

# 데이터마이닝을 활용한 이직의도와 조직몰입의 결정요인에 대한 연구

최영준<sup>1</sup> · 심원술<sup>1</sup> · 백승현<sup>1†</sup>

## A Study of The Determinants of Turnover Intention and Organizational Commitment by Data Mining

Young Joon Choi · Won Shul Shim · Seung Hyun Baek

### ABSTRACT

In this article, data mining simulation is applied to find a proper approach and results of analysis for study of variables related to organization. Also, turnover intention and organizational commitment are used as target (dependent) variables in this simulation. Classification and regression tree (CART) with ensemble methods are used in this study for simulation. Human capital corporate panel data of Korea Research Institute for Vocation Education & Training (KRIVET) is used. The panel data is collected in 2005, 2007, and 2009. Organizational commitment variables are analyzed with combined measure variables which are created after investigation of reliability and single dimensionality for multiple-item measurement details. The results of this study are as follows. First, major determinants of turnover intention are trust, communication, and talent management-oriented trend. Second, the main determining factors for organizational commitment are trust, the number of years worked, innovation, communication. CART with ensemble methods has two ensemble CART methods which are CART with Bagging and CART with Arcing. Comparing two methods, CART with Arcing (Arc-x4) extracted scenarios with very high coefficients of determination. In this study, a scenario with maximum coefficient of determinant and minimum error is obtained and practical implications are presented. Using one of data mining methods, CART with ensemble method. Also, the limitation and future research are discussed.

**Key words** : Organizational commitment, Turnover intention, Data mining, CART, Ensemble

### 요 약

본 논문에서는 조직관련 변수들의 연구를 위해 이직의도와 조직몰입을 목표(종속)변수로서 데이터마이닝 시뮬레이션을 실시하여 접근 방법을 찾고 분석결과 도출을 목적으로 하였다. 데이터마이닝 분석방법 중 CART 앙상블 기법을 활용하였다. 자료는 한국직업능력개발원의 인적자본기업패널조사 1차~3차(2005~2009)데이터를 사용하였다. 조직몰입 변수는 다항목 측정 사항에 대해 신뢰성, 단일차원성 검토를 실행 후 합산척도 변수를 생성하여 분석하였다. 본 연구 결과는 다음과 같다. 첫째, 이직의도에 대한 주요 결정요인은 신뢰, 커뮤니케이션, 인재 중시 풍조 아이템으로 나타났다. 둘째, 조직몰입에 대한 주요 결정요인은 신뢰, 근속기간, 혁신, 커뮤니케이션 아이템으로 나타났다. 데이터마이닝 방법의 CART 앙상블 방법으로 Bagging과 Arcing 알고리즘을 적용한 결과 Arc-x4 방법이 매우 높은 결정계수를 나타낸 시나리오를 추출했다. 본 연구에서는 데이터마이닝 방법 중 하나인 CART 앙상블 시뮬레이션을 통해 최대치의 결정계수, 최소치의 오류를 산출한 시나리오 모델을 도출하고 실무적 시사점을 제시하였으며 한계점 및 향후 연구에 대해 논의되었다.

**주요어** : 조직몰입, 이직의도, 데이터마이닝, CART, 앙상블

접수일(2014년 2월 6일), 심사일(2014년 3월 11일),  
게재 확정일(2014년 3월 20일)

<sup>1)</sup>한양대학교 일반대학원 전략경영학과

주 저 자: 최영준

교신저자: 백승현

E-mail; sbaek4@hanyang.ac.kr

## 1. 서 론

기업경영에서 의사결정은 중요도가 매우 높은 사안이 다. 의사결정이 필요할 때마다 예측 시나리오를 설정하고 시뮬레이션을 실행하는 현상은 이제 SF영화 속의 일이

아니다. 그 예로 미국의 대통령 선거캠프에서는 선거운동 기간 동안 빅데이터 예측모델 시뮬레이션을 약 7만회 실행하며 의사를 결정하고 선거운동을 실시하였다. 분야를 막론하고 기후변화, 해양정보, 국가기관, 기업 등 모두 빅데이터를 분석하고 활용하기 위해 노력하고 있다. 다양한 정보들이 기록, 저장되어 축적된 고용자원들을 활용할 방법 중 하나가 바로 데이터마이닝이다.

목표개체에 대해 한 가지로 통일하는 전체주의(holism) 관점과 다원주의(pluralism) 관점이 서로 양립하며 존재한다. 양립하는 개체를 균형(equilibrium)으로 설명하는 연구자도 있다. 한편 양자론에서는 파동함수를 활용하여 끊임없이 분류한다. 이러한 분류개체는 각각의 층(layer)을 형성하여 다시 정의를 내리고 연구된다. 이는 기업경영의 의사결정 흐름과 유사하다. 의사결정 역시 관점별 분류를 실행한다. 그러나 우리는 아직 어떠한 분류 방법이 과학적(scientific)이며 합리적(rational) 접근인지 아직 해결되지 못한 과제로 남아있다.

정보통신기술(ICT)의 발달로 기업들은 다양한 도구(ERP, MIS, S/W Package, etc)를 활용하여 구성원들이 각자의 프로세스에 해당하는 업무를 진행하고 있다. 연구자들은 이러한 기업정보 현상에 대해 시스템 설계와 관리 내용 부문으로 연구가 편중되어 있다. 기업정보의 틀(frame)과 장(field)을 마련하는 세대에서 기업정보망의 축적된 빅데이터를 분석하는 세대로 전환이 필요한 시기이다. 빅데이터 시대의 조직관련 연구는 다양한 접근법으로 시뮬레이션 실행을 하지 못하고 있다. 마케팅 연구자들이 데이터마이닝을 통한 외부고객 니즈를 파악한다면 조직 연구자들은 데이터마이닝을 통해 내부고객(조직구성원) 니즈를 파악하며 학제간 융합(interdisciplinary convergence)을 시도하며 거시적(macro)관점과 미시적(micro)관점을 통합하는 연구가 필요하다.

기업에서는 훌륭한 일터(Great Work Place, GWP)만들기, 노동생활의 질(Quality of Working Life, QWL) 높이기 등의 다방면으로 이직을 낮추기 위해 노력하고 있다. 외부환경 변화의 속도가 매우 빨라지고 경기가 불황일수록 조직구성원들은 조직 내부보다 외부를 바라보게 된다. 외부를 바라보며 이직의도를 직접 실행하는 구성원이 지속적으로 증가하고 있다. 하지만 이직관련 연구의 일부변수들(3~10개)간의 관계에서 이직의도와 관련하여 명료한 관념적 정의와 타당한 결과를 도출하기는 어려운 실정이다.

사회학, 심리학, 교육학, 경영학 분야의 조직 연구자들은 조직몰입에 대해 그동안 방대한 분량의 개념정의와 실증분석 연구가 이루어졌다. 또한 최근까지 조직연구들에

서 조직몰입은 직무만족과 함께 가장 높은 빈도로 나타나는 연구주제이다. 그렇지만 대부분 조직몰입과 타 변수들과의 인과관계를 횡단적(cross-sectional) 연구로 설명하는데 그치고 있다. 횡단적 연구는 특정 시점의 단면을 해석할 수 있지만 시간의 연속성에 대입할 수 없는 문제를 내포하고 있다. 이러한 이직의도와 조직몰입 연구의 문제를 해결하기 위해서는 다른 접근과 시각이 필요하다. 실무 적용 가능하며 미래상황 예측(prediction)과 전망(prospect)할 수 있도록 다른 변수들과의 관계와 일시적 시점의 분석을 벗어난 연구가 필요하다.

따라서 본 연구의 목적은 다음과 같다. 첫째, 기존 이직의도 및 조직몰입 연구와 차별화된 데이터 기반 방법으로 조직관련 변수들을 목표로 접근하는 CART 시뮬레이션을 실행할 것이다. 둘째, 조직몰입의 종단적(longitudinal) 연구는 그 양이 충분하지 않다. 따라서 종단적 연구모형을 설계하고 분석할 것이다. 셋째, 이직의도와 관련된 주요 예측 아이템을 도출할 것이다. 과거 연구에서 이직의도는 변수들과의 분석 결과에서 통찰력(insight)을 제공하지 못하고 있다. 인과관계 규명과 가설검정은 통계적 검증에서 유의할 수 있으나 점차 실용(적용)가능성에서 문제가 발생하고 있기 때문이다.

## 2. 문헌고찰 및 데이터마이닝

### 2.1 이직의도

이직(turnover)이란 고용되어 있던 조직에서 구성원이 그 외부로 분리되어 나가는 것으로 정의되고 있으며(Hom, 2011)<sup>[8]</sup> Table 1과 같이 외부 관련 요인(External Correlates), 작업관련 요인(Work-related correlates), 개인적 요인(Personal Correlates)으로 세 개의 요인으로 구분된다. 또한 이직으로 인한 비용은 비싸다<sup>[6]</sup>. 이직으로 인해 조직구성원의 재배치에 활용되는 비용은 이직자 월급의 5배에 해당한다<sup>[16]</sup>. 이직은 그 의도에 따라 자발적(voluntary) 이직과 비자발적(involutary) 이직으로 구분된다<sup>[17]</sup>. 이직을 조직의 평가 방향에 따라 역기능적(dysfunctional) 이직과 순기능적(functional) 이직으로 구분하기도 한다.

이직은 자주 연구되는 현상으로 남아있다<sup>[18]</sup>. Porter & Steers(1973)<sup>[22]</sup>는 승진(promotion)이 일반적으로 이직과 관계되어 있음을 보고하였다. 그러나 이직과 관련한 많은 연구들에서 이직의도(turnover intention)와 이직행위(turnover behavior)의 개념을 명확히 구분하지 않고 있는 양상을 보이고 있다<sup>[13]</sup>. 한편 Seo(2008)<sup>[24]</sup>의 연구에서는 조직을 교체하는 과정에만 관심을 갖고 있는 이직 관련

**Table 1.** Correlates of Turnover (Cotton & Tuttle, 1986)<sup>[4]</sup>

External correlates	Work-related correlates	Personal correlates
Employment perceptions Unemployment rate Accession rate Union presence	Pay Job performance Role clarity Task repetitiveness Overall job satisfaction Satisfaction with pay Satisfaction with work itself Satisfaction with supervision Satisfaction with co-workers Satisfaction with promotional opportunities Organizational commitment	Age Tenure Gender Biographical information Education Marital status Number of dependents Aptitude and ability Intelligence Behavioral intentions Met expectations

연구의 문제점을 파악하고 팀교체의도와 조직교체 이직 의도를 구분하여 각각 다른 영향에 대해 보고하였다.

본 연구에서 목표변수로 설정된 이직의도는 조건에 대해 한정시킨 “현재보다 더 좋은 조건 제시의 경우”를 가 정하는 이직의도 아이템이다.

## 2.2 조직몰입

조직몰입이란 개인이 특정조직과 동일시하고, 공헌하 는 것에 대한 상대적인 강도를 의미한다<sup>[20, 21]</sup>. 조직몰입 의 연구는 초기(60년대 이전) 문헌의 양이 충분하지 않고 미시적 요인의 체계적인 연구가 부족하였다. 1960년대 후 반 엘바인 캘리포니아대학(UCI)의 개인-조직 연계 프로 젝트(Individual-Organization Linkage Project)를 시작으 로 장기간 연구하는 소규모 조직들이 확산되었다<sup>[25]</sup>.

조직몰입은 다음과 같은 세 가지의 행위적인 형태로 나 타난다. 첫째, 조직의 목표와 가치에 대한 강한 신뢰와 수용, 둘째, 조직을 위해 열심히 노력하고자 하는 의지, 셋째, 계 속 조직 구성원으로 남고자 하는 강한 욕구이다<sup>[21, 22]</sup>. 세부 적으로 조직몰입은 조직에 대한 구성원의 헌신으로 계산 적(caculative), 정서적(affective), 그리고 규범적(normative) 몰입으로 구분되어 설명된다. 본 연구에서 활용된 목표변 수를 Allen & Meyer (1991)<sup>[1]</sup> 연구에서 정의한 내용으로 구분하면 OC\_01, OC\_03 문항은 정서적 몰입이다. OC\_02 문항은 유지적(continuance) 몰입으로 구분된다(OC 아이 템은 3.2 변수선정과 Table 3을 참고). 유지적 몰입은 계산 적 몰입이 최근화된 것이며 매우 유사하다. 또한 유지적 몰입처럼 정서적 몰입은 최근 고도로 발달한 개념이다<sup>[20]</sup>.

## 2.3 데이터마이닝

다량의 데이터를 탐구하고 분석하여 의미 있는 패턴이 나 규칙들을 찾아내는 과정을 데이터마이닝이라고 한다.

데이터마이닝의 과정은 지식발견(knowledge discovery), 혹은 데이터베이스로부터의 지식발견(knowledge discovery in database, KDD)으로도 불린다. 하지만, 우리는 이보다 는 지식 창조(knowledge creation)라고 부르는 것이 더 적절하다고 생각한다<sup>[9]</sup>. 데이터마이닝의 특징은 직관적 이지 않고, 전에 가설되어지지 않았던 정보를 찾는다. 단 순히 이미 알려진 정보를 찾는 데이터마이닝은 의사결정 과정에 도움을 줄 수 없다. 또한 방대한 데이터 속에서 예 측모형이 새로운 자료에 얼마나 잘 적용되도록 하는 것인 가를 의미한다<sup>[10]</sup>.

### 2.3.1 의사결정나무

의사결정나무는 일련의 단순한 의사결정 규칙들을 적 용시켜 큰 레코드의 집합을 작은 레코드의 집단들로 나누 는 데 쓰이는 구조이다. 하나의 의사결정나무 모형은 특 정한 목표변수에 대하여 큰 다원적인 모집단을 더 작고 일원적인 집단들로 나누는 규칙들의 집합을 포함한다. 목 표변수는 분류에 관련된 것이고 의사결정나무 모형은 주 어진 레코드가 각각의 카테고리에 포함될 확률을 구하거 나, 레코드들을 가장 가능성 높은 카테고리로 구분하는 것이다. 다른 대체기법들이 존재하기는 하지만, 의사결정 나무들은 또한 연속적 변수들의 값을 추출하는 데에도 쓰 일 수 있다<sup>[9]</sup>.

의사결정나무를 형성하는 알고리즘으로 널리 사용되는 것으로 CHAID, CART, QUEST, C4.5등을 이용하는데 이들은 분리기준과 정지규칙 그리고 가지치기 등에서 서 로 다른 형성과정을 갖고 있다<sup>[10]</sup>.

### 2.3.2 CART (Classification and Regression Trees)

CART 알고리즘은 불순도(impurity)를 분리기준으로 사용하여 연속형 또는 범주형의 반응변수들을 분석하는

이진분리(binary split) 방법이다. 포괄 탐색(exhaustive search)을 이용하여 분류변수와 분류집합의 선택을 동시에 결정한다. 분류는 부모노드보다 각각의 자식노드간 불순도의 차이가 뚜렷하게 나타나는 방법을 제안하였다. 즉 범주형 반응변수인 경우 지니 지수, 엔트로피 지수, Twoing 기준을 사용한다. 연속형 반응변수인 경우 분산의 감소량을 사용한다<sup>[10]</sup>. 범주가 많은 범주형 독립변수의 경우 분리기준으로 선택될 확률이 높아 편의가 존재하게 된다<sup>[9]</sup>.

### 3. 연구방법

#### 3.1 데이터 및 분석도구

본 연구는 한국직업능력개발원의 인적자본기업패널「HCCP: Human Capital Corporate Panel, 이하 HCCP」

데이터를 사용하였다. HCCP 데이터는 한국의 기업이 지닌 인력의 양적·질적 수준을 파악하고, 기업 내에서 인적자원을 축적해가는 과정과 내용을 파악하기 위하여 한국직업능력개발원 주관으로 2005년부터 수집하였다. 그 후 동일 기업을 격년(2년) 단위로 추적하는 중장기 패널조사로 되었다<sup>[28]</sup>. 본 연구 분석대상 데이터의 각 년도별 차이점은 Table 2와 같다. 년도별 차이점 주요사항은 다음과 같다. 첫째, 역량요소별 경쟁력 확보 노력(Level of Effort on Competitiveness Component), 분야별 경쟁력이 가장 높은 직급(The Highest Position on Level of Competitiveness in a Field of Staff)에 대한 조사가 2005년도 조사에서만 포함되었다. 둘째, 팀원 인사사고과 결과확인(Team Member's Performance Rating Check), 교육훈련 참여과정(Participation and Effect on HRD Practice), 노동조합

Table 2. Survey Contents of HCCP Dataset

Questionnaire Items	2005	2007	2009	Questionnaire Items	2005	2007	2009
Joining a company YY, MM	O	O	O	Acquirements Function and Technology Usefulness	O	O	O
Starting Current Position YY, MM	O	O	O	The Nature of Take Charge of Work	O	O	O
Joining or Current Position	O	O	O	Determine The Cause of The Problem in Charge of The Work	O	O	X
Current Work	O	O	O	Understanding of The Entire Work	O	O	O
Starting Current Work	O	O	O	Opinion for Team Member	O	O	O
HR department Evaluation	O	O	O	Controlling The Team's Autonomy	O	O	O
Expertise of HR Specialists	O	O	O	Opinion for Company and Work-Related	O	O	O
Competitiveness Component	O	O	O	Job Performance Evaluation of Team	O	O	X
Level of Effort on Competitiveness Component	O	X	X	Controlling the Team on Process Evaluation	O	O	X
Level of Competitiveness in a Field of Staff	O	O	O	Controlling the Team on Member Evaluation	O	O	X
The Highest Position on Level of Competitiveness in a Field of Staff	O	X	X	Distribution of Skill Levels of Manufacturing Workers	O	O	O
Practical Level on Information System	O	O	X	Skilled Manufacturing Workers Level Representative of The Production Line	O	O	O
Team Member's Performance Rating Check	X	O	O	Production Equipment Level Representative of The Production Line	O	O	O
Participation and Effect on HRD Practice	O	O	O	Whether Permanent Worker	O	O	O
Supporting in Retirement and Aging Society	X	X	X	Whether Apprentice	O	O	O
Days of Participation on Company Training	X	X	O	Membership of Unions	X	O	O
Evaluation of Official Training Company	X	X	O	Working Hours	O	O	O
Participate in Training Courses	X	O	O	Salary	O	O	O
After Joining a Company until Now to Perform the Duties	O	O	X	Gender	O	O	O
Previous Job	O	O	X	Birth YY, MM	O	O	O
After Joining a Company until Now the Number of Teams Worked	O	O	O	Marriage	O	O	O
At The Time of Joining a Company and The Current Skill Levels	O	O	O	Education	O	O	O
Certification	O	O	O				

가입여부(Membership of Unions)에 대한 항목이 2007년 조사에서 추가되었다. 셋째, 회사 교육훈련 참여일수(The Days of Participate in Training Courses), 회사차원의 공식훈련에 대한 평가(Evaluation of Official Training Company), 동기부여와 조직몰입 문항에서 일부 보강 항목이 2009년 조사에서 추가되었다<sup>28)</sup>.

이직의도와 조직몰입의 결정요인을 분석하는데 Salford Predictive Modeler (이하 SPM) 7.04 버전을 활용하였다. SPM은 Salford Systems (S/W 개발업체)에서 개발·보급하는 상용 소프트웨어로서 데이터 시나리오 예측 분석과 미래사건의 시뮬레이션 용도에 적합하다. SPM 소프트웨어를 사용해서 가능한 데이터마이닝 기법은 의사결정나무, 데이터 앙상블, 확률예측(random forest), 트리넷(TreeNet) 등이 있다<sup>5)</sup>.

### 3.2 변수 선정

변수 선정 시 훈련데이터를 사용하여 테스트를 실시하고 선정하였다. 종속변수 항목은 Table 3 목표변수(Target Variables)와 같이 구성되어 있다. 이직의도에 대한 항목은 2005년도부터 동일한 변수로 측정되었다. 조직몰입에 대한 항목은 2005년도 설문조사에 조사되지 않아 분석에 포함되지 않았다. Table 3에 사용된 목표변수 조직몰입에 대한 각각의 항목은 다음과 같다. 첫째, OC\_01 항목은 『나는 이 회사의 문제를 내 문제처럼 느낀다.』라고 질문한 문항이다. 둘째, OC\_02 항목은 『회사를 떠난다면 많은 것을 잃게 될 것이다.』라고 질문한 문항이다. 셋째, OC\_03 항목은 『회사에 충성할 만한 가치가 있다.』라고 질문한 문항이다.

오분류(misclassification)의 비율을 최소화하기 위해 년도별 데이터에서 각각 결측치(missing data)를 제거하였다. 결측 제거 후 2005년 데이터는 총 77개의 변수에 대한 13101개의 표본이 활용되었다. 2007년 데이터는 81개의 변수에 대한 11474 표본, 2009년 데이터는 87개의 변수에 대한 10020 표본이 분석대상으로 사용되었다.

### 3.3 앙상블(Ensemble)

일반적으로 의사결정나무 분석법에서 사용되는 단일 분류자(single classifier)를 통한 분석은 다음과 같은 단점을 갖고 있다. 첫째, 데이터가 조금이라도 변하는 경우 모델이 쉽게 변한다. 둘째, 연속형 변수를 비연속적인 값으로 취급하여 분리의 경계 근방에서 예측오차가 높아질 가능성이 있다. 이러한 단일 분류자를 통한 분석의 불안정성을 보완하기 위해 제안된 기법이 앙상블 기법이다<sup>2, 7, 23)</sup>.

Table 3. Target Variables

Target Variable	2005	2007	2009
		Turnover Intention	
			OC_01
			OC_02
			OC_03

앙상블이란 단일모형들보다 더 나은 예측모형을 얻기 위해 여러 모형들을 사용하는 기법으로 더 좋은 예측력과 정확성을 갖는다<sup>11)</sup>.

ROC (Receiver-Operating Characteristic curve) 그래프는 민감도(sensitivity)와 특이도(specificity)를 플로팅(plotting)하여 설정한 모델에 따른 예측값이 실측값과 얼마나 잘 부합되는가를 파악한다<sup>14)</sup>. 본 연구에서는 데이터 앙상블 기법의 ROC 그래프 축에 대한 민감도(X)와 1-특이도(Y)를 설정하고 예측 성공 테이블을 선택한 환경 하에서 분석을 실시하였다.

#### 3.3.1 Bagging 알고리즘

Bagging (Bootstrap AGGREGatING)은 Breiman(1996)<sup>12)</sup>에 의해 처음 소개된 알고리즘으로서 다양한 부트스트랩 표본으로부터 얻은 분류자들을 하나의 분류자로 만드는 기법이다.

Bagging 과정을 다음과 같은 단계로 나타낼 수 있다.

단계 1. 학습용 자료 L에서 n번째 관측치의 확률 $p$ 을 정의한 후 분포  $p(n)$ 으로부터 N개의 표본을 추출한다.

단계 2. 이 과정을 동일하게 T번 반복한다. 여기서 각각의 표본  $L_1, \dots, L_k$  은 부트스트랩 표본이다.

단계 3. T개의 부트스트랩 표본에 대해서 분류자  $C_1, \dots, C_T$  를 계산한다.

단계 4. 각 집단에 대해  $C_1, \dots, C_T$  의 분류자를 결합하여 하나의 분류자  $C^*$  를 생성한다<sup>12)</sup>.

본 연구에서는 앙상블 방법의 Bagging 항목을 선택하고 가지치기(pruning) 방법에서는 교차타당성(cross-validity) 10회 가지치기 검증 방법을 선택하였다. 세부 보고(report detail) 항목에서는 1차(initial) 트리모형, 앙상블 트리모형, 반복사례를 선택한 환경 하에서 분석을 실시하였다.

#### 3.3.2 Arcing 알고리즘

Arcing (Adaptively resample and combine)이란 용어는 Breiman(1998)<sup>3)</sup>이 처음 사용하였다. 이는 순차적으

1)  $p(n) = \frac{1}{N}$

로 표본을 재추출하고 결합시키는 알고리즘의 집합을 의미한다. 즉, 기계학습(machine learning)에서의 부스팅 기법과 유사한 의미이다. 브레이만은 AdaBoost를 다음과 같이 비판했다. AdaBoost를 검증한 결과 이 알고리즘의 성능은 특별한 수식이 아니라 순차적으로 표본을 재추출하는 성질 때문이다. 증가되는 가중치를 가진 관측치에서는 자주 오분류되는 경향이 있다. Freund & Schapire (1996)<sup>7)</sup>의 연구에서 10개의 실제 데이터셋을 분석한 결과 Bagging 알고리즘보다 Arc-x4 알고리즘에서 더 유효한 결과로 분석되었다.

브레이만이 제안한 Arc-x4 알고리즘은 다음과 같다.

단계 1. 학습용 자료로서 표본  $L = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 을 얻는다. 여기서  $x_i$ 는 벡터를 의미한다.  $y_i$ 는 설명변수  $x_i$ 에 의해 영향을 받는 이산형 확률변수를 의미한다.

단계 2.  $k$ 번째 단계 ( $k=1, \dots, T$ )에서 현재 확률  $p^{(k)}(m)$ 을 사용하여 학습용 자료  $L$ 로부터 복원 추출을 하여 새로운 학습용 자료  $L_k$ 를 생성한 후,  $L_k$ 를 사용하여 분류자  $C_k$ 를 생성한다.

단계 3.  $m(n)$ 을  $C_1, \dots, C_T$ 에 의해  $n$ 번째 관측치의 오분류된 개수라고 정의하자.

단계 4. 갱신될  $k+1$ 번째 확률<sup>2)</sup>을 나타낸다.

단계 5.  $k$ 번째 단계를 마친 후,  $C_1, \dots, C_T$ 들은 하나의 분류자로 생성에 의해 하나의 분류자를 생성한다. 즉, 학습용 자료  $L_1$ 은 확률  $p(n)$ 을 갖는 학습용 자료  $L$ 로부터 추출한 것이며, 또 다른 학습용 자료  $L_2$ 도 같은 방식으로 추출된 자료이다. 이런 방식으로 학습용 자료  $L_T$ 까지를 순차적으로 생성한다<sup>12)</sup>.

본 연구에서는 앙상블 방법의 Arcing 항목을 선택하고 가지치기 방법에서는 교차타당성 10회 가지치기 검증 방법을 선택하였다. 세부 보고 항목에서는 1차 트리모형, 앙상블 트리모형, 반복사례를 선택한 환경 하에서 분석을 실시하였다. 또한 Arcing 알고리즘의 여러 가지 종류 중 오분류 경향을 개선한 Arc-x4의 방법에 따라 Arcing 분석의 지수(exponent)는 4로 설정하고 분석되었다.

## 4. 분석결과

### 4.1 신뢰성 및 정확도 검증

본 연구에서 사용한 목표변수에 대한 신뢰성은 크론바흐 알파(cronbach's  $\alpha$  coefficient) 값으로 검증하였다. 조

직몰입 변수는 각각 (2007년 조직몰입의  $\alpha = .782$ ), (2009년 조직몰입의  $\alpha = .802$ )로 나타났다. 통상적으로 바람직하다는 신뢰성계수  $\alpha$ 값의 기준인 .70 이상으로 모두 양호하게 나타났다. HCCP 데이터의 개별 항목들을 합산척도(summated scales)로 생성하기 위해 각 구성개념(construct)별 주성분분석(principal component analysis)을 실시하였다. 주성분분석 결과 2007년 조직몰입 변수의 추출제곱합적재값(Extraction sums of squared loadings)은 전체 2.109 분산은 70.317% 설명력을 보였으며 2009년 조직몰입 변수의 추출제곱합적재값은 전체 2.165 분산은 72.163% 설명력을 나타내고 있다. 주성분 분석 결과를 통해 하나의 주성분(one principal component)으로 묶이게 되어 단일차원성(unidimensionality)을 충족시킨다. 따라서 신뢰성과 단일차원성 검증을 실시한 조직몰입 변수는 합산척도로 재구성되어 목표변수로 분석되었다.

### 4.2 목표변수 예측의 중요점수 산정

본 연구에서는 각 목표변수(Y)의 상대적 중요도 결과를 분석하기 위해 예측변수(Predictors:  $X_1, \dots, X_k$ , Categorical:  $X_i, \dots, X_k$ )를 선택하고 분석되었다. 두 개의 목표변수(TI, OC)는 통합 데이터에서 Y변수에 해당하여 서로의 X변수에서 각각 제거(deleted) 후 분석되었다. 목표변수와 차수마다 상위 20개의 예측변수들의 상대적 중요도 점수를 배열(sort)하고 상위 10개의 변수들이 결과 항목으로 선택되었다.

목표변수 중 이직의도의 중요도 항목은 Table 4 이직의도와 조직몰입의 중요도(Importance of Turnover Intention and Organizational Commitment) 원편과 같이 나타났다(각 아이템의 세부 내용은 부록을 참고). 이직의도에 대한 중요도는 2005년에는 『현재 직장에 만족함』에 대한 항목이 가장 높게 나타났으며 『하는 일에서 성취감과 보람』이 그 뒤를 이어 높은 중요도 점수결과로 나타났다. 2007년에는 『구성원간의 신뢰관계가 돈독함(아는 사람들 외에는 신뢰하기 어려움)』이 가장 높게 나타났다. 『나이(태어난 해)』, 『직원들이 업무에 자발적이라도 유인함』이 그 뒤를 이어 높은 중요도 점수결과로 나타났다. 2009년에는 가장 높은 중요도 항목으로 『회사 경영진을 믿고 따를 수 있음』 항목으로 나타났다. 그 다음으로 『부서간 커뮤니케이션이 잘 됨』, 『평가 및 보상이 공정하게 이루어짐』 항목이 나타났다.

목표변수 중 조직몰입의 중요도 항목은 Table 4 오른편과 같이 나타났다. 조직몰입에 대한 중요도는 2007년에는 『아는 사람들 외에는 신뢰하기 어려움』이 가장 높은

2)  $p(n) = \frac{1+m(n)^4}{\sum(1+m(n)^4)}$

**Table 4.** Importance of Turnover Intention and Organizational Commitment

Target Variable	Relative Importance			Target Variable	Relative Importance	
	2005	2007	2009		2007	2009
Turnover Intention	A_67 (100.00)	B_58 (100.00)	C_65 (100.00)	Organizational Commitment	B_58 (100.00)	C_74 (100.00)
	A_64 (81.17)	B_77 (91.77)	C_62 (99.99)		B_60 (91.947)	C_66 (99.656)
	A_62 (74.49)	B_69 (91.33)	C_64 (99.98)		B_69 (88.961)	C_75 (98.833)
	A_69 (49.19)	B_56 (68.59)	C_60 (99.97)		B_77 (69.883)	C_64 (98.714)
	A_54 (48.03)	B_70 (65.10)	C_61 (99.96)		B_61 (59.797)	C_61 (98.676)
	A_66 (47.30)	B_55 (60.89)	C_70 (99.96)		B_54 (58.452)	C_81 (98.092)
	A_53 (44.60)	B_11 (5.35)	C_58 (0.85)		B_8 (2.436)	C_83 (31.535)
	A_59 (29.01)	B_79 (5.00)	C_59 (0.84)		B_10 (2.390)	C_72 (9.168)
	A_74 (23.74)	B_61 (4.96)	C_29 (0.83)		B_11 (2.340)	C_62 (7.680)
A_8 (22.54)	B_80 (3.63)	C_57 (0.82)	B_13 (2.325)	C_11 (2.390)		

\*inside parenthesis score of importance(percentage)

**Table 5.** Coefficient of Determination

Target Variables	Time	Mode	R <sup>2</sup>	Nodes	Relative Error	RMSE
Turnover Intention	2005	B	0.940	1870	0.060	0.282
		A	0.997	669	0.003	0.195
	2007	B	0.945	1646	0.055	0.257
		A	0.999	375	0.001	0.164
	2009	B	0.945	1394	0.055	0.238
		A	0.999	333	0.001	0.138
Organizational Commitment	2007	B	0.959	1809	0.041	0.141
		A	0.999	769	0.000	0.119
	2009	B	0.965	1563	0.035	0.137
		A	0.999	116	0.001	0.121

\*B=Bagging, A=Arc-x4

중요도 점수로 나타났으며 『회사 경영진을 믿고 따를 수 있음』, 『현재 일에 만족』에 대한 항목이 그 뒤를 이어 높은 중요도 점수결과로 나타났다. 2009년에는 『직원들이 업무에 자발적이라도 유인함』 항목이 가장 높은 중요도 점수로 나타났으며 『변화를 두려워하며 새로운 시도를 가급적 억제함』, 『성과가 부진한 사람은 배겨내기 힘들다』 항목이 그 뒤를 이어 높은 중요도 점수결과로 나타났다.

### 4.3 결정요인 결과

RMSE는 추정 모형이 실제 값을 추정하는데 얼마나 오차를 발생시키는지 평균적으로 나타낸다. RMSE(root mean square error) 수치가 0에 가까울수록 해당 모형이 실제 값을 잘 예측한다고 판단한다<sup>[26]</sup>. R<sup>2</sup>는 결정계수(coefficient of determination)라고 한다. R<sup>2</sup>는 0~1의 값

을 가지며, 회귀모형이 자료에 높은 적합도를 가질수록 1에 가깝게 나타난다. 이는 독립변수 X가 종속변수 Y를 결정하는 정도를 나타내기 때문이다. R<sup>2</sup>가 높을수록 회귀모형은 설명력이 커지며 예측력이 커진다<sup>[15]</sup>. 따라서 SPM 분석을 통한 알고리즘별(Bagging, Arc-x4) 결과에서는 결정계수 모델의 정확도를 높이기 위해 RMSE 수치가 최소값인 모델을 선정하였다.

Table 5 결정계수(Coefficient of Determination)는 RMSE 수치가 가장 낮은 값으로 나타난 모형의 결정계수를 보여준다. B모드와 비교하여 A모드의 R<sup>2</sup> 수치가 모든 모델에서 높게 나타났다(Bagging, Arcing 알고리즘에 대해서는 3.3 양상불 참고). 이직의도 변수의 2005년을 제외하고 나머지 분석 모델에서는 R<sup>2</sup> 수치가 모두 .999로 매우 높게 나타났다. 노드와 상대적 오류는 우하향곡선

형태로 나타난다. 따라서 가지치기를 수행할 경우 노드의 수는 줄어들고 상대적 오류 또한 감소한다. 상대적 오류(relative error) 분석 결과에서도 B모드보다 A모드가 모두 낮게 나타났으며 이직의도 변수의 2005년을 제외한 나머지 모델에서는 .001~.000으로 매우 양호한 결과가 나타났다. 각 목표변수와 시점에 대한 모델 중 종합적으로 분석결과가 가장 적합한 모델은 목표변수 조직몰입의 2007년으로 A모드에서  $R^2$ 는 .999, 769개의 노드와 .000의 상대적 오류, RMSE는 .119로 나타났다.

### 5. 결 론

지금까지 본 논문에서는 데이터마이닝의 의사결정나무 기법 중 CART 양상불 기법의 알고리즘들을 활용 시뮬레이션하여 목표변수의 분석을 실시하고 그 과정을 살펴보았다. 일반적으로 과학철학(philosophy of science)인 논리경험주의(logical empiricism)는 경험론(empiricism)과 합리론(rationalism)을 토대로 하는 것으로 볼 수 있다. 먼저 이론, 과거의 연구성과(finding) 등으로부터 논리를 토대로 현상에 대한 예측을 이끌어 낸다. 이 예측이 곧 연구가설(research hypothesis)로 설정된다. 흔히 가설은 변수들 간의 인과성(causality)으로 표현된다<sup>15)</sup>. 사회적 과정은 사람들의 주관성(subjectivity)과 자유의지가 내재되어 있으며 인간의 행태는 가변성(variability)을 갖고 있어 예측이 매우 어렵다<sup>14)</sup>.

본 논문에서 데이터마이닝 시뮬레이션을 통해 제공하는 시사점은 다음과 같다. 첫째, 이직의도에 대한 결정요인은 신뢰, 커뮤니케이션, 인재중시 풍조로 나타났다. 커뮤니케이션 아이টে은 2005년부터 점차 확대되는 점수로 나타났으며 신뢰는 다수의 아이টে에서 하나의 아이টে으로 축소되는 추세가 나타났다. 또한 2009년 결과에서는 『인재의 중요성 강조』, 『우수한 인재 우대』 그리고 『인적자원 개발에 대한 명확한 비전소유』, 『공식적인 교육훈련 참여일 수』 항목이 높은 중요도 점수로 나타났다. 인재중시 풍조와 인적자원개발(교육) 관련 다수의 아이টে이 두드러지며 변화했다. 둘째, 조직몰입에 대한 결정요인 주요 아이টে은 신뢰, 혁신, 근속기간, 커뮤니케이션 순으로 나타났다. Fig. 1 근로자 이직률(Employee Turnover Rate) 그래프는 고용노동부에서 조사한 고용노동통계자료이다. 사업체노동력조사의 연고용통계자료이며 2009년 수치는 월별데이터의 평균으로 실제 이직률 변화에 대해 작성되었다. 상대적 중요도 점수에서 신뢰 아이টে의 점수가 감소되며 커뮤니케이션 아이টে의 점수는 확대되는 시점인

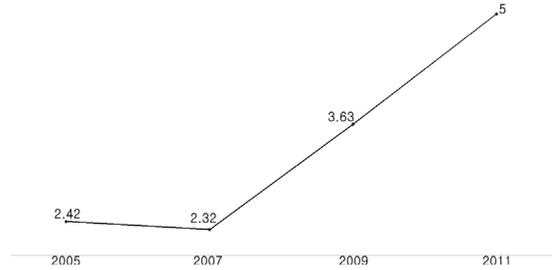


Fig. 1. Employee turnover rate

2007년 이후로 이직률은 매우 높아지고 있다. 2007년의 경우 이직률과 퇴직으로 인해 총 근로자수 또한 감소하고 있다<sup>26)</sup>.

본 연구의 2009년 중요도 점수 결과에서 기업들은 위기를 극복하려 이직률을 낮추고 관리하기 위해 인재의 중요성이 부각되고 교육(인적자원개발) 아이টে이 높은 중요도 점수로 분석되었다. 따라서 본 연구의 데이터마이닝 시뮬레이션의 직접적인 결과인 각 년도별 결과 아이টে과 추세는 간접적 자료인 외부기관 조사자료와 비교분석한 결과 연구의 직접결과와 외부 간접결과들은 연속성을 나타내고 있음을 파악할 수 있다.

본 연구의 한계점은 다음과 같다. 첫째, 본 데이터마이닝 시뮬레이션 결과는 2차 자료 데이터의 10000~14000개의 표본을 바탕으로 실시한 분석으로 2차 자료 내용의 이해와 결과 해석에 유의해야 한다. 둘째, 데이터마이닝 기법을 활용하고 있으나 대규모 용량의 데이터마이닝 시뮬레이션과 분석 전용 하드웨어를 활용하지 않아 미처 발견하지 못한 기술적 오류가 포함되어 있을 수 있다. 셋째, 조직적 관점에서 데이터베이스의 한계로 측정수준(level of measurement)이 고려되지 않았다. 리더와 조직구성원의 결과 차이, 개인과 그룹(팀, 조직)의 차이, 조직과 조직간의 차이를 고려하여 분석되지 않았다.

향후 연구에서 고려해야 할 사항은 다음과 같다. 첫째, 시뮬레이션 설계에서 다차원적 접근이 필요하다. 조직몰입과 이직의도와 직결되는 기업의 비용효율성에 대한 시나리오 설계, 개인의 이직 전환비용 혹은 생산성 변동량에 대한 시나리오 설계가 필요하다. 둘째, 데이터마이닝 시뮬레이션에 적합한 하드웨어(분석 데이터규모를 허용할 수 있는)와 분석도구(프로그램)를 사전에 철저히 준비해야 한다. 셋째, 조직관련 변수 설정에서 조직구성원의 몰입이 팀에 해당하는 몰입인지, 조직에 몰입하는 경우인지를 구분하며 연구해야 할 것이다. 조직 구성원의 이직의도와 기업이 이직에 대비할 수 있는 이직방지 프로그램

과 연결되는 연구, 집단적(팀, 그룹)으로 이직하려는 의도와 실제 집단적 이직이 발생하는 과정, 개인들이 이직의도에서 이직행동으로 연결되는 세부과정에 대한 연구를 고려해야 할 것이다.

## References

1. Allen, N. J., Meyer, J. P., "The Measurement and Antecedents of Affective, Continuance and Normative Commitment to The Organization," *Journal of Occupational Psychology*, Vol.63, No.1, pp. 1-18, 1991.
2. Breiman, L., "Bagging Predictor", *Machine Learning*, Vol. 24, No. 2, pp. 123-140, 1996.
3. Breiman, L., "Arcing Classifiers." *The Annals of Statistics*, Vol. 26, No. 3, 801-849, 1998.
4. Cotton, J. L, Tuttle, J. M., "Employee Turnover. A Meta-analysis and Review with Implications for Research", *Academy of Management Review*, Vol. 11, No. 1, pp. 55-70, 1986.
5. Dan Steinberg, Mikhail Golovnya, CART 6.0 User's Guide Salford Systems, San Diego, pp. 10-16, 2007.
6. Edward E. Lawler III, "The Design of Effective Reward Systems" J. W. Lorsch(Ed), *Handbook of Organizational Behavior*, Englewood Cliffs, NJ: Prentice Hall, pp. 255-271, 1987.
7. Freund, Y., Schapire, R., "Experiments with a New Boosting Algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference* (L. Saitta, ed.) SF: Morgan Kaufmann, pp. 148-156, 1996.
8. Hom, P. W., *Organizational Exit*. In S, Zedeck (Ed), *APA handbook of Industrial and Organizational Psychology*, Washington, DC: American Psychological Association, pp. 325-375
9. Jung, Y. H., "The Study of a Variety of Data Mining Techniques to Improve Performance Based on Ensemble Method", Master Thesis, University of Korea, pp. 5, 2009.
10. Kim, J. H., "Data Mining Case Study Using Decision Tree Analysis", Master Thesis, Ehwa Womans University, pp. 3-5, 1998.
11. Koo, J. E., "A Comparative Study of Random Forests", Master Thesis, University of Korea, pp. 2-6, 2010.
12. Lee, C. Y., "A Study on Comparison for The Methods of Decision Tree: Bagging, Arc-x4", MS Thesis, University of Yonsei, pp. 14-15, 2000.
13. Lee, S. E., "Psychological Empowerment Enterprise that New Employees Aware of The Impact of Employee Turnover on Research" Master Thesis, University of Korea, pp. 9-14, 2012.
14. Lee, H. Y., Roh, S. C., "Advanced Statistical Analysis for Theory and Practice", 2nd Edition, Moonwoosa, pp.403-405, 2013.
15. Lee, H. S., "Regression Analysis for Social Science Research" *Jypheonjae*, pp. 2-6, 2012.
16. Macy, B. A, P. H. Mirvis, "A Methodology for Assessment of Quality of Work Life and Organizational Effectiveness in Behavior-Economic Terms." *Administrative Service Quarterly* Vol. 21 pp. 217-26, 1976.
17. Mobley, W. H. "Employee Turnover: Causes, Consequences, and Control." Reading, Mass.: Addison-Wesley., 1977.
18. Mobley, W. H., Griffeth, R. W., Hand, H. H., Meglino, B. M. "Review and Conceptual Analysis of the Employee Turnover Process", *Psychological Bulletin*, Vol. 86, No. 3, pp. 493-522, 1979.
19. Michael J. A. Berry, Gordon S. Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, 2nd Edition, Wiley Publishing, Indianapolis, pp. 20-21, 2004.
20. Morrow, P. C., "The Theory and Measurement of Work Commitment" London: Jai Press, pp. 71-108, 1993.
21. Mowday, R. T., Porter, L. W., Steers, R. M., "Employee-organizational linkages: The psychology of commitment, absenteeism, and turnover. NY: Academic Press., pp. 21-35, 1982.
22. Porter, L. W., Steers, R. M. "Organizational, Work, and Personal Factors in Employee Turnover and Absenteeism." *Psychological Bulletin*, Vol. 80, No. 2, pp. 151-176, 1973.
23. Roh, J. S., Baek, S. H., Jeon, S. G., "Analysis of Relative Importance of HR Practice Using Data Mining Method: Focus on Manufacturing Companies", *Journal of The Korea Society for Simulation*, Vol. 22, No. 3, pp. 55-69, 2013.
24. Seo, J. H., "Team Change and Turnover Intentions: The Role of Team and Organizational Commitment" *The Korean Journal for Human Resource Development*, Vol. 10, No.3, pp.1-21, 2008.
25. Smith, K. G., Hitt, M. A., "Great Minds in Management -The Process of Theory Development-" Oxford University Press., pp.171-189, 2007.
26. Su, X. and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", *Advances in Artificial Intelligence*, 2009.
27. Federation, K. I., "Employment Trends in 2007 and Recruiting Plans of Major Companies Surveyed", The

Federation of Korean Industries., 2007.

28. KRIVET, "Human Capital Corporate Panel 1st-4th Survey - User Guide -", KRIVET, pp. 1-21., 2013.

## 부 록

- A\_67: 현재 직장에 만족함
- A\_64: 하는 일에서 성취감과 보람을 느낌
- A\_62: 직원들이 업무에 자발적이라도 유인함
- A\_69: 회사의 비전과 가치관을 신봉함
- A\_54: 우수 인재 우대함
- A\_66: 현재 일에 만족함
- A\_53: 회사 사정을 직원들에게 소상하게 알림
- A\_59: 동료가 힘든 상황이면 적극적으로 도움
- A\_8: 현직장 입사시기(년)
- B\_58: 아는 사람들 외에는 신뢰하기 어려움
- B\_69: 직원들이 업무에 자발적이라도 유인함
- B\_56: 상급자에게 자유롭게 의견을 제시함
- B\_70: 성과가 부진한 사람은 배겨내기 힘들
- B\_55: 회사 사정을 직원들에게 소상하게 알림
- B\_11: 현재직급 근무시기
- B\_79: 결혼
- B\_61: 현재 일에 만족
- B\_80: 학력
- B\_60: 회사 경영진을 믿고 따를 수 있음
- B\_61: 변화를 두려워하며 새로운 시도를 가급적 억제함
- B\_54: 인재의 중요성 강조
- B\_8: 현직장 입사시기(년)
- B\_10: 현재직급 근무시기(년)
- B\_13: 현재 직급
- C\_65: 회사 경영진을 믿고 따를 수 있음
- C\_62: 부서간 커뮤니케이션이 잘 됨
- C\_64: 평가 및 보상이 공정하게 이루어짐
- C\_60: 회사사정을 직원들에게 소상하게 알림
- C\_61: 상급자에게 자유롭게 의견을 제시함
- C\_70: 현재 일에 만족
- C\_58: 인적자원개발에 대한 명확한 비전 소유
- C\_59: 인재의 중요성 강조
- C\_29: 공식적인 교육훈련 참여일 수
- C\_57: 우수한 인재 우대
- C\_74: 직원들이 업무에 자발적이라도 유인함
- C\_66: 변화를 두려워하며 새로운 시도를 가급적 억제함
- C\_75: 성과가 부진한 사람은 배겨내기 힘들
- C\_81: 주당 평균 초과 근무시간
- C\_72: 현 직장의 인간관계 만족
- C\_11: 현재직급 근무시기(년)
- A\_74, B\_77, C\_83: 생년(나이)



**최 영 준** (passionyj@hanyang.ac.kr)

2007 한국능률협회컨설팅 전략/HR 컨설턴트  
2009 고려아카데미컨설팅 교육 평가위원  
2012~현재 한양대학교 전략경영학과 박사과정

관심분야 : 빅데이터, 데이터마이닝, 동기유발, 집단역학, Macro OB



**심 원 술** (wsshim@hanyang.ac.kr)

1991 미국 오레곤대학교 경영학과 경영학박사  
1991~1995 미국 보이지주립대학교 경영대학 조교수  
1995~현재 한양대학교 ERICA 캠퍼스 경영학부 정교수

관심분야 : 리더십, 동기유발, 성과관리, 국제경영



**백 승 현** (sbaek4@hanyang.ac.kr)

2000 명지대학교 산업공학과 공학사  
2002 미국 조지아공과대학교 산업공학 공학석사  
2010 미국 테네시대학교 산업공학 공학박사  
2012~현재 한양대학교 ERICA 캠퍼스 경영학부 조교수

관심분야 : 데이터마이닝, 품질경영, SCM(공급사슬관리), 생산운용관리, CRM(고객관계관리), 금융공학