

## 제조업에서의 산업재해 예방을 위한 전문가 시스템 개발

- Development of an Expert System for Prevention  
of Industrial Accidents in Manufacturing Industries -

임 영 문 \*

Leem Young Moon

최 요 한 \*\*

Choi Yo Han

### Abstract

Many researches and analyses have been focused on industrial accidents in order to predict and reduce them. As a similar endeavor, this paper is to develop an expert system for prevention of industrial accidents. Although various previous studies have been performed to prevent industrial accidents, these studies only provide managerial and educational policies using frequency analysis and comparative analysis based on data from past industrial accidents. As an initial step for the purpose of this study, this paper provides a comparative analysis of 4 kinds of algorithms including CHAID, CART, C4.5, and QUEST. Decision tree algorithm is utilized to predict results using objective and quantified data as a typical technique of data mining. Enterprise Miner of SAS and Answer Tree of SPSS will be used to evaluate the validity of the results of the four algorithms. The sample for this work was chosen from 10,536 data related to manufacturing industries during three years(2002~2004) in Korea. The initial sample includes a range of different businesses including the construction and manufacturing industries, which are typically vulnerable to industrial accidents.

**Keywords :** Data Mining, Decision Tree, CHAID, CART, C4.5, QUEST, Expert System

† 본 연구는 산업자원부의 지역혁신 인력양성사업의 연구결과로 수행되었음.

\* 강릉대학교 산업시스템공학과 교수

\*\* 강릉대학교 산업시스템공학과 박사과정 수료

2005년 12월접수; 2006년 2월 수정본 접수; 2006년 2월 게재 확정

## 1. 서론

급속하게 발전되는 고도의 산업화와 다양하게 술현하는 신종의 업종들, 이러한 배경 뒤에는 산업현장에서 본인의 의지와 상관없이 사고를 당하여 재해를 입는 경우가 빈번하게 발생되고 있는 현실이다. 예상치 못한 사고는 때와 장소 등에 관계없이 불특정 다수에게 언제나 발생할 수 있으며, 또한 사고로 인한 재해는 인적, 물적 피해를 발생시키고 있다. 이러한 사고를 방지하기 위한 최선의 방법은 사고 발생 위험 요소를 사전에 제거하는 것이며, 차선의 방법으로는 사고의 예방이다. 그동안 산업재해 예방을 위하여 많은 선행연구[7][8]가 진행되어 왔으며, 많은 양의 데이터[4]가 축적되었다. 그러나 산업재해와 관련된 연구의 대부분은 과거에 발생한 재해에 대하여 단순히 빈도 분석이나 비교 분석을 실시한 결과를 토대로 예방대책을 제시하는 경우가 대부분이었다. 급속하게 변화하는 현실에 있어 대량으로 축적된 데이터는 매우 중요한 재산이며, 대량의 데이터를 분석하여 의미 있는 정보를 찾아내는 것은 미래를 예측할 수 있는 객관적인 방법이 될 수 있을 것이다. 대량의 데이터를 분석하기 위한 기존의 데이터 분석방법은 여러 면에서 많은 한계를 나타내고 있다. 이에 대용량의 데이터를 효과적으로 분석하기 위하여 데이터마이닝(Data Mining)이라는 분야가 부각되어 여러 분야에서 적용[1][2][13][14][15]되고 있으며, 분석된 결과도 다양하게 활용되고 있다. 본 연구에서는 제조업 분야에서 산업재해 예방을 위하여 선행 연구들의 대부분이 산업재해 데이터를 단순히 비교 분석, 빈도 분석 등을 실시한 후 그 결과를 토대로 관리적, 기술적, 교육적 등의 재해예방 대책을 제시한 것과 달리 기존에 발생된 제조업에 관련된 산업재해 데이터를 데이터마이닝의 의사결정나무 기법을 적용하여 알고리즘을 비교 분석한 후 최적 알고리즘을 선정하고자 한다. 또한 선정된 알고리즘을 토대로 노드(Node) 분석 및 데이터베이스(Data Base)를 구축한 후 산업재해 예방을 위하여 정량적, 정성적 예측이 가능한 전문가 시스템(Expert System)[6]을 개발하고자 한다.

## 2. 연구방법

본 연구에서는 강원도 관내 전업종에서 2002년부터 2004년까지 3년간 산업재해 신청을 하여 산재로 결정 통지된 67,278건의 데이터 중 제조업의 10,536건의 데이터를 분석하기 위하여 SAS의 Enterprise Miner[2]와 SPSS의 AnswerTree[1] 소프트웨어를 이용한다. 먼저 데이터마이닝의 기법 중에서 의사결정나무의 대표적인 알고리즘인[5] CHAID[3], CART[11], QUEST[3], C4.5[13]를 각각 제조업에서 발생된 산재 데이터에 적용하여 타당성 평가를 한 후 알고리즘별 정확도, 민감도, 특이도 값을 구하여 비교

분석한다. 또한 이를 토대로 최적의 알고리즘을 선택하여 제조업의 데이터에 대하여 교차타당성(Cross Validation)을 이용한 타당성 평가를 실시한다. 그리고 선택된 최적의 알고리즘을 실행하여 나타난 노드를 분석한 후 데이터테이블을 작성하여 데이터베이스를 구축한다. 이러한 과정들을 토대로 산업재해 예방을 위하여 정량적, 정성적 예측이 가능한 전문가 시스템을 델파이(Delphi) 5.0[9]을 이용하여 개발하고자 한다.

### 3. 데이터 분류 및 분석방법

본 연구의 주된 목적은 기존에 발생된 제조업 관련 산업재해 데이터를 분석하여, 작업조건별로 어떤 유형의 산업재해가 가장 많이 발생되는가를 예측하고, 예측된 재해의 정성적인 위험지수, 또한 예측 모형을 통하여 산업재해 전문가 시스템을 개발하는 것이다. 본 연구에 사용된 원시 데이터는 재해일자, 재해자 구분, 발생형태, 업종, 규모, 직종, 진료 일수, 입원 일수, 통원 일수, 연령, 성별, 요양 기간, 근속기간, 재해월, 재해요일, 재해시간, 근로 손실 일수의 항목 등으로 구성되어 있다. 또한 각 항목들은 세부적인 형태로 분류가 되어있다.

전체 항목들 중 분석에 불필요한 항목들을 제외한 총 9개의 항목을 데이터 분석을 위한 변수로 선택하였다. 목표변수로는 재해자 구분을 재해형태라는 변수 명으로 정의하여 선택하였고, 예측변수로는 발생형태, 규모, 연령, 성별, 근속기간, 재해월, 재해요일, 재해시간으로 구성하였다. 이들 변수들 중에서 재해형태와 연령은 이산형(Binary) 변수로, 발생형태, 재해월, 재해요일, 재해시간은 명목형(Nominal) 변수로, 규모, 연령, 근속기간은 연속형(Interval) 변수로 구성되었으며, 변수들에 대한 설명은 다음의 <표 1>과 같다. 또한 전체 10개의 업종 중에서 제조업을 샘플링으로 선택하여 총 67,278건의 원시 데이터 중에서 10,536건의 데이터를 관찰치로 사용하였다. 이러한 데이터를 SAS의 Enterprise Miner와 SPSS의 AnswerTree를 이용하여 의사결정나무 분석의 대표적인 알고리즘인 CHAID, CART, C4.5, QUEST를 통하여 비교 분석한 후 최적의 알고리즘을 선택하고자 한다. 그리고 제조업 데이터에 대하여 선택된 최적의 알고리즘을 이용하여 노드 분석을 수행하였다

<표 1> 데이터 분석에 사용된 변수

변수명	변수설명	변수값(코딩형식)	변수유형
재해형태	재해자구분	부상=1, 사망=2	이산형(Binary)
발생형태	재해 발생 형태 구분	감김, 끼임=1, 교통사고=2, 낙하 비래=3, 무리한 동작=4, 전도=5, 절단, 베임=6, 직업병=7, 추락=8, 충돌=9	명목형(Nominal)
규모	사업장 근로자수	5인 미만=1, 5이상-9미만=2, 10이상-15미만=3, 16이상-29미만=4, 30이상-49미만=5, 50이상-99미만=6, 100이상-199미만=7, 200이상-299미만=8, 300이상-499미만=9, 500이상-999미만=10, 1000 이상-1999미만=11, 2000이상=12	연속형(Interval)
연령	재해자 연령구분	18-24세=1, 25-29세=2, 30-34세=3, 35-39세=4, 40-44세=5, 45-49세=6, 50-54세=7, 55-59세=8, 60세이상=9	연속형(Interval)
성별	성별 구분	남자=1, 여자=2	이산형(Binary)
근속기간	재해자 근속기간	6월 미만=1, 6월 이상-1년 미만=2, 1년 이상-2년 미만=3, 2년 이상-3년 미만=4, 3년 이상-4년 미만=5, 4년 이상-5년 미만=6, 5년 이상-10년 미만=7, 10년 이상-20년 미만=8, 20년이상=9	연속형(Interval)
재해월	재해 발생월 구분	1월=1, 2월=2, 3월=3, 4월=4, 5월=5, 6월=6, 7월=7, 8월=8, 9월=9, 10월=10, 11월=11, 12월=12	명목형(Nominal)
재해요일	재해 발생요일 구분	월요일=1, 화요일=2, 수요일=3, 목요일=4, 금요일=5, 토요일=6, 일요일=7	명목형(Nominal)
재해시간	재해 발생시간 구분	0-2시=1, 2-4시=2, 4-6시=3, 6-8시=4, 8-10시=5, 10-12시=6, 12-14시=7, 14-16시=8, 16-18시=9, 18-20시=10, 20-22시=11, 22-24시=12	명목형(Nominal)

### 4. 알고리즘 비교분석

의사결정나무 분석의 대표적인 알고리즘인 CHAID, CART, C4.5, QUEST를 비교 분석한 후 최적의 모형을 선택하기 위하여 각 알고리즘별 정분류율 또는 정확도 (Accuracy), 민감도(Sensitivity), 특이도(Specificity) 값을 구하여 비교 분석하였다. 오분류표(Confusion Matrix)란 목표변수의 실제 범주와 모형에 의해 예측된 분류범주 사이의 관계를 나타내는 것으로 다음의 <그림 1>과 같다.

		예측된 변수		
		0	1	
원래 목표 변수	0	실제0 예측0	실제0 예측1	실제0
	1	실제1 예측0	실제1 예측1	실제1
		예측0	예측1	

<그림 1> 오분류표의 구성

위의 그림을 참조로 하여 정분류율 또는 정확도, 오분류율 또는 오차율(Error Rate), 민감도, 특이도의 개념을 다음과 같이 정의할 수 있다[2].

$$\text{정분류율(Accuracy)} = \frac{(\text{실제0} , \text{예측0})\text{의 빈도} + (\text{실제1} , \text{예측1})\text{의 빈도}}{\text{전체빈도}}$$

$$\text{오분류율(Error Rate)} = \frac{(\text{실제0} , \text{예측1})\text{의 빈도} + (\text{실제1} , \text{예측0})\text{의 빈도}}{\text{전체빈도}}$$

$$\text{민감도(Sensitivity)} = \frac{(\text{실제1} , \text{예측1})\text{인 관찰치의 빈도}}{\text{실제 1인 관찰치의 빈도}}$$

$$\text{특이도(Specificity)} = \frac{(\text{실제0} , \text{예측0})\text{인 관찰치의 빈도}}{\text{실제 0인 관찰치의 빈도}}$$

위의 정의를 토대로 본 연구에 사용된 제조업에 관련된 데이터 집합에 대하여 Training set(분석용 데이터)과 Testing set(평가용 데이터)을 50 : 50의 비율로 하여 SAS의 Enterprise Miner와 SPSS의 AnswerTree에서 알고리즘에 대한 분석을 실행하였다. 이러한 과정을 통하여 알고리즘을 비교 분석한 결과 다음의 <표 2>와 같은 값을 얻을 수 있었다.

<표 2> 제조업에 대한 알고리즘 비교

Algorithm	Training set			Testing set		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
CHAID	99.35435	99.70925	82.24299	99.29791	99.63135	84.48276
CART	95.69833	96.32188	64.48598	98.86299	99.86437	54.31034
QUEST	99.14546	99.86431	64.48598	98.95932	99.70981	65.51724
C4.5	97.26651	95.32710	97.30672	96.96279	87.93103	97.16615

위의 <표 2> 제조업에 대한 알고리즘을 비교해 보면 정확도는 CHAID가 Training set과 Testing set에서 99.35435%와 99.29791%로 가장 높은 값으로 나타났고, 민감도는 Training set에서 QUEST가 99.8643%, Testing set에서는 CART가 99.86437%로 가장 높은 값을 나타냈다. 특이도는 C4.5가 Training set과 Testing set에서 97.30672%와 97.16615%로 가장 높은 값을 나타냈으나, 정확도, 민감도, 특이도 전체를 비교하여 볼 때 전반적으로 CHAID가 가장 높은 값을 나타냄으로 제조업관련 데이터를 분석함에 있어서 CHAID가 최적의 알고리즘임을 알 수 있다. 또한 CHAID 알고리즘에 대한 교차타당성(Cross Validation)을 분석하여 보면, <그림 2>에서 볼 수 있듯이 모형구축 오분류 값이 0.0603426, 교차타당성 값이 0.0774704로 나타났다. 일반적으로 이론에 따르면 모형구축 오분류 값과 교차타당성 값의 차이가 10%이내이면 분류된 데이터들의 결과에 대한 타당성을 줄 수 있는 것으로 알려져 있는데, <그림 2>에서 나타난 두 값의 차이는 2%이내의 차이를 나타내고 있어서 본 연구에서 분류된 데이터들은 타당성을 갖고 있다고 할 수 있다.

Misclassification Matrix					
		Actual Category			Total
		2	1		
Predicted Category	2	141	2		167
	1	203	3		3628
	Total	344	3		3795
		Resubstitution Cross-Validation			
Risk Estimate		0.0603426	0.0774704		
SE of Risk Estimate		0.00386537	0.00433962		

<그림 2> 제조업 데이터에 대한 교차타당성(Cross Validation)

이러한 제조업에 관련된 데이터들에 대하여 CHAID, CART, C4.5, QUEST와 같이 네 개의 알고리즘을 비교 분석 한 후 선정된 최적 알고리즘 CHAID에 대하여 전 업종 중 제조업을 선택하여 노드 분석을 실시하였다. 그 결과를 바탕으로 산업재해 예방을 위한 전문가 시스템 개발을 위하여 데이터 테이블을 구성하고자 한다.

## 5. CHAID 알고리즘 분석 결과

SAS의 Enterprise Miner와 SPSS의 AnswerTree를 이용하여 알고리즘별 정확도, 민감도, 특이도 값을 구하여 비교 분석한 결과 선택된 최적의 알고리즘인 CHAID에 대하여 카이제곱 통계량(Chi-Square Statistic)을 구해본 결과 다음의 <그림 3>과 같이 나타났다.

Predictor	Nodes	Split Type	Chi-Square	D,F	Adj.Prob.
재해월	3	Default	127,1009	2	0,0000
규모	4	Default	283,4431	3	0,0000
근속기간	3	Default	109,5113	2	0,0000
재해시간	5	Default	323,1750	4	0,0000
재해요일	3	Default	125,2576	2	0,0000
발생형태	9	Arbitrary	*	*	*

<그림 3> CHAID 알고리즘의 카이제곱 통계량

위의 <그림 3>을 살펴보면 예측변수인 재해월, 규모, 근속기간, 재해시간, 재해요일에 대한 Chi-Square 값은 각각 127.1009, 283.4431, 109.5113, 323.1750, 125.2576으로 높은 적합성을 나타냈다. 이에 본 연구의 주목적인 산업재해 예방을 위한 전문가 시스템 개발을 위하여 변수의 입력순서를 목표변수는 발생형태로 하고, 예측변수는 재해월, 재해요일, 재해시간, 규모, 근속기간, 발생형태 순으로 정하였다. 이러한 과정을 거친 후, SPSS의 AnswerTree를 이용하여 CHAID 알고리즘을 실행한 결과를 요약하면, 뿌리마디는 10,536개의 관측치로 부상 10,313개(97.88%)와 사망 223개(2.12)%로 나타났고, 깊이(Depth)는 7이며, 끝마디는 총 124개로 형성되었다. 변수로는 부상 또는 사망을 나타내는 재해형태, 재해월, 재해요일, 재해시간, 규모, 근속기간, 발생형태가 선택되었다.

CHAID 알고리즘을 이용한 노드 분석 결과 재해형태, 재해월, 재해요일, 재해시간, 규모, 근속기간, 발생형태의 7가지 변수를 모두 만족하는 노드는 전체 124개의 끝마디 중 41개의 끝마디로 나타났다. 이러한 과정을 토대로 산업재해 예방을 위한 전문가 시스템을 개발하고자 <그림 4>와 같이 데이터 테이블을 구성하였다.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	연번	월	요일	시간	경우수	규모	근속기간	발생형태	부상n	사망n	all	부상%	사망%
2	1	3	월(1)	0-2(1)	1	2,3,4,5	1	5	193	7	19574	0.986	0.036
3	2	3	월(1)	0-2(1)	2	2,3,4,5	1	8	293	22	19574	1.497	0.112
4	3	3	월(1)	0-2(1)	3	2,3,4,5	1	1,2,3,4,6,9	551	0	19574	2.815	0.000
5	4	3	월(1)	0-2(1)	4	2,3,4,5	1	7	4	5	19574	0.020	0.026
6	5	3	월(1)	0-2(1)	5	6,7	1,4,7	1,3,4,5,7,9	122	0	19574	0.623	0.000
7	6	3	월(1)	0-2(1)	6	6,7	1,4,7	8	199	15	19574	1.017	0.077

<그림 4> 데이터 테이블

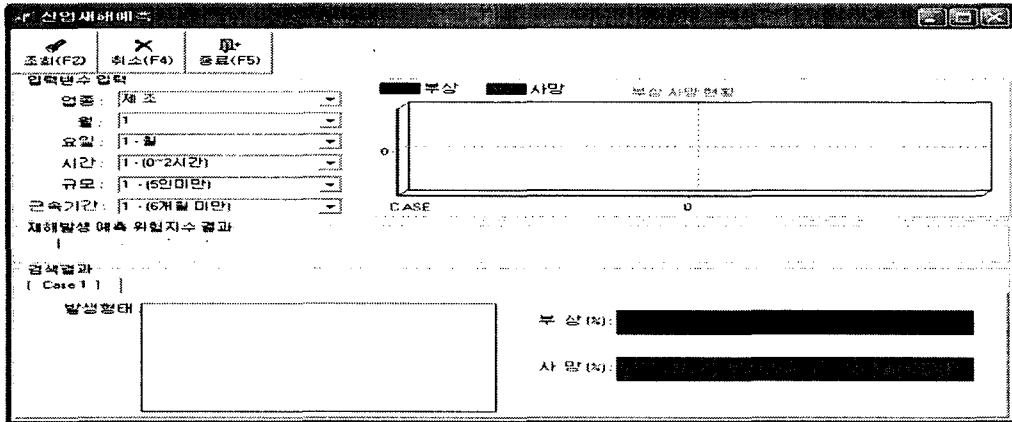
위의 <그림 4> 데이터 테이블을 살펴보면 연번 1의 경우는 재해월은 3월, 재해요일은 월요일(1), 재해시간은 0-2시(1), 규모 5이상-9미만(2), 10-15(3), 16-29(4), 30-49(5)인 사업장에서 근속기간 6개월 미만(1)일 때 발생형태는 전도(5)에 의한 부상이 193건, 사망은 7건으로 전체 19,574건 중 부상은 0.986%, 사망은 0.036%를 나타내고 있다. 이와 같은 방법으로 CHAID 알고리즘을 이용하여 실행한 트리에 대한 분석 결과, 7가지 변수를 모두 만족하는 41개의 끝마디에 대하여 데이터테이블을 이용한 데이터베이스를 구성한 후 산업재해 예방을 위한 전문가 시스템을 개발하고자 한다.

## 6. 전문가 시스템 개발

전문가 시스템이란 특정분야의 전문가 지식을 지식 베이스로 축적하여 상당한 수준의 문제를 취급하는 지식 베이스 시스템으로 전문가 시스템의 특징은 전문가의 지식으로 구성된 지식 베이스를 사용한다는 점, 연역적 추론, 실용성 등을 꼽을 수 있다[6]. 신제품 개발[10], 농업경영[11], 폐수처리운영[12], 혈압조절[16] 등과 같이 다양한 분야에서 전문가 시스템 또는 의사결정지원 시스템이 개발되어 사용되고 있으나 산업재해 관련 분야에서는 전문가 시스템 개발이나 그 활용을 찾아보기가 어렵다. 이에 본 연구에서는 산업재해 예방을 위하여 정량적, 정성적 예측 정보를 제공할 수 있는 전문가 시스템을 개발을 위하여 CHAID 알고리즘의 트리 분석결과를 데이터베이스로 사용하

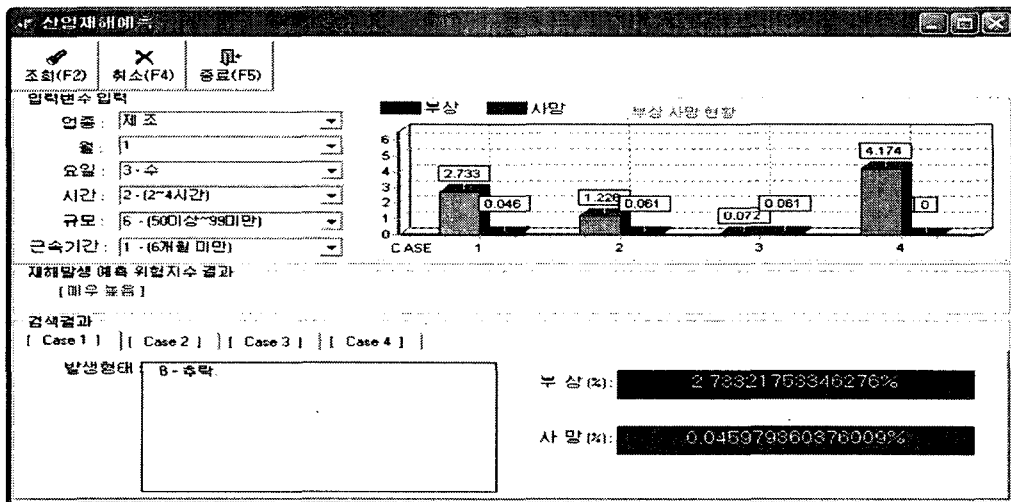


여 델파이(Delphi) 5.0을 이용하였다. 본 연구에서 개발된 전문가 시스템은 아래 <그림 5>와 같다.

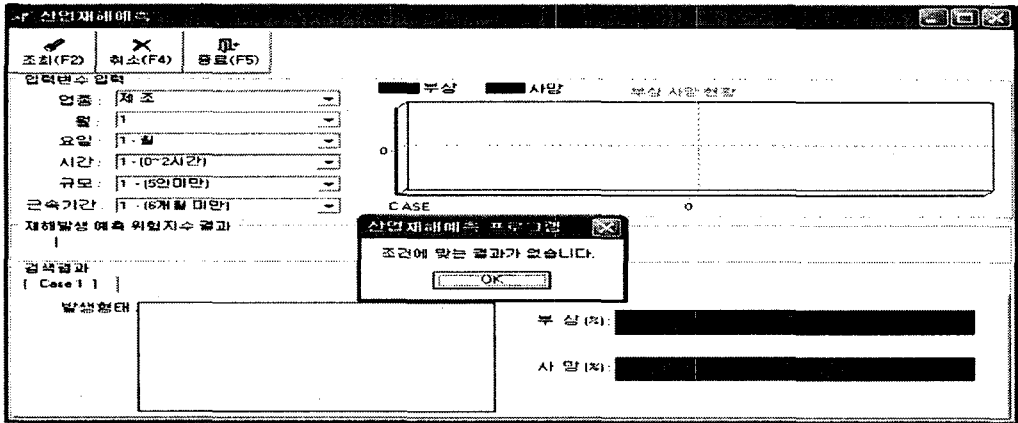


<그림 5> 산업재해 예방을 위한 전문가 시스템 초기화면

위의 <그림 5>를 보면 입력변수인 업종, 월, 요일, 시간, 규모, 근속기간을 입력할 경우 입력변수 값과 동일한 조건의 검색 결과를 발생형태와 발생형태별로 부상과 사망 확률, 경우(Case)별 부상 및 사망 현황을 정량적으로 나타내며, 이에 따른 재해발생 예측 위험지수를 매우 높음, 높음, 낮음, 결과 없음과 같이 정성적으로도 나타낸다. 이러한 과정을 토대로 하여 시스템을 실행하였을 경우 코딩된 값에 의해 다음과 같이 서버에서 프로그램이 실행된다. 결과는 <그림 6-1>과 <그림 6-2>와 같이 출력된다.



<그림 6-1> 산업재해예방을 위한 전문가 시스템 실행 결과 (예 1)



<그림 6-2> 산업재해예방을 위한 전문가 시스템 실행 결과 (예 2)

위의 <그림 6-1>은 입력변수로 업종은 제조업, 월은 1월, 요일은 수요일, 시간은 2~4시 사이, 규모는 50이상~99미만, 근속기간은 6개월 미만의 경우를 입력했을 때, 재해 발생 위험지수는 매우 높음으로 정성적인 예측을 나타내고, 발생형태는 4가지 Case로 나타난다. Case 1의 추락에 대한 부상 발생 위험은 2.733%, 사망 발생 위험은 0.046%와 같이 정량적으로도 예측할 수 있다. 또한 <그림 6-2>는 제조업에서 1월, 월요일, 0~2시 사이, 5인 미만, 6개월 미만을 입력변수로 입력할 경우 조건에 맞는 결과가 없음을 나타내고 있으며 이러한 경우는 재해발생 위험이 없는 것을 의미한다. 이렇게 개발된 전문가 시스템은 다양한 작업조건에서 작업관리책임자 또는 안전 관리자가 현재 상황에 맞는 작업조건을 선택하여 입력하면 작업조건 별로 발생할 수 산업재해를 정량적, 정성적으로 예측하여 결과를 제시함으로 산업재해 예방에 활용될 수 있을 것이다.

## 7. 결론

본 연구는 제조업에서의 산업재해 예방을 위한 전문가 시스템을 개발하고자 선행 작업으로 데이터마이닝 기법 중 의사결정나무의 대표적인 알고리즘인 CHAID, CART, C4.5, QUEST를 SAS의 Enterprise Miner와 SPSS의 AnswerTree를 이용하여 제조업에서 발생한 재해 데이터에 대하여 알고리즘 별 타당성 평가와 정확도, 민감도, 특이도를 구하여 비교 분석하였다. 그 결과 최적의 알고리즘은 CHAID인 것으로 나타났고 이를 토대로 데이터베이스를 구축한 후 델파이 5.0을 이용하여 전문가 시스템을 개발하였는데 결론을 요약하면 다음과 같다.

1. 제조업에 관련된 산업재해 발생 데이터를 가지고 CHAID 알고리즘에서 구현한 결과 뿌리마디는 10,536개의 관측치로 부상 10,313개(97.88%)와 사망 223개(2.12%)로 나

타났고, 깊이(Depth)는 7이며, 끝마디는 총 124개로 형성되었다. 변수로는 부상 또는 사망을 나타내는 재해형태, 재해월, 재해요일, 재해시간, 규모, 근속기간, 발생형태가 선택되었다.

2. CHAID 알고리즘을 이용한 분석 결과에 따라 재해형태, 재해월, 재해요일, 재해시간, 규모, 근속기간, 발생형태의 7가지 변수를 모두 만족하는 노드는 124개의 끝마디 중 41개의 끝마디로 이에 대한 신뢰도를 분석하고 Excel 프로그램을 이용하여 데이터 테이블을 작성 한 후 Access 프로그램으로 변환하여 데이터베이스를 구축하였다.

3. 산업재해 예방을 위하여 산업재해 발생에 관한 정량적, 정성적 예측 정보를 제공할 수 있는 전문가 시스템을 델파이 5.0을 이용하여 개발하였다.

본 연구에서 개발된 전문가 시스템은 다양한 작업조건에서 작업관리책임자 또는 안전 관리자가 현재 상황에 맞는 작업조건을 선택하여 입력하면 작업조건 별로 발생할 수 산업재해를 정량적, 정성적으로 예측하여 결과를 제시함으로써 산업재해 예방에 활용될 수 있을 것이다. 시스템 개발에 사용된 데이터베이스의 데이터양이 제조업의 10,536개를 토대로 하였으므로 전 업종에 확대 적용하기엔 다소 무리가 있다고 생각되며, 더 많은 데이터를 확보하여 본 연구에서 제시한 방법에 의하여 알고리즘 분석 및 이를 토대로 한 데이터베이스 구축이 완성될 수 있다면 전 업종에 적용할 수 있는 전문가 시스템 개발이 가능하리라 생각된다.

## 6. 참 고 문 헌

- [1] 강현철, 최종후, 한상태, 김은석, Answer Tree를 이용한 데이터마이닝, SPSS아카데미, 2001.
- [2] 강현철, 한상태, 최종우, 김은석, 김미경, SAS Enterprise Miner 4.0을 이용한 데이터마이닝, 자유아카데미, 2001.
- [3] 권혜숙, 데이터마이닝 패키지에서 분류나무 알고리즘의 비교 연구, 서울대학교 석사학위논문, 2002.
- [4] 노동부, 산업재해현황분석, 2004.
- [5] 문정호, 사례연구를 통한 데이터마이닝 수행과정 연구, 서울대학교 석사학위논문, 2002.
- [6] 이재규, 전문가시스템 원리와 개발, 법영사, 1996.
- [7] 이종빈, 노민래, 고성식, 재해사례분석을 통한 빌딩공사 재해특성, 한국안전학회지 제19권 제3호, 2004, pp. 101-107.
- [8] 임영문, 황영섭, 최요한, 데이터마이닝 기법을 활용한 산업재해자들에 대한 요인분석, 대한안전경영과학회지 제7권 4호, 2005.
- [9] 정현철, 장성만, 하중훈 공저, Delphi 5, (주)사이버출판사, 2000.

- [10] Balachandra, R., "An expert system for new product development", *Industrial Management & Data Systems*, Vol. 100 No. 7, 2000, pp. 317-24.
- [11] Brayant, K., "ALEES: an agricultural loan evaluation expert system", *Expert Systems with Applications*, Vol. 21 No. 8, 2001, pp. 75-85.
- [12] Carrasco, E.F., Rodriguez, J., Punal, A., Roca, E. and Lema, J.M., "Rule-based diagnosis and supervision of a pilot-scale wastewater treatment plant using fuzzy logic techniques", *Expert Systems with Applications*, Vol. 22 No. 1, 2002, pp. 11-20.
- [13] K. Kirchner, K. -H. Tölle and J. Krieter, "Decision tree technique applied to pig farming datasets", *Livestock Production Science*, Vol. 90, Issues 2-3, November, 2004, pp. 191-200.
- [14] Paul R. Harper, David J. Winslett, "Classification trees; A possible method for maternity risk grouping", *European Journal of Operational Research*, Vol. 169, issue 1, 2006, pp. 146-156.
- [15] S.K. Pal, A. Pal(Eds.), Paul R. Harper, "On learning to predict web traffic", *Decision Support Systems*, Vol. 35, No. 2, 2003, pp. 213-229.
- [16] Srinivas, Y., Timmons, W.D. and Durkin, J., "A comparative study of three expert systems for blood pressure control", *Expert Systems with Applications*, Vol. 20 No. 3, 2001, pp. 267-74.

### 저자소개

**임영문** : 연세대학교에서 학사, 석사학위를 취득하였고, 미국 텍사스주립대학교 산업시스템공학과에서 공학박사를 취득하였으며, 미국 ARRI (Automation and Robotics Research Institute) 연구소에서 선임연구원 및 연구교수를 거쳐 현재는 강릉대학교 산업공학과 부교수로 재직중이다.

**최요한** : 강릉대학교에서 학사, 석사학위를 취득하였고, 현재 강릉대학교 정보전자공학부에서 박사과정을 수료하였으며 관심분야는 데이터마이닝, 산업안전, 정보시스템 등이다.