

# RFID/USN 기반 지능형 가스안전관리 서비스를 위한 자율적 분석 연구

오정석<sup>†</sup> · 최경석 · 권정락 · 윤기봉\*

한국가스안전공사 가스안전연구원 · \*중앙대학교 기계공학부

(2008. 7. 20. 접수 / 2008. 11. 3. 채택)

## A Study on Autonomic Analysis for Servicing Intelligent Gas Safety Management Based on RFID/USN

Jeong Seok Oh<sup>†</sup> · Kyung Seok Choi · Jeong Rock Kwon · Ki Bong Yoon\*

Korea Gas Safety Corporation

<sup>†</sup>Division of Mechanical Engineering, Chung-Ang University

(Received July 20, 2008 / Accepted November 3, 2008)

**Abstract :** As RFID/USN technology is used in the latest industry trend, the information analysis paradigm shifts to intelligence service environment. The intelligent service includes autonomic operation, which select activity by defining itself to the status of industry facilities. Furthermore, information analysis based on IT used to frequently data mining for detecting the meaning information and deriving new pattern. This paper suggest self-classifying of context-aware by applying data mining in gas facilities for serving the intelligent gas safety management. We modify data algorithm for fitting the domain of gas safety, construct context-aware model by using the proposed algorithm, and demonstrate our method. As the accuracy of our model is improved over 90%, the our approach can apply to intelligent gas safety management based on RFID/USN environments.

**Key Words :** RFID/USN, data mining, intelligent service, context-aware model, gas safety management

### 1. 서론

IT 기반의 정보 기술이 발전하면서 축적된 데이터를 바탕으로 의미 있는 정보를 생성하고 의미 있는 정보간에 관련성을 분석하여 새로운 패턴이나 지식을 추출하는 데이터 마이닝 기술이 발전해왔다. 특히, 마이닝 기술은 축적된 데이터를 기반으로 정보 모델을 구축하여 새롭게 발생된 정보의 분류 및 예측 등을 수행할 수 있다<sup>3-6)</sup>.

이와 더불어 정부가 u-It 839, u-Korea, u-City 등의 RFID/USN 기반 전략을 공표하면서 RFID/USN 공동기술 개발과 관련 네트워크 인프라 구축을 위해 다양한 시도를 추진하고 있으며 다양한 산업시설에 RFID/USN 기술을 적용함으로써 이종 분야간의 융·복합을 시도하고 있다<sup>1,2,7,9,10)</sup>.

근래에 이르러 많은 산업시설에 적용되고 있는 자동화 기술이 적용되고 있으며, 최근에 RFID/USN

기술 등 지능화(intelligence) 기술도 요구되었다. 지능화 기술은 시설 주변 환경의 영향과 상태를 스스로 판단하여 행동을 결정하는 자율수행(autonomic operation)의 의미를 포함하고 있다<sup>7,10)</sup>.

가스분야와 같이 이미 산업분야에서는 자동화를 위한 많은 기술과 연구가 수행되었지만 시설의 상황정보를 분석·판단하여 지능적 수행을 요구하는 자율수행에 대한 기술과 연구노력은 미흡하다.

본 연구는 가스안전관리를 위해 가스시설에서 수집되는 상황정보를 데이터 마이닝 기법을 적용하여 자율적으로 상황정보를 분류하는 도대 제시를 목적으로 한다. 이를 위해, 본 연구는 가스안전관리 도메인에 적합하도록 데이터 마이닝 알고리즘을 개선한다. 또한, 가스관련 데이터를 실 데이터(real data)와 가 데이터(simulation data)를 수집/생성하고 변경된 알고리즘을 이용하여 생성된 상황인식 모델을 구축하여 검증을 수행한다. 본 연구의 결과는 가스안전관리의 지능화에 대한 초석을 제시할 수 있다.

<sup>†</sup> To whom correspondence should be addressed.  
dbstar@kgs.or.kr

## 2. 데이터마이닝의 소개

데이터 마이닝 기법을 이용한 모델에서 처리되는 데이터 구조는 Table 1에서와 같다. 데이터 집합을 테이블 형태로 가정할 때, “인스턴스”는 테이블의 행에 해당되며, “속성”은 열에 해당한다. “클래스”는 예측 및 분류를 수행해 결과로 출력되는 속성을 의미하며 적용되는 마이닝 기법에 따라 선택적으로 존재할 수 있다.

데이터 마이닝 모델을 구축하는 방법은 크게 두 가지로 나뉜다. 첫 번째 방법은 클래스 정보를 포함한 마이닝 모델을 구축하는 “supervised” 마이닝이며 대표적으로 분류 혹은 예측 기법이 있다. 다음 방법으로는 클래스 정보를 포함하지 않고 마이닝 모델을 구축하는 “unsupervised” 마이닝이며 대표적으로 군집 기법이 있다. “unsupervised” 마이닝은 데이터간의 유사성을 바탕으로 데이터를 세분화 한다.

데이터 마이닝 모델을 이용하여 분류나 예측을 수행하기 위해서는 수집된 인스턴스에 대응되는 기준정보(클래스)를 정의해야 한다. 기준정보의 정의는 전체 인스턴스에 대한 분석을 요구하며, 시설이나 주변 환경의 다양한 영향을 고려해야 하므로 어렵고 소요시간이 많이 걸리는 작업이다. 기준정보를 너무 세분하게 정의할 시에는 과도하게 분류하여 정확성을 하락시킬 수 있으며, 반대로 너무 극대화시키면 분류가 제대로 수행되지 않아 역시 정확성을 하락시킬 수 있다.

데이터 마이닝 기법을 이용한 모델은 인스턴스를 사용하여 분류 및 예측을 수행할 수 있도록 근사(approximation) 기법하에 구축된다. 마이닝 기법은 모델 구축과정에서 의미 있는 정보를 추출하기 때문에 인스턴스 속성간의 관계나 인스턴스와 기준정보의 관계를 통계 수식 모델처럼 명확하게 규명할 필요가 없다. 분류와 예측을 위한 마이닝 모델은 알고리즘에 의해 기준정보(클래스가) 도출 될 수 있다.

모델의 검증은 크게 새로운 인스턴스를 이용하는 방법과 인스턴스 분할을 이용하는 방법으로 나뉠 수 있다. 새로운 인스턴스를 이용하는 방법은 기존에 모델을 구축하였을 때 사용되었던 인스턴스를 제외하고 새로운 인스턴스의 집합을 구축하여 검증을 정확률과 에러율에 기반하여 수행한다. 인스턴스 분할을 이용하는 방법은 모델에서 사용되었던 인스턴스를 n개의 조각으로 분할하여 교차 검증을 수행한다. 전체 인스턴스의 분할은 가능한 인스턴스 개수와 해당되는 클래스에 대해 균등하게 분포되도록 n개의 집합으로 재배치되어야 한다.

Table 1. Data Structure Example

	속성1	속성2	...	속성i	클래스
인스턴스 1	5.1	87	...	20	c1
인스턴스 2	20.5	71.5	...	40	c3
...	...	...	...	...	...
인스턴스 k	4.5	48	...	20	c1

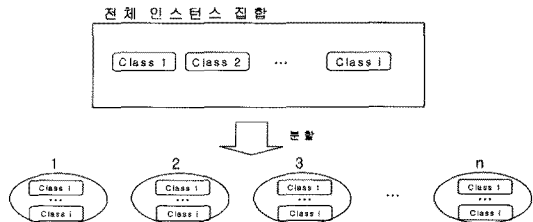


Fig. 1. reassignment process.

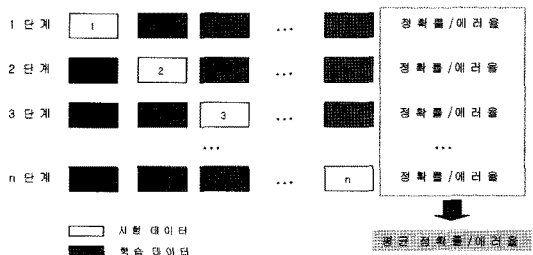


Fig. 2. n cross validation process.

Fig. 1은 전체 인스턴스에 대응되는 클래스가 i개 존재할 때 n개의 군집으로 분할되는 경우를 보인다. 인스턴스 개수와 클래스는 전체 인스턴스가 n개의 군집으로 분할될 때 가능한 균등하게 분포되도록 인스턴스를 재배치한다. 클래스는 모든 조각에서 인스턴스에 대한 클래스가 1부터 i개가 분포되고, 군집당 인스턴스 개수는 전체 인스턴스에서 군집의 나눈 개수로 분포된다.

Fig. 2는 인스턴스를 n개의 조각으로 분할하여 검증 수행하는 n 교차 검증에 대한 수행과정을 보인다. 각 단계는 클래스가 균등하게 분포된 n개의 조각에서 n-1 조각을 이용해서 모델을 생성하고 나머지 한 조각을 시험 데이터로 이용해서 모델에 대한 정확율(에러율)을 계산한다. 최종 정확률(에러율)은 n개의 정확률(에러율)의 평균을 이용한다.

## 3. 정보분석을 위한 데이터 구축

본 연구는 가스시설 특성의 원활한 정보분석을 위해 가데이터(simulation data)를 생성하고 실데이터(real data)를 현장에서 수집하였다. 실 데이터는 두 가지 종류로 분류된다. 첫 번째 실데이터는 CDMA

Table 2. CDMA Data Example

측정일자	최대전위	최소전위	평균전위	상태
2007/10/09 17:06:24	-1554	-1372	-1455	정상
2007/06/27 20:55:05	-879	-741	-818	주의
2007/07/16 14:44:56	-161	-150	-150	경고
...	...	...	...	...

Table 3. WPAN Data Example

온도	습도	조도	적외선	장소
8	60	55	125	지하
7	42	770	167	지하
24	34	803	150	실내
25	33	792	141	실내
14	24	1003	196	실외
8	30	990	186	실외
...	...	...	...	...

방식을 이용해 A 도시가스사 모니터링 시스템으로 전송되는 최대, 최소, 평균 전극치를 이용한다. A사 모니터링 시스템은 이 값들을 받아 A사에서 정한 기준에 의해 정상, 주의, 경고 등의 상태를 판별한다. 본 연구는 Table 2와 같이 A 도시가스사의 협조를 받아 2007년 6월부터 10월까지 수집된 약 6,420건의 데이터를 획득하였다. 또한 ZigBee 센서 Mote를 Mesh 네트워크를 구성해서 실내, 실외, 지하철 공사구간에서 약 18,000여건의 상황정보 데이터를 수집하였고 Table 3에서 보인다. 상황정보 수집 항목은 사용된 Zigbee Mote(기기)에 부착된 센서에서 제공하는 온도(°C), 습도(%) 및 조도와 적외선(빛)에 의해서 변화된 전압값이다.

가스산업시설에서 발생하는 상황인자 값의 범위는 유사하면서 다양하게 발생될 수 있다. 이를 위해, 본 연구에서는 컴퓨터 환경 기반에서 복잡하고 유사한 범위의 데이터가 생성되도록 시뮬레이션 데이터를 생성시켰다. 시뮬레이션 데이터는 OLTP (Online Transaction Processing) 성능평가 표준안인 TPC-C와 Web-Commerce 성능평가의 표준안인 TPC-W 시스템 환경을 구현하여 시스템의 상태를 수집하므로써 구축되었다. 이 인스턴스 데이터는 총 120회의 독립적인 실험으로 약 2,000여개의 시뮬레이션 데이터를 생성하였다.

#### 4. 자율적 상황정보 분석을 위한 마이닝 기법의 개선 필요성

데이터 마이닝에서 예측 및 분석을 위해서는 기

준정보(클래스)가 포함되어 있는 분류(classification) 기법을 이용한다. 본 연구는 마이닝 기법 적용을 위해 공개 연구용 소프트웨어인 WEKA(버전 3.4)를 이용하며, WEKA에서 제공되는 분류기법인 J48 알고리즘을 적용하였다<sup>8)</sup>. J48 알고리즘은 C4.5 알고리즘을 개선하여 의사결정트리(Decision tree)로 결과 모델이 생성된다.

구축된 모델은 검증에 의해 활용여부가 결정되는데, 일반적으로 정확률이 90% 이상일 때 응용분야를 위해 활용될 수 있다. 본 논문은 CDMA, Zig-Bee, 시뮬레이션 데이터를 이용해 상황정보 모델을 구축하고 10교차 검증기법을 적용하여 모델 검증을 수행하였다.

수집된 CDMA 데이터에서 인스턴스 속성은 측정일자를 제거하고 최대전위, 최소전위, 평균전위이고, 기준정보(클래스)는 상태로 정상(Normal), 주의(Notice), 경고(Warning) 값을 가진다. 약 6,400여건의 CDMA 데이터를 기반으로 상황정보 분석 모델을 구축하면 Fig. 3과 같은 트리모델이 생성되었다.

10 교차 검증 기법에 의해 검증을 수행해보니 정확률이 99.98%임을 확인하였다. 이러한 정확률 수치는 구축된 상황정보 분석 모델이 향후 새 데이터가 발생되었을 때 분류나 예측을 위해 수행될 수 있음을 의미한다.

본 연구는 좀 더 복잡한 인스턴스 데이터 집합을 구성하기 위해 ZigBee 센서 네트워크에서 수집된 데이터를 이용하여 상황정보 모델을 구축하였다. 인스턴스 속성은 온도, 조도, 습도, 적외선으로 정의하였고 기준정보(클래스)는 실내, 실외, 지하의 각 센서노드에서 발생된 데이터의 위치를 식별하고 예측할 필요성이 있다는 가정아래 지하에 있는 10 번째 센서노드는 '지하\_10'같이 총 30개의 기준정보를 정의하였다. 생성된 상황정보 모델의 일부분을 Fig. 4에서 보이며 잎 노드(leaf node)만 206개로 복잡하고 판별하기 어려웠다. 10 교차 검증기법에 의해 검증된 모델의 정확률은 88.02%로 구축된 모델이 응용 환경을 위해 사용될 수 없음을 의미하고 있다.

```

avgMV <= -842
| maxMV <= -2565: Warning (160.0)
| maxMV > -2565: Normal (3803.0)
avgMV > -842
| avgMV <= -602: Notice (510.0)
| avgMV > -602
| | maxMV <= -5: Warning (1705.0)
| | maxMV > -5: Normal (26.0)
    
```

Fig. 3. CDMA context-aware model.



되는 인스턴스의 기준정보(클래스)로 할당된다. 예를 들어, Fig. 6과 같이 인스턴스 5는 군집 1로 군집화되었기 때문에 군집1은 인스턴스 5의 기준정보(클래스)가 된다.

재구축된 인스턴스 집합은 고유의 목적에 부합하도록 인스턴스 단위로 분류 및 추천되어야 한다. 그러나 기존 분류 알고리즘에 의한 결과는 군집단위로 분류되고 추천된다. 이러한 결과는 여러 개의 인스턴스들이 포함되어 복수의 시설상태나 환경을 결과로서 반환하기 때문에 세부적인 분석을 위한 결과로서는 부적합할 수 있다. 즉, 새로운 데이터가 입력되었을 때 가장 유사한 인스턴스를 군집에서 탐색하여 분류 및 예측 가능한 기법이 필요하다.

Fig. 7은 인스턴스 단위로 추천하기 위해 군집과 인스턴스간의 매핑정보를 도식화하여 보여준다. 특정환경의 노드위치까지 식별할 수 있는 서비스가 목적일 경우를 가정한 인스턴스 집합은 지하, 지상(실내, 실외)에서 각각 10개의 노드에서 수집된 데이터들이다. n개의 군집으로 기준정보가 생성되었을 때 대응되는 인스턴스의 정보를 예로서 보여준다.

유사한 인스턴스를 찾는 기법은 인스턴스간의 유사도 거리를 측정함으로써 계산할 수 있다. 일반적으로 유사도 측정공식은 유클리디안 공식과 맨하탄 공식이 많이 사용되었으나 다중 차원에서 특정 차원에 의해 발생하는 데이터 편향을 막기 위해 두 인스턴스간의 변이계수 응용 식을 유사도 거리 측정 공식으로 사용하는 것을 제안한다.

	속성 1	속성 2	...	속성 n	클래스	매우 환경
인스턴스 1	5.1	87	...	20	군집 1	Under_node_2
인스턴스 2	20.5	71.5	...	40	군집 2	Inner_node_8
인스턴스 3	27.5	24	...	3	군집 n	Outer_node_3
...	...	...	...	...	...	...
인스턴스 k	4.5	48	...	20	군집 1	Inner_node_2

Fig. 7. information mapping between cluster and instance.

유클리디안	$\sqrt{\sum_{i=1}^n (X_i - X_j)^2}$	새로운 인스턴스 집합	27.00, 94.00, 78.30, 24.00, 85000, 230000, 75000, 160900000, 35
맨하탄	$\sum  X_i - X_j $	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	$\frac{ X_i - X_j }{\sqrt{\frac{(X_i + X_j)}{2}}}$		
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

Fig. 8. similarity calculation procedures among instances.

Fig. 8은 유클리디안, 맨하탄, 변이계수 공식을 이용하여 인스턴스간의 거리를 계산하는 과정을 보여준다. 그림에서 보여주는 인스턴스는 총 9개의 차원을 가지고 있고, 각 차원은 다른값의 범위를 지니고 있다. 새롭게 입력되는 인스턴스 하나와 모델 구축에 사용된 기준정보(클래스)가 지정된 세 개의 인스턴스에 세 개의 거리 공식을 적용하여 거리값을 산출한다. 거리값은 낮을수록 입력된 인스턴스와 유사도가 높음을 의미한다.

유클리디안과 맨하탄 공식은 인스턴스 1, 인스턴스 2, 인스턴스 3 순으로 낮은 거리값이 계산되어 인스턴스 1을 입력 인스턴스와 가장 근접한 데이터로 추천한다. 유클리디안과 맨하탄 공식은 다른 범위를 가지는 각 차원들의 값을 동일하게 취급하여 값의 범위가 적은 전반부 4개 차원의 변경영향이 범위가 큰 후반부 차원의 변경보다 고려되지 못해서 인스턴스 1이 추천되었다. 변이계수는 정규화를 수행하여 값의 범위에 따라 발생하는 편향을 감소시켜 인스턴스 2가 추천되었다.

제안된 방식의 검증을 수행하기 위해 군집기법인 EM 알고리즘을 적용하여 5개의 군집을 기준정보(클래스) 생성하고, 분류모델을 구축하기 위해 인스턴스 단위로 분류/추천으로 수정된 J48알고리즘에 기준정보(클래스)를 활용하여 Fig. 8과 같이 상황 인식 모델을 구축하였다.

Fig. 9와 같이 새로운 수집정보(인스턴스)가 발생되었다고 가정했을 때 분류모델의 분류조건에 따라 추천되는 기준정보(클래스)는 class[2]이고 다섯 개의 인스턴스 집합들이 존재한다. 다섯 개의 인스턴스들은 변이계수 공식을 이용해 다섯 개의 인스턴스들 중 가장 낮은 거리값을 산출하고 그 결과 인스턴스 4를 선별한다.

```

DBLHR <= 80.93
| DBNCHECK <= 88057
| | DBRDS <= 154738896: class[2]
| | DBRDS > 154738896
| | | DBNCHECK <= 80734
| | | | ELAPSEDTIME <= 35: class[2]
| | | | ELAPSEDTIME > 35: class[1]
| | | DBNCHECK > 80734: class[1]
| DBNCHECK > 88057: class[1]
DBLHR > 80.93
| DBBHR <= 70.64
| | DBBHR <= 53.42: class[0]
| | DBBHR > 53.42
| | | ELAPSEDTIME <= 10: class[3]
| | | ELAPSEDTIME > 10: class[0]
| DBBHR > 70.64
| | ELAPSEDTIME <= 40: class[3]
| | ELAPSEDTIME > 40: class[4]
    
```

Fig. 9. The modified context-aware model.

이러한 모델을 10교차 검증기법으로 검증해보니 정확률이 97.8%를 보임으로써 이전보다 정확성이 향상되었다. 실제 가스 중첩 데이터를 가지고 개선된 알고리즘을 적용하여 구축된 상황인식 모델에서도 88.02%에서 96.2%로 정확률이 향상되었다.

## 6. 결론

본 연구는 가스안전관리를 위해 가스시설에서 수집되는 상황정보를 데이터 마이닝 기법을 적용하여 자율적으로 상황정보를 분석하는 토대 제시를 목적으로 한다.

이를 위해, 본 연구는 데이터 마이닝 기법을 적용하여 생성된 상황인식 모델의 정확성을 향상시키기 위해 데이터 축소 방식이 적용된 자율적 기준정보 생성과 인스턴스 단위 기반 분류 및 예측을 수행하도록 모델 구축에 관한 알고리즘을 변경하였다. 자율적 기준정보 생성을 위해서는 군집화 기법이 적용되었고, 인스턴스 단위로 분류결과를 추천하기 위해 유사도 공식으로 변이계수를 적용하였다.

가스안전관리의 상황정보 관리를 위해 개선된 알고리즘과 상황인식모델을 검증하기 위해 실제 데이터와 시뮬레이션 데이터를 수집 및 생성하여 10교차 검증기법을 수행한 결과 분류 및 예측에 관한 정확성이 향상됨을 알 수 있었다.

본 논문의 결과는 가스안전관리를 위한 자율적 정보 분석 서비스를 위한 초석을 제공하며, 향후 연구로는 제시된 알고리즘 이용한 정보 분석 응용 프로그램 개발을 수행할 예정이다.

**감사의 글 :** 본 연구는 지식경제부의 에너지기술혁신 프로그램으로 지원되었으며 이 논문은 “차세대에너지안전연구단”의 연구결과입니다(세부과제번호: 2007-M-CC23-P-03-01-000).

## 참고문헌

- 1) 박승민, “센서 네트워크 노드 플랫폼 및 운영체제 기술 동향”, 전자통신동향분석, 제21권, 제1호, 2006.
- 2) 한국전산원, “USN 기술동향 분석”, 2005.
- 3) S. Elnaffar, P. Martin, and R. Horman “Automatically Classifying Database Workloads”, CKIM Conference, pp. 622 ~ 624, 2002.
- 4) A. Ganek and T. Corbi, “The Dawning of the Autonomic Computing Era”, IBM Systems Journal, Vol. 42, No. 1, pp. 55 ~ 18, 2003.
- 5) D. Moorel, “Statistics Concepts and Controversies”, W. H. Freeman and Company, 2001.
- 6) M. Goeble and L. Gruenwald, “A Survey of Data Mining and Knowledge Discovery Software Tools”, ACM SIGKDD Explorations, Vol. 1, No. 1, pp. 20 ~ 33, 1999.
- 7) Oil & Natural Gas Projects(Transmission, Distribution, and Refining, NETL, 2005.
- 8) Weka Public Mining Software tool, [http://www.cs.waikato.ac.nz/weka]
- 9) G. Zangl and J. Hannerer, “Data Mining Applications in the Petroleum Industry”, IBM Round Oak Publishing, 2003.
- 10) B. Heile, “Emerging Standards: Where does ZigBee fit”, ZigBee Alliance, 2004.