

격자탐색법을 이용한 의사결정나무 분석 최적 모형 선택 : 직장예비군 지휘관의 직장만족도에 대한 영향 요인 분석을 중심으로

정철우[†] · 정원영 · 신다윗
한국국방연구원

Selection of the Optimal Decision Tree Model Using Grid Search
Method : Focusing on the Analysis of the Factors Affecting
Job Satisfaction of Workplace Reserve Force Commanders

Chulwoo Jeong · Won Young Jeong · David Shin
Korea Institute for Defense Analyses

■ Abstract ■

The purpose of this study is to suggest the grid search method for selecting an optimal decision tree model. It chooses optimal values for the maximum depth of tree and the minimum number of observations that must exist in a node in order for a split to be attempted. Therefore, the grid search method guarantees building a decision tree model that shows more precise and stable classifying performance. Through empirical analysis using data of job satisfaction of workplace reserve force commanders, we show that the grid search method helps us generate an optimal decision tree model that gives us hints for the improvement direction of labor conditions of Korean workplace reserve force commanders.

Keywords : Decision Tree, CART, Grid Search, Tree Pruning, Model Tuning

1. 서 론

의사결정나무 분석(Decision Tree Analysis)은 분류 또는 예측을 위한 데이터마이닝 기법 중의 하나로서 방법론적인 측면에서 여러 장점을 가지고 있다. 무엇보다 의사결정나무 분석은 그 결과가 나무구조로 표현되기 때문에 연구자가 쉽게 해석할 수 있고 다른 이들에게 쉽게 설명할 수 있다는 장점이 있다. 또한 의사결정나무 분석은 변수에 대한 선형성(linearity), 정규성(normality) 등의 가정을 필요로 하지 않고 연속형 변수와 범주형 변수를 모두 취급할 수 있다는 장점이 있다. 그리고 사회과학 분야에서 널리 사용되고 있는 회귀분석이나 로짓분석과 같은 모수적(parametric) 기법에서는 상호효과(interaction effects)를 모형에 포함해 분석하는 것이 어렵지만 의사결정나무분석에서는 이러한 효과를 효율적으로 추출해낸다[9].

그러나 이러한 방법론적 장점을 가지고 있음에도 의사결정나무 분석의 단점으로 흔히 지적되는 것이 모형의 불안정성인데, 이는 데이터에 약간의 변화가 있는 경우에도 전혀 다른 결과가 양산될 수 있음을 뜻한다. 따라서 의사결정나무 분석을 수행하기 위해서는 모형의 불안정성을 어떻게 통제하느냐가 매우 중요하다. 의사결정나무 분석 모형의 불안정성을 통제하는 것은 결국 표본 추출에 상관없이 안정적인 결과를 산출해내는 모형을 구축하는 것과 관계되며, 이는 곧 의사결정나무 분석의 결과로서 얻어지는 나무의 크기를 어떻게 통제하느냐와 관련이 있다. 의사결정나무 분석 모형의 나무 크기를 통제하는 것은 일반적으로 최대 뿌리 깊이나 각 노드(node)에서의 분리 가능한 최소 개체 수를 설정하는 것을 통해 이루어진다. 따라서 최대 뿌리 깊이와 각 노드에서의 분리 가능한 최소 개체 수를 어떻게 설정하느냐는 소위 말해 ‘좋은 의사결정나무 분석 모형’, 즉 모형의 설명력 또는 예측력이 높으면서도 그 결과를 일반화할 수 있는 안정적인 의사결정나무 분석 모형을 구축하기 위해 매우 중요하다.

그럼에도 불구하고 의사결정나무 분석을 응용한

많은 연구에서는 모형의 안정성을 확보하기 위한 모형 설계의 과정이 생략되어 있거나[3, 4], 모형의 성과로서 분류정확도 내지는 오분류율만 고려하고 분류정확도 또는 오분류율의 분산을 고려하지 않음으로써[5], 결국 모형의 불안정성이라는 과제는 해결되지 못하는 한계를 보이는 것을 확인할 수 있다. 의사결정나무 분석과 다른 통계 기반 또는 인공지능 기반의 기법을 결합한 통합 기법에 관한 연구에 있어서도 그러한 한계를 동일하게 드러내고 있다[1, 2, 6].

이에 본 연구에서는 최대 뿌리 깊이와 각 노드에서의 분리 가능한 최소 개체 수에 대한 최적 값을 선정함으로써 궁극적으로 의사결정나무 분석의 안정적인 모형을 구축하기 위한 방법으로 격자탐색법(grid search method)을 제안하고 실증 데이터를 이용하여 그 타당성을 보이고자 하였고, 이 점에서 본 연구의 방법론적 측면에서의 의의를 찾을 수가 있다.

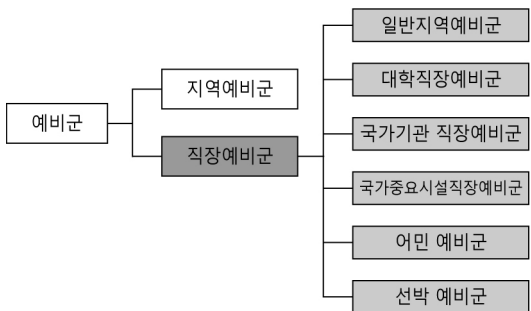
본 연구의 또 다른 의의는 국방 분야의 정책적 측면에서 찾을 수 있다. 즉, 본 연구는 직장예비군 지휘관의 처우에 있어서의 개선 방향 설정을 위한 근거를 마련하는 데 그 의의가 있다. 현재 직장예비군 지휘관은 국방부장관이 임면권을 가지고 보수는 직장의 장이 지급하는 방식으로 관리되고 있다. 그러다보니 직장에 따라 직장예비군 지휘관에 대한 처우가 매우 상이하게 되는 현상이 발생하였고, 이 점에 있어서 개선의 요구가 국회, 국가인권위원회, 국방부 등을 통해 지속적으로 제기되었다. 본 연구는 일반직장예비군 지휘관과 대학직장예비군 지휘관들을 대상으로 설문조사를 하여 그들의 직장만족도를 측정하고 분석하였다. 이를 통해 직장예비군 지휘관의 처우 개선에 대한 요구가 타당성을 가지는지를 살펴보고, 아울러 어떠한 점에 있어 개선이 시급한지에 대해 분석하고자 하였다. 이에 대한 좀 더 자세한 기술은 제 2장에서 다루었다.

본 논문의 구성은 다음과 같다. 제 2장에서는 직장예비군 지휘관 처우 개선의 요구 배경에 대해 소개하고, 제 3장에서는 의사결정나무 분석 및 격자탐색법에 대해 기술하며, 제 4장에서는 실증분석 과

정 및 결과를 제시하며, 마지막 제 5장에서는 결론 및 추후 연구방향에 대해 논한다.

2. 직장예비군 지휘관 처우 개선 요구의 배경

「향토예비군 설치법」 제3조의2에 따르면 예비군은 크게 지역 방어를 목적으로 하는 지역예비군과 직장 방호를 목적으로 하는 직장예비군으로 구분된다. 직장예비군은 다시 [그림 1]에 보는 바와 같이 일반직장예비군, 대학직장예비군, 국가기관 직장예비군, 국가중요시설 직장예비군, 어민예비군, 선박예비군 6가지 유형으로 세분화된다. 본 연구는 이 중에서 부대 수 기준으로 가장 큰 비중을 차지하는 일반직장예비군과 대학직장예비군 부대¹⁾ 소속 지휘관의 직장만족도 차이를 분석하는 데 중점을 두고 진행되었다.



[그림 1] 예비군의 편성

예비군 지휘관은 공통적으로 평시에는 해당 부대 예비군의 훈련 및 관리 등의 임무를 수행하고 전시에는 이들 예비군을 통솔하여 전시 임무를 수행하는 역할을 한다. 단, 직장예비군 지휘관의 경우에는 직장에서 요구하는 추가적인 업무를 수행하기도 한다. 또한, 예비군 지휘관은 소속 예비군 부대의 종류에 상관없이 국방부가 주관하는 동일한 선발 시험을 거쳐 선발된다.²⁾ 단, 시험 성적 및 근무희망

부대에 따라 배정되는 부대가 달라질 뿐이다.

그럼에도 불구하고 예비군 부대의 종류에 따라 소속 부대의 지휘관에 대한 처우는 매우 다르다. 지역예비군 지휘관과 직장예비군 지휘관에 대한 처우가 다르고, 직장예비군 지휘관 사이에서도 소속 직장에 따라 받는 처우가 다르다. 지역예비군 지휘관은 모두 군무원으로 채용되어 동일한 인사관리 규정을 적용받는다. 이에 반해 직장예비군은 직장에 따라 고용형태, 정년, 급여 등의 처우에 있어 상당한 차이를 나타내고 있다.

2013년 10월 국방부가 국회의원 안규백 의원실에서 제출한 자료에 따르면 직장예비군 지휘관 중에서 정규직 비율은 <표 1>에 정리한 바와 같이 95.6%에 달하는 반면, 대학직장예비군 중에서 정규직 비율은 34.0%에 지나지 않는다. 정규직인지의 여부는 예비군 지휘관의 정년과 각종 수당 등에도 영향을 미친다.

<표 1> 직장예비군 지휘관의 고용 형태

	정규직	계약직	합계
일반직장 예비군	520 (95.6%)	24 (4.4%)	544 (100.0%)
대학직장 예비군	83 (34.0%)	161 (66.0%)	244 (100.0%)

또한, 동일한 자료에 따르면 일반직장예비군 지휘관의 평균 연봉은 6,080만 원이고 대학직장예비군 지휘관의 평균 연봉은 3,912만 원으로 약 2,000만 원의 차이가 난다. 최고 연봉과 최소 연봉의 차이는 최대 1억 원까지 나는 것으로 보고되었다. 상황이 이러하다 보니 대학직장예비군 지휘관을 중심으로 처우 개선에 대한 요구가 국회, 국가인권위원회, 국방부 등을 통해 지속적으로 제기되었다.

이러한 배경 하에 본 연구가 시작되었다. 본 연구는 일반직장예비군 지휘관과 대학직장예비군 지휘관을 중심으로 그들의 직장만족도를 분석함으로

1) 2014년 말 기준, 일반직장예비군 부대 및 대학직장예비군 부대는 같이 각각 541개와 264개가 있다.

2) 예비군 지휘관의 선발시험에 관한 자세한 사항은 「예비전력관리 업무담당자 선발 규칙」을 참조하도록 한다.

써 첫째, 직장에비군의 종류에 따라 지휘관의 직장 만족도에 있어 차이가 존재하는지를 확인하고, 둘째, 직장에비군 지휘관의 직장만족도가 어떠한 부분에서 가장 큰 차이를 나타내는지 탐색하며, 셋째, 그 결과를 토대로 직장에비군 지휘관의 처우 개선을 위한 방향을 설정하고자 하였다.

3. 의사결정나무 분석

3.1 CART 알고리즘

의사결정나무 분석은 분류 또는 예측을 위한 데이터 마이닝 기법의 하나로서 의사결정규칙(decision rule)을 도식화하여 관심 대상 집단을 몇 개의 집단으로 분류하거나 예측을 수행한다[7]. 의사결정나무 분석은 분류 또는 예측의 정확도보다는 분석 결과에 대한 해석 또는 설명이 중요할 때 더 유용하게 활용이 된다[17].

의사결정나무 분석을 구현하는 대표적인 알고리즘에는 Hartigan[15]에 의해 제시된 CHAID(Chi-squared Automatic Interaction Detection), Breiman et al.[11]에 의해 제시된 CART, Quinlan[21]의 ID3을 기반으로 한 C5.0 등이 있다.

본 연구에서는 이 중에서도 CART 알고리즘을 이용하여 의사결정나무 모형을 구축하였다. CART 알고리즘은 지니 계수(gini index) 또는 분산의 감소량을 이용하여 이진분리(binary split)를 수행한다. 지니 계수는 각 노드(node)에서의 불순도를 측정하는 지수로서 임의의 한 개체가 목표변수의 j 번째 범주에 속하지만 이를 오분류하는 확률로 표현된다. 지니 계수의 수리적 표현은 식 (1)과 같다.

$$G_i = \sum_{j=1}^c p_{ij}(1-p_{ij}) \quad (1)$$

여기서 G_i : i 번째 노드에서의 지니 계수
 p_{ij} : i 번째 노드에서 임의의 한 개체가 j 번째 범주에 속할 확률
 c : 범주의 개수

식 (1)을 전개해서 정리하면 식 (2)와 같이 된다.

$$G_i = 1 - \sum_{j=1}^c p_{ij}^2 \quad (2)$$

그런데 p_{ij} 는 개념적으로 i 번째 노드에 포함되어 있는 개체 수 대비 j 번째 범주에 속하는 개체 수의 비율을 의미하므로 식 (2)는 식 (3)과 같이 표현이 가능하다.

$$G_i = 1 - \sum_{j=1}^c (n_{ij}/n_i)^2 \quad (3)$$

여기서 n_i : i 번째 노드에 포함되어 있는 개체 수
 n_{ij} : i 번째 노드에서 j 번째 범주에 속한 개체 수

CART 알고리즘은 지니 계수를 최소화하도록 각 노드에서 설명변수 및 그 설명변수의 분리값(cutoff point)을 선정한다.

3.2 격자탐색법을 이용한 가지치기

의사결정나무 분석에 있어 설명력이 높으면서도 동시에 결과를 일반화할 수 있는 모형을 구축하기 위해서는 가지치기를 통해 모형의 복잡성(complexity)을 통제해야 한다. 가지치기의 방법으로는 사전적 방법(prepruning approach)과 사후적 방법(post-pruning approach)이 있다[14]. 사전적 방법은 각 노드에서 분리를 계속 진행할 것인지의 여부를 사전에 정해놓은 기준에 따라 결정하는 방식을 의미한다. 사후적 방법은 각 노드가 더 이상 분리가 되지 않을 때까지 분리를 하여 나무의 가지가 모두 뻗어나가게 한 후 사후적으로 가지를 제거하는 방식이다.

CART 알고리즘에서 이용되는 가지치기 방식은 사후적 방법에 해당한다. 이 방법은 소위 비용 복잡성 가지치기(cost complexity pruning)로 불리는 방식인데, 여기서 비용 복잡성은 나무의 크기에 대한 모형의 오분류율의 함수로서 정의된다[14]. 나무의 크기와 모형의 오분류율은 서로 상충이 되는데 그러기에 적절한 균형이 필요하다. 일반적으로

나무의 크기는 작으면서 모형의 오분류율이 낮은 모형이 최적의 모형으로 선정된다.

본 연구에서는 나무의 크기를 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수를 조정함으로써 통제하였다. 최대 뿌리 깊이가 크면 클수록 나무의 크기는 커지는 경향이 있고, 각 노드에서 분리 가능한 최소 개체 수가 크면 클수록 나무의 크기는 작아지는 경향이 있다. 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수에 대한 적절한 값을 선정하기 위해 본 연구에서는 격자탐색법(grid search method)을 이용하였다.

격자탐색법은 Hsu et al.[16]에서 처음 제시된 기법으로서 SVM(Support Vector Machines)의 최적 모형 선정을 위한 커널 함수의 모수 추정 방법으로서 여러 연구에서 활용되어 그 효과성이 검증된 바 있다[18, 19]. 그러나 의사결정나무 분석의 최적 모형을 선정하기 위한 방법으로 격자탐색법이 활용되어 그 효과성이 밝힌 실증적 연구는 거의 없었다.

본 연구에 있어 의사결정나무 분석 모형 선정을 위한 격자탐색법의 적용 과정은 다음과 같다.

- ① 전체 데이터를 무작위로 k 개의 그룹으로 나눈 후, 1개 그룹을 제외한 나머지 $(k-1)$ 개의 그룹은 모형 구축을 위한 훈련용 데이터(train set)로 이용하고, 나머지 1개 그룹은 구축된 모형의 성과를 테스트하기 위한 데이터(test set)로 이용한다.
- ② 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수 각각에 대해 실험하고자 하는 p 개와 q 개의 수치를 대안으로 설정한다.
- ③ 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수에 대한 $p \times q$ 개의 조합 각각에 대해 훈련용 데이터로 모형을 구축하고, 테스트용 데이터로 모형의 오분류율을 산출해낸다.
- ④ 훈련용 데이터로 이용했던 $(k-1)$ 개 그룹에 대해서도 돌아가면서 1개 그룹씩을 테스트용 데이터로 하고 테스트용 데이터를 제외한 나머지 데이터는 훈련용 데이터로 하여 ③의 과정을 반복한다.
- ⑤ ③과 ④의 과정을 통해 최대 뿌리 깊이와 각

노드에서 분리 가능한 최소 개체 수에 대한 $p \times q$ 개의 조합 각각에 대해 얻어진 k 개의 오분류율의 평균과 표준편차를 계산한다.

- ⑥ 오분류율의 평균과 표준편차가 모두 최소화 되는 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수의 조합을 찾되, 이를 만족시키는 조합이 없을 경우에는 분석자의 판단에 따라 적절한 값을 선정하도록 한다.

4. 실증 분석

본 장에서는 본 연구에서 실시한 설문조사 분석 결과를 기술한다. 먼저, 일반직장예비군 지휘관과 대학직장예비군 지휘관의 직장만족도에 있어 유의한 차이가 존재하는지에 대한 차이 검정 결과를 기술한다. 그리고 그러한 차이에 영향을 미치는 요소에 대한 의사결정나무 분석 결과를 제시한다.

본 연구의 설문조사는 2013년 11월 20일부터 12월 6일까지 전국 직장예비군 지휘관을 대상으로 실시하였고, 국방동원정보체계에 설문지를 올린 후 이메일 또는 팩스로 응답을 받는 방식으로 진행하였다. 전체 805명 중 229명이 응답하였는데 이 중에서 일반직장예비군 지휘관은 138명, 대학직장예비군 지휘관은 91명이었다. 각 설문 문항에 대해 결측치에 대해서는 분석에서 제외하였다. 주된 설문조사 내용은 크게 두 가지로 구분되는데 하나는 직장예비군 지휘관의 직장만족도에 대한 설문이고, 다른 하나는 직장예비군 지휘관의 직장만족도에 영향을 주는 요인에 대한 설문이었다.

4.1 직장만족도 차이 분석 결과

일반직장예비군 지휘관과 대학직장예비군 지휘관의 직장만족도의 차이를 분석하였다. 이를 위해 직장에 대한 전반적 만족도를 측정하는 문항과 지역에 비군 지휘관에 대한 상대적 불만족을 측정하는 문항을 통해 직장예비군 지휘관의 만족도를 측정하였다. <표 2>와 <표 3>은 그 결과를 정리한 것이다.

<표 2> 직장예비군 지휘관 만족도

		평균	표준편차
전반적 만족도	대학직장예비군	2.46	0.923
	일반직장예비군	3.75	1.024
상대적 불만족	대학직장예비군	3.18	1.399
	일반직장예비군	2.48	1.326

<표 2>에 보듯이 일반직장예비군 지휘관이 대학직장예비군 지휘관에 비해 직장에 대한 전반적 만족도에 있어 평균적으로 높은 수준을 기록하였다. 또한 지역에비군 지휘관에 대한 상대적 불만족은 일반직장예비군 지휘관에 비해 대학직장예비군 지휘관이 평균적으로 더 높은 수준을 기록하였다. 이러한 차이에 대한 t-검정 결과, <표 3>에 보는 바와 같이 전반적 만족도에 대해서는 t-통계량이 -9.92이고 p-값이 0.00이 나왔으며, 상대적 불만족에 대해서는 t-통계량이 3.66이고 p-값이 0.00이 나와 유의수준 0.05에서 두 지표 모두 집단 간 통계적으로 유의한 차이가 나는 것으로 분석되었다.

4.2 의사결정나무 분석을 위한 변수 선정

의사결정나무 분석을 위한 변수로서 직장예비군 지휘관의 직장만족도 영향 지표가 이용되었다. 여기서 직장만족도 영향 지표란 직장만족도에 영향을 미치는 요소를 측정하는 지표를 의미한다. 이들 지표를 선정하기 위해 다음과 같은 4단계 과정을 거쳤다.

- Step 1 : 선행연구에서 직장만족도 영향 지표를 찾아 정리하였다. <표 4>는 그 결과를 나타내고 있

<표 4> 선행연구에 나타난 직장만족도 영향 지표

저자	측정 지표	
Brayfield and Rothe [10]	<ul style="list-style-type: none"> • 전반적 만족도 • 업무에 대한 열정 • 일상 업무의 연속성 	<ul style="list-style-type: none"> • 업무의 흥미 • 업무에 대한 불쾌감
Weiss et al. [20]	<ul style="list-style-type: none"> • 업무의 활동성 • 업무의 독립성 • 업무의 사회적 가치 • 상사가 부하를 다루는 방식 • 상사의 의사결정능력 • 도덕적 가치 • 신분의 안정성 • 조력기회 • 권한행사 • 능력 발휘 	<ul style="list-style-type: none"> • 조직의 정책운영방식 • 임금 • 승진가능성 • 자율권 • 재량권 • 업무환경 • 동료와의 관계 • 타인의 인정 • 성취감
Cooper et al. [13]	<ul style="list-style-type: none"> • 급여 • 직장의 전망 • 직원 동료 • 물리적 근무 환경 	<ul style="list-style-type: none"> • 부서 운영 방식 • 능력 발휘 방식 • 관심이나 보유기술을 업무에 발휘 정도
Cass et al. [12]	<ul style="list-style-type: none"> • 업무의 독특성 • 관리자의 직원에 대한 태도 • 동료와의 커뮤니케이션 	<ul style="list-style-type: none"> • 급여 및 복지 혜택 • 주도적으로 일할 수 있는 수준
Al-Rubaish et al. [8]	<ul style="list-style-type: none"> • 자아성취감 • 상사의 감독 • 방침 및 시설 • 업무 자체 	<ul style="list-style-type: none"> • 직원들과의 관계 • 기여감 • 급여 • 업무 부하

다. Brayfield and Rothe[10]는 업무 자체에 대한 만족도를 5개 지표를 이용하여 측정하였다. Weiss et al.[20]과 Cooper et al.[13], Cass et al.[12], Al-Rubaish et al.[8] 등의 연구에서는 업무 자체에 대한 만족도 외에도 상사, 동료와의 관계, 능력 발휘의 기회, 보상, 근무여건, 업무부하 등에 대한 응답자의 인식을 측정하는 지표를 확인할 수 있다.

<표 3> 만족도 지표 차이에 대한 독립표본 t-검정 결과

		등분산 검정		평균의 동일성에 대한 t-검정		
		F	p-값	t	자유도	p-값 (양쪽)
전반적 만족도	등분산 가정	0.023	0.879	-9.714	227	0.000
	이분산 가정			-9.924	206.2	0.000
상대적 불만족	등분산 가정	1.046	0.308	3.705	218	0.000
	이분산 가정			3.660	171.6	0.000

- Step 2 : 선행연구에서 찾아낸 지표 중에서 본 연구의 성격에 적합하지 않은 것을 제외하였다. 이 과정에서 남은 지표는 총 30개였다.
- Step 3 : 이전 단계에서 선정된 지표는 직장예비군 지휘관의 직장만족도 측정에 적합하도록 설문 문항으로 만들었다.
- Step 4 : 설문조사를 통해 얻어진 데이터를 이용하여 독립표본 t-검정을 수행하여 지표별로 일반 직장예비군 지휘관과 대학직장예비군 지휘관 사이에 통계적으로 유의한 차이가 나는지를 확인하였다. 그 결과, 유의수준 0.05에서 통계적으로 유

의한 차이를 나타내는 24개 지표를 선정하였다.

<표 5>는 앞에서 기술한 지표 선정 과정을 모두 거친 후 최종적으로 선정된 측정 지표를 정리한 것이다. ‘자아 성취감’, ‘직속상사’, ‘업무 환경 및 직장의 방침’, ‘동료와의 관계’, ‘급여’, ‘업무 부하’ 등 6개 영역에 대한 총 24개 측정 지표들이 선정되었다. 이와 같이 선정된 측정 지표들은 CART 모형 구축을 위한 설명변수로 이용되었다. 참고로 각 측정 지표에 대한 설문 문항은 <표 5>의 마지막 열에 나타난 바와 같다.

<표 5> 직장만족도 영향 지표 및 설문 문항

구분	지표	설문 문항
자아 성취감	승진 기회	나는 직장에서 승진의 기회가 주어져 있다.
	적절한 보상	나는 업무 성과에 대해 적절한 보상을 받은 편이다.
	전문성 향상	나는 업무를 통해 배움으로써 전문성이 향상되는 것을 느낀다.
	성과 기반 승진	나의 직장은 업무 성과에 따라 승진을 제공한다.
	경쟁심 고취	나의 업무는 경쟁심을 고취시킨다.
	신뢰인지도	나는 직장이 나를 신뢰하고 있음을 느낀다.
직속상사	공정성	나의 직속상사는 직원을 공정하게 대한다.
	신뢰감	나는 나의 직속상사를 신뢰한다.
	업무수행력	나의 직속상사는 업무 수행이 탁월하고 효율적이다.
	긍정적 피드백	나의 직속상사는 긍정적인 피드백을 이용하는 편이다.
업무 환경 및 직장의 방침	사무실	나의 사무실은 편안하고 안전하다.
	비품	사무실 비품은 충분하다.
	가용 인력	업무 추진을 위해 나의 가용 인력이 충분하다.
	방침	직장의 방침들은 합리적이다.
	업무 매뉴얼	나는 예비군 지휘관 업무에 대한 매뉴얼을 제공받았다.
동료와의 관계	피드백 제공	나는 성과에 대해 정기적이고 시기적절하게 피드백을 받는다.
	관계	나는 업무와 관계된 다른 직원들과의 관계에 만족한다.
	커뮤니케이션	나는 다른 직원들과 커뮤니케이션이 잘 되는 편이다.
	협력	나는 업무 수행에 있어 다른 직원들과 협력이 잘 되는 편이다.
급여	친밀성	나는 직장에서 다른 직원들과 친밀하게 지낸다.
	급여	나의 급여는 업무에 비해 적당한 편이다.
	급여 산정 기준	나의 직장은 급여 산정의 기준이 명확하다.
업무 부하	상대적 급여	나는 다른 예비군 지휘관에 비해 급여가 높은 편이다.
	업무 부하	나는 과도한 업무 부하로 인해 성과의 질이 떨어지곤 한다.

*5점 리커트 척도 이용 : 1(매우 아니다), 2(아니다), 3(보통), 4(그렇다), 5(매우 그렇다).

4.3 격자탐색법을 이용한 최적 모형 선정

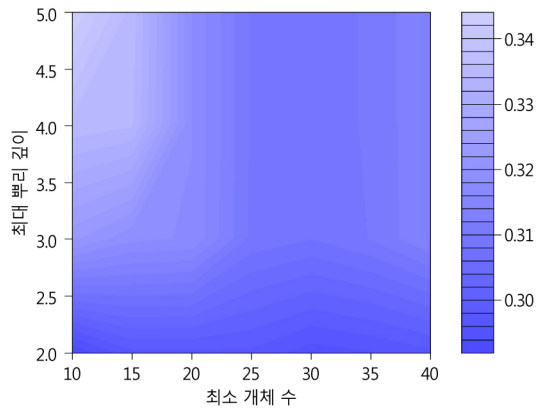
모형의 목표변수는 직장예비군 지휘관의 구분, 즉 일반직장예비군 지휘관인지 대학직장예비군 지휘관인지의 여부로 설정하였고, 설명변수는 앞에서 기술한 만족도 지표 24개로 설정하였다.

모형의 복잡성을 통제하기 위해 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수에 대한 최적 값을 격자탐색법을 이용하여 추정하였다. 우선, 전체 데이터를 무작위로 10개의 그룹으로 나눈 후, 9개의 그룹은 모형 구축을 위한 훈련용 데이터

(train set)로 이용하고, 나머지 1개 그룹은 구축된 모형의 성과를 테스트하기 위한 데이터(test set)로 이용하였다. 훈련용 데이터로 이용했던 9개 그룹에 대해서도 돌아가면서 1개 그룹씩 테스트용 데이터로 이용하고 테스트용 데이터를 제외한 나머지 9개 그룹은 훈련용 데이터로 이용하였다. 다음으로, 최대 뿌리 깊이는 2부터 1씩 증가시켜 최대 5까지 설정하였고, 각 노드에서 분리 가능한 최소 잎 수는 10부터 5씩 증가시켜 최대 40까지 설정하였다. 마지막으로, 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수의 조합 각각에 대해 테스트용 데이터 10개 세트의 오분류율(misclassification rate)을 산출하여 그 평균과 표준편차를 계산하였다. <표 6> 및 [그림 2]는 그 결과를 정리한 것이다.

<표 6> 모형별 오분류율 평균 및 표준편차

연번	모형		오분류율	
	최소 개체 수	최대 뿌리 깊이	평균	표준편차
1	10	2	0.2920	0.0866
2	15	2	0.2976	0.0986
3	20	2	0.2976	0.0986
4	25	2	0.2976	0.0986
5	30	2	0.2949	0.0936
6	35	2	0.2959	0.0939
7	40	2	0.2974	0.0946
8	10	3	0.3225	0.0905
9	15	3	0.3185	0.1050
10	20	3	0.3178	0.1061
11	25	3	0.3095	0.1038
12	30	3	0.3081	0.1019
13	35	3	0.3101	0.1011
14	40	3	0.3140	0.1054
15	10	4	0.3353	0.1077
16	15	4	0.3341	0.1091
17	20	4	0.3181	0.1062
18	25	4	0.3098	0.1038
19	30	4	0.3084	0.1019
20	35	4	0.3104	0.1013
21	40	4	0.3140	0.1054
22	10	5	0.3437	0.1075
23	15	5	0.3352	0.1080
24	20	5	0.3181	0.1062
25	25	5	0.3098	0.1038
26	30	5	0.3084	0.1019
27	35	5	0.3104	0.1013
28	40	5	0.3140	0.1054



[그림 2] 모형별 오분류율 평균

<표 6> 및 [그림 2]에서 보는 바와 같이 최대 뿌리 깊이가 2이고, 각 노드에서 분리 가능한 최소 개체 수가 10일 때 모형의 오분류율 평균이 제일 낮고, 동시에 표준편차 역시 제일 낮은 수치를 보이는 것을 확인할 수 있다. 여기서 오분류율 평균이 제일 낮다는 것은 10개의 데이터 집단에 대해 평균적으로 분류정확도가 가장 높다는 것을 의미한다. 또한, 오분류율의 표준편차가 제일 낮다는 것은 오분류율의 변동성이 가장 낮은 것을 의미하고, 이는 곧 데이터 집단이 달라짐에 따라 모형의 성과가 달라

지는 정도가 가장 낮아 결국, 모형이 가장 안정적이라는 것을 의미한다. 이에 반해 오분류율의 평균과 표준편차가 상대적으로 높은 경우, 예를 들어 최소 개체 수를 10으로 하고 최대 뿌리 깊이를 5로 설정한 모형의 경우, 오분류율의 평균은 최적 모형에 비해 0.0517만큼 높게 나타나고, 표준편차는 0.0209만큼 높게 나타나는 것을 확인할 수 있다. 이러한 결과는 의사결정나무 분석의 모형을 구축하는 데 있어 최대 뿌리 깊이 및 각 노드에서 분리 가능한 최소 개체 수에 대한 최적 값을 찾는 것이 그만큼 중요하다는 사실을 보여준다.

요컨대, 본 연구에서는 최대 뿌리 깊이가 2이고, 각 노드에서 분리 가능한 최소 개체 수가 10일 때 모형이 가장 안정적이고도 높은 분류정확도를 나타내었다. 따라서 최대 뿌리 깊이와 각 노드에서 분리 가능한 최소 개체 수에 대한 최적 값을 각각 2와 10으로 하여 최적 모형을 설계하였다.

4.4 분류 규칙에 대한 논의

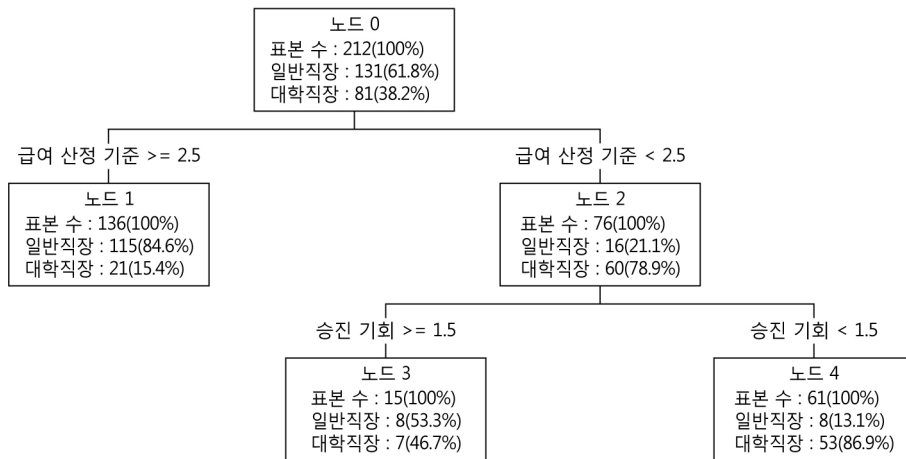
[그림 3]은 앞의 절에서 기술한 분석 결과를 토

대로 선정된 변수 및 최대 뿌리 깊이, 각 노드에서 분리 가능한 최소 개체 수에 대한 최적 값을 이용하여 구축된 의사결정나무 모형을 나타내고 있다.³⁾

이 모형에서 확인할 수 있는 분류 규칙은 총 3개인데 이를 기술하면 다음과 같다.

- 규칙 1 : 급여 산정 기준의 명확성에 대한 동의 정도가 3 이상(보통 이상)이면, 일반직장예비군 지휘관이다(분류정확도 : 84.6%).
- 규칙 2 : 급여 산정 기준의 명확성에 대한 동의 정도가 2 이하('매우 아니다', '아니다')이고, 승진 기회에 대한 인식 정도가 2 이상('아니다', '보통', '그렇다', '매우 그렇다')이면, 일반직장예비군 지휘관이다(분류정확도 : 53.3%).
- 규칙 3 : 급여 산정 기준의 명확성에 대한 동의 정도가 2 이하('매우 아니다', '아니다')이고, 승진 기회에 대한 인식 정도가 1('매우 아니다')이면, 대학직장예비군 지휘관이다(분류정확도 : 86.9%).

세 개의 규칙 중에서 규칙 2는 표본 수가 15개 지나지 않고 이 중에서 일반직장예비군 지휘관과



[그림 3] 직장예비군 분류 모형

3) 여기서 분류 규칙에 있는 1.5나 2.5와 같은 수치는 설문조사에서 5점 척도로 조사한 지표별 만족도를 나타낸다(1 : 매우 불만족, 2 : 불만족, 3 : 보통, 4 : 만족, 5 : 매우 만족). 단, 1.5나 2.5와 같이 정수로 표현되지 않은 것은 CART가 만족도에 대한 전체집합을 실수(real number)로 인식하기 때문이다. 예를 들어 어떠한 지표가 2.5 이상이라고 하면 3 이상을 의미하고, 반대로 어떠한 지표가 2.5 미만이라고 하면 2 이하를 의미한다.

대학직장예비군 지휘관의 수가 8과 7로서 변별력이 없다고 보고 제외한다고 하면 규칙 1과 3을 중심으로 해석이 가능하다. 이 두 가지 분류 규칙을 통해 얻을 수 있는 정책적 시사점은 다음과 같다.

첫째, 일반직장예비군 지휘관과 대학직장예비군 지휘관 사이에 가장 큰 차이를 보이는 직장만족도 영향 지표는 급여 산정 기준의 명확성에 대한 동의 정도이다. 급여 또는 상대적 급여 등의 지표가 있었음에도 급여 산정 기준이 첫 번째 분류 변수로서 선택된 것은 주목할 만한 부분이다. 대학직장예비군 지휘관은 일반직장예비군 지휘관에 비해 급여 산정 기준의 명확성에 대해 동의하지 않는 비율이 높게 나타났다. 이는 대학직장예비군 지휘관의 처우 개선 시 가장 중요한 것이 급여 산정을 위한 합리적인 기준을 마련하고, 이 기준에 따라 급여를 산정하는 것임을 의미한다.

둘째, 대학직장예비군 지휘관은 일반직장예비군 지휘관에 비해 급여 산정 기준의 명확성에 대한 동의 정도 뿐 아니라 승진 기회에 대한 인식 정도 역시 매우 낮은 수준을 나타내고 있다. 이는 급여 산정 기준도 명확하지 않고 승진 기회조차 거의 없는 대학직장예비군 지휘관의 처한 현실이 잘 반영된 결과이다. 한편, 승진 기회가 적다는 것은 대학직장예비군 지휘관의 대부분이 계약직으로 고용되는 사실과 관련이 깊다. 대학직장예비군 지휘관의 처우를 개선한다고 할 때 이러한 고용 형태에 있어서의 개선이 필요하다고 하겠다.

5. 결 론

의사결정나무 분석은 방법론상의 많은 장점에도 불구하고 데이터에 따라 분류 성과가 달라지는 불안정성이 가장 큰 한계점으로 지적되어 왔다. 그럼에도 의사결정나무 분석을 응용한 많은 연구에서 모형의 안정성을 확보하기 위한 과정을 찾아보기가 쉽지 않았다. 이에 본 연구에서는 의사결정나무 분석의 한계점으로 거론되는 모형의 불안정성을 극복하기 위한 방안으로 격자탐색법을 이용한 가지치기

방법을 제시하였다. 또한 실제 데이터를 이용하여 그 타당성을 실증하였다. 실증분석 결과, 최대 뿌리 깊이와 각 노드에서의 분리 가능한 최소 개체 수의 조합이 달라짐에 따라 모형의 오분류율 평균과 표준편차가 상이하게 나타나는 것을 확인할 수 있었고, 오분류율과 표준편차가 모두 가장 낮게 나타나는 최대 뿌리 깊이 및 각 노드에서의 분리 가능한 최소 개체 수의 조합을 선택함으로써 최적의 의사결정나무 분석 모형을 구축할 수 있었다.

본 연구의 분석을 통해 직장예비군 지휘관의 처우 개선을 위한 정책적 방향에 대해서는 다음과 같은 결론을 얻을 수 있었다. 우선, 일반직장예비군 지휘관과 대학직장예비군 지휘관 사이에 직장만족도에 있어 유의한 차이가 존재하는 것으로 판명되었다. 또한, 대학직장예비군 지휘관의 직장만족도를 제고하기 위해서는 급여 산정에 대한 합리적 기준을 마련하는 것이 가장 시급하며, 고용 형태의 개선 또한 필요하다는 것을 확인할 수 있었다.

이와 같은 기여점에도 불구하고 본 연구는 격자탐색법의 타당성 확보라는 측면에서 다음과 같은 한계가 있으며 이를 보완하기 위한 후속 연구가 필요함을 밝혀둔다. 먼저, 격자탐색법의 적용에 있어 본 연구에서는 직장예비군 지휘관 대상의 설문조사를 통해 얻어진 데이터를 이용하였는데 이의 타당성을 확보하기 위해서는 보다 다양한 데이터에 적용하는 것이 필요하다. 향후 연구에서는 보다 다양한 분야의 다양한 데이터를 이용하여 격자탐색법을 이용한 의사결정나무 분석 모형을 구축하고 그 타당성을 밝힐 필요가 있다. 또한, 격자탐색법의 방법론적 타당성을 확보하기 위해서는 의사결정나무 분석 모형의 선정에 관한 기존의 방법론과 그 성과를 비교하여 보여줄 필요가 있다. 예를 들어, 부트스트랩과 같은 통계 기반의 추정방법과 유전 알고리즘 등과 같은 인공지능 기반의 추정방법과의 성과 비교 등이 필요할 수 있다.

참 고 문 헌

- [1] 배재권, 김진화, “지식 결함을 이용한 서로 다

- 른 모델들의 통합”, 『경영과학』, 제24권, 제2호(2007), pp.177-196.
- [2] 이건창, 최 관, “감리지적기업의 분류적 특성에 관한 연구 : 베이지안 망과 C5.0, 그리고 앙상블 방법간의 비교를 중심으로”, 『경영학연구』, 제36권, 제3호(2007), pp.705-737.
- [3] 이성규, “의사결정나무분석을 통한 전문대학의 교원급여 결정구조에 관한 연구”, 『국제회계연구』, 제9권(2003), pp.297-312.
- [4] 이승태, 김성신, “의사결정나무를 이용한 생물의 행동 패턴 구분과 인식”, 『퍼지 및 지능시스템학회』, 제15권, 제6호(2005), pp.682-687.
- [5] 조영빈, 김채복, “온라인 소매상점에서의 효과적인 고객 분류 방법론 : 의사결정나무 기법에의 적용”, 『대한경영학회지』, 제19권, 제6호(2006), pp.2117-2134.
- [6] 정수미, 이건호, “제품별 구매고객 예측을 위한 인공신경망, 귀납규칙 및 IRANN 모형”, 『한국경영과학회지』, 제30권, 제4호(2005), pp.117-130.
- [7] 최종후, 서두성, “데이터마이닝 의사결정나무의 응용”, 『통계분석 연구』, 제4권, 제1호(1999), pp.61-83.
- [8] Al-Rubaish, A.M., S.I.A. Rahim, M.S. Abumadini, and L. Wosornu, “Academic job satisfaction questionnaire : Construction and validation in Saudi Arabia,” *Journal of Family and Community Medicine*, Vol.18, No.1(2011), pp.1-7.
- [9] Berry, M.J. and G. Linoff, *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management*, John Wiley and Sons, New York, 2004.
- [10] Brayfield, A.H. and H.F. Rothe, “An index of job satisfaction,” *Journal of Applied Psychology*, Vol.35(1951), pp.307-311.
- [11] Breiman, L., J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees*, CRC press, 1984.
- [12] Cass, M.H., O.L. Siu, E.B. Faragher, and C.L. Cooper, “A meta-analysis of the relationship between job satisfaction and employee health in Hong Kong,” *Stress and Health*, Vol.19(2003), pp.79-95.
- [13] Cooper, C.L., S.J. Sloan, and S. Williams, *Occupational Stress Indicator : Management Guide*, NFER-Nelson, Winsor, 1998.
- [14] Han, J., M. Kamber, and J. Pei, *Data Mining : Concepts and Techniques*, Elsevier, MA, 2012.
- [15] Hartigan, J.A., *Clustering Algorithm*, Wiley, New York, 1975.
- [16] Hsu, C.W., C.C. Chang, and C.J. Lin, *A Practical Guide to Support Vector Classification*, Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, 2004.
- [17] Kang, H.C., S.T. Han, and J.H. Choi, “Interpretation of data mining prediction model using decision tree,” *The Korean Communications in Statistics*, Vol.7, No.3(2000), pp. 937-943.
- [18] Lee, Y.C., “Application of support vector machines to corporate credit rating prediction,” *Expert Systems with Applications*, Vol.33, No.1(2007), pp.67-74.
- [19] Min, J.H. and Y.C. Lee, “Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters,” *Expert Systems with Applications*, Vol. 28, No.4(2005), pp.603-614.
- [20] Weiss, D.J., R.V. Dawis, and G.W. England, *Manual for the Minnesota Satisfaction Questionnaire*, Minnesota Studies in Vocational Rehabilitation, 1967.
- [21] Quinlan, J.R., *C4.5 : Programs for Machine Learning*, Morgan Kaufmann, San Mateo, 1993.