

# 데이터 마이닝 기법을 활용한 산업재해자들에 대한 요인분석

## - Factor Analysis on Injured People Using Data Mining Technique. -

임영문 \*

Leem Young Moon

황영섭 \*\*

Hwang Young Seob

최요한 \*\*

Choi Yo Han

### Abstract

Many researches have been focused on the analysis of industry disasters in order to reduce them. As a similar endeavor, this paper provides a propensity analysis of injured people from various industries using classification and regression tree(CART), a data mining algorithm. The sample for this work was chosen from 25,157 data related to various industries during one year (2003.2 ~ 2004.1) at Kangwon-Do in Korea. For the purpose of this paper, eight independent variables (injured date, injured time, injured month, type of injured person, continuous service period, sex, company size, age) are taken from injured person group. According to the analysis result, it is found that five out of the eight factors that are predicted as significant have salient effects. Factors of season, time/hour,

† 본 연구는 산업자원부의 지역혁신 인력양성사업의 연구결과로 수행되었음.

\* 강릉대학교 산업시스템 공학과 교수

\*\* 강릉대학교 산업시스템 공학과 박사과정

2005년 7월 접수; 2005년 8월 수정본 접수; 2005년 8월 게재 확정

day of the week, or month which disasters happened do not show any significant effect. This paper provides common features of injured people. The provided analysis result will be helpful as a starting point for root cause analysis and reduction of industry disasters and also for development of a guideline of safety management.

**Keyword** : AnswerTree, CART, Industry disasters, Data mining technique

## 1. 서 론

고도화된 산업의 발달로 인하여 많은 업종들이 생겨나게 되었다. 업종들의 수가 많아지면서 사고에 대비하기 위한 안전장치, 안전장비, 안전교육 그리고 안전법안 등 또한 많아지고 구체화 되었다. 그렇지만 여전히 업종에 따른 크고 작은 사고들이 발생하게 되었다. 이러한 사고의 발생은 부상자, 업무상질병 이환자, 사고 사망자 그리고 업무상 사망자등의 재해자를 만들었다. 사고의 발생으로 인한 재해자를 줄이기 위해 이미 오래전부터 여러 가지 분석 방법과 재해자들의 특성요소를 찾고자 하는 연구가 진행되어 왔으며, 지금도 많은 연구가 진행되고 있는 실정이다. 재해자들의 특성은 업종별로 차이가 있겠지만, 근본적으로 공통적인 특성 요인이 있으리라 짐작된다. 따라서 본 논문에서는 재해자와 관련된 데이터를 바탕으로 재해자들의 성향 및 재해 요인을 분석하고자 하였다. 물론 재해자와 관련된 많은 독립변수들이 존재하지만, 그 많은 변수들을 모두 분석한다는 것은 시간적, 인력적 그리고 자원적 낭비일 것이다.

기존에 이루어졌던 통계 분석을 이용한 재해에 관한 연구를 살펴보면 규모별, 성별, 연령별, 근속기간별, 재해발생 시기별, 발생형태별로 데이터를 분석한 것을 볼 수 있다. 하지만, 이러한 연구의 분석방법을 살펴보면 데이터의 빈도분석을 통해 분석한 내용이 주를 이루고 있으며 빈도분석을 통해 나온 요인별 특성만을 토대로 재해를 줄이기 위한 개선방법에 대해 언급한 것을 볼 수 있다[1][2]. 이에 본 논문에서는 데이터의 효율적이면서 체계적 분석을 위해서 데이터 마이닝 알고리즘 중에서 Classification and Regression Tree(CART) 알고리즘을 사용하였고, 데이터 가공을 위해서 AnswerTree [5]를 사용하였는데, AnswerTree는 가장 보편적인 도구로써 데이터 가공이 쉽고, 결과 분석 역시 쉽게 행할 수 있고[9], 재해자의 성향과 특성을 분석하는데 중요한 도움을 줄 것이기 때문이다.

본 논문에서는 우선, 업종에 따른 많은 독립변수들 중에서 중요한 의미를 가지는 8개의 변수들(재해시간, 재해요일, 재해월, 재해자구분, 근속기간, 성별, 회사규모, 연령)을 선정하여 AnswerTree를 실행 하였고, 결과를 분석하여 의미 없는 변수들은 제거하였다. 그런 다음, 총 10개의 업종 중 해당 종사자들의 수가 전체 데이터에 비해 극히 소수인 업종은 무의미하다고 판단되었기 때문에 제거하여 다시 AnswerTree를 실행해서 재해자의 성향을 분석하였다. 마지막으로, AnswerTree 분석을 통한 데이터의 분석을 하였다.

## 2. 연구 방법

CART(Classification and Regression Trees) 알고리즘은 의사결정나무(Decision Tree)를 형성하는데 있어서 가장 보편적인 알고리즘이며 그룹화를 위해 선택된 기준들을 바탕으로 개별 그룹을 일반화한다[3]. CART는 Breiman 등[8]에 의하여 개발된 것인데, 각 독립변수를 이분화(Binary Split)하는 과정을 반복하여 트리 형태를 형성함으로써, 분류(종속변수가 범주형일 때)와 회귀분석(종속변수가 연속형일 때)을 수행하는 것이다. 이때 변수들은 범주형 또는 연속형 모두에 적용될 수 있으며, 종속변수가 이산형일 때는 지니 지수(Gini Index)[7]를 종속변수가 연속형일 때는 분산의 감소량을 이용하여 이지 분리(Binary Split)[4]를 수행하는 알고리즘이다. 지니 지수는 불순도(Impurity)를 측정하는 하나의 지수이다. 임의의 한 개체가 목표변수  $i$ 번째 범주로부터 추출되었고, 그 개체를 목표변수의  $j$ 번째 범주에 속한다고 오분류(Misclassification)할 확률은  $P(i)P(j)$ 가 된다. 여기에서  $P(i)$ 는 각 마디에서 한 개체가 목표변수의  $i$ 번째 범주에 속할 확률이다. 이러한 오분류 확률은 모두 더하여

$$G = \sum_{i=1}^c \sum_{j \neq i} P(i)P(j)$$

을 얻을 수 있고 이는 위와 같은 분류 규칙 하에서 오분류 확률의 추정치가 된다. 여기서  $c$ 는 목표변수의 범주 수를 말한다. 일반적으로 CART는 범주형 목표변수에 대해서는 지니 지수를 분리기준으로 사용한다. 지니 지수는 각 마디에서의 불순도 또는 다양도(Diversity)를 측정하는 척도 중의 하나이다. 이는 카이제곱 통계량을 사용하는 것과 같은 결과를 갖는다. CART 알고리즘은 지니 지수를 가장 감소시켜 주는 예측변수와 그 변수의 최적분리를 자식마디로 선택하는데, 지니 계수의 감소량은 다음과 같이 계산된다.

$$\Delta G = G - \frac{n_L}{n} G_L - \frac{n_R}{n} G_R$$

여기서  $n$ 은 부모마디의 관측치 수를 말하고  $n_R$ 과  $n_L$ 는 각각 자식마디의 관측치 수를 의미한다. 즉, 자식마디로 분리되었을 때의 불순도가 가장 작도록 자식마디를 형성하는 것이다. 이러한 수학적 이론을 통한 CART는 트리의 형성(Growing the Tree), 가지치기(Pruning), 데이터 분류(Data Classification)와 같이 주요 세 가지 과정을 통해 진행 된다. 본 연구에서는 의사결정을 위해 데이터마이닝 툴인 AnswerTree를 사용하였다. AnswerTree는 최근 각광받고 있는 데이터 마이닝 기법을 구현해 놓은 도구로 SPSS군의 하나이다. AnswerTree의 장점으로서는 첫째로 쉽게 시작할 수 있다는 것이고, 둘째는 결과를 한눈에 볼 수 있다는 것이다. 셋째는 다양한 종류의 데이터 분석이 가능하다는 것이며, 넷째는 지속적인 분석이 가능하다는 것이다. 다섯째로는 신뢰도 높은 최상의 결과를 얻을 수 있으며, 마지막으로 이익(Profit)과 위험(Risk)을 신속히 결정할 수 있다는 것이다. 이러한 장점들 때문에 여러 분야에서 사용되어 지고 있다. 본 연구에서 AnswerTree의 CART 정지규칙(Stopping Rules)[6]을 위해서 최대 트리 깊이(Maximum Tree Depth)는 10으로 정했으며, 경우의 최소수(Minimum Number of

Cases)에서 부모 노드(Parent Node)는 100, 자식 노드(Child Node)는 62로 정했다. 자식 노드(Child Node)를 62로 정한 이유는 본 논문에서 총 10개의 업종으로 데이터를 분석하고자 했으나, 그 중 3개 업종의 재해자 수가 너무 적었기 때문에 10개중 3개의 업종을 제거하였다. 제거된 업종 3개 중 가장 많은 업종수가 62였기 때문에 자식 노드(Child Node)의 최소 향상도는 62가 되어야 한다고 생각했기 때문이다. 그리고 최소 불순도 향상도(Minimum Change in Impurity)는  $0.003 (62/Total \text{ 데이터})$ 값으로 계산되었다.

### 3. 데이터 집합

본 논문에서 분석하고자 하는 데이터는 2003년 2월 ~ 2004년 1월까지 한국산업자원부에서 강원도를 대상으로 집계한 산업재해 발생현황 통계 자료로써, 총 25,157개로 구성되어 있다. < 표 1 >은 업종별 재해자 구분에 따른 분류와 성별에 따른 분류를 나타낸 것이고, < 표 2 >는 업종별 재해자의 빈도와 백분율을 나타낸 것이다.

< 표 1 > 업종별 재해자 구분과 성별에 따른 분류

업종	재해자수			성별		
	질병	사고	합계	남자	여자	합계
농업	0	49	49	39	10	49
어업	0	12	12	12	0	12
전기가스 및 상수도업	4	58	62	62	0	62
기타산업	346	3,814	4,160	2,879	1,281	4,160
건설업	205	6,926	7,131	6,910	221	7,131
금융보험업	36	189	225	153	72	225
임업	11	1,170	1,181	1,112	69	1,181
운수보관통신업	81	864	945	912	33	945
광업	7,140	981	8,121	7,845	276	8,121
제조업	465	2,806	3,271	2,696	575	3,271
총합계	8,288	16,869	25,157	22,620	2,537	25,157

< 표 2 > 업종별 재해자의 빈도와 백분율

업종	빈도	%
농업	49	0.20
어업	12	0.05
전기가스 및 상수도업	62	0.25
기타산업	4,160	16.54
건설업	7,131	28.35
금융보험업	225	0.89
임업	1,181	4.69
운수보관통신업	945	3.76
광업	8,121	32.28
제조업	3,271	13.00
합계	25,157	100

앞의 2장에서 언급한 바와 같이 본 논문은 최초 10개의 업종을 분석하고자 하였으나, 그 중 재해자 수가 너무 적은 3개의 업종은 무의미하다고 판단하여 제거하고 7개의 업종으로 분석하였다. 제거된 3개의 업종의 총 재해자 수는 123명으로 결과적으로 본 논문에서 사용된 총 데이터의 수는 25,034개이다.

## 4. 연구 결과

### 4.1 AnswerTree 결과

뒤에 나오는 < 그림 1 >은 앞 절에서 설명한 데이터들을 바탕으로 AnswerTree 분석을 한 결과이다. 상당히 많은 노드들로 구성되어 있기 때문에 중요한 영향을 미치는 노드들만 정리를 한 것이다. < 그림 1 >에서 볼 수 있듯이 전체 결과 노드는 5 Depth 형태로 구성되어 있다. Depth가 하나하나 깊어질수록 변화하는 오분류 확률은 < 표 3 >과 같다. < 표 3 >을 보면 Depth가 점점 깊어질수록 오분류 확률은 점점 감소되는 것을 볼 수 있다. 이것은 AnswerTree 결과 데이터들을 분류하는 과정이 점점 정확해진다고 판단할 수 있는 근거가 된다. 최상위 노드에서 최하위 노드까지 분류가 진행되어지고 나서 오분류 확률을 확인해 보면 0.65601과 0.303268이다. 이것은 53.8% 정도의 정확성을 보인다는 결과를 나타낸다.

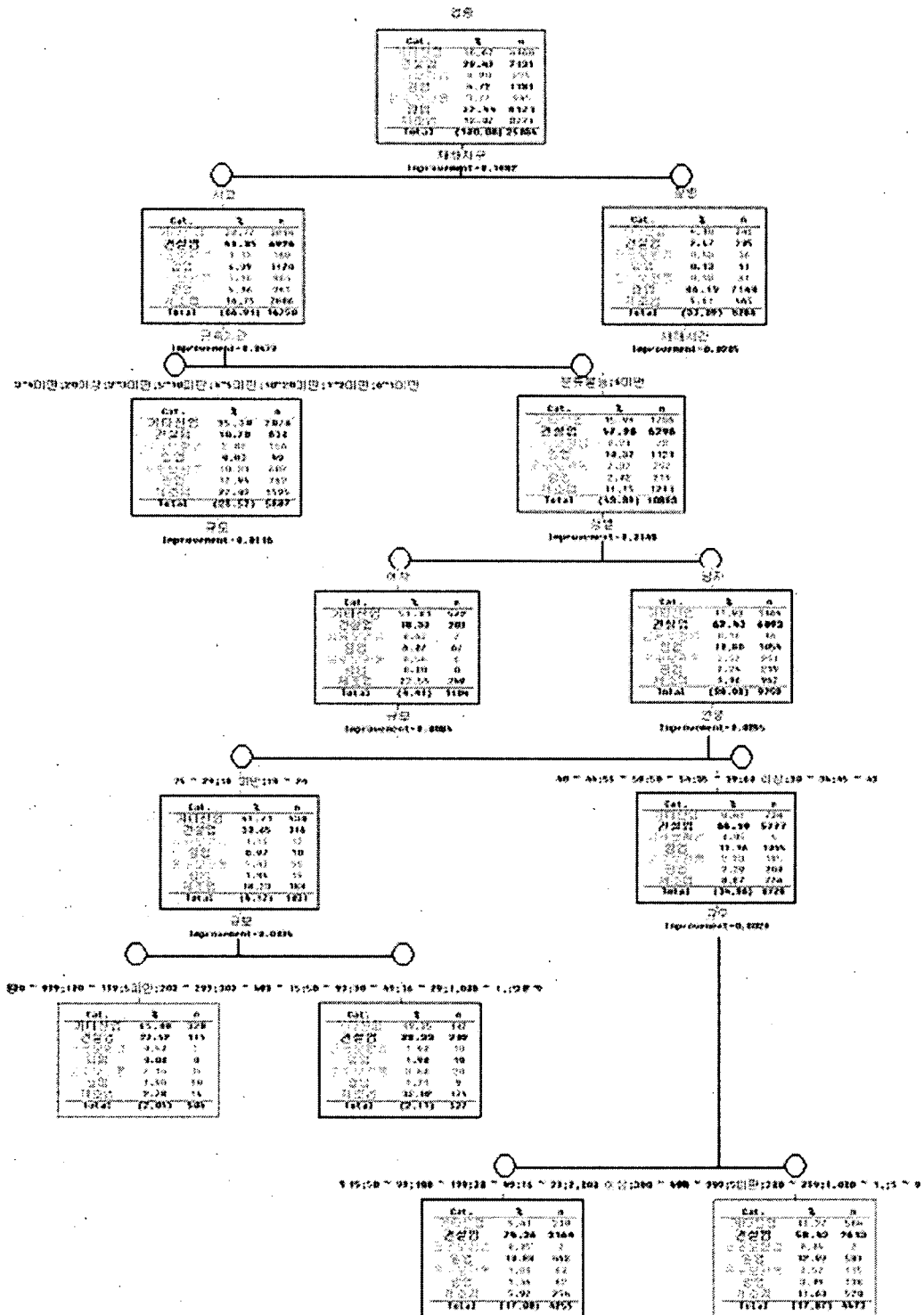
< 표 3 > 의사결정나무 생성후의 오분류 확률

	Risk Estimate (오분류 확률)	Reduced Risk Estimate (감소된 오분류 확률)	
0 Depth	0.675601		
1 Depth	0.438124	0.237477	0 Depth - 1 Depth
2 Depth	0.380283	0.295318	0 Depth - 2 Depth
3 Depth	0.340177	0.335424	0 Depth - 3 Depth
4 Depth	0.322162	0.353439	0 Depth - 4 Depth
5 Depth	0.303268	0.372333	0 Depth - 5 Depth

### 4.2 Training Data 와 Testing Data

하나의 자료에 대해서 적절한 방법을 적용하여 정확하게 모형을 구축했다고 할지라도, 결과가 다른 자료에서도 동일한 결과를 얻을 수 있음을 보장해 주는 것은 아니다. 따라서 하나의 자료로부터 구축된 나무구조가 다른 자료에 대해서도 잘 적용되는가를 확인한 후, 그 나무구조를 그대로 일반화해야 할 것이다. 하나의 자료는 의사결정나무를 구축하는데 사용하고, 이 때 얻어진 결과를 나머지 다른 자료에 적용하여 타당성을 평가해 보는 것이다. 이 경우에 모형을 구축하는데 사용되는 자료를 모형구축자료(Training Data)라고 하고, 얻어진 결과에 적용해 보는 자료를 모형검증자료(Testing Data)라고 한다.

본 논문에서는 AnswerTree를 사용하여 자료의 분할에 의한 타당성 평가를 하기 위해서 분할 데이터 중 모형구축표본(Training Sample)과 모형검증표본(Testing Sample)의 비율을 50% 대 50%로 배분하여 결과를 도출하였다.



< 그림 1 > 주업종에 따른 재해자들의 특성 분류

### 4.2.1 Training Data

아래 < 표 4 >는 총 데이터에서 분리된 모형구축 표본 집단(Training Sample Set)을 나타낸다.

< 표 4 > 업종별 Training Sample의 최상위 노드

cat.	%	n
제조업	13.19	1,653
건설업	28.29	3,546
금융보험업	0.94	118
운수보관통신업	3.68	461
임업	4.79	601
광업	32.32	4,052
기타산업	16.79	2,105
total	(100.00)	12,536

최상위 노드의 오분류 확률의 감소는 < 표 5 >와 같이 0.676771에서 0.318922로 52.9% 감소된 것을 볼 수 있다.

< 표 5 > Training Sample의 오분류 확률

	Root Node	Final Node
Risk Estimate	0.676771	0.318922

### 4.2.2 Testing Data

아래의 < 표 6 >은 총 데이터에서 분리된 모형검증 표본 집단(Testing Sample Set)을 나타낸다.

< 표 6 > 업종별 Testing Sample의 최상위 노드

cat.	%	n
제조업	12.95	1,618
건설업	28.68	3,585
금융보험업	0.86	107
운수보관통신업	3.87	484
임업	4.64	580
광업	32.56	4,069
기타산업	16.44	2,055
total	(100.00)	12,498

< 표 7 > Testing Sample의 오분류 확률

	최상위 노드	최하위 노드
오분류 확률	0.674428	0.323492

최상위 노드의 오분류 확률의 감소는 < 표 7 >와 같이 0.674428에서 0.323492로 52.0% 감소된 것을 볼 수 있다.

즉, 모형구축자료와 모형검증자료를 이용한 타당성 평가에서 모형구축자료와 모형검증자료의 오분류 확률이 거의 차이가 없으므로 본 논문의 의사결정나무는 타당하다고 판단 내릴 수 있다.

### 4.3 교차 타당성

앞의 모형구축자료와 모형검증자료를 이용한 타당성 평가 이외에 또 다른 방법으로 구축된 나무구조를 일반화할 수 있는지를 평가하는 교차타당성(Cross Validation) 평가를 들 수 있다. 이전의 모형구축자료와 모형검증자료에 의한 타당성 평가와 교차타당성 평가가 얼마나 차이를 보이는지 알아보려고 실시하였다. 교차타당성을 평가하기 위해서는 먼저, 전체 자료를 서로 겹치지 않는  $k$ 개의 집단으로 나누어야 한다. 그렇게 되면 어느 하나의 집단을 제외한 다음 나머지 자료를 가지고 모형을 구축하여 최종적으로  $k$ 개의 모형을 구축하게 되는 것이다. 모형 구축이 끝나면  $k$ 개의 위험의 평균을 계산하고 이를 타당성 평가에 의한 위험 추정치로 사용하는 것이다. 이와 같은 방식의 교차타당성 평가를  $k$ -Fold 교차타당성 평가라고 한다.

< 표 8 > 교차타당성 평가후의 오분류 확률

	k개의 위험평균	교차타당성
오분류 확률	0.321203	0.320963

본 논문에서는 교차타당성 평가를 위해 표본 집단의 수를 7(업종의 수)개로 지정하여 분석을 하였다. 그 결과는 위의 < 표 8 >과 같다. 위 < 표 8 >에서 보는 바와 같이  $k$ 개의 위험평균 확률(0.321203)과 교차타당성 확률(0.320963)을 비교했을 때 오분류 확률 차이가 거의 없기 때문에 본 논문의 의사결정나무는 타당하다고 판단 내릴 수 있다.

## 5. 요인분석 고찰

전절의 < 그림 1 >에서 볼 수 있는 것과 같이 업종별 재해자에 가장 큰 영향을 미치는 변수가 재해자 구분인 것을 알 수 있다. 재해자 구분은 재해자 사고와 재해자 질병으로 나뉘는데, 재해자 사고의 경우 16,750명으로 총 재해자 중 66.91%를 차지 한다는 것을 알 수 있다. 이것은 이 데이터 집단에서 재해자 사고가 압도적인 비중을 차지 한다는 것을 의미한다. 다음으로 업종별 재해자가 재해자 구분에서 사



고에 속하는 경우 재해자에게 가장 큰 영향을 미치는 변수는 근속기간이라는 것을 알 수 있다. 근속기간이 6개월 미만과 분류불능인 경우가 43.39%이고, 그 외의 근속기간(6개월~1년 미만·1년~2년 미만·2년~3년 미만·3년~4년 미만·4년~5년 미만·5년~10년 미만·10년~20년 미만·20년 이상)인 경우가 23.52%라는 것을 알 수 있다. 즉, 업종별 재해자에 있어서 가장 큰 영향을 미치는 변수가 재해자 구분과 근속기간이라는 것을 알 수 있다. 재해자 구분이 사고이고, 근속기간이 분류불능·6개월 미만인 경우에 가장 큰 영향을 미치는 독립변수는 성별이라는 것을 알 수 있다. 성별이 여자인 경우에는 4.41%, 남자인 경우에는 38.98로 성별이 남자인 경우가 여자인 경우보다 압도적인 비중을 차지 한다는 것을 알 수 있다. 이것은 강원도라는 특수 지역에서 재해자들의 비중 중 남자가 대부분을 차지한다고 예측할 수 있는 것이다. 재해자의 성별이 남자인 경우 가장 영향을 많이 주는 독립변수는 연령이라는 것을 알 수 있다. 18세 미만·18세~24세·25세~29세인 경우가 4.12%이고, 그 외의 연령대가 34.86%를 차지한다. 즉, 연령대가 높을수록 재해자가 될 확률이 높다고 예측할 수 있다. 연령대별로 18세 미만·18세~24세·25세~29세인 경우 규모로 나누게 되는데 최종 노드 모두 2.01, 2.11로 매우 낮은 비율을 보였다. 그 외의 연령대의 경우에도 회사의 규모로 나누게 되는데 최종 노드 모두 17.00, 17.88로 연령대가 높은 노드인 경우 비율이 높은 것을 볼 수 있다. 결론적으로 업종별 재해자의 특성은 재해자 구분(사고, 질병)이 가장 큰 영향 요인이고, 사고에 따른 근속기간이 분류불능·6개월 미만이며, 성별이 남자이고, 연령이 높은 특성을 띠고 있으며, 회사규모 역시 큰 영향을 미친다는 것을 알 수 있다.

그러나 AnswerTree에 의한 트리 생성 결과 재해시간, 재해요일 그리고 재해월은 포함되어 있지 않았다. 이것은 최초 8개 독립변수 중 재해시간, 재해요일 그리고 재해월은 재해자 특성 분석에 있어서 무의미한 변수로 평가되어졌기 때문이다.

앞에서의 AnswerTree 분리가 얼마나 타당성을 가지고 있는지 확인하기 위해서 본 논문에서는 모형구축자료와 모형검증자료에 대한 비교법과 교차타당성을 사용하였고, 그 결과 모형구축자료와 모형검증자료의 오분류 확률의 차이가 거의 없다고 판단이 되기 때문에 AnswerTree에 의한 분리는 타당성을 보인다는 것을 알 수 있다. 교차타당성 결과 역시 AnswerTree에 의한 분리가 타당성을 지니고 있다고 판단할 수 있다.

## 6. 결 론

데이터 처리를 함에 있어서 최초 10개 업종에 따른 의사결정나무는 재해자의 수가 적은 3개의 업종 때문에 무의미하다고 판단되었다. 그래서 재해자의 수가 극히 적은 3개의 업종을 제외한 7개의 업종을 바탕으로 분석한 의사결정나무에서 재해자들의 특성을 파악 할 수 있었다. 재해자에 영향을 미칠 것이라고 예상한 8개의 독립변수들 중 5개만이 큰 영향을 미친다는 것을 알 수 있었다. 그리고 재해자의 특성은 재해시간, 재해요일 그리고 재해월과는 상관관계가 없다는 것을 알 수 있었다.

결론적으로, 재해자들의 공통된 특성을 정리하면 아래와 같다.

1. 지역적 특성상 질병보다는 사고가 절대적으로 많다.
2. 여자보다는 남자의 비중이 크며, 남자 재해자 수 역시 절대적으로 많다.
3. 근속기간은 종사하는 업종에 숙련도가 부족한 6개월 미만·분류 불능일 때 사고 재해 가능성이 높아진다.
4. 연령대가 낮은 젊은 층은 적응력이 빠르나, 연령대가 높은 고령층은 적응력이 느리기 때문에 연령대가 높을 때 재해 가능성이 높다.
5. 회사규모도 영향을 미치는 변수이지만 특정 규모에 상관없이 재해자들이 사고로 기인할 가능성이 높아진다는 것을 알 수 있다.

기존에 이루어졌던 통계분석을 이용한 재해에 관한 연구는 단지 빈도분석을 통해서 이루어졌다. 이것은 빈도분석을 함에 있어서 본 논문에서 언급한 8개의 독립변수를 분석할 때, 8개 독립변수들 모두 분석하는 것을 의미한다. 즉, 어떤 독립변수가 가장 중요한 영향을 미치는지, 어떤 독립변수가 무의미한지는 알 수 없다는 것이다. 그러나 본 연구에서 제시한 방법을 사용하게 되면 8개 독립변수가 존재하더라도 가장 중요한 변수가 어떤 것인지, 무의미한 변수가 어떤 것인지를 쉽게 찾아낼 수 있다는 것이다.

각 독립변수들 중 범주에 따른 데이터들이 분포된 빈도는 모두 다를 것이다. 여기서 빈도가 높은 범위를 집중적으로 분석을 한 후 중요한 범주들을 새로이 정의하고, 새롭게 정의된 범주들을 이용해서 다시 의사결정나무 분석하게 되면 조금 더 세분화된 정확한 재해자들의 특성을 알 수 있으리라 사료된다.

## 7. 참 고 문 헌

- [1] 김경배, 산업재해의 예방대책에 관한 연구, 관동대학교 석사학위논문, 2004.
- [2] 김종현, 우리나라 산업재해 통계를 이용한 재해실태분석과 통계제도의 개선방향, 경일대학교 석사학위논문, pp. 40-60, 1998.
- [3] 송주미, 의사결정나무 분리기준 알고리즘에 관한 연구, 연세대학교 석사학위논문, pp.1-19, 2004.
- [4] 최종후, 서두성, 데이터마이닝 의사결정나무의 응용, 통계청 [통계분석연구], 1999.
- [5] 최종후 외 3, AnswerTree를 이용한 데이터 마이닝 의사결정나무분석, SPSS 아카데미, pp42-44, pp 53-61, 2000.
- [6] Colombet, I., Ruelland, A., Chatellier, G., Gueyffier, F., Degoulet, P., & Jaulent, M. C., Models to Predict Cardiovascular Risk : Comparison of CART, Multilayer Perceptron and Logistic Regression. Proceedings AMIA Symposium, pp. 156-160, 2000.
- [7] F. Berzal, et al., On the Quest for Easy-to-Understand Splitting Rules, Data and Knowledge Engineering 44, pp. 31-48, 2003.

- [8] L. Breiman, et al., Classification and Regression Trees, Chapman & Hall, London, 1993.
- [9] Pietersma, D., Lacroix, R., Lefebvre, D., Wade, K.M., Induction and Evaluation of Decision Tree for Lactation Curve Analysis. Comput. Electron. Agric. 38, pp. 19-32, 2003.

## 저 자 소 개

**임 영 문** : 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..

**황 영 섭** : 현재 강릉대학교 산업공학과 대학원 재학 중이며 관심분야는 Ubiquitous System, 알고리즘 분석 및 활용 등이다.

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