

# 해양 환경에서의 USN 미들웨어 기반 센서 데이터 마이닝<sup>†</sup>

김성호<sup>0</sup>, 김룡, 이준욱\*, 정재두\*\*, 류근호

충북대학교 데이터베이스 연구실

\*한국전자통신연구원

\*\*육군 본부

e-mail : {kimsungho<sup>0</sup>, kimlyong, khryu}@dblal.chungbuk.ac.kr

\*junux@etri.re.kr

\*\*chungjaedu@hanmail.net

## The Ocean Environment Sensor Data Mining based on USN Middleware

Sungho Kim<sup>0</sup>, Lyong Kim, JunWook Lee\*, Chungjae Du\*\*, Keunho Ryu

Database Laboratory, Chungbuk National University

\*Electronics and Telecommunications Research Institute

\*\*ROK

요 약

유비쿼터스는 간단히 말해서 많은 센서 들로 이루어진 무선 센서 네트워크이며 해양 환경 감시 서비스는 해양에 센서들을 설치함으로써 유비쿼터스 환경을 구축하고 해양 환경 변화를 감시한다. 센서 노드들로부터 수온, 기온, 염도 등을 센서 데이터들이 측정되 이 이러한 데이터를 기반으로 유용한 지식을 탐사해낸다. 그러나 기존의 데이터 마이닝 기법은 이력 데이터에 대해서 마이닝 기법을 적용하지만 센서 데이터들은 아주 빠른 속도로 대 량으로 유입이 되기 때문에 기존의 데이터 마이닝 기법은 적용이 불가능하게 된다. 그러므 로 센서 데이터에 맞는 새로운 센서 데이터 마이닝 기법이 필요하다.

본 논문에서는 센싱된 센서 데이터들을 기반으로 해양 환경 감시 서비스에 제공할 수 있 는 센서 마이닝 기법들을 제안한다.

### 1. 서론

유비쿼터스 센서 네트워크(Ubiquitous Sensor Network, USN)는 유비쿼터스 컴퓨팅 구현을 위한 기반 네트워크로 초 경량, 저전력의 많은 센서들로 구성된 무선 네트워크이다.

해양에 이러한 USN 환경을 구축하여 해양 환경감시 서비스를 도입하면 보다 나은 해양 관리를 진행할 수 있다. 해양 환경 감시 서비스는 해양에서 수온, 기온, 염도, ph 등을 측정하고 센싱 데이터를 기반으로 상황 정보에 대한 이벤트를 감지하고 서버나 현장 작업자 에게 이벤트의 발생(alert)을 알려주어 즉각적인 대응 을 유도하는 이벤트 알림(event alerting)서비스가 주를 이룬다. 그러나 센싱된 데이터에는 질의나 알림 서비스로 알 수 없는 유용한 지식이 존재할 수 있다. 만약 최근의 기후변화를 분석하여 용존산소량의 변화 정도를 예측할 수 있다면 적조 등의 위험한 상황의 발생에 미리 대응할 수 있을 것이다.

기존의 데이터 마이닝은 정적 테이블 데이터 분석 하여 유용한 정보를 찾아낸다. 그러나 센싱된 해양 센서 데이터는 새로운 데이터가 속도가 가변적이고 아주 빠른 속도로 무한히 유입되므로 제한된 용량에 저장하지 못한 데이터는 잃어버리게 되며 다시 볼 수 없다. 그러므로 센서 데이터에 대해서는 요약된 정보 [2]를 가지고 연속적인 질의[1]를 처리하고 분석하게 되며 데이터 마이닝의 결과도 연속적으로 계속 바뀌 어야 하기 때문에 데이터의 양이 많을수록 크게 성능 이 떨어지는 일반적인 데이터 마이닝 기법들을 센서 데이터에 그대로 적용하기는 어렵다.

따라서 본 논문에서는 효율적인 해양 환경 관리를 위한 USN 플랫폼 시나리오와 미들웨어 구조를 제안한 다. 또한 이를 기반으로 보다 유용한 서비스를 제공 하기 위하여 센서 데이터에 대한 분석과 센서 데이터 마이닝 기법들을 제안한다. 전체적인 논문의 구성은 다음과 같다. 2 장에서 해양 관리에 대한 기존의 연 구들을 살펴보고 3 장은 해양 환경 감시 서비스를 위 한 시나리오와 센서 데이터 마이닝에 대한 정의를 내 린다. 다음 4 장에서는 구체적인 센서 데이터 마이닝 기법과 요구사항들을 정리하고 5 장에서는 결론과 향 후 연구에 대해 논의한다.

<sup>†</sup> 이 논문은 한국 전자통신 연구원 “연속 센서 데이 터 마이닝 기법연구” 사업의 연구비 지원에 의해 수 행됨

2. 관련 연구

해양에서 USN 환경을 구축한 사례로는 국내 전북대학교 토목 공학과 해안 항만 실험실 황규남 교수 팀의 광양만 해양 환경 모니터링 시스템을 들 수 있다. 남제주군 미라도 남서쪽에 위치한 이어도에 위치한 이 모니터링 시스템은 조석, 파랑, 조류, 저면고, 풍속, 풍향, 기압, 습도, 일사량, 수온, 염도, 용존 산소량, pH, 탁도, 등 환경 변화에 관련된 각종 요소를 실시간으로 측정하여 해양 환경 오염을 방지한다.

또 다른 한가지로는 한국 해양 연구원 실시간 해양 해상기상 모니터링 시스템이 있다. 이 시스템은 범지구해양관측시스템(Global Ocean Observing System, GOOS)의 시범사업인 동북아시아해양관측시스템(NEAR-GOOS)의 구축에 따라 만들어졌으며 기상, 수온, 염분 및 파고자료는 각각의 센서와 컨트롤러를 통해 독립적으로 측정하고 해양에서의 시, 공간적 변화를 장기간, 연속적으로 실시간 관측하여 그 자료를 사용자(국가기관, 일반 사용자)에게 제공, 활용하게 함으로써 한반도 주변 해양과 한반도의 연안 역에서 연안방재, 해난구조, 해양오염 방제 등 해양재해에 대비하며 나아가 수산, 해운, 해양 개발 등 원활한 해양 산업 활동을 지원한다.

국외에서의 관련 연구를 본다면 Integrated Acoustics Systems for Ocean Observations(IASOO)를 예로 들 수 있다. 이 시스템은 소리가 해양에서 전파될 수 있다는 특성을 이용하여 음성 센서를 바다에 설치해놓고 광범한 센서 네트워크를 구축한다. 바다 깊이 설치된 음성 센서는 바다 생물이 내는 음성들을 측정하고 해양 생물의 활동을 관측한다. 그리고 또한 해양 표면의 바람 소리, 비소리 등 음성 데이터를 분석함으로써 해양 교통도구의 navigation 에도 도움을 준다.

3. 해양환경 서비스를 위한 연속 센서 데이터 마이닝

이 장에서는 해양 환경 서비스를 위한 시나리오와 서비스 유형 및 센서 데이터 마이닝에 대한 정의들을 다룬다.

3.1 해양 환경 서비스 시나리오

본 논문에서 다루고자 하는 해양 환경 감시 서비스는 해양의 부표 또는 인공 지지물에 무선 센서 네트워크를 구축하여 센서로부터 수온, 기온, 염도, pH 등 해양 환경 정보를 자동으로 수집하고 USN 미들웨어 플랫폼의 이벤트 처리기술을 이용하여 해양 환경의 변화와 오염에 신속하게 대처할 수 있도록 서비스를 제공한다. 또한 이러한 해양 환경 감시 시스템을 이용하여 해양 환경 혹은 바다목장을 원격으로 통합적인 관리를 진행한다.

이러한 해양 환경 서비스를 위한 해양 환경 USN 미들웨어 시스템 구조도는 그림 1 과 같다.

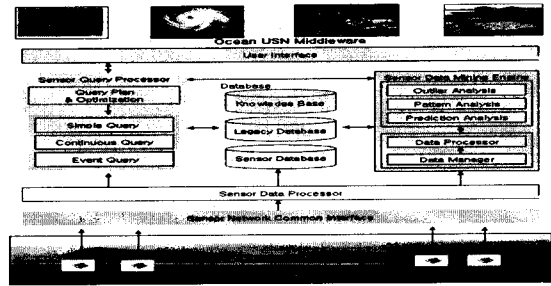


그림 1. 해양 환경 기반 USN 미들웨어 구조도

이 시스템은 센서 노드, 센서 네트워크 공통 인터페이스, 센서 데이터 처리기, 센서 질의 처리기, 데이터베이스, 센서 데이터 마이닝 엔진, User interface 등으로 구성되었으며 기능들을 구체적으로 살펴보면 다음과 같다.

센서 노드는 pH, 용존 산소, 염도, 수온을 측정하며 센서 네트워크 공통 인터페이스를 통해서 데이터 처리기에 전해지고 데이터 처리기는 전해진 데이터를 Sensor Database, 질의 처리기, 센서 데이터 마이닝 엔진에 넘긴다.

센서 질의 처리기는 사용자 인터페이스로부터 입력된 질의를 분석하여 질의 실행계획을 세우고 따라서 어떠한 질의인지를 판단한다. 또한 최종적으로 질의 결과를 사용자 인터페이스에 반환한다.

센서 데이터 마이닝 엔진은 센서 질의 처리기와 연관성을 가지며 요청된 질의와 데이터 처리기로부터 유입된 데이터를 가지고 이상치 분석, 패턴 분석, 예측 분석을 진행하게 되며 그 결과를 데이터 저장소인 Knowledge Base 에 저장한다. 질의 처리기는 최종적으로 Knowledge Base로부터 질의 결과를 사용자 인터페이스에 반환을 한다.

데이터베이스에서 Sensor Database 는 센서 노드들로부터 수집된 센서 데이터를 저장하고 Legacy Database 는 기존에 존재하는 데이터베이스를 가리키며 Knowledge Base 는 센서 마이닝 엔진으로부터 얻어진 결과를 저장한다. 데이터베이스는 센서 질의 처리기와 센서 데이터 마이닝 엔진에 모두 데이터를 제공하게 된다.

해양 환경 감시 서비스는 수집된 수온, 기온, 염도, pH 등 센서 데이터를 이용하여 적조, 태풍, 해양 환경 오염 등의 발생을 분석하고 예측하여 유용한 정보를 제공할 수 있는데 그에 대한 서비스 가능한 서비스 유형들을 정의해 보면 표 1과 같다.

분류	지식의 예
이상치 패턴	지식 1: 오늘 인근 해안의 해수의 밀도가 이상을 보이는 지역을 알려달라
	지식 2: 동해 해안에서 수온 변화가 이상을 보이는 지역을 분류하라.
	지식 3: 바다표물 표면 온도가 평균치보다 2-10도 높은 상태가 지속될 때 엘니뇨가 발생할 확률은 얼마인가?

변화 패턴	지식 4: 지역별 용존 산소 농도의 변화가 3mg/L 이상일 경우 적조의 발생 패턴을 찾는다.
	지식 5: 동해 남에서부터 북으로의 해안 수온변화에 따른 적조의 분포 패턴을 찾는다
패턴 예측	지식 6: 지난 해류의 이상패턴을 분석하여 이상징후가 80%이상인 지역을 알려달라
	지식 7: 용존 산소량, 온도 및 기후 데이터를 기반으로 일주일 후의 지역별 용존 산소량 및 온도를 예측하라

표 1 서비스 유형

### 3.2 센서 데이터 마이닝 정의

해양 센서 네트워크로부터 센싱된 수온, 염도, 용존산소 등 센서 데이터는 단일 센서 스트림(Single Sensor Stream)과 다중 값 스트림(Multi Value Stream)에 대한 마이닝으로 나눌 수 있으며 이러한 측면에서 센서 데이터 마이닝 기법은 크게 이상치 분석, 패턴 분석, 예측 분석과 상관성 분석이 있다. 또한 시간적인 측면에서 이러한 센서 데이터 마이닝 기법들은 구체적으로 다음과 같다. 단일 센서 스트림에 대한 특정된 한 시점의 스냅샷(Snapshot) 센서 데이터 분포에 대한 마이닝 측면에서 일반적인 이상치 분석과 공간 이상치 분석 및 공간 센서 패턴이 있다. 인터벌(Interval) 측면에서 어떤 특정된 기간의 long term 시간 센서 패턴과 슬라이딩 윈도우(Sliding Window)에 의한 마이닝으로서 연속 이상치 분석, 연속 패턴 매칭 및 패턴 기반 예측으로 나눌 수 있다. 센서 데이터 마이닝 기법에 대한 분류는 그림 2 과 같다.

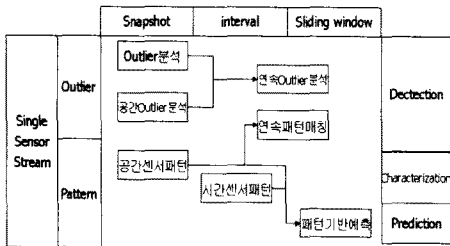


그림 2. 센서 데이터 마이닝 분류

해양 환경을 위한 센서 데이터 마이닝의 구체적인 기법은 4 장에서 다룬다.

## 4. 연속 센서 데이터 마이닝

해양 환경에서 서비스 가능한 유형으로는 크게 센서 데이터 이상치 분석, 센서 데이터 패턴 분석, 센서 데이터 예측 분석 3 가지로 나눈다. 따라서 이 장에서는 각각에 대한 구체적인 마이닝 개념과 요구사항들을 정의하고 서비스 가능한 지식 유형들을 예로 정의하고 설명한다.

### 4.1 센서 데이터 이상치 분석

이상치는 데이터베이스에는 포함되나 데이터의 모델이나 일반적인 행동에 대응하지 못하고 남아있는 데이터 집합에 불일치되거나 심하게 다른 데이터 객체를 가리킨다. 즉 해양 환경에서의 이상치는 평소의 보편적인 현상과 아주 다른 이상징후를 나타내는 것을 가리킨다.

#### 4.1.1 일반 이상치 분석

특정 시점(시각)의 센서 네트워크로부터 수집된 해양 센서 데이터에 대하여 예상되지 않은 센서 값을 검출한다. 이 분석 기법에서 중요시해야 할 점은 사용자는 센서 데이터베이스로부터 특정 시점의 센서 데이터 집합을 한정하거나 특정 시점에 센싱 될 센서 데이터의 집합을 한정하여야 한다. 일반 이상치 분석의 예로 다음과 같다.

**지식 1:** 오늘 인근 해안의 해수의 밀도가 이상을 보이는 지역을 알려달라.

해수의 밀도에 대해 큰 변화폭을 갖지 않은 해양 환경에서 선박들과 같은 기름 유출 사고는 해양에서 심각한 환경오염을 초래하고 또한 어류의 생존 환경에 치명적인 타격을 준다. 따라서 일반 이상치 분석은 이런 사고를 감지하고 대응하도록 유용한 정보를 제공할 수 있다.

#### 4.1.2 공간 이상치 분석

공간 이상치는 특정 시점(snapshot)의 센서 네트워크로부터 수집된 해양 센서 데이터와 해당 센서의 공간 속성을 함께 고려하여 이상치를 분석하는 것이며 주로 위치 속성을 이용하여 그 이웃과의 심각한 차이를 보이는 것이다. 이 분석기법에서 중요시해야 할 점은 사용자는 특정 센서들에 대한 공간 개념 계층의 형태로 사전 지식을 제공할 수 있어야 한다. 그 구체적 예는 다음과 같다.

**지식 2:** 동해 해안에서 수온 변화가 이상을 보이는 지역을 분류하라.

수온은 지역적으로 서로 다른 온도 분포를 나타내며 물고기들은 약간의 수온 변화에도 민감한 반응을 보이며 그들의 서식 환경을 옮긴다. 그러므로 수온 변화에 대하여 알고 있다면 특정 지역에서 특정 물고기들의 분포상황을 알 수 있으며 또한 어부들의 어획량을 대폭적으로 증가시키는데 도움이 될 수 있다.

#### 4.1.3 연속 이상치

연속 이상치는 순간적이 아닌 일정한 시간 간격(time interval)동안 해양 센서 네트워크로부터 수집된 해양 센서 데이터에 대한 이상치이다. 이 분석에서 중요시해야 할 점은 연속적인 시간 윈도우의 경우 윈도우의 크기와 시작시점, 끝시점이 명시되어야 하며 사용자는 분석 결과를 반환할 시기인 슬라이드의 크기를 정의할 수 있어야 한다. 구체적인 예는 다음과 같다.

**지식 3:** 바닷물 표면온도가 평균치보다 2~10 도 현

**상을 이상치로 알려달라.**

이런 지속적인 상황이 발생하게 된다면 엘니뇨가 발생할 가능성이 높다. 엘니뇨는 전 세계적인 현상으로서 인류에게 엄청난 피해를 가져다 주므로 그 발생을 판단하고 대응하는 것이 필요하다.

**4.2 센서 데이터 패턴 분석**

패턴이란 일정한 형태의 양식 또는 유형의 주기적인 행동을 가리키며 해양 환경에서의 센서 데이터 패턴은 시간 패턴과 공간 패턴으로 나누어 볼 수 있다.

**4.2.1 시간 패턴 분석**

시간 패턴 분석이란 특정 시간 간격(time interval) 동안 센서 네트워크로부터 수집된 센서 데이터에 대한 추세 및 반복을 가리킨다. 이 분석에서 중요시해야 할 점은 사용자에 의해 특정 시간간격이 제시되어야 하며 시간 간격의 경우 시작 시점과 끝시점을 명시하여야 한다.

**지식 4: 지역별 용존 산소 농도의 변화가 3mg/L 이상일 경우 적조의 발생 패턴을 찾는다.**

적조의 발생은 해양 생물의 대량 번식으로 하여 용존 산소를 많이 소모하고 따라서 어류의 대량적인 폐사를 초래하므로 발생주기를 판단함으로써 인류에게 주는 피해를 감소시킬 수 있다.

**4.2.2 공간 센서 패턴**

공간 센서 패턴은 특정 시점(snapshot)의 센서 네트워크로부터 수집된 센서 데이터와 해당 센서의 공간 속성을 고려하여 패턴을 분석한다. 이 기법에서 중요시해야 할 점은 사용자는 특정 센서들에 대한 공간 개념 계층과 패턴 모델을 제공할 수 있어야 한다.

**지식 5: 동해 남에서부터 북으로의 해안 수온변화에 따른 적조의 분포 패턴을 찾는다.**

동해 남에서부터 북이라는 위치 속성을 가지고 공간적으로 적조의 발생 패턴을 분석함으로써 발생에 대응한다.

**4.3 센서 데이터 예측 분석**

예측이란 일정한 패턴에 기반하여 미래의 행동에 대한 추세 또는 데이터 수치에 대한 짐작을 가리킨다. 해양 환경하에서의 예측은 연속 패턴 매칭 분석과 패턴 기반 예측으로 나눌 수 있다.

**4.3.1 연속 패턴 매칭**

연속 패턴 매칭은 사전에 탐사된 시간 패턴을 이용하여 특정 시간 간격(time interval) 동안 센서 네트워크로부터 수집된 센서 데이터에 대하여 유사 패턴을 보이는 것이다. 이 분석 기법에서 중요시해야 할 점은 연속적인 시간 윈도우의 경우 논리적으로 식별되는 특정 센서 타입의 센서 데이터와 윈도우의 크기와 시작시점, 끝시점을 함께 명시해야 한다

**지식 6: 지난 해류의 이상패턴을 분석하여 이상징후가 80%이상인 지역을 알려달라**

해류의 흐름은 해양생물이 살수 있는 환경 요소인 수온과 용존 산소량 등의 변화에 직접적인 영향을 끼

친다. 따라서 해류의 이상패턴을 분석함으로써 엘니뇨와 같은 유사한 상황의 발생을 미리 예측할 수 있다.

**4.3.2 패턴 기반 예측**

패턴 기반 예측은 사전에 탐사된 시간 패턴을 이용하여 특정 시점까지 규정된 시간 윈도우(time window)만큼의 연속적인 시간 윈도우(sliding time window) 동안 유사 패턴을 보이는 예측 질의이다. 이 기법에서 중요시해야 할 점은 특정 시점에 따른 예측 질의는 사용자 질의 패턴과 유사도 임계치를 만족하며 신뢰도 임계치를 만족하는 경우 반환 할 값의 미래 시점을 제시하는 것이다.

**지식 7: 용존 산소량, 온도 및 기후 데이터를 기반으로 일주일 후의 지역별 용존 산소량 및 온도를 예측하라**

온도가 높으면 용존 산소량이 작아지고 온도가 낮으면 용존산소량이 높아진다는 패턴이 존재한다면 이러한 패턴에 의해서 미래 특정 시점 특정 지역에서의 물고기의 생존 환경 등을 예측해낼 수 있다.

**5. 결론 및 향후 연구**

현재까지는 스트림 데이터 마이닝으로서 네트워크 모니터링, 교통 감시 시스템 등 분야에서 많이 연구되어 왔지만 해양 환경하에서 센서 데이터에 대한 마이닝 연구는 극히 적었다. 해양 환경에서 센서는 수온, 기온, 염도 등 데이터를 측정하게 되며 이러한 센서 데이터의 특성으로 하여 기존 전통적인 마이닝 보다는 연속 센서 데이터 마이닝이 필요하다. 그리하여 본 논문에서는 구체적인 해양 관련 지식으로 센서 데이터 마이닝 기법을 제안하였다.

향후 연구로서 이러한 마이닝 기법에 대한 구체적인 알고리즘을 구현하려고 한다.

**참고문헌**

[1] D. Terry, D. Goldberg, D. Nichols, and B. Oki. Continuous queries over append-only databases. In Proc. of the 1992 ACM SIGMOD Intl. Conf. on Management of Data, pages 321-330, June 1992.  
 [2] R. Avnur and J. Hellerstein. Eddies: Continuously adaptive query processing. In Proc. of the 2000 ACM SIGMOD Intl. Conf. on anagement of Data, pages 261-272, May 2000.  
 [3] <http://realtime.chonbuk.ac.kr/default/document.jsp?view=description>  
 [4] [http://realtime.kordi.re.kr/realtime/realtime\\_01.asp](http://realtime.kordi.re.kr/realtime/realtime_01.asp)  
 [5] Bruce M. Howe, James H. Miller, and the IASOO Committee, "Integrated Acoustics Systems for Ocean Observatories". <http://www.oce.uri.edu/ao/IASOO.htm>.