기계경비시스템 오경보 이벤트 분석을 위한 데이터마이닝 기법 연구★

김종민* · 최경호** · 이동휘***

요 약

본 연구의 목적은 효율적인 기계경비시스템 오경보 이벤트 분석을 위해 가장 적합한 데이터마이닝 기법을 도출하는데 있다. 이를 위해 기계경비시스템 오경보의 발생원인을 살펴보고, 오경보 시의 출동건수, 오경보율 그리고 오경보원인의 통계자료를 토대로한 데이터를 데이터마이닝 프로그램인 WEKA에 맞게 변환시켜 여러 알고리즘에 적용 및 분석하였다. 본 논문에서는 적합한 데이터마이닝 기법을 찾기 위해 Decision Tree, Naive Bayes, BayesNet Apriori, J48Tree 알고리즘을 활용하였고, 분석을 통해 생성된 가장 높은 값을 도출하여 해당 알고리즘의 적용 가능성을 확인하였다. 이와 같은 연구를 통해 효율적으로 기계경비시스템의 오경보를 예측하고, 오경보에 대한 보다 효율적인 대처방안을 모색할 수 있음을 보여주었다.

A Study of Data Mining Methodology for Effective Analysis of False Alarm Event on Mechanical Security System

JongMin Kim* · KyongHo Choi** · DongHwi Lee***

Abstract

The objective of this study is to achieve the most optimal data mining for effective analysis of false alarm event on mechanical security system. To perform this, this study searches the cause of false alarm and suggests the data conversion and analysis methods to apply to several algorithm of WEKA, which is a data mining program, based on statistical data for the number of case on movement by false alarm, false alarm rate and cause of false alarm. Analysis methods are used to estimate false alarm and set more effective reaction for false alarm by applying several algorithm. To use the suitable data for effective analysis of false alarm event on mechanical security analysis this study uses Decision Tree, Naive Bayes, BayesNet Apriori and J48Tree algorithm, and applies the algorithm by deducting the highest value.

Key words : Data Mining, Mechanical Scurity System, False Alarm, WEKA, Algorithm

접수일(2011년 5월 01일), 수정일(1차: 2011년 5월 15일), 게재확정일(2011년 5월 21일)

★ 이 논문은 2012년도 기상산업 지원 및 활용기술 개발 사업 과제(KMIPA 2012-1202) 지원을 받아 수행 된 연구임.

* 경기대학교 산업보안학과

** 경기대학교 산업기술보호특화센터

*** University of Colorado Denver, Dept. of Computer Science and Engineering (교신저자)

1. 서 론

짧은 역사속에서의 기계경비업무는 1981년에 도입되어 서비스를 시작하였음에도 불구하고 오늘날까지 급속한 성장속도를 보이며 발전하고 있다. 그러나 이 과정에서 기계경비시스템의 오경보로 인해 발생하는 업체의 경제적 손실과 사용자 불편 및 보안업무의 공백이 발생하는 문제점도 함께 대두되고 있다. 따라서 기계경비시스템의 오경보 원인을 규명하고, 그 규명된 요인들을 추출 및 분석하여 향후 발생하는 오경보를 예측할 수 있다면 이러한 악영향을 최소화할 수 있다.

따라서 본 연구에서는 데이터마이닝 프로그램인 WEKA를 이용하여 기계경비시스템의 오경보 원인을 분석 및 예측할 수 있는 최적의 알고리즘을 도출해보고자 한다. 이를 위해 오경보 시의 출동건수, 오경보율 그리고 오경보 원인의 통계자료들을 데이터마이닝에 적용해보고, 효율적으로 기계경비시스템 오경보 이벤트를 분석할 수 있는 연관규칙과 분류기법을 제안하고자 한다.

2. 관련연구

2.1 기계경비시스템의 오경보

기계경비시스템이 이상을 감지하게 되면 현장 또는 출동요원이 이상이 발생한 장소를 확인하고 이상발생 원인에 따른 대처를 하게 된다. 그러나 사용자나 관리자의 실수, 시설장비의 결함, 환경변화에 대한 오경보, 사고 등과 연관이 없는 비정상적인 경보신호 즉, 오경보가 발생하는 경우에는 기계경비 관련회사의 비용적 부담, 고객의 대한 신뢰성 저하 등 부정적인 영향이 크다[1].

2.1.1 기계경비시스템의 오경보 원인

경비시스템의 오경보 원인은 크게 사용자의 실수, 경보기기의 하자, 설비・유지보수상의 하자, 자연적인 영향 등이 있다[2].

<표 1> 오경보의 주요 발생원인[3]

구분	오경보 발생원인
사용자의 실수	① 출입시 지연시간 초과 ② 사용자 조작실수
환경변화	① 급작스런 온도변화 ② 바람에 의한 감지 ③ 진동감지 ④ 안개감지
경비대상 시설의 물리적 변화	① 문 열림 ② 시설물 노후
동물 감지	① 쥐, 고양이 등 소동물 감지 ② 벌레감지
기기이상	① Noise나 서지 발생에 의한 감지기 작동 ② 전자파에 의한 감지기 작동 ③ 기기의 정상적 동작 불능
설계 실수	① 시큐리티 컨설턴트의 SP활동 시 부적합한 감지기 선정과 위치 선정에 의한 잦은 오경보 발생 ② 감지기 규격에 부적합한 형태의 경비설계
공사실수	① 규격에 부적합한 감지기 부착 ② 규정된 자재의 미사용으로 인한 이상 신호 발생
기타	① 취객이 셔터 문에 부딪히는 등 셔터 흔들림으로 인한 셔터 감지기 작동 ② 보수점검의 미비로 배선을 쥐가 갉아먹어 이상 신호 발생 등

<표 1>은 오경보의 주요 발생원인에 대해 나타낸 것이다. 급작스런 온도변화, 쥐, 고양이 등 소동물의 감지, 벌레 감지 등은 열선 감지기의 오경보를 발생시키며, 바람 또는 안개 감지 등은 적외선 감지기의 오경보를, 그리고 대형차량의 진동, 시설물 노후로 인한 흔들림은 진동 감지기의 오작동을 발생시킨다.

2.1.2 기계경비시스템의 오경보 발생 현황

KT텔레캅의 2008년 1분기 기준 경보발생 현황을 보게 되면 최고 49.71%로 고객의 사유로 인한 오경보가 가장 많이 나타났고, 그 뒤를 이어 오동작, 기타, 감지기정비, 시스템정비, 사고 순으로 오경보가 나타났다.

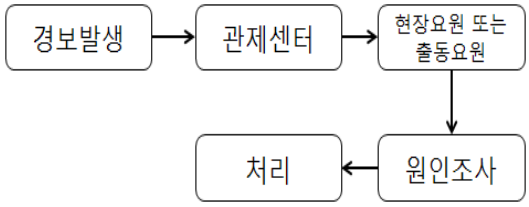
<표 2> KT텔레캅 경보발생 현황[5]

경보사유		건수	백분율
사고	도난, 강도, 파손, 화재 등	2,627	0.78
고객 사유	고객실수,해제실수, 내부점검 등	168,183	49.71
오동작	눈, 비, 바람, 동물 등	31,432	37.45
감지기 정비	감지기철거, 감지기추가	4,672	1.38
시스템 정비	배선 및 PCB, 카드리더 등	4,270	1.26
기타	익일A/S이첩, 모의출동 등	31,880	9.42

기계 및 시스템의 잘못된 설치문제로는 창문에 너무 가까이 설치하거나 적외선을 유동거리가 많은 지역에 설치, 리더기 설치가 출입문과 거리가 멀어서 경계 후 릴레이 타임이 짧아서 경보가 울리는 경우 등이 있다. 그리고 사용자 부주의로 인한 오작동으로는 난로의 사용으로 인한 열선감지가 되는 경우, 셔터문의 불량으로 인한 유격으로 자석감지기가 감지되는 경우 등이 있다[4].

2.1.3 오경보 발생시 대응절차

기계경비 시스템에 경보가 발생하면 관제센터에 알려지고 현장요원이 직접 점검 또는 출동요원이 출동하여 경보에 대한 원인 확인 후 경보에 대한 처리를 하게 된다.



(그림 1) 경보발생 시 출동 절차

2.2 데이터마이닝

데이터 마이닝은 대용량의 데이터로부터 수집한 새로운 정보나 유용한 패턴과 상관관계를 추출하여 의사 결정에 이용하는 작업이다. 데이터 마이닝의 활용 분야로는 기업의 의사결정, 고객관리, 데이터분석, 웹마이닝 등 다양하다[6].

2.2.1 데이터마이닝의 특징 및 기법

데이터마이닝은 몇 가지 관점에서 특징을 가지고 있는데 그 특징을 보면 다음과 같다[7].

(1) 대용량의 관측 가능한 자료를 다룬다.
(2) 컴퓨터 중심적 기법이다.
(3) 경험적 방법에 근거하고 있다.
(4) 일반화에 초점을 두고 있다.
(5) 기업의 다양한 경영상황 하에서 경쟁력 확보를 위한 의사결정을 지원하기 위해서 활용될 수 있다.
(6) 통계학, 컴퓨터과학, 인공지능, 공학과 같은 분야에서 개발된 특징을 가지고 있다.

데이터마이닝 기법을 활용한 연구현황은 <표 3>과 같다.

<표 3> 연구현황

연도	저자	제 목
2012	김성동	주식투자자의 의사결정 지원을 위한 데이터마이닝 도구[8]
2011	신현욱	데이터마이닝 기법을 이용한 지체장애인 취업결정요인 분석 연구[9]
2011	임준호	연관규칙을 이용한 의료 데이터 마이닝[10]
2011	박보은	데이터마이닝을 이용한 실시간 품질관리시스템 구축에 관한연구[11]
2010	이동규 외 1명	데이터마이닝을 활용한 건설 중대재해 분석[12]

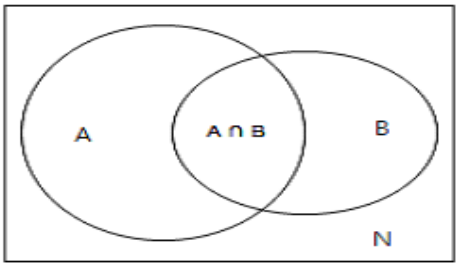
2.2.2 연관규칙 분석(Association Rule)

연관성 규칙발견의 목적은 항목들 간의 연관성의 정도를 수치로 정량화하는데 있다. 이를 위해서는 보통 '지지도'와 '신뢰도'라고 하는 두 가지 기준이 중요한 평가도구로 사용된다[13].

본 논문에서는 요인들 간의 연관성을 알아보기 위해 연관규칙 분석기법 중 Apriori알고리즘을 사용하고자 한다.

$$\text{지지도}(A \Rightarrow B) = \frac{\text{항목}A\text{와}B\text{를 동시에 포함하는 거래의수}}{\text{전체거래수}}$$

$$\text{신뢰도}(A \Rightarrow B) = \frac{\text{항목}A\text{와}B\text{를 동시에 포함하는 거래의수}}{\text{항목}A\text{를포함하는거래의수}}$$



$$Support(A \Rightarrow B) = \Pr(A \cap B) = \frac{n(A,B)}{N}$$

$$Confidence(A \Rightarrow B) = \Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{n(A,B)}{n(A)}$$

(그림 2) 연관성 규칙의 평가기준[13]

2.2.3 분류규칙 기법

분류란 중요한 데이터 클래스를 설명하는 모형을 생성하거나 미래 데이터의 경향을 예측하고자 할 때 사용되는 기법이다. 분류 기법에 대한 연구는 통계, 신경망, 결정트리 등의 분야에서 연구되었으며 여러분야에서 응용된다[14].

BayesNet는 베이스(T. Bayes)가 주창한 통계적 방법으로 불확실한 상황에서 통계적으로 얻은 정보를 토대로 어떤 의사결정을 해야 하는 경제학・경영학・자연자원 개발 등의 분야에서 보다 많이 사용되며 베이스의 정리는 다음과 같이 나타낼 수 있다[15].

$$P(A_j \setminus B_k) = \frac{P(B_k/A_j)P(A_j)}{\sum_{i=1}^{J} P(B_k/A_i)P(A_i)} \quad (1)$$

NaiveBayes는 확률에 근거한 문서 분류 방법으로 임의의 문서에 대해 확률을 계산하여 확률 중 가장 높은 것을 선택하는 것을 말한다. 이때 확률은 다음 식에 의하여 계산된다[16].

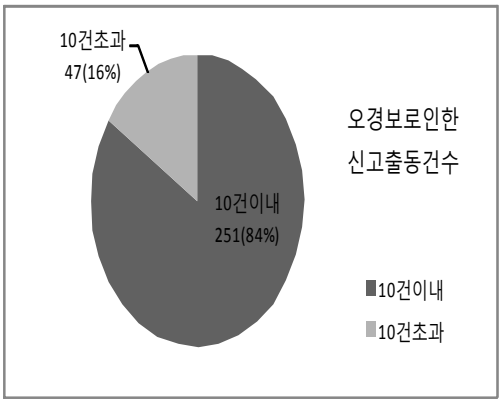
$$P(c_j|d_i) = \frac{P(c_j)P(d_i|c_j)}{P(d_i)} \quad (2)$$

3. 제안하는 방법

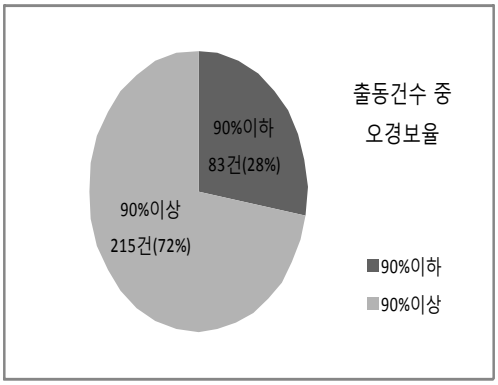
본 장에서는 오경보에 대한 데이터에 데이터마이닝의 기법인 연관규칙을 적용하기 위한 연구의 대상과 연구의 변수를 기술하도록 하겠다.

3.1 연구모형 대상

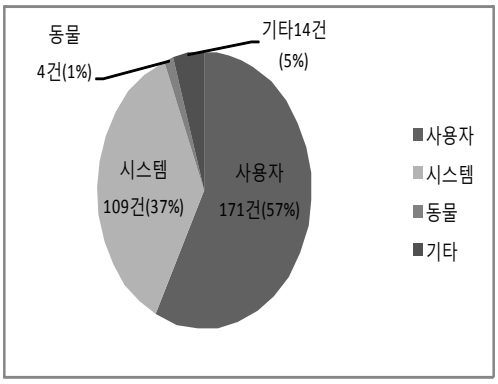
본 연구의 연관규칙 분석에 사용된 데이터는 "기계경비산업과 경찰의 효율적 협력 방안에 관한 연구"에서 사용된 통계자료를 사용하였고 본 연구에 맞도록 재구성 하여 사용하였다.



(그림 3) 오경보로 인한 신고 출동 건수[17]



(그림 4) 출동건수 중 오경보율[17]



(그림 5) 오경보 원인[17]

3.2 연구모형 변수

<표 4> 연관규칙 적용을 위한 리스트

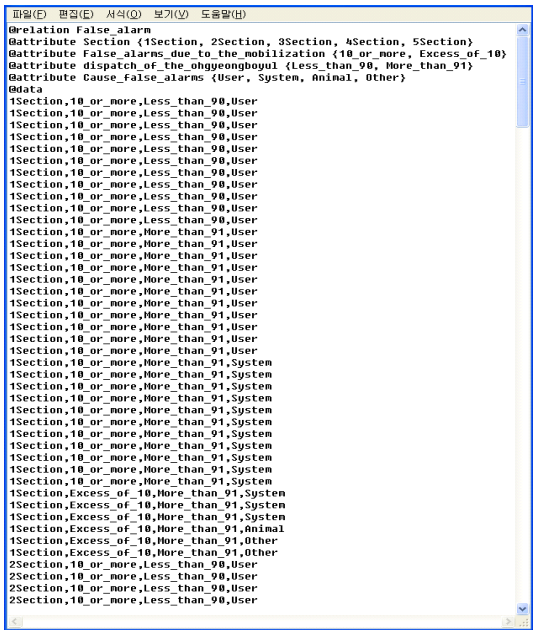
속 성	설 명
구 역	1구역, 2구역, 3구역, 4구역, 5구역
오경보 출동건수	10건이내, 10건초과
건수 중 오경보율	90%이하, 91%이상
오경보원인	사용자, 시스템결함, 동물, 기타

위의 <표 4>는 연관규칙에 사용될 변수 리스트를 정리한 것으로 이 리스트를 EXCEL로 정리한 것은 (그림 6)과 같다.

구역	오경보로인한 출동건수	출동건수중 오경보율	오경보 원인
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	Less_than_90	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	User
1Section	10_or_more	More_than_91	System
1Section	10_or_more	More_than_91	System
1Section	10_or_more	More_than_91	System
1Section	10_or_more	More_than_91	System
1Section	10_or_more	More_than_91	System

(그림 6) 변수데이터

위 파일들을 WEKA에서 분석하기 위해서는 ARFF형식으로 변경해 줘야한다. ARFF 형식으로 변형 시키면 (그림 7)과 같다.

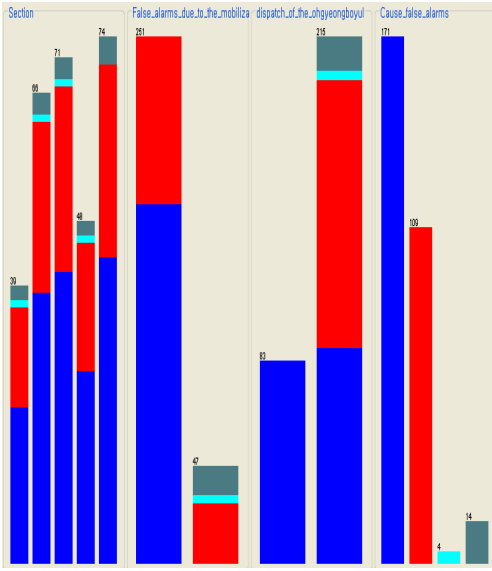


```
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
@relation False_alarm
@attribute Section {1Section, 2Section, 3Section, 4Section, 5Section}
@attribute False_alarms_due_to_the_mobilization {10_or_more, Excess_of_10}
@attribute dispatch_of_the_ohgyeongboyul {Less_than_90, More_than_91}
@attribute Cause_False_alarms {User, System, Animal, Other}
@data
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,Less_than_90,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,User
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,10_or_more,More_than_91,System
1Section,Excess_of_10,More_than_91,System
1Section,Excess_of_10,More_than_91,System
1Section,Excess_of_10,More_than_91,System
1Section,Excess_of_10,More_than_91,Animal
1Section,Excess_of_10,More_than_91,Other
1Section,Excess_of_10,More_than_91,Other
2Section,10_or_more,Less_than_90,User
2Section,10_or_more,Less_than_90,User
2Section,10_or_more,Less_than_90,User
2Section,10_or_more,Less_than_90,User
```

(그림 7) ARFF

3.3 분포형태

ARFF의 파일을 WEKA에서 불러오게 되면 (그림 8)과 같이 본 논문에서 사용된 변수의 분포형태를 나타내게 된다.



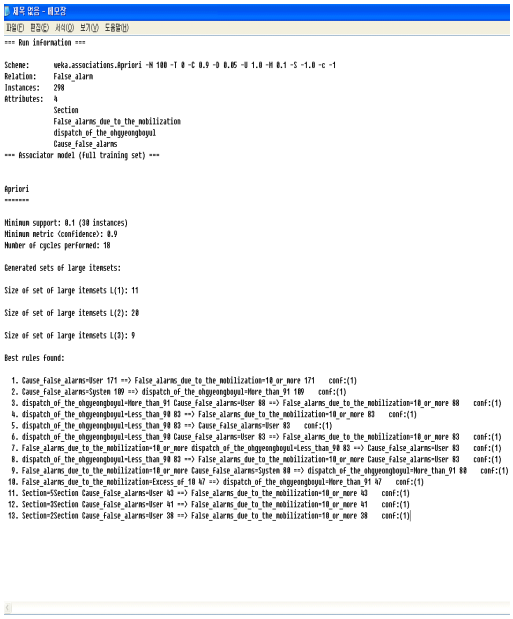
(그림 8) 변수의 분포 형태

4. 연구결과

본 장에서는 WEKA를 이용하여 데이터마이닝 기법인 연관성 분석을 활용하여 연관규칙 방법을 제시하고 연관규칙을 추출하고자 한다.

4.1 데이터마이닝 기법의 적용

연관성 탐사를 위해 3.2절의 ARFF데이터를 Apriori 알고리즘을 적용한다. WEKA의 Associate를 이용하여 “Apriori -N 1000 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1″을 이용하여 연관성 규칙을 적용하면 (그림 9)와 같은 Apriori 알고리즘의 결과를 보이게 된다.



```
새로 열은 - 테모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
=== Run information ===

Scheme:       weka.associations.Apriori -N 100 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     False_alarm
Instances:    290
Attributes:   4
              Section
              False_alarms_due_to_the_mobilization
              dispatch_of_the_ohgyeongboyul
              Cause_false_alarms
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.1 (30 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 20

Size of set of large itemsets L(3): 9

Best rules found:

 1. Cause_False_alarms=User 171 ==> False_alarms_due_to_the_mobilization=10_or_more 171    conf:(1)
 2. Cause_False_alarms=System 109 ==> dispatch_of_the_ohgyeongboyul=More_than_91 109    conf:(1)
 3. dispatch_of_the_ohgyeongboyul=More_than_91 Cause_False_alarms=User 80 ==> False_alarms_due_to_the_mobilization=10_or_more 80    conf:(1)
 4. dispatch_of_the_ohgyeongboyul=Less_than_90 83 ==> False_alarms_due_to_the_mobilization=10_or_more 83    conf:(1)
 5. dispatch_of_the_ohgyeongboyul=Less_than_90 83 ==> Cause_False_alarms=User 83    conf:(1)
 6. dispatch_of_the_ohgyeongboyul=Less_than_90 Cause_False_alarms=User 83 ==> False_alarms_due_to_the_mobilization=10_or_more 83    conf:(1)
 7. False_alarms_due_to_the_mobilization=10_or_more dispatch_of_the_ohgyeongboyul=Less_than_90 83 ==> Cause_False_alarms=User 83    conf:(1)
 8. dispatch_of_the_ohgyeongboyul=Less_than_90 83 ==> False_alarms_due_to_the_mobilization=10_or_more Cause_False_alarms=User 83    conf:(1)
 9. False_alarms_due_to_the_mobilization=10_or_more Cause_False_alarms=System 80 ==> dispatch_of_the_ohgyeongboyul=More_than_91 80    conf:(1)
10. False_alarms_due_to_the_mobilization=Excess_of_10 47 ==> dispatch_of_the_ohgyeongboyul=More_than_91 47    conf:(1)
11. Section=5Section Cause_False_alarms=User 43 ==> False_alarms_due_to_the_mobilization=10_or_more 43    conf:(1)
12. Section=3Section Cause_False_alarms=User 41 ==> False_alarms_due_to_the_mobilization=10_or_more 41    conf:(1)
13. Section=2Section Cause_False_alarms=User 38 ==> False_alarms_due_to_the_mobilization=10_or_more 38    conf:(1)
```

(그림 9) Apriori 알고리즘 결과

(그림 10)은 WEKA의 Classify를 이용하여 “J48-C 0.25 -M 2”이용하여 나타난 결과이다.

```
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     False_alarm
Instances:    298
Attributes:   4
              Section
              False_alarms_due_to_the_mobilization
              dispatch_of_the_ohgyeongboyul
              Cause_false_alarms
Test mode:10-fold cross-validation

--- Classifier model (Full training set) ---

J48 pruned tree
------------------
False_alarms_due_to_the_mobilization = 10_or_more: User (251.0/80.0)
False_alarms_due_to_the_mobilization = Excess_of_10: System (47.0/18.0)

Number of Leaves  :     2

Size of the tree :      3

Time taken to build model: 0 seconds

--- Stratified cross-validation ---
--- Summary ---

Correctly Classified Instances         200               67.1141 %
Incorrectly Classified Instances        98               32.8859 %
Kappa statistic                          0.2835
Mean absolute error                      0.2248
Root mean squared error                  0.3358
Relative absolute error                 83.618  %
Root relative squared error             91.8471 %
Total Number of Instances              298

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   1         0.63       0.681     1         0.81       0.653    User
                   0.266     0.095      0.617     0.266     0.372      0.547    System
                   0         0          0         0         0          0.882    Animal
                   0         0          0         0         0          0.919    Other
Weighted Avg.      0.671     0.396      0.617     0.671     0.601      0.63

=== Confusion Matrix ===
   a   b   c   d   <-- classified as
 171   0   0   0 |   a = User
  80  29   0   0 |   b = System
   0   4   0   0 |   c = Animal
   0  14   0   0 |   d = Other
```

(그림 10) J48 cross-validation

본 논문에서 적용될 분류 기법들에 대한 오경보의 예측 결과 평가를 위한 지표로는 TP(True Positive)와 FP(False Positive), FN(False Negative), 그리고 Recall 및 Precisiondmf 이용하였다. TP의 의미는 분류하고자 하는 클래스를 정확하게 그 클래스로 분류했을 경우를 나타내고 FP는 분류하고자 하는 클래스가 아닌데도 그 클래스로 분류했을 경우의 오류, 즉 양성-오류를 의미한다. Recall과 Precision의 식을 보면 다음과 같다[14].

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} \tag{4}$$

각 알고리즘의 분류 결과에 대한 결과는 다음과 같다.

먼저 Decision Table의 분류기법을 보게 되면 WEKA의 Classify를 이용해 Decision Table -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"를 이용한 결과는 (그림 11)과 같다.

```
=== Summary ===

Correctly Classified Instances         200               67.1141 %
Incorrectly Classified Instances        98               32.8859 %
Kappa statistic                          0.2835
Mean absolute error                      0.2287
Root mean squared error                  0.3359
Relative absolute error                 85.0338 %
Root relative squared error             91.8662 %
Total Number of Instances              298

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   1         0.63       0.681     1         0.81       0.654    User
                   0.266     0.095      0.617     0.266     0.372      0.549    System
                   0         0          0         0         0          0.882    Animal
                   0         0          0         0         0          0.923    Other
Weighted Avg.      0.671     0.396      0.617     0.671     0.601      0.632
```

(그림 11) Decision Table

다음으로 Naive Bayes의 분류기법을 보게되면 WEKA의 Classify를 이용하여 "Naive Bayes"를 적용한 결과는 (그림 12)와 같다.

```
=== Summary ===

Correctly Classified Instances         178               59.7315 %
Incorrectly Classified Instances       120               40.2685 %
Kappa statistic                          0.161
Mean absolute error                      0.1894
Root mean squared error                  0.3101
Relative absolute error                 70.4416 %
Root relative squared error             84.8187 %
Total Number of Instances              298

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   0.848     0.598      0.656     0.848     0.74       0.766    User
                   0.303     0.233      0.429     0.303     0.355      0.662    System
                   0         0          0         0         0          0.889    Animal
                   0         0          0         0         0          0.904    Other
Weighted Avg.      0.597     0.429      0.533     0.597     0.554      0.736
```

(그림 12) Naive Bayes

(그림 13)을 보면 BayesNet의 분류 기법으로 WEKA의 Classify를 이용해 "BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifilers.bayes.net.estimate.SimpleEstimator -- -A 0.5"를 이용한 결과이다.

```
=== Summary ===

Correctly Classified Instances        178               59.7315 %
Incorrectly Classified Instances      120               40.2685 %
Kappa statistic                         0.161
Mean absolute error                     0.1883
Root mean squared error                 0.311
Relative absolute error                70.0107 %
Root relative squared error            85.0451 %
Total Number of Instances             298

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.848     0.598      0.656     0.848     0.74       0.765    User
                 0.303     0.233      0.429     0.303     0.355      0.66     System
                 0         0          0         0         0          0.89     Animal
                 0         0          0         0         0          0.905    Other
Weighted Avg.    0.597     0.429      0.533     0.597     0.554      0.735
```

(그림 13) BayesNet

적용된 분류 기법은 Decision Table, Naive Bayes, BayesNet 그리고 연관 분류 기법으로 J48Tree 알고리즘을 적용하였다. <표 5>는 알고리즘들의 적용한 값을 정리한 것이다.

<표 5> 분류 기법별 평가 결과

Classifier	TP	FP	Precision	Recall	Class
Decision Table	1	0.63	0.681	1	User
	0.266	0.095	0.617	0.266	System
	0	0	0	0	Animal
	0	0	0	0	Other
Naive Bayes	0.848	0.598	0.656	0.848	User
	0.303	0.233	0.429	0.303	System
	0	0	0	0	Animal
	0	0	0	0	Other
Bayes Net	0.848	0.598	0.656	0.848	User
	0.303	0.233	0.429	0.303	System
	0	0	0	0	Animal
	0	0	0	0	Other
J48 Tree	1	0.63	0.681	1	User
	0.266	0.095	0.617	0.266	System
	0	0	0	0	Animal
	0	0	0	0	Other

위의 표에서 분류 결과에 대한 결과를 보게 되면 가장 높은 값을 보인 것은 Decision Table과 J48 Tree 둘 다 같은 값을 나타냈으나 Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error에서 Decision Table보다 J48Tree가 근소한 차이로 J48Tree가 낮게 측정됨으로서 J48 Tree의 알고리즘을 적용하기가 가장 적합하다고 말할 수 있다.

4.2 데이터마이닝분석에 의한 오경보의 효율적인 대처방안

현재의 경보에 대한 대처는 실경보이든 오경보이든 모두 출동을 함으로써, 오경보가 울렸을 시 요원들의 불필요한 출동으로 경비업체의 원가상승 등 경비업체의 막대한 손실을 주고 있다.

본 연구에서 발견한 오경보와 요인들간의 연관규칙을 이용한 연관성 규칙과 분류규칙을 이용한 오경보에 대한 예측을 적절히 활용하여 오경보에 대한 불필요한 출동을 줄여야 할 것이다. 오경보에 대한 연관성 규칙과 분류규칙에 따른 예측이 반드시 그 규칙에 따라 발생했다는 의미는 아니다. 하지만 오경보일 가능성이 높다는 것을 예상할 수 있도록 해준다. 따라서 이와 같은 연관성 규칙을 기반으로 오경보에 대한 출동을 최소화하면서 경비업체의 불필요한 출동을 방지하고 손실을 최소화할 수 있다.

5. 결론

본 논문에서는 기계경비시스템의 오경보에 대한 요인을 구역마다 오경보에 대한 출동건수와 그 건수에 대한 오경보 비율, 오경보 원인으로 설정하여 요인들을 데이터마이닝에 적용시켜 연관규칙을 통해 어떠한 연관규칙들이 있는지 연관성을 찾을 수 있었다. 데이터마이닝 프로그램인 WEKA의 여러 가지 알고리즘들 중 Decision Table, Naive Bayes, BayesNet 그리고 J48Tree 알고리즘을 적용하여 알고리즘들의 값을 구하였다. 그 알고리즘들 중 가장 높은 측정 값은 J48 Tree알고리즘으로 나타났고 그 다음 순서로 Decision Table, BayesNet, Naive Bayes 순으로 나타났다. 이 결과로 기계경비 오경보 분석에 대해 가장 적합한 데이터마이닝 기법은 J48Tree 알고리즘이 가장 적합하다 할 수 있다. 이 결과를 통해 보다 효율적인 오경보에 대한 대처방안을 제시였다는 점에서 연구의 성과

를 얻을 수 있었다.

향후에는 오경보의 축적된 대규모 데이터를 이용하여 의미 있는 분석을 수행하여야 할 것이고 전국을 대상으로 많은 요인들을 추출하여 여러 가지 데이터 마이닝 알고리즘을 적용함으로써 보다 더 적합한 알고리즘의 도출 및 높은 신뢰도를 확보하여야 할 것이다.

참고문헌

[1] 주일엽, , 「기계경비개론」, 진영사, p. 69, 2010.

[2] 김숙진, “무인기계경비시스템의 오경보 대응방안에 관한 연구”, 동국대학교 행정대학원 석사학위논문, pp. 107-109, 2004.

[3] 김태민, 이동제, “시스템 경비 산업의 문제점 및 개선방안”, 한국치안행정학회, p. 57, 2008.

[4] 하승규, “기계경비업의 현황과 발전방향, 연세대학교 법무대학원 석사학위논문 2008.

[5] KT텔레캅 통계자료, 2008.

[6] 임준호, “연관규칙을 이용한 의료 데이터 마이닝”, 고려대학교 석사학위논문, 2011.

[7] 강현철 외, 「데이터마이닝 방법론」, 자유아카데미, 2006.

[8] 김성동, “주식투자자의 의사결정 지원을 위한 데이터 마이닝 도구”, 한국콘텐츠학회논문지, 2012.

[9] 신현욱, “데이터마이닝 기법을 이용한 지체장애인 취업결정요인 분석 연구”, 한국장애인고용촉진공단, 2011.

[10] 임준호, “연관규칙을 이용한 의료 데이터마이닝”, 고려대학교 정보통신대학원 석사학위논문, 2011.

[11] 박보은, “데이터마이닝을 이용한 실시간 품질관리시스템 구축에 관한 연구“, 전북대학교 일반대학원 석사학위논문, 2011.

[12] 이동규 외 1명, “데이터마이닝을 활용한 건설중대재해 분석”, 대한건축학회, 2010.

[13] 최종후 외, 「데이터마이닝 예측 및 활용」, SPSS아카데미, 2002.

[14] 노기용, 이현규, 류근호, “심장 질환 진단을 위한 데이터 마이닝 기법”, 감성과학, vol. 10, No.2, p. 153, 2007.

[15] http://terms.naver.com/entry.nhn?docId=510787, 2012. 04. 23

[16] 이재문, “빈발단어집합을 이용한 Naive Bayes의 정확도 개선, 한국인터넷정보학회지, 2006.

[17] 강영일, 기계경비산업과 경찰의 효율적 협력 방안에 관한 연구, 연세대학교 행정대학원, 석사학위논문, 2007.

[저 자 소 개]

김 종 민 (Jong-Min Kim)

2012년 현재 경기대학교
산업보안학과 박사과정

email : dyuo1004@gmail.com

최 경 호 (Kyong-Ho Choi)

2002년 경기대학교 경제학사
2005년 경기대학교 경제학석사
2008년 경기대학교 정보보호학박사
2012년 경기대학교 연구교수
(산업기술보호특화센터)

email : cyberckh@gmail.com

이 동 휘 (Dong-Hwi Lee)

2000년 경기대학교 컴퓨터과학과
(이학사)
2003년 경기대학교 정보보호기술공학과
(공학석사)
2006년 경기대학교 정보보호학과
(정보보호학박사)
2011년~현재 University of Colorado Denver, Dept, of Computer Science and Engineering

email: dhclub@naver.com