WSN의 응용 분야를 분류하며, 실험을 위한 시뮬레이션 모델과 다변량 데이터 축소 기법을 비교 분석한다. 실험을 위한 다변량 데이터 축소 기법은 웨이블릿(Wavelet), HCL(Hierarchical Clustering), 샘플링(Sampling)과 SVD(Singular Value Decomposition) 기법을 이용하였으며 기본적인 특징과 성능 측면을 간략히 소개한다. 마지막으로 우리는 실험에서 다변량 시계열 데이터, 로봇 센서 데이터와 랜덤 웨이브 생성기에 의해 생성된 데이터 셋을 이용하여 데이터 축소 기법에 대한 상대 에러와 성능 측면을 평가한다. 아울러 각 데이터 타입과 응용 분야에 따라 데이터 수집, 전송 측면에서 다변량 데이터 축소 기법을 어떻게 적용하는 것이 바람직한지 분석한다.

이 논문의 구성은 다음과 같다. 제 2장에서는 관련 연구를 제시한다. 제 3장에서는 WSN의 시스템 구조로 계층/분산형 모델과 데이터 축소 모델을 제시한다. 제 4장에서는 이 연구에서 제시하는 WSN의 시뮬레이션 모델과 다변량 데이터 축소 기법에 대해 비교 분석한다. 제 5장에서는 실험 데이터를 이용하여 다변량 축소 기법의 실험 결과를 평가한다. 제 6장에서는 결론을 제시한다.

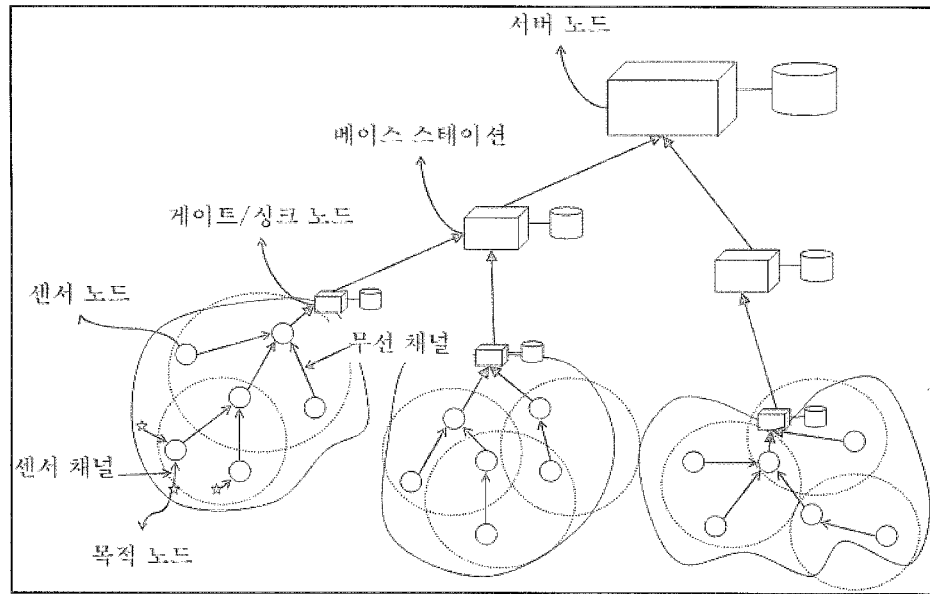
## 2. 관련 연구

일반적으로 센서 장비의 소형화와 근거리 무선 데이터 통신의 제약성을 갖는 WSN에서 끊임없이 수집되는 스트림 데이터는 전송 단계에서 메모리, 프로세스, 전력 측면에서 많은 비용이 소모되는 단점이 있다. 스트림 데이터 축소 전송과 관련된 연구[2, 3, 10, 18]의 대부분은 센서 노드와 싱크 노드 또는 베이스 스테이션 사이에 데이터 집계와 근사적 전송 및 표현에 대한 연구이다. 이 연구들은 센서 네트워크의 하드웨어와 네트워크 제약성 문제를 해결하기 위한 방법이며, 실제 다차원 속성의 센서 데이터를 활용하는 응용 분야의 데이터 요구사항과 표현 방법은 크게 고려되지 않았다.

센서 네트워크의 데이터 분석 및 축소기법에 대한 연구는 클러스터링, 웨이블릿, 히스토그램, 회귀분석, 집계, 샘플링, PCA와 SVD기법이 있다. 집계 기법[2, 3]은 측정된 데이터를 요약(예: Avg(), Max(), Min(), Var())하는 것으로 센서에서 수집된 동일한 속성의 데이터를 요약하여 분석하는데 유용하지만, 상세한 이력 정보가 손실 될 수 있다. 스펙트럼 계열의 기법(예1): DWT, DFT, DCT)은 저주파 데이터 특성에 적합하며 수행 속도와 복잡도가 작은 것이 장점이지만, 다차원 속성에는 시간 및 공간 복잡도 비용이 많이 든다[12, 13, 15]. 클러스터링 계열[18]의 기법은 속성간에 관계를 고려하여 유사성 그룹을 생성하는 것으로 대용량 다차원 속성 분석에 적합하다. 샘플링 기법[12, 13]은 단순하면서 좋은 성능을 가지므로 많이 적용되지만 원본 데이터의 샘플링 비율, 임의의 스카마 조인과 집합 특성 데이터에 대한 축소시 결과가 불명확한 단점이 있다. 회귀분석, PCA와 SVD를 이용한 축소 연구[2, 14, 15, 16]는 데이터의 상호 관계를 이용하여 가장 좋은 특징을 가지는 데이터 속성을 적용하여 근사적 데이터 표현에 많이 이용되고 있다. 특히 다차원 속성 데이터 표현과 축소에 유용하며, 고주파 데이터와 이상치 데이터 특성에도 적합하다[15].

이와 같이 현존하는 데이터 축소 기법은 데이터 특성과 응용 범위에 따라 서로 다른 장단점을 가지고 있다. 따라서 우리는 센서 네트워크의 일반적 구조, 데이터 모델을 분석하고 데이터 특성에 적합한 다차원 속성 축소 기법을 비교 분석 하였다. 분석 결과 동일한 축소 기법을 적용하는 것에 비해 데이터 특성에 따라 선택적으로 축소 기법을 적용하는 것이 에러 비율과 성능 측면에서 우수한 결과를 보였다.

1) DWT (Discrete Wavelet Transform), DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform)



(그림 1) 센서 네트워크의 일반적인 구조

### 3. WSN시스템 구조와 데이터 축소 문제

#### 3.1 계층/분산형 모델

센서 네트워크의 일반적인 모델은 그림 1과 같이 계층/분산형 클러스터링 스키마이며 센서 노드, 싱크 노드, 베이스 스테이션과 서버 노드로 구성된다[4]. WSN의 센서 노드는 온도, 습도와 같이 하나의 값을 수집하는 센서 노드에서 하나의 센서 노드에 다양한 센서를 포함한 다차원 센서 노드로 분류한다. (그림 1)의 각 구성 요소는 다음과 같은 특징을 가진다.

첫째, 센서 노드는 다차원 속성 값을 수집하며, 주기적으로 다른 센서 또는 목적 노드에서 센서 노드로 측정된 값을 전송한다. 데이터 전송은 직접 통신 방식이 아닌 멀티 홉 또는 클러스터 기반의 데이터 중심 통신 방식을 취한다.

둘째, 각 센서는 소형 프로세서와 메모리 기반의 저장소를 가지며, 주기적으로 무선 근거리 데이터 통신을 한다. 싱크 노드는 일정 범위의 센서 노드에서 데이터를 수집하며, 메인 메모리 DB를 가진다.

셋째, 싱크 노드와 베이스 스테이션 사이는 무선 통신이며, 싱크 노드는 수집된 센서 데이터를 베이스 스테이션과 서버에 분석 또는 저장할 필요가 있는 데이터에 대해 원본 데이터 또는 요약 정보를 전송한다.

넷째, 서버 노드와 베이스 스테이션은 기존의 인프라 네트워크를 사용하며, 사용자 또는 애플리케이션에서 이력 정보에 대한 질의 처리를 위해 대용량 데이터를 저장한다. 따라서 일반적으로 서버 노드의 데이터 저장소는 데이터웨어하우스의 다차원 데이터 큐브 형태를 가진다.

#### 3.2 데이터와 질의 모델

센서 네트워크의 단말 노드는 센서 노드(S)로 구성되며, 각 센서 (x)는 온도, 습도와 같은 값(v)을 가진다. 만약 N개

의 센서 노드가 M개의 센서를 가질 때  $S_n = \{(x_1, v_1), \dots, (x_m, v_m)\}$ 이다. 각 센서가 동일한 기능을 하고 일정한 주기로 센서들에서 센서 노드로 데이터를 푸쉬 방식으로 전송한다. 따라서 센서 노드는 큐(queue) 형태의 메모리 버퍼와 소형의 프로세서를 가지며 고정 또는 이동 특성을 가진다. 또한 배터리 기반의 전력에 의존하며, 무선 통신이므로 주기적으로 싱크 노드에 데이터 전송시 데이터 손실이 발생할 수 있다.

싱크 노드(SN)가 N개의 센서 노드로 구성되면  $SN = \{S_1, \dots, S_n\}$ 이고, 메인 메모리 기반의 센서 DBMS을 포함한다. 싱크 노드는 센서에서 수집된 다차원 속성의 데이터를 요약하여 일정기간 저장하거나 질의의 요청에 따라 데이터를 전송한다. WSN의 제약 사항을 고려할 때 데이터 수집단계에서 모든 노드에서 정확한 데이터를 수집하는 것이 불가능하며, 실시간 처리해야 하기 때문에 데이터 요약이나 근사적 데이터 처리를 하는 것이 바람직하다[10].

데이터를 요약하는 방법은 일반적으로 집계 기법을 이용하며 데이터 전송은 다차원 속성에 적합한 축소 기법을 적용한다. 베이스 스테이션과 서버 사이는 기존의 네트워크를 사용하므로 대용량 스트림 데이터를 처리하는데 있어 시스템과 네트워크의 제약성이 크지 않다. 따라서 데이터 처리 및 전송 비용 보다는 다양한 싱크 노드에서 수집된 정보에 대해 사용자의 요구에 따라 빠르게 반응하기 위해 효율적으로 요약 저장하는 기법이 중요하다[25].

센서 노드가 데이터를 수집하거나 데이터의 입력 조건 또는 사건 기반의 데이터 처리시 질의 언어가 중요한 역할을 한다. 질의 언어는 연속 질의(continuous query), 사건 기반 질의(event-based query), 집계(aggregation), 일회성 질의(one-time query)가 있다. 데이터 수집 방식도 샘플링 기반, 통계적 모델 기반과 예측 모델이 있다[1, 2, 10]. 이러한 질의 처리와 수집 방식은 서버에서 센서 노드까지 위치하며,

사용자 또는 애플리케이션에서 질의가 수행되기도 한다. 따라서 질의 및 수집 모델은 애플리케이션 요구에 따라 선택되어야 한다.

대용량 스트림 데이터를 전송하기 위한 방법으로 데이터 축소에 관한 연구가 수행되었다. 대표적인 연구는 집계, 웨이블릿, 히스토그램, 클러스터링, 샘플링, 회귀분석 등의 기법이다. 이러한 축소 기법은 근사적 데이터 표현과 데이터 크기를 축소하여 데이터의 저장공간의 절약과 데이터 전송 비용을 줄이는 장점이 있다. 따라서 WSN의 애플리케이션 요구와 사용자 질의에 따라 데이터 전송과 저장 기법을 적용하는 것이 바람직하다.

### 3.3 다차원 스트림 데이터 축소

(그림 2)는 다차원 스트림 데이터 축소 과정을 표현한다. 먼저 센서 노드에서 싱크 노드로 주기적으로 수집되는 측정값이 메모리 버퍼에 행렬구조로 저장된다. 이때 메모리 크기와 축소기법의 효율성을 위해 최적화된 슬라이딩 윈도우 크기로 분할이 요구된다. 이 논문에서는 데이터 튜닝에 의해 최적의 슬라이딩 윈도우를 설정했다고 가정하며, 각 윈도우는 동일한 크기로 고정되었다고 간주한다.

센서 또는 싱크 노드에서 수집되는 센서 데이터는 다차원 속성들(m)과 일정한 시간간격 동안에 각 시점에 속성의 순서화된 값(n)으로 변형되어 m x n 행렬 구조로 메모리 버퍼에 저장된다. (그림 2)의 송신자는 윈도우 분할 정책에 따라 s[1], ..., s[n] 단위로 분할되며, 수신자는 t[1], ..., t[n]으로 데이터를 주기적으로 수신한다. 이때, 송/수신자는 무손실 원본 데이터를 수신하기 위해 제한된 통신 대역폭과 소형 디바이스의 적은 전력과 컴퓨팅 능력의 한계로 송신자의 데이터 축소와 수신자의 근사적 데이터 수신 및 복원 기술이 필수적으로 요구된다.

(그림 2)에서 입력 스트림 데이터는 윈도우 분할 정책에 따라 S={s1,s2,...,sn}로 분할되고, 하나의 윈도우는 고정된

윈도우 크기 m x n 행렬 구조를 가진다. 각 윈도우는 축소 기법이 적용된 후 T={t1,t2,...,tn}로 대응된다. 대응되는 데이터 값은 손실 축소가 수행되며, 단계적 데이터 복원기법으로 근사치 값을 갖는다. 즉 모든 분할 값에 대해 s[i] ≅ t[i]이고 원본 데이터와 전송 데이터는  $\sum_{i=1}^n s[i] \approx \sum_{i=1}^n t[i]$ 이다.

## 4. WSN 애플리케이션과 다차원 스트림 데이터 축소 기법

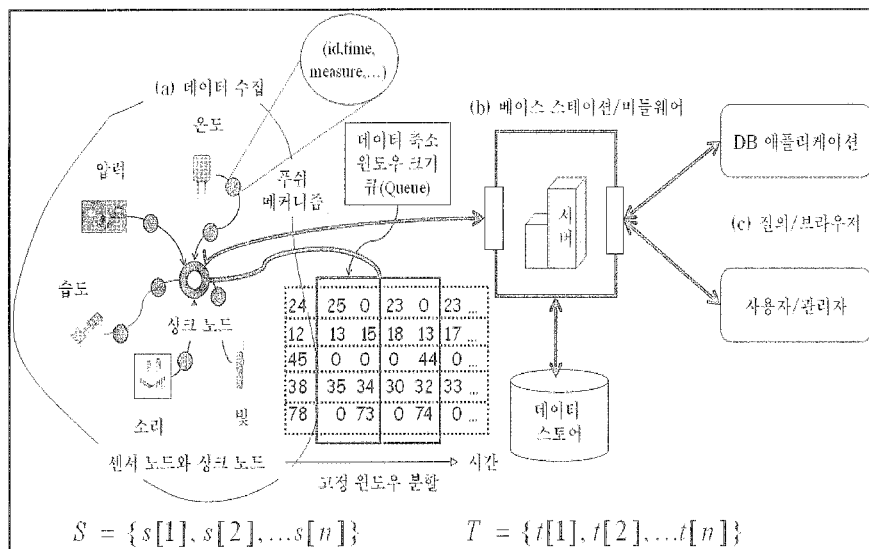
이 장은 센서 네트워크의 애플리케이션 특징을 분석하고 데이터 특성에 따른 질의 및 데이터 전송 방법들을 분류한다. 그리고 다변량 데이터 축소 기법을 적용할 WSN의 시뮬레이션 모델을 제시하고, 적용한 다변량 축소 기법들의 특성과 장단점을 분석한다.

### 4.1 센서 네트워크 애플리케이션 분류

#### 4.1.1 환경 모니터링

응용 분야는 동·식물 생태, 빌딩 모니터링, 홍수 탐지, 토양·습기 모니터링과 대기/해양 기후 모니터링 분야 등이다 [6, 11]. 모니터링 목적은 장기간 데이터 상태 변화를 분석하는 것이며, 데이터를 분석하여 미래를 예측하거나 계획을 세우는데 활용한다. 센서 모델은 지역적으로 고정된 형태로 분포되며, 급격한 데이터 변화가 적기 때문에 배터리를 절약하기 위해 일시적 휴면 상태에 있다가 주기적 또는 비정상적인 값이 측정되면, 수집 주기가 빈발하며 정밀하게 데이터를 수집하는 방식을 가진다.

데이터 타입은 저주파 특성이며 급격한 데이터 변화가 적은 정현파의 특성을 가진다. WSN에서 데이터 전송은 최적화 라우팅 스케줄에 따라 주기적으로 전송하며, 싱크 노드에서 데이터 정제, 집계, 요약하여 서버에 전송한다. 질의 평



(그림 2) 센서 기반의 다변량 스트림 데이터 축소

가는 센서 노드에서 주로 이루어지며, 연속 질의에 “얼마나 자주 또는 얼마 동안”에 대한 정보를 포함하는 질의와 사건 기반 질의가 사용된다.

4.1.2. 객체 추적

응용 분야는 이동 객체(차량, 동물) 추적과 감시, 군사 분야(전장 감시와 전쟁피해 평가), 물류(화물, 차량)분야 등이 다[4, 7]. 모니터링 목적은 객체의 이력 정보를 추적하거나 실시간 이동 객체 감시에 있다. 센서 모델에서 센서 또는 목적 노드가 이동 형태(예: 센서 부착 차량, 군수품 등)이며, 항상 또는 빈번하게 활성화 상태에서 데이터를 수집하여 배터리 소모가 크다.

데이터 타입은 이동 객체이므로 수집 주기가 빈번하며 변화가 크기 때문에 고주파 특성을 가지며 객체와 속성이 매우 복잡한 다차원 속성을 가진다. 많은 양의 스트림 데이터가 수집되고 실시간 데이터를 분석하여 서비스되기 위해 정확한 데이터 분석보다 근사적 데이터 표현 기법을 활용함으로써 WSN의 제약성을 해결한다. 질의 모델은 사건기반 모델, 연속 질의와 데이터의 분포를 고려한 확률과 신뢰도를 고려한 모델 기반(model-based) 질의가 적합하다. 데이터 전송과 저장 측면에서 다차원 속성 데이터의 시간 복잡도가 적은 축소 기법을 적용하는 것이 바람직하다.

4.1.3. 객체 감시

응용 분야는 응급 의료 관리, 침입 탐지, 지진 위험 모니터링, 산림 방재 감시 등이다[4, 11]. 모니터링 목적은 긴급한 환경 또는 이상치 같은 비정상적인 데이터 변화를 분석하는 것으로 이력 데이터와 모니터링 데이터를 함께 분석한다. 예를 들어 현재의 대기 상태와 과거의 이력 데이터를 비교하여 현재의 상태가 어떠한 상태인가를 판단해야 하기 때문이다.

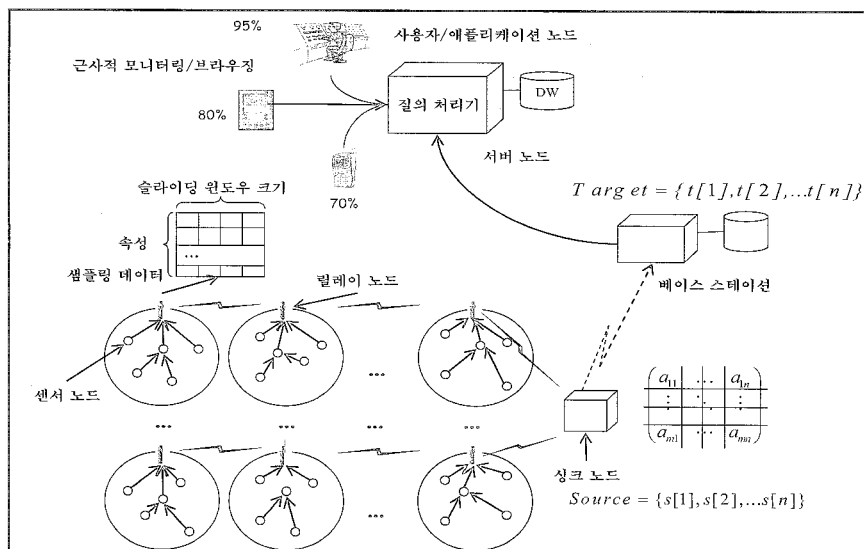
데이터 타입은 저주파, 고주파, 이상치 등 모든 형태가 가능하며, 특정 사건이 발생시 데이터는 비정상적인 변화를

보인다. 객체 감시에서 수집된 데이터는 매우 중요한 속성을 가지므로 근사적 데이터 접근이 아닌 정확한 데이터 전송이 실시간으로 요구된다. 데이터는 다변량이며, 다변량 데이터를 손실 없이 빠르게 전송하는 기법을 적용해야 한다. 질의 모델은 센서 노드에서 직접 데이터를 평가하고 질의를 수행하는 사건기반 모델과 연속 질의가 적합하다.

4.2 시뮬레이션 모델

이 논문에서 제안하는 스트림 데이터의 축소 전송 시뮬레이션 모델은 3장에서 제시한 WSN 구조와 데이터 모델을 따르며, 응용 분야는 4.1절의 3가지 애플리케이션에 공통으로 적용된다. 제안 모델은 환경, 물류, 교통 및 군사 등의 분야에 다차원 속성을 가진 객체를 실시간 모니터링 하거나 근사적 데이터 질의 처리에 적용 가능하다. 시뮬레이션 모델은 그림 3과 같고 사용자 관점에서 센서 노드까지 각 단계는 다음과 같다.

- ① 사용자 애플리케이션 노드 단계: 사용자와 애플리케이션은 센서 노드에서 서버까지 각 레벨에서 데이터를 수집하거나 정보를 얻기 위해 질의를 사용한다. 이 질의는 사전에 정의되어 연속적으로 수행되거나 이력정보와 조인된 결과를 반환한다. 센서 노드에는 연속 질의 또는 일회성 질의를 이용하여 센싱 데이터를 수집하거나 현재의 스냅샷 값을 얻는다. 계층형 네트워크 내부는 요약 및 통계 정보, 서버에는 이력정보와 지식 탐사된 정보에 대한 질의를 요구한다. 요구된 질의에 대한 데이터의 정확도는 응용 애플리케이션과 디바이스의 성능에 따라 근사적으로 데이터를 표현한다. 이는 실제 센서 데이터가 부정확하며, 모든 데이터를 수신하거나 정확하게 분석하는 것이 WSN의 특성과 실시간 데이터 분석 요구에 적합하지 않기 때문이다. 따라서 WSN의 응용 분야와 데이터 특성에 맞는 데이터를 수집하고 전송하는 모델을 적용한다.
- ② 센서·릴레이 노드 단계:사용자에 의해 사전에 정의된 질의 또는 규칙에 따라 센서 노드는 다차원 속성 데이터를



(그림 3) WSN의 스트림 데이터 축소 전송 시뮬레이션 모델

주기적으로 수집한다. 수집된 데이터는 매번 데이터를 전송하는 것이 아니라 일정한 윈도우 시간 간격 동안 데이터를 수집하여 처리한다. 하나의 릴레이 노드는 여러 센서 노드를 가지고 있으며, 센서 노드에서 수집된 다차원 속성 정보를 통합한다. 센서 노드와 릴레이 노드는 적은 메모리와 프로세스를 가지며 배터리 전력이 의존하여 수집된 데이터의 원본 또는 축소 데이터를 주기적으로 싱크 노드에 전송한다. 데이터 저장 형태는 속성을 위한 행과 윈도우 크기 범위에서 값을 저장하는 열로 구성된 행렬 구조이다.

③ **싱크 노드 단계**: 싱크 노드는 릴레이 노드에서 전송하는 단위 데이터를 통합하며 메인 메모리 기반의 DBMS를 가지고 베이스 스테이션과 무선 통신을 한다. 각 릴레이 노드에서 수집되는 다변량 데이터가 대용량이므로 데이터를 처리하거나 전송하는데 샘플링과 다변량 축소 기법을 적용하는 것이 바람직하다.

④ **베이스 스테이션 단계**: 이 노드는 다양한 싱크 노드에서 수집한 데이터를 요약 저장하며, 컴퓨팅 능력이 크며, 기존의 네트워크 방식을 사용한다. 특히 상용 DBMS를 이용하여 대용량 스트림 데이터의 요약 정보 저장이 가능하다. 이 단계는 여러 개의 싱크 노드에서 수집된 데이터를 처리해야 하기 때문에 미들웨어 기능을 수행한다.

⑤ **서버 노드 단계**: 센서 네트워크 응용 분야의 대부분은 실시간 데이터 분석을 요구하지만 센서 노드에서 수집된 정보와 과거 이력 정보를 비교 분석하는 요구사항이 있다. 이를 위해 서버 노드에서는 베이스 스테이션에서 보내온 대용량 데이터를 저장해야 하며, 저장 형태는 데이터웨어하우스의 다차원 데이터 큐브 형태로 구성된다.

### 4.3 다차원 데이터 축소 기법

이 절에서는 4.2절에서 기술된 시뮬레이션 모델을 기반으로 이 논문의 실험에 사용된 다변량 데이터 축소 기법의 고유한 특성과 방법론을 비교한다.

#### 4.3.1 웨이블릿

웨이블릿(DWT)은 선형적 신호 프로세스 기술에 적용하며, 계층형 분해 함수를 사용하는 수학적 도구이다. 웨이블릿은 이산 푸리에 변환(DFT)과 유사하며 사인, 코사인과 같은 정현파 형식의 저주파 데이터 축소에 적합하다. 따라서 입력 신호가 급격히 변하거나 비정상적인 변화를 가지는 신호에는 에러가 증가하는 단점이 있다[12, 13]. 웨이블릿의 일반적인 방법은 각 데이터를 반복적으로 이등분하는 피라미드 방식의 계층 알고리즘이며 빠른 계산 속도를 가진다. 웨이블릿의 장점은 순차 데이터에 대해 빠른 계산 속도와 작은 공간 복잡도이다. 또한 축소 계수 값을 이용하여 단계별 최적화 근사값을 가진다. 복잡도는  $n$ 개의 벡터에 대해  $O(n)$ 이다[12]. 웨이블릿 기법은 많은 연구가[8, 12]에 의해 개선되었지만 다차원 속성의 경우 시간 및 공간 복잡도 측면에서 비효율적이다. 최근 웨이블릿 기법은 [19]에서 최적의 계수를 선택하는 알고리즘이 제안되어 성능이 많이 개선되었다.

#### 4.3.2 계층형 클러스터링(HCL)

클러스터링은 객체들의 유사성과 비유사성 정도를 판별하여 그룹을 구성하는 것이다[13, 18]. 데이터 축소 관점에서 클러스터링은 실제 데이터를 유사한 다른 값으로 대체하는 개념이다. 일반적인 k-means 알고리즘은 수행시 많은 공간이 요구되며, 입력 데이터에 랜덤 접근 방식을 가진다. 스트림 데이터에 대한 클러스터링 연구는 [18]에 의해 연구되었고 선형 공간과 단일 패스로 수행 가능하며 복잡도는  $\tilde{O}(nk)$ 이다. 다차원 속성에 대한 데이터 셋을 반복적으로 병합하는 계층형 클러스터링 방법을 이용한다. 이 기법은 클러스터 생성시 그리드(grid), 하향식(top-down), 상향식(bottom-up) 방식을 이용하여, 단일, 평균, 완전-연결 방법을 적용하여 계층 클러스터를 생성한다. 또한 클러스터링은 다차원 인덱스 트리 형태로 구성하여 요약 데이터를 분석 할 수 있어 질의에 대한 빠른 근사적 결과를 얻는 장점이 있다[12]. TinyDB[20]는 센서 네트워크에서 특정 지역에 산재된 센서에서 수집된 데이터는 동일한 지역에서 수집됨으로 센서 데이터가 유사한 속성을 가지기 때문에 일정한 클러스터 단위로 데이터의 유사성을 검증하는 집계 기법을 센서 네트워크 내부 단계에서 적용한다. 하지만 이 기법은 클러스터의 개수를 사전에 정의하거나 이상치가 많은 데이터 분포를 가질 때 낮은 성능을 가질 수 있다.

#### 4.3.3 샘플링

샘플링(Sampling)은 대용량 데이터 집합에서 더 작은 샘플 데이터로 표현하는 것으로 데이터를 축소 표현한다[12, 13]. 샘플링 방법에 따라 SRSWOR(Simple Random Sample Without Replacement)과 SRSWR(Simple Random Sample With Replacement)로 구분한다. 샘플링 방법은 고정된 크기만큼 샘플을 유지하는 축적 샘플링, 서로 다른 데이터 속성 집합에서 보다 적은 속성을 얻는 클러스터 샘플과 데이터의 크기에 비례해서 비율을 다르게 하는 층화 샘플링이 있다. 샘플링의 복잡도는 데이터 사이즈에 대해 일반적으로 비례하며, 샘플링의 비율을 조절하여 데이터의 에러율을 손쉽게 조절이 가능한 장점이 있다. 하지만 임의의 스키마에 대한 조인 연산으로 데이터의 일관성 성질을 잃을 수 있으며 집합 기반의 근사적 질의에 대해서 비효율적이다[13].

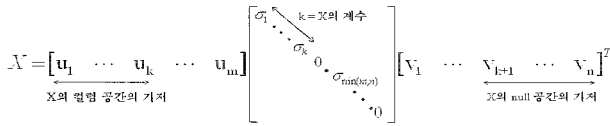
#### 4.3.4 특이값 분해 분석

특이값 분해 분석(SVD)은 행렬 분해를 이용하여 데이터 축소와 근사적 데이터 접근 방안을 제시한다[14, 15]. 다음은 SVD의 정의, 특징과 최적화 근사적 표현에 대해 기술한다. [정의 1] SVD:  $m \times n$ 행렬  $X$ 에 대해서  $X$ 를  $X=U\Sigma V^T$ 로 분해할 수 있다. 이때 정방행렬  $U$ ,  $V$ 와 대각행렬  $\Sigma$ 는 아래의 성질을 가진다[14].

$$U_{m \times m} = UU^T = U^T U = I, \quad V_{n \times n} = VV^T = V^T V = I \quad (4.1)$$

$$\Sigma_{m \times n} = [\Sigma]_{ij} = 0, \quad i \neq j, \quad [\Sigma]_{ii} = \sigma_i \geq 0, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min\{m,n\}} \quad (4.2)$$

U, V는 정방행렬이며, 직교행렬이므로  $UT=U1$ 이고 식 (4.1), (4.2)와 같이 유니터라 행렬이 된다.  $\Sigma$ 가 대각 행렬이고 성분  $\{\sigma_i\}$ 을 특이값(singular value)이라 하며,  $\{\sigma_i\}$ 은 대칭행렬  $XX^T$ 의 고유 값이 되며 이들은 유일하게 정해진다. 계수 k는 X에서 0이 되지 않는 특이 값의 개수를 말하며  $X=U\Sigma V^T$ 는 (그림 4)와 같이 표현된다.



(그림 4) SVD의 열 공간, 계수, 영공간 표현

공간 복잡도 측면에서 원래의 행렬 X는  $N \times M$  데이터 원소로 구성된다. SVD 표현에 따라 k개의 주요 성분으로 데이터를 절단하면 행렬 U에 대해  $N \times k$  데이터 원소가 필요하며, V에 대해  $k \times M$  원소이다. 따라서 SVD의 적용 전과 후의 공간 비율 s\_ratio는 식(4.3)과 같다[15].

$$s\_ratio = \frac{N \times k + k + k \times M}{N \times M} \approx \frac{k}{M} \quad (N \gg M \geq k) \quad (4.3)$$

데이터 전송 측면에서 SVD의 또 다른 장점은 최적화 k 근사적 데이터 표현이다. (그림 3)의 싱크 노드는 슬라이딩 윈도우 크기 분할에 따라  $s[1], \dots, s[n]$  단위로 분할되며 베이스 스테이션은  $t[1], \dots, t[n]$ 으로 원본 데이터를 주기적으로 수신한다. 근사적 데이터 전송 기법은 모든 데이터를 송신하는 것이 아니라 근사적으로 데이터를 축소하여 데이터를 송신한다. 따라서 수신 디바이스는 데이터의 근사적 표현으로 단점은 있지만 소형 프로세서와 메모리를 이용하여 복원이 가능하다.

**[정의 2]** 최적화 k 근사치: 주어진  $m \times n$ 행렬 X에 대해 특이값 k는  $k=\min(m,n)$ 이고, 근사적 계수(r)는  $r < k$ 일 때  $X' = u_1v_1^T + u_2v_2^T + \dots + u_rv_r^T$ 이면  $\min \|X - X'\|_2 = \sigma_{r+1}$ 이다[14].

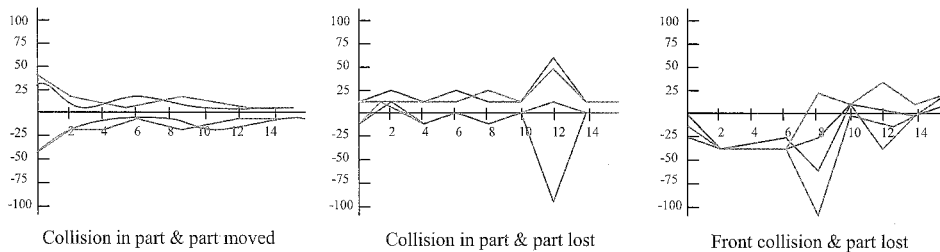
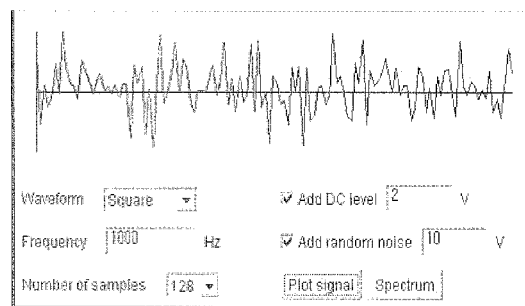
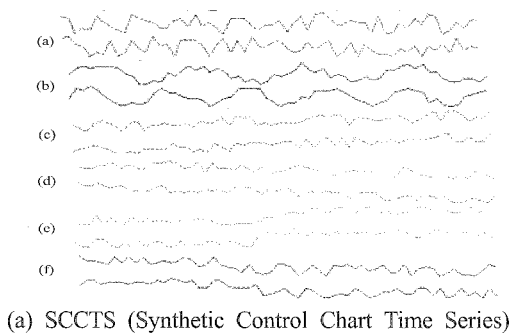
원래의 행렬 X가  $N \times M$  데이터 원소로 구성되었을 때,  $XX^T = u_1\sigma_1v_1^T + u_2\sigma_2v_2^T + \dots + u_rv_r^T$ 이 단계별로 전송되면, 실제 수신 측에서 계산 비용 복잡도는 두 개의 벡터  $u_i$ 와  $v_i$ 와  $\sigma_i$  스칼라의 곱 연산의 합이다. 따라서 SVD의 적용 전과 후의 계산 비율 c\_ratio는 (4.4)와 같다.

$$c\_ratio = \frac{r(M + N + 1)}{N \times M} \quad (N \gg M \geq k) \quad (4.4)$$

예를 들어 특이값이  $5.3 \geq 2.6 \geq 1.2 \geq 0.7 \geq 0$ 으로 계수가 4인 대각 행렬을 이룰 때  $\|X - X'\|_2 = \sigma_5 = 2.6$ 이다. 즉 각 단계는 특이 값만큼 차이가 나며, 단계별로 차이는 점점 작아진다. 데이터 전송 시 패킷(packet) 단위로 2개의 벡터(u,v)와 1개의 스칼라 값( $\sigma$ )으로 분할 전송하고, 수신 측에서 2개의 벡터와 스칼라 곱을 이용하여 복원한다. 따라서 SVD의 근사적 기법은 수신 측의 시스템 성능 및 시간적 요구에 따라 데이터를 최적화된 근사 값을 수신 받을 수 있는 장점이 있다.

### 5. 실험 평가

우리는 (그림 5)와 같이 다변량 시계열 데이터, 랜덤 웨이브 생성기와 로봇 예러 행위 데이터를 이용했다[21, 24]. 이 논문의 실험은 WSN 환경의 애플리케이션에 적용하기 위해 다양한 데이터 셋을 이용하여 다변량 데이터 축소 기법들을 평가하고 비교하는 것이다.



(그림 5) 다변량 시계열 데이터 셋

5.1 데이터 셋

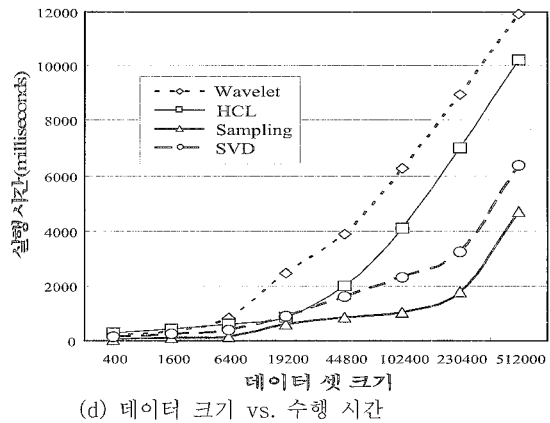
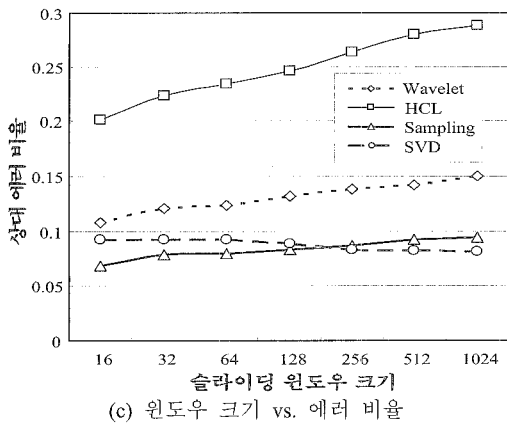
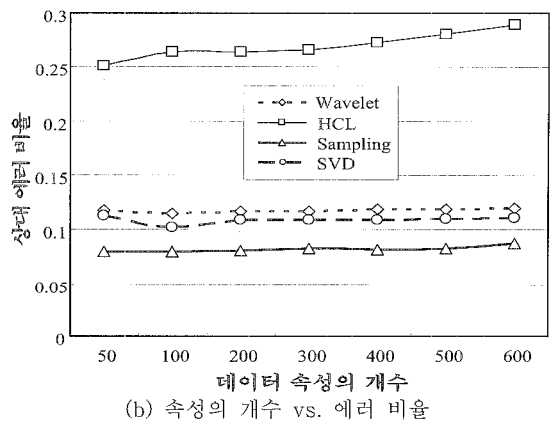
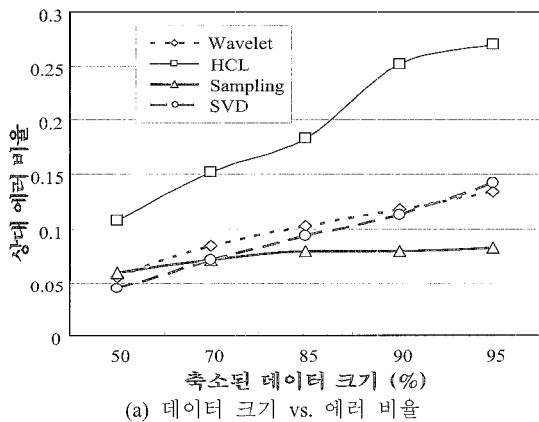
실험 결과는 [21, 24]에서 제공된 3개의 데이터 셋을 수행한 결과를 기반으로 한다. 첫째, (그림 5)(a)는 SCCTS로 Alcock와 Manolopoulos의 프로세스에 의해 인위적으로 생성된 시계열 차트의 600개 예제를 포함한다. 이 데이터는 6개(Normal(a), Cyclic(b), Increasing trend(c), Decreasing trend(d), Upward shift(e), Downward shift(f))의 다른 클래스로 구성된다. 데이터 축소 기법의 성능을 평가하기 위해 우리는 SSCTS를 같은 크기 비율로 구성하였다. 둘째, 랜덤 웨이브 생성기[24]를 이용하여 5 그룹의 연속된 다차원 속성 데이터를 생성했다. (그림 5)(b)와 같이 웨이브 타입(예: 사인, 코사인, 사각형, 톱니형태), 빈도(Hz), DC 레벨과 랜덤 잡음의 매개변수를 이용하여 생성하였다. 셋째, 로봇 행위 분류 데이터로 로봇이 작동 중 외부의 간섭에 의해 에러가 발생한 데이터이며 로봇의 Force/Torque에 의해 측정된 데이터이다 [17, 21]. 각 실패 데이터는 실행 중 실패가 감지된 시점에서 일정한 시간 간격 동안 수집된 15개 Force/Torque 샘플에 의해 특성화 되었다. 데이터 셋은 5개로 구성되며, 이들 각각은 LP1에서 LP5까지 다른 로봇 행위를 분류하고 있다.

데이터 크기와 타입에 대한 상대 비율은 원본 데이터가 행과 열 구조의 행렬  $A$ 이고 축소된 데이터가 행렬  $\hat{A}$  라 가정할 때 상대 에러( $\sigma$ ) 측정치는 (5.1)의 식을 따른다.

$$\sigma = \frac{\|\hat{A} - A\|_F}{\|A\|_F}, \text{ where } \|A\|_F = \left( \sum_{ij} |a_{ij}|^2 \right)^{1/2} \quad (5.1)$$

행렬 구조 데이터에 대해 상대 에러를 비교할 때, 각 행에 대한 열 데이터에 대한 축소 데이터를 복원하는 방법은 다음과 같다. 웨이블릿은 각 행에 대해 각 단계마다 계수 값을 이용하여 복원하였다. 예를 들어 평균값이 8, 4이며 계수가 1, 1일 때 복원된 값은 9, 7, 3, 5로 복원한다. HCL 기법은 각 단계에서 각 행벡터 사이에 최소 편차를 가진 행벡터의 값으로 대치하는 방식으로 데이터를 복원하였다. 샘플링은 시작점에서 시작하여 순서화 했으며 샘플링 되지 않은 시점 값은 시작점과 끝점 사이에 보간법(interpolation)으로 값을 대치하였다. 마지막으로 SVD는 생성된 특이 값의 개수를 복원 비율로 적용하는 방법을 이용하여 두 개의 벡터와 특이 값의 곱을 이용하여 복원하였다.

비교 실험을 위해 사용된 데이터 축소 알고리즘은 [22, 23]에서 제공된 표준화된 알고리즘과 자바 패키지를 이용하였으며 실험 평가를 위해 입력과 알고리즘을 일부 수정하였다. 이 논문에서 적용한 데이터 축소 기법은 기본적인 특성에 기반한 알고리즘을 이용하였으며 각 기법의 최적화된 축소 알고리즘을 적용하여 평가하지 않았다.



(그림 6) 데이터 크기에 대한 축소 기법의 성능 평가



5.2 SSCTS 데이터 (데이터 크기와 성능평가)

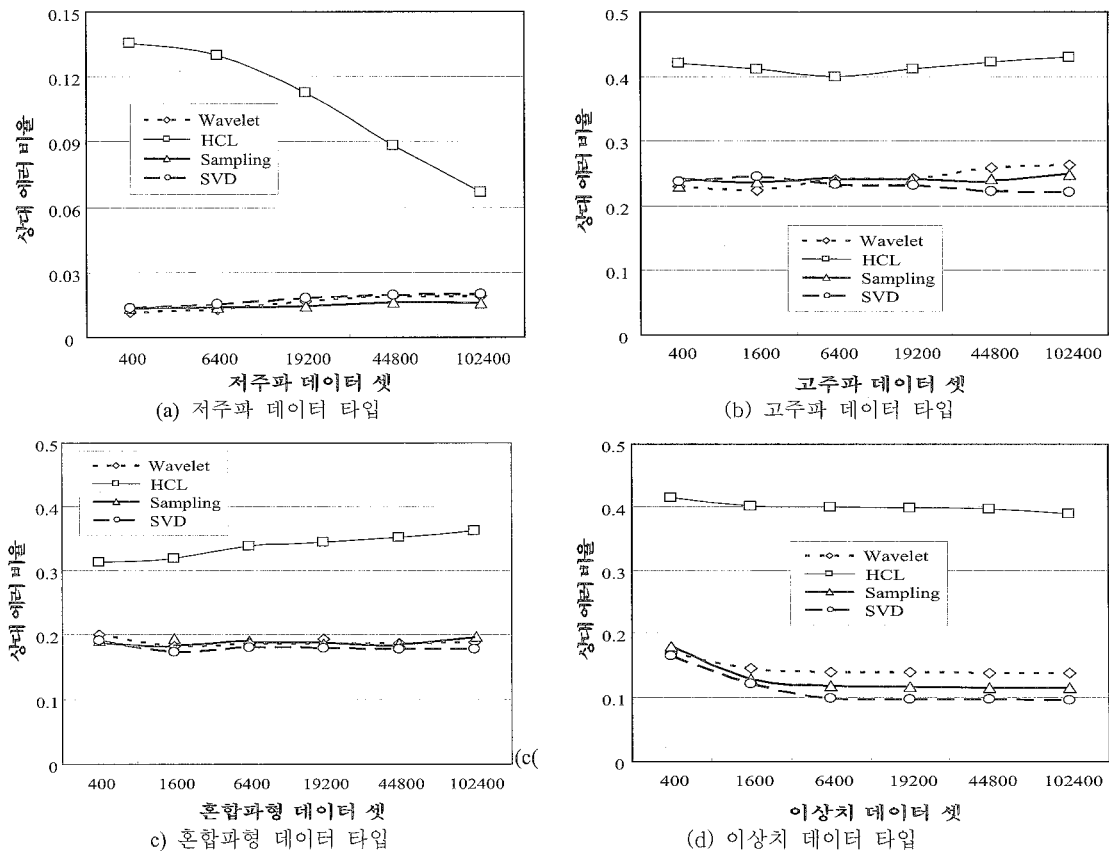
(그림 6)은 SSCTS 데이터 셋에 대해 상이한 데이터 크기에 대해 상대 에러와 수행 시간을 평가하였다. (그림 6) (a)는 축소 비율을 50%에서 95%까지 변화시키면서 각 축소 기법의 상대 에러 비율을 비교하였다. HCL은 샘플링, SVD와 웨이블릿에 비해 상대적으로 에러 비율이 높다. 샘플링 기법의 상대 에러를 계산하기 위해 샘플링 되지 않은 값에 대해 다양한 대처 방법을 시도했다. 실험 결과 샘플링 되지 않은 값에 대해 보간법으로 값을 대처하는 방법이 가장 우수한 결과를 보였다.

(그림 6)(b)와 (그림 6)(c)는 축소 비율이 90%일 때 상대 에러를 비교하였다. (그림 6)(b)는 데이터 속성 개수에 따른 축소 비율을 보인다. 실험 결과 모든 축소 기법은 데이터 속성의 크기에 크게 영향을 받지 않았다. (그림 6)(c)는 슬라이딩 윈도우 크기가 성능에 얼마나 영향을 주는지 평가하였다. 실험 결과 HCL과 웨이블릿은 윈도우 크기가 증가하면서 선형적으로 상대 에러가 증가하였다. 에러의 증가 이유로 웨이블릿은 단계별 축소 시 두 값의 합의 평균을 이용하며, HCL은 가장 유사한 행 벡터로 대처되기 때문에 윈도우 크기가 증가하면서 에러 값이 선형적으로 증가함을 보였다. (그림 6)(d)는 데이터 크기를 증가시키면서 각 기법에 대해 실행 시간을 평가하였다. HCL과 웨이블릿은 반복적 알고리즘 특성 때문에 데이터 크기가 증가하면서 수행 시간이 상대적으로 높게 나타났다.

(그림 6)에 따르면, 샘플링 기법이 SSCTS 데이터의 6개 다른 그룹에 대해 우수한 결과를 보였다. 웨이블릿과 HCL은 윈도우 크기가 증가하면서 에러 비율의 증가와 변수의 개수가 증가하면서 높은 수행 시간을 보였다.

5.3 랜덤 웨이브 생성기 데이터 (데이터 타입과 상대 에러 비율)

(그림 7)은 센서 네트워크 애플리케이션의 다양한 데이터 타입에 따른 데이터 축소 기법을 비교하였다. 랜덤 웨이브 생성 데이터는 (그림 5)(b)와 같다. 객체 추적과 감시 응용 데이터를 생성하기 위해 임의로 이상치를 생성해서 추가하였다. 실험에서 데이터 축소 비율은 80%로 하였으며 윈도우 크기와 속성의 수를 함께 증가시켰다. (그림 7)(a)는 사인과 코사인 파형과 같은 저주파 데이터에 대한 결과이다. 실험 결과 HCL을 제외한 축소 기법은 상대 에러 비율이 작았다. (그림 7)(b)는 고주파 데이터에 적용한 결과이다. HCL과 웨이블릿은 샘플링과 SVD에 비해 상대 에러가 높다. (그림 7)(c)는 고주파와 저주파를 3:2의 비율로 혼합한 데이터에 대한 실험 결과이다. (그림 7)(d)는 비정상적인 데이터와 이상치를 포함한 데이터에 대한 실험 결과이다. 데이터에 비정상적인 상승과 급격한 변화 값을 포함하기 때문에 웨이블릿과 샘플링은 높은 에러 값을 보였다. 이 논문의 실험에서 사용한 데이터는 HCL 기법이 상대적으로 높은 에러 값을 보이지만, 실제 센서 네트워크에서는 공간 지역성 특성 때문에 동일한

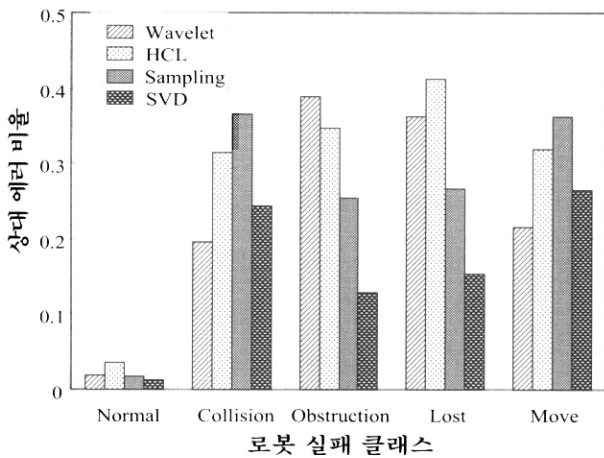


(그림 7) 데이터 타입에 따른 상대 에러 비율

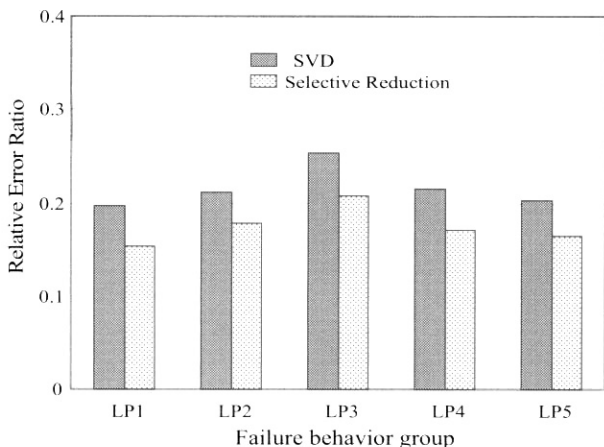
클러스터내의 센서 값에 대한 데이터 집계 및 축소에서 (그림 7)보다 좋은 결과가 예상된다.

5.4 로봇 실행 실패 데이터 (SVD와 선택적 축소 기법 비교)

(그림 8)은 [17, 21]에서 제공한 로봇 실행 실패 데이터에 대해 다변량 데이터 축소 기법을 적용한 결과를 보여준다. (그림 8)(a)는 [21]에서 제공한 데이터 셋의 5개의 클래스 (Normal, Collision, Obstruction, Lost, Move)를 4개의 축소 기법을 이용하여 80%로 축소한 결과이다. 실험 결과 웨이블릿은 “Collision”과 “Move” 클래스에 대해 우수한 결과를 보였으며, SVD 기법은 나머지 클래스에 대해 우수한 결과를 보였다. [21]의 LP1에서 LP5에 대해 (그림 8)(b)은 모든 클래스에 SVD 축소 기법을 적용한 결과와 클래스의 데이터 특성에 따라 최적화된 축소 기법을 선택적으로 적용한 결과를 비교하였다. 실험 결과 선택적 축소 기법이 SVD 기법만을 적용한 결과에 비해 우수함을 알 수 있다. 따라서 WSN 애플리케이션에서 데이터 축소 기법을 적용할 때 데이터 특성에 고려하여 선택적으로 축소 기법을 적용해야 한다.



(a) 로봇 행위에 대한 에러 비율



(b) 선택적 데이터 축소 기법 적용

(그림 8) 로봇 실행 실패 행위의 상대 에러 비율

6. 결 론

이 논문에서는 무선 센서 네트워크의 계층형 네트워크 모델, 데이터 모델과 다변량 스트림 데이터 축소 전송 모델을 제시하였다. 제시된 모델을 기반으로 데이터 특성과 사용자 요구에 적합한 표준 다변량 데이터 축소 알고리즘을 비교 평가하였다.

실험 결과 다차원 데이터 속성이며 고주파 특성인 데이터에 대한 실험에서 샘플링과 SVD기법은 데이터 수집과 전송 측면에서 우수한 결과를 보였다. 또한 센서 노드 정보의 요약에는 HCL기법이 적합하지만 웨이블릿 기법과 함께 슬라이딩 윈도우 크기가 증가하거나, 데이터 특성이 고주파 또는 이상치 데이터 특성에서 에러 값이 증가하는 단점을 보였다. 샘플링과 SVD는 최적화 근사적 데이터 전송이 가능하며, 에러 비율과 공간 복잡도에서 기존의 표준화된 기법에 비해 우수함을 보였다. 또한 데이터 타입에 따라 축소 기법의 성능이 다르기 때문에 데이터 특성에 따라 선택적으로 축소 기법을 적용하는 것이 적합하다. 제안된 데이터 축소 모델은 다차원 센서-데이터가 수집되는 센서 네트워크를 디자인하고 구축하는 응용 분야에서 유용하게 활용될 것이다.

참 고 문 헌

- [1] J. M. Hellerstein, W. Hong, and S. R. Madden, "The Sensor Spectrum: Technology, Trends, and Requirements," In SIGMOD Record. Vol. 32, No.4, pp.22-27, 2003.
- [2] A. Deligiannakis, Y. Kotidis and N. Roussopoulos, "Compressing Historical Information in Sensor Networks," In Conf. of SIGMOD, pp.527-538, 2004.
- [3] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Hierarchical in-Network Data Aggregation with Quality Guarantees," In Conf. of EDBT, pp.658-675, 2004.
- [4] M. J. Franklin and S. R. Jeffery et al, "Design Considerations for High Fan-In Systems: The HiFi Approach," In Conf. of CIDR, pp.290-304, 2005.
- [5] A. Manjeshwar and D. P. Agrawal, "TEEN: A routing protocol for enhanced efficiency in wireless sensor networks," In Proc. of PDPS, pp.2009-2015, 2001.
- [6] A. Mainwaring and J. Polastre et al, "Wireless Sensor Networks for habitat monitoring," In Proc. of WSNA, pp.88-97, 2002.
- [7] B. X. and O. Wolfson, "Time-Series Prediction with Applications to Traffic and Moving Objects Databases," In Proc. of MobiDE, pp.56-60, 2003.

[8] S. Guha, C. Kim, and K. S. Shim, "XWAVE: Approximate Extended Wavelets for Stream Data," In Conf. of VLDB, pp.288-299, 2004.

[9] Y. Chen and G. Dong et al, "Multi-Dimensional Regression Analysis of Time-Series Data Streams," In Conf. of VLDB, pp.323-334, 2002.

[10] A. Deshpande and C. Guestrin et al, "Model-Driven Data Acquisition in Sensor Networks," In Conf. of VLDB, pp.588-599, 2004.

[11] R. C. Oliver and K. Smettem et al, "Field Testing a Wireless Sensor Network for Reactive Environmental Monitoring," In Proc. of ISSNIP, pp.7-12, 2004.

[12] J. Han and M. Kamber, 'Data Mining Concepts and Techniques,' Morgan Kaufmann Publishers, 2000.

[13] M. Garofalakis, and P. B. Gibbons, "Approximate Query Processing: Taming the Terabytes!" In Conf. of VLDB, Tutorial, 2001.

[14] G. Strang, 'Introduction to Linear Algebra,' 3rd Ed., Wellesley-Cambridge Press, 1998.

[15] F. Korn, H. V. Jagadish, and C. Faloutsos, "Efficient Supporting Ad Hoc Queries in Large Datasets of Time Sequences," In Conf. SIGMOD, pp.289-300, 1997.

[16] D. Barbara and W. DuMouchel, et al, "The New Jersey Data Reduction Report," IEEE Data Engineering Bulletin, pp.3-45, 1997.

[17] L. M. Camarinha-Matos, L. S. Lopes, and J. Barata, "Assembly Execution Supervision with Learning Capabilities," In Conf. of ICRA, pp.272-279, 1994.

[18] S. Guha and N. Mishra et al, "Clustering Data Streams," In Conf. of FOCS, pp.359-366, 2000.

[19] A. Deligiannakis, M. Garofalakis, and N. Roussopoulos, "A Fast Approximation Scheme for Probabilistic Wavelet Synopses," Int. Conf. on SSDBM, pp.243-252, 2005.

[20] S. R. Madden, M. J. Franklin, and J. M. Hellerstein, "TinyDB: An Acquisitional Query Processing System for Sensor Networks," In ACM TODS, pp.1-47, 2004.

[21] S. Hettich and S. D. Bay, "The UCI KDD Archive (Synthetic Control Chart Time Series, Robot Execution Failures) [http://kdd.ics.uci.edu], Irvine, CA : University of California, Department of Information and Computer Science, 1999.

[22] "JAMA," A Java Matrix Package, "http://math.nist.gov."

[23] "Multivariate Data Analysis Software," Java Source, "http://astro.ustrasbg.fr/~fmurtagh/mdasw/."

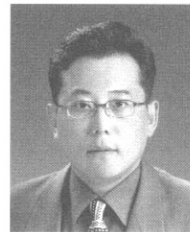
[24] "FFT Spectrum Analyzer," Java Source,

"http://www.dsptutor.freeuk.com/analyser/SA102.html."

[25] S.B. Seo, J.W. Kang, D.W.Lee, and K.H.Ryu, "Multivariate stream data classification using standard text classifiers," In Conf of DEXA, pp.420-429, 2006.

[26] S.B.Seo, J.W.Kang, and K. H. Ryu, "Multivariate Stream Data Reduction in Sensor Network Applications," EUC workshops, pp.198-207, 2005.

**정 훈 조**



e-mail : hjjung@hanseo.ac.kr  
 1984년 동국대학교 통계학과(이학사)  
 1987년 동국대학교 대학원 통계학과  
 (이학석사)  
 1993년 동국대학교 대학원 통계학과  
 (이학박사)

2002년 충북대학교 대학원 전산학과(박사과정수료)  
 1993년~현재 한서대학교 컴퓨터정보학과 교수  
 관심분야 : 데이터마이닝, 스트림데이터 처리, 시공간데이터  
 베이스, 전산통계

**서 성 보**



e-mail : sungboseo@yahoo.co.kr  
 1999년 서원대학교 전산학과(학사)  
 2001년 충북대학교 전산학과(석사)  
 2006년 충북대학교 전산학과(박사)  
 2001년~2003년 한국전자통신연구원  
 우정기술연구센터 연구원

2005년 North Carolina State University 방문 연구원  
 현재 North Carolina State University 박사후 연수  
 관심분야 : 시공간 데이터베이스, 시계열 데이터 분석, 스트림  
 데이터마이닝, USN 미들웨어 시스템

**최 경 주**



e-mail : kjcheoi@chungbuk.ac.kr  
 1996년 충북대학교 컴퓨터과학과  
 (공학사)  
 1999년 연세대학교 컴퓨터과학과  
 (공학석사)  
 2002년 연세대학교 컴퓨터과학·

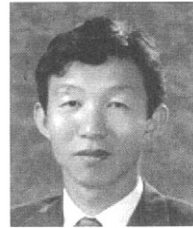
산업시스템공학과(공학박사)  
 2002년 7월~2005년 2월 LG CNS 연구개발센터  
 2005년 3월~현재 충북대학교 전기전자컴퓨터공학부 전임강사  
 관심분야 : 컴퓨터비전, 영상처리, 바이오컴퓨팅, 유비쿼터스컴퓨팅



**박 정 석**

e-mail : jspark@chungju.ac.kr  
1981년 숭실대학교 전자계산학과(공학사)  
1983년 숭실대학교 대학원  
전자계산학과(공학석사)  
2000년 충북대학교 대학원  
전자계산학과(이학박사)

1983년~1996년 한국원자력연구소(전임연구원)  
1996년~2006년 2월 청주과학대학 컴퓨터과학과 부교수  
2006년 3월~현재 충주대학교 전기전자 및 정보공학부 부교수  
관심분야: 이동 객체 데이터 관리, continuous query 처리,  
context database



**류 근 호**

e-mail : khryu@dblab.chungbuk.ac.kr  
1976년 숭실대학교 전산학과(이학사)  
1980년 연세대학교 대학원 전산전공  
(공학석사)  
1988년 연세대학교 대학원 전산전공  
(공학박사)

1976년~1986년 육군군수 지원사 전산실(ROTC 장교), 한국전자  
통신연구원(연구원), 한국방송대 전산학과(조교수)  
근무  
1989년~1991년 Univ. of Arizona Research Staff (TempIS  
연구원, Temporal DB)  
1986년~현 재 충북대학교 전기전자컴퓨터공학부 교수  
관심분야: 시간 데이터베이스, 시공간 데이터베이스, Temporal  
GIS, 유비쿼터스 컴퓨팅 및 스트림 데이터 처리,  
지식기반 정보검색 시스템, 데이터마이닝, 바이오  
인포메틱스