

데이터 시각화를 이용한 취업자 특성분석[†]

조장식¹

¹경성대학교 정보통계학과

접수 2014년 5월 8일, 수정 2014년 6월 17일, 게재확정 2014년 6월 23일

요약

대졸 취업자들의 특성을 분석하기 위해 주로 모수적인 접근방법을 사용해 온 기존의 연구와는 달리, 본 연구에서는 R 프로그램을 이용하여 데이터 시각화에 초점을 맞추어 분석하였다. 이를 위해 취업여부에 미치는 개인특성 변수들의 유사성 분석을 위해 다중대응분석을 실시하였다. 또한 취업여부에 영향을 미치는 개인특성 변수들의 고차 상호작용효과를 분석하기 위해 의사결정나무분석을 실시하였다. 그리고 연관성분석을 이용한 연관성 규칙을 계산하여 개인특성 변수들이 취업여부에 미치는 효과를 분석하고, 연관성규칙의 결과를 사회연결망분석의 연결망 구조로 시각화 하였다. 분석결과 다음과 같은 주요 결과를 얻었다. 첫째, 취업여부에 영향을 미치는 변수들로는 평균평점, 계열, 강의평가점수, 성별 등으로 나타났다. 둘째, 평균평점과 강의평가점수가 보통이상으로 높고 자연계열인 경우 취업 가능성이 높음을 알 수 있다. 또한 수시모집으로 입학한 연령이 낮은 졸업생이 취업가능성이 높게 나타났다. 셋째, 평균평점이 낮고 예체능 계열이며 연령이 높은 대졸자들이 취업가능성이 낮음을 알 수 있다. 또한 예체능 계열의 단일전공을 한 여학생들의 경우도 취업가능성이 높지 않음을 알 수 있다.

주요용어: 다중대응분석, 데이터 시각화, 사회연결망분석, 연관성분석, 의사결정나무분석.

1. 서론

요즘 대졸자들의 취업에 대한 사회적 관심도가 높아지고 있다. 대졸자들의 취업에 대한 어려움은 하향 취업으로 이어지는 경향이 있음을 많은 연구자들은 지적을 하고 있다 (Kim 등, 2012; Lee, 2008). 이런 현실 속에서 대졸자 취업에 관한 연구도 활발하게 진행되었다. 많은 선행연구들은 성, 연령, 학력, 인턴경험, 직업훈련 경험 등 다양한 변인의 영향력을 실증적으로 분석하였다. Park (2011)은 위계적 선형모형을 활용하여 대졸자의 취업에 미치는 개인배경, 개인의 취업준비, 대학특성, 대학의 취업지원변인의 영향력을 분석한 바 있다. 그의 결과에 따르면 성별, 가계소득, 출신고 등의 개인배경과 서열유형, 취업률 등의 대학특성 변인이 영향을 미치는 것으로 나타났다. Jun 등 (2009)은 부산지역 대학의 누리사업단 졸업자를 대상으로 개인의 취업실태에 대한 결정요인을 생존분석과 의사결정나무 분석을 활용하여 연구하였는데, 본인학력, 가구소득, 평균평점, 자격증 등이 미취업 기간 결정에 가장 유의한 변수로 나타났다. Chai (2007)는 로지스틱 회귀분석과 다중회귀분석을 이용하여 신규대졸자의 취업 및 임금수준 결정요인을 분석하였다. 그의 결과를 보면, 학교소재지에서 경기·인천지역, 학교성적이 좋을수록, 취득자격증의 전공 및 취업 관련정도가 높을수록 취업가능성이 높게 나타났다. Lee와 Kim (2003)은 대졸 청년층의 첫 일자리에 초점을 맞추어 노동시장 성과를 결정하는 요인을 학벌이라는 범주와 자격증 취득,

[†] 이 논문은 2014학년도 경성대학교 학술연구비지원에 의해 연구되었음.

¹ (608-736) 부산광역시 남구 수영로 309 번지, 경성대학교 정보통계학과, 교수. E-mail: jscho@ks.ac.kr

재학 중 교육훈련, 일 경험 등 취업을 위한 다양한 스펙들을 구분하여 분석한 바 있다. Cho (2011)은 대졸 취업결정에 미치는 영향력을 분석한 바가 있는데 그의 결과에 따르면, 입학관련 변수들 중에서는 외국어영역 점수가 높을수록, 예체능계열에 비해서 자연계열, 졸업 당시 연령이 작을수록, 졸업시기는 8월 졸업생들에 비해서 2월 졸업생들이 취업확률이 높은 것으로 나타났다. 그 외에도 Hwang과 Baek (2008)은 대졸자의 노동시장 성과 결정요인분석을 연구한 바 있다.

한편 통계그래프 (statistical graphs) 및 데이터 시각화 (data visualization) 기법은 데이터에 대한 직관적이고 쉬운 이해를 제공하며, 데이터에 대한 초기 탐색단계에서부터 고급분석에 이르기까지 매우 중요한 역할을 수행한다. 최근에는 많은 연구자들이 R 프로그램을 이용하여 데이터 시각화를 시도하고 있다.

본 연구에서는 데이터 시각화를 활용하여 취업여부에 관련성이 높은 독립변수들의 범주 수준을 분석하기 위해 다중대응분석 (multiple correspondence analysis)과 의사결정나무분석 (decision tree analysis)을 활용하였다. 또한 데이터에 내재되어 있는 항목들 간의 상호관련성을 찾는 기법인 연관성분석 (association analysis)을 이용하여 취업자와 미취업자의 특성에 미치는 독립변수의 영향력을 연관성규칙으로 표현하였으며, 그 결과를 사회연결망분석 (social network analysis)의 연결망 구조로 시각화를 하였다. 연관성 분석에 대한 연구로는 Liu 등 (1999)과 Cho와 Park (2011a, 2011b) 등이 있다.

본 논문에서는 K-대학 졸업생들의 취업여부에 미치는 개인특성 변수들에 대한 영향력을 데이터 시각화에 초점을 맞추어서 분석하고자 한다. 먼저 제 2절에서는 분석 자료에 대한 설명과 기술통계 결과를 소개하며, 제 3절에서는 다중대응분석과 의사결정나무분석을 이용하여 취업자의 특성과 관련성이 높은 독립변수들의 범주 수준들을 탐색하여 데이터 시각화를 하였다. 그리고 제 4절에서는 취업자 및 미취업자의 특성을 연관성분석의 결과인 연관성 규칙을 이용하고, 그 결과를 사회연결망으로 시각화 하여 분석하였다.

2. 분석자료 및 기초분석

본 연구에서 사용된 데이터는 K-대학교의 2009년 8월과 2010년 2월 졸업생들 중에서 2000년 이전에 입학한 졸업생들과 타 대학으로부터 편입학한 졸업생, 그리고 졸업 후 타 대학 (또는 대학원)으로 진학한 졸업생을 제외한 2,204명만을 분석대상으로 하였다. 또한 취업자의 특성을 분석하기 위해 취업여부 (employment), 입학시기 (recruiting time), 계열 (department category), 평균평점 (gpa), 강의평가점수 (course score), 복수전공 여부 (major), 연령 (age level), 성별 (sex) 등의 변수들로 구성되어 있다. 여기서 평균평점은 졸업생들이 4년간 취득한 모든 교과목들에 대한 총 평균평점을 의미하며, 강의평가점수는 매학기 수강하는 교과목마다 실시되는 강의평가 점수에 대한 총 평균점수를 의미한다.

이들 변수들에 대한 간단한 기술통계를 보면 다음과 같다.

모집시기 (recruiting time)는 정시모집으로 입학한 학생들 (1,409명)이 가장 많고, 그 다음으로 수시모집 (740명), 기타 (55명) 등의 순으로 나타났다. 계열 (department category) 별로는 인문계열 (1,178명)이 가장 많고, 그 다음으로 자연계열 (628명), 예체능계열 (398명) 등의 순으로 나타났다. 또한 복수전공여부 (major)는 단일전공 (1,852명)이 복수전공 (352명) 보다 많았으며, 성별 (sex)은 여학생 (1,315명)이 남학생 (889명) 보다 많았다. 또한 강의평가점수의 평균은 3.99로 나타났으며, 평균평점은 3.49, 연령의 평균은 24.6세로 나타났다.

한편 취업여부와 입학시기, 계열, 복수전공 여부, 성별 등이 관련성이 있는지를 알아보기 위해 교차분석한 결과 다음 Table 2.1과 같다.

Table 2.1에 따르면 입학시기와 계열은 각각 취업여부와 통계적으로 유의한 관련성이 있음을 알 수 있다.

Table 2.1 Crosstable analysis for student characteristics (row %, column %)

variables	category	employment		total	Chi-Square
		no	yes		
recruiting time	early	45.9 (31.1)	54.1 (36.1)	740	8.622 ***
	regular	51.2 (65.8)	48.8 (62.0)	1,409	
	ect	61.8 (3.1)	38.2 (1.9)	55	
department category	humanity	49.9 (53.7)	50.1 (53.2)	1,178	22.934 ***
	natural	43.5 (24.9)	56.5 (32.0)	628	
	art	58.8 (21.4)	41.2 (14.8)	398	
major	single	50.1 (84.7)	49.9 (83.4)	1,852	0.426
	double	47.7 (15.3)	52.3 (16.6)	352	
sex	female	48.6 (58.4)	51.4 (61.0)	1,315	0.214
	male	51.3 (41.6)	48.7 (39.0)	889	

* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$

다음으로 취업여부에 따라 강의평가점수, 평균평점, 연령이 차이가 나는지를 t -검정한 결과 아래 Table 2.2와 같다.

Table 2.2 Results of t -test

variables	employment						t -value
	no			yes			
	N	mean	s.d.	N	mean	s.d.	
course score ¹	1,090	3.98	0.50	1,108	3.99	0.50	0.864
gpa ²	1,095	3.42	0.50	1,109	3.55	0.44	6.450 ***
age	1,095	24.75	2.48	1,109	24.47	2.33	-2.697 ***

* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$

course score¹ means course evaluation score, gpa² means grade point average

Table 2.2의 결과를 살펴보면 강의평가점수를 제외한 평균평점과 연령은 취업여부에 따라 통계적으로 유의한 차이가 남을 알 수 있다.

다음으로 학과별 특성을 분석하고자 한다. 이를 위해 먼저 개인특성변수들을 학과별로 통합하여 학과별로 남학생 비율, 취업비율, 복수전공 비율, 강의평가 평균점수, 평균평점, 평균연령 등을 계산하였다.

먼저 각 학과별 특성을 알아보기 위해 R의 “aplpack” 패키지에서 star() 함수를 사용하여 별 그림(star plot)을 그린 결과 아래 Figure 2.1과 같다. 여기서 r_male은 남학생의 비율이고 r_emp은 취업 비율이며, r_double은 복수전공 비율이다. 그리고 m_score는 강의평가 평균점수를 의미하고 m_gpa는 평균평점이고 m_age는 평균연령을 의미한다. 또한 각 학과별 원 그림에서 학과특성을 나타내는 개별 변수들은 부채꼴의 색깔들로 각각 구분이 되어 있고, 변수 값들의 크기는 부채꼴의 크기로 표현되어 있다.

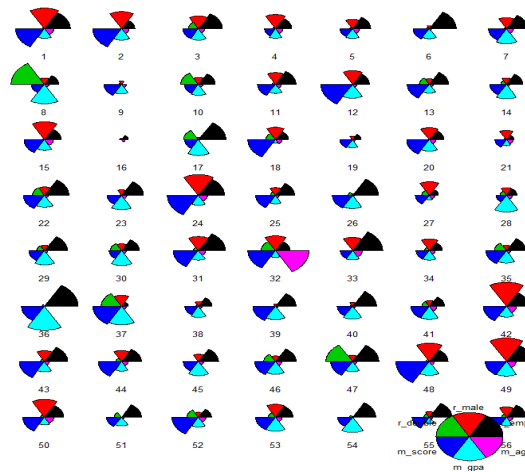


Figure 2.1 Star plot for department characteristics

Figure 2.1의 결과를 보면 타 학과들에 비해서 상대적으로 취업률이 높은 학과들은 대체로 6, 17, 26, 33, 36, 47, 49 등의 학과들이며, 8번과 47번 학과의 경우는 복수전공 비율이 높고, 1, 24, 42, 49번 학과들은 남자비율이 상대적으로 높게 나타났다. 또한 1, 2, 12, 24, 48번 학과들은 강의평가 점수가 높고, 8, 36, 37번 학과들은 평균평점이 높고, 32번 학과는 연령이 상대적으로 높게 나타났다.

한편 학과 특성변수들의 상관계수 행렬을 구하기 위해 R의 `pair()` 함수를 사용한 결과 아래 Figure 2.2와 같다.

Figure 2.2의 결과에 따르면, 남학생 비율과 연령이 0.54의 상관계수로 가장 높게 나타났으며, 그 다음으로 남학생비율과 평균 강의평가점수가 0.48, 평균 강의평가 점수와 평균 연령이 0.28 등의 순으로 나타났다.

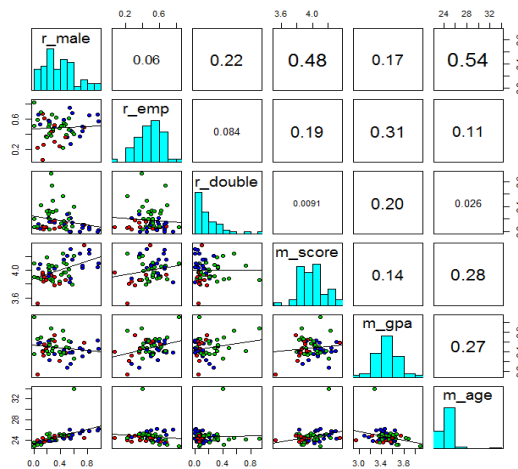


Figure 2.2 Correlation matrix

3. 다중대응분석과 의사결정나무분석

이 절에서는 취업여부와 상호 관련성이 높은 개인특성 변수들을 알아보고, 어떤 특성을 가진 졸업생들이 취업여부에 영향을 미치는지를 알아보기 위해 다중대응분석 (multiple correspondence analysis)과 의사결정나무 분석 (decision tree analysis)을 하고자 한다. 여기서 분석결과에 대한 해석의 용이성을 위해 개인 특성변수들 중에서 평균평점, 강의평가점수, 연령 등 연속형 자료는 동일한 비율을 갖도록 3개의 수준 (low, middle, high)을 갖는 범주형 변수 C_gpa (평균평점), C_score (강의평가점수), C_age (연령)로 만들었다. 다중대응분석에 대한 연구결과로는 Yhang 등 (2007)과 Lee (2008) 등이 있다.

먼저 특성변수들과 취업여부의 상호관련성을 알아보기 위해 각 특성변수들의 범주 수준별로 다중대응분석 (multiple correspondence analysis)을 실시하였다. 다중대응분석은 개체 (케이스)와 범주에 계량적 수치를 부여함으로써 범주형 데이터를 수량화하는 분석기법으로서, 내적 일관성의 원리로부터 범주의 수량화를 실시하는 분석기법이다. 즉 각 변수들의 범주수준별로 유사성의 정도에 따라 좌표평면상에 산점도로 표시되며, 가까운 거리에 위치할수록 서로 유사성의 정도가 높으며, 먼 거리에 위치할수록 서로 유사성의 정도가 낮게 나타난다. R의 “ca” 패키지에서 `mjca()` 함수를 사용하여 다중대응분석을 실시한 결과 아래 Figure 3.1과 같다.

Figure 3.1의 결과에서 알 수 있듯이 취업자 (employ_yes)의 특성과 관련성이 높은 졸업생들의 특

성을 살펴보면, 인문계열 (category_humanity)에서 평균평점이 높고 (C_gpa_high), 복수전공 (major_double)을 하고, 강의평가점수가 높은 (C_score_high) 특성을 갖고 있음을 알 수 있다.

반면, 미취업자 (employ_no)의 특성과 관련성이 높은 졸업생들의 특성을 살펴보면, 단일전공 (major_single), 낮은 강의평가점수 (C_score_low), 예체능계열 (category_art), 그리고 낮은 평균평점 (C_gpa_low) 등의 특성을 갖고 있음을 알 수 있다. 또한 졸업생들에 대한 특성별 분류를 해 보면 전체적으로 4개의 군집으로 분류됨을 알 수 있다.

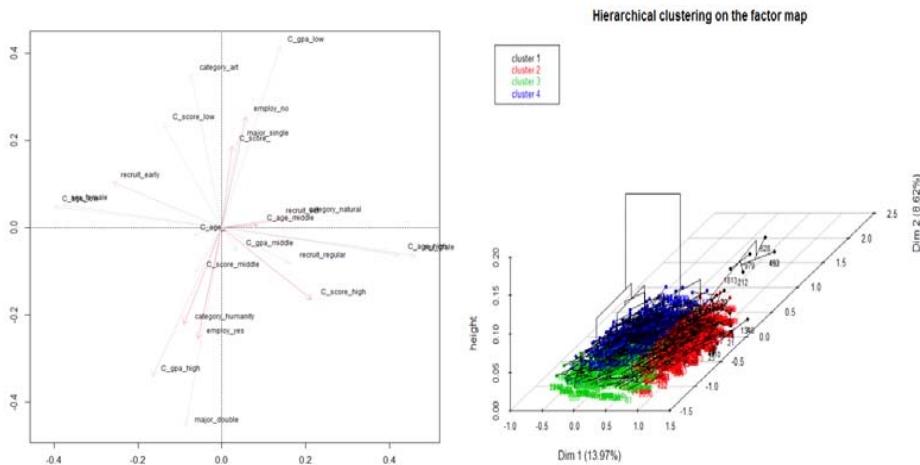


Figure 3.1 Multiple correspondence analysis (left) and cluster analysis (right)

다음으로 대졸자의 특성변수들이 취업여부에 미치는 효과를 분석하기 위해 비모수적인 방법인 의사결정나무 분석을 이용하며, 이를 통해서 취업여부에 영향을 미치는 각 수준별 특성변수들에 대한 고차의 상호작용효과를 분석한다. 의사결정나무분석을 수행하기 위해 지니지수 (Gini index)를 분리기준으로 사용하였으며, 이지분리를 수행하는 CART (classification and regression trees) 알고리즘을 사용하였다. 의사결정나무 분석과 관련된 선행연구로는 Cho (2010, 2012), Cho와 Park (2012), Jung과 Min (2013) 등이 있다. 정지규칙으로는 최대나무깊이 (maximum tree depth)는 3으로 설정하였으며, 최소 케이스 수 (minimum number of cases)에서 부모마디 (parent node)는 300, 자식마디 (child node)는 100으로 설정하였으며, 적절한 가지치기 (pruning)를 병행하였다. R의 “party” 패키지에서 ctree() 함수를 사용하여 의사결정나무분석을 실시한 결과 아래 Figure 3.2와 같다.

Figure 3.2의 결과로부터 취업여부에 가장 많은 영향력이 있는 변수로는 평균평점 (C_gpa)으로 나타났으며, 그 다음으로는 계열 (category), 입학시기 (recruit), 강의평가점수 (C_score)의 순으로 나타났다. 취업률이 상대적으로 높은 대졸자의 특성을 살펴보면, 평균평점이 중간 이상인 경우 자연계열인 경우가 62.3%로 나타났다. 또한 평균평점이 중간이상인 경우 인문 또는 예체능계열이면서 입학시기가 수시모집이고 강의평가 점수가 낮은 대졸자들이 66.9%로 취업률이 높은 것으로 나타났다.

한편, 취업률이 상대적으로 낮은 대졸자의 특성을 살펴보면, 평균평점이 낮은 집단으로 40.4%의 취업률로 낮게 나타났다. 또한 평균평점이 중간이상인 경우는 인문계열 및 예체능계열로서 정시 또는 기타 전형으로 입학한 대졸자들이 48.6%의 취업률로 낮게 나타났다.

참고로 위의 의사결정나무분석 결과에 대한 오 분류비율은 42.8%로 다소 높게 나타났다.

나타나서 상호 연관성의 정도가 높음을 알 수 있다.

다음으로 Table 4.2는 미취업자 (employ=no)와 연관성이 높은 규칙 중, 지지도 0.05 이상, 신뢰도 0.60 이상 되면서 신뢰도 기준 상위 10개를 순서대로 정리한 결과이다.

Table 4.2 Association rules (employ=no)

conditions	support	confidence	lift
gpa=low & score=low	0.078	0.638	1.285
major=single & gpa=low & score=low	0.072	0.637	1.282
recruit=regular & sex=female & category=art	0.054	0.626	1.261
recruit=regular & major=single & category=art	0.074	0.623	1.254
recruit=regular & category=art	0.081	0.616	1.240
gpa=low & age=middle	0.073	0.612	1.232
major=single & gpa=low & age=middle	0.065	0.611	1.230
recruit=regular & gpa=low & age=middle	0.053	0.611	1.229
major=single & gpa=low & age=high	0.053	0.607	1.222
major=single & sex=female & category=art	0.069	0.604	1.216

미취업자 (employ=no)와 관련성의 정도가 높게 나타내는 Table 4.2의 결과를 살펴보면, 평균평점과 강의평가점수가 낮은 경우 지지도는 0.078, 신뢰도는 0.638, 향상도는 1.285로 가장 높은 향상도를 보였다. 즉 이들 간의 연관성의 정도가 미취업자와 높음을 알 수 있다. 또한 정시모집으로 입학한 예체능 계열, 여학생이면서 단일전공을 한 경우 등이 대체로 미취업자와 관련성이 높음을 알 수 있다.

Figure 4.1은 Table 4.1과 Table 4.2의 결과를 사회연결망 분석 (social network analysis)을 이용하여 시각화를 시킨 결과이다. 여기서 노드의 크기가 클수록 지지도가 높음을 의미하고, 노드의 색깔이 진할수록 향상도가 높음을 의미한다.

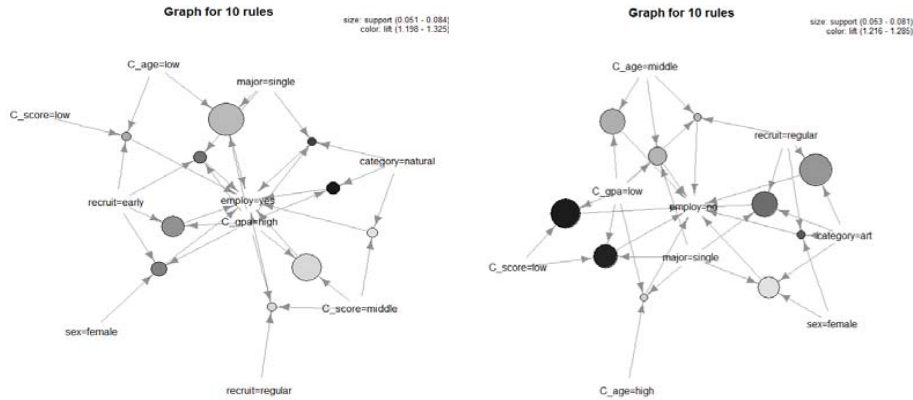


Figure 4.1 Social network analysis based on association rules (left; employ=yes, right; employ=no)

Figure 4.1의 결과에서 특정 노드에서 화살표가 들어오는 부분이 Table 4.1과 Table 4.2에서 대한 조건 (conditions)에 해당되며 화살표가 나가는 부분은 결과에 해당된다. 즉 Table 4.1의 경우 취업자 (employ=yes)가 되며, Table 4.2의 경우는 미취업자 (employ=no)가 된다. 예를 들어 Figure 4.1의 왼쪽 그림에서 연령이 낮고 (C.age=low)이고 단일전공 (major=single)인 경우, 취업자 (employ=yes)로 될 특성에 대한 지지도는 상대적으로 아주 높으며 (노드 크기 큼), 지지도는 상대적으로 낮음 (노드 색깔 옅음)을 알 수 있다.

5. 결론

본 논문에서는 대졸자의 취업특성에 미치는 영향력을 분석하기 위해 학과특성과 졸업생 특성 각각에 대해 다양한 통계적인 방법으로 데이터 시각화에 초점을 두고 분석하였다. 본 논문에서 사용된 통계적 기법으로는 학과특성을 분석하기 위해 상관계수행렬과 별 그림을 이용 하였으며, 다중대응분석, 의사결정나무분석 그리고 연관성분석을 통해 취업여부와 관련성이 높은 개인특성변수들을 탐색하였다. 마지막으로 연관성 분석결과를 시각적으로 표현하기 위해 연관성 규칙을 사회연결망 그림으로 나타냈다. 각각의 분석방법에서 나타난 취업자 및 미취업자들의 특성의 결과는 유사한 부분도 있지만 다소 차이가 나는 부분도 있었다. 여기서는 각 방법들에서 나타난 결과를 바탕으로 개괄적으로 정리하면 다음과 같다.

첫째, 취업여부에 영향을 미치는 변수들로는 평균평점, 계열, 강의평가점수, 성별 등으로 나타났다.

둘째, 평균평점과 강의평가점수가 보통이상으로 높고 자연계열인 경우 취업 가능성이 높음을 알 수 있다. 또한 수시모집으로 입학한 연령이 낮은 졸업생이 취업가능성이 높게 나타났다.

셋째, 평균평점이 낮고 예체능 계열이며 연령이 높은 대졸자들이 취업가능성이 낮음을 알 수 있다. 또 한 예체능 계열의 단일진공을 한 여학생들의 경우도 취업가능성이 높지 않음을 알 수 있다.

한편 본 논문에서 활용된 분석 방법들 간에 나타난 결과의 차이는 수집 가능한 제한된 자료만을 사용하였기 때문에 발생된 결과로 유추된다. 따라서 본 연구의 결과를 지나치게 일반화하기에는 분명한 한계가 있음을 밝혀 둔다. 또한 다양한 변수와 방법론을 활용한 심도 있는 분석은 향후 과제로 남겨둔다.

References

- Chai, G. M. (2007). An analysis of the determinants of employment and wage of new college graduates. *Korean Journal of Social Welfare*, **59**, 35-61.
- Cho, J. S. (2010). A study on equating method based on regression analysis. *Journal of the Korean Data & Information Science Society*, **21**, 513-521.
- Cho, J. S. (2011). Determinants of job finding using student's characteristic information. *Journal of the Korean Data & Information Science Society*, **22**, 849-856.
- Cho, J. S. (2012). Study on the effectiveness of english-medium class. *Journal of the Korean Data & Information Science Society*, **23**, 1137-1144.
- Cho, K. H. and Park, H. C. (2011a). Study on the multi intervening relation in association rules. *Journal of the Korean Data Analysis Society*, **13**, 297-306.
- Cho, K. H. and Park, H. C. (2011b). Discovery of insignificant association rules using external variable. *Journal of the Korean Data Analysis Society*, **13**, 1343-1352.
- Cho, K. H. and Park, H. C. (2012). A study on decision tree creation using marginally conditional variables. *Journal of the Korean Data & Information Science Society*, **23**, 299-307.
- Hwang, Y. J. and Baek, B. B. (2008). Determinants of employment status of university graduates youth. *Journal of Vocational Education and Training*, **11**, 1-23.
- Jun, H. J., Ryu, J. S., Cho, J. S., Park, S. I. and Kim, J. H. (2009). An analysis on the employment determinants of university graduates in NURI Project Teams : The case of the universities in Busan region. *Korean Journal of Labor Economics*, **32**, 31-56.
- Jung, H. J. and Min, D. K. (2013). The study of foreign exchange trading revenue model using decision tree and gradient boosting. *Journal of the Korean Data & Information Science Society*, **24**, 161-170.
- Kim, J. S., Lee, B. H. and Shin, J. Y. (2012). Job seeking method and overeducation of young people. *Quarterly Journal of Labor Policy*, **12**, 51-73.
- Lee, K. J. (2008). An attempt to standardize functional data correspondence analysis. *Journal of the Korean Data Analysis Society*, **10**, 197-210.
- Lee, C. Y. (2008). Behaviors of youth paid-workers in downgrading employment. *Journal of Vocational Education & Training*, **11**, 49-69.
- Lee, K. Y. and Kim, Y. H. (2003). The determinants of youth labor market performance. *Quarterly Journal of Labor Policy*, **3**, 69-93.
- Liu, B., Hsu, W. and Ma, Y. (1999). Mining association rules with multiple minimum supports. *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*, 337-341.

- Park, H. B. (2011). The effects of individual background and school factors on employment outcomes of college and university graduates. *Journal of Vocational Education & Training*, **14**, 1-25.
- Yhang, W. J., Jun, J. K. and Park, M. H. (2007). Image evaluation of Pusan international film festival using corresponding analysis. *Journal of the Korean Data Analysis Society*, **9**, 1381-1392.

Analysis of employee's characteristic using data visualization[†]

Jang Sik Cho¹

¹Department of Informational Statistics, Kyungsoong University

Received 8 May 2014, revised 17 June 2014, accepted 23 June 2014

Abstract

The fundamental concerns of this paper are to analyze the effects of some characteristics on the employment of new college graduated students in viewpoint of data visualization. We use individual and department characteristic data of K-university graduated students in 2010. We apply multiple correspondence analysis, decision tree analysis, association rules and social network analysis for data visualization. The results of the analysis are summarized as follows. First, an analysis of the determinants of employment shows that GPA, department category, age and number of majors, recruiting time affect the employment rate. Second, higher GPA and natural category of department positively affect the employment rate. Finally, low age, single major and early recruiting time also positively affect the employment rate.

Keywords: Association rule, data visualization, decision tree analysis, multiple corresponding analysis.

[†] This research was supported by Kyungsoong University Research Grants in 2014.

¹ Professor, Department of Informational Statistics, Kyungsoong University, Busan 608-736, Korea.
E-mail: jscho@ks.ac.kr