

Ubiquitous 환경을 위한 자율적 상황정보 분석 방법론

오정석, 강승규, 최경석

한국가스안전공사 가스안전연구원

Autonomous Context-Aware Analysis Method for Ubiquitous Environments

Jeong Seok Oh, Seung Kyu Kang, Kyung Suhk Choi

Institute of Gas Safety R&D, Korea Gas Safety Corporation

1. 서론

유비쿼터스 기술은 모든 사물의 지능화 및 네트워크화를 구축함으로써 개인 삶의 질과 기업의 생산성이 향상되고 공공 서비스의 혁신을 가능하도록 한다. 사물의 지능화는 기존의 확실적인 자동수행(automatic operation) 뿐만 아니라 주변 환경의 영향과 사물의 상태에 따라 스스로 판단하여 행동을 결정하는 자율수행(autonomic operation)도 포함하며, 사물의 네트워크화는 유/무선 통신을 이용하여 사물과 사물 또는 사물과 사람간의 통신을 가능하도록 한다. 이를 위해, 정부는 u-it 839전략, u-Korea 전략, u-City 전략 등 다양한 유비쿼터스 기반 전략을 공표하고 유비쿼터스 공통기술개발 및 유비쿼터스 센서 네트워크(USN)와 같은 네트워크 인프라 구축을 위해 다채로운 사업을 수행하고 있다.

해외의 경우에도 일본과 미국 등에서 유비쿼터스 기술을 산업시설에 적용한 사례들이 존재한다. 일본의 MERL (MITSUBISHI Electric Research Lab)은 산업용 공장 모니터링(IPM, Industrial Plant Monitoring)을 위해 유비쿼터스 기술기반 프로토타입 모니터링 시스템 제작을 추진 및 수행하고 있다. 미국은 국가 에너지 기술연구소(National Energy Technology Laboratory, NETL)에서 가스배관 네트워크 센서 시스템인 GASNET 프로젝트를 수행하였다. GASNET 프로젝트의 목적은 시설기기들의 상태정보를 실시간으로 통신할 수 있는 다목적 센서의 분산된 네트워크인 가스배관 네트워크 센서 시스템의 시현 및 개발을 통해 국가의 천연가스 분배 시설구조의 기능을 최적화하는 것이다. 이 프로젝트의 결과는 미국 전역에서 경제적, 환경적, 안정적 측면에서 이익 창출이 가능하며 연간 천만 달러 이상의 비용을 감소할 수 있는 것으로 산정되었다.

이러한 산업시설에 유비쿼터스 기술을 적용하기 위해서는 센서노드의 개발 네트워크 인프라의 구축 및 활용도 필요한 요소이지만 상황정보를 분석·판단하여 지능적 수행을 유도하는 자율수행은 많은 연구적 노력이 필요하다. 기존 연구사례를 살펴보면 산업시설에 위험성 평가모델, RBI(Risk Based Inspection) 등 다양한 자동수행 기술 및 연구에 대한 노력은 많으나 자율적 판단에 근거한 예측 및 분류 수행이 적용되는 기술 및 연구노력은 미흡하다.

본 연구는 유비쿼터스 기반환경 내의 산업시설 등에서 수집되는 상황정보를 IT기법인 데이터 마이닝 기법을 적용하여 상황정보 분류/예측 모델을 구축하고 구축된 모델을 이용하여 자율적으로 상황정보를 분류·판단을 위한 방법론 제시를 목적으로 한다. 이를 위해, 본 연구는 군집화(clustering) 기법을 이용하여 상황정보 데이터의 전처리(preprocessing) 작업인 클래스 정의를 수행하고 개선된 분류(classification) 기법을 이용하여 상황정보 분류/예측 모델을 구축하며 이론적으로 검증하는 방법을 제시하고자 한다. 본 논문의 결과는 자율적인

상황정보의 분석·판단에 대한 방법론을 제시하여 사람의 개입을 최소화시키고 기존 자동화 방식의 기술에 자율적 방식의 기술을 접목할 수 있는 초석을 제공한다.

본 논문의 구성은 다음과 같다. 2장은 유비쿼터스 환경의 도래와 필요성을 기술하는 정부의 유비쿼터스 기반 전략과 USN(Ubiquitous Sensor Network)에 대해 설명한다. 3장은 유비쿼터스 환경에서 데이터 마이닝 기법을 적용된 상황정보 분석·판단 방법론에 대해 기술한다. 4장은 연구 결과에 대한 결론을 맺고 향후계획을 제시한다.

2. 배경

2.1 유비쿼터스 기반 전략

우리나라는 1999년 3월 인터넷 확산 및 디지털 경제촉진 일환으로 Cyber Korea 21 전략을 발표하였으며, 정보 인프라를 지속적으로 고도화하기 위해 2002년 4월 e-Korea Vision 2006 전략을 공표하였고, 정보화를 통한 국가 경쟁력의 혁신적 제고 추진을 위해 2003년 12월 Broadband IT Korea Vision 2007 전략을 제시하였다. 그 결과, 2005년 기준으로 국가 정보화 지수 세계 3위, 전자정부 수준 5위, 초고속 인터넷 보급률 세계 1위 등의 가시적 성과를 얻었다. 그러나 세계 유비쿼터스 시장이 급성장하여 유비쿼터스 IT 관련 사업이 차세대 국가의 핵심 전략 분야로 대두되고 사물의 지능화 및 네트워크화로 다양한 기술 및 서비스 간의 융·복합이 활발하게 진행됨으로써 2006년 u-Korea 전략을 새롭게 발표하였고 그림 1에서 보인다.

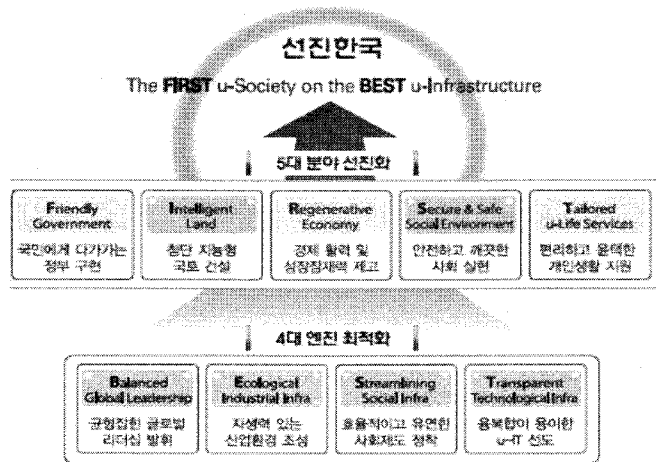


그림 1 u-Korea 전략

u-Korea 전략은 정부, 국토/시설, 경제, 사회, 개인 등의 5대 분야 선진화를 글로벌 리더십, 산업 환경, 사회제도, 융복합 u-IT 등의 4대 엔진 최적화를 통해 달성하는 계획으로서 주요세부 항목으로는 “국가 주요시설물의 지능화를 통한 사회기반 시설의 안전한 관리”, “지역통합관리센터 구축을 통한 u-City의 효율적 운영지원”, “재난·재해 실시간 대응체계 수립으로 국민생활 안전성 강화” 등이 있다.

2.2 USN (Ubiquitous Sensor Network)

USN은 개인의 생활공간 생활기기, 기계 등을 포함한 모든 사물에 컴퓨터 기능과 네트워크 기능을 부여하고 환경과 상황의 자동인지를 통해 인간에게 최적의 기능을 제공함으로써 인간 생활의 편리성과 안전성을 고도화 하는 지능형 네트워크로서 그림 2에서 보인다.

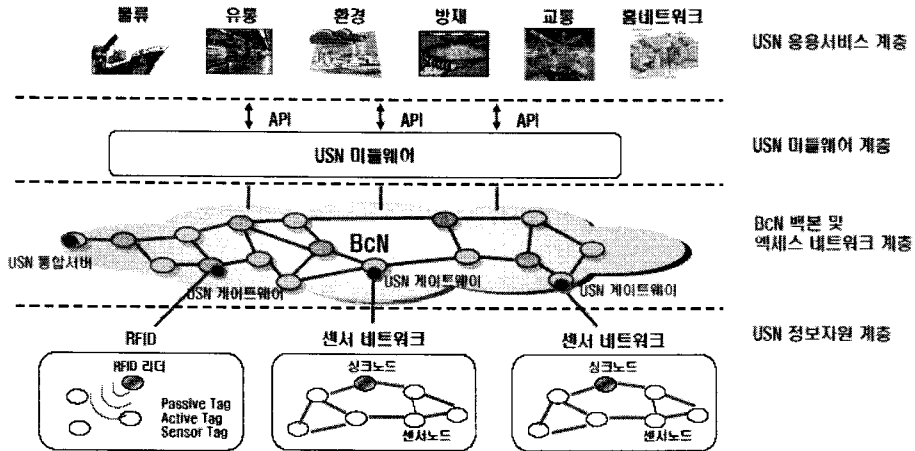


그림 2 USN 구성

USN은 USN 정보자원 계층, 백본 계층, 미들웨어 계층, 응용 서비스 계층으로 분류되며 USN 정보자원계층은 센서노드 및 센서 미들웨어로 구성되며 센서노드는 환경, 물리계에서 센싱된 정보 및 특정 이벤트를 유/무선 통신 기술기반으로 전달하거나 컴퓨팅을 수행하는 센서, 프로세서, 통신소자, 전지등으로 구성되는 시스템으로 데이터 처리, 통신경로 설정 등을 수행한다. 백본 계층은 데이터, 음성, 방송 등 다양한 형식의 정보들을 하나의 표준형식으로 취합하는 BcN(Broadband Convergence Network) 백본 인프라가 점차적으로 구성될 예정이다. USN 미들웨어 계층은 이기종 센서네트워크로부터 수집한 센싱 데이터를 필터링·통합·분석하여 의미 있는 상황정보를 추출·저장·관리·검색하고 응용서비스 전달 및 서비스간 연계 및 통합하는 역할을 담당한다. USN 응용 서비스 계층은 시설 목적에 적합한 서비스를 프로그램을 사용자에게 제공하는 역할을 제공한다.

3. 자율적 상황정보 분석 방법론

산업시설에서 시설의 상태 및 주변 환경에 대한 정보를 습득하여 시설의 건전성이나 위험성 등을 자동적(autonomic)으로 모니터링하고 예측하는 노력은 지속적으로 시도해왔으나 유비쿼터스 시대가 도래하면서 원격 무선 센싱 등을 통해 기하급수적으로 정보가 늘어날 것으로 예상되며, 기존의 방법으로 시설의 상태나 주변 환경을 분석 및 분류하고 예측하기가 더욱 어려워 질 것으로 사료된다.

본 연구는 시설의 유비쿼터스 기반환경에서 발생하는 시설의 상태나 주변환경 정보를 데이터 마이닝 기법을 이용하여 자율적(autonomic)으로 분석 및 예측하는 방법론에 대한 연구를 목적으로 한다. 마이닝 기법의 적용을 위해서는 인스턴스 구축 및 클래스 정의를 선행하고 상황정보 분류 및 예측을 위한 데이터 마이닝 모델 구축 및 검증이 수행되어야 본 연구가 원하는 분류 및 예측이 가능해진다.

인스턴스 구축은 유비쿼터스 환경의 시설에서 수집된 상태 데이터의 정제된 모형을 의미하며 시설의 상황을 잘 설명할 수 있는 속성들을 선별하여야 한다. 클래스 정의는 전체 인스턴스에 대한 분석을 요구하며, 전체 인스턴스 분석은 데이터에 의한 다양한 영향성을 고려해야 하므로 어렵고 소요시간이 많이 걸리는 작업이다. 유비쿼터스 기반환경의 시설물에서 발생하는 데이터는 다양하고 복잡하므로 클래스를 과도하게 분류하여 정확성을 하락시킬 수 있다. 과도한 분류를 억제하기 위한 효과적인 클래스의 생성은 선택정보의 군집화 과정

을 통해 수행될 수 있다. 선택 정보의 군집화는 전체 인스턴스를 몇 개의 작은 군집으로 분할하여 관련성 높은 인스턴스를 동일한 군집으로 재배치되며 생성된 군집은 클래스로서 활용된다. 선택정보의 군집화는 "unsupervised" 마이닝의 군집 기법을 이용해서 수행될 수 있다. 군집 기법은 수행 과정 중에서 전체 인스턴스 분석, 인스턴스의 군집 할당, 추천 정보의 생성에 대해 사람의 개입을 감소시키므로 소요되는 시간을 감소시키고 추천 정보를 효율적으로 생성한다. 마이닝의 군집 기법을 이용해 군집화된 정보는 각 인스턴스에 대한 선택 정보로서 재구축 되어야 한다. 그림 3은 추천 정보의 군집화로 생성된 n개의 군집에서 k개의 인스턴스에 대해 자원 추천을 위해 인스턴스가 재구축 되는 예를 보이고 있다

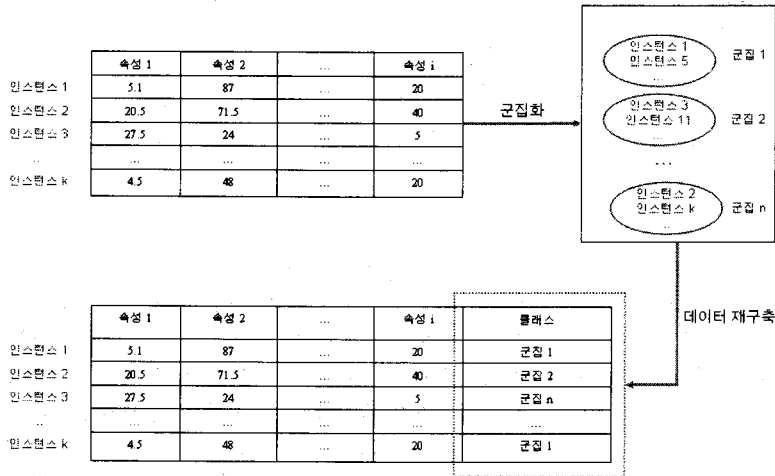


그림 3 인스턴스 구축 방식

재구축된 인스턴스를 기반으로 생성되는 상황정보 모델은 근사(approximation) 기법을 적용하여 클래스에 대한 인스턴스의 분류로서 생성된다. 근사 기법은 가설 및 통계적 수식의 정의가 필요 없는 "supervised" 마이닝을 이용한다. 마이닝 기법은 모델 구축 과정에서 의미 있는 정보를 추출하기 때문에 인스턴스 속성간의 관계나 인스턴스와 추천정보의 관계를 명확히 규명할 필요가 없다.

모델의 검증은 일반적으로 인스턴스 분할을 이용하는 방법을 사용한다. 인스턴스 분할을 이용하는 방법은 추천 모델에 사용되었던 인스턴스를 n개의 조각으로 분할하여 교차 검증을 수행한다. 전체 인스턴스의 분할은 가능한 동일한 인스턴스 개수와 클래스를 가지는 n개의 집합으로 재배치되어야 한다. 그림 4는 인스턴스를 n개의 조각으로 분할하여 검증을 수행하는 n 교차 검증에 대한 수행과정을 보인다. n 교차 검증은 총 n 개의 단계로 구분되어 검증이 수행된다. 각 단계는 클래스가 균등하게 분포된 n개의 조각에서 n-1 조각을 이용해서 상황정보 모델을 생성하고 나머지 한 조각을 시험 데이터로 이용해서 추천 모델에 대한 정확률(에러율)을 계산한다. 최종 정확률(에러율)은 n개의 정확률(에러율)에 대한 평균을 이용한다.

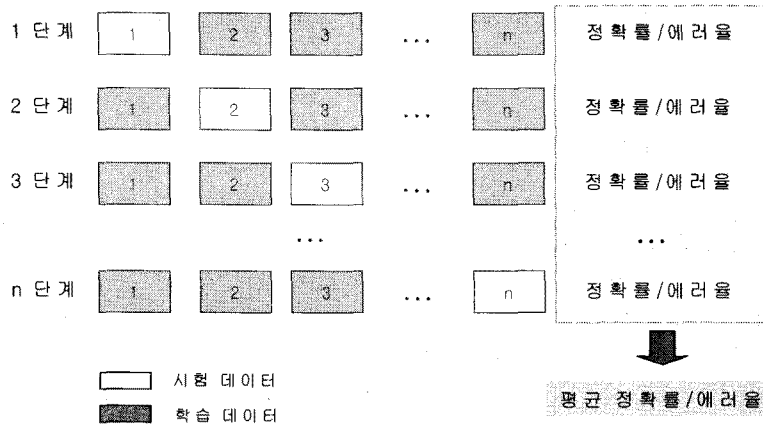


그림 4 n 교차 검증의 수행 과정

구축된 상황정보 모델은 새로운 인스턴스의 입력에 대해 유사한 특성을 가진 인스턴스 집합인 클래스로 결과를 반환한다. 클래스 단위의 결과는 여러 개의 인스턴스들이 포함되어 복수의 자원 환경을 추천 결과로서 반환하기 때문에 모호하다. 상황정보 분류·예측은 입력된 인스턴스에 대해 가장 근사한 인스턴스를 찾아 상황정보 기록을 반환해야 하므로 인스턴스 단위의 추천이 수행되어야 한다. 추천 인스턴스는 입력된 인스턴스와 거리를 계산하여 가장 낮은 거리값을 가진 인스턴스로 선택된다. 일반적으로 거리를 구하는 공식은 유클리디안, 맨하탄 등이 적용되지만 이런 거리 방식들은 고차원을 가진 인스턴스에서 거리를 구할 때 값의 범위가 작은 차원에 대한 편향을 발생시킨다. 본 연구는 고차원에서 특정 차원에 의해 야기되는 편향을 막기 위해 그림 5처럼 두 인스턴스 간의 변이계수를 거리로 사용하는 것을 제안한다.

유클리디안	$\sqrt{\sum_{k=1}^n X_{ik} - X_{jk} ^2}$	새로운 인스턴스 집합	27.00, 94.00, 78.30, 24.00, 85000, 230000, 75000, 160900000, 35
맨하탄	$\sum_{k=1}^n X_{ik} - X_{jk} $	클래스에 있는 인스턴스 집합	1. 26.69, 89.71, 78.23, 23.03, 80251, 440058, 83052, 159735244, 35 2. 27.06, 92.61, 78.17, 22.51, 77326, 315628, 80720, 162347304, 35 3. 26.84, 95.53, 78.23, 22.86, 67573, 190136, 69487, 159093032, 35
변이계수	$\sum_{k=1}^n \frac{ X_{ik} - X_{jk} }{\left(\frac{X_{ik} + X_{jk}}{2}\right)}$		
유클리디안에서 산출된 거리값	맨하탄에서 산출된 거리값	변이계수에서 산출된 거리값	
인스턴스 1 : 1183582.794	인스턴스 1 : 1387620.37	인스턴스 1 : 0.891	
인스턴스 2 : 1449866.418	인스턴스 2 : 1546329.07	인스턴스 2 : 0.574	
인스턴스 3 : 1807500.093	인스턴스 3 : 1869774.9	인스턴스 3 : 0.575	

그림 5 인스턴스간의 거리 공식과 계산 예

4. 결론 및 향후계획

유비쿼터스 기술은 모든 사물의 지능화 및 네트워크화를 구축함으로써 개인 삶의 질과 기업의 생산성이 향상되고 공공 서비스의 혁신을 가능하도록 한다. 사물의 지능화는 기존의

획일적인 자동수행(automatic operation) 뿐만 아니라 주변 환경의 영향과 사물의 상태에 따라 스스로 판단하여 행동을 결정하는 자율수행(autonomic operation)도 포함하며, 사물의 네트워크화는 유/무선 통신을 이용하여 사물과 사물 또는 사물과 사람간의 통신을 가능하도록 한다.

본 연구는 유비쿼터스 기반환경 내의 산업시설 등에서 수집되는 상황정보를 IT기법인 데이터 마이닝 기법을 적용하여 상황정보 분류/예측 모델을 구축하고 구축된 모델을 이용하여 자율적으로 적합한 상황에 대한 분류·판단을 위한 방법론을 제시하였다. 마이닝 기법이 필요한 방법론을 위하여 본 연구는 군집화 기법과 분류 기법을 적용한 구축 모델을 제시하였고, 구축 모델을 이용하여 새로운 상황정보의 분류 및 예측에 대한 정확성을 높이기 위해 인스턴스 단위로 정보를 분석가능하게 변이계수 적용함으로써 클래스 단위내의 세부적인 인스턴스 선택까지 가능하게 하였다. 또한 구축된 모델을 사용성 여부를 확인하기 위해 n 교차검증기법을 적용하고 정확률이 90% 이상일 때만 사용성이 있음을 기술하였다.

본 논문의 결과는 자율적인 상황정보의 분석·판단에 대한 방법론을 제시하여 사람의 개입을 최소화시키고 기존 자동화 방식의 기술에 자율적 방식의 기술을 접목할 수 있는 초석을 제공한다. 향후 연구로는 제안된 방법론을 이용하여 특정 중대시설에서 상황정보를 수집하여 상황정보 모델을 구축하고, 시현한 결과를 제시하며, 기존 산업시설에 적용되었던 자동화 기술과 접목하는 방안을 수행할 예정이다.

5. 참고문헌

1. 강근석; 김충락, 회귀분석, 교우사, 2005
2. 류근관, 통계학, 법문사, 2003
3. 박정식;윤영선, 현대통계학(제4판), 다산출판사, 2005
4. 정보통신부, u-KOREA 기본계획, 2006
5. 정보통신부, u-CITY 구축 활성화 기본계획, 2006
6. Cover, T; Hart, P. Nearest Neighbor Pattern Classification, IEEE Transactions on Information Theory, 1967, 13(1), 21-27
7. Elnaffar, S; Martin, P; Horman, R. Automatically Classifying Database Workloads, CKIM Conference, 2002, 622-624
8. Ganek, A; Corbi, T. The Dawning of the Autonomic Computing Era, IBM Systems Journal, 2003, 42(1), 5-18
9. Moorel, D, Statistics Concepts and Controversies (fifth edition), W.H.Freeman and Company, 2001
10. Goebel, M; Gruenwald, L. A Survey of Data Mining and Knowledge Discovery Software Tools, ACM SIGKDD Explorations, 1999, 1(1), 20-33
11. Oil & Natural Gas Projects (Transmission, Distribution, & Refining), Gasline Network Sensor System
[http://www.netl.doe.gov/technologies/oil-gas/NaturalGas/Projects_n/TDS/TD/T&D_A_41320GASNET.html]
12. Weka Public Machine Learning (Mining) Software tool (version 3.4)
[<http://www.cs.waikato.ac.nz/weka/index.html>]
13. Witten, H; Franke, E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implements, Morgan Kaufmann Publishers, 1999
14. Zangl, G; Hannerer, J. Data Mining Applications in the petroleum industry, IBM Round Oak Publishing, 2003