

## 데이터마이닝 기법을 활용한 대졸자 고용에 미치는 영향요인 분석

김형래\*, 전도홍\*\*

### The Factors that Affects the Employment Type of The Graduates by Data-mining Approach

Hyoungrae Kim\*, Do-hong Jeon\*\*

#### 요 약

데이터마이닝 기법은 대량 데이터로부터 유의미한 정보를 추출하기 위해 고용정보 분야 등 다양한 분야로 폭넓게 활용되고 있다. 대졸자 실업, 여성 재취업, 고령자 취업 등의 불안정한 고용형태가 사회적 문제로 등장함에 따라 고용 안정성을 높여려는 다양한 공공 고용서비스의 노력과 연구가 진행되고 있다. 대졸자의 고용형태(상용직, 무직자, 일용직 등)에 영향을 미치는 주요 요인을 개인생활 환경 또는 학교생활 측면에서 분석한 연구 결과는 효과적인 취업 지원과 대학생의 취업 준비를 지도하는 데에 사용될 수 있다. 이러한 사회 조사 분석은 다양한 요인과 방대한 데이터양으로 인해 일반적 통계적 분석만으로는 한계가 있다. 170여개의 속성들 간의 관계 분석과 2만여 개의 응답데이터를 다루기에는 데이터마이닝 기법이 유용하다. 본 연구는 고용형태에 영향을 미치는 요인을 학교생활 요인, 개인환경 요인, 또는 직장생활 요인으로 구분하고, 의사결정 알고리즘을 통해 연관 관계를 분석하였다. 분석결과 주요 부모님의 소득, 결혼 여부 속성과 같은 개인 환경요인이 개인의 고용형태에 큰 영향을 미치는 것으로 나타났고, 의사결정구조 모델은 87%정확성을 보였다. 학교생활 요인으로는 대학 소재 지역으로 추정되는 학교수준이 주요 요인으로 나타났다.

▶ 주제어 : 데이터마이닝, 대졸자 취업, 데이터 전처리, GOMS

#### ABSTRACT

Data mining technique can be adapted to analysing Employment information in order to discover valuable information out of large data. As the issue employment such as jobless of college graduate, recruitment for women, recruitment for elders etc. became social problem, there are

• 제1저자 : 김형래 • 교신저자 : 전도홍  
• 투고일 : 2012. 01. 17, 심사일 : 2012. 01. 28, 게재확정일 : 2012. 06. 11.  
\* 한국고용정보원(Korea Employment Information Service)  
\*\* 관동대학교(Kwandong University)

many efforts of various public employment services and studies. The factors that affects the college graduate's employment type (regular, temporary, daily) can be used to guide employment and to prepare employment for college students. In analyzing large number of attributes and the huge amount of data elements, regular statistical methods faces their limitation; therefore, data-mining technique is more suitable for the dataset of about 170 attributes and 20,000 elements. We divide the factors that may affect the employment type into personal factor, school factor, company factor, and experience factor; decision tree algorithm is used to find out the interesting relationship between the attributes of the factors and employment type. Personal factors such as the income of parents and marital status were the most affective factors to the employment type. The learned decision tree was able to classify the employment type with 87% of accuracy. We also assume the level of the school affects the employment type of the graduates.

▶ Keywords : Data mining, Graduate employment, Data preprocessing, GOMS

## I. 서 론

대졸자 실업, 여성 재취업, 고령자 취업 등 고용이 사회적 문제로 등장함에 따라 공공 고용서비스 또는 교육 개선 등 다양한 방안을 통해 이러한 문제들을 해결하려는 연구가 진행되고 있다[2,5,6,7]. 최근 국제 경제의 불안으로 인해 고용 안정화를 위한 노력도 요구된다. 이러한 사회적 필요 속에서 구직자의 행태를 파악하고자 패널통계조사(GOMS\*)가 진행되고 있다. 이러한 조사는 흥미로운 패턴이나 영향 요인을 발견하는 것으로 현상을 이해하는 것을 돕는다. 취업 요인들의 관계를 분석할 때 항목들 간의 연관성을 분석하기는 고려해야할 경우의 수가 너무 많아 연산 처리가 쉽지 않다[11]. 이에 따라 고용정보 데이터 분석의 중요성이 점차 강조되고 있다.

본 연구는 통계적 분석을 통해 대졸자의 현재 고용형태(상용직, 무직자, 일용직 등)에 영향을 미치는 학교생활과 가정환경의 요인을 분석한다. 데이터마이닝 접근법을 통해 다양한 원인을 분석하고자 한다. 고용형태에 영향을 미칠 요인을 학교생활 요인, 개인 환경 요인, 직장생활 요인으로 구분하여 측정한다. 이러한 연구는 근로자와 무직자를 구분 짓는 근본 원인을 파악하여, 노동시장을 이해하고 고용과 연관된 공공 서비스와 교육정책의 연구에 도움이 될 것이다.

본 연구의 주요 기여를 열거하면, 첫째 직업경로분석 데이터에 다양한 데이터마이닝 기법을 적용하여 타당성 조사 및 분석 모델을 도출하였다(의사결정트리의 경우 고용형태 분류

를 위한 정확성이 87%임). 둘째, 근로자와 무직자의 현재 고용형태에 영향을 미치는 학창시절 속성을 분석한 결과, 부모님의 소득 등 가정환경이 높은 영향을 미치는 것을 발견하였다. 셋째, 사회과학 데이터를 데이터마이닝 기법에 적용하기 위해 필요한 데이터 전처리과정을 그룹속성, 연계속성, 도메인으로 정리하였다.

본 논문의 이하는 다음과 같이 구성되었다: 섹션 II는 데이터마이닝 기법의 활용 및 직업변화 요인 관련연구를 설명하고, 섹션 III는 연구모형 및 조사에 대한 구체적인 설명이 이어지고, 섹션 IV는 분석에 사용될 데이터에 대한 설명과 평가 방법에 대해 기술하고, 섹션 V는 고용형태를 구분 짓는 요인에 대한 데이터마이닝 분석 결과를 제시하고, 섹션 VI는 본 논문의 연구 내용을 요약한다.

## II. 관련 연구

반정호 외[3]는 청년취업자의 특성을 고용형태별로 분석하고, 안정된 고용형태를 기준으로 불안정한 고용형태를 유지할 가능성에 영향을 미치는 변수를 청년패널자료(2002-2003)로 분석하였다. 취업청년계층이 정규직에서 비정규직이나 실업으로 대거 전환되는 것으로 나타났으며, 특히 여성이나 저학력자의 비정규직으로 전환현상이 두드러지게 나타났다. 정책적 제언으로 청년계층의 교육이나 기능수준에 적합한 직무배치를 통해 고용의 질적 수준을 향상시켜야 한다고 제시하였다. 하성호와 이재신[8]은 데이터마이닝을 고객관리기법 분석에

\*GOMS (Graduates Occupational Mobility Survey): 2년제 대학 이상 대학졸업자의 경력개발 및 직업(직장)이동경로를 장기 추적 조사하여 교육-노동시장 간 인력수급 불일치 완화 도모를 위한 정책수립의 기초 자료제공 추구

적용하여, 온라인 소매상이 직면하고 있는 경영상의 문제를 도출하였다. 본 논문은 차이점으로 청년층이 안정적 고용을 유지하는데 미치는 주요 요인과 항목을 과거 학교생활과 가정 환경과 직장생활을 함께 포함하여 선별한다.

개인의 직업변화의 인과요인에 대해서는 많은 논의가 있어 왔다[10]. 직장 이직의향은 근무여건과 취업 기회의 영향을 받고, 직업 이직의향은 인적자본과 전문경력의 영향을 받는다는 주장에 대한 다양한 실증적 실험이 이루어지고 있지만 통계적인 방법에 주로 의존하고 있다[1]. 송호근[4]은 노동태도(직무만족, 조직몰입)에 대한 인적자본(교육수준, 근무연수 등)의 영향력을 조사하였다. 본 논문은 이러한 항목을 종합적으로 데이터마이닝[11,13] 기법에 적용하여 각 항목별로 영향도를 분석코자 한다.

### III. 연구 방법

#### 3.1 데이터마이닝 기법 선별 및 전처리

최근 데이터마이닝 알고리즘이 사회과학 분야 등으로 폭넓게 적용되고 있으나, 사회과학 설문 조사 결과를 데이터마이닝 알고리즘에 적용하기는 쉽지 않다. 사회과학 분야에 널리 사용되는 잠재변수의 처리, 도메인 값의 해석 등에 있어 데이터마이닝 알고리즘을 적용하기 위해서는 데이터 전처리 기법이 필요하다. 본장에서는 데이터마이닝 적용의 필요성, 적용 알고리즘 선정, 데이터 전처리 방식을 논하고자 한다.

고용형태에 미치는 다양한 속성(약 170여개)의 관계를 파악하기 위해서는 방대한 계산 량을 필요로 한다. 고용형태를 결정 짓는 패턴은 속성의 집합으로 구성될 수 있으며, 이는 다양한 경우의 수를 생성해 낸다. 따라서 데이터마이닝 방식을 적용하는 것이 바람직하다. 데이터마이닝을 통해 분석된 결과는 고용형태와 그에 미치는 속성들 간에 패턴의 의미를 해석 가능하다.

연구에 활용되는 데이터를 분석하기 위한 가장 적합한 데이터마이닝 기법은 의사결정트리(Decision tree)이다. 본 연구에서 구분자를 결정짓는 가장 중요한 속성이 무엇인지 발견하는 것이 주요 목적이며, 의사결정 알고리즘은 여러 속성 중 영향력이 높은 속성의 집합을 발견하기에 적합하다. 데이터마이닝 기법은 학습화 단계가 필요하며, 이는 구분자가 있는 학습과 구분자가 없는 학습으로 구분된다. 본 연구의 데이터에서 구분자는 종속변수인 고용형태이며, 의사결정트리 알고리즘은 구분자가 있는 알고리즘에 해당한다. 또한, 조사 데이터

는 명목척도, 서열척도, 등간척도, 비율척도를 모두 포괄하기 때문에 의사결정트리 알고리즘은 이러한 데이터 속성을 포괄할 수 있어야 한다.

의사결정트리의 학습방식은 탐-다운 방식으로 가장 상위의 질문항목을 선정한 후 다음 항목을 결정한다. 의사결정트리의 계산 원리는 각 질문항목의 도메인 개수에 따라 하위 영역이 구분되면, 각 하위 영역 마다 가장 영향력이 큰 질문항목을 다시 결정한다. 이러한 과정은 응답값이 구분자에 따라 지정한 오류율보다 낮거나 항목수가 없을 때까지 반복한다. 의사결정트리는 쉽고 체계적으로 구분자를 결정하는 속성을 표현하는 장점을 가진다.

데이터마이닝 적용을 위해서는 속성 선택이 선행되어야 하는데, 각 질문항목이 하나의 속성을 대표하지는 않는다. 속성은 개별속성, 그룹속성, 연계속성으로 구분되는데 알고리즘 적용의 유연성 제고를 위해 그룹속성과 연계속성은 그 의미를 해치지 않는 범위에서 개별속성으로 전환한다. 개별속성은 그 속성 하나로서 의미를 가지는 것을 의미 한다(예, 지역, 학력, 평균 임금 등). 그룹속성은 개별속성 군이 하나의 그룹적 의미를 가진다(예, 학교생활 관련 질문 등). 그룹속성은 각 개별속성의 의미와 독립성을 명확히 함으로 개별속성으로 분리할 수 있다. 연계속성은 두개 이상의 속성이 하나의 의미를 가진다(예, 어학연수를 다녀왔습니까? 다녀왔다면 몇 번 다녀왔습니까?). 이러한 연계속성을 개별속성으로 분리하기 위해 가능한 연계질문을 하나의 개별속성으로 재설계 하여야 한다. 가령, ‘어학연수를 다녀왔습니까?’와 ‘다녀왔다면 몇 번 다녀왔습니까?’ 라는 두 가지 질문은 ‘다녀왔다면 몇 번 다녀왔습니까?’라는 질문으로 합친 후 그 결과 값의 도메인을 [0..n]으로 수정하여 두 가지 질문의 의미를 손실하지 않는 범위에서 재설계 한다. 또한 년, 월로 분리된 속성은 기간(월)으로 재설계 한다. 데이터 전처리 과정을 각 속성의 특성에 따라 정리하면 다음과 같다:

- 그룹속성 → 개별속성으로 전환  
(예, 학교생활 관련 속성)
- 연계속성 → 개별속성으로 전환  
(예, 어학연수 여부 와 횟수, 생년과 생월)
- 도메인 → 유의미한 정보로 변환  
(예, 입사년/월을 근무기간으로 변환)

### 3.2 연구모형

청년층의 취업률을 제고하기 위해서는 청년 취업에 영향을 미치는 요인을 우선적으로 분석할 필요가 있다. 이러한 영향요인이 파악될 경우 대학생활 지도를 통해 취업률을 높일 수 있을 것이다. 대학생활의 학점뿐만 아니라 동아리활동, 아르바이트, 어학연수 등과 같은 취업과 연관된 다양한 활동을 종합적으로 고려하여 가장 영향을 크게 미치는 요인을 파악하고자 한다.

고용의 안정성과 취업의 관계는 장기적인 관점에서 바라볼 필요가 있다. 단기적 취업보다는 적성에 맞고 장기적으로 자신의 능력을 발휘할 회사로 취업진로를 교육하는 것이 목적이기 때문이다. 비록, 산업의 다변화에 따라 이직률이 높아지고 있으며, 선진국으로 성장할수록 이러한 현상은 더욱 두드러질 것으로 예상된다. 하지만, 한국의 현실적 측면에서는 비정규직일 경우 직업 안정성 측면에서 부정적인 영향을 미치는 만큼, 정규직 여부는 고용의 안정성 및 회사 내의 전문적 위치를 가능하게 하는 기준이 되고 있다.

가정생활이 대학생활에 영향을 미치므로 대학생활이 고용형태에 미치는 영향을 조사할 때 가정생활도 함께 조사한다. 가령, 대학교 등록금을 스스로 벌기위해 아르바이트를 하는 학생과 그렇지 아니한 학생의 대학생활은 차이가 있을 것이다. 그 외에 결혼 후 학교생활, 부모님의 소득 수준 등 다양한 가정생활 요소가 대학생활에 영향을 미칠 것이다. 첫 직장의 경력 또한 다음 구직활동에 영향을 미칠 수 있으며, 이러한 경험은 장기적인 관점에서 안정적 고용을 이어가는 경력이 될 것이다.

본 연구에서는 정규직, 비정규직과 같은 고용형태를 구분짓는 주요 요인을 파악하기 위해 학교생활, 가정환경, 첫 직장생활의 특성을 분석하고자 하며, 그림 1에 도형화 하였다. 상용직, 임시직, 일용직과 같은 고용형태에 대한 개념적 정의는 설문 응답자의 판단을 따르되, 응답에 대한 신뢰성을 높이기 위해 지난 일주일간 근로한 경험 유무가 확인된다. 상용직은 임금(또는 현물)을 받기로 한 고용계약기간이 1년 이상인 사람으로 분류한다. 임시직은 계약기간이 1개월~1년 미만인 근로자이며, 1개월 미만은 일용직 근로자로 구분한다.

학교생활, 개인 및 가족환경, 직장생활에 대한 질문은 설문지에 각 요인에 대해 명시함으로 응답자에게 인지되었다. 질문항목이 각 요인을 대표하는 지의 여부에 대한 논란을 최소화하고, 고용형태에 영향을 미치는 주요항목을 도출하기 위해 설문지에 사용된 모든 항목을 알고리즘의 입력값으로 동일하게 사용한다.

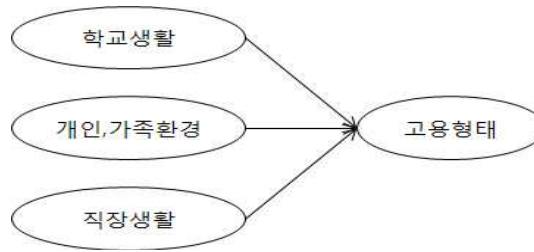


그림 1. 학교생활, 개인 및 가족 환경, 직장생활과 고용형태 간에 패턴 분석을 통한 근로자, 비정규 근로자, 무직자를 구분 짓는 주요 요인 분석  
 Fig. 1. Analyzing the factors that classify full-time, part-time, unemployment out of school life, personal and family environment, company life, and employment status

고용형태에 영향을 미치는 요인은 표1에 나타난 것과 같이 학교생활 관련 질문(47개), 개인환경 관련 질문(50개), 직장생활 관련 질문(73개)으로 구성되며 전체 170개의 질문 항목으로 구성한다. 학교생활에 관한 질문은 일반적 학교생활, 졸업 전 취업목표에 관한 질문, 졸업 전후 취업 준비에 관한 질문, 재학 중 일자리 경험 여부, 어학연수 경험 여부에 관한 질문으로 세분화 된다. 재학 중의 사회생활 경험은 취업에 대한 목표를 뚜렷이 하여 취업에 도움이 될 지, 아니면 학과 수업에 대한 집중력을 떨어뜨려 취업에 부정적인 영향을 줄 지 판단이 어렵다. 최근 대학생에게 인기가 있는 어학연수 경험이 취업에 미치는 영향도 분석 요인으로 추가한다. 학교생활에 관한 질문은 전공, 편입여부, 입사면접시험 횟수 등을 포함한다.

개인 인적사항은 향후 진로에 관한 계획에 관한 질문(3개), 취득 자격증에 관한 질문(2개), 현재 구직활동에 관한 질문(31개), 졸업 후 일자리 경험 여부(1개), 직업훈련 경험 여부(1개), 개인 인적 사항에 관한 질문(12개)으로 세분화되며 전체 50개이다. 앞으로 진로에 대한 뚜렷한 목표가 향후 취업에 긍정적인 영향을 줄 것으로 예측한다. 전문 자격증을 보유하였을 경우 또는 관련 추가 직업훈련을 받은 경력은 취업에 긍정적 영향으로 추정한다. 개인 인적 사항은 응답자의 나이, 결혼여부 등 기타 고용형태를 구분 짓는 속성을 알아본다.

직장에 관한 질문은 현 직장(일자리)에 관한 질문(36개), 첫 직장(일자리)에 관한 질문(37개)로 전체 73개의 질문을 포함한다. 첫 직장에 관한 질문은 입사 연령, 고용형태, 직업, 만족도, 임금, 근로환경, 자기개발 등 다양한 질문을 포함한다. 첫 직장에 대한 만족도가 높고, 자기개발 기회가 높았다면 정규직 직업에 대한 안정성이 높을 것으로 추측한다. 그리

고 직장생활에 임하는 개인적 태도 역시 직업 안정성에 영향을 미칠 것으로 예상된다.

고용형태는 종속변수이며 상용직, 임시직, 일용직, 종업원을 둔 고용주, 종업원이 없는 자영업자, 무급가족종사자, 무직자로 세분화 하였다. 고용형태를 세분화함에 따라 다양한 특성을 발견할 수 있다. 상용직은 정규직 근로자로 직업안정성이 높고, 임시직과 일용직은 회사에서 전문적 경력을 쌓기 어려워 경쟁력의 약화로 이어지는 경향이 있다. 고용형태는 설문조사로 엄밀히 분류하기 어려운 경향이 있다. 임시직의 경우 고임금 전문계약직과 최저임금의 아르바이트 형태를 포괄하기 때문이다. 특급 기술자, 회사 CEO와 같은 고임금 전문 계약직은 취업의 경쟁력을 보유하고 있으므로 수입의 안정성을 보유하고 있다고 볼 수 있다. 하지만 아르바이트 형태는 전문 지식이 낮고 장기적으로는 생계의 불안정을 초래할 수 있다. 본 연구는 이러한 문제점이 있음에도 불구하고 임시직으로 응답한 설문 응답자의 대다수가 사회적으로 이슈가 되고 있는 비정규직 문제의 안고 있다는 전제로 분석을 수행한다. 비정규직 응답자 중 일부 전문직 종사자도 있겠지만 이들의 숫자가 적고, 대다수는 한국의 실정상 고용 불안정성을 가지고 있을 경우 분석 결과의 오류는 오차 범위 내에 머무를 것으로 예상된다. 상세 질문 내역은 별첨의 설문지를 참고한다. 별첨 설문지와 GOMS 설문 자료에서 본 조사에 활용된 질문에 대한 상세 설명을 살펴볼 수 있다.

표 1. 답변 속성  
Table 1. Question list

속성 대분류	속성 중분류	질문 개수
학교생활 (47개)	학교생활에 대한 질문	20
	졸업 전 취업목표에 관한 질문	7
	졸업 전후 취업 준비에 관한 질문	18
	재학 중 일자리 경험 여부	1
	어학연수 경험 여부	1
개인환경 (50개)	앞으로 진로에 관한 질문(1년 이내 계획, 등)	3
	취득 자격증에 관한 질문	2
	현재 구직활동에 관한 질문	31
	졸업 후 일자리 경험 여부	1
	직업훈련 경험 여부	1
직장생활 (73개)	개인 인적 사항	12
	현 직장(일자리)에 관한 질문	36
구분자	첫 직장(일자리)에 관한 질문	37
	고용형태	1

#### IV. 설문 및 조사

대학생활과 직장에서의 고용형태의 관계를 살펴보기 위해서는 졸업 후 2년 이상이 경과한 후의 동태적 정보가 필요하다. 한국고용정보원에서 수행하는 대학교 졸업자 이동 경로조사(GOMS)는 2년제 이상 대학 이상 대학졸업생을 대상으로 학교교육, 직장경험, 직업훈련 등에 대해 조사하는 대졸자 단기 패널 조사이다. 전국 2년제 이상 대학을 졸업한 졸업자 중 약 4% (18,000여명)를 조사하며, 조사 방법은 면접원이 조사 대상자를 직접 방문하여 질문하고 그 응답을 기록하는 면접타계식이다. 조사는 2007년에 수행되었으며, 조사 내용은 교육과정, 현재 일자리, 졸업 후 일자리, 첫 일자리, 졸업 전후 일자리 탐색과정, 재학 중 취업, 훈련과정 및 자격증, 개인사항 등 노동시장 진입 및 정책에 영향을 줄 수 있는 요인으로 구성된다.

데이터 중 부정확한 응답은 분석 결과의 신뢰성 제고를 위해 제외된다. 나이, 졸업 년 월 등의 데이터가 부정확한 경우 제외한다. 나이가 만 22세 미만은 사회적 경험이 아직 적은 관계로 직업경로 분석에 미치는 영향이 적어 제외한다, 직업 시작 나이가 만18세 미만인 경우 통상적 성인 근로기준에 적합하지 않고 정규 근로자와 비정규 근로자의 구분이 불명확하여 제외한다.

의사결정 트리 결과 도출을 위해 J48 알고리즘[13]을 이용하며, 분석 결과의 정확성을 평가한다. 의사결정트리 기법으로 가장 널리 활용 되는 것은 ID3 알고리즘이지만 명목척도만 수용할 수 있다는 단점이 있지만, J48은 명목척도, 서열척도, 등간척도, 비율척도를 모두 포괄할 수 있는 장점을 가지고 있다. 데이터마이닝 프로그램은 널리 사용되고 있는 Weka[14] Version 3.4.5를 이용한다. 결과 평가는 데이터를 10으로 나누어 9를 학습하여 1로 테스트 하는 10분할 평가기법을 적용한다. 모델의 정확성은 정확하게 분류된 응답자수의 비율이 높고 부정확하게 분류된 응답자수의 비율이 낮을수록 정확한 모델이다.

모델이 랜덤(Random)한 경우에 비하여 정확성을 파악하기 위해 분석결과를 랜덤한 경우와 비교한 값을 제시하는 Kappa statistic[14] 분석기법을 활용한다. Kappa statistic은 결과 값이 0.4 초과일 경우 적절한 경우에 포함된다. Kappa statistic은 예측한 결과 값을 랜덤한 경우에서 뺀 후 랜덤한 경우가 아닐 확률로 나눈다.

## V. 결과 분석

### 5.1 학교유명 및 전공별 고용형태 분석

대학교 계열별, 유형별 특성이 고용형태에 영향을 미치는지 조사하기 위해, 대학교 계열별, 유형별로 분류하여 조사하였다. 대학교 특성은 계열별, 유형별로 구분하였다. 계열별로 상용직 근로자 수를 비교하였을 때 공학계열(79%)이 가장 높게 나타났으며 의학계열(78%)이 뒤를 이었다. 상용직 비율이 낮은 계열은 인문(66%), 예체능(59%)로 나타났다. 인문계열은 무직자 비율이 16%로 가장 높게 나타나 계열별 차이가 있었다. 공학계열에 대한 전공 인기도가 낮아지는 것으로 언론에서 언급되고 있으나, 상용직 취업률이 타전공계열과 비교하여 가장 높게 나타난 것은 고무할 내용이다.

대학교 유형은 교육대, 4년제, 전문대로 구분하였다. 교육대가 상용직 근로자 비율이 92%로 가장 높고, 무직자는 2%로 가장 낮았다. 교육대 졸업자의 경우 자영업이나 일용직 선택 비율은 상대적으로 낮았다. 전문대 졸업자가 4년제 졸업자에 비해 상용직 비율은 유사한 반면, 무직자 비율은 낮고, 자영업(중업원) 비율은 높아 보아 현재 취업 현황이 양호하게 나타났다. 일반적으로 전문대졸자가 4년제에 비해 낮아 상용직 비율 및 고용유지율이 낮을 것으로 예상되나, 본 조사에는 전문대졸자의 상용직 비율이 동일하게 나타났다. 따라서 향후 전문대졸자의 고용유지율에 대한 추적조사 비교가 필요하다.

### 5.2 고용형태에 미치는 주요 요인 분석

현재 근로 상태를 결정 짓는 학교생활의 주요 속성 평가는 향후 근로자를 늘이고 무직자를 줄일 수 있는 원인 파악에 도움이 되어 매우 흥미 있는 분석이다. 이러한 분석을 위해 의사결정트리 분석이 사용되었다. 학교생활 속성은 전공, 전공선택 이유, 전공에 대한 전망, 대학학비 마련 방법, 휴학횟수, 졸업학점, 고등학교 계열 등을 포함한다.

현재 고용형태를 결정짓는 가장 주요 속성으로는 현재 부모님의 수익이었다. 현재 부모님의 수익이 100만 원 초과인 경우 상용직일 확률이 가장 높았으며, 100만 원 이하일 경우 무직일 확률이 높았다. 부모님의 수익이 100만 원 이상인 경우 나이가 32.9세(32세11개월)이하인지 초과인지가 고용형태에 영향을 미치는 주요한 속성으로 나타났다. 이러한 결과는 나이가 많을수록 상용직을 구하는 기회가 줄어드는 사회적 현상 때문으로 추정된다.

부모님의 수익이 100만원 미만인 경우 결혼상태로 나타났다. 결혼한 경우가 미혼이나 이혼 상태보다 안정적 직업을 가지려는 의욕이 강하기 때문으로 해석된다. 미혼인 경우 주간대학을 나온 경우 무직자 응답수가 우세하였고, 야간대학을 졸업한 경우 상용직 응답수가 우세하였다. 학교유형별 조사결과 전문대학교 졸업자가 4년제 대학교 졸업자와 차이가 없었고, 야간대학 학력 소유자가 주간대학 학력 소유자 보다 근로 상태에서는 오히려 높은 결과를 나타냈다. 결혼하여 배우자가 있는 경우, 대학교시절 학교 취업지원 활동에 만족도 정도가 중요한 영향 속성으로 나타났다. 학교취업지원활동에 긍정적으로 답한 응답자는 상용직이 우세하여, 학교 취업 지원활동이 대학생의 취업에 영향을 미치는 것으로 나타났다. 의사결정 트리에서 나타난 주요 속성으로는 대학교와 고등학교 지역이었지만, 본 연구에서는 지역 요인을 제외한다. 국내 현실상 우수 대학이 서울 지역에 집중 분포되어 있는 것을 고려하여 대학교 수준이 졸업생의 취업에 영향을 미치는 요인으로 추정할 수 있다.

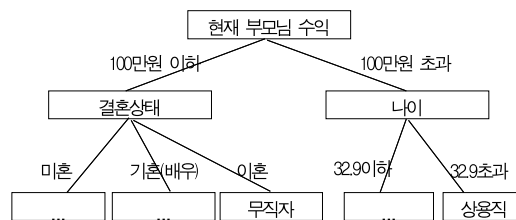


그림 2. 의사결정트리  
Fig. 2 Decision tree

학교생활뿐만 아니라 일반 직장생활의 만족도 및 생활환경의 모든 속성을 데이터마이닝 알고리즘에 적용하였을 경우, 현 직장을 들어가게 된 주요 방법(예, 고용안정센터 등 공공 직업알선기관 등록, 민간 직업알선기관 등록, 교수님의 추천으로 등) 등이 높게 나타났다. 현 직장에 들어가게 된 주요 방법은 답변항목으로 “교수님의 추천으로”, “개인 사업 시작(가족사업 참여)”, “취업박람회를 통해”, “민간 직업알선기관 등록” 등 매우 세분화 되어있어 어느 정도 고용형태를 예측할 수 있기 때문에 선정된 것으로 추정된다. 흥미 있는 속성으로 공무원 시험 준비 경험이 있는 경우 임시직 비율이 높았다. 현재 결혼 상태와 결혼 기간은 경제력과 결혼 생활은 밀접한 관련이 있기 때문에 상위에 나타난 것으로 추정된다. 위의 모델의 예출률은 87%로 나타났으며, 랜덤한 경우와 비교하는 Kappa statistic 분석 결과는 0.68로 모델이 충분한 설명력이 있음을 나타냈다.

## VI. 결 론

노동시장에서 고용에 대한 이해를 높이고 취업을 제고를 위한 학교교육과 고용서비스 개선을 위한 연구의 참고자료를 만들기 위해 고용형태에 미치는 영향요인을 의사결정트리 기법을 통해 분석하고자 하였다. 데이터마이닝 기법에 적용하기 위한 전처리 기법으로 정보항목을 그룹속성, 연계속성, 도메인으로 분류하여 각각에 대한 데이터 조작 방법을 정리하였으며 이는 사회과학 분야의 데이터 정제에 용이하게 활용될 수 있을 것이다.

대학교 계열을 이용하여 통계 현황을 분석한 결과, 고용형태(상용직, 무직자, 임시직 등)는 대학교 계열별, 대학교 유형별로 평균적 차이를 보였다. 대학교 계열별 분석 결과 인문과 예체능 계열은 상용직 근무자 비율이 각각 66%, 59%로 가장 낮고, 인문계열 졸업자의 경우 무직자 비율이 16%로 가장 높아 계열별 고용형태 차이가 크게 나타났다. 이러한 결과는 취업재교육(예, 국문학 졸업자에게 통계 교육을 지원하여 사회조사 분석 전문가로 육성) 등을 통한 향후 인문 계열 졸업자에 대한 취업 지원의 필요성을 나타내었다.

학교생활과 개인 및 가족환경 중 현재 고용형태를 결정짓는 가장 주요 속성으로는 현재 부모님의 수익이었다. 현재 부모님의 수익이 100만 원 초과인 경우 상용직일 확률이 가장 높았으며, 미만인 경우 무직일 확률이 높았다. 부모님 수익이 100만원 초과일 경우는 다음 주요 항목으로는 연령으로서 나이가 32살 11개월 이상인 경우 상용직 비율이 높게 나타났다. 부모님의 수익이 100만원 미만인 경우 다음으로 가장 주요 항목으로는 결혼상태로 나타났다. 이러한 결과는 결혼한 경우가 미혼이나 이혼 상태보다 직업을 가지려는 의욕이 강하기 때문으로 해석된다. 의사결정 트리를 통해 현재 고용형태를 정확하게 추정할 확률은 87%이며, 랜덤한 경우와 비교하는 Kappa statistic 분석 결과는 0.68로 모델이 충분한 설명력이 있음을 나타냈다.

정책적 또는 교육 개선 측면에서의 제언으로는 부모의 수익과 같은 가정환경 요인이 고용 형태에 미치는 영향이 크다는 점이다. 그리고 수익이 낮은 경우 결혼 생활을 유지하는 것이 고용형태와 밀접한 관련이 있음을 보여주었다. 따라서 고용의 질을 높이기 위해서는 국가에서는 가구당 기초소득이 보장되도록 지원하고, 결혼생활 유지로 가정을 지킬 수 있는 지원이 강화되어야 할 것으로 예상된다.

본 연구의 한계로는 데이터의 임시직에 대한 세부적인 분류가 부족하여 임시직을 고용이 불안한 취약계층으로 통칭하

였다. 이러한 임시직에 대한 불명확한 한계를 해결하기 위해 향후 임금수준에 따라 고임금 임시직과 저임금 임시직으로 구분하여 패턴을 분석할 필요가 있다. 정규직이 고용안정성이 높다고 추정할 점도 유의하여 결과를 해석하여야 한다.

## 참고문헌

- [1] S. Kim, and Y. Seo, "Organizational and Occupational Mobilities in Firm and Occupational Internal Labor Markets: The Case of Medical Profession," Korean Sociological Association, Vol 37 No 4, pp. 1-29, 2003.
- [2] H. Kim, and I. Jeong, "Bi-directional Fuzzy Matching Algorithm," Korean Society for Internet Information, Vol. 12 No. 3, 2011.
- [3] J. Ban, K. H. Kim, K. H. Kim, "A Study on Factors Affecting Youth Employee's Labor Mobility and Employment Status Transition," Korean Academy of Social Welfare, Vol 57 No. 3, pp. 73-103, 2005.
- [4] H. Song, "A Comparative Analysis of Labor market Structure: Korea and Taiwan," Korean Sociological Association, Vol. 34, pp. 951-979, 2000.
- [5] J. Yoon, "Structural Changes in Korean Labor Market and Active Labor Market Policy," Master thesis, Chungnam National Univ., pp. 40-56, 2004.
- [6] H. Jang, "A Study on the Programs for vitalization of the "Employment Center" in Ministry of Labor : With the Focus on the 4 "employment center" in Incheon," Master thesis, Incheon Univ., pp. 62-72, 2004.
- [7] K. Jeong, and H. Kim, "Quality Status Comparison and Analysis for the Service Development Direction of Domestic Job Information Site," Journal of the Korean Society of Computer and Information, Vol. 13 No. 5, Korean Society of Computer and

Information, 2008.

[8] S. Ha, and J. Lee, "Customer Relationship Management Techniques Based on Dynamic Customer Analysis Utilizing Data Mining)," Korea Intelligent Information System Society, Vol.9 No. 3, 2003.

[9] Agrawal, R. & Srikant, R, "Fast algorithms for mining association rules," The International Conference on Very Large Databases, pp. 487-499, 1994.

[10] Bridges, W. P. & W. J. VILLEMEZ, "Employment Relations and the Labor Market: Integrating Institutional and Market Perspectives," American Sociological Review, 56, pp. 748-764, 1991.

[11] J. Han, and M. Kamber, "Data Mining," Elsevier, CA. USA. pp. 279, 2001.

[12] H. Lee, "Analysis of Business Attributes in Information Technology Environments," Journal of Information Processing System, Vol 7, No. 2, pp. 385-396, 2011.

[13] E. Rosch, "Cognitive Representations of Semantic Categories," Journal of Experimental Psychology, Vol 8, pp. 431-455, 1995.

[14] I. H. Witten, and E. Frank, "Data mining: practical machine learning tools and techniques," Diane Cerra, CA. USA, 2005.

저 자 소 개



김형래(Hyoung-Rae Kim)

1997년 관동대학교 전자계산 학사

2001년 Florida Institute of Technology, 전산석사

2005년 Florida Institute of Technology, 전산박사

8/2006 - present: Research Fellow  
Korea Employment Information Service

관심분야 : Employment Information Systems, Data Mining

E-mail: goddoes8@gmail.com



전도홍(Do-hong Jeon)

1985 미국 Oklahoma City University 컴퓨터과학 학사

1987 미국 Florida Institute of Technology 컴퓨터과학 석사

1990년 미국 Florida Institute of Technology 컴퓨터교육학 박사

1990년 - present : 관동대학교 컴퓨터학과 교수

관심분야 : Computer graphics, Data mining

E-mail : dhjeon@kdac.kr